Strategic planning and sensitivity-enhancing tactics for detecting low-mass particle dark matter with phonon-mediated detectors

Thesis by
Osmond Wen

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

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
Defended October 16, 2024

Osmond Wen
ORCID: 0000-0002-6334-6813

Acknowledgments

Funding sources

This work is supported by the following organizations and programs:

- the DOE Office of High Energy Physics Cosmic Frontier and Advanced Detector Research programs;
- the Fermilab Laboratory Directed Research and Development program;
- the NASA Space Technology Graduate Research Opportunity program;
- and the Caltech Division of Physics, Mathematics, and Astronomy.

Personal acknowledgments

Sunil, I am forever grateful for the boundless support, care, and attention that you give to your students and their projects. You truly go the extra mile with your impressive stamina during our in-depth weekly group meetings, your attention to detail toward written documents, including this thesis, and your many, many, many thoughtfully written emails that powered me through graduate school. I did not personally pay for my tuition (see above), yet somehow, working with you, I feel like I got my money's worth and more.

To my committee—Prof. Bradley Filippone, Prof. Jonas Zmuidzinas, and Prof. Kathryn Zurek: thank you for being a part of my committee and for following along during my graduate school journey, through pandemic and all. Brad, it was a pleasure to TA Ph 2c with you back in 2019. Your enthusiasm for physics and teaching is contagious. I very much enjoyed working with and learning from you. Kathryn and Jonas, I thank both of you for being such strong leaders in your respective fields given their relevance to this thesis. Without your leadership, I doubt that the fields of low-mass particle dark matter and kinetic inductance detectors would be as rich and mature as they are today. Thanks to the groundwork you laid (and of course that Sunil and many others laid as well), I was given an exciting graduate school project that has only gained in relevance since I took it on.

To the unofficial mentors of my graduate career—Karthik Ramanathan, Noah Kurinsky, Dylan Temples, and Bruce Bumble: you have made me a stronger and more capable scientist, and I thank you for that. Karthik, we went through a lot together with the B101 Oxford fridge, but I would not have had it any other way. You push me to be creative and resourceful. Many of the results discussed in this thesis would not have been reached without your dedication and good ideas. Noah, you have a knack for the bigger picture and thinking about the next step. I admire both you and the work that you have done and feel grateful to have been able to share my work with you since the very beginning. Dylan, you are the perfect example of someone who leads by example. Your management of and passion for the KIDs effort at NEXUS led the project to grow and prosper beyond what I had thought possible. I feel that we work well together. Bruce, thank you for teaching me everything that I know about device fabrication. I consider myself lucky to have been able to fabricate the devices that we test, and that would not have been possible without your

patience and guidance. Not only did you teach me about the ways of the cleanroom, but you also taught me how to conduct myself in moments of failure and disappointment. I thank you for that.

To the Caltech crew working on KIDs for dark matter, both new and old—Yen-Yung Chang, Taylor Aralis, Brandon Sandoval, Yann Sadou, and Robin Xiong: thank you for joining me through these formative years of graduate school. I have learned a lot from each of you and hope to keep in touch as we go our separate ways. Yen-Yung, thank you for teaching me about device design, simulation, and fabrication. I use your fabrication procedure as my cleanroom Bible. Taylor, thank you for teaching me how to use the fridge, take data, and measure an energy resolution. Each of those things is a hugely significant undertaking, and you taught me all of it. Brandon, thank you for all your hard work in managing the fridge and preparing the detectors. Getting a result out of B240103 would not have been possible without you. Yann, thank you for taking on the task of completing the fridge upgrades. That project has been a long time coming, and I am glad you are taking it to the finish line. Robin, thank you for the enthusiasm you bring to SuperCDMS and KIPM detectors. Between you, Brandon, and Yann, I feel that I am leaving the project in good hands.

To my NASA Space Technology Graduate Research Opportunity officer, Thomas Stevenson: thank you for your reliability, diligence, and wisdom over the four years of my graduate fellowship. You offered sage advice and discerning critiques every time we met. Additionally, thanks to both you and Wonsik Yoon for providing hospitality and an interesting project during my Visiting Technologist Experience at Goddard.

To the Long-Term Planning Task Force—Sunil Golwala, Tarek Saab, Tyler Reynolds, Eleanor Fascione, Harrison Coombes: thank you for all the hard work in producing the SuperCDMS Snowmass 2021 contribution. Lots of man- and core-hours went into producing that document. I feel proud to have been a part of that effort.

To the members of the SuperCDMS collaboration who supported me during my Run 4 analysis project—Ziqing Hong, Miriam Diamond, Matthew Wilson, Valentina Novati, Stefan Zatschler, Ata Sattari, Sudipta Das, Enze Zhang, Imran Alkhatib, Huanbo Sun, Matt Pyle, Enectalí Figueroa-Feliciano, Kyle Kennard, Rohan Shenoy: thank you for your comments, suggestions, and questions. Kyle, thank you for working with me on investigating the KS tests and the Cs-137 data. I am excited to hear more about HVeV at CUTE. Rohan, thank you for helping me complete the project with your diligence in producing the correct HV NRDM signal model as well as setting optimum interval limits. I can truly say that this thesis would not be complete today if we had not crossed paths.

To the students I have had the pleasure of mentoring—Chi Cap, Mira Menezes, Garrison Chan, Monic Moy, Hanna Park, and Emily Xu: thank you for giving me your time and effort. It was rewarding getting to mentor each of you, and I look forward to hearing more about your future endeavors.

To my physics teachers and mentors from high school and college—Mr. David Carroll, Prof. Frank Toffoletto, Prof. Mustafa Amin, and Prof. Douglas Natelson: thank you for instilling in me a love of physics. Through conversations, lectures, and answers to my pesky questions, you shaped me into the physicist I am today.

To my friends from graduate school—Magel Su, Adriano Testa, Léo Borrel, Bailey Gu, Jacob Shen, Alex Buser: thank you for the late-night board games, day hikes through the mountains, and pandemic-era Zoom calls. You kept me sane through these past six years.

To my family—Mom & Dad, and Oriana & Eric: thank you for the love and support throughout all these years. You have made me the person I am today.

To Sarah and our cats, Mel & Manuka: thank you for being just the way you are. You bring joy to my life. I love you so very much.

To the next generations of graduate students:
this thesis is for you.

Abstract

A non-baryonic matter beyond the framework of the Standard Model is required to explain a vast set of astrophysical and cosmological phenomena in our universe; it is referred to as dark matter and comprises 85% of all matter. Previously centered on particle candidates in the 1 GeV to 10 TeV mass range, dark matter model building has expanded to masses well beyond that range, with an emphasis toward low-mass particles below 1 GeV. Low-mass dark matter models have invoked new and creative mechanisms for producing the relic abundance of dark matter and in doing so have provided a variety of new laboratory-testable hypotheses about the early universe.

Direct detection experiments seek to directly measure a dark matter particle interaction from the Milky Way dark matter halo with ultra-sensitive detector technologies in low-background environments. As the paradigm has shifted toward lower-mass particle candidates, detector technologies have followed suit: single-charge-sensitive detectors and low-threshold, purely phonon-mediated detectors are among the best detector architectures for probing the most immediately accessible theoretical models of low-mass dark matter.

The Super Cryogenic Dark Matter Search (SuperCDMS) has used and developed detector technologies on both of these fronts. On the axis of single-charge-sensitive detectors, the High Voltage (HV) detector program of the SuperCDMS Collaboration has demonstrated gram-scale, single-charge sensitive detectors known as HVeV detectors. Recent advances in the collaboration's understanding of single-charge backgrounds have enabled much improved sensitivity to low-mass dark matter parameter space with even these gram-scale detectors. HVeV detectors are a prototype version of the HV kg-scale detectors to be deployed at the flagship SuperCDMS experiment in SNOLAB. HV detectors are projected to test vast regions of unconstrained parameter space for both electron- and nuclear-recoiling dark matter, as shown in Chapter 4 of this thesis among many other SuperCDMS sensitivity projections.

A potentially limiting background for SuperCDMS detectors at SNOLAB is the zero-charge low energy excess, which is characterized by an exponentially rising spectrum of background phonon events below about 100 eV to 1 keV recoil energy. Chapter 5 of this thesis presents a data-driven technique to subtract the zero-charge low energy excess (0QLEE) as observed in HVeV detectors. A search for charge-producing, nuclear-recoiling dark matter is performed with this background-subtraction technique. The resultant exposure-limited constraint on the nucleon-dark-matter cross section is nearly a factor of 10× stronger than the background-limited constraint and is within tens of percent from unconstrained parameter space. The two dominant sources of systematic uncertainties for this search are (1) the uncertainty on the total rate and spectral shape of zero-charge low energy excess events and (2) the completely unknown behavior of the ionization yield function in silicon for nuclear recoils below 100 eV.

On the axis of low-threshold phonon-mediated detectors, the SuperCDMS Collaboration must improve phonon energy thresholds to below 1 eV in order to attain sensitivity to sub-GeV nucleon-coupled dark matter. Presently, SuperCDMS has achieved detector phonon thresholds in

the range from 10 eV to 200 eV depending on detector size. In Chapters 6, 7, and 8, we present a radically different phonon sensor architecture that may provide long-term gains in sensitivity: the kinetic inductance detector.

There are two main quantities that constrain the capacity of kinetic inductance detectors to be effective phonon sensors: the detector readout noise and the phonon collection efficiency. Chapter 7 explores the former, detailing the variety of noise sources in kinetic inductance detectors and how they may impact a sensor's energy resolution using both theoretical calculations and experimental measurements. In general, resolution on energy absorbed in the sensor is presently limited to a range from 1 eV to 5 eV. Chapter 8 then reports on the overall detector energy performance of three different KID-based phonon-mediated (KIPM) detectors, each of which suffers from percent-scale phonon collection efficiencies. An empirical model is then built to parametrize and understand the reasons for the poor phonon collection efficiencies, thereby outlining a path forward to lowering energy thresholds in KIPM detectors.

Published Content and Contributions

Ramanathan, Karthik, Taylor Aralis, et al. (Nov. 2022). "Identifying drivers of energy resolution variation in a multi-KID phonon-mediated detector." In: *Journal of Low Temperature Physics* 209.3, pp. 457–463. DOI: 10.1007/s10909-022-02753-5.

For this conference proceeding, I contributed to the development of the code used for KID readout and to the analysis of the results.

Ramanathan, Karthik, Osmond Wen, et al. (2024). *Significant noise improvement in a kinetic inductance phonon-mediated detector by use of a wideband parametric amplifier.* DOI: 10.48550/arXiv.2402.05419.

For this under-review article, I am the second corresponding author. I fabricated the KID-based phonon-mediated detector that was used for this measurement, performed and analyzed the noise measurements, and wrote a section of the article.

SuperCDMS Collaboration (2023). *A strategy for low-mass dark matter searches with cryogenic detectors in the SuperCDMS SNOLAB facility.* DOI: 10.48550/arXiv.2203.08463.

For this SuperCDMS white paper submitted to the Snowmass 2021 process, I was a member of a team that operated and upgraded a body of code to perform sensitivity forecasts for future SuperCDMS-style experiments.

Temples, Dylan J. et al. (Oct. 2024). "Performance of a phonon-mediated kinetic inductance detector at the NEXUS cryogenic facility." In: *Physical Review Applied* 22 (4), p. 044045. DOI: 10.1103/PhysRevApplied.22.044045.

For this published article, I fabricated the detector and performed noise measurements on a sister device that informed the performance of this detector. I then contributed greatly to the formulation of the data-taking protocol and advised closely during the analysis and interpretation of results.

Wen, Osmond et al. (Nov. 2022). "Performance of a phonon-mediated detector using KIDs optimized for Sub-GeV dark matter." In: *Journal of Low Temperature Physics* 209.3, pp. 510–517. DOI: 10.1007/s10909-022-02764-2.

For this conference proceeding, I fabricated the detector and then measured, analyzed, and modeled its noise performance.

Table of Contents

Acknowledgments . . . 3
Abstract . . . 6
Published Content and Contributions . . . 8
Table of Contents . . . 9
List of Figures . . . 11
List of Tables . . . 14

I Dark matter **15**
Preface . . . 16
Chapter 1: The evidence for dark matter . . . 17
1.1 The matter content of the universe . . . 17
1.2 Galaxy clusters . . . 20
1.3 Galaxy rotation curves . . . 22
Chapter 2: Low-mass dark matter particle candidates . . . 25
2.1 The thermal history of the universe . . . 25
2.2 The matter power spectrum: a gauge on relativistic dark matter . . . 28
2.3 WIMPs and freeze-out: 1 GeV to 10 TeV . . . 29
2.4 The hidden sector: a few keV to 10 GeV . . . 32
2.5 New dark matter histories in the hidden sector . . . 33
2.6 Defining sharp hidden sector targets for direct detection . . . 36
Chapter 3: Detector architectures for low-mass particle dark matter . . . 40
3.1 Typical recoil energies . . . 40
3.2 Charge-producing nuclear recoils for WIMP searches . . . 47
3.3 Single-charge-sensitive detectors . . . 50
3.4 Low-threshold phonon-mediated detectors . . . 51

II SuperCDMS **59**
Preface . . . 60
Chapter 4: SuperCDMS SNOLAB and beyond . . . 61
4.1 The basics of SuperCDMS SNOLAB . . . 61
4.2 SuperCDMS SNOLAB sensitivity projections . . . 65
4.3 NRDM sensitivity projections . . . 70
4.4 ERDM sensitivity projections . . . 73
4.5 Forecasting for next-generation SuperCDMS detectors at SNOLAB . . . 75

4.6 Next-generation: low-threshold 0V detectors . . . 78
4.7 Next-generation: low-leakage HV detectors . . . 81
4.8 Strategies for detecting sub-GeV dark matter with 0V and HV detectors . . . 83
Chapter 5: HVeV Run 4 NRDM search with background subtraction of the low energy excess: a prototype SuperCDMS SNOLAB analysis . . . 85
5.1 Background information, key concepts, and motivation . . . 85
5.2 HVeV Run 4 . . . 89
5.3 Live-time cuts . . . 92
5.4 Data-quality cuts . . . 93
5.5 Comparing the 0V and HV spectra . . . 96
5.6 Background estimation for NFC1 . . . 98
5.7 Constraining the DM-nucleon cross section . . . 99

III Kinetic inductance detectors for dark matter 103
Preface . . . 104
Chapter 6: Superconducting microwave resonators: signal calibration and responsivity . . . 105
6.1 Absorbed energy → quasiparticles: the Cooper pair . . . 105
6.2 Quasiparticles → complex conductivity: Mattis-Bardeen theory . . . 105
6.3 Complex conductivity → surface impedance: a lesson in penetration depth . . . 111
6.4 Surface impedance → resonator parameters: the kinetic inductance fraction . . . 113
6.5 Resonator parameters → δS_{21}: the RF responsivity . . . 114
6.6 Summary of equations and Mattis-Bardeen fits . . . 117
6.7 Non-linear pulse shapes in KIDs . . . 118
Chapter 7: Noise sources in superconducting microwave resonators . . . 121
7.1 Noise on resonator readout . . . 121
7.2 Generation-recombination noise . . . 124
7.3 Two-level system noise . . . 129
7.4 Johnson, amplifier, and readout chain noise . . . 134
7.5 Noise measurements in KIPM detectors . . . 137
7.6 Electronics $1/f$ noise and correlated-noise removal . . . 138
7.7 An amplifier-noise-dominated KIPM detector . . . 143
7.8 TLS-noise-dominated KIPM detectors . . . 145
7.9 KIPM detector operation with a KI-TWPA . . . 153
7.10 Summary of noise sources . . . 155
Chapter 8: Optimizing KIPM detectors for low-threshold particle detection . . . 156
8.1 Previous measurement of energy resolution . . . 156
8.2 Design considerations for a low-threshold detector . . . 158
8.3 OW200127, a single resonator device: results and limitations . . . 160
8.4 B240103, a single resonator device with reduced dead metal . . . 167
8.5 Phonon collection efficiency modeling . . . 173
8.6 Summary and outlook toward sub-eV energy thresholds . . . 176

IV Appendices 178
Appendix A: Computing dynamic structure factors . . . 179
Appendix B: USRP noise timestream procedure . . . 182
Appendix C: KID debugging . . . 184

List of Figures

Number		*Page*
1.1	Apparent magnitude versus redshift of supernovae	18
1.2	CMB power spectrum as it varies with Ω_k and Ω_m	19
1.3	The Bullet Cluster	20
1.4	The Coma Cluster	21
1.5	Galaxy rotation curves for 21 spiral galaxies	22
1.6	Rotation curve of the Milky Way and its sub-components	23
2.1	Matter power spectrum	28
2.2	Freeze-out dynamics	30
2.3	Schematic of the hidden and visible sectors	32
2.4	ELDER dark matter evolution in the early universe	35
2.5	ELDER/SIMP/WIMP scenarios that produce the relic abundance	36
2.6	Theory targets for MeV-scale electron-recoiling dark matter	38
3.1	The kinematics of free nuclear recoils	41
3.2	The kinematics of free electron recoils	43
3.3	Recoil energy for silicon nuclei, germanium nuclei, and electrons versus dark matter mass	46
3.4	Xenon time projection chamber	48
3.5	Charge-coupled devices for dark matter	50
3.6	SuperCDMS phonon sensor design	52
3.7	Single-phonon dispersion relations	54
3.8	Multi-phonon dynamic structure factor	56
3.9	Expected reach of a GaAs phonon-mediated detector to electron-recoiling dark matter with a light mediator.	57
4.1	E-field configurations of SuperCDMS detectors	62
4.2	Phonon sensor configurations of SuperCDMS detectors	63
4.3	The SuperCDMS SNOLAB experiment	64
4.4	Ionization yield models $Y(E_r)$	67
4.5	NRDM sensitivity of SuperCDMS SNOLAB and silicon detector spectra	71
4.6	NRDM sensitivity of SuperCDMS SNOLAB and germanium detector spectra	72
4.7	SuperCDMS SNOLAB ERDM sensitivity and related detector spectra	73
4.8	Exposures and background rates for SuperCDMS SNOLAB HV detectors	74
4.9	`Det A/B/C` scenarios: expected energy thresholds, mass reach, and ionization leakage rates	76
4.10	NRDM sensitivity of next-generation 0V detectors and related detector spectra	79
4.11	ERDM sensitivity of next-generation 0V detectors and related detector spectra	80

4.12 ERDM sensitivity of next-generation HV detectors and related detector spectra . . . 81
4.13 NRDM sensitivity of next-generation HV detectors and related detector spectra . . . 82
5.1 HVeV Run 3 versus HVeV Run 4 . . . 86
5.2 The zero charge low energy excess as observed in various experiments . . . 87
5.3 Ionization yield models for nuclear recoils and HV NRDM signal models . . . 88
5.4 HVeV Run 4 detector payload . . . 90
5.5 Raw event rates in HVeV Run 4 . . . 91
5.6 Example events identified by the anti-coincidence-based live-time cut . . . 93
5.7 Cs-137 source data: χ^2 versus pulse amplitude and event types . . . 94
5.8 Zoom-in of the post-cut χ^2 versus amplitude-reconstructed energy of HV and 0V Cs-137 data. . . . 95
5.9 Post-cut unblinded background spectra of the HVeV Run 4 detectors. . . . 97
5.10 Kolmogorov-Smirnov tests of the 0V and HV data in the ROI . . . 97
5.11 Maximum-likelihood-based background model parameter estimation for HVeV Run 4 99
5.12 HVeV Run 4 NRDM constraints . . . 101
6.1 Basic machinery of Mattis-Bardeen theory . . . 106
6.2 Quasiparticle density and complex conductivity versus temperature. . . . 108
6.3 Quasiparticle density and complex conductivity versus a reciprocal temperature axis. 109
6.4 The Mattis-Bardeen equations for κ_1 and κ_2 as a function of temperature and photon frequency . . . 110
6.5 Schematics of thin film and bulk superconductors . . . 112
6.6 Examples of $S_{21}(f)$ for resonators in the overcoupled, optimally coupled, and undercoupled regimes . . . 116
6.7 Example Mattis-Bardeen fit . . . 118
7.1 The transformation from raw voltage data to idealized S_{21} . . . 122
7.2 Generation-recombination noise as a stochastic Poisson point process . . . 126
7.3 TLS basics . . . 130
7.4 Signal and noise power levels as they evolve along an example readout chain . . . 135
7.5 Correlated noise cleaning in KIDs . . . 140
7.6 Multi-tone cleaning in resonator readout . . . 142
7.7 YY180726.2 device performance . . . 144
7.8 OW200127 Nb resonator noise PSDs . . . 147
7.9 OW221031 TiNx resonator noise PSDs . . . 148
7.10 YY180726.2 Al resonator noise PSDs . . . 149
7.11 OW200127 Al resonator with niobium-capped capacitor noise PSDs . . . 150
7.12 TLS comparison across different materials . . . 152
7.13 Improved KIPM detector noise performance with first-stage amplification from a KI-TWPA . . . 154
8.1 Phonon collection efficiency of DMLE2 resonators . . . 157
8.2 Designs and photographs of OW200127 . . . 158
8.3 Schematic depicting possible avenues for phonon energy loss . . . 159
8.4 A collection of all published plots illustrating the TLS noise in the aluminum resonators of OW200127 . . . 162
8.5 LED pulses: setup and examples . . . 163
8.6 LED pulse histograms and σ^2 versus μ for photon energy calibration . . . 166
8.7 B240103 device design . . . 167
8.8 Optical fiber and coax cable setup from 300 K to 4 K in the Cahill B101 Oxford Kelvinox 25 from July 2021 installation . . . 168

8.9 LED pulses in B240103 . . . 169
8.10 B240103 optimal filter components . . . 170
8.11 B240103 pulse histograms and resolution modeling . . . 172
8.12 Three KIPM detector designs used to build a model for η_{ph} . . . 173
8.13 η_{ph} volume-scaling modeling . . . 175
8.14 Pie charts of phonon losses . . . 176

List of Tables

Number		*Page*
2.1	The thermal history of the universe	26
4.1	Anticipated exposures and detector parameters for SuperCDMS SNOLAB	64
4.2	Challenges and opportunities for next-generation 0V and HV detectors in search of sub-GeV dark matter	84
7.1	Summary of noise sources in KIPM detectors	155
8.1	Comparison of noise results for OW200127	161
8.2	Volume-scaling model inputs	174
8.3	Summary of KIPM detector performance	177

Part I

Dark matter

Preface

In Part I of this thesis, I provide a short review of dark matter evidence, particle candidates, and detector architectures. Essentially all of the figures and equations are reproduced from some other source with the notable exception of those in Section 3.1, which provides a general framework for both Chapter 3 as well as the strategic planning that is discussed in Chapter 4. There is an attempt to be broadly encompassing in all three chapters, but they are by no means comprehensive.

Chapter 1

The evidence for dark matter

The evidence for dark matter is multipronged and spans a wide range of both physical scales and fields of astronomical study. We start at the largest scales for the evidence of dark matter and work toward smaller scales.

1.1 The matter content of the universe

The Einstein field equations govern the relationship between energy and spacetime:

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{1.1}$$

$G_{\mu\nu}$ is the Einstein tensor and contains information about the spacetime curvature of our universe and its evolution. $T_{\mu\nu}$ is the stress-energy tensor and contains values for the densities and fluxes of both energy and momentum in our universe. G is the gravitational constant.

The G_{00} component of the Einstein field equations gives rise to the first Friedmann equation, which relates the expansion of our universe to its energy density:

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

a is the scale factor of the universe relative to the present day ($a = 1$ today). k is the curvature of the universe: $k = 0$ for flat Euclidean space; $k = 1$ for open or spherical space; and $k = -1$ for closed or hyperbolic space. ρ is the energy density of the universe. $H = \dot{a}/a$ is referred to as the Hubble parameter.

The first Friedmann equation can be recast in terms of individual energy density components and their relative abundances at the present day:

$$\frac{H^2}{H_0^2} = \Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda, \tag{1.3}$$

where an individual component $\Omega_i = \rho_i/\rho_{\text{crit}}$, $\rho_{\text{crit}} = 3H_0^2/8\pi G$, and H_0 is the Hubble parameter in today's universe, also known as the Hubble constant. The subscripts r, m, k, and Λ respectively denote radiation, matter, curvature, and dark energy. Dark energy, also known as the cosmological constant, is believed to drive the accelerated expansion of the universe.

For a heuristic understanding of the matter and radiation power laws that arise in Equation 1.3, consider the following: ρ_m exhibits $\propto a^{-3}$ because the energy per unit volume shrinks for every dimension in volume; ρ_r exhibits $\propto a^{-4}$ because there is an additional redshifting of the

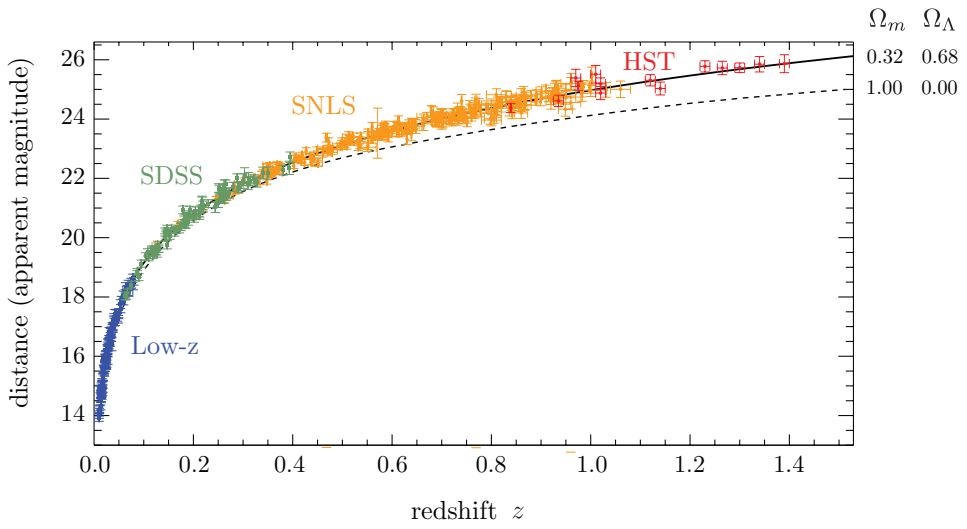

Figure 1.1: **Apparent magnitude versus redshift of supernovae.** Brightness decreases with increasing apparent magnitude. This data allows for a fit of Ω_m and Ω_Λ, assuming a flat $k = 0$ universe. Reproduced from Baumann (2022).

energy, $E \propto a^{-1}$. Matter-radiation equality occurred about 60,000 years after the Big Bang, so the radiation energy density component has been subdominant since then and is about one part in 10,000 today.

An essential fact of modern cosmology is that the data prefer a flat universe, predominantly composed of dark energy and matter. Measurements of Type 1a supernovae and the cosmic microwave background provide compelling motivation for this statement, among many other measurements.

Type 1a supernovae

With the assumption of a flat universe ($\Omega_k = 0$), the brightnesses of Type 1a supernovae may be used to infer Ω_m and Ω_Λ. Type 1a supernovae are a standard candle of cosmology; their luminosity is observed to be predicted by their age, so their relative brightness can be combined with their intrinsic luminosity to infer their distance. When this distance estimate is plotted against the redshift of each supernova, the observed relation implies that the universe is undergoing an accelerated expansion. Figure 1.1 shows the universe that the best fits these data are $\Omega_m = 0.32$ and $\Omega_\Lambda = 0.68$.

Baryon acoustic oscillations in the cosmic microwave background

The CMB is a background 2.726 K blackbody spectrum seen across the sky. Temperature fluctuations at one part in 10,000 have been measured in the CMB, and their power spectrum as a function of angular scale is plotted in Figure 1.2.

The angular position and amplitude of the first peak in the CMB power spectrum are controlled by the dynamics of baryon acoustic oscillations (BAO) in the early universe. BAO occurred before photons decoupled from the primordial plasma[1] and after overdensities began to form from local matter density perturbations. There are two relevant forces that govern baryon acoustic oscillations: the inward force of gravity, which depends on the total mass of the overdensity, and the outward radiation pressure, which only affects the baryonic matter since it is coupled to

[1] See Chapter 2 for a more detailed description of the decoupling process.

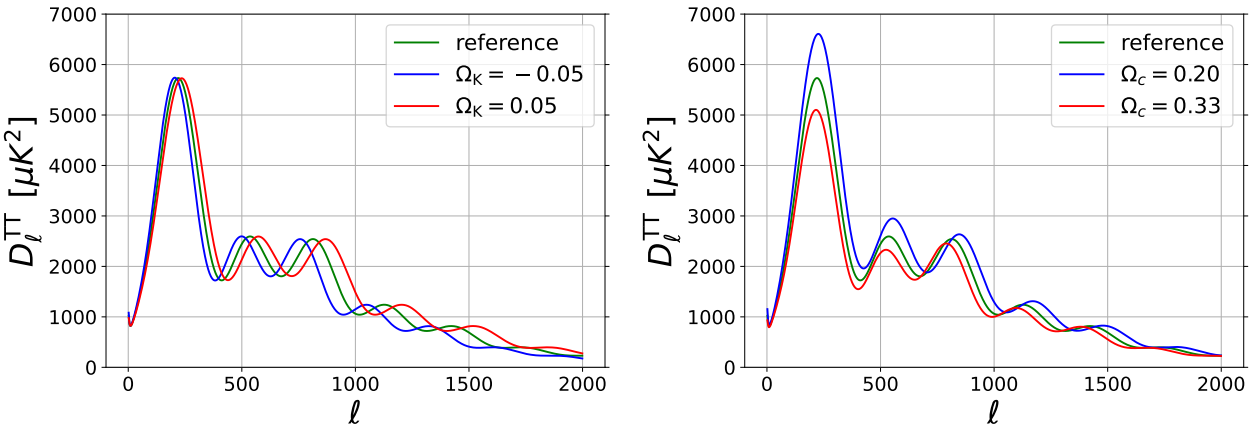

Figure 1.2: **CMB power spectrum as it varies with Ω_k and Ω_m.** Left: how varying Ω_k from the Planck measured "reference" value impacts the power spectrum. Right: how varying Ω_c from the Planck measured "reference" value impacts the power spectrum. Figure produced by Robin Wen.

photons. When photons decoupled from baryonic matter during recombination and began their journey toward becoming the cosmic microwave background (CMB), the oscillations stopped, and their pattern was imprinted upon the angular power spectrum of the CMB. The physical size of the oscillations at the time of recombination serves as a standard ruler for cosmologists.

The location of the first peak plays an important role in constraining the universe to be consistent with a Euclidean, flat geometry. The relative energy density abundances of matter, curvature, and dark energy were measured by the Planck Collaboration using the cosmic microwave background and were reported in Planck Collaboration (2020):

$$\Omega_m = 0.321 \pm 0.013, \quad \Omega_k = 0.0007 \pm 0.0019, \quad \Omega_\Lambda = 0.679 \pm 0.013.$$

The impact of changing Ω_k from the above reference quantity is shown in Figure 1.2.

Ω_m can be broken down into two sub-components that dominate the total relic abundance of matter in our universe:

$$\Omega_m = \Omega_b + \Omega_c, \tag{1.4}$$

i.e., a baryonic component Ω_b and a non-baryonic component Ω_c. Baryonic matter in the context of cosmology refers mainly to protons and neutrons, the three-quark particles that compose "normal" matter. Non-baryonic matter is a generic term to indicate a component of Ω_m that does not interact with photons. This component of the energy density in the early universe is thus not subject to radiation pressure during the era of BAO. Consequently, the BAO amplitude depends on the ratio of Ω_c to Ω_b: for larger Ω_c, the BAO amplitude is smaller. Figure 1.2 Right shows this effect. There are two curves with different Ω_c quantities and Ω_b held fixed.[2] The presence of non-baryonic matter in the early universe serves to smooth out the fluctuations we see in the CMB. Drawn in green is the CMB power spectrum corresponding to the Ω_c value measured by Planck Collaboration (2020):

$$\Omega_b h^2 = 0.02212 \pm 0.00022, \quad \Omega_c h^2 = 0.1206 \pm 0.0021,$$

i.e., dark matter composes 85% of the matter content in our universe and 27% of the total energy content.

[2]The dark energy abundance Ω_Λ is modified so as to maintain $\Omega_r + \Omega_m + \Omega_k + \Omega_\Lambda = 1$ for the sake of the figure.

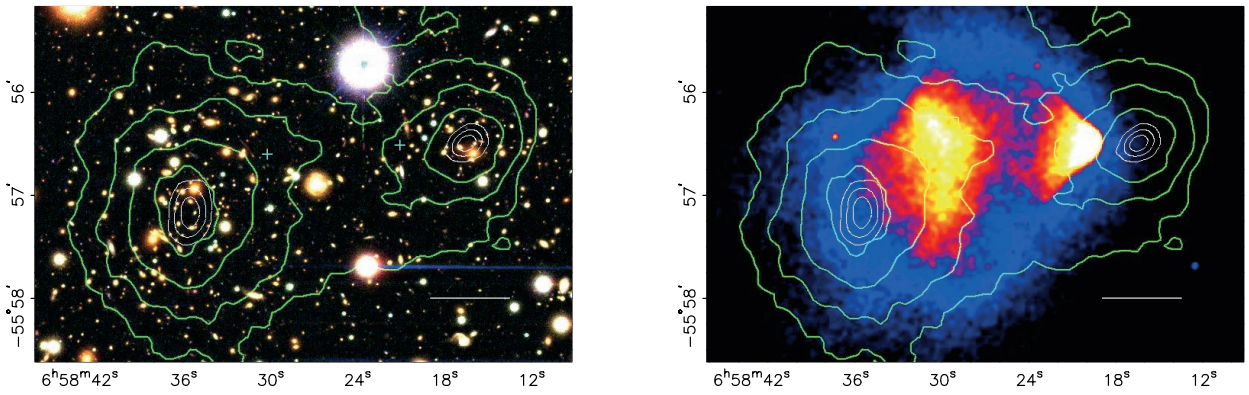

Figure 1.3: **The Bullet Cluster.** Both: the white bars indicate 200 kpc, and the green contours are weak-lensing κ reconstructions. Left: optical image from Magellan Telescopes of the Bullet Cluster, X-ray plasma peaks marked with blue crosses. Right: Chandra Observatory image of the X-ray-emitting plasma. Figures and caption reproduced from Clowe et al. (2006).

1.2 Galaxy clusters

Galaxy clusters are made of hundreds to thousands of galaxies, are usually 10^{14} to 10^{15} solar masses, and are about 1 Mpc to 10 Mpc in size. For reference, the observable universe is about 29 Gpc in diameter. Galaxy clusters are the largest gravitationally bound structures in the universe. The Bullet Cluster and Coma Cluster are two famous examples of galaxy clusters that exhibit evidence for dark matter.

The Bullet Cluster

The Bullet Cluster, or cluster 1E 0657-558, is a striking example of the need for dark matter. Two smaller galaxy clusters constitute the Bullet Cluster and passed through each other approximately 100 million years ago. By way of weak-lensing and X-ray measurements reported in Clowe et al. (2006), astronomers were able to separately image the matter and plasma distributions of both the larger and smaller sub-clusters of the Bullet Cluster.

In the two images of Figure 1.3, the larger cluster is on the left and the smaller cluster is on the right. The discrepancy between the plasma distributions and the matter distributions for both clusters is shown in the right image. The X-ray distributions have lagged behind the matter distributions in the collision between clusters because of baryonic effects. Without dark matter, the matter distribution as inferred from the weak-lensing measurements should coincide with the X-ray-emitting plasma. Instead, the matter distributions of both clusters are offset from the X-ray peaks by approximately 8σ. This offset favors collisionless models of dark matter.

The original Bullet Cluster paper (Clowe et al., 2006) reports average magnification κ values around the brightest cluster galaxy and around the plasma peak for each galaxy cluster. κ is measured from the elongation of background stars and is directly proportional to the surface mass density of the gravitational lens. In the larger galaxy cluster, $\kappa = 0.36 \pm 0.06$ near the center of the matter distribution (with the smaller cluster's κ subtracted out), and $\kappa = 0.05 \pm 0.06$ at the peak of the X-ray emission (with both clusters' κ subtracted out). Thus, there is a 7:1 ratio in the surface matter density between these two locations. The paper furthermore reports that the total mass of plasma and stars is roughly comparable near the galaxy cluster matter distribution center and at the X-ray emission peak. Therefore, the dark matter to baryonic matter relative abundance is 7:1, which is not far from the 6:1 ratio inferred from the CMB.

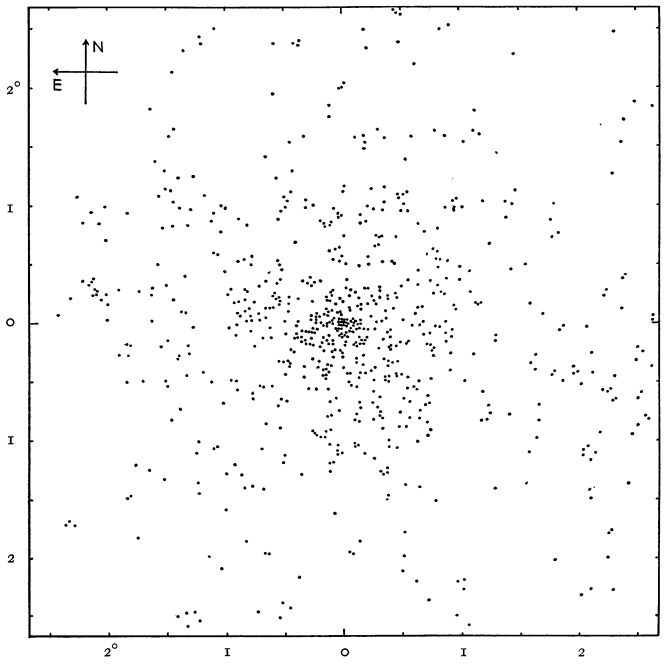

Figure 1.4: **The Coma Cluster.** Each dot represents a galaxy. Galaxies are referred to as nebulae in the original publication. Degrees are demarcated along the axes. Reproduced from Zwicky (1937).

The Coma Cluster

In the 1930s, Swiss astronomer Fritz Zwicky published two articles (Zwicky, 1933) (Zwicky, 1937) on the existence of a "*dunkle Materie*," or dark matter, to explain the higher-than-expected mass of the Coma Cluster. The articles used data from an 18-inch Schmidt telescope at the Palomar Observatory on Mount Palomar in San Diego County. The locations of the various galaxies that make up the Coma Cluster are shown in Figure 1.4. The size of the Coma Cluster is approximately 2×10^6 light years (600 kpc) in radius, according to Zwicky.

Zwicky (1937) performs a virial theorem calculation to estimate the mass of the Coma Cluster as a stable many-body gravitational system:

$$-2\langle T\rangle = \langle U\rangle \tag{1.5}$$

$$-\left\langle \sum m_i v_i^2 \right\rangle > -\frac{5GM^2}{R}$$

$$M\langle\langle v^2\rangle\rangle < \frac{5GM^2}{R}$$

$$M > \frac{R\langle\langle v^2\rangle\rangle}{5G}. \tag{1.6}$$

$\langle T\rangle$ and $\langle U\rangle$ are the total kinetic and potential energies averaged over time. m_i and v_i are the masses and velocities of individual galaxies. M is the total mass of the cluster, and R is the radius of the cluster. $\langle\langle v^2\rangle\rangle$ is the galaxy velocity, averaged over time and mass. The inequality in the second line is due to $-5GM^2/R$ being an underestimate of the potential energy of the system (overestimate of the absolute value).[3] As Figure 1.4 shows, most of the galaxies are clumped toward the center of the cluster.

[3]Recall that the gravitational potential energy of a uniform sphere is $-3GM^2/5R$. In the case of the Coma cluster, the galaxies are clumped toward the center, so the total potential energy is certainly less than $-3GM^2/5R$ (larger in

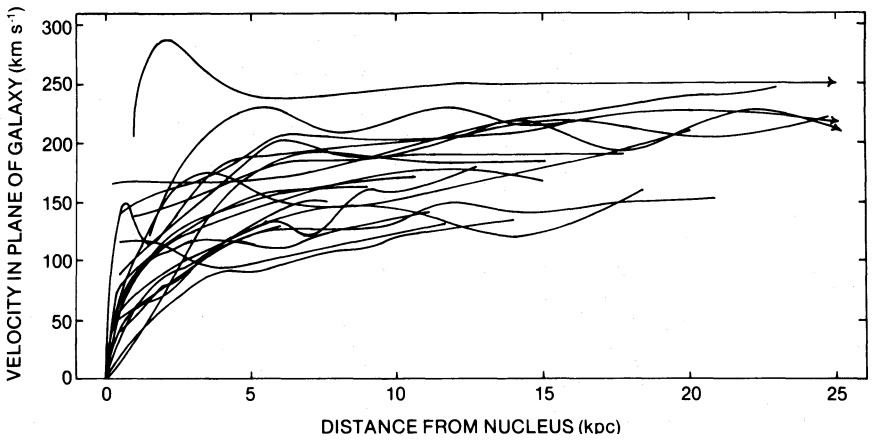

Figure 1.5: **Galaxy rotation curves for 21 spiral galaxies.** Reproduced from Rubin et al. (1980).

Zwicky evaluates the inequality in Equation 1.6 with the following quantities:

$$M > 4.3 \times 10^{13}\ M_\odot \left(\frac{R}{2 \times 10^6 \text{ light years}}\right)\left(\frac{\langle\langle v^2\rangle\rangle}{1.5 \times 10^{16}\ \text{cm}^2\text{s}^{-2}}\right). \tag{1.7}$$

Since there are about 1000 galaxies in the Coma cluster according to Zwicky, he estimates that the average galaxy weighs no less than $4.3 \times 10^{10}\ M_\odot$. This number was about two orders of magnitude in excess of his expectation[4] and provided the first clue for unseen matter in the universe.

1.3 Galaxy rotation curves

The need for dark matter within galaxies was first posited in Rubin et al. (1980). The argument is derived from introductory physics principles concerning a point-like body of mass m that undergoes centripetal acceleration a_c due to the force of gravity F_g:

$$\begin{aligned} F_g &= ma_c \\ \frac{GM_{<r}m}{r^2} &= m\frac{v^2}{r} \\ v &= \sqrt{\frac{GM_{<r}}{r}}. \end{aligned} \tag{1.8}$$

$M_{<r}$ is the total mass of the galaxy within some radius r from the center of a galaxy. $F_g = GM_{<r}m/r^2$ holds for spherical distributions of M and is a consequence of the r^{-2} dependence of F_g. v is the velocity and depends on the matter distribution in the galaxy as given by Equation 1.8. In the case where most of the mass is located close to the center of the galaxy, the velocity should fall as $r^{-0.5}$ at large distances.

A galaxy rotation curve is a plot of the orbital velocity versus the radial distance from the center of the galaxy. Orbital velocity can be obtained from line-of-sight velocities, which are

absolute value). However, for the potential energy to be less than or equal to $-5GM^2/R$, the majority of the mass would need to be concentrated within a radius $3R/25$. The Coma Cluster is not so centrally peaked, as is roughly visible in Figure 1.4.

[4]Zwicky expected each galaxy to have a luminosity of about $8.5\times10^7 L_\odot$, which means an M/L ratio of 500. In the local Kapteyn stellar system, the M/L ratio is 3.

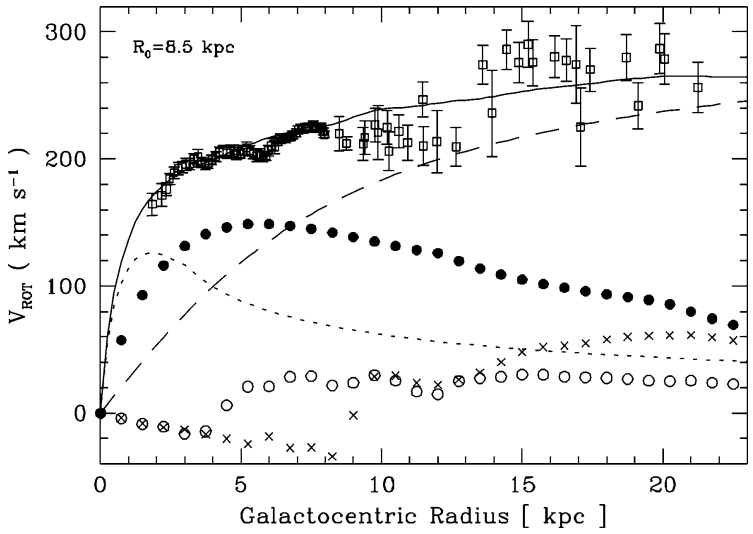

Figure 1.6: **Rotation curve of the Milky Way and its sub-components.** Stellar velocity versus galactocentric radius is plotted with the squares. The differently sized error bars correspond to $r < R_0$ and $r > R_0$ radius estimation, where $R_0 = 8.5\,\mathrm{kpc}$ is the assumed galactocentric radius of Earth. The total best-fit rotation curve is the solid line. The different contributions to the rotation curve from various elements of the Milky Way are the best-fit dark matter halo (long dashed), the stellar disk (filled circles), the bulge (short dashed line), molecular hydrogen gas (empty circles), and atomic hydrogen gas (crosses). Negative values indicate an outward pointing force. Reproduced from Sikivie (2003); caption and figure originally from Olling et al. (2000).

measured via redshift. Figure 1.5 shows galaxy rotation curves from 21 different spiral galaxies. None of the rotation curves exhibits $r^{-0.5}$ behavior in the plotted region. Instead, they all seem to maintain an elevated, flat velocity at large radii.

The elevated and flat rotation curves at large radii can be explained by a dark matter density profile that exhibits the appropriate scaling relation with r. In Figure 1.6, a best-fit rotation curve for the Milky Way is shown as the solid line. The contributions to the rotation curve from the stellar disk, the galactic bulge, molecular hydrogen gas, and atomic hydrogen gas are displayed, and none of these contributions are large enough to explain the elevated rotational velocity of stars $\gtrsim 10\,\mathrm{kpc}$ from the galactic center. In Olling et al. (2000), the following dark matter halo profile $\rho(r,z)$ as a function of galactocentric radius r and elevation above the galactic plane z was used to model and fit the Milky Way's rotation curve:

$$\rho(r,z) = \rho_h \frac{R_h^2}{R_h^2 + r^2 + (z/q)^2}, \tag{1.9}$$

where ρ_h is the central density, R_h is the halo core radius, and q is a halo flattening parameter.

$M_{<r}$ in Equation 1.8 can be expressed as an integral of $\rho(r)$: $M_{<r} = \int_0^r 4\pi r'^2 \rho(r') dr'$. In the regime where $r \ll R_h$ and the halo flattening term can be ignored, the integral is $\propto r^3$ and the rotation curve from Equation 1.8 goes as r. For $r \gg R_h$, the integral is $\propto r$, and the rotation curve is therefore flat.

Although Equation 1.9 is an effective minimal model needed to fit the data in Figure 1.6, the total mass of the galaxy is not easily computed because the integral equation for $M_{<r}$ diverges as r. A popular halo density profile that ameliorates this divergence and fits simulations well is the

Navarro-Frenk-White (NFW) profile from Navarro et al. (1997). It involves higher powers of r in the denominator:

$$\rho(r) = \frac{\rho_0}{\frac{r}{R_s}\left(1+\frac{r}{R_s}\right)^2}, \tag{1.10}$$

where ρ_0 and R_s provide the scale for the galaxy's density and radius. For $r \gg R_s$ the integral for $M_{<r}$ is $\propto \ln(r)$, and the rotation curve falls slowly. The mass integral of the NFW profile is only logarithmically divergent in r.

The essential picture is that the visible matter of the flat and spiral-shaped Milky Way is embedded in a spherical halo of dark matter. Visible matter clumps into compact objects like stars and planets whereas dark matter is diffusely distributed. The galactic dark matter halo is the foundation for direct detection experiments: we seek rare interactions of the dark matter flux through earth using low-background settings and ultra-sensitive detectors.

Chapter 2

Low-mass dark matter particle candidates

In Section 2.1 and Section 2.2 of this chapter, we explore how dark matter fits within the larger thermal history of the universe. In Section 2.3, we touch on dark matter histories that invoke weak-scale interactions and masses. Then, in Section 2.4 and Section 2.5, we motivate and explore the hidden sector, wherein many newly developed dark matter histories involve hidden sector extensions to the Standard Model at mass ranges below the weak scale. These extensions are known as low-mass dark matter particle candidates. Finally, in Section 2.6, we study how these new dark matter models can provide sharp theory targets for direct detection experiments.

2.1 The thermal history of the universe

This section follows Chapter 3 of Baumann (2022).

The ΛCDM model contends that the universe began as a primordial plasma where everything was in thermal equilibrium at a temperature T. The equilibrium of the primordial plasma is maintained through various particle processes such as scatterings, annihilations, and creations. As the universe expands, which is a consequence of the first Friedmann equation, the temperature decreases with size: $T \propto a^{-1}$. The increasing universe size and decreasing universe temperature are the main drivers in determining when major events in the early universe occur. The major events that evolve the primordial plasma into the universe that we know today are outlined in Table 2.1.

Each of the events in Table 2.1 can be related to one of the following three phenomena:

1. The temperature falls below the mass of a certain particle. There is not enough thermal energy to produce the particle from creation interactions. If the particle is not stable, its abundance decays away. Certain event depend on the relative abundance of certain particles.
   - On a related note, how relativistic a particle of mass m is in the early universe is determined by m/T (setting $c = \hbar = k = 1$ here and throughout this chapter): if $m/T \ll 1$, the particle's total energy equals its kinetic energy, which scales with T, and the particle is said to be relativistic; if $m/T \gg 1$, then the particle's total energy is given by its rest mass, and the particle is non-relativistic.
2. An interaction rate Γ for a certain process becomes small compared to the expansion rate H. The interaction rate for a particular process goes as $\Gamma = n\sigma v$, where n is the number density of particles,[1] σ is the interaction cross section, and v is the average velocity of the particles.

[1] For a process of the form $1 + 2 \leftrightarrow 3 + 4$, we would write the interaction rate of species 1 as $\Gamma_1 = n_2 \sigma v$, where n_2

Event	time t	redshift z	temperature T
EW phase transition	20 ps	10^{15}	100 GeV
QCD phase transition	20 μs	10^{12}	150 MeV
Dark matter freeze-out	?	?	?
Neutrino decoupling	1 s	6×10^{9}	1 MeV
Electron-positron annihilation	6 s	2×10^{9}	500 keV
Big Bang nucleosynthesis	3 min	4×10^{8}	100 keV
Matter-radiation equality	60 kyr	3400	0.75 eV
Recombination, CMB forms	380 kyr	1100	0.26 eV
Present	13.8 Gyr	0	0.24 meV

Table 2.1: **The thermal history of the universe.** Modified from Baumann (2022).

When $\Gamma \ll H$, that interaction essentially stops happening, and any physical condition that relied on its occurrence (e.g., maintenance of thermal coupling) no longer holds.

3. A relevant energy scale, such as a binding energy, becomes large compared to the temperature. Particles fall into bound states instead of remaining free within the primordial plasama.

We now go event-by-event in Table 2.1 and connect each event to one of the three reasons stated above.

- The electroweak phase transition occurs around $T = 100\,\text{GeV}$. At this point, the W and Z bosons, which have masses from 80 GeV to 90 GeV, are no longer abundantly produced. Electroweak particle interactions weaken with temperature and enter the regime of an effective field theory wherein the strength of particle interactions is given by the unit-ful Fermi constant G_F.
  - Technically, the phase transition occurs because the Higgs field is no longer thermally prevented from rolling to the bottom of its sombrero potential. When it does, the Higgs field acquires its non-zero vacuum expectation value, which gives mass to all particles.
- The QCD phase transition occurs around $T = 150\,\text{MeV}$. Above this temperature, quarks are asymptotically free. Below this temperature, the strong interaction mediated by gluons binds quarks into baryons and mesons. On a related note, pions become the effective field theory mediator to describe interactions between baryons and have mass around this energy scale, 135 MeV to 140 MeV.
- Neutrino decoupling occurs when the weak-scale interaction rate becomes much less than the expansion rate, $\Gamma_{\text{weak}} \ll H$. Via a heuristic dimensional analysis argument,[2] we can assume $n \sim T^3$, $\sigma \sim G_F^2 T^2$, and $H \sim \sqrt{\rho}/M_{\text{pl}} \sim T^2/M_{\text{pl}}$, where M_{pl} is the Planck mass and is equal to

is the density of the target species 2 and v is the average relative velocity of 1 and 2. The interaction rate of species 2 would be $\Gamma_2 = n_1 \sigma v$. We have used the expectation that at high energies $n_1 = n_2 \equiv n$. Footnote taken directly from Baumann (2022).

[2]The heuristic argument is that energy units can be used to infer the correct expression for the various quantities as they depend on T. When $c = \hbar = k = 1$, the units of a quantity can be expressed in terms of the equivalent energy unit exponent. For example, the number density n has units of inverse volume, which in energy units is eV^3. Thus, $n \sim T^3$ (this expression could also be reached from the $1/a^3$ dilution of the universe). Next, for a weak-scale interaction below 100 GeV, σ is proportional to G_F^2, i.e., a squared matrix element. G_F has energy units of eV^{-2}. To recover the eV^{-2} energy units of cross section, $\sigma \sim G_F^2 T^2$ is required. Finally, $\rho \sim T^4$ is the expression for n with an additional factor of T given that ρ is an energy density. The T^4 dependence can also be recovered from radiation redshift of the primordial plasma, as mentioned in Section 1.1.

$\sqrt{1/G}$. We combine these expressions to find when $\Gamma_{\text{weak}} \ll H$:

$$\begin{aligned} n\sigma v &\ll \frac{T^2}{M_{\text{pl}}} \\ T^3 \times G_F^2 T^2 &\ll \frac{T^2}{M_{\text{pl}}} \\ T &\ll \left(\frac{1}{G_F M_{\text{pl}}}\right)^{1/3} \\ T &\ll 1\,\text{MeV}. \end{aligned} \tag{2.1}$$

The temperature of neutrino decoupling as shown in Table 2.1 is derived from this calculation.

- Below the mass of the electron, around $T = 500\,\text{keV}$, electron-positron annihilations dominate over pair creations; photons do not have enough energy to create electrons. After this point and before photons decouple from the plasma, the energy of electrons and positrons is transferred to photons, which is why the cosmic microwave background temperature 2.726 K is warmer than the cosmic neutrino background temperature of roughly 1.95 K.
- The formation of nuclei during Big Bang nucleosynthesis occurs around $T = 100\,\text{keV}$ and is sparked by the production of deuterium nuclei D, which are composed of one neutron and one proton:

  $$n + p^+ \leftrightarrow \text{D} + \gamma. \tag{2.2}$$

  The deuterium nucleus has a binding energy of 2.2 MeV, so, at that temperature, the formation of deuterium nuclei becomes energetically favorable over free protons and neutrons in the primordial plasma. However, the relative paucity of baryons with respect to photons in the early universe inhibits the forward process in Equation 2.2 until the temperature is well below the 2.2 MeV binding energy. The observed baryon-to-photon ratio is about 10^{-9}, so it is required that $e^{2.2\,\text{MeV}/T}$ greatly exceed 10^9 in order for deuterium to be abundantly produced.[3]
- Recombination of electrons and protons to form hydrogen is the forward process of the following reaction:

  $$e^- + p^+ \leftrightarrow \text{H} + \gamma. \tag{2.3}$$

  The binding energy of the hydrogen atom is 13.6 eV. Like with the formation of deuterium, in order for the forward reaction to dominate over the backward reaction, $e^{13.6\,\text{eV}/T} \gg 10^9$ is required, occurring around $T = 0.26\,\text{eV}$; electrons and protons (of which there are believed to be an equal amount in our universe) are much less abundant than photons. For both the formation of deuterium and hydrogen, the temperature of the universe must be some factor of 20 to 40 below the corresponding bonding energy so that the high-energy tail of photons does not cause the corresponding backward reaction to dominate.

Matter-radiation equality is also included in Table 2.1 for its relevance in the next section concerning relativistic dark matter. We then provide an extensive discussion of various possible dark matter evolutions in the early universe, focusing first on freeze-out in Section 2.3 and then on other possibilities within the hidden sector framework in Section 2.5.

[3] The exponential expression is derived from a Boltzmann equation.

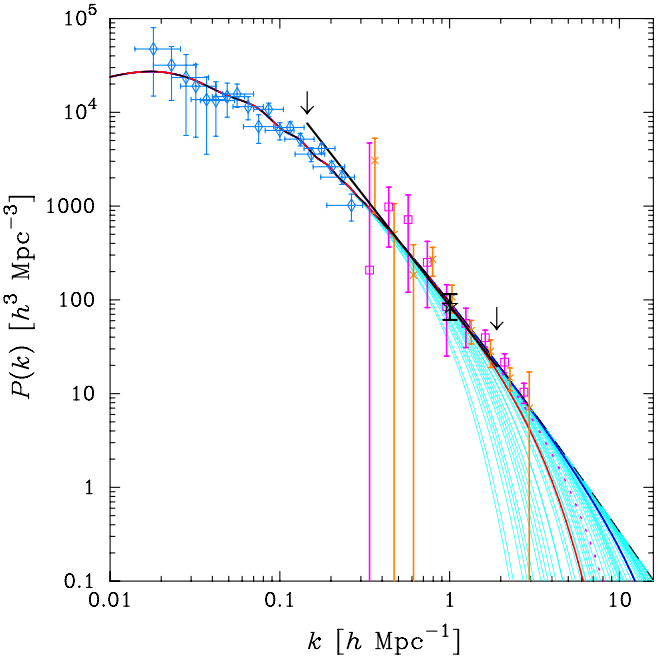

Figure 2.1: **Matter power spectrum.** Shown are the resulting linear matter power spectra $P(k)$ for a standard flat cosmological model $\Omega_c = 0.26$, $\sigma_8 = 0.9$, $\Omega_b = 0.04$, and $h = 0.7$ at $z = 0$, and with sterile neutrino warm to cold dark matter in the mass range $0.3\,\mathrm{keV} < m_s < 140\,\mathrm{keV}$ (cyan). The corresponding cold dark matter case is dashed (black). Small-scale clustering data used here are the SDSS 3D power-spectrum of galaxies (diamonds), the inferred slope and amplitude of the matter power spectrum from SDSS Lyman-α forest observations (star point and slope between arrows), the inferred matter power spectrum from Lyman-α forest observations from Croft et al. (2002) (cross points) and the LUQAS (square points), as interpreted by Viel et al. (2004). Lyman-α forest measures are evolved to $z = 0$ by the appropriate growth function. The solid (blue) line at high-k is $P(k)$ for upper limit $m_s = 8.2\,\mathrm{keV}$ from observations of Virgo (Abazajian et al., 2001), the solid (red) line at low-k is $P(k)$ for the lower limit from the SDSS Lyman-α forest ($m_s = 1.7\,\mathrm{keV}$), and the dotted line is $P(k)$ for the lower limit using high-resolution Lyman-α forest data ($m_s = 3.0\,\mathrm{keV}$). Figure and caption reproduced from Abazajian (2006).

2.2 The matter power spectrum: a gauge on relativistic dark matter

The matter power spectrum $P(k)$ as a function of inverse length scale, represented as the mode wavevector k, constrains how long a thermally produced dark matter particle can be relativistic in the early universe and consequently also constrains the dark matter mass. Scale-invariant inflationary models predict that the $P(k) \propto k$ for matter fluctuations. In the radiation-dominated era (up to about 60,000 years after the Big Bang; $z \approx 3400$), matter fluctuation modes that are smaller than the Hubble radius (within the horizon; sub-horizon) are suppressed by a factor k^{-2} for a duration of time $\propto k^{-2}$ while the mode is sub-horizon, i.e., disturbable by radiation-driven expansion. This suppression of matter fluctuations due to radiation-driven expansion is known as the Mészáros effect. The net result is a k^{-3} power spectrum dependence for all modes that are sub-horizon before matter-radiation equality. Once matter domination begins, the expansion rate is reduced, and modes on all scales grow as a^1. Thus, the matter power spectrum is frozen in shape at matter-radiation

equality and simply grows in amplitude from then on.

The constraint on how relativistic dark matter was in the early universe is provided by the high-k behavior of the matter power spectrum. $P(k) \propto k^{-3}$ over sub-horizon scales holds for non-relativistic matter. Relativistic dark matter can free stream out of overdensities due to its high kinetic energy, suppressing the amplitudes of these overdensities. The more relativistic dark matter is, i.e., the lower mass it is, the longer its free-streaming length. Modes smaller than the free-streaming length undergo greater suppression than the k^{-3} scaling noted above. Such dark matter that is relativistic or near-relativistic at the time of matter-radiation equality is called "hot" or "warm" while non-relativistic dark matter is called "cold."

Various $P(k)$ measurements are shown in Figure 2.1. They show agreement with $P(k) \propto k^{-3}$. Figure 2.1 shows the expected effect of dark matter with mass $0.3\,\text{keV} < m_s < 140\,\text{keV}$, in this case for a dark matter candidate modeled as a sterile neutrino. The data in Figure 2.1 were used to set lower limits $m_s > 1.7$ keV or $m_s > 3.0$ keV, depending on the selection of the data that was used. The resulting matter power spectra for warm dark matter at these lower limits are plotted in red, both solid and dashed, in the Figure 2.1. The matter power spectrum from cold dark matter, which is dark matter that would be non-relativistic throughout all relevant eras during radiation domination, is also shown in Figure 2.1 and exhibits no deviation from the k^{-3} power law at small scales.

The key takeaway from this discussion is that dark matter as a thermal relic must be greater than one to a few keV in order to be consistent with measurements of the matter power spectrum at small scales.

2.3 WIMPs and freeze-out: 1 GeV to 10 TeV

If dark matter is a thermal relic of the primordial plasma, it would undergo a process known as "freeze-out" to produce the relic abundance that we observe today. This process is shown in Figure 2.2. When a particle becomes non-relativistic, i.e., $m \gg T$, the co-moving number density is exponentially suppressed by a Boltzmann factor $e^{-m/T}$ relative to the co-moving number density of relativistic particles. If the annihilation process for a particular particle decouples ($\Gamma_{\text{ann}} \ll H$), then the co-moving number density of that particle cannot decrease, and the particle falls out of equilibrium with the primordial bath. The resultant fixed abundance of that particle is referred to as the relic density.

The relic density for any particle that undergoes freeze-out can be calculated using a Boltzmann equation with the cross section of the annihilation interaction and the mass of the particle. The earlier a particle decouples, the larger its relic density is expected to be.

A good dark matter candidate χ has a relic density that matches the observed relic density Ω_c from the CMB. The thermally averaged cross section $\langle\sigma v\rangle$ at the time of freeze-out that produces the correct relic abundance may be estimated in the following way:

$$\begin{aligned}
\Gamma(T_f) &= H(T_f) \\
n_\chi(T_f)\langle\sigma v\rangle &= \sqrt{\frac{8\pi G}{3}\rho_r(T_f)} \quad \text{radiation-dominated era} \\
\langle\sigma v\rangle &= \sqrt{\frac{8\pi G}{3}\frac{\pi^2}{30}g_*(T_f)T_f^4}\frac{m_\chi}{\rho_\chi(T_f)} \quad \text{radiation} \propto T^4 \\
\langle\sigma v\rangle &= \sqrt{8\pi G}\sqrt{\frac{g_*(T_f)}{10}}\frac{\pi}{3}T_f^2\frac{m_\chi}{\rho_{0,\chi}}\frac{T_0^3}{T_f^3} \quad \text{matter} \propto T^3
\end{aligned}$$

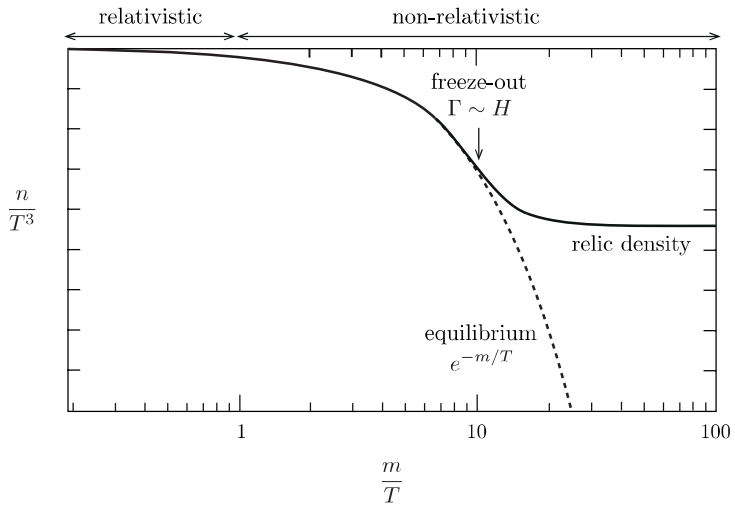

Figure 2.2: **Freeze-out dynamics.** n is the number density, m is the mass, and T is the temperature. In the case of radiation domination, $T \propto a^{-1}$, so n/T^3 can be interpreted as the co-moving number density. Time increases to the right as the temperature falls. Reproduced from Baumann (2022).

$$\langle \sigma v \rangle = \sqrt{8\pi G} \sqrt{\frac{g_*(T_f)}{10}} \frac{\pi}{3} \frac{T_0^3}{\Omega_c \rho_{\text{crit}}} \frac{m_\chi}{T_f} \tag{2.4}$$

$$\langle \sigma v \rangle = 3 \times 10^{-9}\ \text{GeV}^{-2} = 3 \times 10^{-26}\ \text{cm}^3/\text{s}. \tag{2.5}$$

ρ_r and ρ_χ are energy densities for radiation and dark matter. 0 and f subscripts indicate quantities at the present and at freeze-out, respectively. $g_*(T_f)$ is the effective number of relativistic degrees of freedom at the freeze-out temperature, or in other words, the number of particles that are relativistic at a given temperature. m_χ/T_f is the position on the x-axis in Figure 2.2 at which freeze-out begins. It must be calculated from a Boltzmann equation; 10 is a good order-of-magnitude estimate.

The quantity calculated in Equation 2.5, $\langle \sigma v \rangle = 3 \times 10^{-9}\ \text{GeV}^{-2}$, is extremely important. If dark matter is a thermal freeze-out relic of the primordial plasma, as is the case for all known matter, then it must have this annihilation cross section to produce the correct relic abundance.

An example of a particle candidate that could produce $\langle \sigma v \rangle = 3 \times 10^{-9}\ \text{GeV}^{-2}$ is a particle that annihilates via the Z boson. Such a candidate falls under the classification of a weakly interacting massive particle (WIMP). For this Z-boson-annihilating candidate, the WIMP would be more massive than the Z boson, which is about 90 GeV. The thermally averaged annihilation cross section would follow the standard form for a point-like scattering since the mass of the Z boson is small compared to the WIMP mass:

$$\begin{aligned} \langle \sigma v \rangle &= \frac{g^4}{256\pi^2 m_\chi^2} \\ &= 3 \times 10^{-9}\ \text{GeV}^{-2} \left(\frac{g}{0.65} \right)^4 \left(\frac{1.1\ \text{TeV}}{m_\chi} \right)^2 . \end{aligned} \tag{2.6}$$

The plugged-in value of $g = 0.65$ is the coupling constant for the weak force. The fact that such a coupling constant combined with a larger-than-Z-boson mass would yield a cross section that is consistent with the relic abundance of dark matter has been termed the "WIMP miracle" and so motivates the search for TeV-scale freeze-out-produced dark matter. This regime has received

special attention from the particle physics community because it is a natural energy regime for new physics. One such new physics candidate is supersymmetry. In general, the above argument applies not just for the Z boson, but for any massive mediator, such as the Higgs boson or a squark, which is a hypothetical particle from supersymmetry.

The input values for g and m_χ are certainly not the only combination of these parameters that would produce the appropriate value for $\langle\sigma v\rangle$. We can increase both m_χ and g, but we need to maintain $g \lesssim 1$ so that the theory remains perturbative. This constraint on g places an upper bound on m_χ in this simplistic model at roughly 10 TeV. We can also decrease both m_χ and g and achieve the appropriate value for $\langle\sigma v\rangle$, but then we are required to use a less massive boson to mediate the interaction so that m_χ remains larger than the mass of the mediator. Such combinations of parameters would no longer involve weak-scale physics and so have been referred to as WIMPless models. A contemporary version of a WIMPless model is the "Elastic Scalar" model described in Section 2.4.

There is an alternative WIMP regime below the mass of the Z boson wherein the hypothetical strength of a Z-boson-mediated annihilation now depends on the mass of the Z boson; in this regime, the electroweak symmetry is broken, and the weak force takes its familiar form, mediating nuclear processes like beta decays. Recall that it is the relatively large masses of the W and Z bosons that give the weak force its historical name; the mass of the bosons becomes the dominant term in the denominator of the propagator. For WIMPs that interact only via the weak force, the relevant interaction parameter is the Fermi constant $G_F = \sqrt{2}/8 \cdot g^2/M_W^2$, the coupling constant of an effective field theory at these low energies, where M_W is the mass of the W boson.

In this regime below the mass of the Z boson, a thermally averaged cross section can be recast as an annihilation rate per unit particle number density and so is computed with Fermi's Golden Rule:

$$\begin{aligned}\langle\sigma v\rangle = \frac{\Gamma}{n_\chi} &= 2\pi|\mathcal{M}|^2 D(E) \\ &= 2\pi|G_F|^2 \frac{1}{8\pi^2} E^2 N \\ &= \frac{N}{4\pi} G_F^2 m_\chi^2 \\ &= 3\times10^{-9}\ \text{GeV}^{-2} \left(\frac{G_F}{1.166\times10^{-5}\ \text{GeV}^{-2}}\right)^2 \left(\frac{m_\chi}{5\ \text{GeV}}\right)^2 .\end{aligned} \tag{2.7}$$

In the above calculation, $D(E)$ is the density of states, which is the number of energy states per unit energy per unit volume, and thus carries units of GeV^2. N is the number of relativistic species that the dark matter particle could annihilate to and is assumed to roughly cancel the factor of 4π.[4] The key result here is that a weakly interacting dark matter particle around a few GeV (5 GeV is the result of the calculation) may also produce the correct relic abundance.

Without invoking Fermi's Golden Rule, we can also use dimensional analysis to motivate the $\propto m_\chi^2$ scaling: G_F is a coupling constant for an effective field theory and carries units of GeV^{-2}; in order to get the appropriate GeV^{-2} units for the cross section, which involves G_F^2, we can infer two additional powers of energy from m_χ. The $\propto m_\chi^2$ proportionality further reminds us that it does not hold for large m_χ. Once the WIMP mass becomes comparable to the W boson mass, we leave the regime of effective field theory and can no longer use G_F as the coupling constant.

The key takeaway from this discussion is that there is a wide potential range for the dark matter mass m_χ from 1 GeV to 10 TeV that would produce the correct relic abundance if dark

[4] When the temperature falls from roughly 100 MeV to 0.5 MeV, there are about 10 to 20 relativistic species present in the universe.

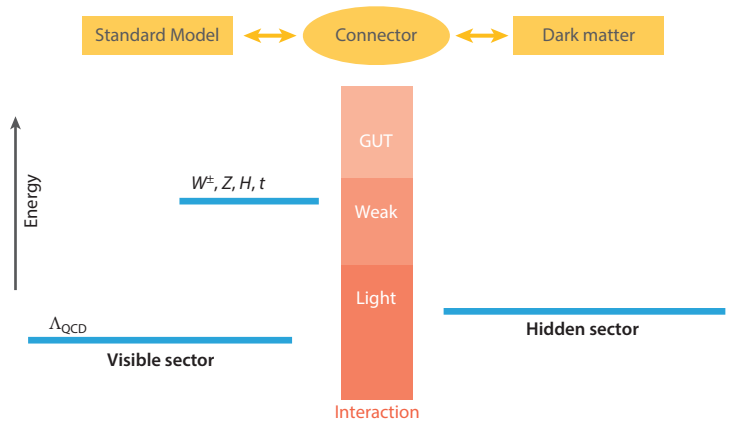

Figure 2.3: **Schematic of the hidden and visible sectors.** Figure reproduced from Zurek (2024).

matter interacts with normal matter via the weak force or a force mediated by electroweak-scale particles.

2.4 The hidden sector: a few keV to 10 GeV

To produce the relic abundance without the use of weak-scale physics, new interactions and theories are required. The vast possibilities for these new interactions and theories are referred to as the hidden sector, in contrast to the visible sector. In general, hidden sector theories may span up to the 100 TeV unitarity limit, but we focus on extensions to the Standard Model below the weak scale, $m_\chi \lesssim 10$ GeV.

Figure 2.3 presents a generic schematic for some overarching structure to the hidden sector. The horizontal lines represent the mass gap for the various regimes: the masses of the W and Z bosons, the Higgs, and the top quark correspond to the electroweak-symmetry-breaking scale; and the mass of the proton, which is a set of confined quarks, provides a rough scale for $\Lambda_{\rm QCD}$. The barrier in the middle of the figure represents the interaction between the visible and hidden sectors: a larger barrier would mean a weaker interaction. The goal of dark matter direct detection experiments is to find an interaction of dark matter with the Standard Model beyond just the force of gravity.

Extensions to the Standard Model that produce the correct relic abundance of dark matter are very loosely constrained. There may be many dark valleys with complicated internal structure across a wide range of energies. What is known though is that, for a fermionic thermal relic, the small-scale matter power spectrum measured from the Lyman-α forest constrains the dark matter mass to be above a few keV; see Section 2.2. However, there could be other hidden sector particles that do not satisfy this mass constraint or could even be massless, such as a "dark photon" mediator between the hidden and visible sectors.

Interactions between the visible and hidden sectors may be mediated by a scalar boson field or a vector boson field. The unique normalizable interactions of an SM-neutral boson compatible with all SM symmetries are the vector portal and the Higgs portal (Battaglieri et al., 2017). In the vector portal, there is essentially a mirror copy of the gauged U(1) symmetry that gives us the photon. The vector field that arises from this hypothetical symmetry is called the dark photon, which can kinetically mix with the SM photon via a coupling term of the form $-\frac{1}{2}\epsilon F_{\mu\nu}F'^{\mu\nu}$, where

ϵ is known as a kinetic mixing parameter. In the Higgs portal, the minimal model mixes a scalar field with a minimal SM Higgs, but these couplings can become more complicated. Since the Higgs field is intimately connected to the mass of particles, many constraints on this particular portal come from particle physics, such as the decay of the muon.

In the vector portal, there are important differences in the phenomenology depending on whether the dark matter particle is a complex scalar, a Dirac fermion, or a Majorana fermion. In the Lagrangian, the mass terms for each of the particles are different, which then changes how the vector field interacts with the U(1) symmetry. Without going through the details of calculating Feynman diagrams, the annihilation cross sections of the Majorana fermion and complex scalar are suppressed by the velocity of the incoming particles. This kind of annihilation is called p-wave annihilation and corresponds to interactions with angular momentum quantum number $\ell = 1$. There is no such velocity suppression for Dirac fermion dark matter; it undergoes s-wave annihilation, which corresponds to interactions with angular momentum quantum number $\ell = 0$; see Battaglieri et al. (2017). These ideas are revisited in Section 2.6.

2.5 New dark matter histories in the hidden sector

Some realizations of the portal to the hidden sector are not compatible with the freeze-out mechanism in that they cannot produce the appropriate relic abundance. Here are three alternatives to the freeze-out mechanism that may produce the correct relic abundance.

Freeze-in

It is possible for dark matter to have never been in thermal equilibrium with the primordial plasma, and it is only very weak interactions with the plasma that cause the dark matter abundance to slowly accrue towards the relic abundance. The dark matter particle would need to have extremely weak couplings to the Standard Model and so is referred to as a Feebly Interacting Massive Particle (FIMP) (Hall et al., 2010). Creation interactions of the FIMP produce FIMPs until they stop because of expansion. If the creation interaction that creates the FIMP has a larger coupling constant, then one would expect more FIMPs to be created and a higher relic density to be produced. Note the qualitative difference from WIMP freeze-out, which is controlled by an annihilation cross section.

Part of the appeal of the freeze-in mechanism is that it is easily reconcilable with supersymmetry. In many models, the particle that creates the FIMP is called the Lightest Observable Sector Particle (LOSP) (Hall et al., 2010). Within the framework of supersymmetry, the LOSP is the lightest superpartner in thermal equilibrium. However, supersymmetry has recently fallen out of favor due to null results from CERN.

Asymmetric Dark Matter

If the dark matter particle is different from its antiparticle, then there may be an asymmetry in the number of particles and antiparticles that could determine the relic abundance (Zurek, 2014). If, say, there were more particles than antiparticles, then once all of the antiparticles have annihilated away, the remaining particles would comprise the relic abundance. In an asymmetric dark matter evolution, annihilations cease because there are no more antiparticles with which particles can annihilate.

The source of this asymmetry may be from a coupling to the Standard Model where there are plenty of asymmetries. There are more baryons than anti-baryons, and more leptons than anti-leptons, so either of these two symmetries could be the source of the dark matter asymmetry. This model was originally motivated by the fact that the energy abundances for baryons and dark

matter are of the same order of magnitude: ∼4% for baryons and ∼26% for dark matter (Zurek, 2014). Heuristically, this factor of 5 difference could be attributed to a factor of 5 difference in the masses of the dark matter particle and the proton mass, so this model would predict a dark matter particle mass around 5 GeV.

ELDERs and SIMPs

All of the above dark matter histories focus on the annihilation or creation interaction of a dark matter particle χ into or from Standard Model particles as the main process that controls the relic abundance. The ELDER/SIMP framework introduces two other interactions that could affect the evolution of dark matter: elastic scattering with SM particles and self-annihilations. The following discussion of the ELDER/SIMP framework comes from Kuflik et al., 2016 and Kuflik et al., 2017.

The three relevant interactions are as follows:

1. annihilations to SM: $\chi + \chi \rightarrow \text{SM} + \text{SM}$
2. elastic scattering with SM: $\chi + \text{SM} \rightarrow \chi + \text{SM}$
3. "$3 \rightarrow 2$" self-annihilations: $\chi\chi\chi \rightarrow \chi\chi$.

"$2 \rightarrow 2$" scatterings between χ particles also occur and keep the χ particles in equilibrium with each other.

With three interactions, the order in which the interactions decouple becomes important. The rates of the first two interactions depend on a kinetic mixing parameter ϵ that couples the hidden sector to the SM. ϵ was first introduced in the previous section as a way to couple the SM U(1) symmetry with a hypothetical hidden sector U(1) symmetry. Self-annihilations depend on a coupling constant α that resides in the hidden sector. The interaction rates also depend on the prevailing number densities of the interacting particles: if the number densities are higher, then the interaction occurs more. We assume that the number densities of the χ particles are exponentially suppressed by a Boltzmann factor of $e^{-m_\chi/T}$ because the universe has cooled to a point where the χ particles are no longer relativistic. The number of χ particles on the "reactants" side of the process then determines the number of Boltzmann factors to include: self-annihilations are most suppressed, followed by annihilations to SM, then elastic scattering. Annihilations to SM always decouple before elastic scattering with SM because annihilations to SM have two Boltzmann factors of suppression.

We first cover the dark matter evolution of the elastically decoupling relic (ELDER). The critical assumption here is that the self-annihilations coupling constant α is large enough that self-annihilations decouple last of the three interactions (order of decoupling: $1 \rightarrow 2 \rightarrow 3$). Consequently, before self-annihilations decouple, χ is no longer in equilibrium with the SM (all interactions with it have decoupled), so χ would be at a different temperature T' from the SM temperature T. This period of time in which self-annihilations are dominant is called cannibalization. Critically, to conserve energy, self-annihilations must inject kinetic energy into the χ particles now that there are fewer of them. These self-annihilations increase the temperature T' relative to T, modifying the Boltzmann suppression to $e^{-m_\chi/T'}$. Applying conservation of entropy to the now decoupled dark matter reveals that T'/T increases quite rapidly; see Figure 2.4 Left. The number density of χ falls much more slowly than it would have if it had remained in equilibrium with the SM primordial plasma. Once self-annihilations decouple and the relic density is frozen in, the relic density has been tightly constrained by the modified Boltzmann suppression, i.e., *the relic density depends very weakly on the exact moment of freeze-out and thus the coupling constant* α. This weak dependence on α is illustrated in the inset of Figure 2.4 Right. In this plot of the co-moving dark matter number density with time, the relic density has a strong dependence on ϵ and the precise moment at which the elastic scatterings decouple, hence the name elastically decoupling relic. After freeze-out, the

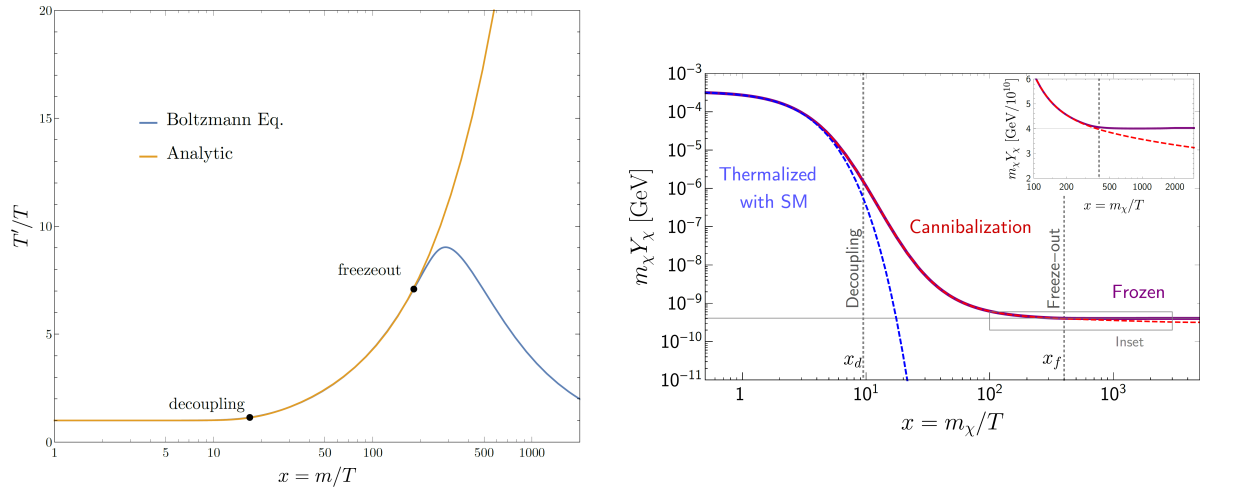

Figure 2.4: **ELDER dark matter evolution in the early universe.** Left: T' of dark matter as a function of the primordial plasma T, which deviate due to self-annihilations of dark matter. The analytic curve does not take into account the decoupling of the self-annihilations. Reproduced from Kuflik et al., 2017. Right: mass density versus time in the ELDER framework. $Y \equiv n/s$, where n is the number density and s is the entropy density. Since $s \propto T^3 \propto a^{-3}$, Y can roughly be interpreted as the co-moving number density. Reproduced from Kuflik et al., 2016.

self-annihilations stop injecting energy into the dark matter, so the temperature falls as a^{-2} (the dark matter is non-relativistic; this relation results from the first Friedmann equation) whereas the temperature of the SM falls as a^{-1}, so the temperature of the SM and χ approach each other after freeze-out of χ, as shown by the Boltzmann solution in Figure 2.4 Left. The main takeaway is that the relic abundance is strongly dependent on ϵ but weakly dependent on α.

For the strongly interacting massive particle (SIMP), α is small enough so that the self-annihilations decouple before the elastic scatterings but after the annnihilations to the SM (order of decoupling: $1 \to 3 \to 2$). When the final annihilation interaction decouples and χ freezes out, χ is still in equilibrium with the SM, so the number density is Boltzmann suppressed by the usual $e^{-m_\chi/T}$. The moment at which self-annihilations decouple has a large impact on the final relic density. This evolution is similar to the process depicted in Figure 2.2, except that self-annihilations govern the freeze-out. Consequently, the relic density is strongly dependent on α and weakly dependent on ϵ.

The so-called "SIMP miracle" is the fact that an m_χ around 100 MeV (which is the QCD symmetry-breaking energy scale) and an α around the strong interaction scale give the appropriate relic density. It is important to note the qualitative difference with the WIMP miracle: here, the strong interaction would be between dark matter particles residing in a mirror copy of the SM strong interaction in the hidden sector, whereas WIMPs interact with the SM via the weak force.

The final possible permutation of the three interactions has the self-annihilations decoupling first (order: $3 \to 1 \to 2$). The last annihilation interaction to decouple is the annihilation to the SM, which means that ϵ determines the relic density. Qualitatively, the dark matter evolution is very similar to the WIMP process, where an annihilation interaction with the SM controls the freeze-out process; refer to Figure 2.2. In this case, it is possible to produce the correct relic abundance without invoking weak-scale couplings. These models are described as "WIMPless."

Figure 2.5 synthesizes the ELDER/SIMP/WIMP(less) relationship into one plot. Each curve represents different values of α and ϵ that give the appropriate relic density. When the curve is

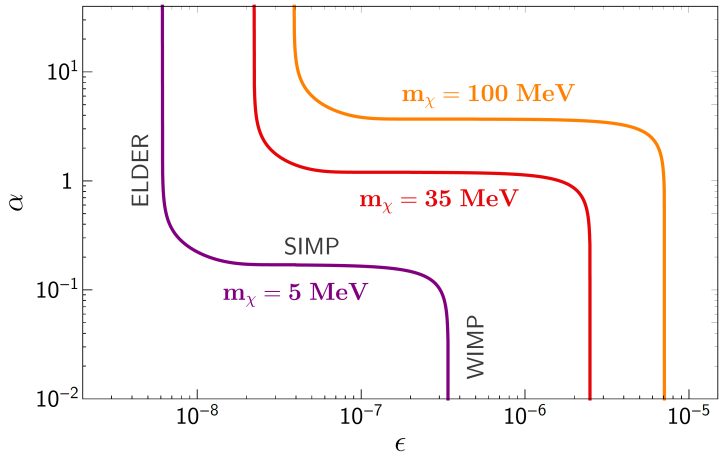

Figure 2.5: **ELDER/SIMP/WIMP scenarios that produce the relic abundance.** α is the coupling constant that governs the self-annihilations and ϵ is the coupling constant that governs the kinetic mixing between the hidden sector and the SM. Reproduced from Kuflik et al. (2017).

vertical, there is a strong dependence on ϵ, and when the curve is horizontal, there is a strong dependence on α.

Summary

As the zeitgeist has shifted away from weak-scale model building, a variety of new theoretical dark matter histories have been developed that hypothesize connections between the Standard Model and the hidden sector. Many of these models are built around the inclusion of a hidden U(1) symmetry.

There has been exploration beyond freeze-out of early universe mechanisms that can produce the correct relic abundance of dark matter. Freeze-in is a mechanism by which the dark matter relic abundance is set by the creation cross section. For ELDERs, the relic abundance is set by the elastic scattering cross section with the SM. For SIMPs, it is the self-annihilation cross section. Some of these models, like the WIMPless models and asymmetric dark matter, are still focused on how the annihilation cross section with the Standard Model may set the relic abundance.

2.6 Defining sharp hidden sector targets for direct detection

The scattering cross sections for a dark matter particle with an electron or a nucleon, $\bar{\sigma}_e$ and $\bar{\sigma}_n$, are hypothetically measurable in a direct detection experiment. For all of the dark matter histories outlined in the previous section, except for SIMPs, we can draw a sharp theory line in the $\bar{\sigma}_e$–m_χ plane under the assumption that electrons are the Standard Model annihilation product (WIMPless or asymmetric dark matter), scattering target (ELDERs), or annihilation parent (freeze-in). Conversion from one of these cross sections to the $\bar{\sigma}_e$ scattering cross section in a direct detection experiment is possible for all the aforementioned interactions because they all arise from the same Feynman diagram. The Feynman diagram would also include the propagator for the mediator, which depends on the mass of the mediator. Following Battaglieri et al. (2017), we only look at the dark-photon vector portal, where the vector field mediator has a mass $m_{A'}$ that is set to $3m_\chi$ for concreteness. We assume electrons for this section, but the following theory targets may also be converted into the $\bar{\sigma}_n - m_\chi$ plane.

Example: annihilation cross section → direct detection cross section

We follow a calculation from Zurek (2024). The thermally averaged annihilation cross section for two m_χ particles to annihilate into electrons through into a vector mediator is given by

$$\langle \sigma v \rangle_V = \frac{16\pi \alpha_\chi \alpha_e m_\chi^2}{\left(4m_\chi^2 - m_{A'}^2\right)^2}, \tag{2.8}$$

where α_χ and α_e are the couplings of the mediator A' with the dark matter and the electron. Specifically, α_e is the fine structure constant. Note that the denominator in Equation 2.8, which arises from the propagator of this theory, is equal to (the available momentum)^2 for the mediator. There are two dark matter particles of mass m_χ that contribute $(2m_\chi)^2$ total energy, and then we subtract $m_{A'}^2$ to determine (the available momentum)^2.

In the case of a heavy mediator, the interaction cross section with an electron is

$$\bar{\sigma}_e = \frac{16\pi \alpha_e \alpha_\chi \mu_{e\chi}^2}{m_{A'}^4}, \tag{2.9}$$

where $\mu_{e\chi}$ is the reduced mass between the electron and the dark matter. Because both the expression for $\langle \sigma v \rangle_V$ and the expression $\bar{\sigma}_e$ contain the product $\alpha_e \alpha_\chi$, the direct detection interaction cross section can be expressed in the following way:

$$\bar{\sigma}_e = \langle \sigma v \rangle_V \left(1 - \frac{4m_\chi^2}{m_{A'}^2}\right)^2 \frac{\mu_{e\chi}^2}{m_\chi^2}. \tag{2.10}$$

The critical concept to glean from this equation is that $\bar{\sigma}_e$ can be completely specified given a dark matter mass m_χ, a mediator mass $m_{A'}$, and an annihilation cross section $\langle \sigma v \rangle$. This last quantity is an input from cosmology and is repeated here:

$$\langle \sigma v \rangle = 3 \times 10^{-9}\,\mathrm{GeV}^{-2} = 1.1 \times 10^{-36}\,\mathrm{cm}^2, \tag{2.5 revisited}$$

where the annihilation cross section has been converted into SI units.

In order to draw a sharp theory line in the $\bar{\sigma}_e - m_\chi$ plane, we must also specify $m_{A'}$. In Figure 2.6, $m_{A'}$ is set equal to $3m_\chi$, thus removing any m_χ dependence from within the parentheses in Equation 2.10. For the model derived in Equation 2.10, we therefore expect $\bar{\sigma}_e \propto m_\chi^{-2}$, which is true for the curves labeled Elastic Scalar and Majorana NR. The curves deviate from m_χ^{-2} at larger masses where m_χ is comparable to the mass of the electron, and the cross section calculations become more complicated. Putting $\langle \sigma v \rangle = 1.1 \times 10^{-36}\,\mathrm{cm}^2$, $m_{A'} = 3m_\chi$, and $m_\chi = 1\,\mathrm{MeV}$ in Equation 2.10, $\bar{\sigma}_e$ is approximately $10^{-37}\,\mathrm{cm}^2$, which is towards the upper end of Figure 2.6 Left.

We provide brief comments on each of the various curves that are drawn in Figure 2.6.

Elastic Scalar

A complex scalar χ may interact with the SM through $\chi + \chi^* \leftrightarrow A' \rightarrow \mathrm{SM} + \mathrm{SM}$. This curve corresponds to the "WIMPless" scenario (Battaglieri et al., 2017). Technically, s-wave annihilations of a complex scalar have been ruled out by the CMB. The drawn theory target is for a p-wave annihilation, which would be adequately velocity-suppressed by the time that recombination occurs and further velocity-suppressed in present-day direct detection experiments.

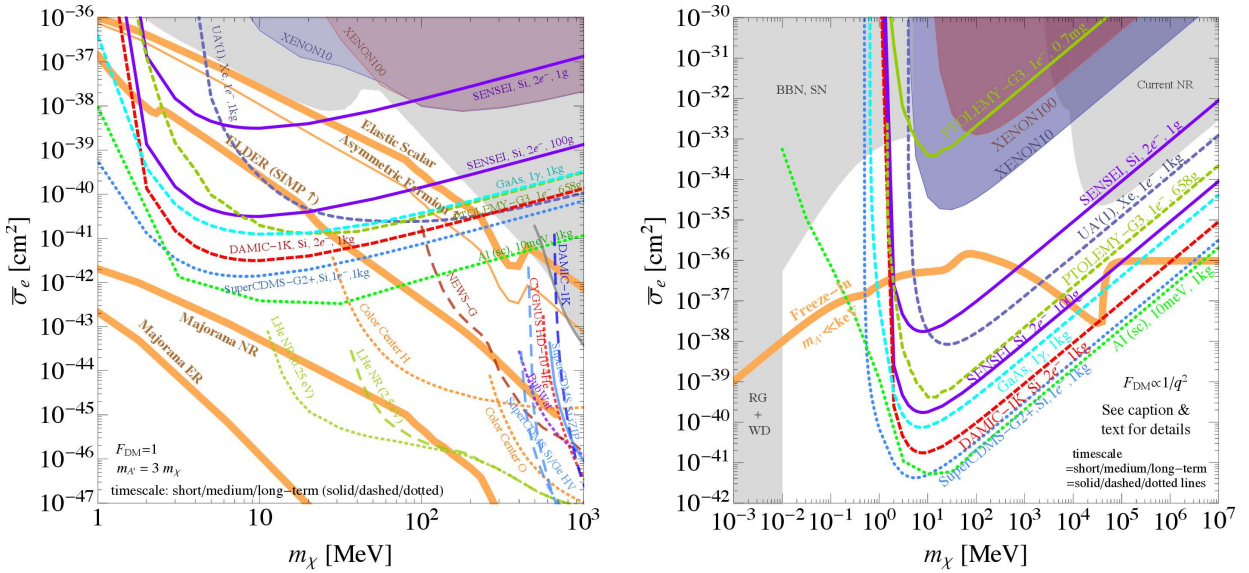

Figure 2.6: **Theory targets for MeV-scale electron-recoiling dark matter.** Left: $\bar{\sigma}_e$–m_χ plane for a heavy vector mediator, $m_{A'} = 3m_\chi$. Right: $\bar{\sigma_e}$–m_χ plane for a massless vector mediator, $m_{A'} \ll m_\chi$. Both: the orange lines show different combinations of m_χ and $\bar{\sigma}_e$ that are compatible with the observed relic density for different dark matter candidates. Note the arrows on the SIMP and Asymmetric Fermion lines, for these represent lower bounds. The shaded regions are current constraints, and the other colored lines are projected constraints for upcoming experiments. Reproduced from Battaglieri et al., 2017.

ELDER/SIMP

The ELDER line falls below the elastic scalar line because the ELDER relic abundance is set by scatterings instead of annihilations. Scatterings are not as Boltzmann suppressed as annihilations (only one dark matter particle is required instead of two), so the correct relic-abundance-producing scattering cross section must be correspondingly lower. The region in between the ELDER line and the elastic scalar line is the region that is compatible with the SIMP cosmology. Given m_χ and $m_{A'}$, a range of scattering cross sections produces the correct relic density because the relic density depends weakly on couplings to the SM. The SIMP region is bounded above by the "WIMPless" elastic scalar line, where annihilations to the SM decouple after self-annihilations (Battaglieri et al., 2017).

Asymmetric Dirac Fermion

If dark matter is a Dirac fermion, then the annihilations would be s-wave and there would be no velocity suppression. Constraints from *Planck* data on the power injected into the CMB by such annihilations imply that the correct relic density can no longer be reached by freeze-out. The correct relic density can be achieved if there is an asymmetry between particle and antiparticle dark matter. The annihilations stop when there are no more antiparticles with which the particles can be annihilated. These annihilations must stop early enough (the annihilation rate must be high enough) so that they do not influence the CMB, which is why the constraint plot shows a lower bound in Figure 2.6 Left as an upward pointing arrow.

Majorana Fermion

If dark matter is a Majorana Fermion, recall from Section 2.4 that it can only undergo p-wave annihilations, which are velocity-suppressed and thus are not constrained by the CMB. In a direct detection experiment, the cross section is suppressed by the velocity of dark matter in the galactic halo. The Majorana line is the Elastic Scalar line multiplied by $2(\mu^2_{\chi,e/n}/m^2_\chi)v^2_\chi$, where $\mu^2_{\chi,e/n}$ is the reduced mass of the χ with either an electron or a nucleon. Both scattering processes are shown in Figure 2.6 Left (Battaglieri et al., 2017).

Freeze-in

If dark matter followed a freeze-in evolution, has a light mediator, and has a mass less than the mass of the electron, the direct detection cross section $\bar{\sigma}_e$ scales as

$$\bar{\sigma}_e \approx 10^{-39}\,\mathrm{cm}^2 \frac{m_\chi}{1\,\mathrm{keV}}. \tag{2.11}$$

This result follows from solving a Boltzmann equation that tracks the production of dark matter in the early universe (Zurek, 2024). This scaling is visible in Figure 2.6 Right below 1 MeV.

Experimental exclusion limits and sensitivity projections

There are a multitude of direct detection experiments that can probe the 1 MeV to 1 GeV mass region. The basic mechanism of detection is very similar to the nuclear recoil measurements for WIMP candidates. Here, because of the low mass of the dark matter particles, detectors search for electron recoils instead of nuclear recoils. If a given hidden sector model includes a mediator that couples to both the electron and nucleus, as is the case for the dark photon, then nuclear recoil exclusion curves can be converted into electron recoil exclusion curves. In the following chapter, we dive deeper into the fundamental principles of direct detection experiments.

Chapter 3

Detector architectures for low-mass particle dark matter

In this chapter, we provide an abbreviated review of the various kinds of particle detector architectures that can be used to probe particle dark matter. It is by no means comprehensive. We discuss these architectures in order of detector maturity; the most mature technologies are those that can be operated at the high exposures and low backgrounds needed to reach some desired level of dark matter sensitivity. In Section 3.1, we lay the groundwork for deciding upon a particular detector architecture given the dark matter mass and kinematics of a particular signal mechanism (nuclear-recoiling or electron-recoiling). In Section 3.2, we discuss particle detection technologies for weak-scale dark matter candidates, otherwise known as WIMPs, in the 10 GeV and above mass range. In Section 3.3, we study single-charge-sensitive detector architectures for low-mass electron-recoiling dark matter. Lastly, in Section 3.4, we explore the burgeoning field of low-threshold phonon-mediated detectors from a few different angles, including sensor architectures, phonon calculations, and material choice.

3.1 Typical recoil energies

Nuclear recoils

The transfer of kinetic energy $\frac{1}{2}m_\chi v^2$ from an incoming dark matter particle of mass m_χ and velocity $\vec{v}$ ($v \approx 10^{-3}$; $c = 1$ throughout) to the recoil energy ΔE_{NR} of some target nucleus with mass m_N can be computed with non-relativistic kinematics. In the lab frame, the dark matter particle imparts some momentum $\vec{q}$ upon the essentially motionless nucleus, providing the nucleus with recoil energy

$$\Delta E_{\mathrm{NR}} = q^2/2m_N. \tag{3.1}$$

Note that, throughout this section, the $\vec{}$ notation indicates a 3-momentum, and the absence of the $\vec{}$ symbol for a vector quantity indicates a Euclidean magnitude; in other words, for the velocities of dark matter interactions, we can ignore relativistic effects.

Starting from conservation of energy, we can parametrize q in terms of the reduced mass $\mu_{N\chi}$ and the scattering angle in the lab frame θ_{qv} (the angle between $\vec{q}$ and $\vec{v}$), and then insert the result back into Equation 3.1:

$$\begin{aligned}\Delta E_{\mathrm{NR}} &= -\Delta E_\chi \\ \frac{q^2}{2m_N} &= \frac{1}{2}m_\chi v^2 - \frac{(m_\chi \vec{v} - \vec{q})^2}{2m_\chi}\end{aligned}$$

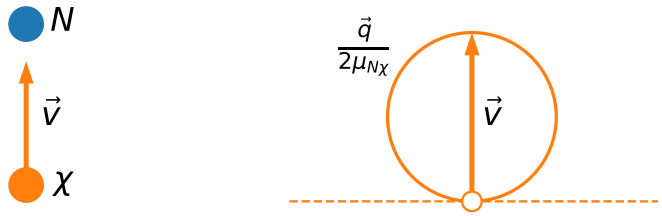

Figure 3.1: **The kinematics of free nuclear recoils.** Left: schematic of a dark matter particle χ about to scatter off a nucleus N. Right: the allowed values of $\vec{q}/2\mu_{N\chi} = v\cos\theta_{qv}$ for a nuclear recoil, given by Equation 3.3. The exclusion of $\vec{q} = 0$ is indicated by the empty circle at the origin. All values of $\vec{q}$ have some component that points in the direction of $\vec{v}$, i.e., $\vec{q}\cdot\vec{v} > 0$. The dashed line is tangent to the circle at the origin and has been drawn so as to highlight its differences with the tangent line drawn in Figure 3.2.

$$\frac{q^2}{2m_N} = \vec{q}\cdot\vec{v} - \frac{q^2}{2m_\chi}$$

$$\frac{q^2}{2\mu_{N\chi}} = \vec{q}\cdot\vec{v} \tag{3.2}$$

$$q = 2\mu_{N\chi} v\cos\theta_{qv} \qquad -\frac{\pi}{2} < \theta_{qv} < \frac{\pi}{2} \tag{3.3}$$

$$\longrightarrow \Delta E_{\rm NR} = \frac{2\mu_{N\chi}^2 v^2}{m_N}\cos^2\theta_{qv} \qquad -\frac{\pi}{2} < \theta_{qv} < \frac{\pi}{2} \tag{3.4}$$

$$\longrightarrow \Delta E_{\rm NR} = \frac{1}{2}m_\chi v^2 \frac{4m_\chi m_N}{(m_\chi + m_N)^2}\frac{1+\cos\theta_{\rm CM}}{2} \qquad -\pi < \theta_{\rm CM} < \pi, \tag{3.5}$$

where $\theta_{\rm CM}$ is the scattering angle in the center-of-momentum frame. In this case, where one of the particles is motionless in the lab frame, $\theta_{\rm CM} = 2\theta_{qv}$. The constraint $-\pi/2 < \theta_{qv} < \pi/2$ arises from the fact that momentum must be transferred to the nucleus in the direction of the incoming dark matter particle; the constraint is also necessary to satisfy Equation 3.2. The complete solution to Equation 3.2 is given by Equation 3.3 and is plotted in Figure 3.1 in the form of $q/2\mu_{N\chi} = v\cos\theta_{qv}$, a circle with radius v along the direction of $\vec{v}$. All possible values of $\vec{q}$ in the collision of a dark matter particle with a motionless nucleus transfer kinetic energy to the nucleus. We note that the final line of the derivation, Equation 3.5, is the same as Equation 3.8 in Lewin et al. (1996).

In order to compute a typical recoil energy, we must specify a typical scattering angle. $\theta_{qv} = \theta_{\rm CM} = 0$ corresponds to head-on collision, the maximal transfer of kinetic energy, and the top of the circle in Figure 3.1 Right. The boundaries of the scattering angle constraints correspond to small-angle scattering of the dark matter particle, the vanishing transfer of kinetic energy, and the bottom of the circle in Figure 3.1 Right. The typical scattering is chosen to be the middle of these two regimes, corresponding to 45° scattering angles in the lab frame and 90° scattering angles in the center-of-momentum frame. $\cos^2 45° = 1/2$ is also the mean value of $\cos^2\theta_{qv}$. This angle is inserted into Equation 3.5:

$$\Delta E_{\rm NR} = \frac{1}{2}m_\chi v^2 \frac{2m_\chi m_N}{(m_\chi + m_N)^2}, \quad v = 10^{-3} \tag{3.6}$$

$$\xrightarrow{m_\chi \approx m_N} \frac{1}{4}m_\chi v^2 = 5\ {\rm keV}\left(\frac{m_\chi}{20\ {\rm GeV}}\right) \tag{3.7}$$

$$\xrightarrow{m_\chi \gg m_N} m_N v^2 = 20 \text{ keV} \left(\frac{m_N}{20 \text{ GeV}}\right) \tag{3.8}$$

$$\xrightarrow{m_\chi \ll m_N} \frac{m_\chi^2 v^2}{m_N} = \frac{1}{2} \text{ eV} \left(\frac{m_\chi}{100 \text{ MeV}}\right)^2 \left(\frac{20 \text{ GeV}}{m_N}\right). \tag{3.9}$$

We are primarily interested in the final regime, $m_\chi \ll m_N$, given the range of low-mass dark matter particles that are discussed in Chapter 2 and the range of available nuclear targets. The $E_{\text{NR}} \propto m_\chi^2$ scaling relation in this regime is critical for determining the type of detector architecture that is appropriate for detecting a certain dark matter mass.

Electron recoils

Electrons are well suited for low-mass dark matter detection because of their much smaller mass relative to nuclei. A free electron recoil extracts more kinetic energy from an MeV-scale dark matter particle than a free nuclear recoil. Furthermore, the kinematics of electron recoils must be treated differently from nuclear recoils because of the electron velocity; electrons are not motionless like nuclei. The final deposited energy cannot be characterized as $q^2/2m_e$ because the electron has some initial momentum $\vec{p} = m_e \vec{v}_e$ (non-relativistic expression because $v_e \ll 1$, as we see below). Instead, the electron recoil energy ΔE_{ER} can be expressed in the following way, reproduced and slightly modified from Hochberg, Pyle, et al. (2016):

$$\begin{aligned} \Delta E_{\text{ER}} &= \frac{(m_e \vec{v}_e + \vec{q})^2}{2m_e} - \frac{1}{2} m_e v_e^2 \\ &= \frac{q^2}{2m_e} + \vec{q} \cdot \vec{v}_e. \end{aligned} \tag{3.10}$$

An important assumption that is made in Equation 3.10 is that there is no binding energy for the electron to overcome or at least that the binding energy is small compared to the transferred kinetic energy. A binding energy is introduced later in this section.

As discussed in Hochberg, Pyle, et al. (2016), a key result of Equation 3.10 is that the deposited energy depends on the initial velocity of the electron. To see the implications of this fact more explicitly, we work through a similar calculation as was done for a free nuclear recoil in Equation 3.3 to arrive at the momentum transfer for a free electron recoil:

$$\Delta E_{\text{ER}} = -\Delta E_\chi$$

$$\frac{q^2}{2m_e} + \vec{q} \cdot \vec{v}_e = \frac{1}{2} m_\chi v^2 - \frac{(m_\chi \vec{v} - \vec{q})^2}{2m_\chi}$$

$$\frac{q^2}{2m_e} + \vec{q} \cdot \vec{v}_e = \vec{q} \cdot \vec{v} - \frac{q^2}{2m_\chi} \tag{3.11}$$

$$\frac{q^2}{2\mu_{e\chi}} = \vec{q} \cdot (\vec{v} - \vec{v}_e) \tag{3.12}$$

$$q = 2\mu_{e\chi} v_{\text{rel}} \cos\theta_{q v_{\text{rel}}} \qquad -\frac{\pi}{2} < \theta_{q v_{\text{rel}}} < \frac{\pi}{2}, \tag{3.13}$$

where $\vec{v}_{\text{rel}} \equiv \vec{v} - \vec{v}_e$ is the relative velocity between the dark matter particle and the electron, and $\theta_{q v_{\text{rel}}}$ is the angle between $\vec{q}$ and $\vec{v}_{\text{rel}}$. Note the similarity of Equation 3.12 and Equation 3.13 with Equation 3.2 and Equation 3.3, the main difference being $\vec{v} \to \vec{v}_{\text{rel}}$. The circle of allowed $\vec{q}/2\mu_{e\chi}$ in Figure 3.2 Right is now centered around $\vec{v}_{\text{rel}}$ instead of just $\vec{v}$. As argued in Essig et al. (2016) and Hochberg, Pyle, et al. (2016), the velocity of the outer electrons in semiconductors or of the electrons near the Fermi energy in metals is about $\alpha \approx 1/137 \approx 10^{-2}$, i.e., the Fermi velocity. v_e for

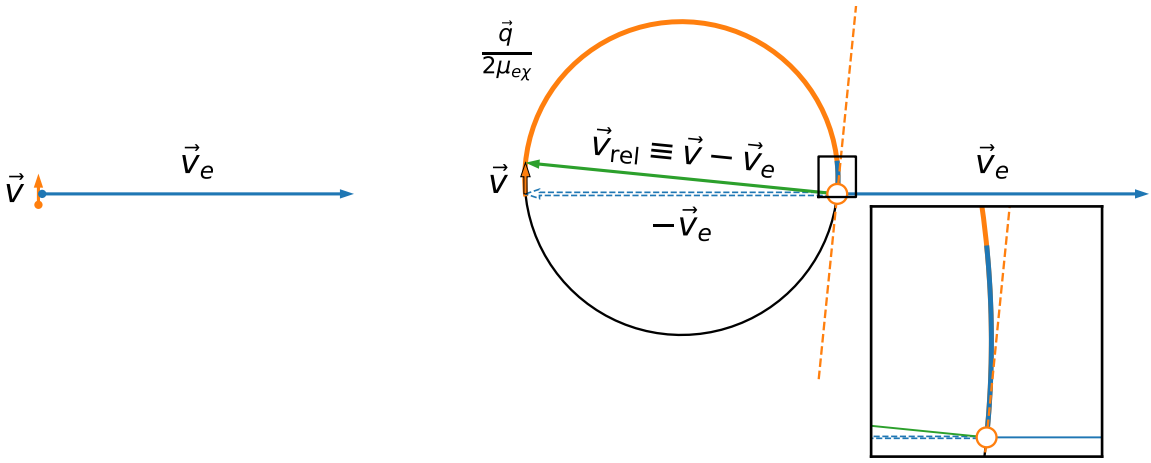

Figure 3.2: **The kinematics of free electron recoils.** Left: a schematic of an electron recoil wherein the dark matter velocity is perpendicular to the electron velocity. Right: the allowed values of $q/2\mu_{e\chi}$ for this electron recoil, now centered on $\vec{v}_{\rm rel} \equiv \vec{v} - \vec{v}_e$ and given by Equation 3.13. The colored segments over the top half of the circle indicate that $\vec{q} \cdot \vec{v} > 0$. The smaller blue segment highlighted by the inset indicates the limited range of $\vec{q}$ such that $\vec{q} \cdot \vec{v}_e > 0$. As before, the empty circle at the origin corresponds to zero momentum transfer, i.e., no collision occurs, and the dashed line is the tangent to the circle at the origin, now rotated by almost 90° from the dashed line in Figure 3.1 due to the rapid velocity of the electron.

these electrons thus dominates over the dark matter velocity $v \approx 10^{-3}$, which means that $\vec{v}_{\rm rel} \approx -\vec{v}_e$. In the following paragraphs, we study under what circumstances $\Delta E_{\rm ER}$ is positive and therefore detectable for different regimes of m_χ.

Large dark matter mass: $m_\chi \gg m_e$

In this regime, the expression $\mu_{e\chi} v_{\rm rel}$ in Equation 3.13 is simply the electron momentum, which can be estimated from the Fermi momentum or $m_e v_e$: 3 keV to 5 keV. The typical deposited energy $\Delta E_{\rm ER}$ is most easily estimated from the right-hand side of Equation 3.11, using $\mu_{e\chi} v_{\rm rel} \approx m_e v_e = 4\,{\rm keV}$ and Equation 3.13 for q:

$$\Delta E_{\rm ER} = \vec{q} \cdot \left(\vec{v} - \frac{\vec{q}}{2m_\chi} \right) \approx \vec{q} \cdot \vec{v} = 2 \times 4\,{\rm keV} \times \cos\theta_{q v_{\rm rel}} \times 10^{-3} \times \cos\theta_{qv} \approx 4\,{\rm eV}, \tag{3.14}$$

where we assume that $m_\chi \gg q/2v = \mu_{e\chi} v_{\rm rel} \cos\theta_{q v_{\rm rel}}/v \approx m_e v_e \cos\theta_{q v_{\rm rel}}/v \approx 4\,{\rm MeV}$ and that a typical value of $\cos\theta_{q v_{\rm rel}}$ is $1/\sqrt{2}$. Note that $\cos\theta_{qv}$ would then approximately be $1/\sqrt{2}$ as well. Furthermore, note that $\vec{q} \cdot \vec{v}$ is analogous to Equation 3.8, the typical nuclear recoil energy when $m_\chi \gg m_N$: $m_N v^2$, except that $m_N v$ is replaced by $m_e v_e$. The accessible energy is boosted by a factor of v_e/v. It is important to emphasize that, unlike in the case of free nuclear recoils, not all collisions of dark matter with free electrons bring about an increase in electron energy. This fact is indicated by the partial coloring of the allowed $\vec{q}/2\mu_{N\chi}$ circle in Figure 3.2 Right where $\vec{q} \cdot \vec{v}$ is positive. The intuition for why only about half of the $\vec{q}/2\mu_{N\chi}$ circle provides the electron with kinetic energy is the following: in the dark matter rest frame, all electron scatters lose kinetic energy (as is the case for dark matter particles that scatter off motionless nuclei). Only the electrons that are scattered in the direction of the dark matter's lab frame velocity receive boosting to their kinetic energy when transformed back to the lab frame. The kinematics are the same as for spacecraft that

undergo gravity-assisted acceleration, which was performed on the two Voyager probes. On the other hand, if $\vec{v}$ and $\vec{v}_e$ point in opposite directions, then all values of $\vec{q}$ would produce $\vec{q} \cdot \vec{v} > 0$.

Small dark matter mass: $m_\chi \ll m_e$

In this regime, $\mu_{e\chi} v_{\rm rel} \approx m_\chi v_e$. The deposited energy can now be estimated from the left-hand side of Equation 3.11, using $\mu_{e\chi} v_{\rm rel} \approx m_\chi v_e$ and Equation 3.13 for q:

$$\Delta E_{\rm ER} = \vec{q} \cdot \left(\frac{\vec{q}}{2m_e} + \vec{v}_e \right) \approx \vec{q} \cdot \vec{v}_e = 2 m_\chi v_e^2 \cos\theta_{q v_{\rm rel}} \cos\theta_{q v_e} \approx 2 m_\chi v_e^2 \sin^2 \frac{1}{20} \approx \frac{1}{2} m_\chi v^2, \quad (3.15)$$

where $\theta_{q v_e}$ is the angle between $\vec{q}$ and $\vec{v_e}$. Equation 3.15 uses the limit $m_e \gg q/2v_e \approx m_\chi \cos\theta_{q v_{\rm rel}}$ to ignore the $\vec{q}/2m_e$ term in the parentheses. Only a limited range of $\vec{q}$ values produces positive $\vec{q} \cdot \vec{v}_e$ as indicated by the blue arc in Figure 3.2 Right. For Equation 3.15, we use $\theta_{q v_{\rm rel}} = \theta_{q v_e} = \pi/2 - 1/20$ so that $\vec{q}$ points halfway between vertical and the orange dashed line in Figure 3.2 Right, maximizing the product $\cos\theta_{q v_{\rm rel}} \cos\theta_{q v_e} = \sin(\pi/2 - \theta_{q v_{\rm rel}}) \sin(\pi/2 - \theta_{q v_e})$. This direction of $\vec{q}$ is also halfway between $\vec{v}_{\rm rel}$ and $\vec{v}_e$. Notably, Equation 3.15 recovers the full kinetic energy of dark matter, which is the same conclusion as is reached in Hochberg, Pyle, et al. (2016) and broadly motivates the use of electron recoils to search for sub-MeV dark matter.[1] However, there are stringent kinematic conditions for which the maximal transfer of kinetic energy to electrons is possible: $\vec{v}$ cannot be collinear with $\vec{v}_e$, and the collision must be glancing. The transfer of kinetic energy from dark matter to electrons is much less frequent than the other way around. This fact is indicated by the relatively short length of the blue arc in Figure 3.2 Right where $\vec{q} \cdot \vec{v}_e > 0$.

Ultimately, the depressed rate of $\Delta E_{\rm ER} > 0$ when $m_\chi \ll m_e$ and $v_e \gg v$ means that a strongly constraining dark matter search in this regime requires greater exposure and lower background rates than if (mass of target) $= m_\chi$ and (velocity of target) $= 0$. The large exposure ensures that the just-glancing perpendicular scatterings with $\Delta E_{\rm ER} > 0$ occur frequently enough for detection; a low background rate is then required from the larger exposure. An alternative strategy for probing sub-MeV dark matter would be to use an electron target with $v_e \ll v$, an idea explored in Section 3.2 of Hochberg, Kahn, Lisanti, et al. (2018a) and the "Novel Materials" section of Hochberg, Kahn, Kurinsky, et al. (2021); both examples invoke the need for a low Fermi velocity material. If $v_e \ll v$, then the kinematics would have the same behavior as free nuclear recoils: maximum transfer of kinetic energy when $m_\chi \approx m_e$ and a $\Delta E_{\rm ER} \propto m_\chi^2$ scaling when $m_\chi \ll m_e$. So, less kinetic energy can be transferred to the electron when $v_e \ll v$, but $\Delta E_{\rm ER} > 0$ occurs more frequently than when $v_e \gg v$. In other words, a dark matter search with a $v_e \ll v$ material has less stringent requirements on exposure and background levels but has more stringent requirements on energy threshold for a given dark matter mass, with the additional requirement that the background levels near those lower energy thresholds are sufficiently low.

We briefly revisit the assumptions made in the previous sections. For $m_\chi \gg m_e$, the term $-q^2/2m_\chi$ was dropped. The typical energy transfer calculated and the range of $\vec{q}$ for which $\Delta E_{\rm ER}$ is positive should be considered as overestimates. In contrast, for $m_\chi \ll m_e$, the term $+q^2/2m_e$ was discarded. Thus, the range of $\vec{q}$ for which $\Delta E_{\rm ER}$ is positive should be considered an underestimate.

Electron binding energies

We now focus on the impact of electron binding energies $E_{\rm bound}$ that are comparable to the free recoil energies calculated above. A large $E_{\rm bound}$ introduces inelasticity to particle collisions,

[1] Technically, it is not possible for the full kinetic energy to be transferred from the dark matter to the electron in the case of perpendicular velocities. There needs to be some component of $\vec{v}_e$ in the direction of $\vec{v}$ to completely stop the dark matter particle. The example provided in Section 2 of Hochberg, Pyle, et al. (2016) does so.

which means that, when an electron is liberated from a bound state, its free energy E_{free} cannot be written as $(\vec{p}+\vec{q})^2/2m_e$ as it was for a free electron. If we instead parameterize the electron's absorbed energy as $E_{\text{free}} + E_{\text{bound}}$, then the kinematic equation for conservation of energy becomes the following, reproduced and slightly modified from Equation 2.23 of Trickle et al. (2020) and Equation A.16 from Essig et al. (2016):

$$\begin{aligned} \Delta E_{\text{ER}} = E_{\text{free}} + E_{\text{bound}} &= -\Delta E_\chi \\ &= -\frac{|m_\chi \vec{v} - \vec{q}|^2}{2m_\chi} + \frac{1}{2} m_\chi v^2 \\ &= \vec{q}\cdot\vec{v} - \frac{q^2}{2m_\chi} \\ \longrightarrow v_{\text{min}}(E_{\text{free}} + E_{\text{bound}}, q) &= \frac{E_{\text{free}} + E_{\text{bound}}}{q} + \frac{q}{2m_\chi}. \end{aligned} \tag{3.16}$$

In Equation 3.16, the angle between $\vec{q}$ and $\vec{v}$ is set equal to 0, corresponding to head-on collision and maximal transfer of kinetic energy from the dark matter to the electron. Thus, v_{min} is the minimum dark matter velocity that could transfer momentum $\vec{q}$ and energy $E_{\text{free}} + E_{\text{bound}}$ to produce a free electron with energy E_{free}.

It is again useful to see how the $(E_{\text{free}} + E_{\text{bound}})/q$ term of Equation 3.16 behaves in the two regimes of earlier discussion: $m_\chi \gg m_e$ and $m_\chi \ll m_e$. When the dark matter mass is large, the typical momentum transfer is given by the electron momentum, as stated before. For $E_{\text{bound}} \sim O(1\,\text{eV})$, which is the case for semiconductor bandgaps between the valence and conduction bands,[2] $q \approx 3\,\text{keV}$ to $5\,\text{keV}$ would constrain v_{min} to be 2×10^{-4} to 3×10^{-4}. The majority of the dark matter particles in the galactic halo have greater speed than this range. In other words, the rate of electron recoils is not dramatically suppressed by the presence of $O(1\,\text{eV})$ bandgaps when the dark matter mass is large. When $m_\chi \ll m_e$, the typical momentum transfer is given by $q \approx m_\chi v_e$. For $m_\chi = 50\,\text{keV}$, $v_{\text{min}} \approx 1/500$ is required to free an electron from a $1\,\text{eV}$ binding energy. This constraint depresses the rate of detectable electron recoils given a dark matter velocity distribution that is centered around $v \approx 10^{-3}$. In short, semiconductor bandgaps do not greatly inhibit the ability of direct detection experiments to search for $m_\chi \gg m_e$ dark matter but are a significant impediment to the search for $m_\chi \ll m_e$ dark matter. In the latter case, both sub-eV electron recoil energy thresholds and sub-eV binding energies are required.

Detector architectures for low-mass dark matter

The key ideas of this chapter are summarized in Figure 3.3. The typical and maximum nuclear recoil energies as a function of dark matter mass for silicon and germanium are plotted in blue and orange, as computed from Equation 3.6. The regime $m_\chi \approx m_N$ is towards the right edge of the plot and corresponds to when $E_{\text{NR}} \propto m_\chi$. The regime where $m_\chi \gg m_N$ and $E_R \propto m_N$ is to the right of the plot limits. The maximum nuclear recoil energy is $2\times$ the typical nuclear recoil energy and is plotted with the dashed curves. Only when $m_\chi = m_N$ does the maximum recoil energy curve make contact with the red dashed line, indicating full transfer of dark matter kinetic energy and optimal kinematic matching.

When the nuclear recoil energy is greater than tens of eV, the nucleus is ejected from its site, and some amount of charge is produced. This energy regime is indicated in green in Figure 3.3.

[2] For reference, in silicon and germanium, the bandgaps are 1.11 eV and 0.67 eV. These are the minimum energies required to produce an electron-hole pair. The *average* energies to produce an electron-hole pair are 3.6 eV and 2.9 eV. See Section 5 of Essig et al. (2016).

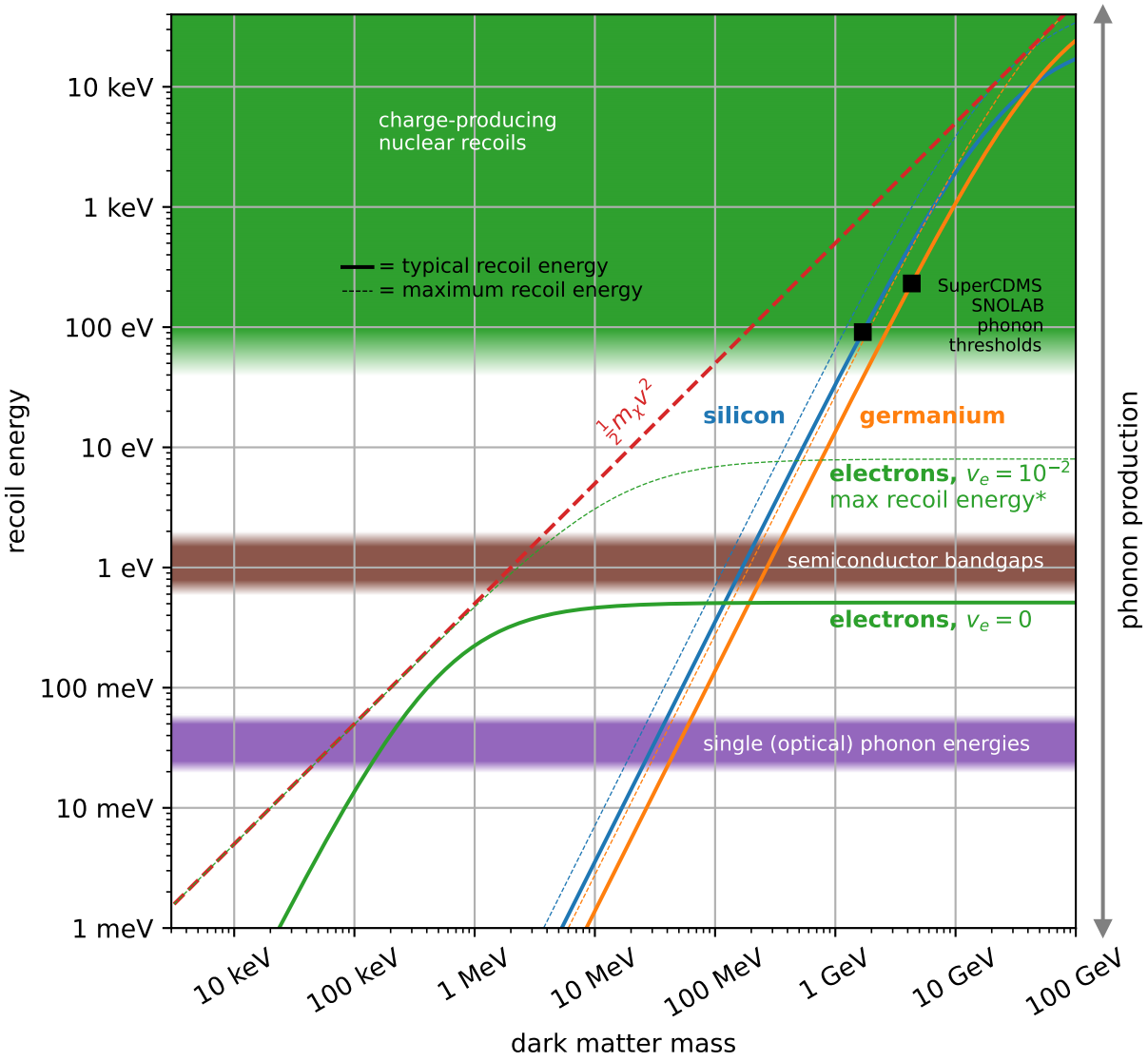

Figure 3.3: **Recoil energy for silicon nuclei, germanium nuclei, and electrons versus dark matter mass.** Typical and maximum recoil energies are compared against the dark matter kinetic energy. * indicates that the green dashed curve is constructed in an ad hoc way so as to match the limiting behaviors that were computed earlier in Section 3.1. Relevant energy regimes are drawn with different colored bands. Phonon production spans the entire plotted nuclear recoil energy range. SuperCDMS SNOLAB phonon thresholds are indicated.

The energy required to displace the nucleus is referred to as the displacement threshold energy and was measured to be 19.7 eV in germanium (SuperCDMS Collaboration, 2018) and is believed to range from 20 eV to 40 eV for silicon (Holmström et al., 2008). SuperCDMS SNOLAB targets recoil energies at hundreds of eV and above, as shown by the black points in Figure 3.3. An ideal detector architecture for these energy regimes would be able to measure charge production and use it for discrimination of nuclear recoils from background electron recoils. This technique is described in Section 3.2.

For energies below about 20 eV to 40 eV in silicon and germanium, nuclear recoils no longer

produce charge; instead, all of the nuclear recoil energy goes directly into the production of phonons. Thus, to probe nucleon-coupled dark matter below about 1 GeV, there is only the phonon channel. For sensitivity to ∼10 MeV dark matter mass in silicon and germanium, meV nuclear recoil thresholds are required. Detector architectures focused on the lowest possible energy thresholds are discussed in Section 3.4. It is important to keep in mind that, for every decade in dark matter mass reach that is desired, two decades of improvement in energy threshold are required.

The typical energy of a single optical phonon is shown as the purple band in Figure 3.3. Due to the constant dispersion relation of optical phonons, they may be a particularly fruitful phonon channel; see Knapen, Lin, Pyle, et al. (2018) and Section 3.4. At recoil energies below about 100 meV, discretized phonon production becomes relevant. The rate calculations of single-phonon and multi-phonon excitations from Trickle et al. (2020) and Campbell-Deem et al. (2022) are examined in Section 3.4.

The electron recoil energy curves are plotted in green. As a reference, the typical $v_e = 0$ electron recoil energy calculated by Equation 3.6 is drawn as the solid curve in Figure 3.3 to indicate how electrons have better kinematic matching than nuclei; nearly five orders of magnitude separate the masses of electrons and the plotted nuclei. In practice, electrons have velocity $v_e \approx 10^{-2}$, which dramatically changes the kinematics from the $v_e = 0$ case. There are two main differences to highlight, as illustrated by the dashed green curve. First, the dashed curve is not $2\times$ the solid curve as is the case for nuclear recoils. In the $m_\chi \gg m_e$ regime, the maximum accessible recoil energy is boosted by a factor of v_e/v. The curve is drawn to level out at 8 eV, corresponding to collinear electron and dark matter velocities and head-on collision. Second, when $m_\chi \ll m_e$, the entire kinetic energy of the dark matter is accessible to the rapidly moving electron, but only for specific kinematic scenarios: roughly perpendicular electron and dark matter velocities and a glancing collision. As indicated in the caption of Figure 3.3, the dashed curve for electrons is drawn in an ad hoc way so as to match the two limiting behaviors that are computed earlier in this section. Across all regimes, the kinematics of an electron recoil allows for greater transfer of kinetic energy, but the fractional rate of maximal energy transfer is less than the fractional rate of maximal energy transfer from a nuclear recoil due to the stringent kinematic requirements for maximal energy transfer from an electron recoil.

Lastly, only certain electron recoils produce enough energy to overcome certain electron binding energies. When $m_\chi \gg m_e$, the recoil energy boost v_e/v is sufficient to liberate the maximally recoiled electrons given semiconductor bandgaps. Charge-sensitive semiconductor detectors with > 1 eV thresholds are well suited for MeV–GeV dark matter detection and, furthermore, are generally accessible for fabrication. They are studied in Section 3.3. In contrast, when $m_\chi \ll m_e$, smaller electronic bandgaps are required in addition to meV energy thresholds. Such materials include Dirac materials and superconductors, explored in Hochberg, Kahn, Lisanti, et al. (2018b) and Hochberg, Zhao, et al. (2016), and are much less readily accessible for detector fabrication.

3.2 Charge-producing nuclear recoils for WIMP searches

Liquid-noble-element detectors: scintillation & charge

Liquid-noble-element detectors have proven to be very effective for WIMP searches above 10 GeV. In the case of the LUX-ZEPLIN (LZ) experiment, which operates a liquid xenon detector, the combination of ultra-radiopurity down to 10^{-5} dru[3] (LZ Collaboration, 2023a) and experimental scalability up to ton·yr exposures (LZ Collaboration, 2023b) has been the key to the sensitivity of the LZ detector.

[3] 1 dru=1 event/kg/day/keV

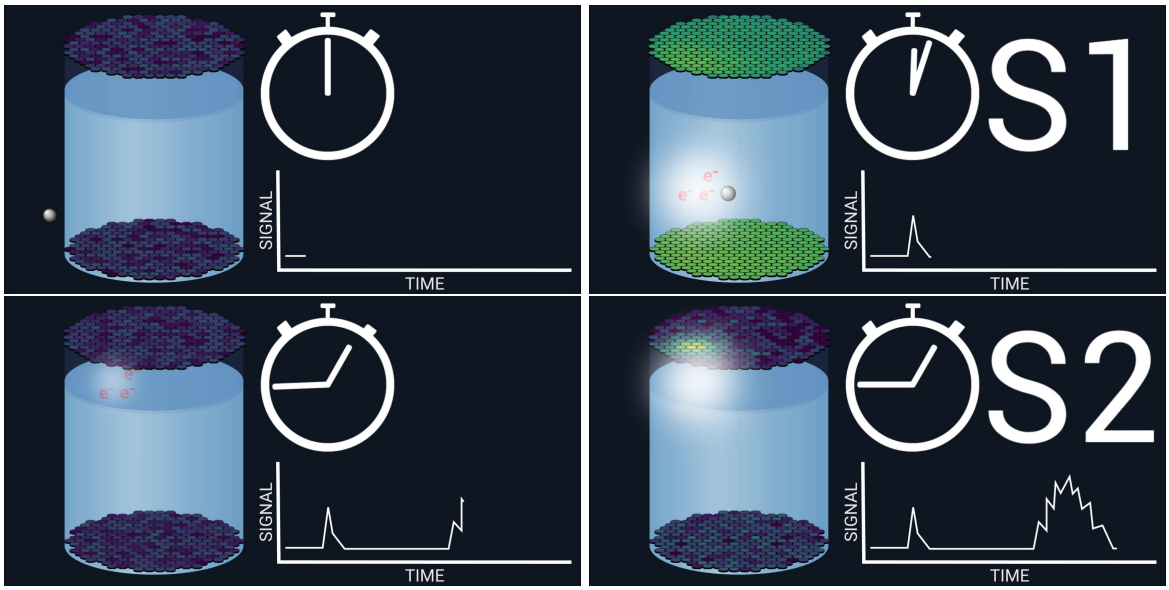

Figure 3.4: **Xenon time projection chamber.** Figure panels are screenshots of a video from XENONnT Collaboration (2024).

A 30 GeV nuclear-recoiling dark matter interaction would cause an approximately 10 keV nuclear recoil in xenon, producing both scintillation photons and ionization. Liquid noble experiments leverage both forms of signal in a detector architecture referred to as a time projection chamber. The production and detection of these two signals are illustrated in Figure 3.4. Both scintillation and ionization signals are produced immediately upon impact of an incoming particle, as shown in Figure 3.4 Top Right. The initial scintillation photons, illustrated as the faint white aura, propagate immediately to the photomultiplier tubes at the top and bottom of the detector; this measurement is called the S1 signal. Over the following milliseconds, a vertical electric field drifts the electrons to the top of the liquid xenon, where there is an interface between gaseous and liquid xenon. Here, a stronger electric field extracts the electrons from the liquid and accelerates them through the gaseous xenon where proportional electroluminescence converts the ionization signal to light; see Figure 3.4 Bottom Left. The measurement of these secondary photons is called the S2 signal, which is localized to the photomultiplier tubes that are closest to the electrons. It provides a measure of the charge production from the initial particle event. The relative strengths of the S1 and S2 signals can be used to discriminate between nuclear recoils and electron recoils; an electron recoil produces a fractionally stronger S2 signal than a nuclear recoil.

Argon is another liquid noble element that has been used in time projection chambers. The main experiment that operates an argon time projection chamber is the DarkSide Collaboration. For searches above 10 GeV dark matter mass, argon-based searches benefit from additional electron recoil discrimination via the S1 pulse shape, which is independent of the S2 to S1 ratio (DarkSide-50 Collaboration, 2018). For low-mass dark matter searches, argon has better kinematic matching than xenon because of its lower atomic mass ($\sim$40 amu versus $\sim$131 amu). Currently, the world-leading exclusion limit in the range from 2 GeV to 5 GeV for low-mass WIMPs is from Collaboration (2023), which uses an S2-only analysis.

The typical mode of detector operation for low-mass dark matter searches in liquid-noble-element detectors is high-voltage, S2-only (charge-only) readout, analogous to the high-voltage operation of semiconductor detectors that are discussed in Chapters 4 and 5. The main obstacle to

these kinds of searches in liquid-noble-element detectors are single-electron backgrounds, consisting of spontaneous extraction of single electrons from the liquid noble element or emission from other sources (e.g., wires). These backgrounds tend to be worse in xenon than in argon.

Semiconductor detectors: phonons & charge

In silicon and germanium, nuclear recoils greater than about 50 eV dislocate the nucleus from its lattice site in the crystal. The following description for what happens next is taken directly from Particle Data Group (2014): "The recoiling particle deposits energy along its track, with the majority going directly into phonons. A minority of the energy goes directly into the production of charge. Energy deposited in ionization is recovered when the carriers recombine." Similar to liquid-noble experiments that separately measure scintillation and charge, many semiconductor detectors are constructed to separately measure the production of phonons and charge. A nuclear recoil produces fractionally less charge than an electron recoil.

The charge is collected by applying an electric field throughout the detector. The collected charge is read out by transconductance amplifiers attached to the ionization bias electrodes. SuperCDMS uses high electron mobility transistors (HEMTs) in these amplifiers (SuperCDMS SNOLAB TDR, 2018) and anticipates charge resolutions of about 50 electrons rms (SuperCDMS Collaboration, 2023). In the past, SuperCDMS has used junction field-effect transistors (JFETs) for charge sensing. Current HEMT-based techniques for charge sensing have reached charge resolutions as low as 7 to 10 electrons by minimizing parasitic capacitances; see Anczarski et al. (2024) and Ricochet Collaboration (2024).

There are many types of phonon sensors that can be used for phonon-mediated particle detection. Thermal phonon sensing offers one such avenue for detection. The initial recoil energy thermalizes over the course of milliseconds and raises the temperature of the crystal's blackbody phonon spectrum over that timescale. CDMS I and EDELWEISS I–III were two experiments that used neutron-transmutation-doped (NTD) germanium thermistors to sense the temperature change. In both cases, the resistance change of the thermistor was amplified with a JFET-based voltage amplifier. The simultaneous detection of charge and phonon signals enabled nuclear recoil discrimination from electron-recoiling backgrounds. In a review of sub-Kelvin detectors in Particle Data Group (2014), the ultimate phonon energy resolution of a 10 mK, 1 kg detector that senses thermal phonons is calculated to be about 5 eV and is in practice usually about an order of magnitude worse due to readout noise or unwanted power dissipation. This energy resolution is still sufficient for detecting dark matter with a mass greater than 10 GeV.

Athermal phonon sensing is perhaps a preferred alternative to thermal phonon sensing. There are two main reasons: (1) sensing athermal phonons before they down-convert via anharmonic scattering to the thermal phonon population enables position reconstruction of the phonon-producing event and consequent rejection of backgrounds; and (2) athermal phonon-sensing architectures are not subject to the crystal thermal noise, only the intrinsic noise of the sensor. Superconducting sensors are a prime candidate for the detection of athermal phonons; the minimum quanta of detectable energy in a superconductor is the superconducting bandgap, which for most superconductors ranges from 10 μeV to 1 meV. The current and future possibilities for athermal-phonon-sensing architectures are discussed in Section 3.4.

We also provide some background information for the energy and frequency scale of athermal phonons. 1 THz ($\approx$ 4 meV) is the frequency at which phonons begin to propagate ballistically over a cm-sized chip. The phonons produced during the down-conversion cascade and recombination of electron-hole pairs have much higher energies (optical phonons in silicon are about 60 meV; see Griffin, Inzani, et al., 2020) but then undergo frequent anharmonic scatterings and down-convert to lower energies. The rate of anharmonic scatterings goes as ν^5. At frequencies around 1 THz,

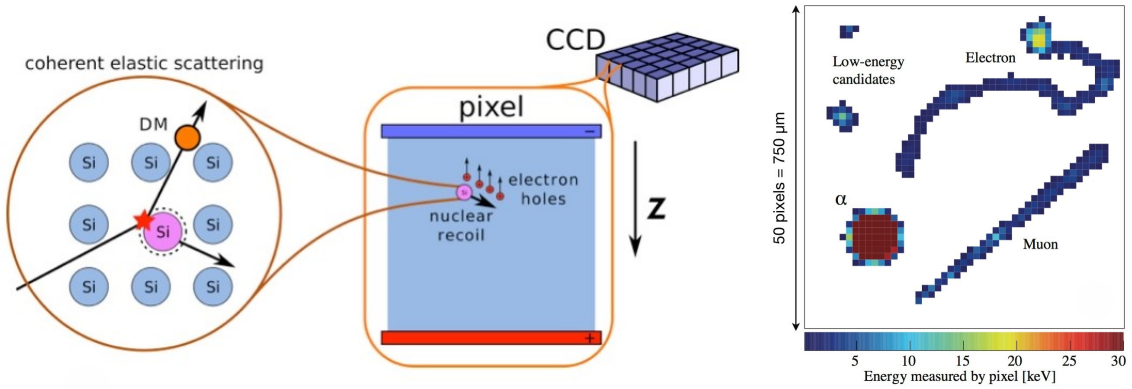

Figure 3.5: **Charge-coupled devices for dark matter.** Figures are from DAMIC-M Collaboration (2024).

the rates of anharmonic and isotopic scatterings (which change the direction of phonons but do not change their energy; the rate of isotopic scatterings goes as ν^4) decrease such that ballistic propagation through the crystal is possible. See Pyle (2012), Martinez et al. (2019), Particle Data Group (2014), and the references therein for more information.

3.3 Single-charge-sensitive detectors

For signal models of electron-recoiling dark matter with mass greater than the electron, about 1 to 10 free electrons are expected to be produced (Essig et al., 2016). The rapid velocity of electrons enables enhanced transfer of kinetic energy, as described in Section 3.1. The many theory targets that would yield this range of charge production motivate a detector architecture that is sensitive enough to detect single electrons.

A silicon charge-coupled device (CCD) is an excellent detector architecture that is capable of single-electron sensitivity. In the CCD schematic shown in Figure 3.5, it is organized into 15μm$\times 15\mu$m pixels, each with a 3-phase polysilicon gate that can hold and transfer charge to adjacent pixels. After the charge is collected in the various pixels, it is brought to an on-chip charge-to-voltage amplifier via an assembly-line process called clocking, which essentially is a voltage control protocol of the 3-phase gates. The final product is a picture of the charge-producing events in a silicon chip, as seen in Figure 3.5 Right. Techniques for discriminating nuclear recoils from electron recoils with CCDs are possible, as is evident in Figure 3.5 Right.

The Skipper CCD is an advanced version of the CCD and has provided unprecedented charge-resolving capabilities in direct detection experiments. Skipper CCDs were originally developed in the 1990s for use in the Hubble Space Telescope and other such applications; see Chandler et al. (1990). Skipper CCDs can non-destructively re-measure the charge quantity in a single pixel in order to integrate down readout noise. This resolution-enhancing tactic has proven immensely successful at circumventing $1/f$ low frequency noise that plagues low noise experiments. The result is a truly remarkable baseline charge resolution of $0.068\,e^-$/pixel (rms) (Tiffenberg et al., 2017).

The SENSEI and DAMIC-M collaborations have used CCDs with masses ranging from 2 g to 10 g to search for electron-recoiling dark matter. Presently, SENSEI and DAMIC-M have set the two most stringent constraints on electron-recoiling dark matter in a range around 10 MeV dark matter mass (the size of the range is either one or three decades depending on the choice of mediator mass), both with about 100 g·day exposures (SENSEI Collaboration, 2023) (DAMIC-M Collaboration, 2023). Both collaborations are pursuing larger exposure searches within their low-background

underground laboratories: SENSEI at SNOLAB and DAMIC-M at the Laboratoire Souterrain de Modane. Beyond that, the two collaborations plan to join forces in the near future as the Oscura collaboration.

Notably, SuperCDMS has also developed gram-scale single-charge-sensitive detectors using only phonon sensors. They are known as HVeV detectors. These detectors are covered in Chapters 4 and 5.

3.4 Low-threshold phonon-mediated detectors

Phonon sensor architectures

As indicated by Figure 3.3, detecting nuclear-recoiling dark matter below 1 GeV with a semiconductor detector requires phonon-mediated detectors with $O(10\,\text{eV})$ thresholds. Such thresholds have already been achieved in prototype gram-scale detectors and have enabled world-leading dark matter searches in the sub-GeV mass range: see SuperCDMS Collaboration (2021) and SuperCDMS Collaboration (2022a).

SuperCDMS phonon sensors are known as quasiparticle-trapping-assisted electrothermal-feedback transition-edge sensors (QETs). The design of the SuperCDMS QETs is shown in Figure 3.6. Beginning with CDMS II, the SuperCDMS collaboration has used this sensor architecture to detect athermal phonons. A transition edge sensor (TES) is a superconducting wire that is operated at its critical temperature T_c; a slight change in the temperature of the TES as a result of some absorbed energy results in a large change in the resistance of the wire. The TES is the sensitive element of this architecture.

To maximize sensitivity, a single TES's volume is kept small, around tens of μm^3 ($200\,\mu\text{m} \times 2\mu\text{m} \times 40\,\text{nm}$ are the dimensions of a SuperCDMS High Voltage detector's TES, and is coupled to multiple phonon collection fins, which are responsible for efficient collection of the athermal phonon energy in the substrate. The phonon fins are usually around $10{,}000\,\mu\text{m}^3$ in volume ($700\,\mu\text{m} \times 500\,\mu\text{m} \times 500\,\text{nm}$ are rough dimensions of the aluminum fins of a SuperCDMS High Voltage detector). For efficient phonon collection, hundreds of such cells are wired in parallel over the surface of the detector.

When phonons are absorbed in the collection fins, Cooper pairs are broken and quasiparticles are created. Quasiparticles can be thought of as free electrons in the superconductor, but their precise treatment is more complicated: they are momentum eigenstates of a diagonalized BCS Hamiltonian (Tinkham, 2004). The material for a phonon collection fin is chosen to have a higher T_c than the TES T_c. When there are two materials with different T_cs in contact with each other, randomly diffusing quasiparticles from the higher-T_c material may enter the lower-T_c material and then emit a phonon to become trapped in the lower-T_c material. In this case, the trapping of quasiparticles from the collection fin and into the TES enables the TES to be sensitive to pair-breaking phonons from the substrate. The longest dimension for a phonon fin is usually kept around $500\mu\text{m}$ to 1 mm to accommodate the typical diffusion length of quasiparticles in aluminum.[4] The best achieved energy resolutions with this sensor architecture range from 375 meV (TESSERACT Collaboration, 2024) to 700 meV (Romani et al., 2024), corresponding to $O(\text{eV})$ energy thresholds.

These detectors, the aforementioned $O(10\,\text{eV})$ threshold detectors, and many other low-threshold phonon-mediated detectors in rare event searches have uncovered a new and dominant phonon-only background at these low energies (EXCESS workshop, 2022). This background has been referred to as the low energy excess (LEE) or more specifically the zero-charge low energy excess (0QLEE).[5] We describe in more detail and explore tactics to circumvent this background in

[4] The diffusion length $\ell = \sqrt{D\tau_{\text{qp}}}$, where D is the diffusion coefficient and τ_{qp} is the quasiparticle lifetime.

[5] Low-energy ionizing backgrounds are also discussed in EXCESS workshop (2022).

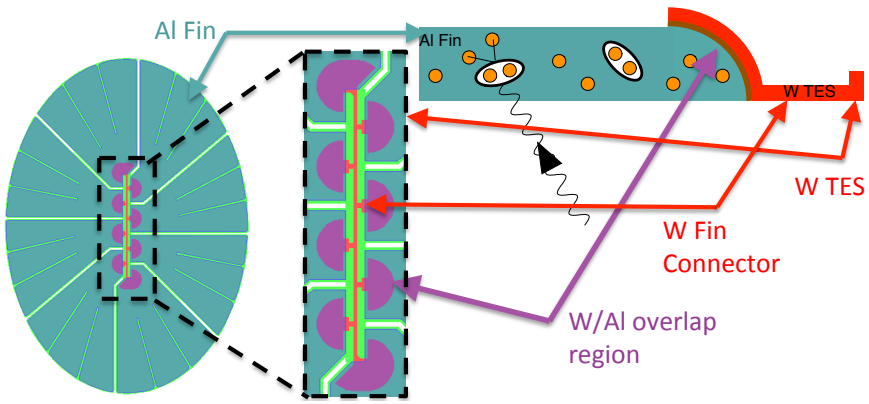

Figure 3.6: **SuperCDMS phonon sensor design.** The entire sensor is referred to as a quasiparticle-trapping-assisted electrothermal-feedback transition edge sensor (QET). The TES in red is the sensitive element and is made from tungsten. The collection fins in teal are responsible for phonon collection and are made from aluminum. Reproduced from SuperCDMS SNOLAB TDR (2018).

Chapters 4 and 5.

New phonon sensor architectures are being explored to achieve sub-eV energy thresholds. Kinetic inductance detectors (KIDs) are an attractive option; they are directly sensitive to quasiparticle density in the superconductor and so could provide long-term benefits as a unified phonon absorber and sensitive element. This idea is explored in great detail in Chapters 6, 7, and 8. Two other alternatives have been proposed in the last year that involve quasiparticle counting with superconducting qubits: superconducting quasiparticle-amplifying transmons (SQUATs) (Fink et al., 2024) and quantum parity detectors (QPDs) (Ramanathan, Parker, et al., 2024).

The transition from free nuclear recoils to single phonons

At low momentum transfer q and low recoil energies ω (previously labeled ΔE_{NR}; note that ω is not a frequency), a nuclear interaction transitions from the realm of free nuclear recoils to the quantized excitation of phonons. We work through some of the main results in Trickle et al. (2020). The article details the rate calculation for both free nuclear recoils and single phonon excitations, provides intuition for what regimes each calculation holds, and also builds a theoretical framework for computing dark matter scattering rates across a wide variety of excitations.

Rate calculation

Trickle et al. (2020) writes the spin-independent differential rate (per target mass) energy spectrum as

$$\frac{dR}{d\omega} = \frac{1}{\rho_T}\frac{\rho_\chi}{m_\chi}\frac{\pi\bar{\sigma}}{\mu^2}\int d^3v f_\chi(\vec{v})\int \frac{d^3q}{(2\pi)^3}\mathcal{F}^2_{\mathrm{med}}(q)S(\vec{q},\omega_{\vec{q}})\delta(\omega-\omega_{\vec{q}}). \tag{3.17}$$

Notes and observations:

- $\rho_T = m_T/V$ is the density of a detector with volume V, and ρ_χ is the local density of dark matter, usually set to $0.3\,\mathrm{GeV/cm^3}$. $V\rho_\chi/m_\chi$ is the number of dark matter particles present in the detector.
- $\int d^3v f_\chi(\vec{v})$ is the integral over the dark matter's velocity distribution, generally assumed to be a Maxwell-Boltzmann distribution, shifted by the Earth's velocity and truncated by the dark matter escape velocity.

- $(\pi\bar{\sigma}/\mu^2)\mathcal{F}_{\rm med}^2(q) = \overline{|\mathcal{M}(q)|^2}$ is the vacuum matrix element for 2→2 scattering. $\bar{\sigma}$ is the scattering cross section that may be constrained in a direct detection experiment.[6] μ is the reduced mass of dark matter with whatever the target is. $\mathcal{F}_{\rm med}$ is the q dependence of $\mathcal{M}$, which comes from the mediator propagator for tree-level scattering:

$$\mathcal{F}_{\rm med}(q) = \begin{cases} 1 & \text{heavy mediator} \\ (q_0/q)^2 & \text{light mediator,} \end{cases} \tag{3.18}$$

  where q_0 is a reference momentum and is equal to $m_\chi v_0$ for nucleon scattering and αm_e for electron scattering. In essence, $\mathcal{M}(q)$ characterizes the strength of the fundamental particle physics coupling of dark matter with the Standard Model.
- $S(\vec{q},\omega)$ is the dynamic structure factor and captures the target's response to an energy deposition ω and a momentum transfer $\vec{q}$. It is given by

$$S(\vec{q},\omega) \equiv \frac{1}{V}\sum_f |\langle f|\mathcal{F}_T(\vec{q})|i\rangle|^2 2\pi\delta(E_f - E_i - \omega), \tag{3.19}$$

  where the dark matter triggers a transition $|i\rangle \to |f\rangle$ from some initial state with energy E_i to a final state with energy E_f. $\mathcal{F}_T(\vec{q})$ is a target form factor composed of weighted contributions from protons p, neutrons n, and electrons e:

$$\mathcal{F}_T(\vec{q}) \propto f_p(\vec{q})\tilde{n}_p(-\vec{q}) + f_n(\vec{q})\tilde{n}_n(-\vec{q}) + f_e(\vec{q})\tilde{n}_e(-\vec{q}), \tag{3.20}$$

  where $f_{p,n,e}$ are the effective in-medium form factors and $\tilde{n}_{p,n,e}(-\vec{q})$ are the Fourier transforms of the number densities for the different components. $\mathcal{F}_T(\vec{q})$ may have a factor of $1/f_n$ or $1/f_e$ if the cross section to the neutron or the electron is used as the reference. Notably, the response of the target has been factorized from the fundamental particle physics coupling given by $\overline{|\mathcal{M}(q)|^2}$.
- The transfer of energy is constrained by the momentum transfer such that

$$\omega_{\vec{q}} = \frac{1}{2}m_\chi v^2 - \frac{(m_\chi\vec{v} - \vec{q})^2}{2m_\chi} = \vec{q}\cdot\vec{v} - \frac{q^2}{2m_\chi}. \tag{3.21}$$

  The delta function $\delta(\omega - \omega_{\vec{q}})$ selects the values of ω that are on this curve. S is evaluated at $\omega_{\vec{q}}$.

Dynamic structure factor

The key difference in the rate calculation between nuclear recoil production and single-phonon production lies in the dynamic structure factor $S(\vec{q},\omega)$. We reproduce and slightly modify the dynamic structure factors for these two production mechanisms as written in Equation 59 and Equation 90 from Trickle et al. (2020):

$$\text{Nuclear recoils:} \quad S(\vec{q},\omega) = 2\pi\frac{\rho_T}{m_N}\frac{f_N^2}{f_n^2}F_N^2(\vec{q})\delta\left(\omega - \frac{q^2}{2m_N}\right) \tag{3.22}$$

$$\text{Single phonons:} \quad S(\vec{q},\omega) = \frac{\pi}{\Omega}e^{-2W(\vec{q})}\frac{f_N^2}{f_n^2}F_N^2(\vec{q})\sum_\nu \frac{\left|\vec{q}\cdot\vec{\epsilon}^*_{\nu,\vec{k}}\right|^2}{m_N\omega_{\nu,\vec{k}}}\delta(\omega - \omega_{\nu,\vec{k}}). \tag{3.23}$$

The single phonon expression has been simplified from Trickle et al. (2020) to the case of a single atom in the primitive cell.

There is a noteworthy correspondence between each of the terms in Equation 3.22 and Equation 3.23:

[6] Generally, the constrained cross section is either with the nucleon or the electron; in the latter case an additional factor $\sigma_{\rm nucleus} = A^2\sigma_{\rm nucleon}$ is needed to take into account coherence effects.

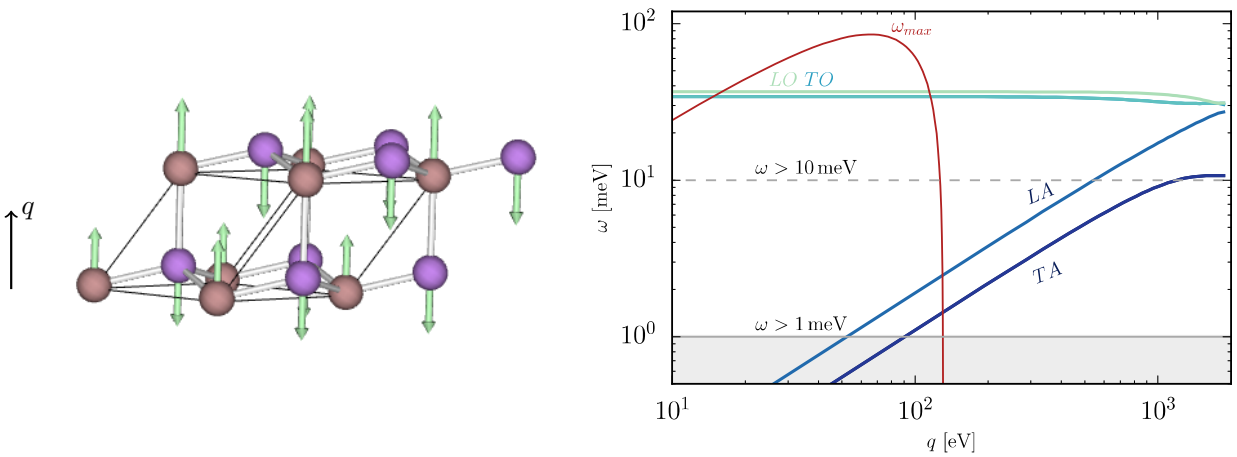

Figure 3.7: **Single-phonon dispersion relations.** Left: a visualization of longitudinal optical phonons in a polar material such as GaAs. Reproduced from Griffin, Knapen, et al. (2018). Right: phonon modes in GaAs, truncated at $q_{\rm BZ} \approx 2\,\text{keV}$. The phase space for a dark matter particle with mass $m_\chi = 25\,\text{keV}$ is indicated in red. Phonon thresholds at 10 meV and 1 meV are also shown. Reproduced from Knapen, Lin, Pyle, et al. (2018).

- $\rho_T/m_N \longleftrightarrow 1/\Omega$. Ω is the volume of the primitive cell.
- $(f_N^2/f_n^2)F_N^2(\vec{q}) \longleftrightarrow (f_N^2/f_n^2)F_N^2(\vec{q})$. This expression, identical for the two cases, is a reorganization of the $f_{p,n,e}$ terms for $\mathcal{F}(T)$ in Equation 3.20. $\mathcal{F}_T$ has been expressed in terms of a weighted average nucleon form factor $f_N \equiv f_p Z + f_n(A-Z)$ and a nucleus form factor $F_N(\vec{q})$ that is given by the Helm form factor (Helm, 1956).
- $1 = q^2/2m_N\omega \longleftrightarrow \sum_\nu |\vec{q}\cdot\vec{\epsilon}^*_{\nu,\vec{k}}|^2/m_N\omega_{\nu,\vec{k}}$. The former does not appear in Equation 3.22 but may be inserted because it is equal to unity. The equality is enforced by the δ function. The latter is a sum over the phonon branches ν, which are populated with phonon polarization vectors $\vec{\epsilon}_{\nu,\vec{k}}$ and phonon energies $\omega_{\nu,\vec{k}}$. The phonon momentum $\vec{k}$ satisfies $\vec{q} = \vec{k} + \vec{G}$, where $\vec{G}$ is the reciprocal lattice vector in the first Brillouin zone. The correspondence of note here is that both terms take the form $q^2/m\omega$.
- $\delta(\omega - q^2/2m_N) \longleftrightarrow \delta(\omega - \omega_{\nu,\vec{k}})$. The δ functions enforce conservation of energy. In the case of nuclear recoils, the nucleus receives energy $q^2/2m_N$; in the case of a single phonon production, the phonon has energy $\omega_{\nu,\vec{k}}$, indexed by momentum and energy.

More details on the rate calculation for free nuclear recoils and single phonons can be found in Appendix A.

The main difference between Equation 3.22 and Equation 3.23 is the term $e^{-2W(\vec{q})}$, which is known as the Debye-Waller factor. It arises from phonon annihilation-creation commutation relations. As reported in Trickle et al. (2020), $W(\vec{q}) \sim q^2/4m_N\omega$. If $q \gg \sqrt{2m_N\omega_{\rm ph}} \approx 100\,\text{keV}$, where $\omega_{\rm ph} \approx 10$–$100\,\text{meV}$ is the energy of a single optical phonon, then the direct production of phonons is greatly suppressed, and the free nuclear recoil rate is higher. Furthermore, if $q \gg \sqrt{2m_N\omega_{\rm ph}}$, then $\omega = q^2/2m_N \gg \omega_{\rm ph}$, i.e., large momentum transfers correspond to large energy depositions that are well in excess of typical single-phonon energies. Trickle et al. (2020) makes a heuristic argument on the topic of energy regimes using the uncertainty principle: hard scattering occurs on the timescale of $1/\omega$ whereas atomic vibrations occur on the timescale of $1/\omega_{\rm ph}$. If $1/\omega \ll 1/\omega_{\rm ph}$, i.e., $\omega \gg \omega_{\rm ph}$, then the interaction is well approximated as an instantaneous nuclear recoil.

For determination of the actual phonon energies that may be produced from a dark matter

interaction, we must relate the phonon energy with the transfer of momentum, i.e., $\omega_{\nu,\vec{k}}$ as it depends on $\vec{q}$, also known as the dispersion relation. Different phonon branches have different dispersion relations. Acoustic phonons follow a linear dispersion: $\omega \approx c_s q$, where c_s is a sound speed. In contrast, for optical phonons, $\omega \approx$ constant, which means that even low momentum transfers can produce an optical phonon so long as there is sufficient energy transfer. Optical phonons arise when the atoms in the primitive cell of a crystal oscillate out of phase, as shown in Figure 3.7 Left. When $\vec{q} = 0$, an optical phonon can be thought of as a standing, non-propagating wave which stores a finite amount of energy (Knapen, Lin, Pyle, et al., 2018). Examples of the longitudinal and transverse modes of the optical and acoustic branches are drawn in Figure 3.7. The available phase space of a dark matter interaction is also drawn, given by $qv - q^2/2m_\chi$. The intersections of the dark matter phase curve with the phonon dispersion relations are the loci at which energy and momentum are conserved, i.e., the interaction is allowed. As a result of the constant dispersion, low-mass dark matter is more easily detectable through the optical phonon channel than through the acoustic phonon channel. Importantly, the calculation presented in Figure 3.7 Right is confined to the region where (1) the momentum transfer is much less than the crystal momentum, i.e., $q \ll q_{\mathrm{BZ}} = 2\pi/a \approx 2\,\mathrm{keV}$, where q_{BZ} is roughly the size of the first Brillouin zone and is given by the lattice spacing a and (2) the energy transfer is at the single-phonon level (10–100 meV).

Multi-phonon calculations

We now look to bridge the gap between the single-phonon regime where $q < q_{\mathrm{BZ}} \approx 2\,\mathrm{keV}$ and the free nuclear recoil regime where $q \gg \sqrt{2m_N\omega_{\mathrm{ph}}} \approx 100\,\mathrm{keV}$. We follow Campbell-Deem et al. (2022) in this section, which connects between the two regimes by performing multi-phonon-production calculations in the crossover energy range. This calculation is important for the somewhat narrow range of 1 MeV to 30 MeV dark matter. It should be noted that the calculation in Campbell-Deem et al. (2022) only computes rates for the instantaneous production of n phonons from a dark matter scattering and does not incorporate anharmonic contributions, so the calculations should be seen only as an approximation. As before, the calculation is centered on the behavior of the dynamic structure factor $S(\vec{q},\omega)$.

In Campbell-Deem et al. (2022), they take the approximation that the n phonon modes are not coupled to each other and thus are incoherently activated; this approximation is referred to as the n-phonon incoherent approximation. With this approximation, the dynamic structure factor contains the following expression:

$$S(\vec{q},\omega) \sim \sum_n \left(\frac{q^2}{2m_N}\right)^n \frac{1}{n!} \prod_i^n \int d\omega_i \frac{D(\omega_i)}{\omega_i}, \tag{3.24}$$

where $D(\omega_i)$ is a density of states and serves to weight the various contributions of $1/\omega_i$. The scattering dark matter particle would couple to and excite each of these n phonons simultaneously. The expression in Equation 3.24 is analogous to the $q^2/2m\omega$ expression from Equation 3.23. The structure of $S(\vec{q},\omega)$ for GaAs in this regime, calculated from about 2 keV to 100 keV, is very clearly peaked in Figure 3.8 Right at specific phonon energies, and the extension of the longitudinal optical phonon mode in this regime above q_{BZ} is also clear. The multi-phonon production of longitudinal optical phonon modes is visible as the brighter horizontal bands at integer multiples of the ω_{LO} band.

When $q \gtrapprox 100\,\mathrm{keV}$, the calculation enters a regime where the multi-phonon production is well characterized as a Gaussian envelope around the free-nuclear-recoil regime. The dynamic structure factor can be calculated with the "impulse approximation." Campbell-Deem et al. (2022) asserts that this regime is entered when $q > 2\sqrt{2m_d\bar{\omega}_d}$, where m_d and $\bar{\omega}_d$ are the mass and average phonon

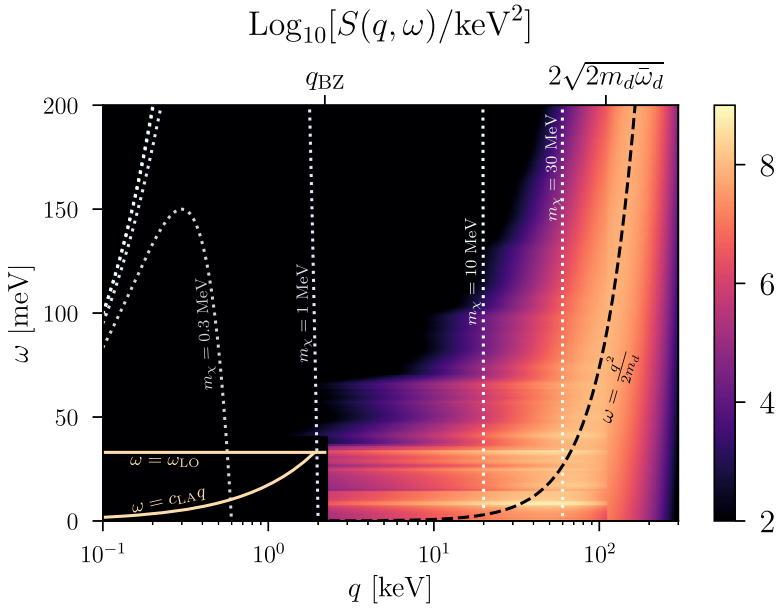

Figure 3.8: **Multi-phonon dynamic structure factor.** $S(\vec{q},\omega)$ for GaAs follows the free-nuclear-recoil approximation when $q > 2\sqrt{2m_d\bar{\omega}_d}$. Below that, the multi-phonon production calculation is performed, leading to a peaked dynamic structure factor corresponding to the integral production of various phonon channels. Below q_{BZ}, the single phonon dynamic structure factor is shown. The dark matter phase space and free-nuclear-recoil approximation are shown as the dotted and dashed curves.

frequency at some lattice site d, very similar to the $q \gg \sqrt{2m_N\omega_{\mathrm{ph}}}$ constraint from the Debye-Waller factor. This particular boundary for the impulse approximation is derived from the fact that the multi-phonon expansion of the dynamic structure factor includes a factor of $(q^2/2m_d\bar{\omega}_d)^n/n!$ for n phonon production (Equation 3.24). The higher n terms are more dominant in the expansion when q is large, or specifically when $q^2/2m_d\bar{\omega}_d \sim n$. $S(\vec{q},\omega)$ versus ω is more Gaussian-shaped for higher phonon modes, which are more dominant for larger q. m_d is generally around 30 GeV for semiconductors of interest, and $\bar{\omega}_d$ can be estimated from the single-phonon energy scale, which is around 30 meV. The free-nuclear-recoil calculation therefore holds when $q \gtrapprox 2\sqrt{2m_d\bar{\omega}_d} \approx 100\,\mathrm{keV}$. It is interesting that this momentum scale is given by the geometric mean of the nucleus mass and the phonon frequency. From Figure 3.8, the $\omega = q^2/2m_d$ curve crosses $q = 100\,\mathrm{keV}$ around $\omega = 80$,meV.

Also shown in Figure 3.8 is the available phase space of energy deposition and momentum transfer for various dark matter masses. The curves are given by $\omega = qv - q^2/2m_\chi$. Figure 3.8 shows that the multi-phonon calculations are important in the regime between 1 MeV and 30 MeV and that the single-phonon calculations are important for the regime below 1 MeV.

Phonon-mediated detection with polar materials

Polar materials are an interesting and novel platform for next-generation phonon-mediated detectors and were initially proposed for low-mass dark matter searches in Knapen, Lin, Pyle, et al. (2018) and Griffin, Knapen, et al. (2018). In a polar material, the unit cell is composed of a positive and negative ion. A free charge in the material would distort the lattice and do so in opposite directions for the differently charged ions, consequently coupling to the longitudinal optical modes of the material (optical phonons correspond to two adjacent atoms in a unit cell oscillating in opposite

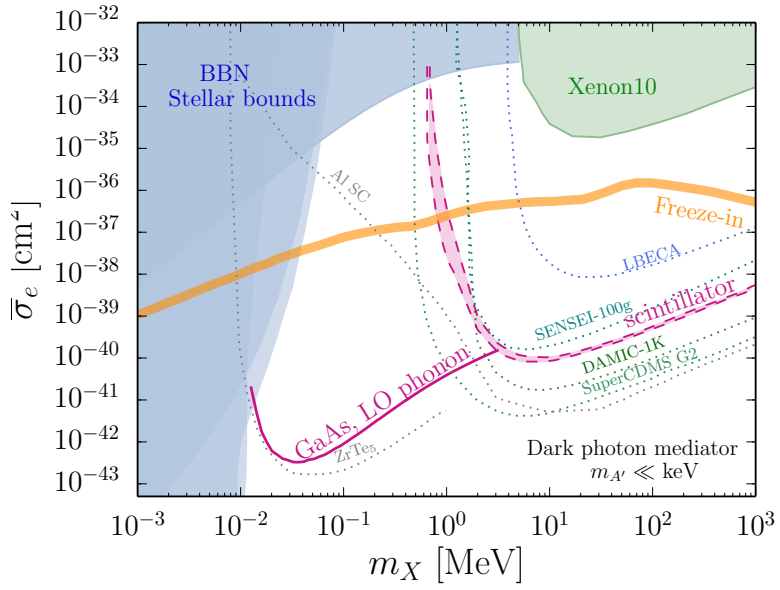

Figure 3.9: **Expected reach of a GaAs phonon-mediated detector to electron-recoiling dark matter with a light mediator.** Projected exclusion curves assume 3 events for a kg·yr exposure. Existing limits, charge-channel sensitivities, and other next-generation detector architecture sensitivities are shown. Reproduced from Knapen, Lin, Pyle, et al. (2018).

directions; see Figure 3.7 Left). This coupling between charge and optical phonons is governed by the Frölich Hamiltonian.

In the case of a dark photon A' that mediates interactions with the Standard Model via some kinetic mixing ϵ, the dark matter current is coupled to the Standard Model current. To show this, Knapen, Lin, Pyle, et al. (2018) starts from the Lagrangian for a hidden U$'$(1) symmetry that has some kinetic mixing with the Standard Model:

$$\mathcal{L} \supset \bar{\chi}(\not{\partial} - ig_\chi \not{A}')\chi - \frac{1}{4}F'^{\mu\nu}F'_{\mu\nu} - \frac{\epsilon}{2}F^{\mu\nu}F'_{\mu\nu} - \frac{m_{A'}^2}{2}A'^{\mu}A'_{\mu}. \tag{3.25}$$

If A' is massive, a diagonalization to the mass basis can be performed, and the electron has a charge κe under the dark photon. If the mediator is massless, the final term disappears and we can perform a field redefinition $A' \to A' - \epsilon A$. The result is

$$\mathcal{L} \supset \bar{\chi}(\not{\partial} - ig_\chi \not{A}' - i\epsilon g_\chi \not{A})\chi - \frac{1}{4}F'^{\mu\nu}F'_{\mu\nu}. \tag{3.26}$$

The final term in the parentheses shows that the dark matter particle χ is milli-charged under the Standard Model photon A, with charge $e' = \epsilon g_\chi$. It is of course also charged under the dark photon with coupling g_χ. In both the massive A' and massless A' cases, the Standard Model current is coupled to the dark matter current with strength $\epsilon g_\chi e$.

Since dark matter is milli-charged in the massless mediator model, we can apply the Frölich Hamilitonian directly: in a polar material, milli-charged dark matter would source a small electric field and induce oscillations of the ions, which are just optical phonon modes. The only modification to the Hamiltonian is a re-scaling by the charge ratio of dark matter to the electron: $\epsilon g_\chi / e$. Single optical phonons are a particularly attractive low energy signal in phonon-mediated detectors because of their dispersion relation $\omega \sim$ constant; the full kinetic energy of a dark matter particle is accessible to the detector even when the momentum transfer is arbitrarily small.

The essential advantage of phonon-mediated polar material detectors is that optical phonons can be created by electron-coupled DM below the mass range that can create electron-hole pairs. This advantage is shown in Figure 3.9; the reach of phonon-mediated polar material detectors is two orders of magnitude lower in dark matter mass than the charge- and scintillation- mediated detectors, providing unparalleled sensitivity to low-mass electron-recoiling dark matter. More precisely, the optical phonon signal can be used to place constraints on $\bar{\sigma}_e$, the scattering interaction cross section between dark matter and the electron. In practice, a constraint on e' is placed using phonon scattering rates, and that constraint can then be transformed into a constraint on $\bar{\sigma}_e$:

$$\bar{\sigma}_e = \frac{4e'^2 \alpha \mu_{\chi e}^2}{(\alpha m_e)^4}. \tag{3.27}$$

Other interesting opportunities for next-generation detector architectures are shown in Figure 3.9, such as low-bandgap Dirac materials (the $ZrTe_5$ curve), as well as superconducting targets (aluminum is shown as an example).

As a side note, polar materials can also place constraints on the parameter space of sub-MeV dark matter mediated by a heavy dark photon, but this mass region for heavy mediators is excluded by Big Bang nucleosynthesis (BBN), specifically constrained by the abundance of hydrogen, deuterium, and helium in the universe (Knapen, Lin, and Zurek, 2017). Current data is consistent with a Standard Model photon and three neutrino species in equilibrium at the time of BBN. Zurek (2024) points out that, for a heavy mediator, the couplings between dark matter and the Standard Model are strong enough for dark matter to be in equilibrium with the primordial plasma at the time of BBN and would thus be constrained by the relativistic degrees of freedom that are presently inferred from the data. Thus, there is no unconstrained sub-MeV heavy-dark-photon-mediated parameter space for polar materials to probe.

Part II

SuperCDMS

Preface

In Part II of this thesis, I describe my main contributions to the SuperCDMS Collaboration. There are two chapters. Chapter 4 is a reproduction of some key results from SuperCDMS Collaboration (2023), which presents a long-term strategy for the SuperCDMS Collaboration to search for low-mass particle dark matter using the SuperCDMS SNOLAB infrastructure. Two core elements of this strategy are emphasized in Chapter 4 and throughout the rest of the thesis: high-voltage (HV), single-charge-sensitive detector operation in Chapter 5, and zero-voltage (0V), low-threshold detector operation in Part III, which encompasses Chapters 6, 7, & 8.

I was one of three members of the SuperCDMS Long-Term Planning Task Force who operated, upgraded, and interpreted results from `SensitivityEstimate`, the body of code that was used to build the long-term strategy described in SuperCDMS Collaboration (2023). Significant upgrades were made to the code to best inform our strategic planning. My contributions to these upgrades include but are not limited to the following: implementing a performance-driven model for forecasting future detector energy and timing resolution, tactically modifying analysis regions of interest to circumvent HV detector leakage, and enabling of the code to compute two-sided allowed regions for production of discovery potentials.

In Chapter 5, I describe the results of a nuclear-recoiling dark matter search performed on data from an experiment known as HVeV Run 4. This search includes characterization and subsequent subtraction of a zero-charge low energy excess. I was part of a team of analyzers who investigated the HVeV Run 4 data for many different purposes; various analysis techniques and protocols were shared among this team and are pointed out in Chapter 5 when originating from another analyzer or team of analyzers. Outside of these shared techniques and protocols, I was the sole analyzer for the dark matter search described in Chapter 5.

Chapter 4

SuperCDMS SNOLAB and beyond

4.1 The basics of SuperCDMS SNOLAB

SuperCDMS SNOLAB is an under construction experiment to search for nuclear-recoiling and electron-recoiling dark matter with cryogenically operated silicon and germanium detectors. There are two detector architectures that will be used: interleaved Z-sensitive Ionization- and Phonon-mediated detectors (iZIPs) and High Voltage (HV) detectors. There are thus four different types of dark matter searches that will be performed by SuperCDMS SNOLAB, identified by detector material (Si or Ge) and detector architecture (iZIP or HV): Si iZIP, Si HV, Ge iZIP, and Ge HV.

Essential detector characteristics

When a nucleus or electron recoils within these semiconductor substrates, the recoil energy E_r immediately goes towards phonon energy or the production of free electron-hole pairs. For the case of an electron recoil, all of the initial energy of the electron recoil goes toward the production of electron-hole pairs. For nuclear recoils in the energy regime of interest (below 100 keV), most of the initial energy goes into phonon production, but a significant fraction still goes immediately into exciting charge above the bandgap (1.11 eV in silicon, 0.67 eV in germanium) and producing one or many free electron-hole pairs. The electron-hole pairs that are produced from electron and nuclear recoils quickly fall back to the bottom of the conduction band by shedding phonons, and when all the electron-hole pairs finally recombine and the electron re-enters the valence band, phonons are again produced with energy equal to the bandgap. The net result is that the final total phonon energy E_{P_t} is very close to the initial recoil energy E_r.

A fundamental aspect of the current SuperCDMS detector architectures is that a bias voltage V_b is applied to the detector. The impact of placing a bias voltage on such a detector is twofold: (1) the electron-hole pairs move to the surface of the detector where a charge sensor may be placed to sense the arriving charge; and (2) the phonon energy is boosted by an amount given by the work performed by the electric field upon the electron-hole pairs. This second phenomenon is known as the Neganov-Trofimov-Luke (NTL) effect (see Neganov et al., 1985 and Luke, 1988), and the phonons resulting from this effect are referred to as NTL phonons and have energy E_{NTL}. The total phonon energy E_{P_t} for a given event becomes

$$\begin{aligned} E_{P_t} &= E_r + E_{\mathrm{NTL}} \\ &= E_r + N_{\mathrm{eh}} e V_b \\ &= E_r + \frac{Y(E_r) E_r}{\epsilon_{\mathrm{eh}}} e V_b, \end{aligned} \tag{4.1}$$

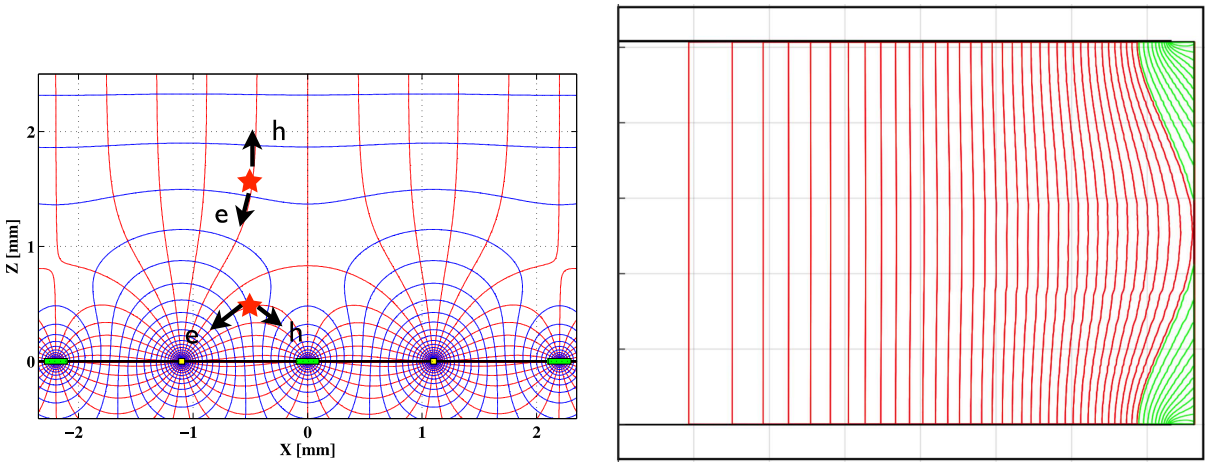

Figure 4.1: **E-field configurations of SuperCDMS detectors** Left: E-field configuration for iZIPs in red and equipotential curves in blue. The electrodes shown in yellow are held at a positive bias while the electrodes shown in green are grounded; they are interleaved with each other. Two example paths of electron-hole pairs are suggestively depicted with the stars and arrows. Right: E-field configuration for HV detectors with lines that terminate at the top and bottom shown in red and lines that terminate at the edge in green. Figures reproduced from Golwala (2016).

where N_{eh} is the number of electron recoils that are produced in an event and is equal to $Y(E_r)E_r/\epsilon_{\mathrm{eh}}$. $Y(E_r)$ is the yield function that represents the fraction of the total initial recoil energy that eventually goes toward charge production, and ϵ_{eh} is the average amount of energy required to produce an electron hole pair. It is evident how the total phonon energy E_{P_t} has been boosted beyond the initial recoil energy E_r by an amount E_{NTL}.

iZIPs use the first mentioned effect of the bias voltage, point (1) from the previous paragraph: a voltage brings charge to the surface of the detector. Via direct measurement of charge, iZIPs can be used to discriminate between nuclear and electron recoils, as well as between surface and bulk events. iZIPs have both phonon and charge sensors on their surface. The applied electric field ($\pm 3\,\mathrm{V}$ in Ge, $\pm 4\,\mathrm{V}$ in Si) moves the charge that is produced by a particle event to the charge sensors on the surface. Nuclear and electron recoils produce different amounts of charge for equal amounts of produced phonon energy; $Y(100\,\mathrm{keV}) < 0.5$ for nuclear recoils, whereas $Y = 1$ for all electron recoils. This discrepancy in $Y(E_r)$ enables discrimination between nuclear and electron recoils on an event-by-event basis. The low-energy bound of this discrimination is given by the charge resolution of the charge sensors.

In contrast, HV detectors use point (2) from the aforementioned effects: a bias voltage boosts the total phonon energy to achieve a lower recoil energy threshold. They are operated at $V_{\mathrm{b}} = 100\,\mathrm{V}$, much higher than the bias voltage applied to iZIPs, and are only instrumented with phonon sensors. The HV phonon sensors collect primarily the NTL phonon energy, which is now much larger than the recoil energy because $Y(E_r)eV_b/\epsilon_{\mathrm{eh}} \gg 1\,eV_b/\epsilon_{\mathrm{eh}} \gg 1$. The relatively elevated bias voltage boosts the energy of recoil events that may be well below the phonon energy threshold. Given a charge-producing particle event and a fixed energy threshold, a higher voltage detector is sensitive to lower energy recoils. With good enough phonon energy resolution and large enough bias voltage, charge-producing events may become discretely resolved in phonon energy, where discretized peaks in the phonon spectrum correspond to integer quantities of electron-hole pairs. The remarkable feature of such an architecture is that events that create even a single electron-hole pair in the substrate can be identified using only phonon sensors.

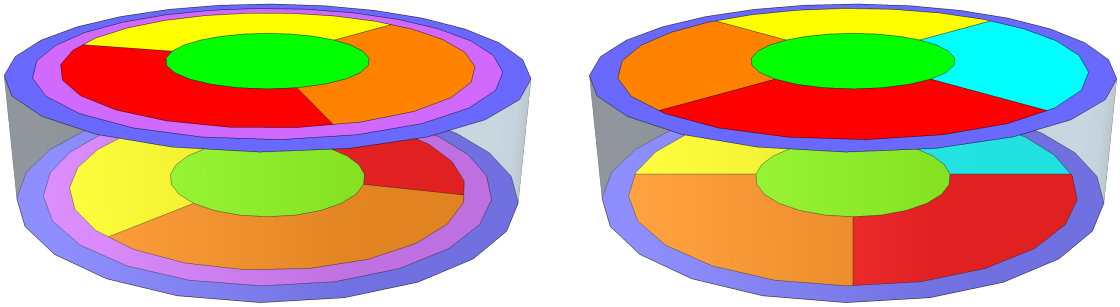

Figure 4.2: **Phonon sensor configurations of SuperCDMS detectors.** Left: HV detector phonon sensor configuration, optimized for sensitivity to the narrow NTL phonon column. Right: iZIP phonon sensor configuration, optimized for sensitivity to the homoegenous initial recoil energy of phonons.

Detector position sensitivity

Both iZIPs and HV detectors utilize segmented phonon sensors, six on each face, to recover event position information. Some of the sensors are located at the outer rim of the detectors, allowing for discrimination of surface events from bulk events with a cut in event radius. High-r events may also suffer suppressed charge collection due to the electric field lines curving outward near the edges, as is depicted in Figure 4.1 Right (the bowing of field lines near the outer edges is also true for iZIP detectors but is not shown). Dark matter events are expected to occur uniformly across the bulk of the detector and so can be discriminated from the subset of backgrounds that predominantly interact near the surfaces.

A z-direction fiducial volume cut is also possible with both detectors, but the iZIP phonon sensors are much more sensitive to z-position. The reason for this fact is twofold. First, since the iZIP is operated at a lower voltage than an HV detector, a fractionally higher component of the total phonon energy belongs to the initial production of phonons. For the isotropically distributed initial phonons, z-position sensitivity is given by the $1/z^2$ scaling of the solid angle that a certain phonon sensor subtends with respect to an event in the bulk. Second, the interleaved nature of the bias electrodes in an iZIP causes the NTL phonons for a surface event to be generated primarily near the surface, as is evident in Figure 4.1 Left. In contrast, the HV detector schematically depicted in Figure 4.1 Right has its NTL phonons generated uniformly over the charge's drift path without much dependence on the z-position (only via what fraction of NTL phonons is generated by electrons and by holes).

iZIPs are additionally equipped with charge-based position sensitivity in both the r- and z-dimensions. The interleaved nature of the iZIP charge sensors enables the discrimination of surface events via charge collection asymmetry; see Figure 4.1 Left. Furthermore, since at least the era of CDMS I BLIP detectors, CDMS charge-mediated detectors have had an outer annular charge electrode to enable rejection of high radius events (Golwala, 2000).

Experimental layout

The SuperCDMS SNOLAB detectors are organized into four detector towers, two for each detector architecture, iZIP or HV. An example of a detector tower is shown in Figure 4.3. Each tower can hold up to six cylindrical detectors. Each detector is 10 cm in diameter and 3.3 cm in height, which corresponds to 1400 g of germanium and 610 g of silicon for each detector. The towers will be housed in what is called the SNOBOX, which is a set of nested cryogenically cooled cans, also

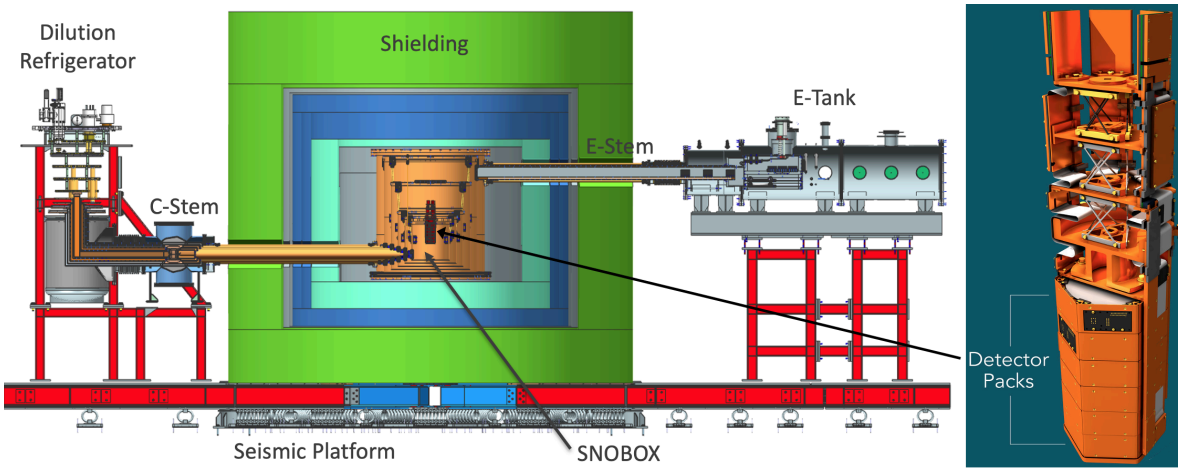

Figure 4.3: **The SuperCDMS SNOLAB experiment.** Left: SuperCDMS SNOLAB. The SNOBOX will house the detector towers in a set of nested, cryogenically cooled cans. The shielding attenuates external particle backgrounds. The dilution refrigerator is connected to the rest of the experiment via the C-stem. Readout cables are housed in the E-tank, which receives additional cooling. Right: a detector tower, with the detectors indicated at the bottom.

visible in Figure 4.3. The SNOBOX resides in three layers of external shielding in order to attenuate external particle backgrounds. The innermost and outermost shields, depicted in Figure 4.3 in cyan and green, attenuate the neutron background. The middle shield in blue attenuates gamma rays. In the region corresponding to the gray, a thin radon barrier maintains a nitrogen atmosphere within the interstices of the inner shield, and a μ-metal shield provides adequate rejection of the Earth's magnetic field for detector operation.

The SNOBOX can accommodate seven towers, but only four will be installed for the initial run of SuperCDMS SNOLAB. To provide some insight about why this number of detector towers and the consequent exposure listed in Table 4.1 is sufficient given the backgrounds that are expected, we examine the background levels shown in Figure 4.6 Bottom and Figure 4.5 Bottom. A dark matter search is exposure-limited if it observes zero events over the duration of its exposure. With zero events, a 90% confidence level (CL) upper limit on the dark matter cross section may be set at the cross section that would produce 2.3 events over the duration of the exposure; at this cross section, 90% of experiments would have observed more than 0 events. The search becomes background-limited when the expected number of background events exceeds 2.3 events.

For iZIP detectors, the background levels range from 10^{-5} dru above 5 keV phonon energy to 10^{-1} dru below 1 keV phonon energy. At low energies, only a few tens of kg·day exposures are needed to exceed 2.3 expected background events. At high energies, say up to 100 keV, a few thousand kg·day exposures are needed to become background-limited. The 45 kg·yr exposure for Ge iZIPs reported in Table 4.1 is thus sufficient for a search to be background-limited up to 100 keV. HV background levels range from 10^{-3} dru to $10^{-1.5}$ dru below 1 keV depending on detector material. These background levels would require 100 kg·day to a few 1000s of kg·day of exposure. The 7.8 kg·yr and 36 kg·yr exposures

	iZIP		HV	
	Ge	Si	Ge	Si
Number of detectors	10	2	8	4
Total exposure [kg·yr]	45	3.9	36	7.8
Phonon resolution [eV]	33	19	34	13
Ionization resolution [eV$_{ee}$]	160	180	–	–
Voltage Bias ($V_+ - V_-$) [V]	6	8	100	100

Table 4.1: **Anticipated exposures and detector parameters for SuperCDMS SNOLAB.**

listed in Table 4.1 are once again sufficient given the expected background levels. Germanium HV detectors are expected to have lower background levels, which is why the germanium detectors are set to have more exposure.

General outline for the rest of the chapter

In Section 4.2, Section 4.3, and Section 4.4, we discuss the projected sensitivity of SuperCDMS SNOLAB, which is planned to probe unexplored parameter space of both nuclear-recoiling dark matter and electron-coupled dark matter. In the SuperCDMS Snowmass 2021 contribution, three different classes of electron-coupled dark matter are explored: low-mass sub-GeV electron-recoiling dark matter, dark photon dark matter, and axion-like particles. Low-mass sub-GeV electron-recoiling dark matter produces scattering events whereas the latter two types of dark matter candidates produce absorption events. In this thesis, we focus on nuclear-recoiling dark matter (NRDM) and low-mass sub-GeV electron-recoiling dark matter (ERDM).

The organization of the experiment into detector towers provides modularity and flexibility for future experiments to deploy different detector types. In Section 4.5, Section 4.6, and Section 4.7, we discuss future possibilities for SuperCDMS SNOLAB and how future detector towers composed of next-generation detector architectures can leverage the existing infrastructure to probe parameter space that is not accessible to SuperCDMS SNOLAB. We introduce the 0V detector architecture as a means to probe low-mass dark matter. The SuperCDMS Snowmass 2021 contribution includes forecasts for iZIP-, piZIP-, HV-, and 0V-style detectors. This thesis focuses only on HV- and 0V-style detectors.

4.2 SuperCDMS SNOLAB sensitivity projections

Snowmass 2021 was a grassroots study to plan for US particle physics ventures in the decade from 2025 to 2035 with an eye toward 2035 to 2045 (APS DPF, 2021). The SuperCDMS contribution to the Snowmass 2021 process is presented in SuperCDMS Collaboration (2023). The first main component of this contribution is an update to the sensitivity projections for the SuperCDMS SNOLAB experiment; the previous set of sensitivity projections were presented in SuperCDMS Collaboration (2017). The main modifications relative to these previous projections are (1) a discretized treatment of charge production for HV detectors and (2) an updated understanding of backgrounds and detector resolution performance.

The assumptions about background levels and detector performance in SuperCDMS Collaboration (2023) are informed by literature, simulation, and preliminary data from the SuperCDMS SNOLAB detectors as well as other prototype SuperCDMS detectors. The four limits, using both the optimum interval (OI) and profile likelihood ratio (PLR) techniques, are shown in Figure 4.5 and Figure 4.6. The optimum interval limit setting technique is an exclusion-only method while the profile likelihood ratio technique is a background-subtracting method. Presently, the MATLAB code used to perform these calculations is located in the `LTPFeatures` branch of

`git@gitlab.com:supercdms/Limits/SensitivityEstimate.git`.

More details about the code and how it works can be found in the SuperCDMS-internal webpage Reynolds (2020).

Procedure for projecting sensitivity

1. Generate raw background spectra, which are broadly classified into the four categories below. All raw background spectra are generated using GEANT simulations and normalized by assay

results. The one exception is the ionization leakage background, the formula for which is given below.

- Bulk electron recoils: Compton scatters from detector external radioactivity; tritium (^{3}H) beta decays within the detectors; for silicon detectors only, ^{32}Si and ^{32}P beta decays; and for germanium detectors only, cosmogenic activation lines. Except for Compton scatters, all of these backgrounds arise from detector or material exposure to high-energy cosmic-ray secondaries or their spallation products.
- Surface electron recoils: ^{210}Pb and ^{210}Bi beta decays, which result from radon exposure and subsequent ^{210}Pb plate-out onto detector surfaces and detector housing materials.
- Bulk nuclear recoils: coherent scattering of neutrons, neutrinos, and photons with nuclei. The latter process (Robinson, 2017) is not modeled in GEANT, so we include an approximate analytic form that is normalized to the Compton scattering rate.
- Surface nuclear recoils: Recoiling ^{206}Pb from ^{210}Po alpha decays. ^{210}Po is present due to radon decay and ^{210}Pb implantation.
- Ionization leakage (HV detectors only): the tunneling of individual charge carriers from the electrodes into the bulk. This background is treated as a Poisson process with rate normalized by prototype HV detectors known as HVeV detectors and scales with the detector surface area. The following Poisson distribution equation provides the rates for any number of leakage-induced electron-hole pairs, given the rate of single electron-hole pairs R_{dark} and a time resolution t_{sep}:

$$R_j = R_{\text{dark}} \frac{(R_{\text{dark}} t_{\text{sep}})^{j-1}}{(j-1)!} e^{-R_{\text{dark}} t_{\text{sep}}}. \tag{4.2}$$

  R_{dark} is scaled from HVeV Run 4. t_{sep} is computed using the time resolution one can infer from the optimal filter formalism; see Section 1.2.3.2 of Reynolds (2020). The former is assumed for the sensitivity projections presented in the SuperCDMS Snowmass 2021 contribution and this thesis.

2. Generate signal spectra; options are:
   - WIMP signal model, which is a nuclear-recoiling signal;
   - `QEDark` code signal model, which is an electron-recoiling signal;
   - other signal models not further discussed in this thesis: dark photon dark matter and axion-like particle dark matter, both of which produce charge signals via direct absorption by an electron and creation of an electron-hole pair. These hypothetical signals are monochromatic because the entire mass of the particle would be absorbed.

3. Transform all recoil energy spectra into phonon energy spectra by way of Equation 4.1 and apply relevant detector cuts.
   - The yield functions chosen for this particular run of SuperCDMS projections and forecasts are the IMPACT model for silicon (SuperCDMS Collaboration, 2022b) and the Sarkis model for germanium (Sarkis et al., 2020). The former is based on the IMPACT experimental measurement while the Sarkis model is a theoretical one that accounts for the binding energy of an ion to its lattice site. These models and various others are shown in Figure 4.4. Cutoff energies were applied to the yield function: 27.8 eV for silicon and 22.7 eV for germanium.
     - Ionization yield cuts are performed for iZIPs. The amount of charge produced by an event of some recoil energy E_r is computed by way of the yield function $Y(E_r)$. The charge threshold, which is given by the ionization resolutions specified in Table 4.1,

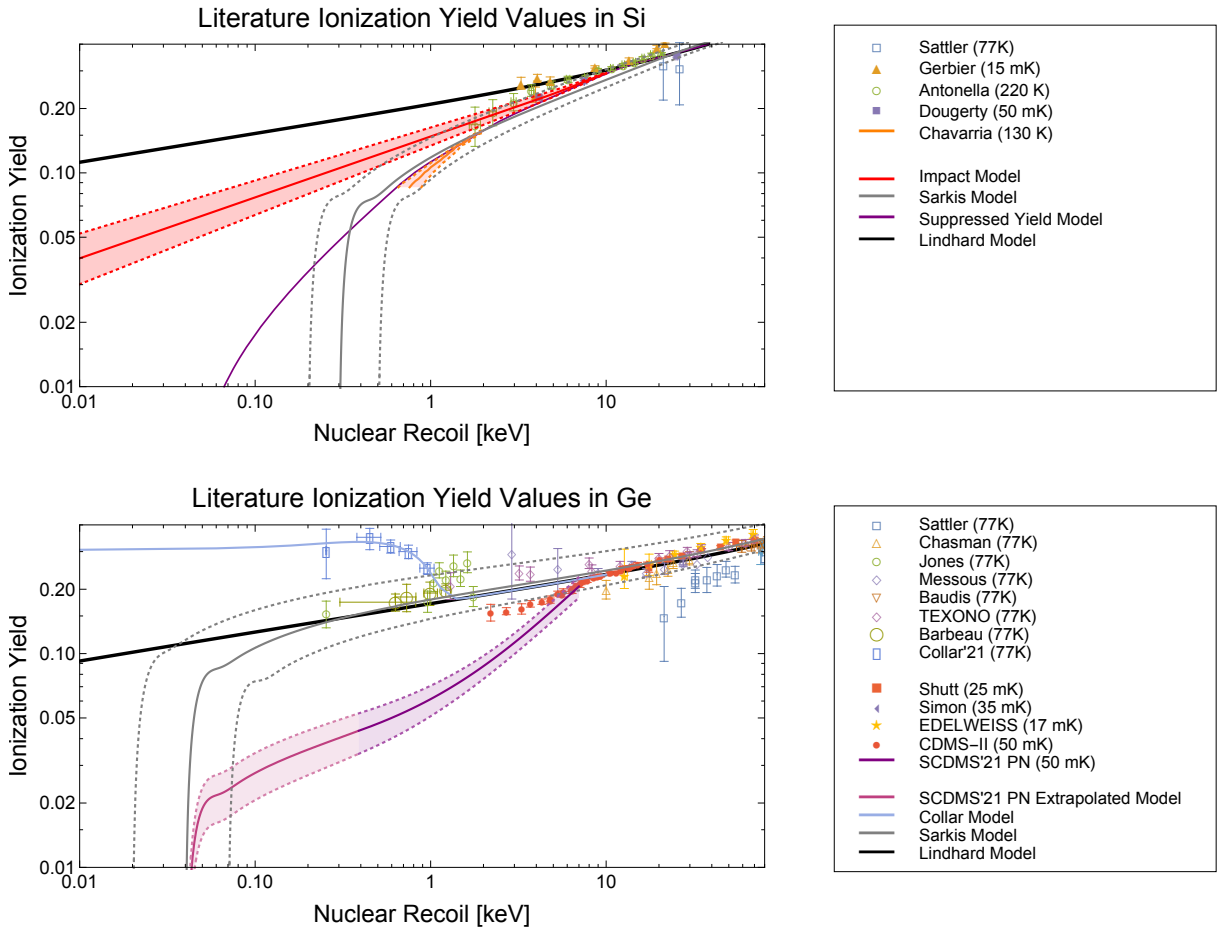

Figure 4.4: **Ionization yield models** $Y(E_r)$**.** Top: ionization yield data and models for silicon, shown for different experiments and associated extrapolations, alongside two different theoretical models: the Lindhard model and the Sarkis model (Sarkis et al., 2020). Bottom: same, except for germanium. Figure is from SuperCDMS SNOLAB TDR (2018) and was produced by Tarek Saab.

sets the phonon energy above which electron recoils are efficiently rejected. The rejection ability of these cuts turn off smoothly with energy. The events that are accepted by it are everything below the ionization threshold, and the fraction of events falling below that threshold increases smoothly (though steeply) as the mean yield approaches that threshold.

- For iZIPs , use “continuous” code that treats charge production and ensuing NTL boost to the phonon spectra as continuous transformations.
- For HV detectors, use “discrete” code that treats charge production discretely at low energy.[1] This discretization incorporates the different expected Fano factors for electron

[1] The “continuous” and “discrete” codes are actually quite distinct bodies of MATLAB code that organize the background and signal spectra and related MATLAB objects in different ways and perform the steps of this procedure in slightly different orders. There are some shared helper functions, such as everything related to setting dark matter parameter constrains, but the background and signal spectra generation are performed completely differently, in both physics and code structure. The choice between the "continuous" and "discrete" codes must be made before any other

and nuclear recoils and for silicon and germanium. It furthermore incorporates charge trapping and impact ionization. These two effects "fill in" the spectra between integer electron-hole pair peaks; see SuperCDMS Collaboration (2020) for more details. This effect is relevant for all detector spectra, including leakage, electron recoils, and nuclear recoils.

- Convolve the phonon energy spectra with the detector energy resolution. There is a fixed baseline resolution as well as an energy-linear degradation for position dependence. See Section 1.2.3.3 of Reynolds (2020). In the continuous code, the convolution is performed on the entire phonon energy spectra that have been transformed from recoil energy using Equation 4.1. In the discrete code, each bin in recoil energy is treated separately in the production of quantized charge distributions, which ultimately transform into phonon energy with Equation 4.1. After the phonon spectrum from a particular bin in recoil energy is calculated and convolved with the energy resolution, the spectra are summed across the recoil energy bins.
- Position fiducialization cuts: the continuous code uses a simple model to build a distribution of events in r and z, derived from prior data. The discrete code performs r- and z-fiducialization by way of geometrical models for the propagation of phonons to produce full distributions of fiducialization paremeters.
  - r-direction fiducial volume cut; both iZIPs and HV exhibit r-sensitivity due to the presence of inner and outer electrodes.
  - z-direction fiducial volume cut; this cut is only applicable for iZIPs because it relies on the interleaved bias electrodes.

4. On the basis of the above spectra, generate 100 sets of fake background-only events with Monte Carlo methods and the expected exposure. We refer to these events as d, and as in an actual experiment, they take the form of a list of event energies: there are N_{obs} total events, and d_j is the energy of the jth event.
5. Set 90% confidence level exclusion limits on the dark matter interaction cross section (i.e., on the normalization of the signal spectrum) with two different limit-setting techniques for all 100 data sets and compute the median limit for each technique.
   - Upper-limit-only analysis with limits given by the optimum interval (OI) technique; see Yellin (2002) and Yellin (2007). The fundamental question that an upper-limit-only analysis seeks to answer is the following: for what combinations of dark matter mass and cross section would 90% of experiments observe more events than the observed number of events? On top of that, the optimum interval technique identifies the energy gap between events that is most constraining for a given DM mass and then pays a trials penalty for it.
   - Profile likelihood ratio (PLR) technique, wherein the generated data is modeled as the sum of the signal and background spectra and a constraint is set on how much signal could be present in the presence of a known, subtractable background. The forecasting code assumes three separate subtractable backgrounds: neutrinos, neutrons, and the sum of all electron recoils. A more specific walkthrough of the PLR method is carried out in Chapter 5 and may provide additional intuition for the technique.
     - The following likelihood function was used:

$$\mathcal{L}(N_s,\{N_b\}) = P_{\text{Poisson}}(N_{\text{obs}} \mid N_s + \sum_{k=1}^{3} N_{b,k}) \quad \begin{array}{l}\text{Poisson probability of}\\ \text{total \# of events}\end{array}$$

step and depends on the type of detector architecture for which sensitivity calculations are desired.

$$\times \prod_{j}^{N_{\rm obs}} f(d_j \,|\, N_s,\{N_b\}) \qquad \text{probability of each event given a particular spectral shape}$$

$$\times \prod_{k=1}^{3} G(N_{b,k} \,|\, N_{b,k,\rm exp}), \qquad \text{Gaussian priors on background rates} \tag{4.3}$$

$$\text{where} \quad f(d_j \,|\, N_s,\{N_b\}) = \frac{N_s f_s(d_j) + \sum_{k=1}^{3} N_{b,k} f_{b,k}(d_j)}{N_s + \sum_{k=1}^{3} N_{b,k}}, \tag{4.4}$$

N_s and $\{N_b\}$ are the signal and background normalizations (the three backgrounds are indexed by k), f_s and $\{f_b\}$ are the signal and background spectral shapes, and $G(N_{b,k} \,|\, N_{b,k,\rm exp})$ is a Gaussian prior on the background rates given the expected number of background events $N_{b,k,\rm exp}$ for a particular background.

– 10% systematic uncertainty is assumed for each subtracted background. In other words, 10% of the mean value is taken as the rms of each Gaussian prior G. This assumption is intended to stand in for uncertainties on contamination assays, simulation systematics, and uncertainty on empirical rate normalizations (e.g., from high energies and/or signal sidebands). For signals and backgrounds that have the same spectral shape, i.e., for data that has no constraining ability on a particular signal, this assumption limits the gains to be had with exposure, which would indefinitely scale with $\sqrt{\text{exposure}}$ if it were only the Poisson fluctuations on the background rates that contribute to the uncertainty on the background subtraction.

– By Wilk's theorem, the following D-statistic is assumed to follow a χ^2 distribution:

$$D(N_s) = -2 \ln \frac{\mathcal{L}(N_s,\{\hat{\hat{N}}_b\})}{\mathcal{L}(\hat{N}_s,\{\hat{N}_b\})} \tag{4.5}$$

The $\hat{}$ notation in the denominator indicates that the likelihood function is maximized over both variables simultaneously. The $\hat{\hat{}}$ notation in the numerator indicates that the the likelihood is maximized over $\{N_b\}$ at fixed N_s. Since the data are generated only from the background spectra, the maximum likelihood set of parameters $\hat{N}_s$ and $\{\hat{N}_b\}$ should have $\hat{N}_s$ at or near 0. Given that $D(N_s)$ is χ^2-distributed, the PLR technique defines an X% confidence level allowed region by accepting values of N_s for which D is below the X% point of its cumulative distribution function—i.e, the experimental outcome is among the 90% most likely outcomes for the computed values of N_s and $\{\hat{\hat{N}}_b\}$.

– The fundamental question that the profile likelihood ratio technique seeks to answer is the following: given the data, for what values of N_s is the observed experimental outcome within the set of X% most likely outcomes?

The neutrino fog

The solar neutrino background is an irreducible background for SuperCDMS SNOLAB and limits the reach to about $10^{-45}\,\text{cm}^2$ to $10^{-44}\,\text{cm}^2$ dark-matter-nucleon cross section. The exposures required to reach this level are around 100 kg·yr. SuperCDMS is not expected to be neutrino-background-limited, but the most sensitive detector architectures may detect one to a few neutrinos over the course of the experiment.

Single neutrino sensitivity and the neutrino fog are shown in the various forecasts that show sensitivity to the dark-matter-nucleon cross section. The neutrino fog is determined by a PLR

method: the neutrino background is assumed to be subtractable and could potentially be identified by the overall rate and spectral shape, but the uncertainty on this subtraction is the ultimate limiting factor on the capacity for experiments to explore beyond $10^{-45}\,\mathrm{cm}^2$ to $10^{-44}\,\mathrm{cm}^2$.

The un-modeled zero charge low energy excess

An extremely important caveat in these projected sensitivity forecasts is that the zero charge low energy excess (0QLEE) has not been modeled as a background. The 0QLEE background is a rising low energy spectrum that has been observed below 100 eV to 1 keV and greatly inhibits low-mass dark matter searches. Details about the low energy excess across many different detector types are reported in EXCESS workshop (2022). The impact of the 0QLEE background depends on the phonon threshold of a given experiment. See Section 4.6 and Section 5.1 for more discussion.

4.3 NRDM sensitivity projections

The above procedure is followed to project SuperCDMS SNOLAB dark matter sensitivity across four combinations of detector architecture and material combinations. The limits are shown in Figure 4.5 and Figure 4.6, with accompanying plots that show the signal and background spectra used to compute each limit. Observations:

- High-level descriptions of the shapes of the spectra:
  - The iZIP detector spectra exhibit downward turns corresponding to either electron recoil rejection or surface event rejection. Bulk electron recoils in red are efficiently rejected above ∼1 keV because the ionization threshold is 7×180 eV, or 7 times the ionization resolution. The efficiency of this rejection improves as the recoil energy increases beyond this threshold.
  - The HV detector spectra exhibit resolved peaks corresponding to integer quantities of electron-hole pairs produced for both electron and nuclear recoils. For electron recoils, the peaks are located at $N_{\mathrm{eh}} \times (100\,\mathrm{eV} + \epsilon_{\mathrm{eh}})$. For nuclear recoils, the peaks are shifted upward in energy due to the lower ionization yield for nuclear recoils: for a given charge produced, a nuclear recoil must deposit more recoil energy than the corresponding electron recoil, and that extra recoil energy appears as primary phonons.
- HV detectors have better mass reach than iZIP detectors because the greater NTL gain results in a lower recoil energy threshold. For example, the dashed magenta curve in the silicon iZIP spectrum plot is the signal spectrum for a 1.6 GeV dark matter particle and reaches roughly 1 keV phonon energy. In contrast, the dashed magenta curve in the silicon HV spectra plot is for a lower mass 0.5 GeV dark matter particle and reaches roughly 1 keV phonon energy.
- HV detectors are limited by ionization leakage rather than by the phonon energy threshold. To the left of the dash-dotted line in the spectra, leakage rates are very high; the analysis conservatively takes the dash-dotted line as the lower bound on the analysis region.[2] HV detectors would be further enhanced in mass reach were it not for the leakage threshold. The silicon leakage threshold is above the second electron-hole pair peak. The germanium threshold is above the third.
- Both the iZIP and the HV detector spectra incorporate rejection of surface events above a certain energy. This rejection can be characterized with fiducial volume cuts in the r- and z-directions of the detector. The mustard and green curves are surface nuclear recoils and surface electron recoils, respectively.

[2] The analysis could have been done down to the phonon energy threshold, treating the leakage as a background, but the computational cost of the profile likelihood ratio fit, which is done using an unbinned maximum likelihood technique, is prohibitive. The energy region dominated by the leakage offers little additional sensitivity.

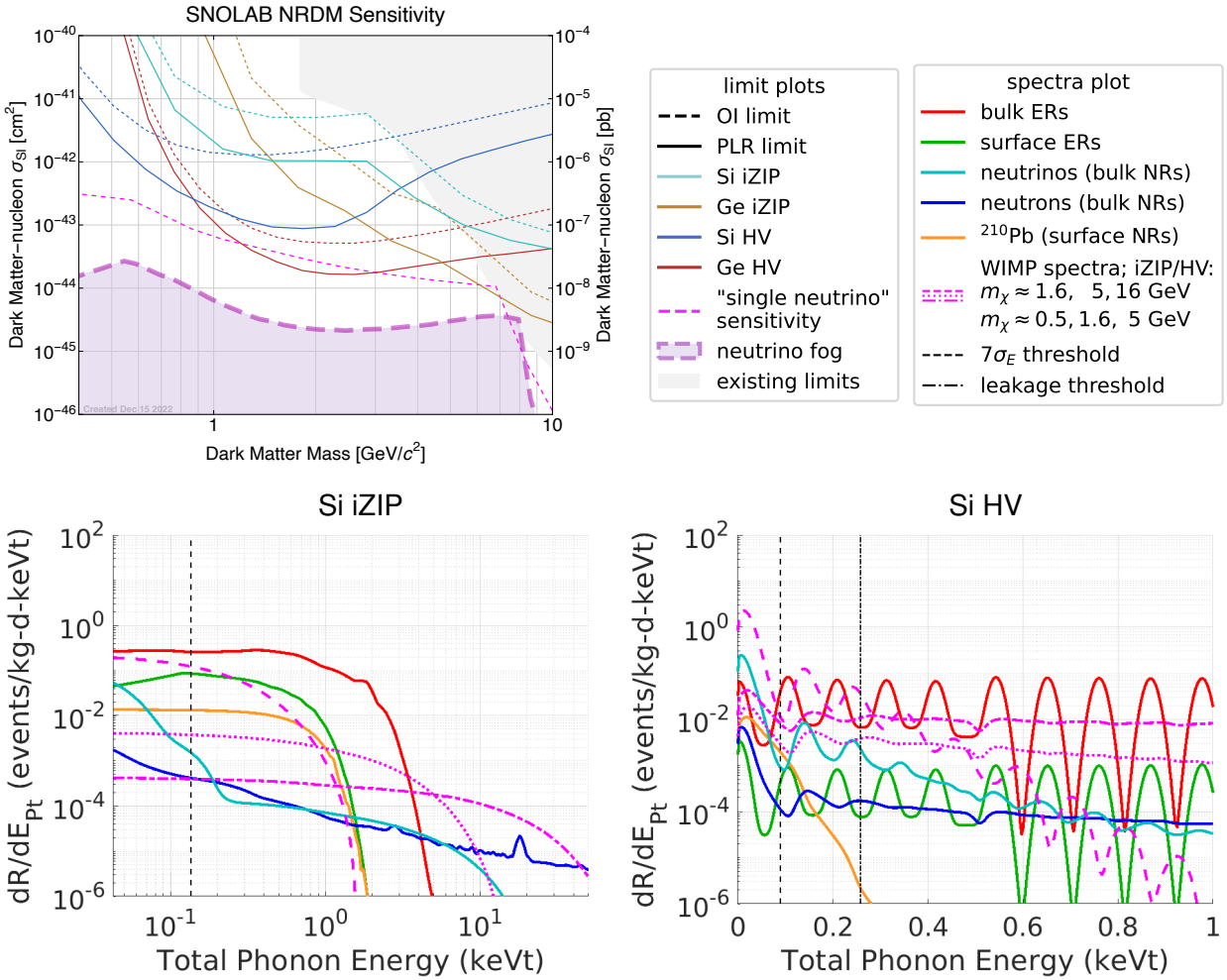

Figure 4.5: **NRDM sensitivity of SuperCDMS SNOLAB and silicon detector spectra.** Top: projected 90% CL exclusion limits for both OI and PLR limit-setting techniques for all four detector architecture and material combinations along with current limits, the sensitivity at which a single neutrino event is observed, and the neutrino fog. Bottom: Background and signal spectra at the exclusion cross section in phonon energy after all analysis cuts for Si iZIPs (Left) and Si HV (Right) along with phonon energy (dashed) and leakage thresholds (dash-dotted), where applicable.

- The silicon sensitivity projections reach lower dark matter masses than the germanium projections because the silicon detectors are projected to have lower phonon energy thresholds than the germanium detectors, as evidenced by the lower $7\sigma_E$ threshold for silicon, which are delineated in the spectra plots. Furthermore, silicon has a lower atomic mass than germanium and thus has better kinematic matching at lower dark matter mass; for equal energy thresholds, silicon is more sensitive to smaller dark matter masses.
- The better resolution of the silicon detector is visible in the HV detector spectra. The germanium HV spectra with 34 eV energy resolution are much more smeared than the silicon HV spectra with 13 eV energy resolution.
- The germanium sensitivity projections reach to lower dark matter cross section because the background rate is lower for germanium than it is for silicon. This latter fact can be seen

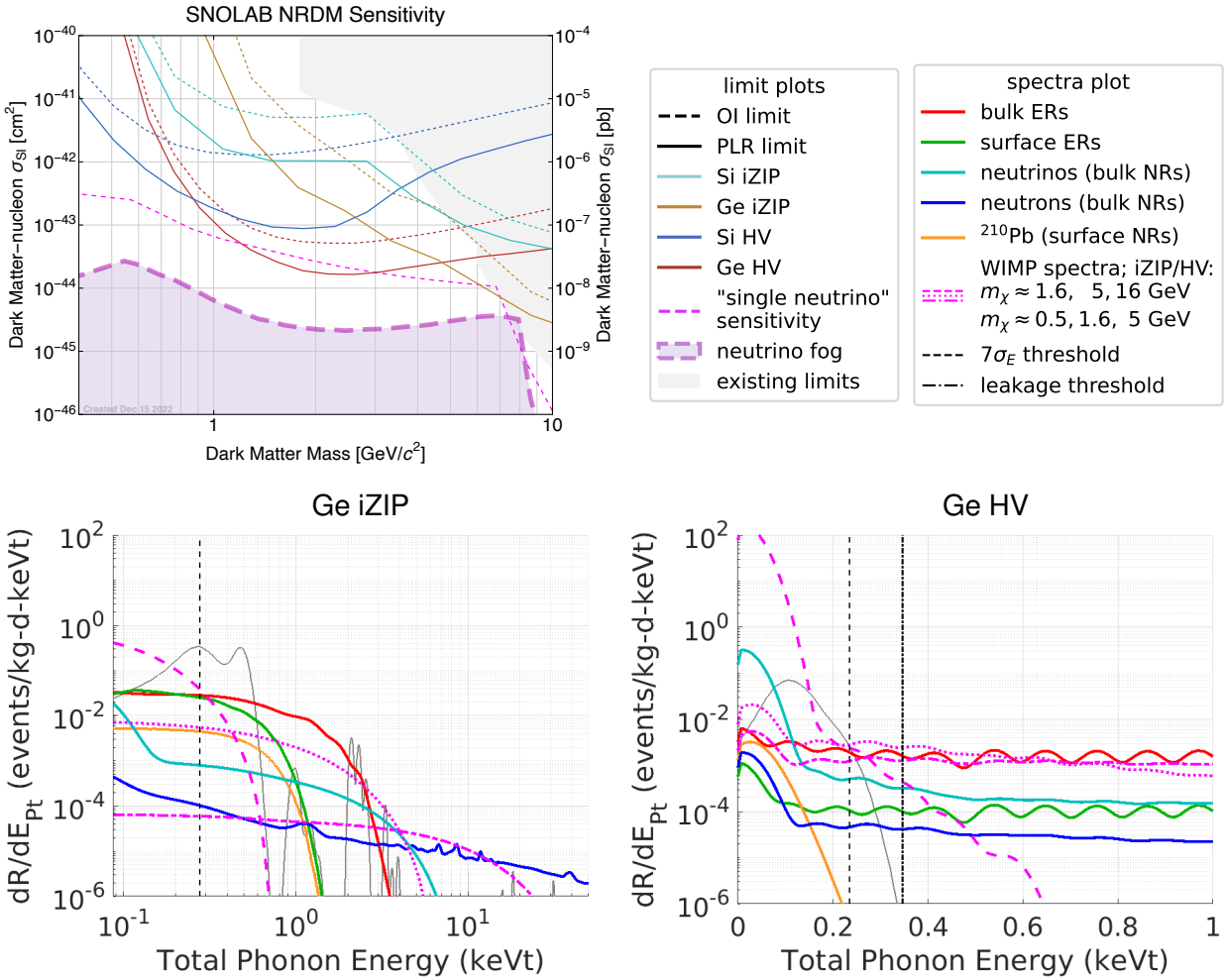

Figure 4.6: **NRDM sensitivity of SuperCDMS SNOLAB and germanium detector spectra.** Caption is the same as Figure 4.5, except for germanium signal and background spectra.

from the spectra of all background components; the overall level of the backgrounds, which are generally dominated by bulk ERs displayed in red, are lower for germanium than for silicon.

Both the NRDM projections in this section and the ERDM projections in the next section may suffer from worse than projected sensitivity due to the 0QLEE background. Given the 100 eV to 200 eV $7\sigma_E$ thresholds of SuperCDMS SNOLAB, the background rate near the phonon energy threshold may be higher due to the 0QLEE background, which has been observed to rise above the Compton background from anywhere between 100 eV and 1 keV. iZIPs are not expected to provide the best mass reach for SuperCDMS, so excess rates near threshold are not as much of a concern. At higher energies, previous CDMS experiments have shown excellent capacity to reject zero charge events in the 1 keV to 10 keV lower threshold analysis regions that have been investigated; see SuperCDMS Collaboration (2010). For HV detectors, the 0QLEE background is expected to be sub-dominant to the ionization leakage background at energies below the dash-dotted lines in the spectra plots. Above the leakage threshold, it may be possible to subtract the 0QLEE background, as described in great detail in Chapter 5.

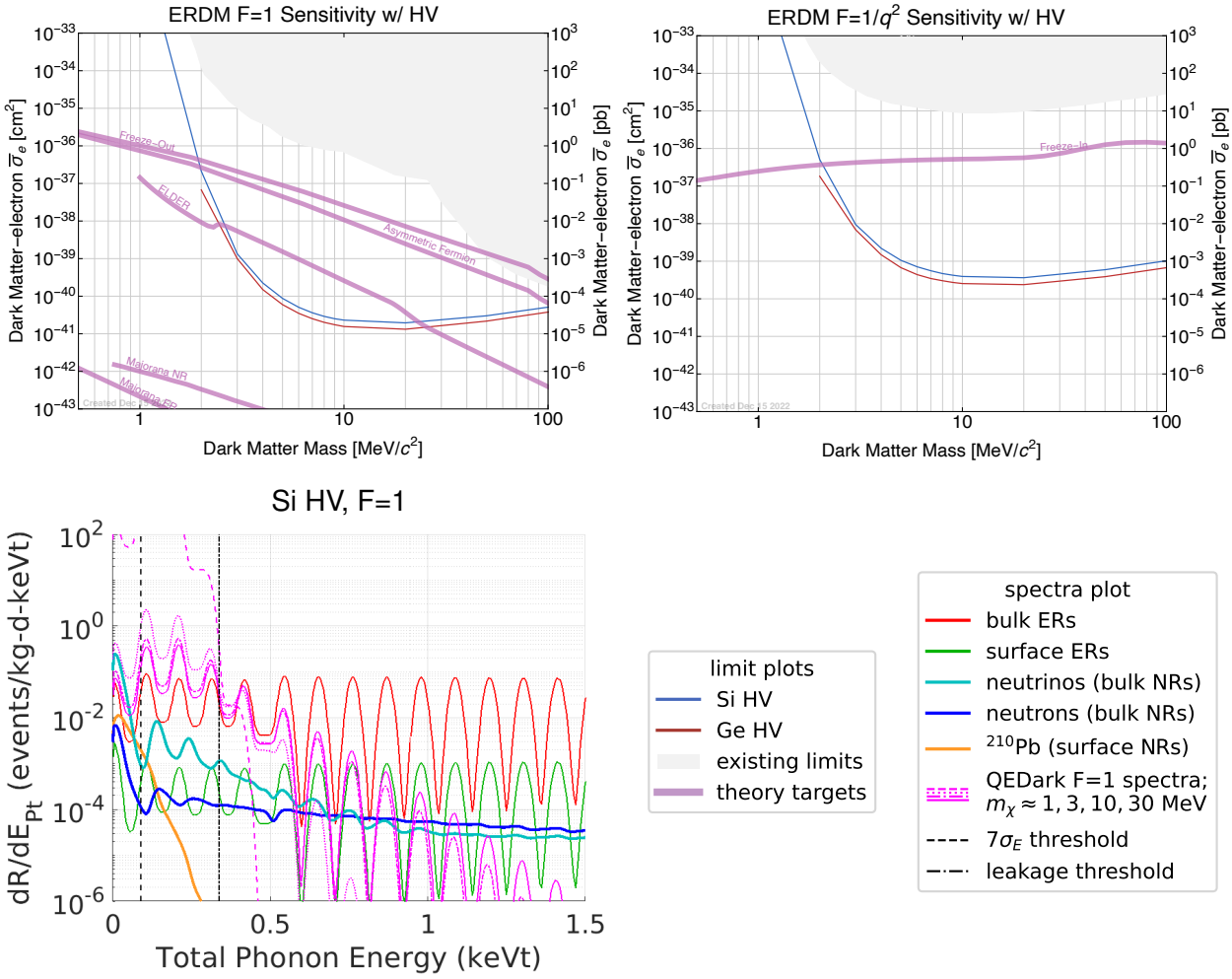

Figure 4.7: **SuperCDMS SNOLAB ERDM sensitivity and related detector spectra.** Top Left: projected PLR-generated 90% CL exclusion limits for electron-recoiling, heavy mediator, low-mass dark matter with SuperCDMS SNOLAB HV detectors, compared with current limits and various theory targets. Top Right: same as Top Left, except for a light mediator. Bottom: background and heavy mediator signal spectra at the exclusion cross section in phonon energy after all analysis cuts for Si HV detectors, along with phonon energy and leakage thresholds.

4.4 ERDM sensitivity projections

SuperCDMS SNOLAB HV detectors will provide extraordinary reach beyond current limits for electron-recoiling low-mass dark matter, as shown in Figure 4.7. At the projected sensitivity, SuperCDMS SNOLAB will probe cross sections that are more than four orders of magnitude smaller than current limits at 10 MeV dark matter mass for both heavy and light dark photon mediators. This expectation is attributable to (1) the 4 to 5 additional orders of magnitude in exposure for SuperCDMS SNOLAB in comparison to the experiments that set the currently world-leading limits and (2) the low backgrounds that are expected for dark matter searches around 10 MeV particle mass, in particular for the energy regime where 4 or more recoil-generated electron-hole pairs are produced. Exposure versus background rate is plotted in Figure 4.8 for the SuperCDMS SNOLAB projections, as well as the recent results from DAMIC-M (DAMIC-M Collaboration, 2023) and

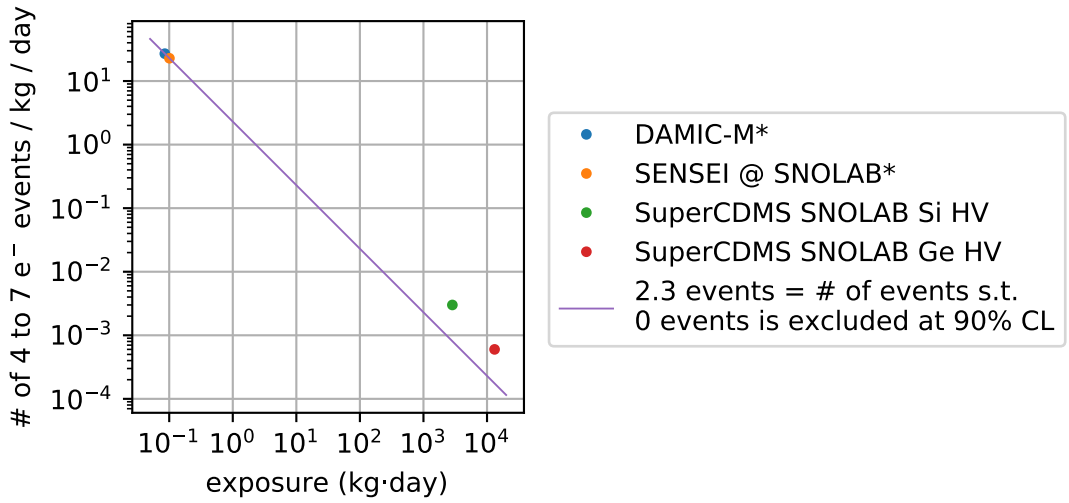

Figure 4.8: **Exposures and background rates for SuperCDMS SNOLAB HV detectors.** Comparison with DAMIC-M and SENSEI results are provided. The * denotes the fact that DAMIC-M and SENSEI observed zero events in the 4 to 7 e^- range as reported in their most recent results, so the rate at which zero events over the reported exposure would be excluded at 90% CL is chosen instead; as mentioned in the legend, the number of events corresponding to that rate is 2.3.

SENSEI (SENSEI Collaboration, 2023) to convey the capacity of SuperCDMS SNOLAB to explore beyond current exclusion limits.

For electron recoils that produce four or more electron-hole pairs, the background rate is projected to be about 2×10^{-3} dru for the germanium HV detectors, dominated by a tritium ^{3}H decay background and low energy Compton scatters. The background rate is expected to be about $10\times$ higher for the silicon HV detectors due to the presence of ^{32}Si. To achieve these low background rates, SuperCDMS has made numerous design choices regarding detector fabrication, detector handling, and cryostat material makeup to minimize the impact of detector internal contamination and experimental material radioactivity.

A key limitation for SuperCDMS HV detectors in their search for electron-recoiling low-mass dark matter is ionization leakage, which dominates background rates at one to a few electron-hole pairs. The origin of ionization leakage is the tunneling of individual charge carriers from the electrodes into the bulk. The spectrum plot for the ERDM projection includes a "leakage threshold" below which the leakage rates are dominant; Events in that energy regime are ignored, as was done for the nuclear recoil sensitivities.

As a reminder, one important background that has not been modeled for these projections and that may impact the sensitivity of the experiment is the 0QLEE background. This background is not expected to be dominant over ionization leakage in HV detectors, which dominate SuperCDMS SNOLAB's sensitivity for electron-recoiling low-mass dark matter. Above the leakage thresholds, it may be possible to subtract the 0QLEE background in search of electron-recoiling dark matter, as is mentioned for NRDM searches. The subtraction of the 0QLEE background for an NRDM search is done in Chapter 5.

HV detectors at SuperCDMS SNOLAB are especially exciting for dark matter searches because of their capacity to probe sharp theory targets in the space of electron-recoiling low-mass dark matter. In the parameter space presented in Figure 4.7, the theory targets are the combinations

of scattering cross section and dark matter mass that produce the correct relic abundance of dark matter. The projected sensitivities surpass the freeze-out and asymmetric fermion target curves over two decades of dark matter mass from roughly 2 MeV to 100 MeV for a heavy mediator ($F = 1$). They also surpass the ELDER theory curve across a single order of magnitude in dark matter mass from 2.5 MeV to 25 MeV. For the case of a light mediator ($F = 1/q^2$), the projected SuperCDMS SNOLAB HV detectors are projected to be sensitive to the freeze-in target band over two decades of mass.

4.5 Forecasting for next-generation SuperCDMS detectors at SNOLAB

The second major component of the SuperCDMS contribution to the Snowmass process (SuperCDMS Collaboration, 2023) is to lay out a plan for future SuperCDMS-style experiments. The SuperCDMS Long-Term Planning Task Force systematically studied future possibilities that could leverage the under-construction SuperCDMS SNOLAB infrastructure. There are six core components to forecasting the performance of future experiments: signal models of interest, detector architecture, detector material, detector size, detector upgrade scenario, and background upgrade scenario.

- Four signal models of interest: nuclear-recoiling WIMP-like dark matter, electron-recoiling low-mass dark matter (both heavy and light mediator), dark photon dark matter, and axion-like particles. Only the first two are presented in this thesis.
- Four detector architectures: iZIPs, HV, 0V, and piZIP. This thesis only discusses the forecasts that were performed with HV and 0V detectors.
- Two detector materials: silicon and germanium.
- Three different detector sizes: SNOLAB-sized, 10 cm^3, and 1 cm^3. Each forecast is performed with the assumption of two towers of detectors, each of which provides 72 readout channels. Limited by readout channels, a single tower accommodates six SNOLAB-sized detectors, 72 10 cm^3, or 72 1 cm^3 detectors. Occasionally, when a 10 cm^3 or 1 cm^3 forecast is seen to be exposure-limited, 20× exposure forecasts are performed, which would require additional cabling beyond what is currently planned.
- Three detector upgrade scenarios: `Det A`, `Det B`, and `Det C`, which correspond to tiers of improvement in detector resolution and other detector performance parameters, with `Det C` involving the greatest detector advancement. `Det A` are upgrades that are already possible today but are too late to be implemented for SuperCDMS SNOLAB.
- Three background upgrade scenarios: `Bkg 1`, `Bkg 2`, and `Bkg 3`, which correspond to tiers of improvement in background levels, with `Bkg 3` involving the greatest decrease in backgrounds. `Bkg 1` are upgrades that are already possible today but are too late to be implemented for SuperCDMS SNOLAB.
  - A major finding of the long-term planning forecasts is that upgrades beyond `Bkg 1` are generally not required to probe large amounts of unexplored parameter space. Detector upgrade scenarios are shown to be fully capable of exploring new parameter space with only minimal background improvements. All of the presented sensitivity forecasts are thus shown for the `Bkg 1` scenario. An important caveat in this claim is that there are likely to be poorly understood and perhaps unforeseen backgrounds (e.g., 0QLEE, background radiation) that almost certainly require careful study and mitigation.

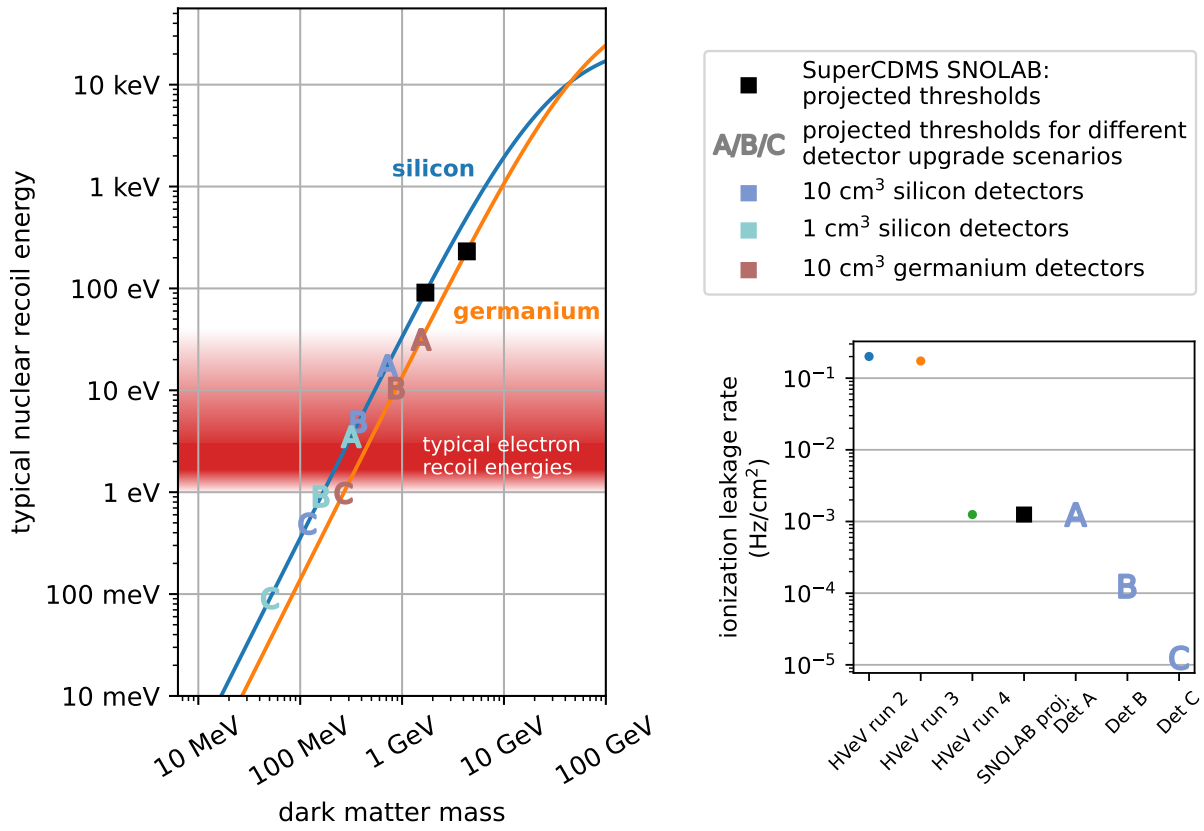

Figure 4.9: `Det A/B/C`**: expected energy thresholds, mass reach, and ionization leakage rates.** Left: elastic nuclear recoil energy versus dark matter mass, with expected thresholds and ergo mass reach demarcated along the curve. Bottom Right: measured and expected ionization leakage rates for both past and future SuperCDMS experiments.

The `Det` scenarios: projecting future energy resolutions and leakage performance

The results of Section 4.6 and Section 4.7 make apparent the impact of detector upgrade scenarios `Det A/B/C` on experimental sensitivity. Thus, we are motivated to present additional details on how the various `Det` scenarios are chosen and projected.

The main difference between the different `Det` scenarios is the projected detector phonon energy resolution σ_E for each scenario. σ_E of a hypothetical detector is projected based on current detector performance and appropriate scaling laws for σ_E as it depends on phonon sensor parameters and detector dimensions. Specifically, all projected energy resolutions are scaled from the SuperCDMS PD2 results (SuperCDMS Collaboration, 2021); the key parameters of the phonon sensor are T_c, which is the critical temperature of the TES, and $t_{\text{W-Al}}$, which is the single attempt probability of a quasiparticle to transmit from the aluminum collection fin to the tungsten TES (see Section 3.4). From PD2, $T_c = 41.5\,\text{mK}$ and $t_{\text{W-Al}} = 10^{-4}$. The following formula is used toe forecast σ_E:

$$\sigma_E = \sigma_{\text{E (PD2)}} \left(\frac{T_c}{T_{c\,(\text{PD2})}} \right)^3 \frac{t_{\text{W-Al (PD2)}}}{t_{\text{W-Al}}} \sqrt{\frac{f_{\text{Al}} A_{\text{tot}}}{f_{\text{Al (PD2)}} A_{\text{tot (PD2)}}}} X(\tau_{\text{coll}}, \tau_{\text{TES}}), \tag{4.6}$$

where f_{Al} is the fractional area coverage of the aluminum fins, A_{tot} is the total surface area of the detector, and $X(\tau_{\text{coll}}, \tau_{\text{TES}})$ is an additional scaling law that depends on the phonon collection time

τ_{coll} and TES fall time τ_{TES} and arises from an optimally filtered pulse height estimation. This final scaling law is represented as $X(\tau_{\text{coll}},\tau_{\text{TES}})$ for simplicity; see Pyle (2012) for the full expression.

The scalings of σ_E with T_c^3 and $t_{\text{W-Al}}^{-1}$ are well known to the SuperCDMS collaboration. The former arises from the scaling of thermal fluctuation noise with temperature, and the latter is derived from the fact that $t_{\text{W-Al}}$ is proportional to the efficiency of energy collection within the TES in the limit of $t_{\text{W-Al}} \ll 1$. The projected values for T_c in the `Det A/B/C` scenarios are 40 mK, 30 mK, and 20 mK. For $t_{\text{W-Al}}^{-1}$ the projected values are 10^{-4}, $10^{-3.5}$, and 10^{-3}. The T_c values are set by what can be run in the cryostat (the base temperature goal is 15 mK), and the $t_{\text{W-Al}}$ values are ad hoc but seem realistic. The resultant thresholds for the various `Det` scenarios are shown in Figure 4.9 Left, overlain on the curve from Equation 3.6 to show the expected mass reach for each hypothetical detector upgrade scenario.

There is an additional scaling with $(f_{\text{Al}}A_{\text{tot}})^{-1/2}$ that has not been addressed. $f_{\text{Al}}A_{\text{tot}}$ is the total phonon sensor area. Fundamentally, the thermal fluctuation noise in a TES scales as the square root of TES volume. With the forecasting choice that the phonon sensor cell design does not change for each of the projected sensitivities, $f_{\text{Al}}A_{\text{tot}}$ scales linearly with the TES volume. In short, $(f_{\text{Al}}A_{\text{tot}})^{-1/2}$ is a recast form of the typical $V_{\text{TES}}^{-1/2}$ scaling for σ_E. The upshot of the $(\sqrt{f_{\text{Al}}A_{\text{tot}}})^{-1/2}$ scaling is that, at fixed aluminum collection fin fractional coverage f_{Al}, energy resolution improves as detector size decreases, which is visible in the Figure 4.9 Left from the much better resolution that the 1 cm^3 detectors provide over the 10 cm^3 detectors.

For HV detectors, beyond energy resolution, the different `Det` scenarios also project different levels of ionization leakage, as well as different levels of impact ionization and charge trapping. The projected leakage improvements are shown in Figure 4.9 Bottom Right. The leakage current has been seen to improve with pre-biasing the detector at some elevated bias voltage. Improvements in ionization leakage might also come from the implementation of an insulating blocking layer between the electrodes and the substrate.

Discovery potential

To illustrate dark matter discovery potential, the toy Monte Carlo was expanded to include injection of a signal and the PLR code was modified to look for lower and upper boundaries to the 90% CL allowed region. We revisit Equation 4.5:

$$D(N_s) = -2\ln\frac{\mathcal{L}(N_s,\{\hat{\hat{N}}_b\})}{\mathcal{L}(\hat{N}_s,\{\hat{N}_b\})}. \qquad \text{(4.5 revisted)}$$

When the maximum likelihood $\hat{N}_s$ is at or near zero, the χ^2-distributed D statistic only reaches the 90% value of its CDF for a single positive value of N_s. If the maximum likelihood $\hat{N}_s$ is far enough from 0, then there may be a second positive N_s that deviates $D(N_s)$ to a 90% confidence level lower bound on the dark matter cross section.

The forward-looking sensitivity forecasts in the following sections show allowed regions. To perform this calculation, we assume a hypothetical signal model with a cross section that is a factor of a few greater than the 90% confidence level background-only upper limit and inject its spectrum into the data. The maximum likelihood set of N_s and $\{N_b\}$ should then recover N_s from the injected signal model, and an allowed region can be computed around that N_s. Additionally, in this case of an injected signal, we compute the 99.7% confidence level allowed region, corresponding to a 3σ discovery. Such an exercise will be necessary if and when actual experimental data reflect consistency with a dark matter signal model.

These 99.7% allowed regions are shown throughout the following sections. They are referred to as "allowed regions" in the various legends. In some cases, especially for the ERDM `QEDark`

signal model, the discretization of the mass range is poor, and the allowed regions do not close properly around the injected signal.

4.6 Next-generation: low-threshold 0V detectors

Cryogenic phonon-mediated detectors are well positioned to probe unexplored parameter space beyond the reach of SuperCDMS SNOLAB. The fundamental advantage of phonon-mediated detectors over charge- or scintillation-mediated detectors is that the latter have relatively high single-quantum production thresholds (eV and tens of eV).

The effect of improved thresholds

A straightforward path toward new parameter space involves improving the energy threshold to sub-eV and eventually meV-scale energies. In Section 3.1, we work through the example of an elastic collision between a motionless nucleus with mass m_N and dark matter particle of mass m_χ and velocity $\vec{v}$ ($v = 10^{-3}$). The recoil energy imparted on the nucleus is given by

$$\begin{aligned} E_{\mathrm{NR}} =& \frac{1}{2} m_\chi v^2 \frac{4 m_\chi m_N}{(m_\chi + m_N)^2} \frac{1 + \cos\theta_{\mathrm{CM}}}{2} && \text{(Equation 3.5 revisited)} \\ =& \frac{1}{2} m_\chi v^2 \frac{2 m_\chi m_N}{(m_\chi + m_N)^2} && \text{(Equation 3.6 revisited)} \\ \xrightarrow{m_\chi \ll m_N}& \frac{m_\chi^2 v^2}{m_N}, \end{aligned}$$

where θ_{CM} is the angle of deflection in the center-of-mass frame and isotropic scattering is assumed at the second equality, i.e., $\cos\theta_{\mathrm{CM}}$ is a uniformly distributed around 0.

In the limit $m_\chi \ll m_N$, Equation 3.6 becomes $E_R = m_\chi^2 v^2 / m_N$; the recoil energy decreases with the square of the dark matter mass. For each decade of mass sensitivity that is desired in an experiment with fixed target mass, two decades of threshold improvement are required. SNOLAB projects thresholds in the range from 100 eV to 300 eV, providing sensitivity to 1 GeV dark matter. For sensitivity to 100 MeV dark matter, a threshold of ∼1 eV is needed, which is projected for the `Det C` upgrades of 10 cm^3 silicon and germanium detectors, as well as the `Det B` upgrade of a 1 cm^3 silicon detector. According to Figure 4.10, these three scenarios would be sensitive to 100 MeV dark matter to better than $10^{-42}\,\mathrm{cm}^2$ dark matter-nucleon cross section, right around the level of single neutrino sensitivity.

For electron-recoiling low-mass dark matter searches with phonon-mediated silicon or germanium 0V detectors, there is no gain in mass reach by improving the threshold below the bandgap. In silicon and germanium, the band gaps are $\epsilon = 1.1$ eV and $\epsilon = 0.76$ eV. Thus, 1 cm^3 `Det C` upgrade scenarios are superfluous for electron-recoiling dark matter searches with a 0V detector. Figure 4.11 displays this cutoff in the signal spectrum.

Lower cross sections may be probed by increasing the exposure beyond just two towers of 1 cm^3 silicon 0V detectors. Figure 4.11 shows three higher exposure possibilities: 20× exposure versions of the `Det A` and `Det B` 1 cm^3 silicon detectors as well as the nominal two towers of `Det C` 10 cm^3 silicon 0V detectors. As described in the previous paragraph, since both $7\sigma_E$ values for `Det B` 1 cm^3 and `Det C` 10 cm^3 silicon detectors are below the silicon bandgap of 1.1 eV, these detectors have the same mass reach.

0V detector architecture and the low energy excess

As has been mentioned for all the sensitivity projections that are presented in SuperCDMS Collaboration, 2023, these sensitivities do not take into account the low energy excess. The impact

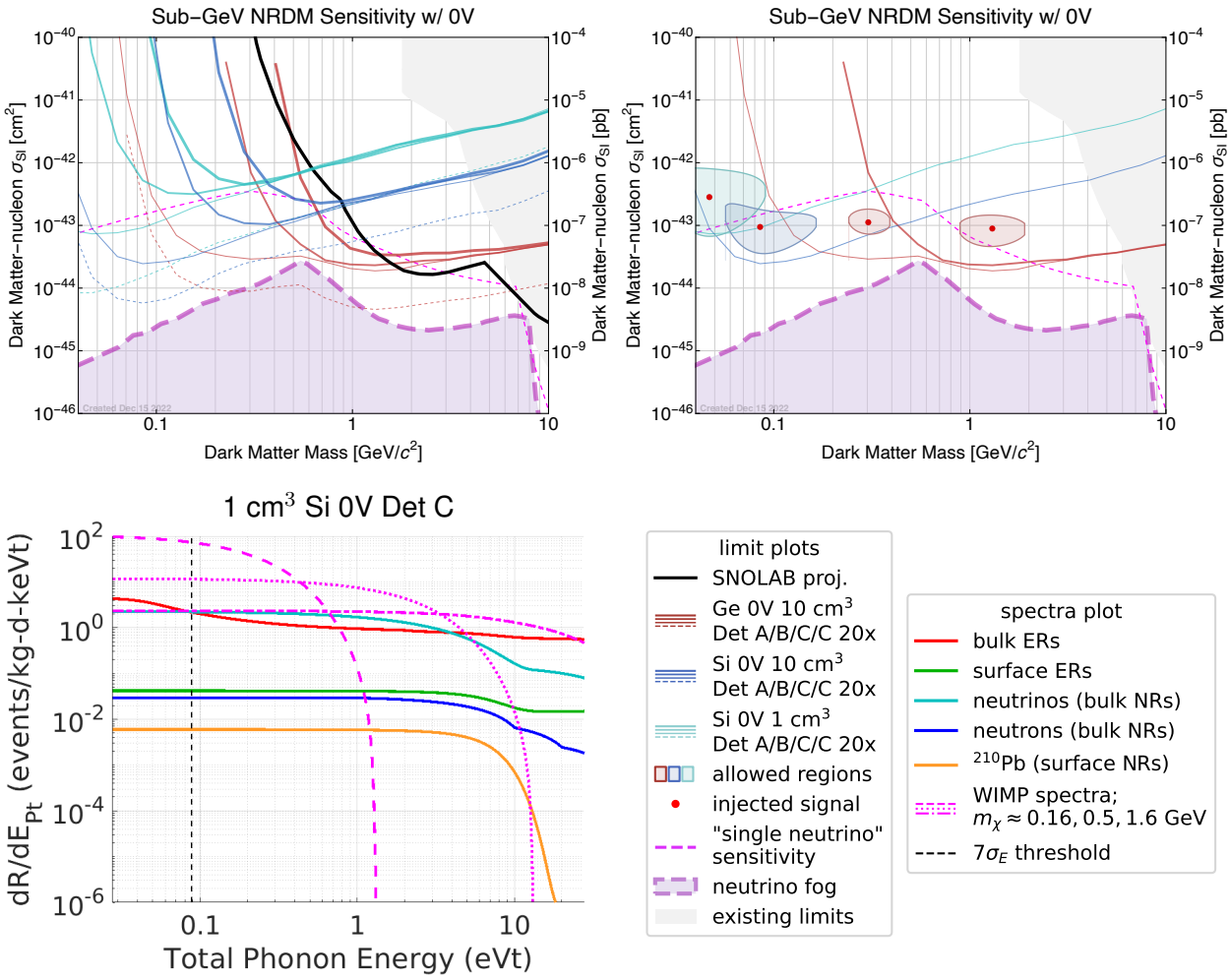

Figure 4.10: **NRDM sensitivity of next-generation 0V detectors and related detector spectra.** Top Left: PLR 90% CL exclusion sensitivity for nuclear-recoiling dark matter with a selection of improved-threshold, next-generation 0V detectors, compared with existing limits, projected SuperCDMS SNOLAB limits, and the neutrino fog. Top Right: The most sensitive projections (that are not 20× exposure) along with 99.7% CL allowed regions for an example injected signal. Bottom Left: background spectra, signal spectra at the 90% CL sensitivity curve, and phonon energy threshold for a two tower 0V silicon 1 cm^3 projection.

of 0QLEE for eV-threshold 0V projections is particularly acute because the excess is seen to rise steeply at low energies. In each of the experiments shown in Figure 5.2, the low energy excess below 200 eV is seen to rise some four to five orders of magnitude above the flat background spectrum at higher energies, peaking at the lowest energies below 50 eV. Understanding and mitigating the 0QLEE background is the foremost issue for low threshold 0V detector dark matter searches.

In a SuperCDMS-style 0V phonon-mediated silicon detector operated above ground (SuperCDMS Collaboration, 2021), the observed event rate at about 20 eV was about 8 events/(g·day·eV), or 8 million dru, which is seven orders of magnitude greater than the expected SuperCDMS SNOLAB Compton background at 20 eV: approximately 0.6 dru, according to Figure 4.10 Bottom Left. 8 million dru is also nearly eight orders of magnitude greater than the beryllium-7 neutrino nuclear

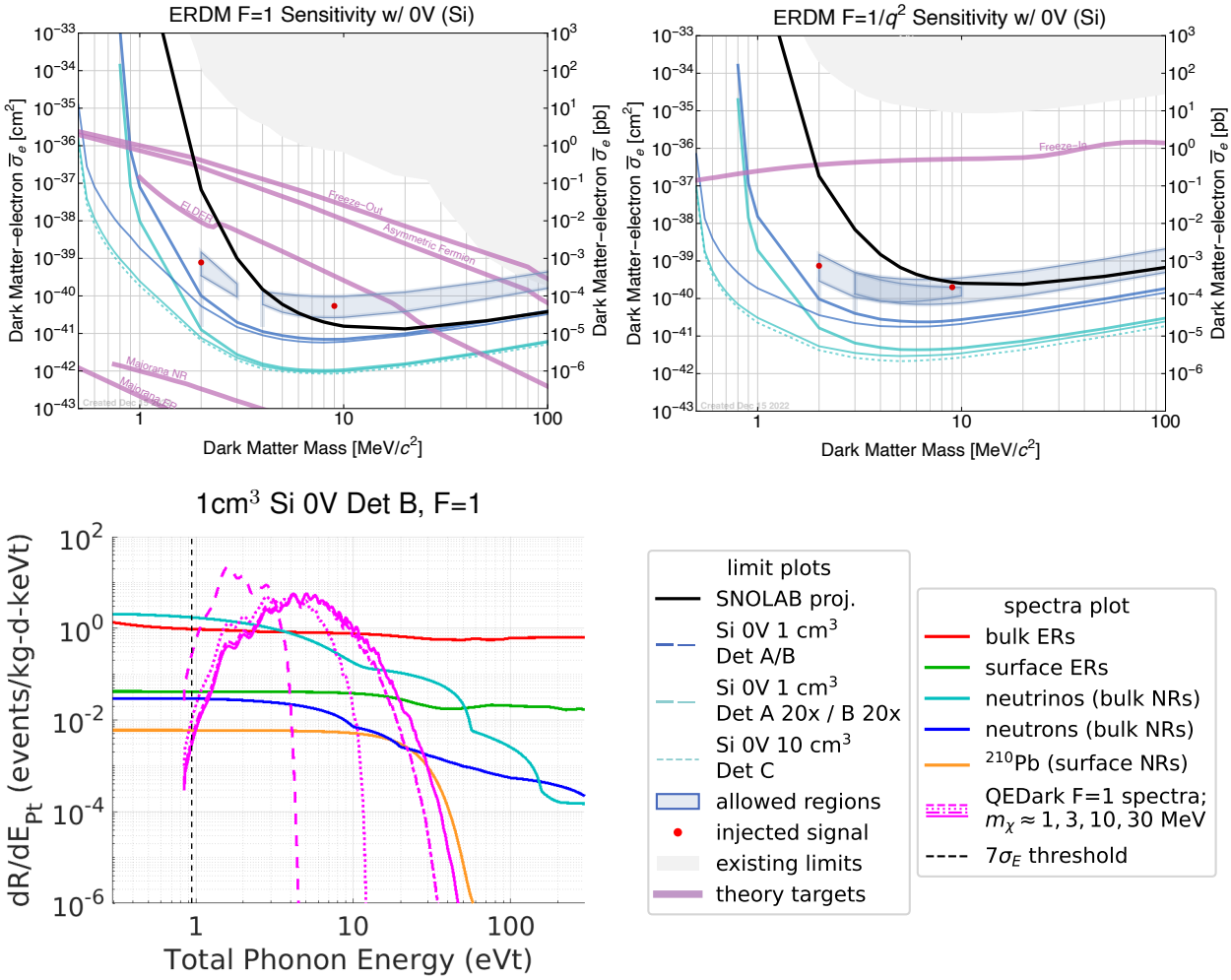

Figure 4.11: **ERDM sensitivity of next-generation 0V detectors and related detector spectra.** Top Left: PLR 90% CL exclusion sensitivity and 99.7% CL allowed regions for electron-recoiling, heavy mediator dark matter with a selection of improved-threshold, next-generation 0V detectors, compared with existing limits, projected SuperCDMS SNOLAB limits, and various theory targets. Top Right: same as Top Left, except for a light mediator. Bottom Left: background spectra, heavy mediator signal spectra at the 90% CL sensitivity curve, and phonon energy threshold for a two tower 0V silicon 1 cm^3 projection.

recoil rate at 20 eV.

Given that the observed 0QLEE rate is enormous compared to the projected background rates that were used to forecast the sensitivity of future SuperCDMS detectors, it is imperative to mitigate or eliminate the 0QLEE background. An effort to identify the source of the 0QLEE background is detailed in Anthony-Petersen et al. (2024). The article demonstrates that stress-induced backgrounds made up a significant fraction of the background rate at 20 eV. Changing the detector mounting scheme to a lower stress version reduced the detector's background rate at 20 eV by almost two orders of magnitude. Despite immense success in identifying a component of and decreasing the 0QLEE background, the rate is still at least five orders of magnitude higher than the expected Compton and neutrino backgrounds at SNOLAB. Film stress is viewed as the likely

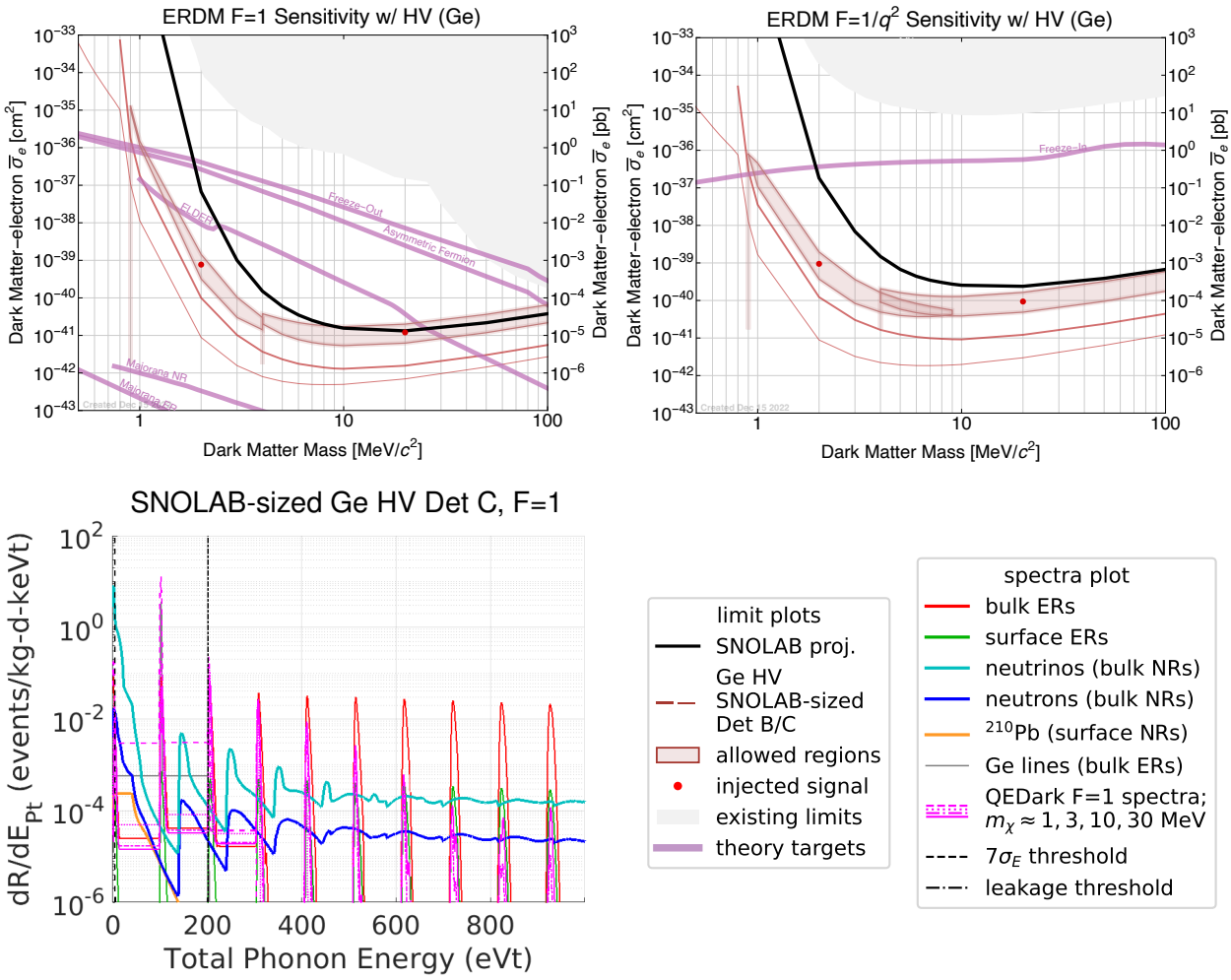

Figure 4.12: **ERDM sensitivity of next-generation HV detectors and related detector spectra.** Top Left: PLR-generated 90% CL exclusion limits and 99.7% CL allowed regions for electron-recoiling, heavy mediator dark matter with two future HV detector payloads, compared with existing limits, projected SuperCDMS SNOLAB limits, and various theory targets. Top Right: same as Top Left, except for a light mediator. Bottom Left: background and heavy mediator signal spectra at the exclusion cross section in phonon energy after all analysis cuts for a two tower projection of Ge HV SNOLAB-sized detectors, along with phonon energy and leakage thresholds.

dominant remaining source (Anthony-Petersen et al., 2024). More work beyond the forecast upgrade scenarios described in Section 4.5 is necessary to bring the projected sensitivities to fruition.

4.7 Next-generation: low-leakage HV detectors

Next-generation HV detectors will benefit greatly from lower leakage backgrounds for searches of electron-recoiling dark matter. Figure 4.12 Top Left and Right show projected gains in ERDM sensitivity in both cross section and mass for SNOLAB-sized `Det B` and `Det C` upgrade scenarios. Both sensitivity projections are shown for germanium detectors, which offer lower radioactive backgrounds than a same-size silicon detector. The fact that SNOLAB-sized HV detectors show better sensitivity than smaller HV detectors implies that exposure is more important than threshold

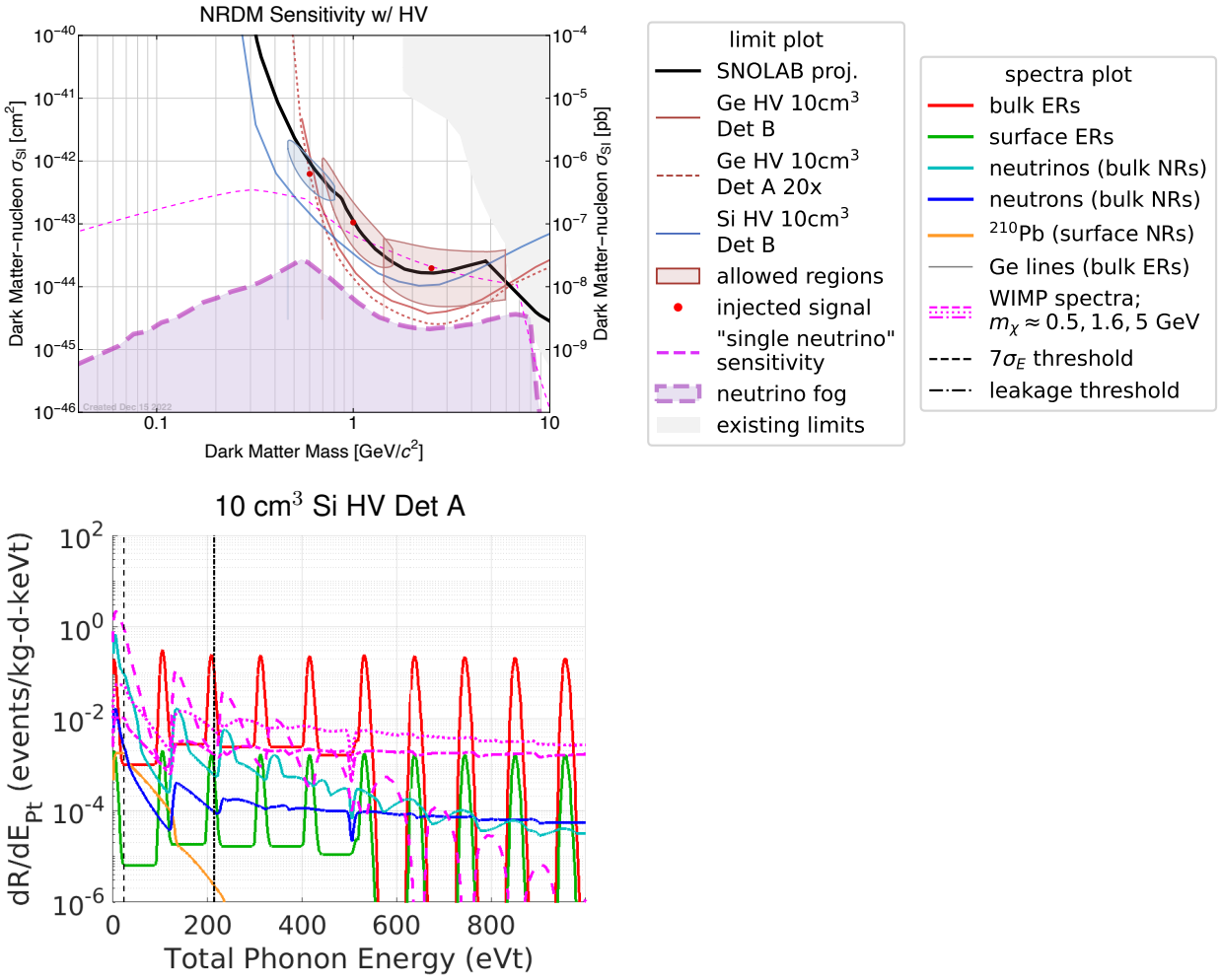

Figure 4.13: **NRDM sensitivity of next-generation HV detectors and related detector spectra.** Top Left: PLR-generated 90% CL exclusion limits and 99.7% CL allowed regions for three future payloads, compared with existing limits, projected SuperCDMS SNOLAB limits, and various theory targets. Bottom Left: background and WIMP signal spectra at the exclusion cross section in phonon energy after all analysis cuts for a two tower projection of Ge HV SNOLAB-sized detectors, along with phonon energy and leakage thresholds.

for ERDM.

In the example signal and background spectra shown in Figure 4.12 Bottom Left, the leakage threshold is now below the second electron-hole pair peak. There are thus three additional electron-hole pair peaks to use for the dark matter search as compared to the SuperCDMS SNOLAB projections, now excluding only the first electron-hole pair peak in the analysis region. These additional peaks provide an extended mass reach for next-generation HV detectors, as lower mass `QEDark` models produce fractionally more signal at lower numbers of electron-hole pairs than at higher ones. It is also interesting to note the sharp detector resolution that is plotted in Figure 4.12 Bottom Left for a `Det C` detector: the electron recoil peaks are very finely resolved around their $N_{\mathrm{eh}} \times (100\,\mathrm{eV} + \epsilon_{\mathrm{eh}})$ positions; likewise, the charge-producing nuclear recoils from neutrinos and neutrons are clearly visible in between the peaks.

For searches of nuclear-recoiling dark matter, lower leakage rates and lower leakage thresholds are useful but not necessarily as critical for enhanced sensitivity. The leakage threshold shown in Figure 4.13 Bottom Left excludes the second electron-hole pair peak, yet the sensitivity projections are still able to reach the neutrino fog. Since HV NRDM searches rely on charge-producing nuclear recoils being at higher energies than equal-charge-producing electron recoils, an HV detector's capacity to resolve in-between-peak events from events in the peaks is key to its sensitivity. $10\,\mathrm{cm}^3$ `Det B` detector scenarios are sufficient in energy resolution to enable sensitivity down to the neutrino fog, as seen in Figure 4.13 Top Left; `Det C` upgrades are not needed. Even a `Det A` upgrade scenario at $20\times$ the nominal exposure is sufficient to reach similar levels of sensitivity; more exposure helps to resolve nuclear recoils from the electron-recoiling background peaks.

Once again, the 0QLEE background is not modeled for these sensitivity projections. Compared to the SuperCDMS SNOLAB projections, these sensitivity projections are more susceptible to degradation from the 0QLEE background due to the lower leakage thresholds that are hypothetically achieved in these `Det` scenarios. It is still possible that the 0QLEE background is suppressed enough by 200 eV by its exponential shape to not have a significant impact on the validity of these projections. In this way, HV detectors are not as crippled in sensitivity by the 0QLEE background compared to 0V detectors because the leakage thresholds are in a regime where the low energy excess may be sub-dominant. In other words, leakage backgrounds are the foremost issue for HV detectors whereas the 0QLEE background is the foremost issue for 0V detectors.

As a reminder of how the leakage background is modeled, it is assumed to scale with the surface area of the electrodes, which would be true if the source of the leakage current is the tunneling of individual charge carriers from the electrodes into the bulk. See Figure 4.9 for the expected level of ionization leakage in the first electron hole pair peak.

4.8 Strategies for detecting sub-GeV dark matter with 0V and HV detectors

Table 4.2 provides a summary matrix for the variety of potential paths to investigate the unexplored parameter space beyond the reach of SuperCDMS SNOLAB. This thesis is only concerned with the 0V and HV detector architectures, but the SuperCDMS Snowmass 2021 contribution includes forecasts for next-generation iZIPs and the yet-to-be-demonstrated piZIP detector architecture. Furthermore, as mentioned earlier, the Snowmass contribution includes projections for sensitivity to dark photon and axion-like particle dark matter in addition to the projections for nuclear-recoiling and electron-recoiling sub-GeV dark matter that are presented in this thesis. There are two main paths toward improved sensitivity that are outlined in this thesis: low-threshold 0V detectors and low-leakage HV detectors.

The first path involves developing better phonon energy resolution and lower threshold detectors. This can be done with TES-based detectors by lowering the critical temperature T_c and improving the transmission of quasiparticles between the phonon-absorbing fins and the TES itself, the projections of which were shared in Section 4.6. Current efforts to this effect have further demonstrated the need for greater shielding of low T_c detectors from pair-breaking IR.

An alternative path toward more sensitive detectors is to use a completely different phonon sensor architecture. Kinetic inductance detectors (KIDs), quantum capacitance detectors (QCDs), and superconducting quasiparticle amplifying transmons (SQUATs) are three possible future sensor architectures that may pave the way toward meV energy thresholds. Fundamentally, they are all sensitive to the quasiparticle number in the sensor; absorbed energy can be computed in terms of the number of broken Cooper pairs, which is the fundamental quanta of superconducting energy. In contrast, TESs rely on the conversion from a temperature change to absorbed energy and, therefore,

	nuclear-recoiling dark matter	electron-recoiling dark matter
0V	• limited by 0QLEE • 1 decade of mass reach for 2 decades of threshold improvement • need better phonon sensors	• limited by 0QLEE • no benefit beyond sub-bandgap thresholds • sharp theory targets
HV	• potentially limited by 0QLEE • large uncertainty in yield function • theoretically benefits from spectral discrimination of nuclear recoils from electron recoils and leakage backgrounds	• current exclusion curves are exposure-limited and thus motivate 10s kg·year exposures • limited by leakage backgrounds and/or sub-gap IR for one to a few e^-h^+ pairs • potentially limited by 0QLEE • sharp theory targets

Table 4.2: **Challenges and opportunities for next-generation 0V and HV detectors in search of sub-GeV dark matter.**

are dissipative in a way that may limit long-term gains.

A second path toward unexplored parameter space involves using existing single-charge sensitive detectors. These detectors have significant overlap in projected sensitivity with other charge-sensing technologies, most notably charge-coupled devices and their Skipper variety; see Section 3.3. The path toward increased sensitivity involves tens of kg·year exposures as well as background rates below 0.01 dru. Single-charge backgrounds are also a significant impediment toward increased sensitivity, and the mechanism for their production needs to be better understood. Most importantly, single-charge sensitive detectors provide the clearest path forward to testing those dark matter models for which there is a possible mechanism to produce the appropriate relic abundance in the early universe.

The zero charge low energy excess (0QLEE) has the potential to limit the sensitivity of any low-mass dark matter search that uses a phonon-mediated detector. It is particularly limiting for 0V detectors, the main advantage of which is their ability to probe at low energies where the 0QLEE is dominant. Current evidence points to stress-induced events as a dominant contributor to this excess, but the rate of events still significantly exceeds other standard particle backgrounds.

Chapter 5

HVeV Run 4 NRDM search with background subtraction of the low energy excess: a prototype SuperCDMS SNOLAB analysis

This chapter is temporarily embargoed.

5.1 Background information, key concepts, and motivation

This chapter is temporarily embargoed.

Figure 5.1: **HVeV Run 3 versus HVeV Run 4.**

This chapter is temporarily embargoed.

Figure 5.2: **The zero charge low energy excess as observed in various experiments.**

This chapter is temporarily embargoed.

Figure 5.3: **Ionization yield models for nuclear recoils and HV NRDM signal models.**

This chapter is temporarily embargoed.

5.2 HVeV Run 4

This chapter is temporarily embargoed.

Figure 5.4: **HVeV Run 4 detector payload.**

This chapter is temporarily embargoed.

Figure 5.5: **Raw event rates in HVeV Run 4.**

This chapter is temporarily embargoed.

5.3 Live-time cuts

This chapter is temporarily embargoed.

Figure 5.6: **Example events identified by the anti-coincidence-based live-time cut.**

5.4 Data-quality cuts

This chapter is temporarily embargoed.

Figure 5.7: **Cs-137 source data: χ^2 versus pulse amplitude and event types.**

This chapter is temporarily embargoed.

Figure 5.8: **Zoom-in of the post-cut χ^2 versus amplitude-reconstructed energy of HV and 0V Cs-137 data.**

This chapter is temporarily embargoed.

5.5 Comparing the 0V and HV spectra

This chapter is temporarily embargoed.

Figure 5.9: **Post-cut unblinded background spectra of the HVeV Run 4 detectors.**

Figure 5.10: **Kolmogorov-Smirnov tests of the 0V and HV data in the ROI.**

This chapter is temporarily embargoed.

5.6 Background estimation for NFC1

This chapter is temporarily embargoed.

Figure 5.11: **Maximum-likelihood-based background model parameter estimation for HVeV Run 4.**

5.7 Constraining the DM-nucleon cross section

This chapter is temporarily embargoed.

This chapter is temporarily embargoed.

Figure 5.12: **HVeV Run 4 NRDM constraints.**

Part III

Kinetic inductance detectors for dark matter

Preface

In this final part of my thesis, I explore kinetic inductance detectors (KIDs) as a new and potentially more sensitive phonon sensor. The detector architecture is termed a KID-based phonon-mediated (KIPM) detector. In Chapter 6, I first explain how kinetic inductance detectors work, going step by step from the absorbed energy in a superconducting film to the modulation of some resonator-coupled readout signal. This chapter is intended to be an introductory guide for kinetic inductance detectors and so predominantly draws on reference texts regarding superconductivity and microwave engineering as well as various seminal works on kinetic inductance detectors. The one exception is Section 6.7, which presents a new parametrization that I developed to understand the nonlinear response of kinetic inductance detectors to large deviations in resonator parameters.

In Chapter 7, I work through the different sources of noise in kinetic inductance detectors, first providing calculations for each noise's expected contribution to the resolution on absorbed energy, and then reporting their measured contribution to the resolution from various experiments. I was either the sole or primary data taker and data analyzer for all of the reported measurements, all but one of which was performed at Caltech.

In Chapter 8, I examine the energy performance of three different KIPM detectors. The first detector, known as DMLE2, was originally studied in Moore et al. (2012) and Moore (2012); they were the first publications on a KID-based phonon-mediated detector. The energy performance of the second detector, known as OW200127, was measured in NEXUS at Fermilab and published in Temples et al. (2024). As mentioned in the foreword on published content, I designed and fabricated this detector, studied its noise performance at Caltech, developed the LED data-taking protocol for measurement of energy resolution, and provided close supervision to the experiment at NEXUS that measured the energy resolution. The third and final detector that is examined in Chapter 8 is known as B240103, which was fabricated by Bruce Bumble at the Jet Propulsion Laboratory. I was the lead member of the Caltech data-taking team that measured the energy performance of this detector, using the same protocol as was used for OW200127. A notable improvement to the analysis that I developed for B240103 was the implementation of a nonlinear pulse coordinate system, allowing for calibration to the energy in the resonator across a much broader range of energy. With the energy performance of these three detectors as data points, I then build an empirical model in Chapter 8 to describe the phonon collection efficiency for KIPM detectors.

Chapter 6

Superconducting microwave resonators: signal calibration and responsivity

Kinetic inductance detectors are made with superconducting microwave resonators. During operation of a kinetic inductance detector, the resonator is probed with a microwave readout tone at the resonant frequency f_r. Modulations in the complex transmission δS_{21} of the resonant tone can be used to measure the energy absorbed $E_{\rm abs}$ by the resonator. The physical processes that govern $E_{\rm abs}$ becoming δS_{21} are the topic of the first section of this chapter. This discussion also provides the conversion of the measured δS_{21} into the desired $E_{\rm abs}$. In broad strokes, the conversion can be conceptualized in the following way:

absorbed energy	quasiparticle density	complex conductivity	surface impedance	resonator parameters	complex transmission
$E_{\rm abs}$	$n_{\rm qp}$	$\sigma = \sigma_1 - i\sigma_2$	$Z_s = R_s + i\omega L_s$	Q_r, f_r	S_{21}

6.1 Absorbed energy → quasiparticles: the Cooper pair

Below the critical temperature of a superconductor, the majority of free electrons enter a paired state known as the Cooper pair. An unpaired free electron is referred to as a quasiparticle. The energy required to break a Cooper pair is 2Δ, where Δ is the superconducting bandgap. A broken Cooper pair increases the quasiparticle population $N_{\rm qp}$ by two. Thus, the average energy to create one quasiparticle is Δ. Kinetic inductance detectors respond to the density of quasiparticles $n_{\rm qp}$, so we divide by the superconducting volume $V_{\rm sc}$. We can write

$$\boxed{\frac{dn_{\rm qp}}{dE_{\rm abs}} = \frac{1}{V_{\rm sc}\Delta}.} \tag{6.1}$$

6.2 Quasiparticles → complex conductivity: Mattis-Bardeen theory

Mattis-Bardeen theory (Mattis et al., 1958) is concerned with the response of superconducting electronic states upon application of an electromagnetic wave with frequency ω, an idea succinctly expressed as Ohm's law: $\mathbf{J} = \sigma\mathbf{E}$, where $\mathbf{J}$ is the current density, $\mathbf{E}$ is the electric field, and

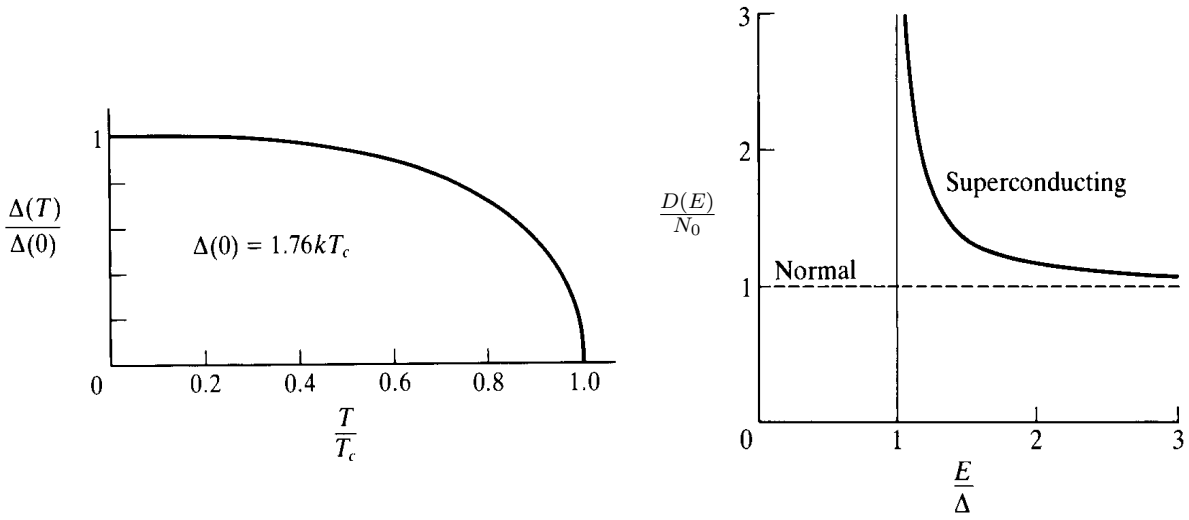

Figure 6.1: **Basic machinery of Mattis-Bardeen theory.** Left: The dependence of Δ on temperature. Figure is from Tinkham (2004). Right: $D(E)$ vs E for a superconductor. The flat density of states for a normal conductor is also noted. Figure is modified from Tinkham (2004).

$\sigma = \sigma_1 - i\sigma_2$ is the complex conductivity. Both $\mathbf{J}$ and $\mathbf{E}$ are taken to be phasors with time dependence $e^{i\omega t}$. The real component σ_1 corresponds to in-phase, dissipative oscillations in $\mathbf{J}$ with respect to $\mathbf{E}$, and the imaginary component σ_2 corresponds to the out-of-phase and reactive oscillatory response.

Mattis-Bardeen theory builds on BCS theory and invokes the following terms:

$$D(E) = N_0 \frac{E}{\sqrt{E^2 - \Delta^2}}, \quad F(\Delta, E, E') = \frac{1}{2}\left(1 \mp \frac{\Delta^2}{EE'}\right), \quad f(E) = \frac{1}{1 + e^{E/kT}}. \tag{6.2}$$

- Δ is an order parameter that characterizes the phase transition of a metal from normal conductor to superconductor. The transition occurs at some critical temperature T_c. In the case of superconductivity, Δ is the gap that appears in the density of states. Above T_c, the order parameter is 0. At T_c, the order parameter follows a $\sqrt{T_c - T}$ curve and its derivative diverges, marking a phase transition that is characteristic of mean-field theories. Below T_c, Δ can be numerically computed in a self-consistent way such that for a particular value of Δ there is a corresponding n_{qp}, i.e., if there are fewer quasiparticles then Δ must be larger. As a result, Δ asymptotically reaches some Δ_0 in the limit where n_{qp} and T approach 0. This behavior is plotted in Figure 6.1 Left. Details about this self-consistency calculation can be explored further in Tinkham (2004). One result of the calculation is that $\Delta_0 \approx 1.76kT_c$.
- $D(E)$ is the quasiparticle density of states and is plotted in Figure 6.1 Right. N_0 is the single-spin density of electron states at the Fermi energy. There is a 2Δ energy gap between the lowest-energy quasiparticles and the Cooper pair energy. The function peaks near the gap and asymptotes to 1 away from the gap.
- $F(\Delta, E, E')$ is a coherence term that ranges from 0 to 1 and results from the wavelike nature of quasiparticles and Cooper pairs: 0 is destructive interference and 1 is constructive interference. For $E, E' \gg \Delta$, the gap becomes irrelevant and the wavelike properties are diminished: $F = \frac{1}{2}$. For electromagnetic absorption, $E' = E + \hbar\omega$.

  - The negative sign in the expression for $F(\Delta, E, E')$ is used for processes that are even under time reversal of electronic states; this case is referred to as case I in the BCS and Mattis-Bardeen literature. The positive sign is used for processes that are odd under time reversal, and these situations are referred to as case II. The canonical example of a case II process in a superconductor is nuclear-spin relaxation by interaction with quasiparticles; the spin interaction is the source of the odd parity under time reversal.
- $f(E)$ is the standard Fermi-Dirac distribution for spin-$\frac{1}{2}$ particles.

The thermal population of quasiparticles n_{qp} can be computed from these terms and evaluated in the limit where $kT \ll \Delta$:

$$n_{\text{qp}} = 4\int_{\Delta}^{\infty} D(E)f(E)dE \quad \overset{kT\ll\Delta}{=} \quad 2N_0\sqrt{2\pi kT\Delta_0}e^{-\Delta_0/kT}. \tag{6.3}$$

Equation 6.3 is plotted in Figure 6.2 with $N_0 = 1.72\times10^{10}$ eV$^{-1}\mu$m^{-3} and $\Delta_0 = 180\,\mu$eV, values for aluminum. Because we are interested in detecting perturbations in n_{qp}, the goal of Mattis-Bardeen theory as it applies to kinetic inductance detectors is to derive expressions for the dependence of the complex conductivity on n_{qp}.

With the basic machinery described, we return to the original question of Mattis-Bardeen theory: how does a superconductor respond to an external electromagnetic field. There are two natural regimes of Mattis-Bardeen theory, given a photon with energy $\hbar\omega$ and a superconducting bandgap of 2Δ: $\hbar\omega < 2\Delta$ and $\hbar\omega > 2\Delta$. Kinetic inductance detectors are generally probed at frequencies well below the gap frequency f_Δ: $f_\Delta = 2\Delta/h \approx 74\text{ GHz}\times(T_c/1\text{ K})$. $\hbar\omega < 2\Delta$ is assumed throughout this section.

The real component σ_1 of the conductivity can be interpreted as a measure of electromagnetic absorption. σ_1 signifies the capacity for in-phase power from an electromagnetic field to be delivered to the conductive material. For the sub-gap photons that are used to probe kinetic inductance detectors, only the quasiparticles can be excited into higher energy states. σ_1 may thus be computed from quasiparticle scattering rates, which depend on the relative occupation of $D(E)$ at the initial and final energies of the quasiparticle; the essential building blocks of the calculation are underscored with brackets below. The reported quantity is a ratio of σ_1 to the normal state conductivity σ_n, which may be computed with a similar integral as below and setting $\Delta = 0$.[1]

$$\frac{\sigma_1}{\sigma_n} = \frac{2}{\hbar\omega}\int_{\Delta}^{\infty} \underbrace{\frac{E}{\sqrt{E^2-\Delta^2}}}_{D(E)} \underbrace{\frac{E+\hbar\omega}{\sqrt{(E+\hbar\omega)^2-\Delta^2}}}_{D(E+\hbar\omega)} \underbrace{\left(1+\frac{\Delta^2}{E(E+\hbar\omega)}\right)}_{\text{coherence term}} \underbrace{(f(E)-f(E+\hbar\omega))}_{\text{F-D statistics}} dE \tag{6.4}$$

Observations:

- The integral is performed only on the quasiparticles, from Δ to ∞. In general, there is another integral from $-\Delta-\hbar\omega$ to Δ for direct Cooper pair to quasiparticle creation. It is 0 for $\hbar\omega < 2\Delta$.
- The coherence term is for a case II interaction (positive sign) because the interaction Hamiltonian for electromagnetic absorption is proportional to $\mathbf{p}\cdot\mathbf{A}$,[2] which is odd under time reversal.
- The net scattering rate is computed as (forward scattering rate − backward scattering rate). A forward scattering rate from $E_1 \to E_2$ takes the form $D(E_1)f(E_1)\times D(E_2)\,(1-f(E_2))$; it depends on the occupation of the initial state and the vacancy of the final. Subtraction of the backward scattering from $E_2 \to E_1$ results in $\propto f(E_1)-f(E_2)$.

[1] There is a factor of N_0 and other material parameters that cancel out between numerator and denominator. See Tinkham (2004).

[2] This interaction term comes from the kinetic term in the Lagrangian for a charged particle in an electromagnetic field, which has a modified canonical momentum $(\mathbf{p}-e\mathbf{A})^2/2m$.

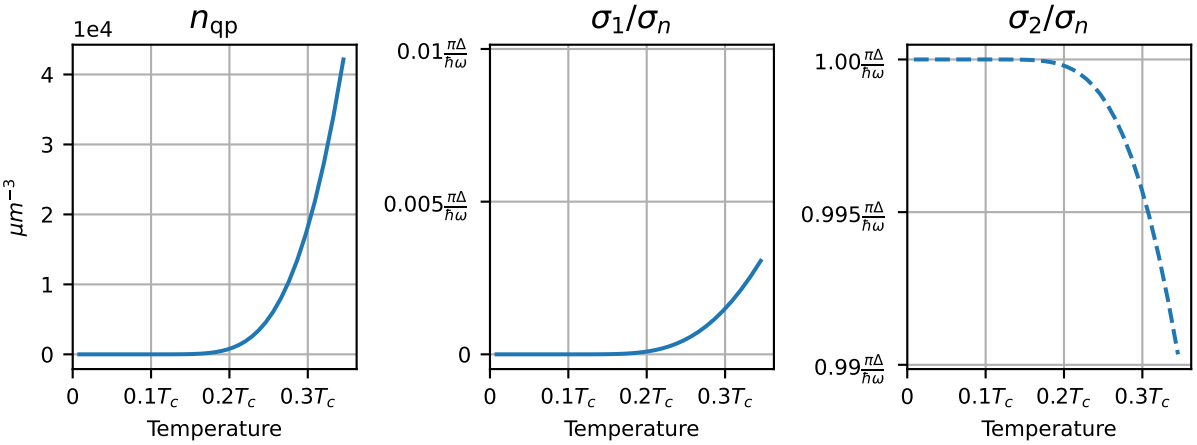

Figure 6.2: **Quasiparticle density and complex conductivity versus temperature as a fraction of $T_c \approx \Delta/1.76k$.** The span across the y-axis is identical across the right two plots for ease of comparison.

- When $\Delta = 0$, the integrand simplifies to $f(E) - f(E + \hbar\omega)$.

The imaginary component σ_2 of the conductivity represents the strength of out-of-phase oscillations to be induced by an external electromagnetic wave. σ_2 can be computed in a similar way to σ_1 but has a few key differences:

$$\frac{\sigma_2}{\sigma_n} = \frac{1}{\hbar\omega}\int_{\Delta-\hbar\omega}^{\Delta} \underbrace{\frac{E}{\sqrt{\Delta^2 - E^2}}}_{D(E)} \underbrace{\frac{E+\hbar\omega}{\sqrt{(E+\hbar\omega)^2 - \Delta^2}}}_{D(E+\hbar\omega)} \underbrace{\left(1 + \frac{\Delta^2}{E(E+\hbar\omega)}\right)}_{\text{coherence term}} \underbrace{(1 - 2f(E+\hbar\omega))}_{\text{F-D statistics}} dE, \tag{6.5}$$

Observations:

- Equation 6.5 is reproduced as written from Mattis et al. (1958). Zmuidzinas (2012) provides another expression for Equation 6.5 that integrates from Δ to $\Delta + \hbar\omega$ after applying a change of variables $E \to E + \hbar\omega$. Under that change of variables, an interpretation for Equation 6.5 is that it is an integral over virtual quasiparticle states that range in energy from Δ to $\Delta + \hbar\omega$.
  - Note the bounds of the integral and the first $D(E)$ term. The integral is performed over states *in the gap!* A speculative interpretation of the integral is that a photon has a non-zero probability of spontaneously creating quasiparticles above the gap by interacting with virtual states inside the gap. This process is similar to a photon that drives the stimulated emission of an excited electron in an atom, but in this case the outgoing quasiparticle is out of phase with the initial photon.
- Because the photon interacts with virtual states, the Fermi-Dirac term only cares about the available states at the final quasiparticle energy: $1 - f$. The photon makes 2 quasiparticles, hence the factor of 2.
- The different Fermi-Dirac terms between σ_1 and σ_2 lead to an important qualitative difference: $\sigma_1 \to 0$ as $T \to 0$, whereas $\sigma_2 \to$ constant.

If we consider the limit $\hbar\omega \ll \Delta$ and $kT \ll \Delta$, then these integrals can be done analytically

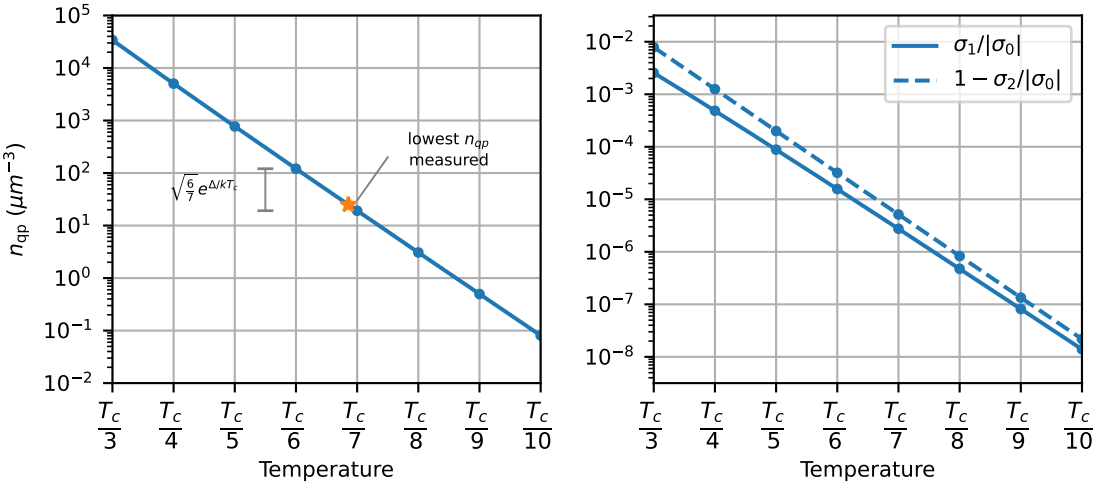

Figure 6.3: **Quasiparticle density and complex conductivity versus a reciprocal temperature axis.** Fractional deviations from $|\sigma_0| = \pi\Delta/\hbar\omega$ are computed. The intersections with the vertical grid lines have been marked as a visual guide. The vertical spacing between two adjacent markers has been pointed out to emphasize the exponential scaling of n_{qp} with $1/T$.

and have been reported in various documents; see Gao (2008). They are

$$\frac{\sigma_1}{\sigma_n} = \frac{4\Delta}{\hbar\omega} e^{-\Delta/kT} \sinh(x) K_0(x) \quad \text{and} \quad \frac{\sigma_2}{\sigma_n} = \frac{\pi\Delta}{\hbar\omega}\left(1 - 2e^{-\Delta/kT} e^{-x} I_0(x)\right), \tag{6.6}$$

where $x = \hbar\omega/2kT$ and I_n and K_n are nth order modified Bessel functions of the first and second kinds. It is useful to calculate the low temperature limiting value of the conductivity $\sigma_0 = \lim_{T\to 0}\sigma(T)$. Taking this limit in Equation 6.6, we find that $\sigma_0 = -i(\pi\Delta/\hbar\omega)\sigma_n$; it is completely reactive.

Equation 6.6 is plotted in Figure 6.2. They show qualitatively similar exponential responses to temperature as n_{qp}. The response of σ_2 as a fraction of σ_n is larger than that of σ_1, visually aided by the equal reaches across the y-axis in Figure 6.2 Middle and Right. Furthermore, the deviations of σ_1 and σ_2 have been expressed as a fraction of $\pi\Delta/\hbar\omega$, which provides the scale for the deviations.

It is interesting to plot Equation 6.3 and Equation 6.6 on a logarithmic y-axis and a reciprocal x-axis as is done in Figure 6.3. Doing so illustrates the approximately linear relationship between the logarithmic quantities on the y-axis with the reciprocal temperature, e.g., $\ln n_{\text{qp}} \propto 1/T$. The key point of Figure 6.3 is that the other temperature-dependent terms in the expressions for n_{qp}, σ_1/σ_n, and σ_2/σ_n are small compared to $\exp(-\Delta/kT)$.

The lowest n_{qp} that has been robustly measured in the literature to our knowledge is from de Visser et al. (2011)[3] and has been pointed out in Figure 6.3 Left at $25\,\mu m^{-3}$. Measurements of n_{qp} bottom out at whatever the quiescent quasiparticle population is, as shown in Figure 4 of de Visser et al. (2011). Likewise, σ_1/σ_n and σ_2/σ_n (as well as $\delta f_r/f_r$ and $\delta(1/Q_i)$, which we introduce later) cannot actually be measured below a certain temperature because σ_0, the conductivity at zero temperature, cannot be measured in experiment. So, in practice, it is instead necessary to

[3] Temples et al. (2024) infers a quiescent quasiparticle population of $18.5\,\mu m^{-3}$, but this measurement is not as robust as de Visser et al. (2011). See Section 7.2.

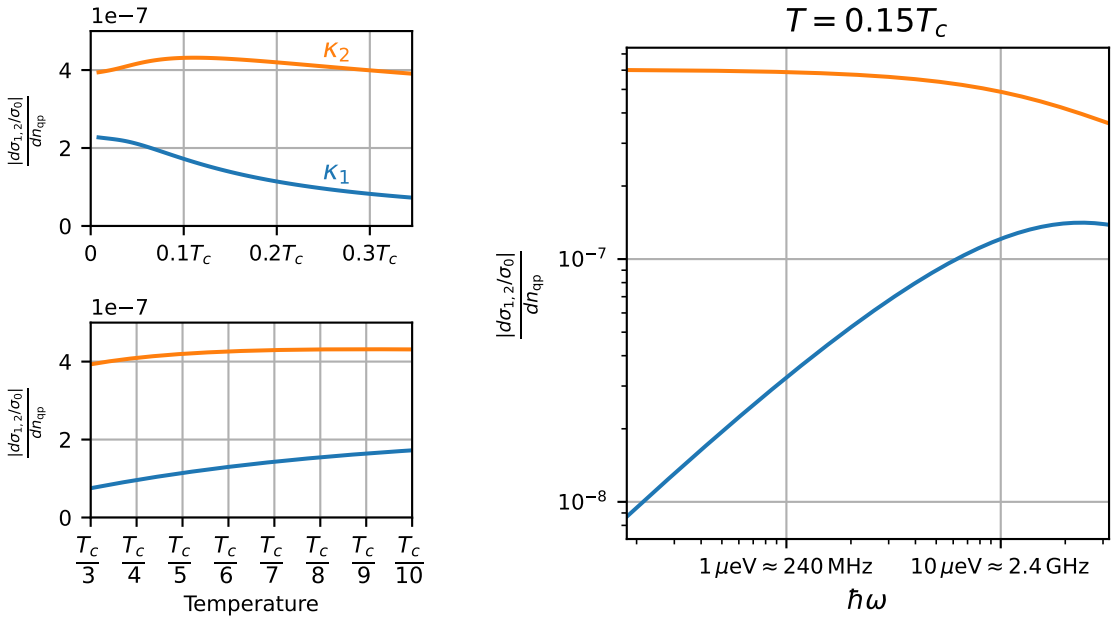

Figure 6.4: **The Mattis-Bardeen equations for κ_1 and κ_2 as a function of temperature and photon frequency.** Top Left: κ_1 and κ_2 versus a linear temperature axis. Bottom Left: κ_1 and κ_2 versus a reciprocal temperature axis. Right: κ_1 and κ_2 as they depend on the electromagnetic wave frequency ω, assuming $\Delta = 0.18\,\text{meV}$ as it is for aluminum.

define some baseline reference value σ_{ref} (or $f_{r,\text{ref}}$ or $Q_{i,\text{ref}}$) that is measured at the lowest possible temperature of the experiment. Deviations from the reference quantity may instead be plotted.

We desire to calculate the derivatives of Equation 6.6 with respect to n_{qp} (i.e., how the y-axis of Figure 6.3 Right depends on the y-axis of Figure 6.3 Left), which involves the subtlety that σ and n_{qp} both depend on temperature under certain assumptions. Gao (2008) describes two methods for computing this integral: (1) if n_{qp} is completely specified by the thermal population as determined by Equation 6.3, then $d\sigma/dn_{\text{qp}} = \partial\sigma/\partial T \div \partial n_{\text{qp}}/\partial T$ by the chain rule; and (2) if n_{qp} is independent of T, as would be the case if the n_{qp} population is dominated by external pair-breaking sources, then the expression for n_{qp}, Equation 6.3, can be substituted into Equation 6.6, and the partial derivative $\partial\sigma/\partial n_{\text{qp}}$ can be taken as the total derivative that is desired. In both cases, the temperature dependence of Δ must also be included. Gao (2008) shows that these two methods lead to very similar numerical results and thus draws an equivalence between thermal quasiparticles and excess quasiparticles. We report the latter result below:

$$\frac{1}{\sigma_n}\frac{d\sigma_1}{dn_{\text{qp}}} = \frac{1}{N_0\hbar\omega}\sqrt{\frac{2\Delta}{\pi kT}}\sinh(x)K_0(x) \quad \frac{1}{\sigma_n}\frac{d\sigma_2}{dn_{\text{qp}}} = -\frac{\pi}{2N_0\hbar\omega}\left(1+\sqrt{\frac{2\Delta}{\pi kT}}e^{-x}I_0(x)\right). \tag{6.7}$$

The previous equations may then be expressed in terms of fractional deviations from $|\sigma_0|$, allowing us to define

$$\boxed{\kappa_1 \equiv \frac{d\sigma_1/|\sigma_0|}{dn_{\text{qp}}} = \frac{1}{\pi N_0}\sqrt{\frac{2}{\pi kT\Delta}}\sinh(x)K_0(x) \quad \kappa_2 \equiv \frac{d\sigma_2/|\sigma_0|}{dn_{\text{qp}}} = -\frac{1}{2N_0\Delta}\left(1+\sqrt{\frac{2\Delta}{\pi kT}}e^{-x}I_0(x)\right).} \tag{6.8}$$

We remind the reader that $x = \hbar\omega/2kT$ and that K_0 and I_0 are modified Bessel functions. Equation 6.8 is plotted in Figure 6.4. Only the absolute value is plotted; the κ_2 response is technically negative. The key observations about the plot are (1) the κ_2 response is greater than the κ_1 response, as suggested by Figure 6.2 and Figure 6.3, and (2) κ_1 and κ_2 do not change much over the typical operating temperatures of a kinetic inductance detector. If this second point were not true, kinetic inductance detectors would be much more difficult to calibrate. Equation 6.8 is the ultimate result of this section and can be interpreted as the fractional response of the conductivity to changes in the quasiparticle density.

Also plotted in Figure 6.4 Right is the dependence of $|d\sigma_{1,2}/\sigma_0|/dn_{\mathrm{qp}}$ on the applied electromagnetic wave frequency ω. The κ_2 response plateaus at low frequency. The κ_1 response goes to 0, which is expected for the dissipative response to a low-frequency electromagnetic wave. At the right edge of the plot, the inequality $\hbar\omega \ll \Delta$ no longer holds.

6.3 Complex conductivity → surface impedance: a lesson in penetration depth

The next physical quantity for conversion of E_{abs} into δS_{21} is the superconducting film's surface impedance Z_s, which governs its electrical response and is usefully parametrized as $R_s + i\omega L_s$, the sum of impedances due to surface resistance and surface inductance. The relationship between the conductivity and surface impedance of a material is easiest to understand when Ohm's law is local: $\mathbf{J} = \sigma\mathbf{E}$. For a superconductor, the relevant length scales that determine locality are the penetration depth of the magnetic field λ from the Meissner effect, the coherence length of the Cooper pairs ξ, and the mean free path of the charge carriers ℓ. Locality occurs when $\ell \ll \lambda$ and $\ell \ll \xi$. In other words, the charge carriers scatter on a small enough scale that the long-length-scale coherence effects of superconductivity are suppressed. The result is that the normal forms for impedance and resistance in terms of σ can be adopted with the replacement of σ by $\sigma_1 - i\sigma_2$.

Scenario 1: thin film

We start with the most basic case: a thin superconducting film of dimension $L \times W \times t$, where the thickness $t \ll \lambda$, thus enforcing $t \approx \ell \ll \lambda,\xi$. The fact that the thickness is much smaller than the penetration depth implies that the fields (and thus the current) are uniform across the film thickness rather than decaying with distance from the surface. The resistance adopts the familiar form

$$R = \frac{\rho L}{A} = \frac{\rho L}{Wt} = R_s \frac{L}{W}; \quad R_s = \frac{\rho}{t} = \frac{1}{\sigma t}, \tag{6.9}$$

where R_s is the sheet resistance and carries units of Ω/square. For the rectangular geometry in this example, the number of squares can be calculated with L/W, which can be recognized as the aspect ratio of the film. To obtain the desired relation between Z_s and σ, we promote $R_s \to Z_s$ and substitute $\sigma \to \sigma_1 - i\sigma_2$, with the assumption that the sheet inductance of the film is due solely to the complex conductivity. The final equation is

$$Z_s = R_s + i\omega L_s = \frac{1}{\sigma t} = \frac{1}{(\sigma_1 - i\sigma_2)t}. \tag{6.10}$$

As with the complex conductivity, we are interested in fractional perturbations about the low temperature impedance $Z_{s,0}$ and the low temperature conductivity σ_0, recalling that $\sigma_1 \to 0$ at low temperatures:

$$\frac{\delta Z_s}{Z_{s,0}} = -\frac{\delta\sigma}{\sigma_0}. \tag{6.11}$$

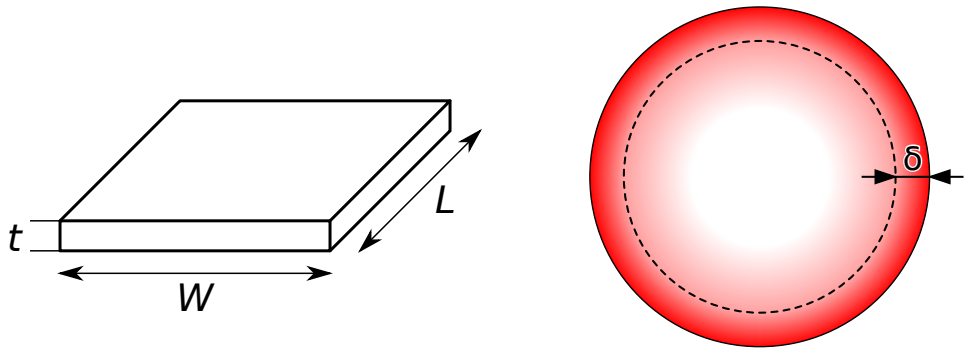

Figure 6.5: **Schematics of thin film and bulk superconductors.** Left: thin film superconductor. Right: bulk superconductor with penetration depth δ. In the text, penetration depth is referred to as λ. Both images are from the Wikimedia commons.

Scenario 2: bulk superconductor in the local limit

If we remain in the local regime but now allow for the incident **E** and **B** fields to decay over a length scale given by λ, then we start from the formula for the intrinsic impedance η of an EM wave inside of a good conductor:

$$\eta = (1+i)\frac{1}{\sigma\lambda} = (1+i)\frac{1}{\sigma}\sqrt{\frac{\omega\mu\sigma}{2}} = \sqrt{\frac{i\omega\mu}{\sigma}}, \tag{6.12}$$

where ω is the frequency of the incident wave and μ is the magnetic permeability of the material. In the second equality, the expression for the skin depth in a good conductor is used. There is an interesting conceptual link between the expression for R_s in Equation 6.9 and the first expression for η above: they are equivalent upon exchange of $t \leftrightarrow \lambda$, modulo the factor of $(1+i)$; the surface layer through which the field penetrates can be interpreted as the effective thickness of a thin conducting film in which the current flows. A change in σ includes a change in λ, which conspires to produce an $\eta \propto \sigma^{-1/2}$ relationship. As before, we take η to be Z_s and substitute $\sigma \to \sigma_1 - i\sigma_2$:

$$Z_s = \sqrt{\frac{i\omega\mu}{\sigma}} = \sqrt{\frac{i\omega\mu}{\sigma_1 - i\sigma_2}}. \tag{6.13}$$

The fractional response of the impedance about the low temperature impedance is now suppressed due to the square root:

$$\frac{\delta Z_s}{Z_{s,0}} = -\frac{1}{2}\frac{\delta\sigma}{\sigma_0}. \tag{6.14}$$

Scenario 3: a non-local superconductor and the anomalous skin effect

When $\xi \gg \lambda$ and $l \gg \lambda$, then the current is coherent (does not suffer scattering) over a much larger length scale than the decay length of the fields, so the current at a given point can no longer be related only to the field at the point, i.e., the superconductor is no longer in the local limit. Instead, we must adopt a non-local version, first suggested by Chambers and Pippard and motivated by the phenomenological London equations, $\mathbf{J} = -1/\lambda^2\mathbf{A}$; see Tinkham (2004) and Gao (2008).

The fractional response of the impedance about the low temperature impedance is now further suppressed over the local superconductor case due to the anomalous skin effect:

$$\frac{\delta Z_s}{Z_{s,0}} = -\frac{1}{3}\frac{\delta\sigma}{\sigma_0}. \tag{6.15}$$

Summary

The results of the three scenarios are collected here:

$$\frac{\delta Z_s}{Z_{s,0}} = \gamma\frac{\delta\sigma}{\sigma_0}; \quad \gamma = \begin{cases} -1 & \text{thin film, local;} \\ -\frac{1}{2} & \text{thick film, local} \\ & \text{i.e., “dirty” limit;} \\ -\frac{1}{3} & \text{thick film, non-local} \\ & \text{i.e., extreme anomalous limit.} \end{cases} \tag{6.16}$$

The main takeaway is that, of these three scenarios, thin films are the most responsive to fractional changes in the complex conductivity.

The response of the surface impedance can be broken down into the dissipative and reactive components.

$$\begin{gathered} \frac{\delta Z_s}{Z_{s,0}} = \frac{\delta R_s + i\omega\delta L_s}{i\omega L_s} = \gamma\frac{\delta\sigma_1 - i\delta\sigma_2}{\sigma_0} \Rightarrow \\ \boxed{\frac{\delta R_s}{\omega L_s} = \gamma\frac{\delta\sigma_1}{|\sigma_0|} \quad \text{and} \quad \frac{\delta L_s}{L_s} = \gamma\frac{\delta\sigma_2}{|\sigma_0|}.} \end{gathered} \tag{6.17}$$

In the low temperature limit, it is assumed that R_s and σ_1 are both 0. The final form of Equation 6.17 has stripped all factors of i and has had the absolute value taken where appropriate so that all quantities are real.

6.4 Surface impedance → resonator parameters: the kinetic inductance fraction

The essential ingenuity of a kinetic inductance detector is that changes to the surface impedance of a superconductor (which arise from changes to the complex conductivity which in turn arise from quasiparticle density) can be measured by fashioning a resonator out of the superconductor and then measuring changes to the resonant frequency f_r and internal quality factor Q_i of that resonator. For a resonator fashioned out of an RLC circuit, these quantities are

$$f_r = \frac{1}{2\pi}\frac{1}{\sqrt{LC}} \quad \text{and} \quad Q_i = \frac{2\pi f_r L}{R} = \frac{1}{2\pi}\frac{1}{f_r RC}. \tag{6.18}$$

Following the theme of the rest of this chapter, we explore fractional perturbations to these quantities[4] as a result of $\delta Z_s = \delta R_s + i\omega\delta L_s$, reproducing derivations that are found in Siegel (2016):

[4] $1/Q_i$ is already a fractional quantity; it represents the fraction of stored energy lost in the resonator over a radian of oscillation.

$$\begin{aligned}\frac{\delta f_r}{f_r} &= \frac{\sqrt{L}}{\sqrt{L+\delta L_s}} - 1 \\ &= \left(1 + \frac{\delta L_s}{L}\right)^{-1/2} - 1 \\ &= -\frac{1}{2}\frac{\delta L_s}{L} \end{aligned} \tag{6.19}$$

$$\begin{aligned}\delta\frac{1}{Q_i} &= \frac{R+\delta R_s}{\omega(L+\delta L_s)} - \frac{R}{\omega L} \\ &= \frac{R+\delta R_s}{\omega L}\left(1 - \frac{\delta L_s}{L}\right) - \frac{R}{\omega L} \\ &= \frac{\delta R_s}{\omega L} - \frac{R\delta L_s}{\omega L^2} - \frac{\delta R_s \delta L_s}{\omega L^2} \\ &= \frac{\delta R_s}{\omega L}. \end{aligned} \tag{6.20}$$

In the above derivation, we have used the fact that $\delta L_s/L$ and $R/\omega L$ are small, only keeping linear terms of these quantities in the final result.

The perturbations are expressed as a fraction of the total inductance L. The total inductance includes a component L_{geom} that depends on the geometry, allowing us to define α, the fraction of the total inductance that is due to the surface inductance:

$$L = L_{\text{geom}} + L_s; \quad \alpha \equiv \frac{L_s}{L}. \tag{6.21}$$

α is usually referred to as the kinetic inductance fraction. With this definition, we can rewrite 6.19 and 6.20 as

$$\boxed{\delta\frac{1}{Q_i} = \alpha\frac{\delta R_s}{\omega L_s} \quad \text{and} \quad \frac{\delta f_r}{f_r} = -\frac{1}{2}\alpha\frac{\delta L_s}{L_s}.} \tag{6.22}$$

6.5 Resonator parameters $\rightarrow \delta S_{21}$: the RF responsivity

S_{21} is measured across a feedline that is inductively or capacitively coupled to a microwave resonator. The resonator has some internal quality factor Q_i. The coupling to the feedline is a path for energy to leave the resonator as well as for energy to enter; therefore, there is a coupling quality factor Q_c associated with this loss of energy. Q_c is associated with non-dissipative energy loss, whereas Q_i is associated with dissipative energy loss. The loaded quality factor Q_r that is associated with a resonator embedded into the readout circuitry is thus

$$\frac{1}{Q_r} = \frac{1}{Q_i} + \frac{1}{Q_c}. \tag{6.23}$$

For some additional intuition regarding this equation, recall that quality factors in general may be defined as $2\pi \times$ (energy stored in the resonator $\div$ energy lost per cycle). Equation 6.23 can be interpreted as a statement of energy conservation: the total energy lost per cycle is equal to the energy that leaks out of the resonator to the feedline plus the energy lost to dissipation within the resonator itself. For further discussion of loaded and unloaded quality factors, please see Chapter 5 of Pozar (2005).

The quality factor for a parallel RLC circuit can be derived from the following calculation, where we use $\omega_r = 1/\sqrt{LC}$ and $\omega = \omega_r + \Delta\omega$, with $\Delta\omega \ll \omega_r$:

$$\begin{aligned}\frac{1}{Z_{\text{in}}} &= \frac{1}{R} + \frac{1}{i\omega L} + i\omega C \\ &= \frac{1}{R} + \frac{1}{i(\omega_r + \Delta\omega)L} + i(\omega_r + \Delta\omega)C \end{aligned} \tag{6.24}$$

$$
\begin{aligned}
&= \frac{1}{R} - i\frac{1}{\omega_r L}\left(1 - \frac{\Delta\omega}{\omega_r}\right) + i(\omega_r + \Delta\omega)C \quad \Delta\omega \ll \omega_r \\
&= \frac{1}{R} - i\left(\frac{1}{\omega_r L} - \frac{\Delta\omega}{\omega_r^2 L}\right) + i(\omega_r + \Delta\omega)C \\
&= \frac{1}{R} + 2i\Delta\omega C \\
&= \frac{1}{R}\left(1 + 2iQ\frac{\Delta\omega}{\omega_r}\right). \qquad (6.25)
\end{aligned}
$$

In the last line, we define $Q \equiv \omega_r RC = R/\omega_r L$. When the resonator is probed on resonance, i.e., $\Delta\omega = 0$, the imaginary component of the input impedance is 0, Z_{in} is maximized, and power $\propto |Z_{\text{in}}|^2$ is maximally delivered to the resonator. Equation 6.25 motivates another definition for the quality factor of a resonator that is useful for measurement: $Q_r = f_r/\text{FWHM}$, where FWHM is the full width at half maximum. If we plug in this definition for Q and set ω such that $\Delta\omega/\omega_r = \Delta f/f_r = \text{FWHM}/2f_r$, then we find that $1/Z_{\text{in}} = (1+i)/R$, which is a factor of $\sqrt{2}$ larger in magnitude than when $\Delta\omega = 0$. The power delivered to the resonator is then a factor of 2 smaller than when $\Delta\omega = 0$, thus motivating the "half maximum" part of the definition.

We write without derivation the transmission across a feedline that is coupled to the resonator:

$$
S_{21}(f) = 1 - \frac{Q_r}{Q_c}\frac{1}{1 + 2iQ_r\frac{f - f_r}{f_r}}. \qquad (6.26)
$$

It is fully derived in Gao (2008). Observations:

- The denominator $1 + 2iQ_r(f - f_r)/f_r$ matches the expression in the parentheses of Equation 6.25.
- When the input impedance of the resonator is minimized in magnitude by setting $f = f_r$, the transmission S_{21} is minimized. Maximal power is delivered to or reflected at the resonator.
- When the circuit is operated on resonance, the transmission is $1 - Q_r/Q_c$, giving the fraction of the RF power that is transmitted across the resonator. This expression can be rewritten as $1 - (1/Q_c)/(1/Q_r)$; the latter term has been recast in terms of energy loss mechanisms and is literally the ratio of energy lost through the feedline coupling to the total lost energy. There are three scenarios shown in Figure 6.6.
  1. $Q_i \gg Q_c$; the resonator is overcoupled to the feedline. When the coupling to the feedline is too strong, energy leaks out of the resonator too quickly before it can dissipate in the resonator. This situation corresponds to a large reflected wave and a smaller transmitted wave, the amplitudes of which can be extracted directly from Figure 6.6 Top Left for the scenario where $Q_i = 3Q_c$. The fraction of supplied feedline power (P_g in most of the KID literature) that is transmitted, reflected, and dissipated at the resonator is $(1/4)^2$, $(3/4)^2$, and $1 - (1/4)^2 - (3/4)^2 = 3/8$, respectively. As stated, the power dissipated at the resonator is less than maximal because most of the power is reflected.
  2. $Q_i = Q_c$; the resonator is optimally coupled to the feedline. At optimal coupling, the amplitudes of the transmitted and reflected waves are equal and opposite, and the powers of the transmitted and reflected waves are 1/4 of the supplied power, as shown in Figure 6.6 Middle. The dissipated power is thus half of the supplied feedline power, the maximal value it can reach.
  3. $Q_i \ll Q_c$; the resonator is undercoupled to the feedline. When the coupling to the feedline is too weak, energy is not efficiently delivered to the resonator, and the $|S_{21}(f)|^2$ curve becomes very shallow, as shown in Figure 6.6 Right for $Q_i = Q_c/3$. In this case, the fraction of P_g that is transmitted, reflected, and dissipated is $(3/4)^2$, $(1/4)^2$, and $1 - (3/4)^2 - (1/4)^2 = 3/8$, respectively, which is the converse of the overcoupled case.

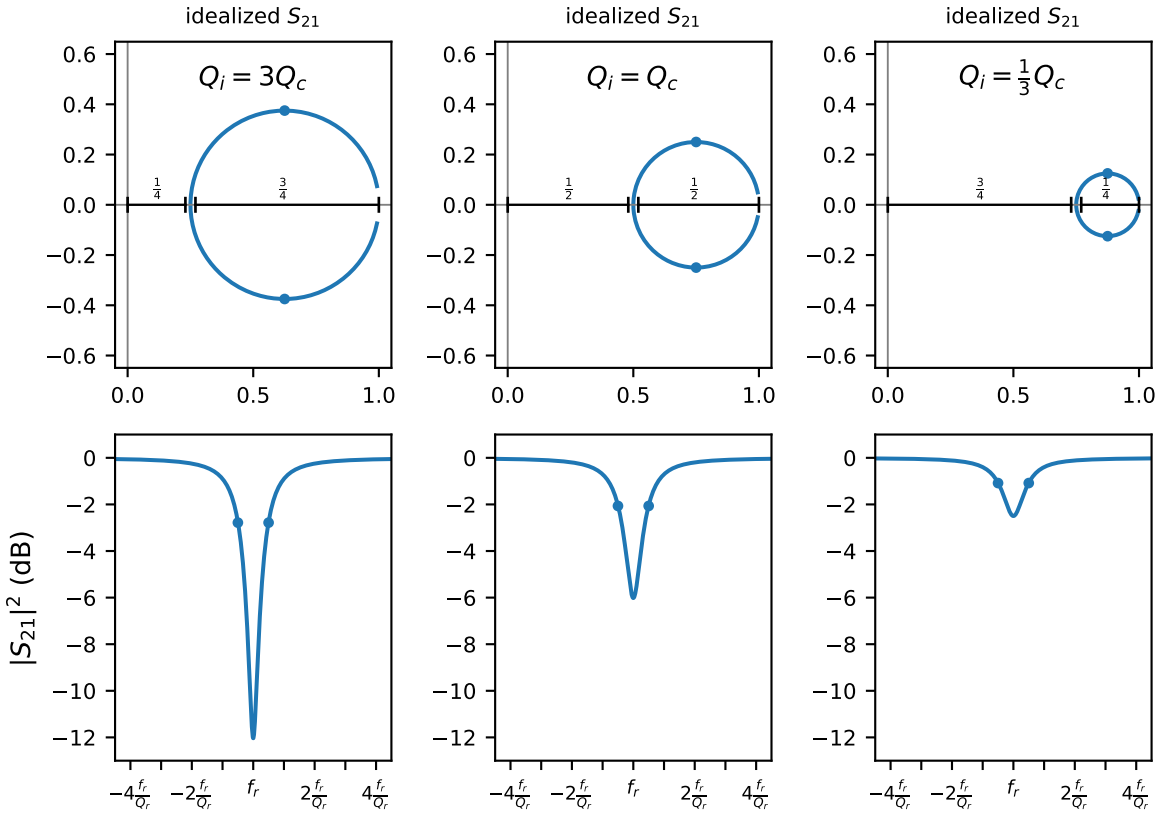

Figure 6.6: **Examples of $S_{21}(f)$ for resonators in the overcoupled, optimally coupled, and undercoupled regimes.** Top: complex $S_{21}(f)$ for three different ratios of Q_i to Q_c. The distance of the on-resonance S_{21} point from the origin and from the off-resonance transmission are indicated. Bottom: the power transmission across the resonator for the three different Q ratios.

- When $f - f_r = \pm\text{FWHM}/2 = \pm f_r/2Q_r$,

$$S_{21} = 1 - \frac{Q_r}{Q_c}\frac{1}{1 \pm i} = \left(1 - \frac{1}{2}\frac{Q_r}{Q_c} \mp i\frac{1}{2}\frac{Q_r}{Q_c}\right). \tag{6.27}$$

  So, $S_{21}(f_r \pm \text{FWHM}/2)$ are separated by Q_r/Q_c, as indicated by the blue dots in Figure 6.6 Top and Bottom.

We are interested in how S_{21} changes in response to changes in f_r and Q_i. We take the derivative of Equation 6.26 with respect to these two quantities and evaluate at the resonant frequency, where the response is maximized:

$$f_r \frac{dS_{21}}{df_r}\bigg|_{f=f_r} = -2i\frac{Q_r^2}{Q_c} \quad \text{and} \quad \frac{dS_{21}}{d(1/Q_i)}\bigg|_{f=f_r} = \frac{Q_r^2}{Q_c}. \tag{6.28}$$

For $\delta(1/Q_i)$, we use the fact that $d/d(1/Q_i) = d/d(1/Q_r)$ as a result of Equation 6.23. We also use $d/d(1/Q_r) = -Q_r^2 d/dQ_r$. These two equations can be succinctly expressed as

$$\boxed{\delta S_{21} = \frac{Q_r^2}{Q_c}\left(\delta\frac{1}{Q_i} - 2i\frac{\delta f_r}{f_r}\right).} \tag{6.29}$$

The regime of validity for Equation 6.29 can be roughly approximated as $\delta S_{21} \ll Q_r/Q_c$, i.e., the S_{21} deviation is much smaller than the diameter of the resonance circle.

There are two very important simplifications that have been made to Equation 6.26: (1) it assumes an off-resonance transmission of unity and (2) it assumes a real Q_c. In practice, the transmission of a resonance as measured with a vector network analyzer is better modeled with the following equation:

$$S_{21}(f) = ae^{2\pi i\tau(f-f_r)}\left(1 - \frac{Q_r}{\hat{Q}_c}\frac{1}{1+2iQ_r\frac{f-f_r}{f_r}}\right), \tag{6.30}$$

where a is the off-resonance transmission, τ is a measure of the cable delay, and $1/\hat{Q}_c = e^{i\phi}/|\hat{Q}_c|$ is a re-parametrization of the coupling quality factor when there are impedance mismatches along the feedline, which are commonly seen in actual data; see Khalil et al. (2012). When the derivatives are taken as before, we get

$$\delta S_{21} \propto |a|\frac{Q_r^2}{|\hat{Q}_c|}\left(\delta\frac{1}{Q_i} - 2i\frac{\delta f_r}{f_r}\right), \tag{6.31}$$

where the overall complex phase has been ignored. Note that the responsivity in terms of raw S_{21} depends on the magnitude of a and $\hat{Q}_c$. This expression is relevant in the next chapter about the noise of resonator readout.

6.6 Summary of equations and Mattis-Bardeen fits

The boxed equations from this section are collected here:

$$\begin{aligned}
\delta S_{21} &= \frac{Q_r^2}{Q_c}\left(\delta\frac{1}{Q_i} - 2i\frac{\delta f_r}{f_r}\right) && \text{(6.29 revisited)}\\
&= \frac{Q_r^2}{Q_c}\alpha\left(\frac{\delta R_s}{\omega L_s} + i\frac{\delta L_s}{L_s}\right) && (6.32)\\
&= \frac{Q_r^2}{Q_c}\alpha\gamma\left(\frac{\delta\sigma_1}{|\sigma_0|} + i\frac{\delta\sigma_2}{|\sigma_0|}\right) && (6.33)\\
&= \frac{Q_r^2}{Q_c}\alpha\gamma(\kappa_1 + i\kappa_2)\delta n_{\mathrm{qp}} && (6.34)\\
&= \frac{Q_r^2}{Q_c}\alpha\gamma(\kappa_1 + i\kappa_2)\frac{1}{V_{\mathrm{sc}}\Delta}E_{\mathrm{abs}}. && (6.35)
\end{aligned}$$

To actually use Equation 6.35, one needs to measure or compute all the relevant quantities that appear. The Q_r and Q_c are computed with $S_{21}(f)$ fits, which are typically performed directly with Equation 6.30. The combination $\alpha\gamma$ and the superconducting bandgap Δ are measured with what are known as Mattis-Bardeen fits. An approximate Mattis-Bardeen fit may take the following form, derived from Equation 6.29 and Equation 6.34 and reproduced from Aralis (2024):

$$\frac{d(1/Q_i)}{dn_{\mathrm{qp}}} = \alpha\gamma\kappa_1(T) \qquad \frac{df/f_r}{dn_{\mathrm{qp}}} = \alpha\gamma\kappa_2(T)$$

$$\frac{1}{Q_i(T)} - \frac{1}{Q_{i,0}} = \alpha\gamma\kappa_1(T)\left(n_{\mathrm{qp}}(T) - n_{\mathrm{qp},0}\right) \tag{6.36}$$

$$\frac{f_r(T) - f_{r,0}}{f_{r,0}} = \alpha\gamma\kappa_2(T)\left(n_{\mathrm{qp}}(T) - n_{\mathrm{qp},0}\right). \tag{6.37}$$

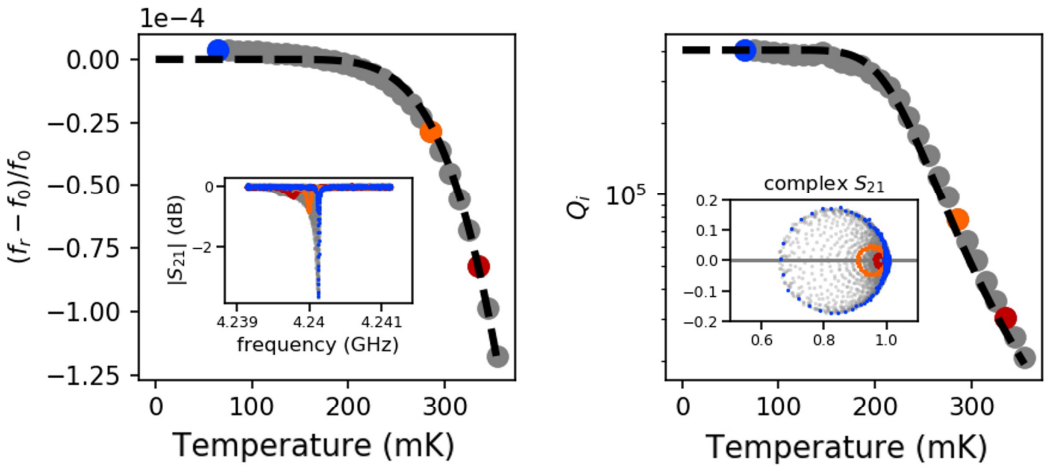

Figure 6.7: **Example Mattis-Bardeen fit.** Left: fractional frequency shift as a function of temperature with the fit to Equation 6.37 shown as the black dashed curve. Right: internal quality factor as a function of temperature with the fit to Equation 6.36 shown as the black dashed curve. Insets show complex S_{21} data from which the resonator parameters are extracted. For illustrate purposes, colors indicate specific temperatures at which S_{21} was measured. Fit values are $\Delta = 0.18\,\text{meV}$ and $\alpha = 3.8\%$. Baseline resonator parameters are $f_0 = 4.2401\,\text{GHz}$; and $Q_{i,0} = 410{,}000$. Figure reproduced from Wen et al. (2022).

Going from the first line to the second line in the above two equations, an integral over n_{qp} is performed from some quiescent quasiparticle population $n_{\text{qp},0}$ at some base temperature to a quasiparticle population at a temperature T. In this approximate Mattis-Bardeen fit, the temperature dependence of κ_1 and κ_2 are ignored in the integral since they are rather slow functions of temperature in comparison with $n_{\text{qp}}(T)$ from Equation 6.3:

$$n_{\text{qp}}(T) = 2N_0\sqrt{2\pi kT\Delta_0}e^{-\Delta_0/kT}. \qquad \text{(6.3 revisited)}$$

The general procedure of a Mattis-Bardeen fit is to measure the left-hand sides of Equation 6.36 and Equation 6.37 and then to fit for $\alpha\gamma$ and Δ_0 on the right-hand sides. The measurement consists of extracting the resonator parameters f_r and Q_r at elevated temperatures where the thermal population of quasiparticles is dominant, allowing us to ignore $n_{\text{qp},0}$. Technically, this procedure fits for the superconducting bandgap at zero temperature Δ_0, but Δ is what enters in Equation 6.35 and Equation 6.1 and is slightly smaller than Δ_0 for non-zero temperatures. It is assumed that $\Delta \approx \Delta_0$ for our purposes.

Once $\alpha\gamma$ and Δ_0 are determined from the above fitting procedure, a $\delta S_{21}(f_r; t)$ timestream can be converted to a $\delta n_{\text{qp}}(t)$ timestream using Equation 6.34. An example of a Mattis-Bardeen fit is shown in Figure 6.7.

6.7 Non-linear pulse shapes in KIDs

For small perturbations in δn_{qp}, the δS_{21} response is linear:

$$\delta S_{21} = \frac{Q_r^2}{Q_c}\alpha\gamma(\kappa_1 + i\kappa_2)\delta n_{\text{qp}} \qquad \text{(6.34 revisited)}$$

For large enough deviations $\delta f_r/f_r$ and $\delta(1/Q_r)$, the S_{21} response is no longer well approximated as linear. An effort to parameterize the non-linear response of KIDs is reported in Zobrist et al.

(2021). In this section, we show that the quantities $\delta f_r/f_r$ and $\delta(1/Q_r)$ are always extractable for any deviation δS_{21} measured by a tone held fixed at some original resonant frequency.

The complete S_{21} response can be expressed in terms of the initial resonant frequency $f_{r,0}$, the initial quality factor $Q_{r,0}$, and the fractional changes about these quantities: $\delta f_r/f_{r,0}$ and $\delta(1/Q_i)$. The relevant terms in the S_{21} formula Equation 6.26 transform in the following way:

$$\frac{1}{Q_r} = \frac{1}{Q_{r,0}} + \delta\frac{1}{Q_i} \tag{6.38}$$

$$f_r = f_{r,0} + \delta f_r \Rightarrow \frac{f - f_r}{f_r} = \frac{-\delta f_r}{f_{r,0} + \delta f_r}. \tag{6.39}$$

These transformations may then be plugged into the full expression for $\delta S_{21} = S_{21}(f_r) - S_{21}(f_{r,0})$, based on Equation 6.26:

$$\begin{aligned}
\delta S_{21} &= 1 - \frac{Q_r}{Q_c}\frac{1}{1 + 2iQ_r\frac{f-f_r}{f_r}} - \left(1 - \frac{Q_{r,0}}{Q_c}\right) \\
&= 1 - \frac{1}{Q_c}\frac{1}{\frac{1}{Q_r} + 2i\frac{f-f_r}{f_r}} - \left(1 - \frac{Q_{r,0}}{Q_c}\right) \\
&= 1 - \frac{1}{Q_c}\frac{1}{\frac{1}{Q_{r,0}} + \delta\frac{1}{Q_i} + 2i\frac{-\delta f_r}{f_{r,0}+\delta f_r}} - \left(1 - \frac{Q_{r,0}}{Q_c}\right) \\
&= 1 - \frac{Q_{r,0}}{Q_c}\frac{1}{1 + Q_{r,0}\delta\frac{1}{Q_i} - 2iQ_{r,0}\frac{\delta f_r}{f_{r,0}+\delta f_r}} - \left(1 - \frac{Q_{r,0}}{Q_c}\right) \\
&\approx \frac{Q_{r,0}}{Q_c}\left(1 - \frac{1}{1 + Q_{r,0}\delta\frac{1}{Q_i} - 2iQ_{r,0}\frac{\delta f_r}{f_{r,0}}}\right).
\end{aligned} \tag{6.40}$$

The final line assumes $\delta f_r \ll f_{r,0}$. The value of Equation 6.40 is that the pulse has been completely expressed in terms of the initial resonator frequency and quality factor, $f_{r,0}$ and $Q_{r,0}$, as well as the *total* change in fractional resonant frequency $\delta f_r/f_r$ and the *total* change in internal quality factor $\delta(1/Q_i)$.

Inverting the final equation, we get

$$\delta\frac{1}{Q_i} - 2i\frac{\delta f_r}{f_{r,0}} = \frac{1}{Q_{r,0}}\left(\frac{1}{1 - \frac{Q_c}{Q_{r,0}}\delta S_{21}} - 1\right). \tag{6.41}$$

The fact that we have an invertible equation means that conversion from δS_{21} to $\delta f_r/f_{r,0}$ and $\delta 1/Q_i$ is now completely tractable for any deviation in the idealized S_{21} plane with $\delta f_r \ll f_{r,0}$. The regime of validity for this expression is much greater than $|\delta S_{21}| \ll Q_r/Q_c$, the regime of validity for Equation 6.29.

It may useful to define $\beta_t \equiv (\delta f_r/f_{r,0})/\delta(1/Q_i)$ and recast Equation 6.40:

$$\delta S_{21} = \frac{Q_{r,0}}{Q_c}\left(1 - \frac{1}{1 + Q_{r,0}\frac{\delta f_r}{f_{r,0}}(\frac{1}{\beta_t} - 2i)}\right). \tag{6.42}$$

This substitution is important because as f_r shifts in the resonator, so does Q_i. We now take the derivative as we did in Equation 6.28:

$$f_{r,0}\frac{dS_{21}}{df_r} = \left(\frac{1}{\beta_t} - 2i\right)\frac{Q_{r,0}^2}{Q_c}\frac{1}{\left(1 + Q_{r,0}\frac{\delta f_r}{f_{r,0}}(\frac{1}{\beta_t} - 2i)\right)^2}. \tag{6.43}$$

For clarity, we reiterate that $\delta S_{21} = S_{21}(f_r) - S_{21}(f_{r,0})$ and that $\delta f_r = f_r - f_{r,0}$. We have taken the derivative of S_{21} with respect to f_r. Equation 6.43 implicitly assumes that β_t is constant with $\delta f_r/f_r$. Comparing Equation 6.43 with Equation 6.28,

$$f_r \frac{dS_{21}}{df_r}\bigg|_{f=f_r} = -2i\frac{Q_r^2}{Q_c}, \qquad \text{(6.28 revisited)}$$

we see that there are two main differences. First, the coefficient is modified from $-2i \to (1/\beta_t - 2i)$. The implication is that the derivative takes into account changes in both f_r and Q_i through the β_t parametrization. Second, the derivative is shown for any deviation $\delta f_r/f_{r,0}$ such that $\delta f_r \ll f_{r,0}$. We use Equation 6.43 to derive the following propagation of error formula:

$$\sigma_{\delta f_r/f_{r,0}} = \frac{Q_c}{Q_{r,0}^2}\frac{1}{\left|\frac{1}{\beta_t} - 2i\right|}\left|1 + Q_{r,0}\frac{\delta f_r}{f_{r,0}}\left(\frac{1}{\beta_t} - 2i\right)\right|^2 \sigma_{S_{21}}. \tag{6.44}$$

We highlight the term in $|\cdot|^2$ as the "energy-dependent" term:

$$\sigma_{\delta f_r/f_{r,0}} = \left|1 + Q_{r,0}\frac{\delta f_r}{f_{r,0}}\left(\frac{1}{\beta_t} - 2i\right)\right|^2 \sigma_{\delta f_r/f_{r,0}}\Big|_{f_r=f_{r,0}}. \tag{6.45}$$

This relationship is particularly important for converting amplifier noise, which is fixed in S_{21}, to noise in $\delta f_r/f_{r,0}$ or quasiparticle density for any δS_{21} such that $\delta f_r \ll f_{r,0}$. See Section 8.4 and Figure 8.9 Left in particular for an application of the above nonlinear formulae.

An important caveat to the previous paragraph is the assumption that β_t is constant. β_t, which is the ratio of the *total* changes in $\delta f_r/f_{r,0}$ and $\delta(1/Q_i)$, should depend on the integral of κ_2/κ_1, which is the ratio of the *instantaneous* $\delta f_r/f_r$ and $\delta(1/Q_i)$ at a particular temperature. κ_1 and κ_2 are plotted in Figure 6.4 Top Left. Variations in κ_1 and κ_2 can theoretically be incorporated into an expression for β_t through integrals of κ_1 and κ_2.

Chapter 7

Noise sources in superconducting microwave resonators

There are many sources of noise in superconducting microwave resonators. In the first half of this chapter, we describe and detail the following:

- generation-recombination noise (not yet observed in KIPM detectors);
- two-level system noise;
- Johnson, amplifier, and readout chain noise;
- and electronics $1/f$ noise.

In the second half of the chapter starting with Section 7.5, we present measurements and characterizations of the latter three as have been observed in KIPM detectors.

7.1 Noise on resonator readout

We start with a basic description of how the noise on S_{21} readout manifests during operation of a kinetic inductance detector and how that noise is converted into physical units using equations from Chapter 6. $S_{21}(t;\, f_r)$ can be measured from the input to the output of some cryostat operating the resonator and is proportional to the voltage measured at the output. In other words, $S_{21} = V_2/V_1$, where V_2 is the complex voltage at the output and V_1 is the complex voltage at the input. For this section, we refer to the raw data in terms $V(f)$ and $V(t;\, f_r)$, the voltage as measured by some instrument at the output of the cryostat. A may be divided by gains and/or attenuations that appear on the readout chain so that it is instead measured at some other location along the readout chain, but doing so does not affect the following results. A voltage may even be divided by the voltage of the input tone to the cryostat so as to produce a raw S_{21}, which is usually directly measured by a vector network analyzer, abbreviated VNA.

Figure 7.1 Left shows how the raw voltage data may appear during data-taking. $V(f)$ is shown in gray and is schematically illustrated as a circle in the complex V space. $V(t;\, f_r)$ is illustrated as the blue noise blob with size σ_V. a may be thought of as the off-resonance voltage, or the voltage that would be measured at this frequency if there were no resonator present. Notice that the axis of the circle that includes $V(t;\, f_r)$ is rotated away from a, parametrized by an angle ϕ. This parametrization is described in Khalil et al. (2012) and arises from impedance mismatches of the transmission line near the resonator. In the parametrization by Khalil et al. (2012), the

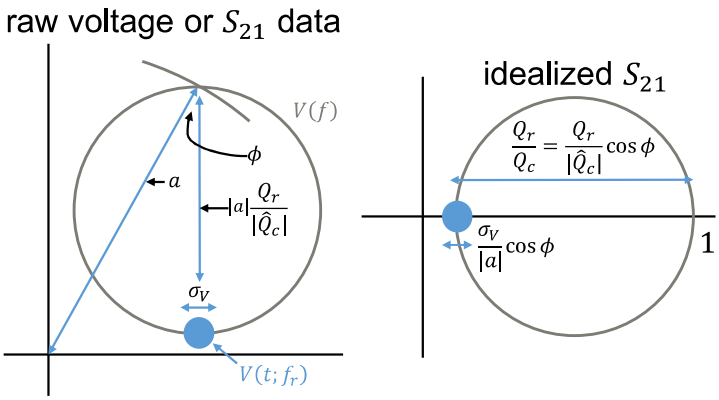

Figure 7.1: **The transformation from raw voltage data to idealized S_{21}.** Left: an example resonator's $V(f)$ scan in gray and the on-resonance readout noise in blue. Right: "idealized" version of the resonator, where $V(f)$ have been scaled such that the off-resonance transmission is equal to unity and where the resonance circle is projected along the real axis in a manner consistent with Khalil et al. (2012).

diameter of the circle is $|a|Q_r/|\hat{Q}_c|$ as can be inferred from Equation 6.30, reproduced below:

$$S_{21}(f) = ae^{2\pi i\tau(f-f_r)}\left(1 - \frac{Q_r}{\hat{Q}_c}\frac{1}{1+2iQ_r\frac{f-f_r}{f_r}}\right). \qquad \text{(6.30 revisted)}$$

$1/\hat{Q}_c$ may be parametrized as $e^{i\phi}/|\hat{Q}_c|$; note that $1/|\hat{Q}_c|$ is the magnitude of the parametrized complex quantity, not $|\hat{Q}_c|$ as one might naively assume. Furthermore recall that τ is the cable delay and is the reason for the arc shape at the off-resonance transmission in Figure 7.1 Left.

We desire to transform the raw data into what is known as "idealized S_{21}," which corresponds to the following resonator equation from the previous chapter:

$$S_{21}(f) = 1 - \frac{Q_r}{Q_c}\frac{1}{1+2iQ_r\frac{f-f_r}{f_r}}. \qquad \text{(6.26 revisited)}$$

We desire to operate in the idealized S_{21} space where the equations from Chapter 6 hold. In this space, the off-resonance transmission a and resonance-circle rotation ϕ due to impedance mismatches are scaled out. The transmission can now be studied purely as it relates to the resonance parameters f_r, Q_r, and Q_c.

Figure 7.1 Right also shows how the transformation to the idealized S_{21} plane is done. There are two steps: (1) division of all points by a so that the the off-resonance transmission is equal to 1 and (2) projection along the real axis as is described in detail by Khalil et al. (2012), resulting in a natural definition for $1/Q_c \equiv \mathrm{Re}[1/\hat{Q}_c] = (1/|\hat{Q}_c|)\cos\phi$.[1] The on-resonance idealized transmission is then equal to $1 - Q_r/Q_c$, i.e., the circle is centered on the real axis.

Crucially, both $V(f)$ and σ_V are scaled by $\cos\phi$ during this transformation into the idealized space. For $V(f)$, the idealized resonance circle has diameter Q_r/Q_c, which is equal to $(Q_r/|\hat{Q}_c|)\cos\phi$; the transformation is $|a|Q_r/|\hat{Q}_c| \rightarrow (Q_r/|\hat{Q}_c|)\cos\phi$. Likewise, for σ_V, transformation into the

[1] Note that the inverse quantities $1/Q_c$ and $1/|\hat{Q}_c|$ are what provide a measure for $S_{21}(f)$ and $V(f)$ circle size in the idealized S_{21} and raw voltage planes; $\cos\phi$ projects $1/|\hat{Q}_c|$ into $1/Q_c$. The inverse nature of the $\cos\phi$ projection is very important for its ensuing cancellation in Equation 7.2!

idealized space follows $\sigma_V \to (\sigma_V/|a|)\cos\phi$. This transformed noise quantity is referred to as $\sigma_{S_{21}}$. Since both signal and noise are multiplied by $\cos\phi$, it has no effect on the final SNR and the resultant resolution on $\delta f_r/f_r$ and absorbed energy. We refer to equations from the previous chapter for the following calculation:

$$\sigma_{S_{21}} = \frac{\sigma_V}{|a|}\cos\phi \tag{7.1}$$

$$\sigma_{\delta f_r/f_r} = \sigma_{S_{21}}\frac{Q_c}{2Q_r^2} = \frac{\sigma_V}{|a|}\cos\phi\frac{|\hat{Q}_c|}{2Q_r^2}\frac{1}{\cos\phi} = \frac{\sigma_V}{|a|}\frac{|\hat{Q}_c|}{2Q_r^2} \tag{7.2}$$

$$\sigma_{n_{\rm qp}} = \sigma_{\delta f_r/f_r}\frac{1}{\alpha\gamma\kappa_2} = \sigma_V\frac{|\hat{Q}_c|}{2|a|Q_r^2}\frac{1}{\alpha\gamma\kappa_2} \tag{7.3}$$

$$\sigma_{E_{\rm abs}} = \sigma_{n_{\rm qp}}V_{\rm ind}\Delta = \sigma_V\frac{|\hat{Q}_c|}{2|a|Q_r^2}\frac{1}{\alpha\gamma\kappa_2}V_{\rm ind}\Delta. \tag{7.4}$$

A key result from the above calculation is the cancellation of $\cos\phi$ that occurs in Equation 7.2. The resolution on physical quantities, such as $\delta f_r/f_r$, $n_{\rm qp}$, and $E_{\rm abs}$, only depends on the size of the noise blob divided by diameter of the resonator circle in the raw data plane, with an additional factor of Q_r for the frequency spacing around the resonant frequency: $(\sigma_V/|a|)/(Q_r/|\hat{Q}_c|)/Q_r$. The resolutions on these physical quantities do not depend on the resonator rotation directly, but it is possible that larger ϕ rotations are correlated with worse transmission $|a|$.

Noise on S_{21} readout may be converted into any of these units, but certain noise sources are naturally expressed in certain units. For example, generation-recombination noise arises from stochastic fluctuations in quasiparticle number, so it is usually measured in units of quasiparticle density $\sigma_{n_{\rm qp}}$ then converted into $E_{\rm abs}$. Amplifier noise due to cryogenic readout produces a voltage noise at the input of the amplifier and so can be measured in terms of σ_V and then converted into $\sigma_{E_{\rm abs}}$ using the above calculation.

Throughout this chapter, we refer to the noise levels and spectral shapes from various noise sources in terms of their power spectral density $J(f)$. For some real-valued continuous timestream $x(t)$, its power spectral density can be computed from the Fourier transform $\tilde{x}(f)$:

$$J(f) = \lim_{T\to\infty}\frac{1}{T}|\tilde{x}(f)|^2. \tag{7.5}$$

T is the duration of the timestream. PSDs can also be computed from the Fourier transform of the autocorrelation $R(\tau)$:

$$\begin{aligned} J(f) &= \int_{-\infty}^{\infty} R(\tau)e^{-i2\pi f\tau}d\tau \\ &= \int_{-\infty}^{\infty}\left[\lim_{T\to\infty}\int_0^T x(t-\tau)x(t)dt\right]e^{-i2\pi f\tau}d\tau \\ &= \int_{-\infty}^{\infty}\langle x(t-\tau)x(t)\rangle e^{-i2\pi f\tau}d\tau. \end{aligned} \tag{7.6}$$

Autocorrelation is the correlation of a signal with itself that is delayed by some time τ. Note that the average in the last line is performed over the time variable t as concretely expressed in the second line.

By Parseval's theorem, the total variance of the time stream is equal to the integral of power spectral density:

$$\sigma_x^2 = \langle x(t)^2\rangle = \int_{-\infty}^{\infty} J(f)df. \tag{7.7}$$

In other words, PSDs written as $J(f)$ decompose the size of the noise blobs into the component frequency contributions. For actual detector operation, what really matters is the optimally filtered variance or resolution. Optimal noise filters generally take the shape of low-pass filters, cutting away the contribution of noise from higher frequencies that do not contribute to the shape of the signal. If $J(f)$ rises at low frequencies, like with two-level system noise or electronics $1/f$ noise, the optimal filter down-weights the contribution from those frequencies. The optimal filter $\phi(f)$ for a particular signal is precisely the Fourier coefficients of that signal in Fourier space, $\tilde{s}(f)$, divided by $J(f)$:

$$\phi(f) = \frac{\tilde{s}^*(f)}{J(f)}. \tag{7.8}$$

The optimally filtered variance is then

$$\sigma_x^2 = \left[\int_{-\infty}^{\infty} df \frac{|\tilde{s}(f)|^2}{J(f)}\right]^{-1}. \tag{7.9}$$

Throughout this chapter and for simplicity, we apply the optimal filter corresponding to a single-sided single-pole exponential with fall time $\tau_{\rm qp}$ in order to compute expected energy resolutions:

$$\tilde{s}(f) = \frac{\tau_{\rm qp}}{1 + i2\pi f \tau_{\rm qp}}. \tag{7.10}$$

To calculate the contribution to $\sigma_{E_{\rm abs}}$ of each type of noise, we assume the following values for the various relevant parameters: $V_{\rm ind} = 30{,}000\,\mu{\rm m}^3$, $\Delta = 180\,\mu{\rm eV}$, $\alpha = 0.038$, $\gamma = 1$, $\kappa_2 = 4.3 \times 10^7 \mu{\rm m}^3$, and $Q_r \approx Q_c = 10^5$. This final assumption is a design choice to overcouple the resonator to the feedline so that the resonator rise time $Q_r/\pi f_r$ is short enough to see the phonon rise time information. The values for Δ and α come from Figure 6.7.

7.2 Generation-recombination noise

Generation-recombination (GR) noise is the fundamental limit of noise in a kinetic inductance detector. GR noise results from fluctuations in the number of quasiparticles present at any one time due to the stochastic process of Cooper pairs breaking to yield quasiparticles and quasiparticles recombining to form Cooper pairs. GR noise is analogous to thermal fluctuation noise in TESs. The key distinction is that, in principle, GR noise in superconductors should be exponentially suppressed (with $1/T$) while thermal fluctuation noise in TESs scales as a power law in T_c (precisely what power law depends on some design choices). In practice, anything that creates a quiescent quasiparticle population causes GR noise to saturate at the level determined by that population density.

GR noise formalism

To derive an expression for the GR noise PSD, we start with the following differential equation:

$$\begin{aligned} \frac{dn_{\rm qp}}{dt} &= -2\Gamma_R \\ &= -Rn_{\rm qp}^2. \end{aligned} \tag{7.11}$$

Γ_R is the rate of recombination events per unit volume. For each recombination event, two quasiparticles become Cooper pairs, hence the factor of 2. R is known as the recombination constant and characterizes the strength of the recombination process. Heuristically, two quasiparticles must find each other in order to recombine and decrease the total population. The number of pairs of

quasiparticles per unit volume scales with $n_{\rm qp}^2/2$. R is the proportionality constant between the rate of recombination events and the number of pairs of quasiparticles: $\Gamma_R = Rn_{\rm qp}^2/2$. See Wilson et al. (2004) for further context.

The steady-state solution to the above differential equation is $n_{\rm qp} = 0$, but in practice, a generation rate per unit volume $\Gamma_{\rm G}$ should be introduced to the right-hand side to account for the generation of quasiparticles from their thermal distribution (Equation 6.3) or external loading. The latter is usually dominant for aluminum resonators below about 200 mK. If we set $2\Gamma_{\rm G} = Rn_{\rm qp,0}^2$ so that $n_{\rm qp} = n_{\rm qp,0}$ is a steady-state solution, then the differential equation becomes

$$\begin{aligned}\frac{dn_{\rm qp}}{dt} &= 2\Gamma_G - 2\Gamma_R \\ &= Rn_{\rm qp,0}^2 - Rn_{\rm qp}^2.\end{aligned} \tag{7.12}$$

It is useful to insert $n_{\rm qp} = n_{\rm qp,0} + \delta n_{\rm qp}$ into Equation 7.12:

$$\frac{d\delta n_{\rm qp}}{dt} = -2Rn_{\rm qp,0}\delta n_{\rm qp} - R\delta n_{\rm qp}^2. \tag{7.13}$$

There are two regimes to point out: $\delta n_{\rm qp} \ll n_{\rm qp,0}$ and $\delta n_{\rm qp} \gg n_{\rm qp,0}$. In the first regime, the linear term on the RHS of Equation 7.13 dominates; solutions take the form $\exp(-2Rn_{\rm qp,0}t)$; and the expression $1/2Rn_{\rm qp,0}$ can be seen as a decay time or a quasiparticle recombination time $\tau_{\rm qp}$. Henceforth, we equate

$$\frac{1}{2R} = n_{\rm qp,0}\tau_{\rm qp}.^{2} \tag{7.14}$$

On the other hand, if $\delta n_{\rm qp} \gg n_{\rm qp,0}$, the quadratic term on the RHS dominates and the differential equation is non-linear. Solutions take the form $1/t$ instead of $e^{-t/\tau_{\rm qp}}$. The complete solution to Equation 7.13 is derived in Appendix A.3 of Chang (2023) and compared to data in Appendix D of Aralis (2024).

To incorporate the stochastic effects of shot noise, we replace $2\Gamma_R$ and $2\Gamma_G$ with $I_R(t)$ and $I_G(t)$, which in turn are expressed as the sum of delta functions:

$$\begin{aligned}\frac{dn_{\rm qp}}{dt} &= I_G(t) + I_R(t) \\ &= \sum_{k_G} \frac{2}{V_{\rm ind}}\delta(t - t_{k_G}) - \sum_{k_R} \frac{2}{V_{\rm ind}}\delta(t - t_{k_R}) \\ &\approx 2(\Gamma_G + \delta\Gamma_G(t)) - 2(\Gamma_R + \delta\Gamma_R(t)).\end{aligned} \tag{7.15}$$

There are a few comments to make.

- The variable $I(t)$ is chosen to represent the "currents" of recombination and generation. There is a direct correspondence between GR noise and electrical current shot noise, which was first introduced in Schottky (1918).
- The delta functions are located at time points t_k. Integrating both sides over a small window around a particular t_k shows that an individual delta function serves to increase or decrease the quasiparticle density by $2/V_{\rm ind}$, which is the "shot size" of each event. The integral of a delta function is the step function.

[2] $\tau_{\rm qp}$ as measured from a pulse fall time may saturate to some maximum value $\tau_{\rm max}$ less than the $\tau_{\rm qp}$ as inferred from a quasiparticle density measurement. For this calculation, we assume that $n_{\rm qp}$ is what sets $\tau_{\rm qp}$ as given by Equation 7.14, using R as a constant derived from material parameters (see below). One way to take into account $\tau_{\rm max}$ is to set $\tau_{\rm qp}^{-1} = 2Rn_{\rm qp,0} + \tau_{\rm max}^{-1}$.

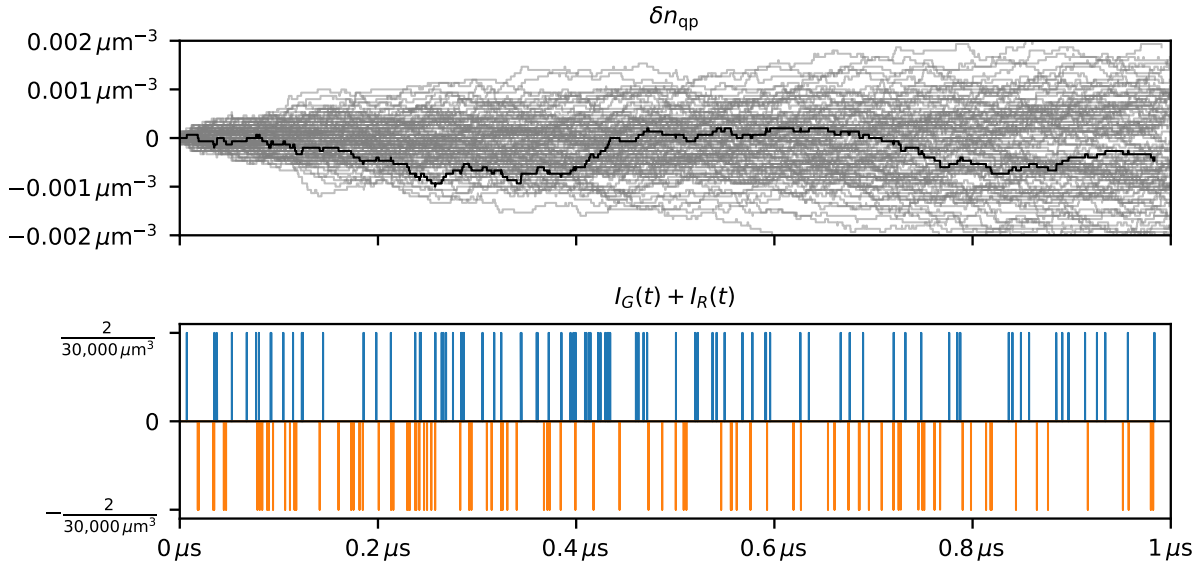

Figure 7.2: **Generation-recombination noise as a stochastic Poisson point process.** Top: 100 realizations of $\delta n_{\text{qp}}(t)$. Each step corresponds to a delta function in $I(t)$, and the size of each step is $2/V_{\text{ind}}$, where we choose $V_{\text{ind}} = 30{,}000\mu\text{m}^3$. The black realization corresponds to the $I(t)$ curves shown in the bottom panel. Bottom: $I_G(t)$ in blue and $I_R(t)$ in orange. The average rate for both processes is set to $V_{\text{ind}}\Gamma = V_{\text{ind}}Rn_{\text{qp},0}^2/2$, where we set $R = 10\,\mu\text{m}^3/\text{s}$ and $n_{\text{qp},0} = 25\,\mu\text{m}^{-3}$. Note that the example shown in this figure is a fully Poissonian point process, i.e., there are no correlations on τ_{qp} timescales as there is for true GR noise.

- $\{t_{k_G}\}$ and $\{t_{k_R}\}$ describe separate and independent Poisson point processes for the recombination and generation of quasiparticles. The Poisson point processes are constructed to have average occurrence rates of $V_{\text{ind}}\Gamma_G$ and $V_{\text{ind}}\Gamma_R$. Thus, taking the average over t of both sides in the second line of Equation 7.15 produces $\Delta n_{\text{qp}}/\Delta t = 2\Gamma_G - 2\Gamma_R$. Figure 7.2 provides an example of such a process.
- Fluctuations about the average rates are parametrized in the final line of Equation 7.15 with $\delta\Gamma_G(t)$ and $\delta\Gamma_R(t)$. Given the Poissonian construction of $I(t)$, it can be shown that the autocorrelation and power spectral density are equal to

$$\langle I(t)I(t-\tau)\rangle = 4\Gamma^2 + \frac{4\Gamma}{V_{\text{ind}}}\delta(\tau) \quad \text{and} \quad J_I(f) = 4\Gamma^2\delta(f) + \frac{4\Gamma}{V_{\text{ind}}}. \tag{7.16}$$

  See Section 2.3.2.1.2 of Siegel (2016) for a derivation. Recall that the power spectral density and autocorrelation form a Fourier transform pair. $4\Gamma^2$ is the steady-state value of $\langle I(t)\rangle^2$ and is equal to (shot size)$^2\times$(rate of events)$^2 = (2/V_{\text{ind}})^2(V_{\text{ind}}\Gamma)^2$. $4\Gamma/V_{\text{ind}}$ corresponds to the fluctuations in the rate as a result of the irregular occurrence of generation and recombination events—in other words, shot noise. It is equal to (shot size)$^2\times$(rate of events) $= (2/V_{\text{ind}})^2 V_{\text{ind}}\Gamma$. As one would expect for a Poissonian process, the variance of the rate scales with the rate itself, and the noise is completely uncorrelated—it has a white PSD. Given the factors of 2 in Equation 7.15, we identify Γ/V_{ind} as the autocorrelation and power spectral density of $\delta\Gamma(t)$:

$$\langle \delta\Gamma(t)\delta\Gamma(t-\tau)\rangle = \frac{\Gamma}{V_{\text{ind}}}\delta(\tau) \quad \text{and} \quad J_{\delta\Gamma}(f) = \frac{\Gamma}{V_{\text{ind}}}. \tag{7.17}$$

The GR noise power spectral density for $\delta n_{\rm qp}$ fluctuations can be computed from Equation 7.15 with the following derivation. We allow Γ_R to fluctuate by setting $2\Gamma_G - 2\Gamma_R = Rn_{\rm qp,0}^2 - Rn_{\rm qp}^2 \approx -2Rn_{\rm qp,0}\delta n_{\rm qp} = -\delta n_{\rm qp}/\tau_{\rm qp}$. As a reminder, this approximation uses $n_{\rm qp} = n_{\rm qp,0} + \delta n_{\rm qp}$ and sets $\delta n_{\rm qp} \ll n_{\rm qp,0}$.

$$
\begin{aligned}
\frac{d\delta n_{\rm qp}}{dt} &= 2(\Gamma_G + \delta\Gamma_G(t)) - 2(\Gamma_R + \delta\Gamma_R(t)) && \text{Equation 7.15}\\
\frac{d\delta n_{\rm qp}}{dt} &= -\frac{1}{\tau_{\rm qp}}\delta n_{\rm qp} + 2\delta\Gamma_G(t) - 2\delta\Gamma_R(t) && 2\Gamma_G - 2\Gamma_R \approx -\delta n_{\rm qp}/\tau_{\rm qp}\\
i2\pi f\delta\tilde{n}_{\rm qp} &= -\frac{1}{\tau_{\rm qp}}\delta\tilde{n}_{\rm qp} + 2\delta\tilde{\Gamma}_G - 2\delta\tilde{\Gamma}_R && \text{Fourier transform}\\
\delta\tilde{n}_{\rm qp} &= \frac{\tau_{\rm qp}}{1+2\pi i f\tau_{\rm qp}} 2\left(\delta\tilde{\Gamma}_G - \delta\tilde{\Gamma}_R\right) && \text{re-arrange}\\
|\delta\tilde{n}_{\rm qp}|^2 &= \frac{\tau_{\rm qp}^2}{1+(2\pi f\tau_{\rm qp})^2} 4\left(|\delta\tilde{\Gamma}_R|^2 + |\delta\tilde{\Gamma}_G|^2\right) && \text{multiply by conjugate}\\
J_{\rm GR}(f) &= \frac{\tau_{\rm qp}^2}{1+(2\pi f\tau_{\rm qp})^2} 4\frac{\Gamma_G + \Gamma_R}{V_{\rm ind}} && \text{shot noise result}\\
&= \frac{4Rn_{\rm qp,0}^2\tau_{\rm qp}^2/V_{\rm ind}}{1+(2\pi f\tau_{\rm qp})^2} && \Gamma_G + \Gamma_R = Rn_{\rm qp,0}^2\\
\boxed{J_{\rm GR}(f) = \frac{2n_{\rm qp,0}\tau_{\rm qp}/V_{\rm ind}}{1+(2\pi f\tau_{\rm qp})^2} = \frac{1/RV_{\rm ind}}{1+(2\pi f\tau_{\rm qp})^2}.} && R = \frac{1}{2n_{\rm qp,0}\tau_{\rm qp}} && (7.18)
\end{aligned}
$$

$J_{\rm GR}(f)$ is the power spectral density of GR noise in units of (quasiparticle density)^2/Hz. This expression is for a two-sided power spectral density. For comparison, the numerator of the second to last line matches the sum of Equations 6.7 and 6.10 from Gao (2008). Note that Siegel (2016) and de Visser et al. (2011)[3] report one-sided PSDs, so their expressions are larger by a factor of 2.

Characteristics of GR noise

Here are a few important characteristics of the power spectral density for GR noise.

- The spectral shape of GR noise is flat up to $1/\tau_{\rm qp}$. Note in Equation 7.18 that the roll-off at $1/\tau_{\rm qp}$ arises from the $-\delta n_{\rm qp}/\tau_{\rm qp}$ term, which in turn arises from letting the $2\Gamma_R$ term from Equation 7.15 fluctuate away from its equilibrium value, separate from the $\delta\Gamma_R(t)$ shot-noise fluctuation. GR noise is different from pure shot noise with a fixed average rate (such as electrons entering a diode) because shot noise fluctuations can cause the underlying rate of GR events to also fluctuate; more or fewer quasiparticles at a given point in time directly impacts the instantaneous rate of GR events. The $-\delta n_{\rm qp}/\tau_{\rm qp}$ term brings the quasiparticle population back to equilibrium on timescales around $\tau_{\rm qp}$, meaning that $\delta n_{\rm qp}$ is correlated with itself on timescales around $\tau_{\rm qp}$. For true Poissonian shot noise as is shown in Figure 7.2, there is no roll-off at high frequencies and no correlation between events on any timescale.
- For a thermal population of quasiparticles, the flat level $n_{\rm qp,0}\tau_{\rm qp}$ does not change with temperature. $\tau_{\rm qp}$ can be derived from an integral similar to Equation 6.3, which is restated here:

$$
n_{\rm qp,0} = 4\int_{\Delta}^{\infty} D(E)f(E)dE \quad \overset{kT\ll\Delta}{=} \quad 2N_0\sqrt{2\pi kT\Delta_0}e^{-\Delta_0/kT}, \qquad \text{(6.3 revisited)}
$$

[3] See de Visser (2014) for the derivation and comment on the single-sidedness of the PSD in de Visser et al. (2011).

where N_0 is the single spin density of states at the Fermi level. Note that the integrand is $\propto D(E)f(E)$ (see Section 6.2). When two quasiparticles recombine, they emit a phonon. So, as stated in the second line of Equation 8 from Kaplan et al. (1976), $\tau_{\rm qp}(E)$ can be computed from a similar integral over emitted phonon energies Ω, with integrand proportional to $D(\Omega - E)f(\Omega - E)$. For quasiparticles at the gap edge ($E = \Delta$), Kaplan et al. (1976) computes the following expression for $\tau_{\rm qp}$:

$$\tau_{\rm qp}^{-1} = \tau_0^{-1}\sqrt{\pi}\left(\frac{2\Delta_0}{kT_c}\right)^{5/2}\left(\frac{T}{T_c}\right)^{1/2} e^{-\Delta_0/kT}, \tag{7.19}$$

where τ_0 is a characteristic electron-phonon interaction time. The key point is that, as a result of the similar integrands for computing the thermal population $n_{\rm qp,0}$ and the recombination time $\tau_{\rm qp}$, both $n_{\rm qp,0}$ and $1/\tau_{\rm qp}$ are proportional to $\sqrt{T}e^{-\Delta_0/kT}$. Thus, the product $n_{\rm qp,0}\tau_{\rm qp}$ is a constant given by material parameters:

$$n_{\rm qp,0}\tau_{\rm qp} = \tau_0\frac{2N_0(kT_c)^3}{(2\Delta_0)^2} = \frac{1}{2R}. \tag{7.20}$$

For aluminum, $R \approx 10\,\mu{\rm m}^3/{\rm s}$.

- The bandwidth of the noise increases with temperature because $\tau_{\rm qp}$ decreases with temperature.

Expected energy resolution

GR-noise-dominated $\sigma_{n_{\rm qp}}$ can be computed from Equation 7.9:

$$\begin{aligned}
\sigma^2_{n_{\rm qp}} &= \left[\int_{-\infty}^{\infty}\frac{|\tilde{s}(f)^2|}{J_{\rm GR}(f)}df\right]^{-1} \qquad J_{\rm GR}(f) = \frac{2n_{\rm qp,0}\tau_{\rm qp}/V_{\rm ind}}{1+(2\pi f\tau_{\rm qp})^2} \quad \tilde{s}(f) = \frac{\tau_{\rm qp}}{1+i2\pi f\tau_{\rm qp}} \\
&= \left[\int_{-\infty}^{\infty}\frac{\tau_{\rm qp}^2}{1+(2\pi f\tau_{\rm qp})^2}\frac{1+(2\pi f\tau_{\rm qp})^2}{2n_{\rm qp,0}\tau_{\rm qp}/V_{\rm ind}}df\right]^{-1} \\
&= \left[\int_{-\infty}^{\infty}\frac{V_{\rm ind}\tau_{\rm qp}}{2n_{\rm qp,0}}df\right]^{-1} \\
&= \frac{2n_{\rm qp,0}}{V_{\rm ind}\tau_{\rm qp}}\left[\int_{-\infty}^{\infty}df\right]^{-1} \\
\longrightarrow \sigma^2_{n_{\rm qp}} &= \frac{2n_{\rm qp,0}}{V_{\rm ind}\tau_{\rm qp}}\frac{1}{2\Delta f} = \frac{2Rn^2_{\rm qp,0}}{V_{\rm ind}}\frac{1}{\Delta f}.
\end{aligned} \tag{7.21}$$

In the final line, the integral over f has been replaced with some effective bandwidth Δf. The key concept with GR noise is that the signal-derived optimal filter rolls off in the same way as GR noise, enabling integration of signal-to-noise up to arbitrary frequencies. In practice, there comes a frequency at which the GR noise becomes subdominant to some other source of noise. For concreteness, we evaluate Equation 7.21 at $\Delta f = 1/2\pi\tau_{\rm qp} = Rn_{\rm qp}/\pi$, which corresponds to the 3 dB cutoff frequency of the signal-derived low-pass filter.

We use $n_{\rm qp,0} = 25\,\mu{\rm m}^{-3}$, the lowest observed quasiparticle density for aluminum, from de Visser et al. (2011). As mentioned in a footnote in Section 6.2, Temples et al. (2024) infers a quasiparticle density of $18.5\,\mu{\rm m}^{-3}$, but the pulse lifetime measurement in that article is not as robust as the combined pulse lifetime and noise PSD measurement from de Visser et al. (2011).

Putting in the expression for Δf and the value for $n_{\rm qp}$,

$$\sigma^2_{n_{\rm qp}} = \frac{2\pi n_{\rm qp,0}}{V_{\rm ind}} \tag{7.22}$$

$$\longrightarrow \sigma_{E_{\text{abs}}} = \sigma_{n_{\text{qp}}} V_{\text{ind}} \Delta = \sqrt{2\pi n_{\text{qp},0} V_{\text{ind}}}\,\Delta = 390\,\text{meV}. \tag{7.23}$$

Equation 7.22 can be computed as the variance in the quasiparticle number N_{qp} by multiplication of V_{ind}^2. The result is $\sigma_{N_{\text{qp}}}^2 = 2\pi N_{\text{qp}}$. The proportionality of variance to N_{qp} makes sense, given that GR noise arises from a Poissonian process.

From Equation 7.21, one might argue that we can integrate up to an arbitrary frequency and so can do better than 390 meV. Doing so relies on other sources of noise being subdominant to GR noise at higher frequencies than $1/2\pi\tau_{\text{qp}}$ (faster timescales), where the GR noise PSD rolls off as $1/f^2$. The interpretation here is that sudden generation events of quasiparticles provide an infinite signal-to-noise ratio over the τ_{qp}-correlated GR noise. From a Poissonian counting perspective, $\sigma_{N_{\text{qp}}}^2 = 2N_{\text{qp}}$ provides a rough scale beyond which further resolution improvement may be difficult. The 2 is a result of the fact that the fundamental underlying stochastic process is the emission and absorption of 2Δ phonons during the generation-recombination process; quasiparticles are created and destroyed in pairs. So, we argue that 390 meV may be improved by a factor of $\sqrt{\pi} \approx 1.77$ so that $\sigma_{E_{\text{abs}}} = 220\,\text{meV}$, assuming that other sources of noise can be rendered subdominant up to $1/2\tau_{\text{qp}}$ (instead of up to $1/2\pi\tau_{\text{qp}}$). If the other sources of noise can be brought to even lower levels beyond $1/2\tau_{\text{qp}}$, further improvements are theoretically possible but exceedingly difficult given the $1/f^2$ roll-off of the GR noise PSD.

In the readout of KID-based phonon-mediated detectors, the GR noise level has been unobserved and is expected to be much lower than the noise level of other sources, largely due to the large volume resonators that are used. Although more realistic values for n_{qp}, say around $200\,\mu\text{m}^{-3}$ to $1000\,\mu\text{m}^{-3}$, would degrade the GR noise contribution to $\sigma_{E_{\text{abs}}}$ by a factor of 3 to 6, the contribution is still subdominant to other sources of noise.

7.3 Two-level system noise

Exposure of a superconducting microwave resonator to the atmosphere inevitably leads to the formation of surface oxides, such as aluminum oxide or silicon dioxide. Two-level systems (TLS) are present in surface oxides as a result of their amorphous atomic structure. Figure 7.3 Left illustrates how atoms that tunnel between adjacent available sites can form a two-level system. The energy separation between the two levels is denoted as E. See Phillips (1987) for a review of how these tunneling states affect the material properties of amorphous materials.

Each atom-vacancy pair that forms a two-level system carries an electric dipole moment, as indicated in Figure 7.3 Left with the arrows. The electrical response of an amorphous material to an external electromagnetic wave depends on the aggregate polarizability of these dipole moments or, in other words, on the electric permittivity ϵ_h of the TLS-hosting material. For superconducting microwave resonators, fluctuations in the capacitor's permittivity manifest as fluctuations in the resonant frequency.

Here are two ways in which two-level systems may mediate fluctuations in the permittivity of an amporphous material.

- Two-level systems with transition energy E roughly equal to the readout photon energy hf_r can randomly switch states via absorption or emission of a phonon. When an EM-resonant TLS switches states, the dipole moment flips, and the reactive load of the TLS on the resonator switches sign. Random state switching as a result of thermal phonons occurs frequently if $kT \gg hf_r$ (note that $100\,\text{mK} \approx 2\,\text{GHz} \times h/k$). However, even if $kT \ll hf_r$, the resonant TLS may absorb the photon energy and reside in the upper state, and spontaneous emission can still occur.
- Two-level systems with $E \sim hf_r$ may fall out of resonance with the EM wave as a result

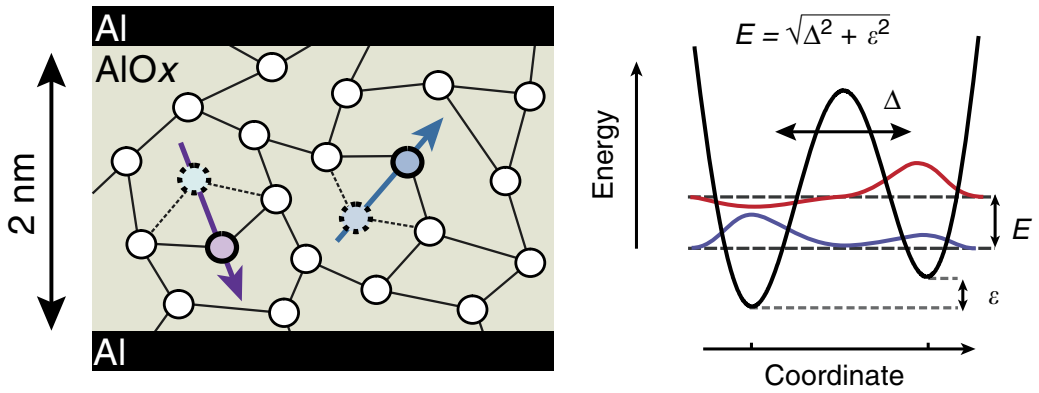

Figure 7.3: **TLS basics.** Left: schematic of two-level systems in an aluminum-oxide Josephson junction. For resonators, oxides may form on their surfaces. The two-level systems are shown as the purple and blue atoms that may tunnel between two nearby sites in the amorphous material, creating electric dipole moments as indicated by the arrows. Right: a two-level system where the levels are separated by an energy ε and the tunneling amplitude is given by Δ (we use δ throughout instead).[4] The red and blue curves are the two eigenstates of the diagonalized Hamiltonian with eigenenergies separated by $E = \sqrt{\delta^2 + \varepsilon^2}$. Both are reproduced from Lisenfeld et al. (2015).

of phonon interactions with nearby TLSs that have $E \lesssim kT$. When a nearby TLS switches states, it influences the strain field experienced by an EM-resonant TLS. This effect can also be interpreted as the coupling between the elastic dipole moments of the TLSs. The EM-resonant (but not thermally active) TLS may fluctuate in and out of resonance as a result of the thermally active (and noisy) TLSs, causing the permittivity at the readout frequency to fluctuate. These fluctuations can occur even if $kT \ll hf_r$, so long as there are TLSs with $E \lesssim kT$. This case is explored in Faoro et al. (2015).

TLS noise formalism

We now provide a brief overview of the mathematical formalism for TLS noise, following Phillips (1987), Gao (2008) and Faoro et al. (2015). The basic two-level Hamiltonian H_0 with energy splitting ε and tunneling amplitude δ can be expressed in terms of Pauli matrices $\sigma_{x,y,z}$:

$$H_0 = \frac{1}{2}\begin{pmatrix} \varepsilon & \delta \\ \delta & -\varepsilon \end{pmatrix} = \frac{1}{2}\sigma_z\varepsilon + \frac{1}{2}\sigma_x\delta = \vec{\sigma}\cdot\vec{a}, \tag{7.24}$$

where $\vec{a} = (\delta, 0, \epsilon)$. The eigenenergies for this Hamiltonian are separated by energy $E = |\vec{a}| = \sqrt{\varepsilon^2 + \delta^2}$. The eigenstates of this system are linear combinations of the no-tunneling basis states,[5] as reflected by the red and blue curves of Figure 7.3 Right (the figure uses Δ for the tunneling amplitude whereas we use δ). We refer to these eigenstates as the tunneling basis states.

A two-level system may interact with a number of things in its environment. We list some interaction possibilities with the following Hamiltonian in the no-tunneling basis and provide

[4] The depiction of the tunneling amplitude Δ as the width of the tunneling barrier is perhaps misleading. As discussed in Phillips (1987), the tunneling amplitude is approximately proportional to $\exp(-d(2mV_0/\hbar^2)^{1/2})$, where d is the well separation, V_0 is the tunneling barrier height, and m is the particle mass.

[5] In other words, the vector of no-tunneling basis states is multiplied by a 2×2 rotation matrix with angle θ to produce a vector of tunneling basis states. θ satisfies $\tan 2\theta = \delta/\varepsilon$.

descriptions for each term:

$$H = H_0 + H_{\text{int}}^{\text{phonon}} + H_{\text{int}}^{\text{field}} + H_{\text{int}}^{\text{TLS}} = H_0 + \sigma_z \gamma e + \sigma_z \vec{d}_0 \cdot \vec{\mathcal{E}} + \frac{1}{2} \sum_{i,j} \frac{u_{i,j}}{r_{i,j}^3} \sigma_z^i \sigma_z^j. \tag{7.25}$$

Descriptions:

- $H_{\text{int}}^{\text{phonon}} = \sigma_z \gamma e$: the TLS is embedded within the amorphous atomic structure and therefore can emit or absorb phonons when it transitions between its two eigenstates. e is the strain field, and γ is a coupling constant and can be conceptualized as an elastic dipole moment. The tensorial and vector nature of e and γ have been ignored. The σ_z term reflects the fact that the physical displacement is maximized between the no-tunneling basis states.
- $H_{\text{int}}^{\text{field}} = \sigma_z \vec{d}_0 \cdot \vec{\mathcal{E}}$: the material may be probed by an external electromagnetic wave with an electric field $\vec{\mathcal{E}}$ that couples to the TLS's electric dipole moment $\vec{d}_0$. As calculated in Gao (2008) (in the tunneling basis) and mentioned in Faoro et al. (2015), the dynamics produced by $H_0 + \sigma_z \vec{d}_0 \cdot \vec{\mathcal{E}}$ is equivalent to the dynamics of a spin-$\frac{1}{2}$ particle in a magnetic field with a large time-constant component and a small time-varying component. The latter has been studied in great detail for its application to nuclear magnetic resonance and MRI.

  The calculation essentially involves solving the Bloch equations for an ensemble average of TLS/spin-$\frac{1}{2}$ particles that are thermally coupled to some bath at a temperature T. We highlight the main results and point out correspondences in parentheses: (1) T_1 is an aggregate relaxation time for the ensemble of TLSs (spin-$\frac{1}{2}$ particles) to emit thermal phonons (photons) and reach an equilibrium configuration, given by the vector on the Bloch sphere that is collinear with $\vec{a}$ (the static component of the magnetic field); and (2) T_2 is the transverse relaxation time or the dephasing time between TLSs. Conceptually, vectors on the Bloch sphere precess around the equilibrium direction, and the precession rate of a given TLS depends on $E = \sqrt{\varepsilon^2 + \delta^2}$. For EM-resonant TLSs, interactions with other TLS may cause fluctuations in E and therefore the phase between them. T_1 and T_2 as decoherence times are also relevant in qubit technologies.
- $H_{\text{int}}^{\text{TLS}} = \frac{1}{2} \sum_{i,j} \frac{u_{i,j}}{r_{i,j}^3} \sigma_z^i \sigma_z^j$: the two-level system might interact with nearby two-level systems via phonon exchange, where $u_{i,j}$ is an interaction strength and $r_{i,j}$ is the distance between two TLSs. Faoro et al. (2015) points out that the maximum radius of nearby TLSs to an individual TLS can be expressed as $R_0 = (U_0/\Gamma_2)^{1/3}$, where U_0 is the interaction strength and Γ_2 is a T_2-affiliated energy loss linewidth resulting from phonon interactions with a single thermally active TLS. As the temperature drops, Γ_2 decreases, causing R_0 to increase and the number of nearby TLSs to increase. The result is that TLS noise actually decreases with temperature, which is atypical for a noise source.

Through the aforementioned mechanisms, TLS can alter the permittivity of a TLS-hosting material by an amount $\delta\epsilon_h = \delta\epsilon_1 - i\delta\epsilon_2$. By solving a set of Bloch equations and relating the net dipole moment to the applied electric field, an average $\delta\epsilon_h$ can be expressed in terms of d_0, f_r, T_2, T, and $E_{\max}$; see Gao (2008) for more details and a full expression, which involves a digamma function. T_1 is irrelevant for a TLS that resonates at $f_r \sim$ few GHz since most $T_1 \gtrsim 1\,\mu$s for $T \lesssim 1\,$K (the T_1-affiliated energy loss linewidth Γ_1 for thermally active TLSs, however, is important and determines Γ_2 for the EM-resonant TLS). The resulting fractional frequency change of a cavity resonator can be derived from cavity perturbation theory (Pozar, 2005), where $\delta\epsilon_h$ resides in a volume V_h:

$$\frac{f_r - f_{r,0}}{f_{r,0}} = -\frac{U_h}{U} = -\frac{\int_{V_h} \delta\epsilon_1(\vec{r},t) |\vec{\mathcal{E}}(\vec{r})|^2 d^3\vec{r}}{2 \int_V \epsilon |\vec{\mathcal{E}}(\vec{r})|^2 d^3\vec{r}}. \tag{7.26}$$

U_h is the electrical energy stored in the TLS-hosting material, and U is the electrical energy stored in the unperturbed resonator with volume V. The factor of 2 arises from the energy stored in the magnetic field. $\delta\epsilon_h(\vec{r},t)$ also depends on the many other physical parameters listed previously. The imaginary component $\delta\epsilon_2$ causes the Q_i of the resonator to shift.

For spectral information about $\delta f_r/f_r$ noise, the autocorrelation function of the permittivity would need to be calculated:

$$\begin{aligned}
\left\langle \frac{\delta f_r}{f_r}(t-\tau)\frac{\delta f_r}{f_r}(t)\right\rangle &= \frac{\left\langle \int_{V_{h,1}} \delta\epsilon_1(\vec{r}_1,t)|\vec{\mathcal{E}}(\vec{r}_1)|^2 d^3\vec{r}_1 \times \int_{V_{h,2}} \delta\epsilon_1(\vec{r}_2,t-\tau)|\vec{\mathcal{E}}(\vec{r}_2)|^2 d^3\vec{r}_2 \right\rangle}{4\left(\int_V \epsilon|\vec{\mathcal{E}}(\vec{r})|^2 d^3\vec{r}\right)^2} \\
&= \frac{\int_{V_{h,1}}\int_{V_{h,2}} \langle \delta\epsilon_1(\vec{r}_1,t-\tau)\delta\epsilon_1(\vec{r}_2,t)\rangle\, |\vec{\mathcal{E}}(\vec{r}_1)|^2|\vec{\mathcal{E}}(\vec{r}_2)|^2 d^3\vec{r}_1 d^3\vec{r}_2}{4\left(\int_V \epsilon|\vec{\mathcal{E}}(\vec{r})|^2 d^3\vec{r}\right)^2} \\
&= \langle \delta\epsilon_1(t-\tau)\delta\epsilon_1(t)\rangle \frac{\int_{V_h}|\vec{\mathcal{E}}(\vec{r})|^4 d^3\vec{r}}{4\left(\int_V \epsilon|\vec{\mathcal{E}}(\vec{r})|^2 d^3\vec{r}\right)^2} \qquad (7.27)
\end{aligned}$$

$$\text{where} \quad \langle \delta\epsilon_1(\vec{r}_1,t-\tau)\delta\epsilon_1(\vec{r}_2,t)\rangle = \langle \delta\epsilon_1(t)\delta\epsilon_1(t-\tau)\rangle\, \delta^3(\vec{r}_1-\vec{r}_2). \qquad (7.28)$$

From the first line of Equation 7.27, we assume that $\delta\epsilon_1$ and $\delta f_r/f_r$ are stationary in time so that their autocorrelation functions only depend on the time difference τ. The delta function in Equation 7.28 implies that the fluctuations in $\delta\epsilon_1$ are not correlated in position and enables one of the volume integrals to be evaluated. Furthermore, we assume that the material is uniform and that TLS fluctuations are not a function of position, allowing the autocorrelation function to be taken out of the volume integral in the final line of Equation 7.27. To actually compute the autocorrelation function, remember that the average $\langle\cdot\rangle$ is performed over the time variable t. Cross-TLS time correlations as mediated by phonon interactions would need to be calculated, but doing so is highly non-trivial.

Since the power spectral density $J_{\text{TLS}}(f)$ can be computed from the Fourier transform ($\tau \to f$) of the autocorrelation function, we present the following parametrization from Gao (2008):

$$\begin{aligned}
J_{\text{TLS}}(f) &= \int_{-\infty}^{\infty} \left\langle \frac{\delta f_r}{f_r}(t-\tau)\frac{\delta f_r}{f_r}(t)\right\rangle e^{-i2\pi f\tau} d\tau \\
&\propto \int_{-\infty}^{\infty} \langle \delta\epsilon_1(t-\tau)\delta\epsilon_1(t)\rangle\, e^{-i2\pi f\tau} d\tau \\
&= J_\epsilon(f) = \frac{\kappa(f)}{\sqrt{|\vec{\mathcal{E}}|^2+\mathcal{E}_c^2}} \qquad (7.29)
\end{aligned}$$

In the final line, $J_\epsilon(f)$ is parametrized in terms of frequency dependence $\kappa(f)$ and electric field dependence. For field strengths beyond some critical field $\mathcal{E}_c$, the PSD obeys a $J_{\text{TLS}} \propto |\vec{\mathcal{E}}|^{-1} \propto P_g^{-1/2}$ relationship and is constructed in this way to match experimental observations, reflecting the semi-empirical nature of Equation 7.29. Faoro et al. (2015) describes one technique for computing $\kappa(f)$ in the case of EM-resonant two-level systems that interact with thermally active two-level systems; they derive a $1/f$ dependence.

Characteristics of TLS noise

TLS noise has been observed and studied in many superconducting microwave resonators. We report some basic observations about these measurements, gathered from Zmuidzinas (2012) and Gao (2008).

- TLS noise is only present in the $\delta f_r/f_r$ direction. TLS noise PSDs are commonly expressed in units of $(\delta f_r/f_r)^2/\text{Hz}$; note the unitless numerator.
- The TLS noise power spectral density generally follows this form:

$$J_{\text{TLS}}(f) = \frac{Af^{-n}}{1+(f/f_{\text{roll-off}})^2}. \tag{7.30}$$

  There are two observations to make about the shape of TLS noise:
  - The shape of the noise as reflected in the power spectral density follows a power law f^{-n} where n is usually less than 1 and generally around 0.5. The power law reflects the TLS density of states. $f^{-0.5}$ corresponds to a log-uniform density of states, i.e., $D(\delta) \propto 1/\delta$. Faoro et al. (2015) concludes that, for resonant TLSs that are coupled to thermally active TLSs, the PSD goes as f^{-1}.
  - The noise rolls off with the resonator ring down time: $f_{\text{roll-off}} = f_r/2Q_r = 1/2\pi\tau_{\text{res}}$.
- Figure 14 from Zmuidzinas (2012) compares the TLS noise level from many different superconducting resonators. The TLS noise PSD at 1 kHz is used as the fiducial value for comparison between resonators, yielding

$$J_{\text{TLS}}(f) = \frac{J_{\text{TLS}}(1\,\text{kHz})\left(\frac{|f|}{1\,\text{kHz}}\right)^{-n}}{1+(f/f_{\text{roll-off}})^2}. \tag{7.31}$$

- TLS noise varies with readout power as $P_g^{-0.5}$ and with temperature as $T^{-\beta}$ where $\beta = 1.5$ to 2. At low temperatures, Figure 5.19 of Gao (2008) shows a flattening of the noise level, suggestive of the $\tanh hf_r/2kT$ dependence seen in other quantities like Q_i (Equation 5.75 of Gao, 2008).
- TLS noise depends on the capacitor geometry. This dependence comes from the electric field strength, Equations 5.78 to 5.80 of Gao (2008).
- TLS noise is material dependent. Different materials have different propensities for the formation of surface oxides and those oxides may have different levels of TLS.

Expected energy resolution

Given the form of TLS noise expressed in Equation 7.31, the impact on the intrinsic resolution of the detector may be computed. For this calculation, we assume that the fall time of the pulse is longer than the ring down time of the resonator; in other words, we assume that the signal bandwidth is smaller than the noise bandwidth. We also assume that the slope of the TLS power spectral density is $n = 1/2$.

$$\begin{aligned}
\sigma^2_{\delta f_r/f_r} &= \left[\int_{-\infty}^{\infty} df \frac{|\tilde{s}(f)|^2}{J_{\text{TLS}}(f)}\right]^{-1} \qquad J_{\text{TLS}}(f) = \frac{J_{\text{TLS}}(1\,\text{kHz})\left(\frac{|f|}{1\,\text{kHz}}\right)^{-1/2}}{1+(f/f_{\text{roll-off}})^2} \qquad \tilde{s}(f) = \frac{\tau_{\text{qp}}}{1+i2\pi f\tau_{\text{qp}}} \\
&= \left[\frac{1}{J_{\text{TLS}}(1\,\text{kHz})\sqrt{1\,\text{kHz}}}\int_{-\infty}^{\infty} df \frac{\tau_{\text{qp}}^2|f|^{1/2}}{1+4\pi^2f^2\tau_{\text{qp}}^2}\right]^{-1} \qquad x = 2\pi f\tau_{\text{qp}} \quad df = \frac{dx}{2\pi\tau_{\text{qp}}} \\
&= \left[\frac{\tau_{\text{qp}}^{1/2}}{(2\pi)^{3/2}J_{\text{TLS}}(1\,\text{kHz})\sqrt{1\,\text{kHz}}}\int_{-\infty}^{\infty} dx \frac{|x|^{1/2}}{1+x^2}\right]^{-1} \qquad 2\int_0^{\infty} dx \frac{x^{1/2}}{1+x^2} = \sqrt{2}\pi
\end{aligned}$$

$$= 2\sqrt{\pi}\frac{J_{\mathrm{TLS}}(1\,\mathrm{kHz})}{\sqrt{\tau_{\mathrm{qp}}\cdot 1\,\mathrm{ms}}}. \tag{7.32}$$

Note that the denominator of Equation 7.32 is the geometric mean of the pulse fall time τ_{qp} with 1 ms. In Figure 14 of Zmuidzinas (2012), the lowest reported $J_{\mathrm{TLS}}(1\,\mathrm{kHz})$ is $10^{-21}\,(\delta f_r/f_r)^2/\mathrm{Hz}$. We input that value, as well as a quasiparticle lifetime of 2.2 ms from de Visser et al. (2011). Values for α ($= 0.038$), γ ($= 1$), and κ_2 ($= 4.3\times10^7\,\mu\mathrm{m}^3$) that have been observed for KIPM detectors (Wen et al., 2022) are also plugged in for conversion to absorbed energy.

$$\sigma_{\delta f_r/f_r} = 1.5\times10^{-9}\left(\frac{J_{\mathrm{TLS}}(1\,\mathrm{kHz})}{10^{-21}\,(\delta f_r/f_r)^2/\mathrm{Hz}}\right)^{\frac{1}{2}}\left(\frac{2.2\,\mathrm{ms}}{\tau_{\mathrm{qp}}}\right)^{\frac{1}{4}} \tag{7.33}$$

$$\longrightarrow \sigma_{E_{\mathrm{abs}}} = \sigma_{\delta f_r/f_r}\frac{1}{\alpha\gamma\kappa_2}V_{\mathrm{ind}}\Delta = 510\,\mathrm{meV}. \tag{7.34}$$

$V_{\mathrm{ind}} = 3\times10^5\mu\mathrm{m}^3$ and $\Delta = 0.18\,\mathrm{meV}$ have also been used. So, even in the optimistic scenario where the TLS noise level is at the lowest that has been observed, energy resolutions are limited to the $O(100\,\mathrm{meV})$ level for current design values. Furthermore, the improvement to the energy resolution with the quasiparticle lifetime τ_{qp} is a slower function of τ_{qp} and thus n_{qp} than in the GR-noise dominated case (quartic root versus linear).

7.4 Johnson, amplifier, and readout chain noise

Kinetic inductance detectors require very small amounts of feedline readout power P, typically less than 10 pW ($= -80\,\mathrm{dBm}$) and sometimes as low as 1 fW ($= -120\,\mathrm{dBm}$). Signals of this power can be produced at room temperature, but the 50 Ω output impedance of the source residing at 300 K would emit Johnson (blackbody) noise, which is given by

$$\sigma_P = kT\Delta f, \tag{7.35}$$

where T is the noise temperature and Δf is the readout bandwidth. Equation 7.35 expresses power fluctuations σ_P in terms of the equivalent power fluctuations that arise from a resistor at temperature T due to electron thermal movement; thus arises the term "noise temperature." Throughout this section, noise is referenced in units of temperature. As a starter example, the Johnson noise at 300 K over a readout bandwidth of 1 MHz[6] is about 4 fW and would actually exceed the desired signal at the low end of the aforementioned readout power range.

Improved signal-to-noise ratio at the desired readout power can be achieved by producing a signal tone at room temperature with a larger power than needed and then attenuating the signal with cryogenic resistors for which the Johnson noise is smaller than room temperature noise. For this section, we assume that signals of any size can be produced with room temperature readout noise, but this assumption is certainly not true for all signal powers.

Attenuators and amplifiers are the basic tools with which we can control the signal and noise level of our readout. The equations that describe the effect of these components on the signal and noise of a tone are the following:

$$P_{\mathrm{out}} = \frac{P_{\mathrm{in}}}{L_{\mathrm{atten}}} \qquad T_{\mathrm{out}} = \frac{T_{\mathrm{in}}}{L_{\mathrm{atten}}} + T_L\left(1-\frac{1}{L_{\mathrm{atten}}}\right) \tag{7.36}$$

$$P_{\mathrm{out}} = P_{\mathrm{in}}G_{\mathrm{amp}} \qquad T_{\mathrm{out}} = (T_{\mathrm{in}} + T_{\mathrm{amp,in}})\,G_{\mathrm{amp}}, \tag{7.37}$$

[6] 1 MHz readout bandwidth corresponds to a sampling rate of 2 MHz, or two samples every 1 μs. 1 μs is about the shortest time interval over which we expect there to be relevant signal information.

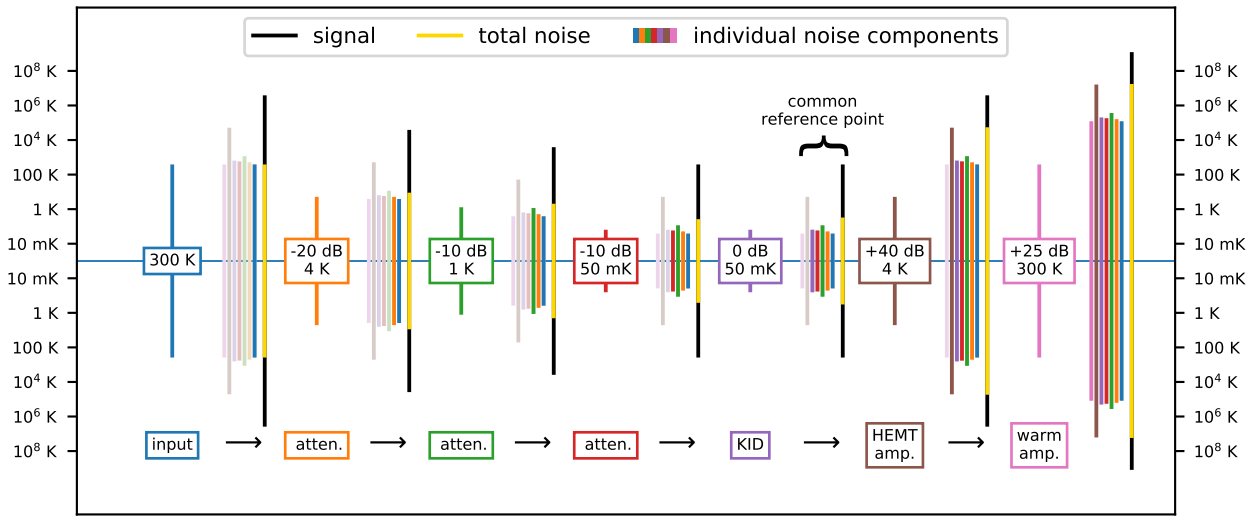

Figure 7.4: **Signal and noise power levels as they evolve along an example readout chain.** Full color individual noise component lines represent physically realized sources of noise as they are incrementally introduced along the readout chain; their sum is shown in gold. Faded color component lines represent the various components of the final readout noise as referred to different locations along the readout chain.

where P_{in} & P_{out} are the signal power at the input and output of the component, T_{in} & T_{out} are the total noise level of the signal tone at the input and output of the component, L_{atten} & G_{amp} are the loss of the attenuator and the gain of the amplifier, T_L is the temperature of the attenuator, and $T_{\text{amp, in}}$ is the input noise of the amplifier. Observations of these equations:

- Equation 7.36 shows that input power and noise are attenuated by a factor of L_{atten}. The second term of the noise equation indicates that the attenuator adds noise $T_L(1 - 1/L_{\text{atten}})$. This term vanishes when $L_{\text{atten}} = 1$; i.e., a lossless line adds no noise.
- Equation 7.37 shows that the input power and noise are amplified by a factor G_{amp}. The noise that the amplifier adds has been expressed in terms of the noise at the input of the amplifier.

  Noise from a particular component can always be referred to a different spot on the readout chain by multiplying or dividing by the gain or loss along the chain. In this case, the physical noise from the amplifier that is realized at its output is $T_{\text{amp, in}}G_{\text{amp}}$; the gain is then divided out to infer the noise temperature of the amplifier as referred to the input of the amplifier $T_{\text{amp, in}}$. In practice, amplifier gain is difficult to measure precisely, so the input noise of an amplifier is instead usually measured using a Y-factor measurement.
- Attenuators and amplifiers always degrade the signal-to-noise ratio along a readout chain. The level of degradation depends on the relative values of T_L or $T_{\text{amp, in}}$ versus T_{in}.

Figure 7.4 presents an example readout chain and the evolution of the signal and noise as they progress through the chain, using Equation 7.36 and Equation 7.37 to introduce new noise contributions for each additional component along the chain. Each new component's contribution is plotted in full color and contributes to the total noise. The example begins with 300 K input noise and an input signal that is 10,000× larger than the noise, which would correspond to about 40 pW or roughly −74 dBm; see Equation 7.35. Since the plot has a logarithmic vertical axis, the signal-to-noise ratio at a given point in the chain is given by the length of the signal line in black

that is in excess of the noise line in gold. As Equation 7.36 and Equation 7.37 foreshadowed, the signal-to-noise ratio degrades after each step along the readout chain due to the introduction of new noise from each component.

At each point, the noise contributions from all components of the chain are plotted. If the bar is in full color, then it represents a physically realized noise source that contributes to the total noise. If the bar is in faded color, then it represents the noise of that source as referred to that point in the chain, i.e., the gains and/or losses of the intervening components have been divided out to put each source of noise on the same footing (see second observation above). We comment that the ratios of the sources of noise are independent of the position along the readout chain.

The three stages of cryogenic attenuation total 40 dB, reducing the signal to roughly −114 dBm. Importantly, the noise has been reduced to the sub-1 K level, a huge improvement over the non-cryogenic option of producing −114 dBm directly at room temperature. Cryogenic attenuation is a key strategy in producing low-noise readout tones and is relatively easy to implement into a readout chain. It should be noted that a single 40 dB attenuator located at the mixing chamber would have a similar effect of decreasing 300 K noise to sub-1K levels, but in this case, most of the 40 pW signal power would be dissipated at the mixing chamber. If higher readout powers are desired, then the power dissipation at the mixing chamber may become prohibitive. Since similar noise attenuation can be achieved with the chain of attenuators shown in Figure 7.4, the attenuator chain provides a greater dynamic range for signal power than a single attenuator at the mixing chamber.

After the attenuation, the signal reaches the kinetic inductance detector. Noise may be added to the readout tone as has been described in the previous two sections. That noise has been shown in purple and is drawn as a "0 dB" attenuator. 0 dB attenuators technically do not add any noise, per Equation 7.36. The KID has been drawn in Figure 7.4 as a 0 dB attenuator because the noise that it adds does not quite fit into the framework described by Equation 7.36 and Equation 7.37, but it can certainly add readout noise that is measurable in units of temperature via the mechanisms described in the previous two sections.

The dominant noise contributor appears after the KID: the noise of the first amplifier, in this case a high-electron-mobility transistor (HEMT) amplifier. Typical input noise values for HEMTs are between 2 K and 5 K. In Figure 7.4, the HEMT noise has been set to 4 K and dominates over other noise sources in the readout chain. Furthermore, even though the warm amplifier just downstream of the HEMT has a larger input noise, the HEMT gain ensures the HEMT noise remains dominant. Thus, the warm amplifier degrades the signal-to-noise as referred to the output of the HEMT by a negligible amount. This scenario is referred to as HEMT-dominated noise.

Improving the noise contribution from the first amplifier of a readout chain is in general quite difficult. For improvements beyond an order of magnitude, new technologies are needed. One such example that are explored in Section 7.9 is the kinetic inductance traveling wave parametric amplifier, commonly referred to as a KI-TWPA.

Expected energy resolution

We now compute the expected energy resolution for amplifier-dominated noise. For this subsection, it is useful to work in voltage units instead of power units. To convert from power fluctuations on a transmission line into voltage fluctuations, we modify Equation 7.35 in the following way:

$$\sigma_V^2 = \frac{1}{4} kTZ_0\Delta f, \tag{7.38}$$

where Z_0 is the characteristic impedance of the transmission line and V is an rms quantity. For a sine wave, $V_{\text{rms}} = V_{\text{peak}}/\sqrt{2}$. In our notation, $V = V_{\text{rms}}$. In some texts, $4kTR\Delta f$ is shown to be the

Johnson noise for a resistor of resistance R at a temperature T. Here is a full accounting of the various discrepant factors that lead to a factor of 16:

- For a transmission line, the Z_0 impedance source is in a circuit with a Z_0 impedance load. The voltage fluctuations appear across a resistance of $2Z_0$, which reduces the variance by 4.
- Δf is one-sided in these other texts, i.e., only positive frequencies are used.
- The total noise power is a sum of noise in two different quadratures, such as gain and phase. For the following calculation, we only examine the noise in one quadrature of readout.

By Parseval's theorem, the power spectral density for these voltage fluctuations is

$$J_{T_N}(f) = \frac{1}{4} k T_N Z_0. \tag{7.39}$$

We now work through the optimal filter calculation for amplifier-dominated noise:

$$\begin{aligned}
\sigma_V^2 &= \left[\int_{-\infty}^{\infty} df \frac{|\tilde{s}(f)|^2}{J_{T_N}(f)}\right]^{-1} \qquad J_{T_N}(f) = \frac{1}{4} k T_N Z_0 \quad \tilde{s}(f) = \frac{\tau_{\mathrm{qp}}}{1 + i 2\pi f \tau_{\mathrm{qp}}} \\
&= \left[\frac{4}{k T_N Z_0} \int_{-\infty}^{\infty} df \frac{\tau_{\mathrm{qp}}^2}{1 + 4\pi^2 f^2 \tau_{\mathrm{qp}}^2}\right]^{-1} \qquad x = 2\pi f \tau_{\mathrm{qp}} \quad df = \frac{dx}{2\pi \tau_{\mathrm{qp}}} \\
&= \left[\frac{4}{k T_N Z_0} \frac{\tau_{\mathrm{qp}}}{2\pi} \int_{-\infty}^{\infty} dx \frac{1}{1 + x^2}\right]^{-1} \\
&= \frac{k T_N Z_0}{2\tau_{\mathrm{qp}}} \\
\sigma_V &= \sqrt{\frac{k T_N Z_0}{2\tau_{\mathrm{qp}}}} = 3.8\,\mathrm{nV} \sqrt{\frac{T_N}{4\,\mathrm{K}} \frac{Z_0}{50\,\Omega} \frac{100\,\mu\mathrm{s}}{\tau_{\mathrm{qp}}}}.
\end{aligned} \tag{7.40}$$

This time, for amplifier noise, we first compute $\sigma_{S_{21}}$ (remembering that the $\cos\phi$ eventually cancels!) and then convert into $\sigma_{E_{\mathrm{abs}}}$ as usual. $\sigma_{S_{21}}$ is given by Equation 7.1, where $|a|$ is the rms voltage amplitude. For a $-80\,\mathrm{dBm}$ signal on a $50\,\Omega$ impedance feedline, $|a| = \sqrt{P_g Z_0} = 22.4\,\mu\mathrm{V}$.

$$\sigma_{S_{21}} = \frac{\sigma_V}{|a|} \cos\phi = 3 \times 10^{-4} \cos\phi \tag{7.41}$$

$$\sigma_{E_{\mathrm{abs}}} = \frac{\sigma_V}{|a|} \frac{|\hat{Q}_c|}{2Q_r^2} \frac{1}{\alpha\gamma\kappa_2} V_{\mathrm{ind}} \Delta = 280\,\mathrm{meV} \tag{7.42}$$

$Q_r^2/|\hat{Q}_c| = 100{,}000$ is chosen for this calculation and reflects a common design choice KID-based phonon-mediated detectors (see Section 7.7). The result of the calculation is that HEMT-dominated amplifier noise limits the resolution on absorbed energy to the $O(100\,\mathrm{meV})$ range, but the exact value is very sensitive to readout power, V_{ind}, and α.

7.5 Noise measurements in KIPM detectors

Experimental setup

In the following three sections, we report on a variety of noise measurements that were made over many different KIPM detectors. All noise measurements were performed in Cahill B101 in an Oxford Kelvinox 25 dilution refrigerator with operating temperatures ranging from 60 mK to 100 mK. Many different devices are discussed in the following sections; we provide a basic description of each of the resonators and relevant device parameters.

The readout circuit that was used for each of these measurements is the same as what is shown in Figure 7.4. The expectation with this readout circuit is that the noise is HEMT-dominated, but, in many cases, the resonator-internal TLS noise is also a significant contributor.

For all noise measurements that were performed in the Cahill B101 dilution refrigerator, we used a layer of μ-metal magnetic shielding, thermally sunk to the 1 K still stage and surrounding everything at the mK stage. Shielding from the Earth's magnetic field is critical for achieving high Q_i superconducting resonators.

All data in this section and the following chapter were taken with an Ettus Research Labs universal software radio peripheral (USRP). The USRP is flexible and relatively easy to use compared to its counterparts. The firmware and code used to operate the USRP was developed by Lorenzo Minutolo and is described in Minutolo et al. (2019).

Power spectral densities

Power spectral densities (PSDs) $J(f_n)$ of noise timestreams $x(t)$ for some real-valued variable x are computed with the function `welch` from the `scipy.signal` library, wherein the user provides the sample frequency f_s and window size N. The highest frequency bin produced by `welch` is $f_s/2$, the Nyquist frequency. The lowest frequency bin, which is also equal to the frequency spacing, is given by the lowest frequency wave that fits a full cycle in the duration of the window size: $1/(N/f_s) = f_s/N = \Delta f$. When N is even, discrete power spectral densities span from $-f_s/2$ to $f_s/2 - \Delta f$. Because all of the PSDs in this thesis are computed for real-valued timestreams, $J(-f_n) = J(f_n)$.[7] `welch` chooses to return one-sided PSDs $J_+(f_n)$ in this case, spanning from 0 to $f_s/2$, where the positive frequency contributions have been multiplied by 2: $J_+(f_n) = 2J(f_n)$. All PSDs produced in this thesis are technically one-sided PSDs as directly produced by `welch` but are referred to as $J(f_n)$, i.e., the one-sided notation is suppressed when referring to $J(f_n)$.

Parseval's theorem states that the variance of the noise timestream equals the integral of its PSD over frequency:

$$\langle x(t)^2 \rangle = \sum_{f_n=-f_s/2}^{f_s/2-\Delta f} J(f_n)\Delta f = \sum_{f_n=0}^{f_s/2} J_+(f_n)\Delta f. \tag{7.43}$$

In other words, the variance of a timestream receives independent and orthogonal contributions from the different frequency components that form a basis for the real-valued timestream space. As a result of Parseval's theorem, variance may be estimated quickly from a PSD plot by estimating some average value over all frequencies and then multiplying by the bandwidth $f_s/2$.

Two final notes on the PSDs plotted in this thesis: (1) `welch` averages the PSDs across M blocks of length N that make up a timestream of length $M \times N$. Prior to averaging, a Hann window is applied to each N-sample block. (2) All plotted PSDs omit the DC bin where $f = 0$.

7.6 Electronics $1/f$ noise and correlated-noise removal

Electronics $1/f$ noise is a relatively generic term that can arise from many different sources. One possible source for electronics $1/f$ noise is drifts in the total gain and/or transmission of a readout circuit. In general, this kind of drift has no roll-off frequency—it rises to as low a frequency as one can measure.

More specific to this application, the dominant electronics $1/f$ noise arises from the Ettus USRP readout system used for KIPM detector readout in this thesis. The $1/f$ noise from the

[7] This equality is a result of the fact that the Fourier coefficient at negative frequency is equal to the complex conjugate of the Fourier coefficient at positive frequency for a real-valued timestream. PSDs are computed from the square of the fourier coefficients.

USRP is believed to be caused by the different local oscillators for up- and down-conversion between the USRP baseband and the few GHz RF band of our resonators, in particular the phase noise engendered by using two different LOs. This noise manifests in the "electronics arc length" direction.

For both gain/transmission and USRP phase noise, the noise is correlated between on-resonance and off-resonance tones. Through simultaneous readout of on- and off-resonance tones, the correlated noise can be tracked and removed ("cleaned").

Defining the "radial" and "arc length" directions

In general, $S_{21}(t, f_r)$ or $V(t, f_r)$ timestreams are complex and thus have two quadratures. The USRP reports voltages in terms of complex quantities with real part I and complex part Q, so $S_{21}(t, f_r)$ or $V(t, f_r)$ chart out trajectories in the complex IQ-plane. Another natural basis is the gain and phase basis, which is equivalent to converting the complex number $I + iQ$ into $re^{i\theta}$, where r corresponds to gain and θ to phase.

The USRP noise under study was found to be well-aligned along this latter basis and was particularly egregious in the phase direction. In this thesis, we transform the gain and phase quadratures into radius and arc length timestreams such that the two orthogonal timestreams carry the same units as the IQ-plane. The cleaning algorithm is then performed separately on the radius and arc length directions.

The USRP reports voltages as a fraction of the ADC full scale voltage, which are referred to as "ADC units" in this thesis; we perform no conversions of ADC units into voltage units. All radius and arc length units are thus reported in the generic "ADC units" or "ADCu."

The basic cleaning procedure

KIDs provide a natural way to remove correlated electronics noise by the simultaneous readout of off-resonance tones. In the simple case where there is a single on-resonance tone with real timestream $D(t)$ and a single off-resonance tone with real timestream $S(t)$, the on-resonance tone can be cleaned of the correlated noise with the off-resonance tone using the following transformation:

$$D'(t) = D(t) - xS(t), \tag{7.44}$$

where x is referred to as a cleaning coefficient. In this case, $D(t)$ and $S(t)$ may either be radius or arc length timestreams. The variance of a "cleaned" tone $D'(t)$ can then be minimized in a way that subtracts the correlated noise between $D(t)$ and $S(t)$:

$$\begin{aligned}
\mathrm{Var}(D'(t)) &= \mathrm{Var}(D(t) - xS(t)) \\
&= \mathrm{Var}(D'(t)) + x^2\mathrm{Var}(S(t)) - 2x\mathrm{Cov}(D(t), S(t)) \\
\longrightarrow \frac{\partial \mathrm{Var}(D'(t))}{\partial x} &= 2x\mathrm{Var}(S(t)) - 2\mathrm{Cov}(D(t), S(t)) \\
\longrightarrow x &= \frac{\mathrm{Cov}(D(t), S(t))}{\mathrm{Var}(S(t))},
\end{aligned} \tag{7.45}$$

where in the final line, $\partial \mathrm{Var}(D'(t))/\partial x$ has been set to 0, and the equation is then solved for x. As mentioned earlier, this procedure is performed separately for the radius and arc length basis directions.

This cleaning procedure is demonstrated in Figure 7.5. The data shown are for a niobium resonator of the same design as from Chang (2023) and Aralis (2024). The raw data from the USRP is shown in Figure 7.5 Left, for both the VNA data $V(f)$ and the timestream data $V(t; f)$ for both the on-resonance and off-resonance tone. The radius and arc length basis directions have been

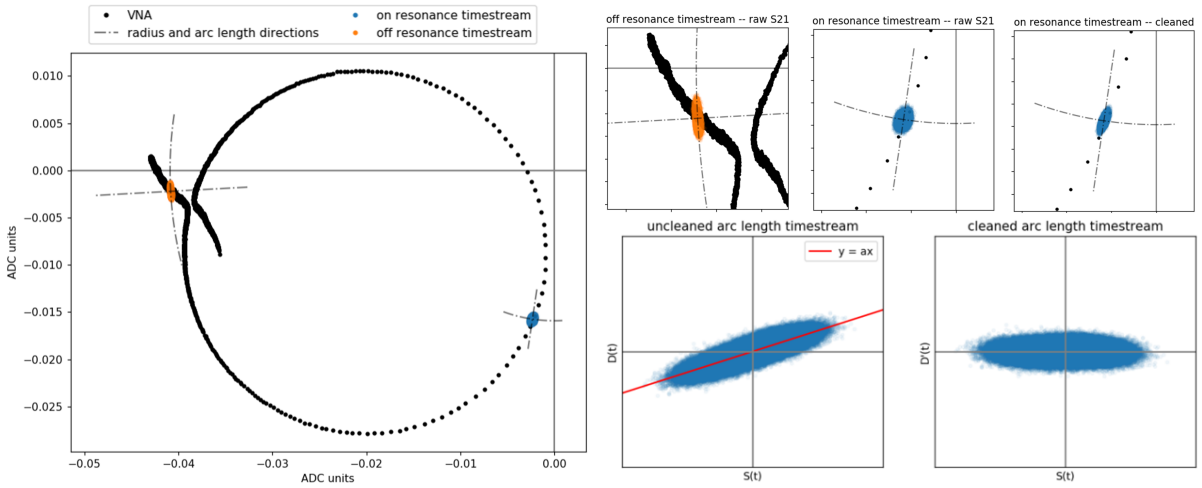

Figure 7.5: **Correlated noise cleaning in KIDs.** Left: raw USRP $V(f)$ in black and $V(t;\, f)$ in orange and blue in the complex plane. The radius and arc length directions are marked as well. Top Right: the same data, zoomed into the two noise blobs. The plot in the far upper right shows the on-resonance timestream after it has been cleaned of the correlated noise with the off-resonance timestream and has the same scale as the plot to its left. Bottom Right: a representation of the cleaning algorithm in the $D(t)$ versus $S(t)$ and $D'(t)$ versus $S(t)$ planes for the arc length direction of the complex timestreams.

indicated and are defined with respect to the origin of the IQ plane, which is located at 0 ADC units.

Figure 7.5 Bottom Right then shows how the correlated noise is removed in a particular direction, in this case the arc length direction. The on- and off-resonance timestreams in the arc length direction $D(t)$ and $S(t)$ are seen to be strongly correlated. The slope of the correlation can be computed with Equation 7.45 and is shown to be well aligned with the major axis of the noise blob in the $D(t)$ versus $S(t)$ plane. The plot of $D'(t)$ versus $S(t)$ then shows how the subtraction of $xS(t)$ from $D(t)$ leads to substantial noise reduction of $D'(t)$ in comparison to $D(t)$. In Figure 7.5 Top Right a reduction in the size of the noise blob in the arc length direction is visible.

Cleaning in the radius direction for this niobium resonator had little impact. The fact that the remaining dominant noise is primarily in the frequency direction and uncorrelated with off-resonance tones is evidence that it is TLS.

A sometimes critical step in the calculation of the cleaning coefficient is additional decimation (low pass filtering & down-sampling). This may be required so that the high frequency noise does not dominate in the calculation of the cleaning coefficient. The $1/f$ noise is largest at low frequencies, so it is important that the noise is dominated by the correlated noise in order to correctly calculate the cleaning coefficient. One can calculate x with decimated data and then use that coefficient x for the subtraction of the un-decimated $S(t)$ from the un-decimated $D(t)$.

Correlated-noise removal with multiple cleaning tones

For validation of the cleaning algorithm, it may be desired to use multiple off-resonance tones, so that the off-resonance tones can be used to clean themselves, thus providing a resonator-agnostic procedure for validation of the cleaning algorithm. When cleaning the off-resonance tones, the on-resonance tone is not used. With N off-resonance tones S_n that require N cleaning coefficients

x_n, the cleaning algorithm can be generalized in the following way (t dependence is now suppressed):

$$\begin{aligned} D' =& D - \sum_n x_n S_n \\ \mathrm{Var}(D') =& \mathrm{Var}(D - \sum_n x_n S_n) \\ =& \mathrm{Var}(D) + \mathrm{Var}(\sum_n x_n S_n) - 2\mathrm{Cov}(D, \sum_n a_n S_n) \\ =& \mathrm{Var}(D) + \sum_n x_n^2 \mathrm{Var} S_n + \sum_{n\neq m} \sum_m x_n x_m \mathrm{Cov}(S_n, S_m) - 2\sum x_n \mathrm{Cov}(D,S_n) \\ \frac{\partial \mathrm{Var} D'}{\partial x_n} =& 2x_n \mathrm{Var}(S_n) + 2\sum_{m\neq n} x_m \mathrm{Cov}(S_n, S_m) - 2\mathrm{Cov}(D, S_n) \qquad (7.46) \\ \longrightarrow \nabla \mathrm{Var} D' =& 2\mathbf{K_{S,S}x} - 2\mathbf{b} \qquad (7.47) \end{aligned}$$

where $\mathbf{K_{S,S}}$ is the covariance matrix of the N cleaning tones, $\mathbf{x}$ is a vector of the N cleaning coefficients, and $\mathbf{b}$ is a vector whose elements are $\mathrm{Cov}(D, S_n)$. The gradient in the final line is taken with respect to the cleaning coefficients x_n. To minimize the variance of D' under the cleaning transformation, we find a local minimum by setting the gradient equal to 0:

$$\mathbf{K_{S,S}x} = \mathbf{b}, \qquad (7.48)$$

which can be solved for $\mathbf{x}$. In the case where there is only one cleaning tone S for which we desire the optimal cleaning coefficient x, Equation 7.48 simplifies to the single-tone equation for the cleaning coefficient written in Equation 7.45.

The results of a multiple-cleaning-tone algorithm are displayed in Figure 7.6. The resonator under study in this example is a TiNx resonator of the same design as B240103 described in Section 8.4, but moved up in frequency. For this example, three cleaning tones were chosen, the minimum number such that the multi-tone cleaning algorithm could be used for the off-resonance tones. In all plots, the gray line shows the frequency at which all timestreams were decimated for the determination of cleaning coefficients: down to 1 kHz, or a 500 Hz Nyquist frequency. This level of decimation was chosen so that the variances of the noise timestreams were dominated by the low-frequency correlated noise.

In the on-diagonal plots, power spectral densities (PSDs) are plotted before and after cleaning in both the radius (blue) and arc length (orange) directions. The off-resonance tones show a capacity to clean below $10^{-13}\mathrm{ADCu}^2/\mathrm{Hz}$ in power spectral density, exhibiting flat noise characteristic of amplifier-dominated white noise. The on-resonance PSDs show excess noise above $10^{-13}\mathrm{ADCu}^2/\mathrm{Hz}$ in both the radius and arc length directions, the source of which is believed to be two-level systems in the resonator as detailed in Section 7.8.

The only off-resonance timestream that did not demonstrate cleaning below $10^{-13}\mathrm{ADCu}^2/\mathrm{Hz}$ was the arc length direction timestream of off-resonance tone 1, which also showed the greatest transmission of the four tones. Assuming that the $1/f$ noise is proportional to the carrier amplitude, cleaning a tone with amplitude a_1 using a tone with amplitude a_2 results in a_2's white noise σ_2 appearing in the cleaned a_1 with amplitude $(a_1/a_2)\sigma_2$. If $a_1 \gg a_2$, then the white noise of a_2 is greatly amplified.

The off-diagonal plots show the coherence between the uncleaned and cleaned tones. Coherence is defined as

$$C(f;\, S_1, S_2) = \frac{|J_{S_1,S_2}|^2}{J_{S_1} J_{S_2}} \qquad (7.49)$$

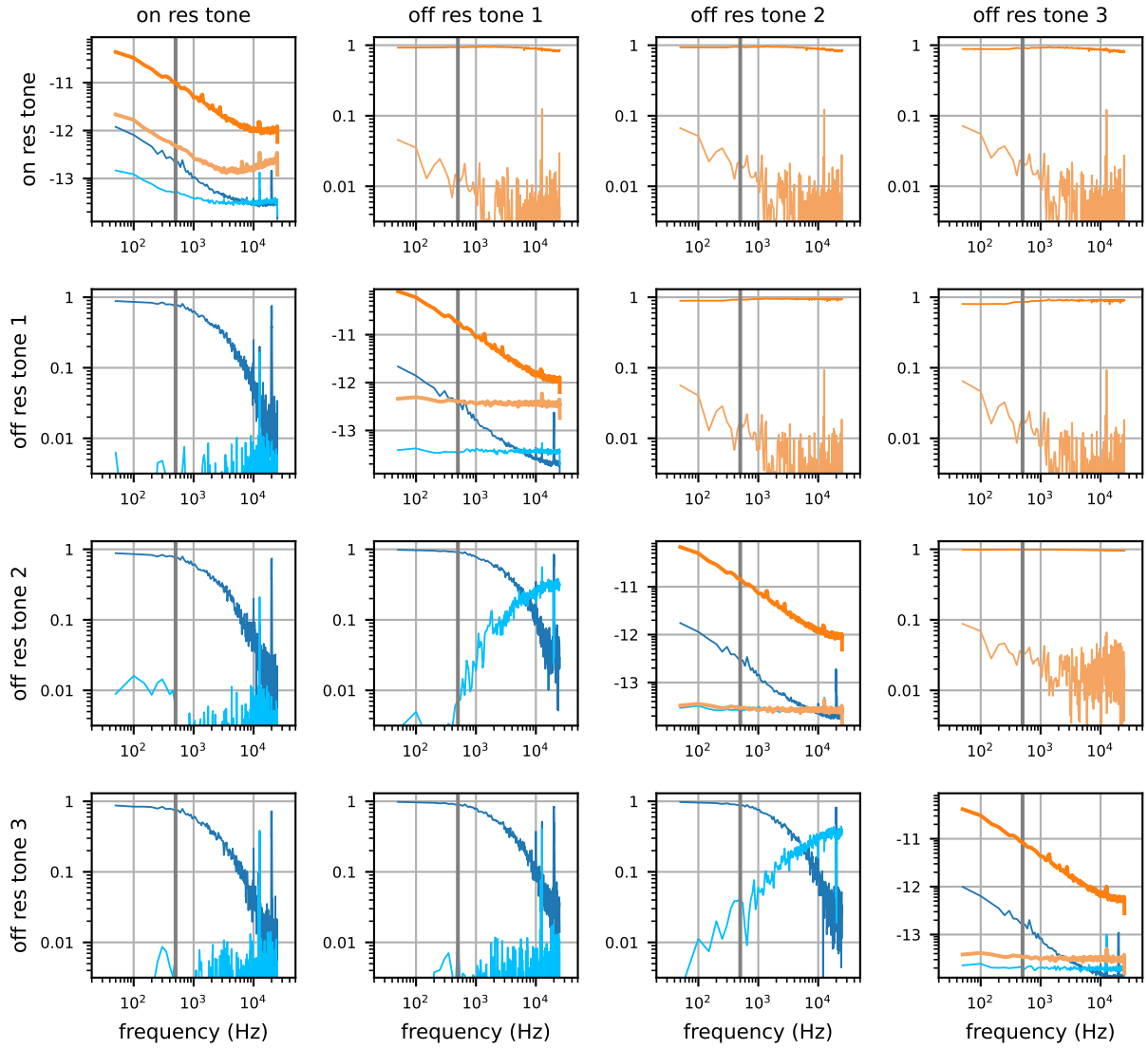

Figure 7.6: **Multi-tone cleaning in resonator readout.** On-diagonal plots show power spectral densities in units of $\log_{10}(\mathrm{ADCu}^2/\mathrm{Hz})$, and off-diagonal plots show coherence (defined in text) between two particular tones, as specified by their row and column headers. Both off- and on-diagonal plots show the effects of cleaning for both the radius and arc length timestreams: dark blue corresponds to uncleaned radius timestreams; light blue are their cleaned counterparts; dark orange corresponds to uncleaned arc length timestreams; light orange are their cleaned counterparts.

and is a useful quantity for evaluating the level of correlation between two timestreams. A single subscript J is a power spectral density, and a double subscript J is a cross spectral density. In general, the cleaned tones are much less coherent with the other tones after cleaning is performed. In fact, the pre-cleaning coherence values are near 1 across nearly the entire plotted bandwidth for all pairs of tones in Figure 7.6.

The key takeaway from Figure 7.6 is that off-resonance tones can be used to clean each other down to white noise, which implies that any residual $1/f$ noise in the on-resonance timestream after cleaning is intrinsic to the resonator. The remaining noise in Figure 7.6 Top Left is uncorrelated with the off-resonance tones. As we see in Section 7.8, the remaining noise is entirely in the frequency direction and therefore can be identified as TLS noise.

7.7 An amplifier-noise-dominated KIPM detector

YY180726 is an 80-resonator device on a 75 mm diameter, 1 mm thick silicon wafer. The resonators are made from 30 nm aluminum and the feedline from 300 nm niobium. The resonant frequencies are located between 3.1 GHz and 3.55 GHz, intended to align with a ROACH-1 ADC/DAC board readout architecture used for MUSIC(Duan et al., 2010). For more details on the design of this device, see Chang (2023) and Aralis (2024).

YY180726 exhibited hugely varying transmission across its operating frequencies. Figure 7.7 shows transmission variation of around 60 dB. Given the few cm size of the device box and the results of an Eccosorb test described in Chang (2023), it is believed that the coupling of the RF modes of the feedline to the RF modes of the box is the main cause of the transmission non-idealities. The effect of these excitations may have been worsened by the somewhat atypical grounding scheme that was used for the coplanar waveguide transmission line. Specifically, the ground plane does not cover the entire wafer and so is perhaps poorly grounded to the box given its meandering structure over the wafer; see Chapter 8 for a description of a similarly designed detector. Chang (2023) discusses these effects in more detail as well as potential strategies to improve transmission non-idealities.

Since the vast majority of resonators exhibited poor transmission and poor Q_r, the readout noise was dominated by the amplifier. Recalling Equation 7.2,

$$\sigma_{\delta f_r/f_r} = \sigma_{S_{21}} \frac{Q_c}{2Q_r^2} = \frac{\sigma_V}{|a|} \cos\phi \frac{|\hat{Q}_c|}{2Q_r^2} \frac{1}{\cos\phi} = \frac{\sigma_V}{|a|} \frac{|\hat{Q}_c|}{2Q_r^2} \tag{7.2 revisited}$$

$\sigma_{\delta f_r/f_r}$ can be expressed in terms of $|a|$, $|\hat{Q}_c|$, and Q_r, three quantities that are fit for and extracted from raw $V(f)$ data or raw $S_{21}(f)$ data. Furthermore, if σ_V is mostly fixed in the raw $V(f)$ plane across probing tone frequencies and thus fixed across different resonators, the performance of a particular resonator in terms of $\sigma_{\delta f_r/f_r}$ is dominantly dependent on the quantity $|a|Q_r^2/|\hat{Q}_c|$. We define this quantity:

$$R_Q \equiv |a| \frac{Q_r^2}{|\hat{Q}_c|}, \tag{7.50}$$

and we refer to R_Q as the "RF responsivity." We remark again that a can be extracted from either $V(f)$ or S_{21} data. In the case of the former, $|a|$ is larger for larger readout powers and thus can be used to facilitate cross-readout-power comparisons; for the latter, $|a|$ does not depend on readout power. A quantity similar to R_Q that omits the dependence on $|a|$ was first introduced in Ramanathan, Aralis, et al. (2022).

The key result in Figure 7.7 is that the variation in resonator noise performance is due to variation in RF responsivity. The full bandwidth level of $\delta f_r/f_r$ noise is shown in Figure 7.7 Bottom Left and displays the tightest trend between noise level and R_Q of the three plots in Figure 7.7

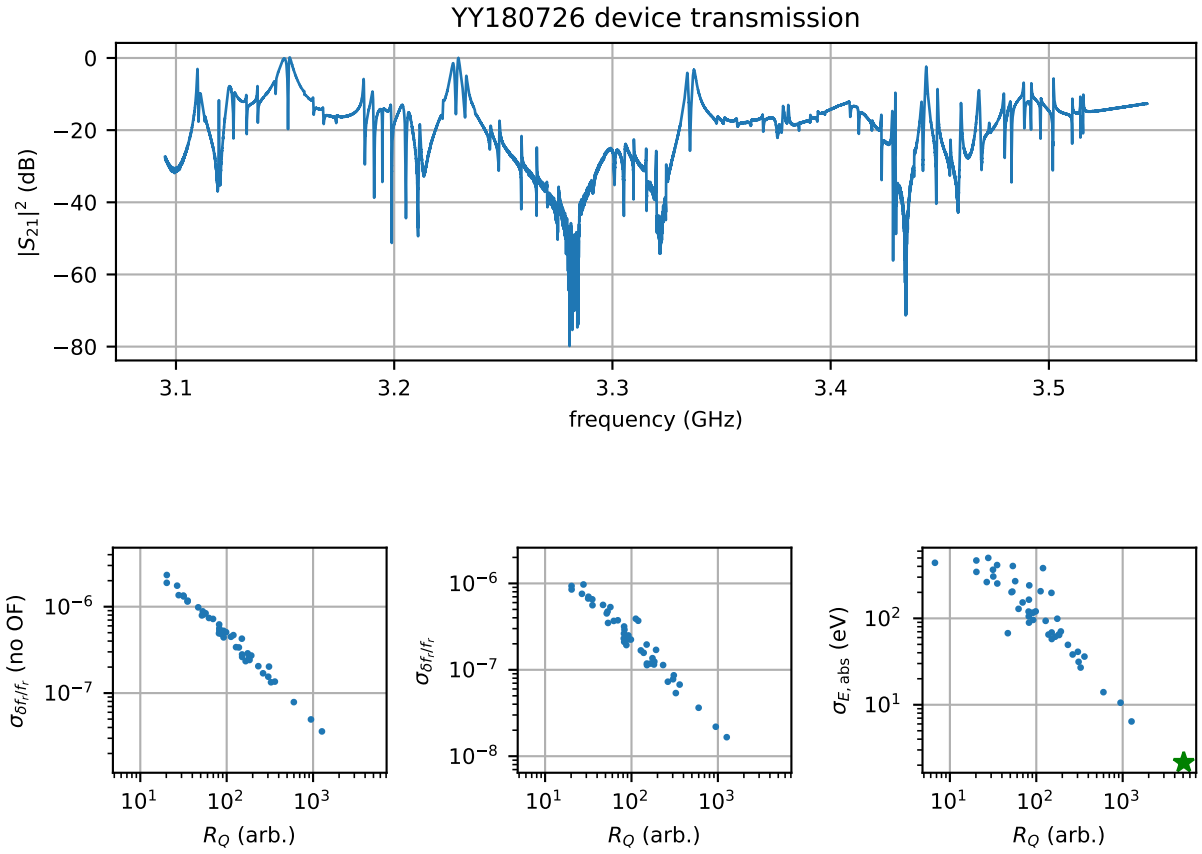

Figure 7.7: **YY180726.2 device performance.** Top: transmission of the device. Bottom: resonator noise performance in various units versus R_Q. The green star corresponds to the expected noise performance given the desired R_Q described in the text. Bottom Left: resonator $\delta f_r/f_r$ noise over the full bandwidth. Bottom Center: resonator $\delta f_r/f_r$ noise after application of the optimal filter. Bottom Right: resonator resolution on absorbed energy.

Bottom. Figure 7.7 Bottom Center displays the R_Q dependence of the optimally filtered $\delta f_r/f_r$ resolution, which shows greater variation because of pulse shape variation among resonators. The optimally filtered resolution is computed in the following way:

$$\sigma_0 = \sqrt{\frac{1}{\sum_{f_n} \frac{T|\tilde{s}(f_n))|^2}{J(f_n)}}} \tag{7.51}$$

Finally, Figure 7.7 Bottom Right displays the R_Q dependence of $\sigma_{E_{\mathrm{abs}}}$, which has additional scatter due α and Δ variations.

In Figure 7.7 Bottom Right, a green star is shown to indicate the expected noise performance for a resonator with R_Q given by the best transmission observed in the data and $Q_r \approx Q_c = 10^5$.[8] It is located around 2 eV, about a factor of 10 worse than the expectation calculated in Section 7.4. This

[8] Setting $Q_c = 10^5$ and overcoupling the resonator are desirable for this particular detector architecture because phonon arrival time information is maintained if the resonator ring up time is not too long. The ring-up time of the resonator is given by $Q_r/2\pi f_r$.

discrepancy may be due to uncertainties in the operating readout power at the device, worse than expected HEMT noise, or, perhaps most likely, unaccounted for attenuation in the readout chain between the device and the amplifier (recall that attenuation always degrades signal-to-noise). The RF performance of the device was very poor, and its overall transmission is not well characterized.

7.8 TLS-noise-dominated KIPM detectors

In this section, we report the contributions of amplifier noise, electronics $1/f$, and TLS noise in the readout of a lumped element resonator design that was originally designed and engineered by Chang (2023). The original design for 3 GHz resonators was later modified for 4 GHz and 5 GHz resonant frequencies, with the desire to better match the gain bands of kinetic inductance traveling wave parametric amplifiers. Many different materials were also explored. The list of devices studied here is as follows:

- Niobium resonator from OW200127,
- TiNx-Al bilayer resonator from OW221031p3,
- Al resonator from YY180726p2,
- Al resonator with Nb-capped capacitor from OW200127.

After an introduction to the various bases used for PSD analysis, we present the data for each resonator in each basis and study how the noise depends on readout power.

PSD units

The PSDs in this section, shown in Figures 7.8 to 7.11, are displayed in three different bases: the electronics basis, the fractional electronics basis, and the resonator basis. PSDs in each basis carry different units. Each basis has two directions, and PSDs are computed for both directions. We explain the difference between the different bases first.

Electronics basis

The raw data from the USRP in ADC units is shown in the left column of Figure 7.8 to 7.11. The gross effect of readout power on the signal-to-noise ratio can be seen in Figure 7.8 to 7.11 Top Left, where the data are shown in ADC units. Higher readout powers lead to larger voltages at the ADC, which is illustrated by the fact that the VNA circles get larger with readout power. The readout powers are separated by factors of 5 dB, which corresponds to a factor of $\sqrt{\sqrt{10}} \sim 1.78$ difference in circle size between adjacent powers.

In the electronics basis, PSDs are computed in the radius and arc length directions of the complex plane; these directions are illustrated with gray dashed lines and are collinear with the gain and phase directions, which are more often quoted in the literature. The $S_{21}(t; f_r)$ timestreams in the top left panel of each figure do not change much in size with different readout powers. The noise of these timestreams is generally dominated by amplifier noise, which is additive noise and stays fixed as the carrier tone amplitude is changed. An elongation of the $S_{21}(t; f_r)$ noise blob is visible along the tangent direction to the $S_{21}(f)$ circle; this excess noise is TLS noise, as we discuss in more detail below.

Fractional electronics basis

In the middle columns of Figure 7.8 to 7.11, all data have been divided by the resonator fit parameter a for each readout power, converting $V \to S_{21}$. In the complex S_{21} space, all VNA circles lie on top of each other now, modulo small changes to f_r and Q_r with readout power. The

noise blobs decrease in size with increasing readout power, reflecting the improved signal-to-noise ratio with higher readout powers.

This basis, where the off-resonance transmission has been divided out, is referred to as the "fractional electronics basis," to reflect the fact that the noise PSDs are computed in fractional units with respect to the amplitude of the incoming wave. In other words, $S_{21} = V_2/V_1$, its usual definition, where V_2 corresponds to an output port just after the resonator and V_1 to an input port just before the resonator.

There are two observations about the noise PSDs in the fractional electronics basis:

- The white noise level in the PSDs decreases as P_g^{-1} (recall that the readout powers for the PSDs are separated by 5 dB). This dependence is most evident in the arc length direction PSDs at the lowest readout powers, which has the greatest contribution from amplifier white noise: the blue, orange, and green curves are each separated by about a half decade.
- The uncleaned noise shows a prominent $1/f$ noise at the highest readout powers, which is roughly fixed at $10^{-8}\,(\delta S_{21})^2/\mathrm{Hz}$ at 100 Hz in the arc length direction. This constancy reflects the fact that the electronics $1/f$ noise is believed to be a multiplicative gain and phase noise and therefore is fixed in this fractional electronics basis.

Resonator basis

In the right columns of Figures 7.8 to 7.11, the S_{21} data are projected onto the real axis using the resonant fit parameter ϕ; the resonance circle and timestream data are multiplied by $\cos\phi$. This transformation is the same as the transformation performed in Figure 7.1 Right; the resonator basis and the "idealized S_{21}" basis are the same.

In this idealized S_{21} space, the deviation of S_{21} from its quiescent value is described, for small changes (linear approximation), by

$$\delta S_{21} = \frac{Q_r^2}{Q_c}\left(\delta\frac{1}{Q_i} - 2i\frac{\delta f_r}{f_r}\right), \tag{7.52}$$

This expression allows us to extract $\delta f_r/f_r$ and $\delta(1/Q_i)$ timestreams from δS_{21}.

Calibration to the resonator basis

Depending on the readout system of choice, there may be systematic differences between the VNA scan that measures $S(f)$ and the noise timestream that measures $S_{21}(t;\, f_r)$. The USRP has exhibited misalignment and offsets between noise timestreams and VNA scans, so a scheme was devised to correct for such deviations.

The crux of the scheme is to, just before and after data acquisition, and sometimes during, measure S_{21} at two tones slightly off resonance to determine the frequency direction. Typical frequency values used for these off-resonance tones are $f_r \pm f_r/5Q_r$. Currently, the alignment of the noise data with the VNA is done by fitting a line in the complex plane that passes through $[V(t;\, f_r - f_r/5Q_r),\, V(t;\, f_r),\, V(t;\, f_r + f_r/5Q_r)]$ and rotating these points so that they are parallel to the line that passes through the corresponding points from the VNA: $[V(f_r - f_r/5Q_r),\, V(f_r),\, V(f_r + f_r/5Q_r)]$.

For better confidence in the efficacy of the calibration, it would be good to have more calibration tones, as shown in Ramanathan, Wen, et al. (2024). At some point, with noise timestreams at enough frequency points, the resonator basis calibration begins to resemble a VNA scan. With more just-off-resonance noise timestreams, the on-resonance timestream can be more confidently converted into the resonator basis.

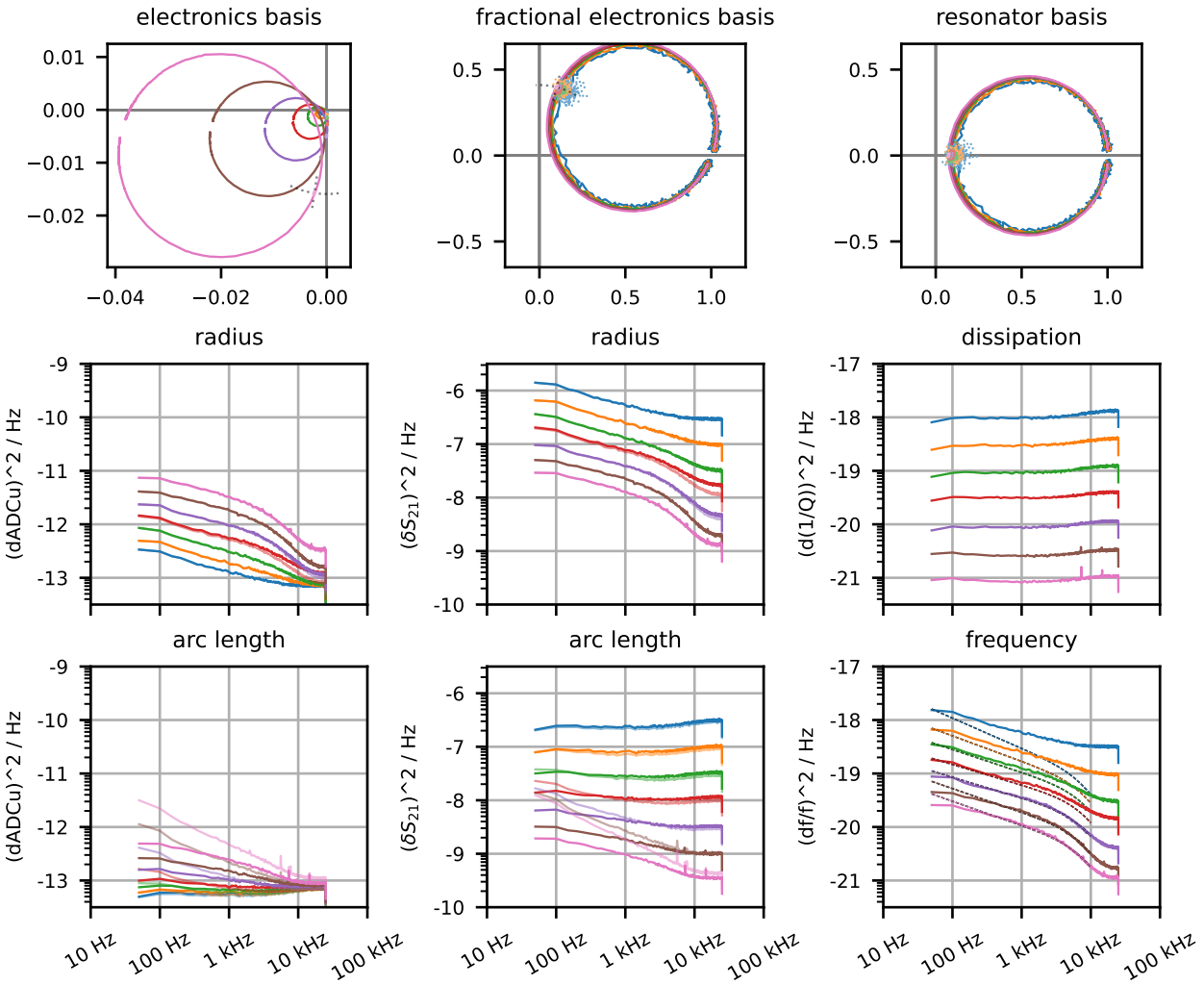

Figure 7.8: **OW200127 Nb resonator noise PSDs.** The resonant frequency is 4.206 GHz. Top: $S_{21}(f)$ are depicted as lines, and $S_{21}(t; f_r)$ are depicted as points. The different colors correspond to different readout powers, separated by 5 dBm. The "radius" and "arc length" directions are drawn with the dotted gray lines. Middle and Bottom: $\log_{10}$ of power spectral densities of the various timestreams in different units. The faded curves are for uncleaned timestreams. The dashed lines in the bottom right are the TLS component of a TLS + white noise fit.

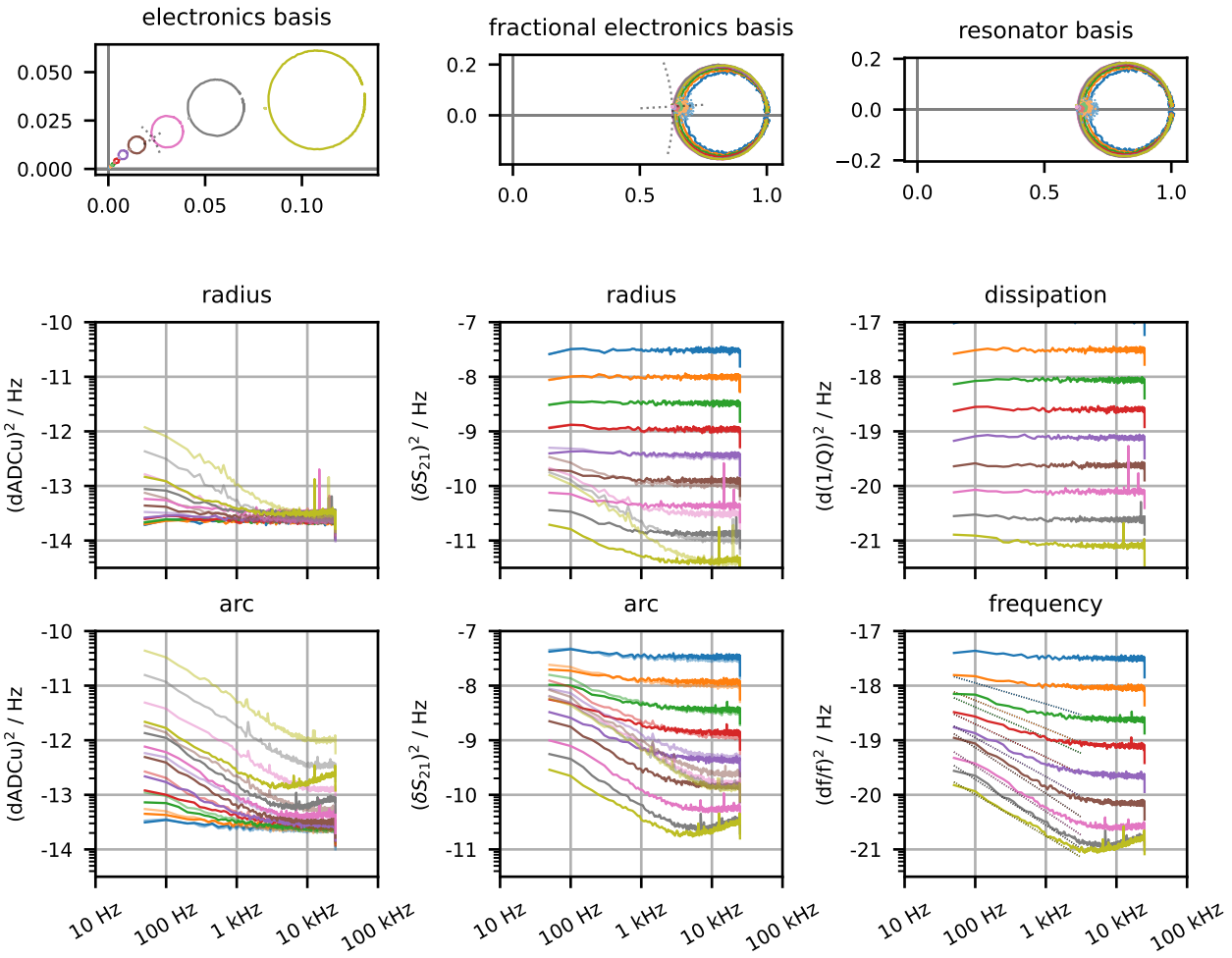

Figure 7.9: **OW221031 TiNx resonator noise PSDs.** The resonant frequency is 4.345 GHz. Caption is the same as the caption in Figure 7.8.

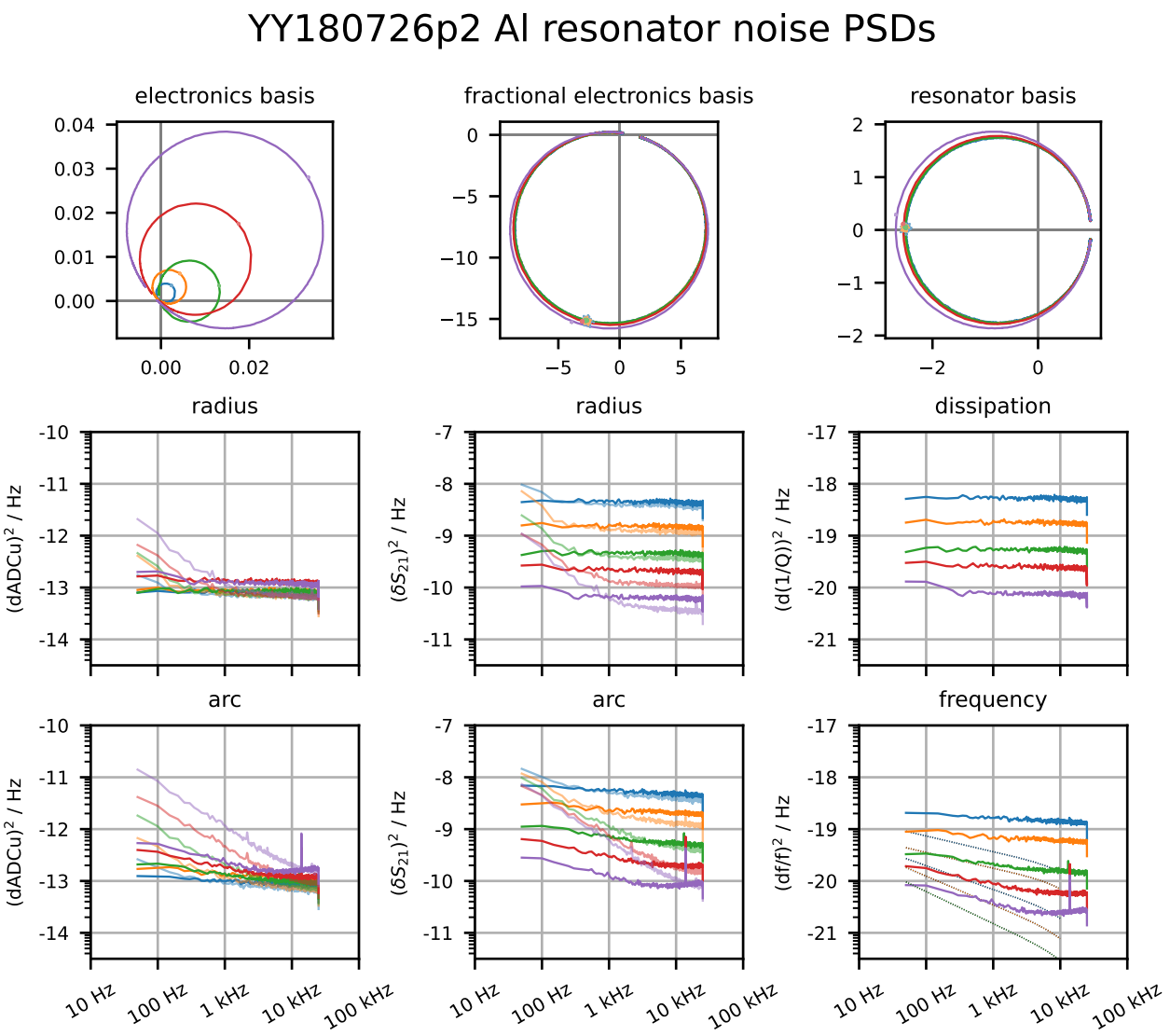

Figure 7.10: **YY180726.2 Al resonator noise PSDs.** The resonant frequency is 3.423 GHz. Caption is the same as the caption in Figure 7.8.

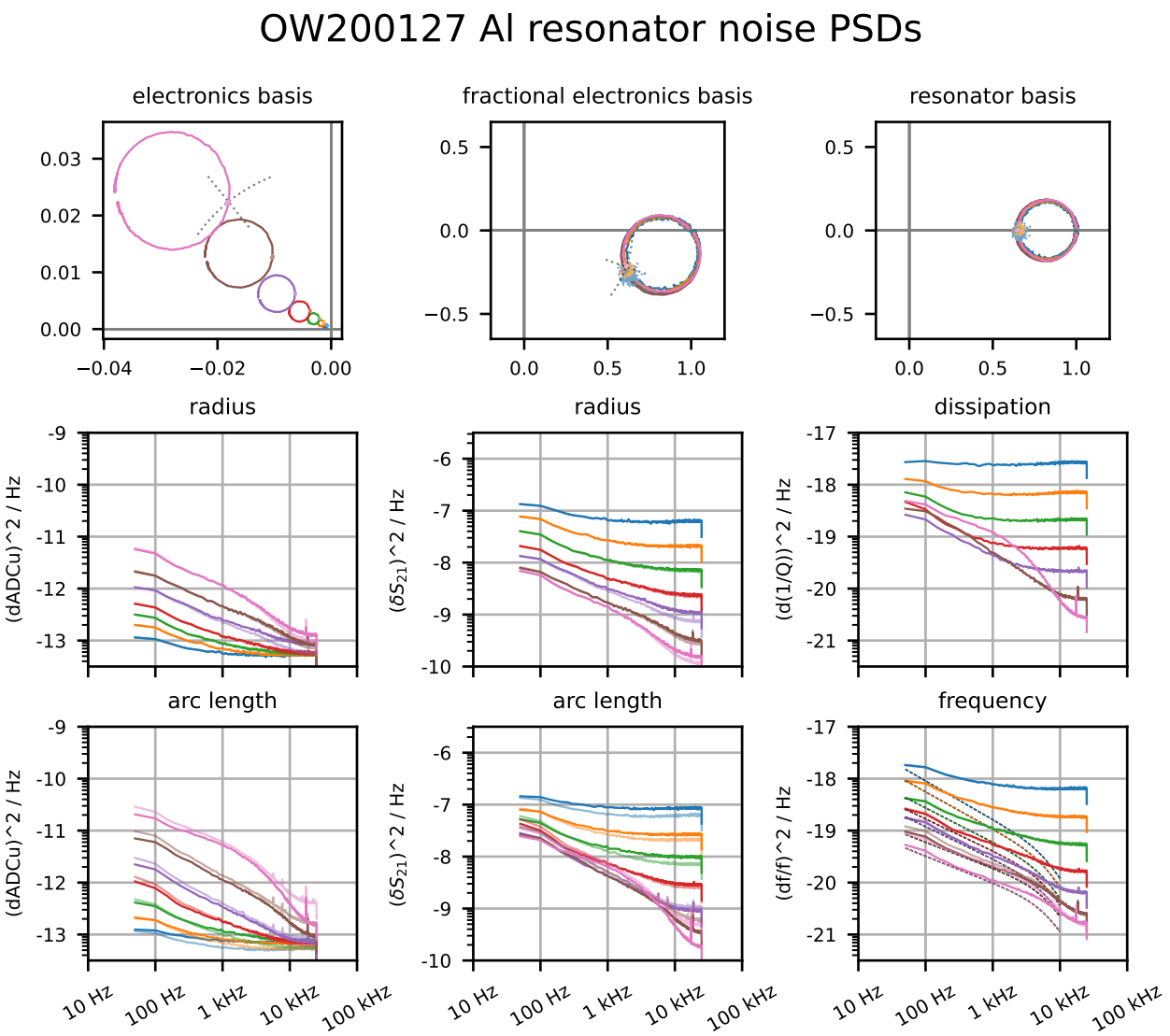

Figure 7.11: **OW200127 Al resonator with niobium-capped capacitor noise PSDs.** The resonant frequency is 4.240 GHz. Caption is the same as the caption in Figure 7.8, with the exception of an additional κ_1–κ_2 correlated noise cleaning in the third column; the post-cleaning PSD is shown as the solid color curves. In the third column only, the faded colors are for PSDs that are not cleaned of the κ_1–κ_2 noise (but are cleaned of the electronics noise).

Example 1: niobium

Figure 7.8 shows the noise performance in different readout units of a niobium resonator with a 4 GHz lumped element design as a function of readout power. The excess noise in the frequency direction in comparison to the dissipation direction provides clear evidence for TLS noise in this niobium resonator. The frequency direction PSDs are fit with a model consisting of TLS and white components. The dashed curves in the frequency direction PSDs are the TLS component of the fit. For the TLS component, the frequency roll-off is fixed by the resonator bandwidth, and the amplitude and power law of the TLS component are floated. As expected with TLS noise, the TLS noise level decreases as $P_g^{-1/2}$. Recall that the power levels are separated by 5 dB, and observe that the TLS components of the fit are roughly separated by a quarter decade along the y-axis.

Example 2: TiNx-aluminum bilayer

Figure 7.9 shows a similar set of data for a TiNx-aluminum bilayer resonator from OW221031, which was fabricated with the intention of reducing TLS noise in the capacitor. TiNx was believed to exhibit less TLS noise than niobium (McRae et al., 2020) while also serving as a higher gap material (target T_c around 4 K) for the dead metal region of the capacitor. The "x" in the name refers to the stoichiometric nature of TiNx. The 30 nm aluminum resonator with a 100 nm TiNx layer capping the capacitor was never observed, but the resonators with the 100 nm TiNx layer capping the entire aluminum resonator (that is, a TiNx-aluminum bilayer) were used to measure the TLS noise of the proximitized TiNx resonators. The TLS noise slopes were better described by power laws closer to -1 than to -0.5, the latter being observed for the niobium resonator.

Example 3: aluminum

Resonators composed only of aluminum were studied in YY180726, an 80-resonator, 3 in diameter device with a 300 nm niobium feedline and 30 nm aluminum resonators. The aluminum resonator reported in Figure 7.10 is one of the resonators whose transmission is shown in Figure 7.7. This resonator had the highest RF responsivity R_Q and thus the lowest contribution from amplifier white noise. One oddity about this particular resonator is its shape in the IQ-plane: the resonance circles encircle the origin, suggesting a Q_c that is less than Q_r and by inference a Q_i that is negative.

TLS noise was marginally visible above the amplifier noise. The noise undergoes the same fitting procedure for the previous resonators, and the TLS component of the noise shows a rough agreement with the $P_g^{-1/2}$ scaling expected for the TLS noise. Of the 70 resonators observed on YY180726, we focused on this one because it had the largest R_Q and thus the TLS noise was most easily visible above the amplifier noise.

Example 4: aluminum resonator with a niobium-capped capacitor

OW200127 had ten 30 nm niobium resonators and a single 30 nm aluminum resonator that also had a 30 nm niobium capping layer on the capacitor. The noise of this resonator was studied and reported in Wen et al. (2022) and plotted in Figure 8.4.

The dissipation direction exhibited enormous noise in excess of the white noise level, the source of which is currently unknown. The noise was observed to be correlated between the frequency and dissipation directions. The noise was referred to as κ_1–κ_2 correlated noise. To remove this noise from the frequency direction, a similar cleaning procedure as was done for the electronics noise was followed:

$$\delta n_{qp,f,\text{cleaned}}(t) = \delta n_{qp,f}(t) - A_{Q,f}\delta n_{qp,Q}(t), \quad \text{where} \quad A_{Q,f} = \frac{\text{Cov}(\delta n_{qp,Q}, \delta n_{qp,f})}{\text{Var}(\delta n_{qp,f})}. \tag{7.53}$$

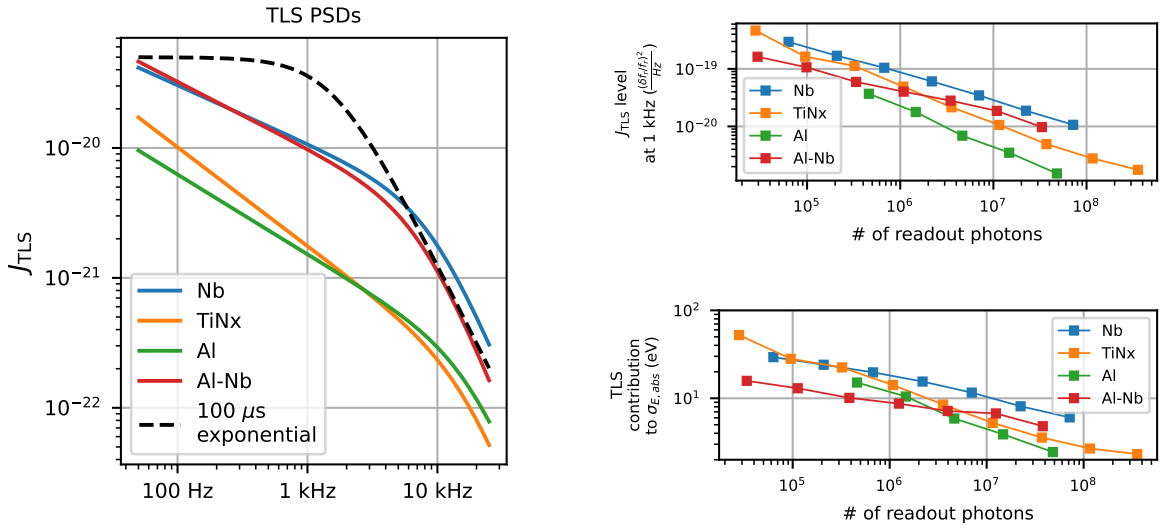

Figure 7.12: **TLS comparison across different materials.** Left: fit-for TLS noise PSD in the four different material compositions that were measured at the highest readout power for each material's dataset. It is important to note that relative noise levels should not be compared with this plot because the stored energy in the resonators are not equal. For comparison, an arbitrarily normalized Fourier coefficient signal template is shown to illustrate the relevant signal bandwidth for these detectors. Top Right: $J_{\mathrm{TLS}}(1\,\mathrm{kHz})$ versus number of readout photons, which follows a power law with the stored energy in the resonator. Bottom Right: the calculated impact on the intrinsic energy resolution $\sigma_{E_{\mathrm{abs}}}$.

In the frequency direction, the PSDs that include the κ_1–κ_2 correlated noise are shown in the faded colors, and the solid colors show the result of the κ_1–κ_2 correlated noise cleaning.

After the κ_1–κ_2 correlated noise is cleaned away in the frequency direction, TLS noise remains present and is fit for in combination with white noise, as was done for the other three resonators.

Comparing PSDs for different materials

In order to compare the TLS levels of the different materials under study, an approximation to the number of readout photons at each readout power was computed for each resonator. The conversion from readout power to number of photons was done by first calculating the stored energy in each resonator:

$$E_{\mathrm{stored}} = \frac{\chi_c}{2}\frac{1}{1+4Q_r^2x'^2}\frac{Q_iP_g}{2\pi f_r}, \tag{7.54}$$

where $\chi_c = 4Q_cQ_i/(Q_c+Q_i)^2$ is a coupling efficiency, x' is the detuning of a resonator due to elevated readout power, and P_g is the readout power on the feedline at the resonator. We operate in a low readout power regime where x' is 0. It is interesting to note that $\chi_cQ_i/2 = 2Q_r^2/Q_c$; the latter is the same expression used in conversion from δS_{21} to $\delta f_r/f_r$ and $\delta(1/Q_i)$ in the linear approximation of δS_{21}. Finally, E_{stored} is converted into photon number by $N = E_{\mathrm{stored}}/h\nu$. It should be remarked that there are large systematic uncertainties on the stored energy of the resonator due to uncertainties regarding the measurement of Q_i, especially when the resonator is greatly rotated by some angle ϕ. There are further complications when the baseline transmission itself is not precisely known.

The extracted TLS noise level at 1 kHz as a function of photon number is shown in Figure 7.12 Top Right. Compared with the data in Figure 14 of Zmuidzinas (2012), the aluminum and TiNx TLS noise levels are comparable to the lowest levels seen in the literature, $J_{\mathrm{TLS}}(1\,\mathrm{kHz}) \approx 10^{-21}\,\mathrm{Hz}^{-1}$.

The impact of TLS noise on intrinsic energy resolution $\sigma_{E_{\mathrm{abs}}}$ is then shown in Figure 7.12 Bottom Right. The noise PSDs used in the calculation of the optimally filtered resolution (Equation 7.51) are shown in Figure 7.12 left, where it is apparent that, for a typical signal template, the signal bandwidth is smaller than the TLS noise bandwidth, hence the assumption made in Section 7.3 that $1/\tau_{qp} \ll f_{\text{roll-off}}$. Figure 7.12 Bottom Right shows that TLS noise limits the intrinsic energy resolution to about 1 eV to 10 eV, depending on the material and the operating readout power. The calculation from Section 7.3 computed a TLS-limited intrinsic energy resolution for this resonator design of around 500 meV. The discrepancy between the calculation and the data is probably due to the 2.2 ms lifetime that was assumed in Equation 7.32 whereas the signal template assumed here has a fall time of 100 μs.

7.9 KIPM detector operation with a KI-TWPA

In this section, we briefly discuss the noise improvement results of Ramanathan, Wen, et al. (2024) given its relevance to the discussion of the KIPM detector noise sources in this chapter. We show how a kinetic inductance traveling-wave parametric amplifier (KI-TWPA) enabled resonator-noise domination instead of HEMT-noise domination.

KI-TWPAs are a burgeoning technology in the field of quantum-limited amplifiers that leverage non-linearity in the kinetic inductance of superconductors to provide quantum-limited cryogenic amplification for RF circuits. As shown in Section 7.4, cryogenic amplifiers are a common and significant source of noise for KID readout. Widely accessible HEMT amplifiers are limited to about 2 K to 4 K of noise. KI-TWPAs, by virtue of their superconductivity, can in theory reach the standard quantum limit (SQL) of added noise for linear amplifiers, which is $\frac{1}{2}h\nu$.[9]

KI-TWPAs are dispersion-engineered transmission lines that transfer pump power into signal power. When the pump is supplied to the transmission line, the non-linearity in the kinetic inductance enables the transfer of power. A notable feature of the KI-TWPA is that, since it is just a feedline, the readout circuit transmits the signal with unity gain (modulo insertion and reflection losses) even if the KI-TWPA is not supplied with a pump tone. This feature greatly facilitates gain and noise measurements of the amplifier and is in general not a property of other amplifiers.

Ramanathan, Wen, et al. (2024) reports on a 40-resonator, 75 mm diameter prototype KIPM detector that was operated with a KI-TWPA as the first cryogenic amplifier. The experiment was performed in a dilution refrigerator located in the Microdevices Laboratory of the Jet Propulsion Laboratory. Some of the 40 resonators were made of aluminum and others of niobium; only six aluminum resonators were investigated. Like YY180726, which was discussed in Section 7.7, this 75 mm diameter device exhibited box modes that were strongly coupled to the feedline and degraded both transmission and resonator performance. Further details on the cryostat, the KI-TWPA circuitry, and the absolute noise calibration can be found in Ramanathan, Wen, et al. (2024).

The key result KIPM detector readout with a KI-TWPA was a 5× improvement in the baseline intrinsic energy resolution over the HEMT-dominated noise. This result is shown in Figure 7.13. The figure is reminiscent of Figure 7.7 Bottom in that R_Q dominates the noise performance across the six different resonators, indicated with different markers. The y-axis is the optimally filtered noise performance transformed into substrate energy resolution σ_E for the case of a hypothetical single-resonator device with an assumed phonon collection efficiency of 30%.

[9] There is an additional $\frac{1}{2}h\nu$ of vacuum zero-point motion presented at the input to the amplifier even if there were no white noise from the input side chain or the KID. The SQL input noise for an amplifier is thus $h\nu$.

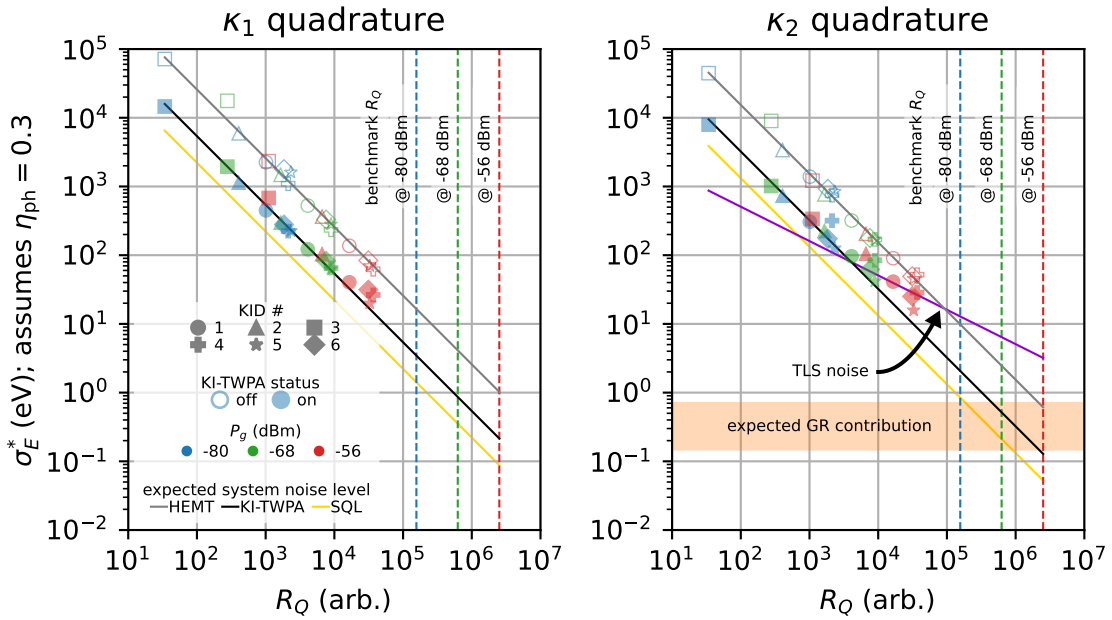

Figure 7.13: **Improved KIPM detector noise performance with first-stage amplification from a KI-TWPA.** Left and Right: the inferred κ_1 and κ_2 energy resolutions of a hypothetical single-resonator device for six different resonators as a function of R_Q, readout power, and KI-TWPA status. Dashed lines indicate benchmark R_Q values given design goals for Q_r and optimal transmission performance. Solid lines indicate the average noise level for different amplifier scenarios. The SQL noise level is calibrated for using a Y-factor measurement. In the κ_2 quadrature, an additional trend is drawn suggestively in purple to show the impact of TLS on the KI-TWPA-improved readout noise. The expected contribution from GR noise is also shown.

In contrast with Figure 7.7, Figure 7.13 shows the noise performance of each resonator for three different readout powers; higher readout power leads to higher values of a and therefore higher values of R_Q and lower values of σ_E. Benchmark R_Q values are shown for the different readout powers assuming optimal transmission of the detector and $Q_r \approx Q_c = 10^5$.

The 5× improvement in noise is visible in the difference between the empty and filled points, as well as in the difference between the gray and black lines, which, respectively, are fits to the differently styled points. R_Q has been referred to the input of the KI-TWPA so that a is independent of whether the KI-TWPA is on or off, i.e., supplied with a pump or not. The noise has clearly improved, but is still about a factor of 3 away from the standard quantum limit, which was calibrated using an *in situ* Y-factor measurement.

The improvement in amplifier noise with increasing R_Q does not apply to the inferred final detector energy resolution in the κ_2 direction at the highest R_Q values because the resonators become TLS-noise limited. The TLS noise demonstrates good agreement with a $P_g^{-1/2}$ trend line that is suggestively drawn in Figure 7.13 Right. So, as the contribution of the amplifier noise is reduced by improved R_Q, the next limiting factor to reach the desired sub-eV energy resolutions becomes the noise in the resonator, which in this case is TLS noise. The TLS noise for this prototype device is quite elevated, probably due to suboptimal fabrication conditions. Beyond TLS noise, the contribution of GR noise is also shown in Figure 7.13 Right and corresponds to a quasiparticle population of 25 μm^{-3} to 475 μm^{-3}.

7.10 Summary of noise sources

	GR noise	TLS noise	amplifier noise	electronics noise
source	n_{qp} fluctuations	two-level systems in surface oxides	Johnson-Nyquist thermal noise	transmission & gain drift, USRP phase noise, etc.
relevant basis	quasiparticle basis, signal direction	only present in the $\delta f_r/f_r$ direction	fixed in electronics basis	fixed in fractional electronics basis
spectral shape	flat, with roll-off at $1/2\pi\tau_{\mathrm{qp}}$	f^{-n} for $n < 1$, with roll-off at $f_r/2Q_r$	flat	$1/f$ at low frequencies (other slopes possible)
P_g dep.; elec. basis	P_g^*	$P_g^{1/2}$	fixed	P_g
P_g dep.; res. basis	fixed*	$P_g^{-1/2}$	P_g^{-1}	fixed
limit on $\sigma_{E_{\mathrm{abs}}}$ with current design	220 meV – 390 meV	510 meV	280 meV	–
how to improve?	reduce quiescent n_{qp} with IR shielding, both free-space and in-line	• modify capacitor geometry • use metals with lower propensity for surface oxide formation	quantum-limited amplifiers	correlated-noise removal

Table 7.1: **Summary of noise sources in KIPM detectors.** The table includes descriptions of the the noise PSDs, the dependence on power, current limitations, and ways to improve the noise. The *s in the GR noise entries for P_g dependence indicate the assumption that n_{qp} is not P_g-dominated.

Chapter 8

Optimizing KIPM detectors for low-threshold particle detection

8.1 Previous measurement of energy resolution

Moore et al. (2012) and Moore (2012) report on the energy resolution of a KIPM detector known as DMLE2 ("dark matter lumped element"). DMLE2 was a 2.2 cm×2 cm×1 mm silicon chip with twenty 25 nm thick aluminum resonators instrumented on one surface of the detector. For measurement of energy resolution, DMLE2 was illuminated with an iodine-129 source, which provides X-ray lines near 30 keV. Moore et al. (2012) and Moore (2012) also report roughly 1 mm position resolution using position reconstruction based on relative pulse heights.

The baseline energy resolution σ_E of DMLE2 was measured to be 380 eV, and the phonon collection efficiency $\eta_{\rm ph}$ was measured to be 7.0%±1.1%. These quantities can be used to infer the intrinsic energy resolution $\sigma_{E_{\rm abs}}$, also known as the resolution on absorbed energy in the sensor:

$$\sigma_E = \frac{\sqrt{N_r}}{\eta_{\rm ph}}\sigma_{E_{\rm abs}}. \tag{8.1}$$

The above equation assumes that $\sigma_{E_{\rm abs}}$ is the same across the N_r resonators and that the total phonon collection efficiency summed over all the resonators is $\eta_{\rm ph}$. In the above references, $\eta_{\rm ph}$ is computed by measuring the peak number of quasiparticles observed in each resonator and summing over the resonators. We note that Equation 8.1 applies to any $\sigma_{E_{\rm abs}}$, regardless of whether it is amplifier-noise-dominated, TLS-noise-dominated, or GR-noise-dominated. Using Equation 8.1, $\sigma_{E_{\rm abs}}$ for a single resonator in DMLE2 is around 5.9 eV.

When we confine our study of DMLE2 to a single resonator, we find that the phonon collection efficiency for particular events is resonator dependent: resonators that are closer to the event detect more energy; see Figure 8.1 Left. As we explain later, the relevant quantity is the phonon collection efficiency of the single resonator that is closest to the event. From the relative pulse heights in Figure 8.1 Left and the reported quantity of 7.0%±1.1% for all resonators, we infer a phonon collection efficiency of 1.1%±0.25% for the closest-resonator phonon collection efficiency.

Thin-film KIPM detectors

Three different thicknesses were explored for the aluminum film in DMLE2-style devices: 15 nm, 25 nm and 75 nm. The 15 and 25 nm devices performed better in baseline energy resolution than the 75 nm device by about a factor of 3. In Figure 8.1 Right, a reproduction of Figure 7.23 in Moore (2012), the collection efficiency is shown to improve with film thickness, but other factors

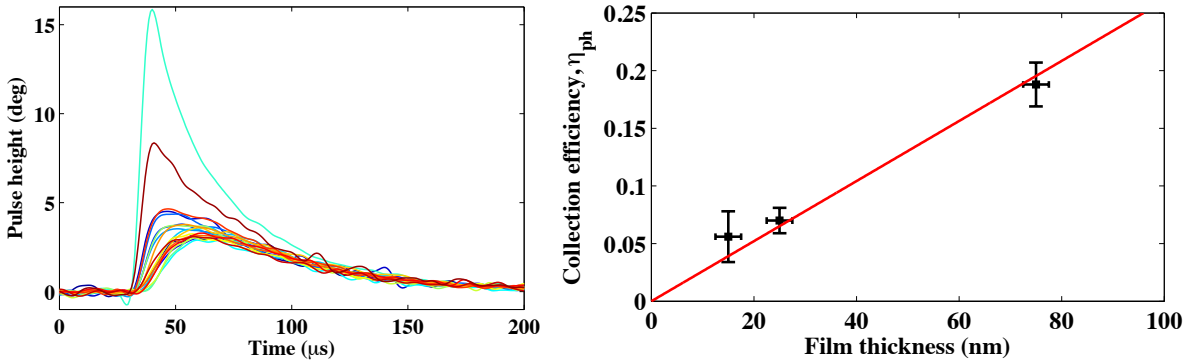

Figure 8.1: **Phonon collection efficiency of DMLE2 resonators.** Left: relative pulse heights of the twenty different resonators in response to a 200 keV event in the substrate. Reproduced from Figure 7.6 in Moore (2012). Right: phonon collection efficiency versus film thickness for DMLE2-style devices. Reproduced from Figure 7.23 in Moore (2012).

caused the detector energy resolution to degrade overall. The discussion of why the overall detector energy resolution degrades with thickness is reproduced from Section 7.5.4 of Moore (2012):

> As shown in Fig. 7.23, there is an approximately linear increase in the collection efficiency with film thickness, as expected in the limit that the film thickness $t \ll \lambda_{\mathrm{pb}}$, where $\lambda_{\mathrm{pb}} \approx 1\,\mu$m is the characteristic pair-breaking length in Al. As the films are made thinner, the volume is also reduced linearly, so the quasiparticle density remains approximately unchanged. In addition, α increases with decreasing thickness, giving higher responsivity for the same quasiparticle density and improving the signal-to-noise of the device. However, thinner films also saturate at lower powers, decreasing the maximum readout power that can be used and increasing the amplifier noise contribution. The quasiparticle lifetime was also found to decrease from $\tau_{\mathrm{qp}} = 73.7\,\mu$s for the 75 nm thick device to $\tau_{\mathrm{qp}} = 15.4\,\mu$s for the 15 nm thick device. Taken together, these tradeoffs led to a factor of 3 improvement in the baseline resolution of the 25 nm device relative to the 75 nm device, but no further improvement in the resolution when decreasing the film thickness to 15 nm. From these tests, the Al film thickness that maximizes signal-to-noise was found to be 15–25 nm.

The key takeaway is that thinner films optimized the confluence of detector parameters for energy resolution. The subsequent detectors described in this chapter use 30 nm aluminum films. For comparison, the aluminum films used for phonon collection by QET phonon sensors are usually 600 nm thick and thus comprise a much larger collection volume. Given the smaller collection volume of the lumped-element resonator design, which combines the roles of the sensitive element and the phonon absorber, we expect this design to be more sensitive to other loss mechanisms in the substrate. QET phonon sensors have demonstrated collection efficiencies around 30% (see Ren et al., 2021) or greater (see Romani et al., 2024 or TESSERACT Collaboration, 2024). In principle, collection efficiencies at this level should be achievable as long as the aluminum film remains the dominant sink for greater-than-$2\Delta_{\mathrm{Al}}$ phonons in the substrate. One of the main thrusts of research presented in this chapter is an exploration of the dependence of phonon collection efficiency on metal volumes and other loss mechanisms. This analysis provides guidance on how to modify the KIPM detector design so that the collection efficiency may be improved to the tens of percent level that is expected for a phonon-mediated detector.

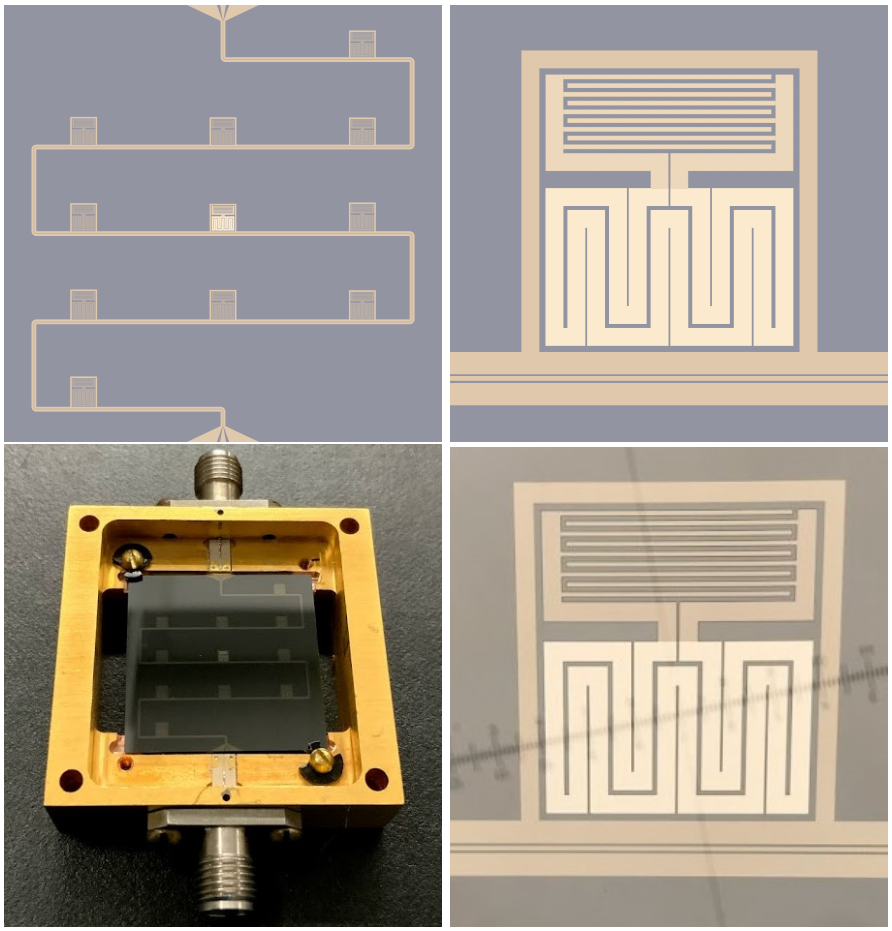

Figure 8.2: **Designs and photographs of OW200127.** Top Left: the design of the 2.2 cm × 2.2 cm OW200127 chip with 11 total resonators: 10 Nb resonators and one Al resonator in the middle. Each resonator is coupled to a 300 nm thick Nb co-planar waveguide (CPW) feedline (dark beige) that meanders over the chip and terminates in wirebond pads at the top and bottom. The signal line is visible and separated from the the ground planes by 10.75 μm. Top Right: the central resonator is fashioned out of 30 nm thick Al, shown as a light beige, with 30 nm of Nb on the capacitor (the same Nb that is used for the other 11 resonators), shown as an intermediate beige. Bottom Left: Photograph of an OW200127 chip mounted in its device housing. Bottom Right: Micrograph of the central resonator in OW200127.

8.2 Design considerations for a low-threshold detector

We initiated work on a new KIPM design to improve σ_E of DMLE2. The new device design is referred to as OW200127 (initials of the fabricator and date of fabrication, YYMMDD). With DMLE2 as a reference point, there are three ways to improve σ_E.

First, the phonon collection efficiency of DMLE2 almost certainly had room for improvement. Given the much higher phonon collection efficiencies ($\gtrsim 30\%$) for other QET-based phonon-mediated detectors mentioned previously, it is likely that there were loss mechanisms in DMLE2 that limited its phonon collection efficiency. It was hypothesized that the non-signal-contributing aluminum components, such as the resonator capacitors, the feedline, and other auxiliary components, were the main source of phonon loss.

Given this hypothesis, the first major design choice for OW200127 was that niobium, a

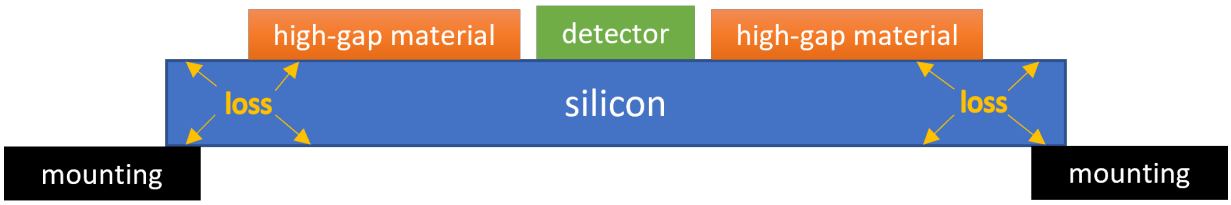

Figure 8.3: **Schematic depicting possible avenues for phonon energy loss.**

higher gap material, would be used for the "dead metal" on the device, i.e., any metal that does not contribute to the signal. In Figure 8.2 Top Left and Right, the dead metal is shown in dark beige, and the inductor is shown in lighter beige. The dead metal in OW200127 includes the feedline, other resonators that are used for phonon creation via readout power (see Aralis, 2024), and the capacitor of the phonon-sensitive resonator. For the capacitor, we used the proximity effect to lift its gap; a Nb film is placed on top of an Al film. In Figure 8.2 Top Right, the proximitized aluminum-nioibium bilayer is shown as an intermediate shade of beige. Overall, by lifting the superconducting bandgap of the dead metal, we prevent phonons between $2\Delta_{\mathrm{Al}}$ and $2\Delta_{\mathrm{Nb}}$ (or whatever the proximitized gap is) from being absorbed and down-converting in these non-signal-contributing regions.

The second way to improve energy resolution is to confine all the phonon energy to just one resonator, i.e., set N_r to 1. As shown in Equation 8.1, this choice is motivated by the fact that the readout noise of such KID-based detector scales with $\sqrt{N_r}$ because the noise from each readout channel must be added in quadrature upon the summation of collected energies. In some circumstances, decreasing N_r tends to decrease η_{ph} and thus degrade σ_E. The naive intuition is that η_{ph} scales with the surface area coverage of "live metal," which is the metal where absorbed phonons contribute to the signal, i.e., our inductors. However, if the live metal is the dominant down-converting loss mechanism of the phonon energy in the substrate, then fewer resonators do not necessarily imply a lower η_{ph}. This idea that η_{ph} may not degrade as N_r decreases motivates the design choice $N_r = 1$.

The third way to improve σ_E would involve improving $\sigma_{E_{\mathrm{abs}}}$. This process involves a careful understanding of the noise sources involved in the readout of the resonator; these noise sources and some strategies for reducing their contributions are discussed in detail in Chapter 7.

Given the results that are shared in this chapter, we comment in advance on two ways in which these design choices might not have the desired effect.

- Other down-converting and lossy surfaces, such as mounting interfaces or diced, unpolished edges, can dominate loss of phonons; see Figure 8.3. Simply making the dead metal of the chip inert to phonons may not be enough if there are other significant loss mechanisms. In fact, Figure 8.1 Right provides a clue to this effect. The figure shows that the collection efficiency of the DMLE2 design improved linearly with film thickness. This result demonstrates that DMLE2 operated in the regime in which there were dominant mechanisms of phonon loss beyond the aluminum film. The contrapositive here is that, if the aluminum film was indeed the dominant sink for phonons in the silicon, then there should be no improvement in phonon collection efficiency with thickness because the volume of dead aluminum and sensitive aluminum both increase linearly with thickness.
- Enough of the initial phonon energy may be above $2\Delta_{\mathrm{Nb}}$ such that it is then absorbed in the Nb, and this absorption may promote down-conversion of that energy to $< 2\Delta_{\mathrm{Al}}$.

A comment on sub-percent surface coverage designs

The dimensions of the inductor are roughly $700\,\mu\text{m} \times 1\,\text{mm}$, corresponding to $0.007\,\text{cm}^2$. The surface area of the $2.2\,\text{cm} \times 2.2\,\text{cm} \times 1\,\text{mm}$ chip is $10.56\,\text{cm}^2$. The surface area coverage is very low, about 0.07%. It is believed that most of the detectable phonon energy can still be absorbed even at such low surface area coverage. We follow the argument presented in Golwala and Figueroa-Feliciano (2022). The article notes that the majority of the athermal phonon energy can be collected if

$$f_{\text{surf}} f_{\text{abs}} N_{\text{surf}} \sim 1, \tag{8.2}$$

where f_{surf} is the fraction of the surface covered with sensors, f_{abs} is the probability of absorption for a single interaction between an athermal phonon and a phonon sensor, and N_{surf} is the number of times an athermal phonon strikes the surface of the crystal before it down-converts to sub-2Δ energy. Heuristically, $f_{\text{surf}} N_{\text{surf}}$ is the average number of interactions between an athermal phonon and a phonon sensor, and $f_{\text{abs}} f_{\text{surf}} N_{\text{surf}}$ is the fraction of such athermal phonons that are absorbed by the phonon sensor before being lost. An essential realization in the early days of CDMS was that if $f_{\text{abs}} N_{\text{surf}}$ is large enough, then $f_{\text{surf}} < 1$ could still allow highly efficient phonon collection. In general, smaller f_{surf} tends to lead to improved intrinsic sensor resolution (by reducing the volume of the phonon sensor itself), so operating near the lower bound of f_{surf} is usually beneficial.

Hochberg, Pyle, et al. (2016) bounds N_{surf} between 10^3 and a few $\times 10^5$ for bare and polished semiconductor crystals. f_{abs} can be approximated as (the thickness of the phonon sensor) / (its pair-breaking length). Taking into account the acoustic mismatch between the sensor and the crystal, $f_{\text{abs}} \approx 0.01$–$1$ is presented as a reasonable range. For $N_{\text{surf}} \approx 10^3$, a conservative lower bound on f_{surf} is thus 0.001-0.1. The above provides a rationale for $f_{\text{surf}} = 0.07\%$ in OW200127. As the results of this chapter show, the aluminum inductor is not the dominant phonon-loss mechanism in the detector, and consequently, the collection efficiency suffered from such a low-coverage design.

8.3 OW200127, a single resonator device: results and limitations

The performance of the OW200127 design has been reported twice: Wen et al. (2022) and Temples et al. (2024). The two devices reported in the two articles are different chips from the same wafer and have the same design, so are referred to as OW200127 Device 1, or just "D1", and OW200127 Device 2, "D2." In Wen et al. (2022), the noise performance of D1 as a function of readout power was reported and shown to be limited by TLS noise. These tests were performed in an Oxford Kelvinox 25 dilution refrigerator at Caltech, as described in Chapter 7. In Temples et al. (2024), an absolute energy calibration, i.e., a calibration on the energy deposited in the substrate, was performed for D2, in addition to a characterization of the noise performance. These tests were performed at the Northwestern Experimental Underground Site in Fermilab, in parallel with HVeV Run 4, which is described in Chapter 5. Details about the NEXUS experimental setup concerning the cryostat and external shielding can be found there.

The readout chain at NEXUS was slightly different than the readout chain at Caltech and, in general, provided better environmental shielding from parasitic loads. Instead of an attenuator chain consisting of $20\,\text{dB}_{@4\,\text{K}}$–$10\,\text{dB}_{@1\,\text{K}}$–$10\,\text{dB}_{@100\,\text{mK}}$ as described in Chapter 7 for the Caltech readout chain, the input chain at NEXUS had $20\,\text{dB}_{@4\,\text{K}}$–$20\,\text{dB}_{@10\,\text{mK}}$; both are adequate for suppressing room temperature blackbody noise below the HEMT noise of about 3 K to 4 K. Furthermore, a 10 GHz low pass filter was deployed at the 1 K stage to prevent pair-breaking high-frequency RF waves from entering the detector. Further shielding in this regard was provided by in-line Eccosorb filters both at the input and output of the device; these filters are intended to attenuate infrared pair-breaking radiation and are described in Spahn (2021). Lastly, the NEXUS readout chain

Comparison of noise results for OW200127		
	D1 @ Caltech	D2 @ FNAL
$J_{\mathrm{TLS}}(1\,\mathrm{kHz})$ $\left(\frac{(\delta f_r/f_r)^2}{\mathrm{Hz}}\right)$	2×10^{-19} to 10^{-20}	2×10^{-19} to 3×10^{-20}
feedline power P_g	$-105\,\mathrm{dBm}$ to $-75\,\mathrm{dBm}$	$-106.5\,\mathrm{dBm}$ to $-91.5\,\mathrm{dBm}$
photon number	3×10^4 to 4×10^7	5×10^4 to 2×10^6
τ_{qp}	50–100μs	6.5 ms
$\sigma_{E_{\mathrm{abs}}}$	6 eV	2.1±0.2 eV

Table 8.1: **Comparison of noise results for OW200127,** including TLS noise level and the feedline & resonator powers (in units of photon number) at which those TLS noise levels are measured.

included an isolator at the mK stage between the device output and the HEMT amplifier at 4 K. The isolator has two purposes: (1) prevent reflections and mitigate standing waves that can degrade R_Q and (2) prevent emission from the HEMT from reaching the detector and degrading its performance.

Both Caltech and NEXUS readout schemes used an Ettus Research Labs USRP X300 for measuring $S_{21}(f)$ frequency scans and $S_{21}(t;\,f_r)$ timestreams. Thus, both readout schemes required cleaning of the correlated electronics $1/f$ noise that was described in Section 7.6.

TLS noise

In Wen et al. (2022), TLS noise was shown to dominate the readout noise of the aluminum resonator. The previous chapter provides additional context for this result and describes the κ_1–κ_2 correlated-noise cleaning that was performed. After that noise was removed, the TLS noise was fit for (also described in the previous chapter). This procedure was followed in Temples et al. (2024), and the results of both devices are shown in Figure 8.4

Figure 8.4 Top shows that the TLS noise in both D1 and D2 becomes visible above the amplifier white noise when the white noise is below approximately $10^{-2}\,(\mu\mathrm{m}^{-3})^2/\mathrm{Hz}$. As the readout power is pushed higher, the white noise level in quasiparticle density units decreases, causing the TLS noise to become even more visible above the white noise.

Figure 8.4 Middle shows how the white and TLS noise levels were extracted from PSD fits. The residuals of the fits are also shown. Both fits show good agreement with the $f^{-0.5}$ power law that is generally expected for TLS noise PSDs.

Lastly, Figure 8.4 Bottom shows how TLS noise levels compare across a range of readout powers for the two devices. For the measurements of D1, the nominal attenuation in the readout chain was $-65\,\mathrm{dB}$.[1] Thus, in terms of feedline power P_g, the two data sets are roughly aligned along the x-axis at the lower end of readout powers that were measured in D1. A comparison of the noise performances is provided in Table 8.1 and shows good consistency across the range of readout powers at which both devices were studied.

[1]There was an additional $-25\,\mathrm{dB}$ of attenuation on top of the intentional $-40\,\mathrm{dB}$ for this particular data set because of a faulty copper coax at the input to the device.

Device 1 at Caltech

Device 2 at FNAL

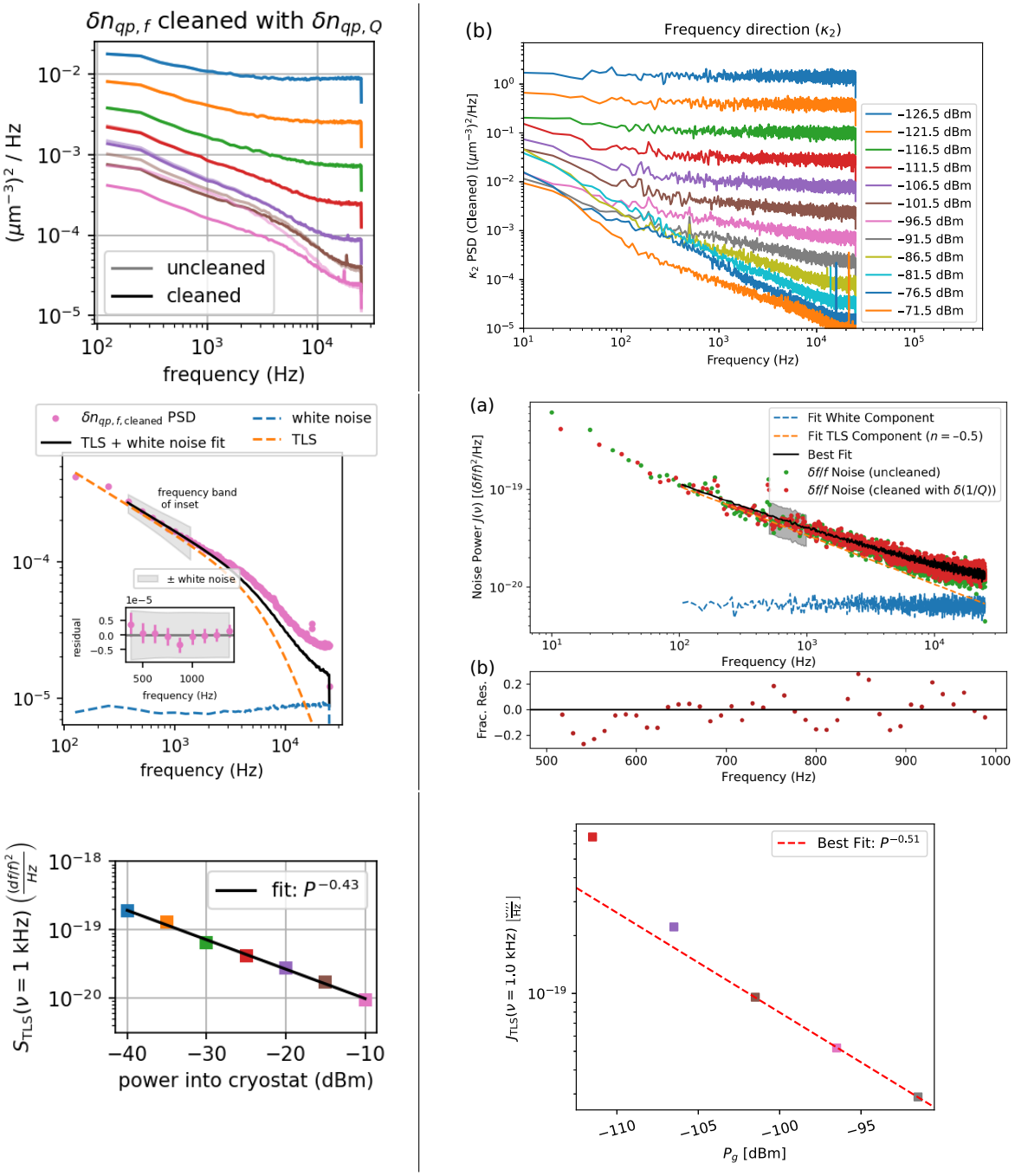

Figure 8.4: **A collection of all published plots illustrating the TLS noise in the aluminum resonators of OW200127.** Left Column: OW200127 D1 TLS results from testing at Caltech. Right Column: OW200127 D2 TLS results from testing at FNAL. Top Row: κ_2 direction PSDs at different readout powers. Middle Row: example of TLS noise + white noise PSD modeling. Bottom Row: TLS noise level as a function of readout power.

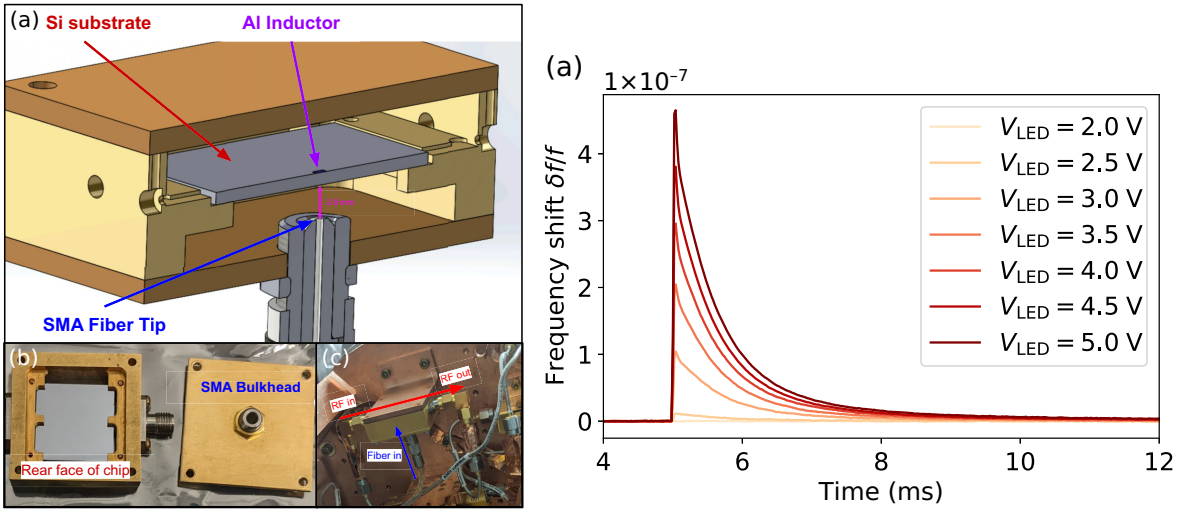

Figure 8.5: **LED pulses: setup and examples**. Left: images of the LED mounting scheme with the device box. The photons are deposited on the rear face of the detector. Right: examples of phonon-mediated detection of LED photons with kinetic inductance detectors. The different LED voltages correspond to different LED intensities and thus different pulse amplitudes. Each plotted timestream is averaged over several hundred pulses. Figures reproduced from Temples et al. (2024).

Measuring phonon collection efficiency using LED illumination

Temples et al. (2024) reports two measurements of η_{ph}: 0.66±0.1% and 0.89±0.11%, which are compatible to within 1.7σ. $\eta_{\mathrm{ph}} = E_{\mathrm{abs}}/E_{\mathrm{dep}}$ is measured from separate calibrations of E_{abs}, the energy absorbed by the phonon sensor, and E_{dep}, the energy deposited in the substrate. The intrinsic energy response is calibrated from δS_{21} to E_{abs} with Mattis-Bardeen theory. E_{dep} is measured by Poisson counting statistics as inferred from photon shot noise:

$$\begin{aligned}\sigma^2 &= \sigma_0^2 + \sigma_{\mathrm{Poisson}}^2 \\ &= \sigma_0^2 + \left(\sigma_{N_{\mathrm{photons}}} r\right)^2 \\ &= \sigma_0^2 + N_{\mathrm{photons}} r^2 \\ &= \sigma_0^2 + \mu r\end{aligned} \tag{8.3}$$

where μ and σ are the mean and width of a histogram corresponding to a pulse height μ, σ_0 is the optimally filtered baseline energy resolution, and r is the responsivity of the detector, i.e., the response of the detector in readout units per photon. r is used to convert photon number into readout units: $\mu = N_{\mathrm{photons}} r$ and $\sigma_{\mathrm{Poisson}} = \sqrt{N_{\mathrm{photons}}} r$ as used in Equation 8.3. This technique for inferring the photon number from Poissonian fluctuations was first demonstrated in Cardani et al. (2018).

Setup and procedure for taking LED pulse data

The setup for depositing LED photons in the KIPM detector is shown in Figure 8.5. 2.6 eV photons are deposited on the rear face of the detector. 2.6 eV is greater than the bandgap of silicon (1.11 eV) and so is sufficient to excite electrons into the conduction band, eventually creating phonons as the electron-hole pairs recombine.

The 2.6 eV photons were routed from a room-temperature fiber-coupled LED that was controlled by an arbitrary waveform generator. The voltage level on the arbitrary waveform

generator controlled the LED intensity, which set the number of photons that reached the detector. Different LED voltages yielded different pulse heights, as shown in Figure 8.5 Right. The timestreams in Figure 8.5 Right were averaged over several hundred pulses. Averaging was facilitated by the synchronization of the arbitrary waveform generator with the USRP: the USRP generates a single pulse per second (PPS) from its internal clock and initiates readout commands in sync with this PPS signal, so this signal was also used to trigger the arbitrary waveform generator. More details can be found in Temples et al. (2024).

An optical fiber coupled the LED photons to the detector. The fiber was thermalized at multiple temperature stages below 50 K and interrupted with a KG3 infrared filter at 100 mK to mitigate undesired blackbody photon radiation.

The LED was pulsed at 5 Hz for 200 seconds with 2 μs pulse width. The narrow pulse width ensured that the timescale of phonon production by the incident photons was much less than any timescale of the resonator.[2] Since the observed pulse fall times were about 6 ms, 5 Hz pulse repetition rate (200 ms between pulses) was deemed sufficient for the pulse to fully decay and recover to baseline before the next pulse began. The most essential part of the procedure is the repeated measurement of the detector response at many different LED intensities so that the photon number can be inferred from increasing pulse height fluctuations for larger pulse heights.

The readout of the KIPM detector was performed synchronously with the pulsing of the LED, as described above. In this experiment, regular checks of the LED-off timestreams were also performed so that calibration to the resonator basis was ensured. This calibration is described in Section 7.8. Lastly, since the RF readout was performed with an Ettus USRP, which has shown prominent phase noise above other noise sources (see Section 7.6), off-resonance tones were necessary to remove the correlated $1/f$ electronics noise. Importantly, cleaning coefficients must be computed with LED-off data and then applied to the LED-on data.

We now provide an analysis procedure for computing the number of photons deposited in the detector using the above LED-pulsing technique. This procedure references the LED-off data and cleaning tones that are mentioned above.

Procedure for cleaning and analyzing pulse data

1. Compute the cleaning coefficients from the LED-off data in the electronics basis, i.e., the radius and arc length directions. Confirm that the cleaning algorithm works by examining pre- and post-cleaning PSDs and cross-tone coherences (defined in Section 7.6) of the various tones, as was done in Section 7.6.
2. Clean the pulse data by using the cleaning coefficients computed in the previous step. It may be necessary to compute and subtract the quiescent point of the pulses first so that the cleaning tones and the to-be-cleaned tones are both centered around zero baseline. Confirm that the cleaning algorithm works by examining pre- and post-cleaning PSDs and cross-tone coherences of the non-pulse regions.
3. Average the pulses together. If the LED is synchronized with the S_{21} readout, leverage the fact that the pulse positions are known in time to perform the averaging.
4. Convert the complex S_{21} pulse shapes into some units that are linear with LED energy, such as $\delta f_r/f_r$, $\delta(1/Q_i)$, δn_{qp}, or E_{abs}.
5. Build amplitude histograms for each pulse height.
   a) Compute $\tilde{s}(f)$ from the largest energy, averaged pulse.
   b) Compute $J(f)$ from regions of the timestream in between pulses. Confirm the stability of $J(f)$ for each data set; $J(f)$ should not change between different pulse amplitudes.

[2] The shortest such timescale is the resonator ring-up time, given by $Q_r/\pi f_r \approx 19\,\mu$s.

Select a $J(f)$ that is representative of the typical $J(f)$, such as the median. The mean may suffer from distortion by outliers.

c) Compute the optimally filtered baseline resolution:

$$\sigma_0 = \sqrt{\frac{1}{\sum_{f_n} \frac{T|\tilde{s}(f_n))|^2}{J(f_n)}}}. \tag{8.4}$$

d) Apply the optimal filter to each pulse in the timestream to estimate the pulse's amplitude:

$$A = \frac{\sum_{f_n} \frac{\tilde{s}^*(f_n))\tilde{v}(f_n)}{J(f_n)}}{\sum_{f_n} \frac{|\tilde{s}(f_n))|^2}{J(f_n)}}, \tag{8.5}$$

and populate histograms with A.

e) Fit for the the mean μ and variance σ^2 of each histogram.

6. Fit the σ^2 vs. μ data with Equation 8.3 to measure the shot noise of the photons and infer the number of photons arriving at the detector.

Figure 8.6 Top shows the result of step 5 of this procedure. The data show good agreement with the Gaussian distribution fits, as indicated by the residuals plot and the χ^2 plot of Figure 8.6 Top. Figure 8.6 Bottom shows the results of step 6, which includes a fit of Equation 8.3. The legend shows the main result of the fit: the baseline energy of the detector $\sigma_E = 318.3 \pm 28.3\,\text{eV}$, which can be computed from $\sigma_0/r \times 2.6\,\text{eV}$. For the measurement shown in Figure 8.6, $\sigma_{E_{\text{abs}}} = 2.1 \pm 0.2\,\text{eV}$.

Once calibration to the total photon energy is performed, calculation of the phonon collection efficiency is now possible: $\eta_{\text{ph}} = E_{\text{abs}}/E_{\text{dep}} = \sigma_{E_{\text{abs}}}/\sigma_E$. As stated at the beginning of the section, Temples et al. (2024) reports two measurements of η_{ph}: 0.66±0.1% and 0.89±0.11%.

Discussion

Since the phonon collection decreased from the 1.1%±0.25% that was measured for a single resonator in DMLE2, the hypothesis that the niobium would be inert to phonons in the substrate was incorrect. There are two mechanisms to consider. First, niobium may more efficiently than expected down-convert the incident phonon energy into sub-$2\Delta_{\text{Al}}$ phonons due to the shedding of phonons by quasiparticles. Second, phonons may stay in the niobium too long and effectively become trapped. However, it is notable that the size of the resonator has decreased and the size of the dead volume (now considering Nb as dead volume) has increased enormously relative to DMLE2; the phonon collection efficiency has not scaled simply with (live metal) / (dead metal). This fact means that the niobium dead metal is likely less susceptible to phonon down conversion than the aluminum dead metal of DMLE2, but the sheer volume of niobium countered any gains from replacement of dead aluminum by Nb or the niobium-aluminum bilayer.

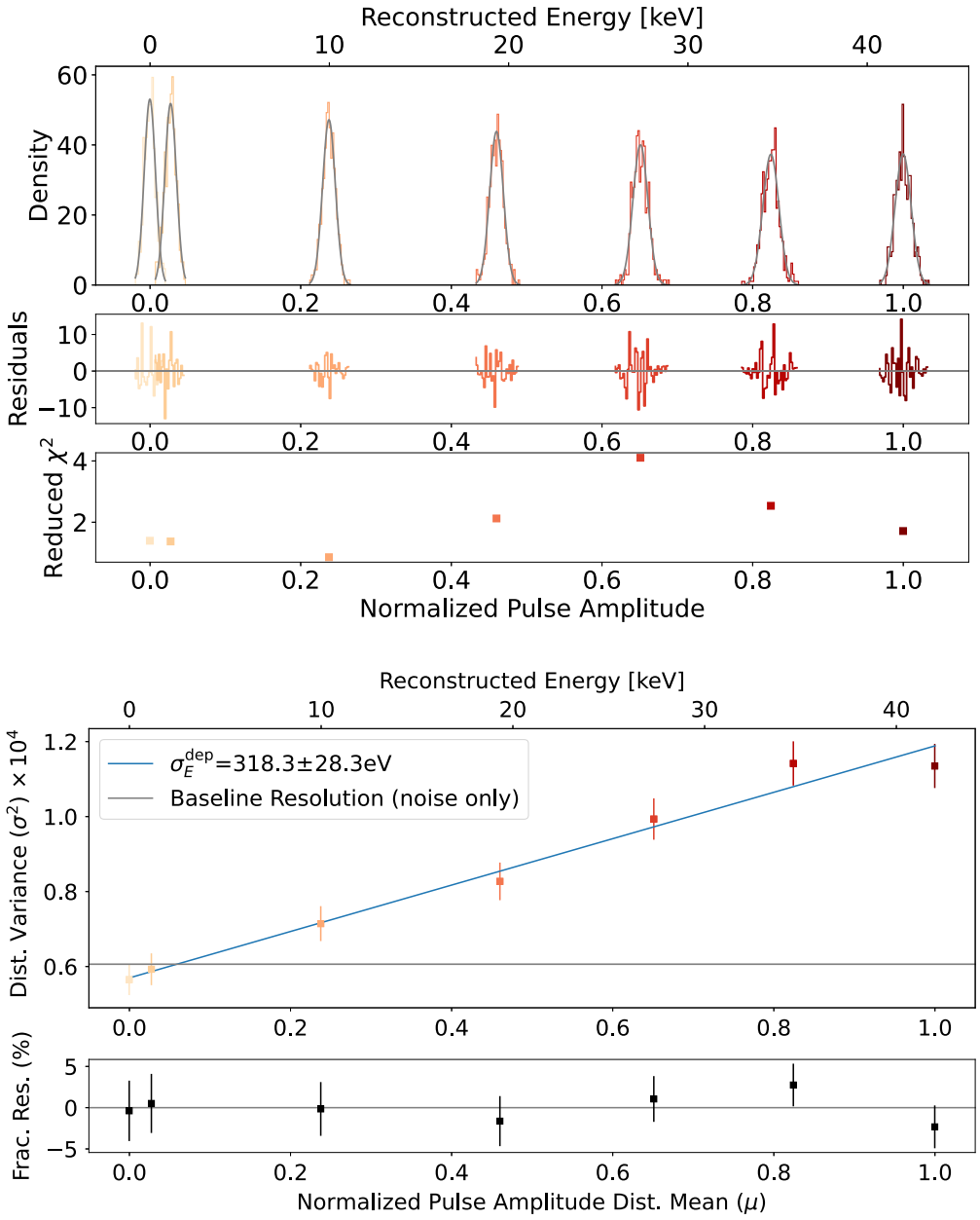

Figure 8.6: **LED pulse histograms and σ^2 versus μ for photon energy calibration.** Top: pulse height histograms for different LED intensities. The data are fit to Gaussian distributions shown in gray, and the residuals and χ^2 values of these fits are shown. Bottom: σ^2 versus μ for the different LED intensity pulse histograms and the fit for Equation 8.3 shown in blue. The fractional residuals of this fit are also shown.

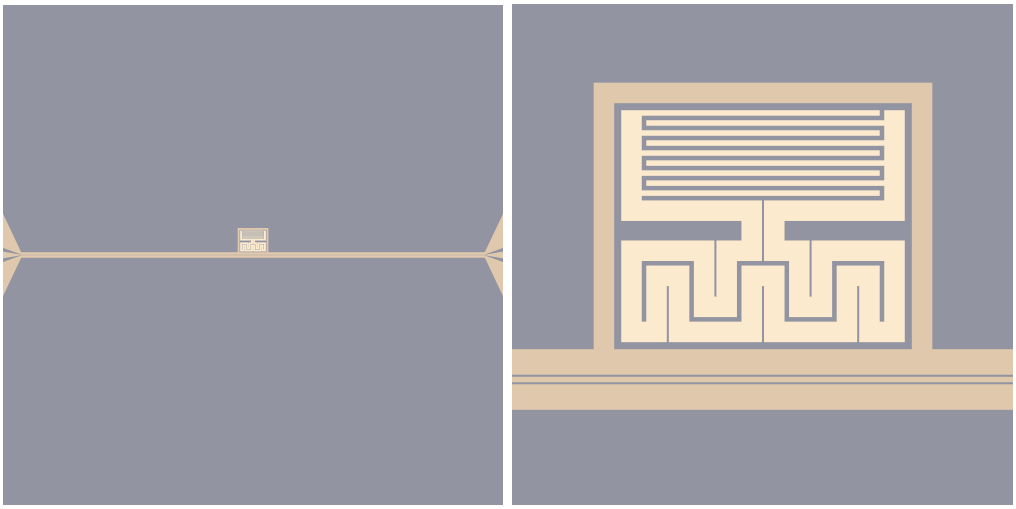

Figure 8.7: **B240103 device design.** The darker beige is niobium, and the lighter beige is aluminum. The device is 2.2 cm by 2.2 cm.

8.4 B240103, a single resonator device with reduced dead metal

Design considerations

OW200127 showed percent-level phonon collection efficiency despite all dead metal being replaced with niobium. It was hypothesized that the niobium on the chip could still prevent detectable phonon energy from reaching the aluminum. To test this hypothesis, our next device iteration (termed "B240103") vastly reduced the amount of Nb. The additional Nb pulsing resonators were eliminated. This change also allowed the feedline to be shortened because it only needs to serve one resonator. Furthermore, the thickness of the niobium decreased from 300 nm to 70 nm. The niobium thickness was not further reduced out of concern for elevated kinetic inductance, which could lead to feedline impedance mismatches and thus transmission non-idealities. Furthermore, previous experience from Cornell (2018) had indicated that the fabrication yield of 30 nm niobium feedlines was poor. The volume of niobium thus decreased from 8,570,000 μm^3 to 530,000 μm^3.

We also changed the resonant frequency of the Al KID from 4 to 5 GHz to better match the high-gain, low-T_N band of kinetic inductance traveling wave parametric amplifiers; see Figure 4 of Faramarzi et al. (2024) and the discussion in Section 7.9. To increase the resonant frequency, the inductor height was changed from 720 μm to 420 μm.

Experimental setup

The Cahill B101 Oxford Kelvinox 25 at Caltech possessed an LED calibration setup similar to the one in NEXUS at Fermilab; see Figure 8.8. The fiber running from 300 K to 4 K has connector breaks at two places, coinciding with two of the blackbody blocker disks in the dilution refrigerator access tube. The connector breaks were put in to prevent blackbody radiation from leaking through the holes through which the fiber passes.[3] There is a fiber break at the 4 K stage to allow the optical path to pass through an IR filter. The wavelength of the LED is 475 nm, which corresponds to 2.6 eV. As stated previously, this energy is well above the silicon bandgap, so these photons are efficiently absorbed in silicon (aside from the 30% interface reflectance due to index mismatch).

[3]Two other blocker disks do just have pass-through holes for the fiber.

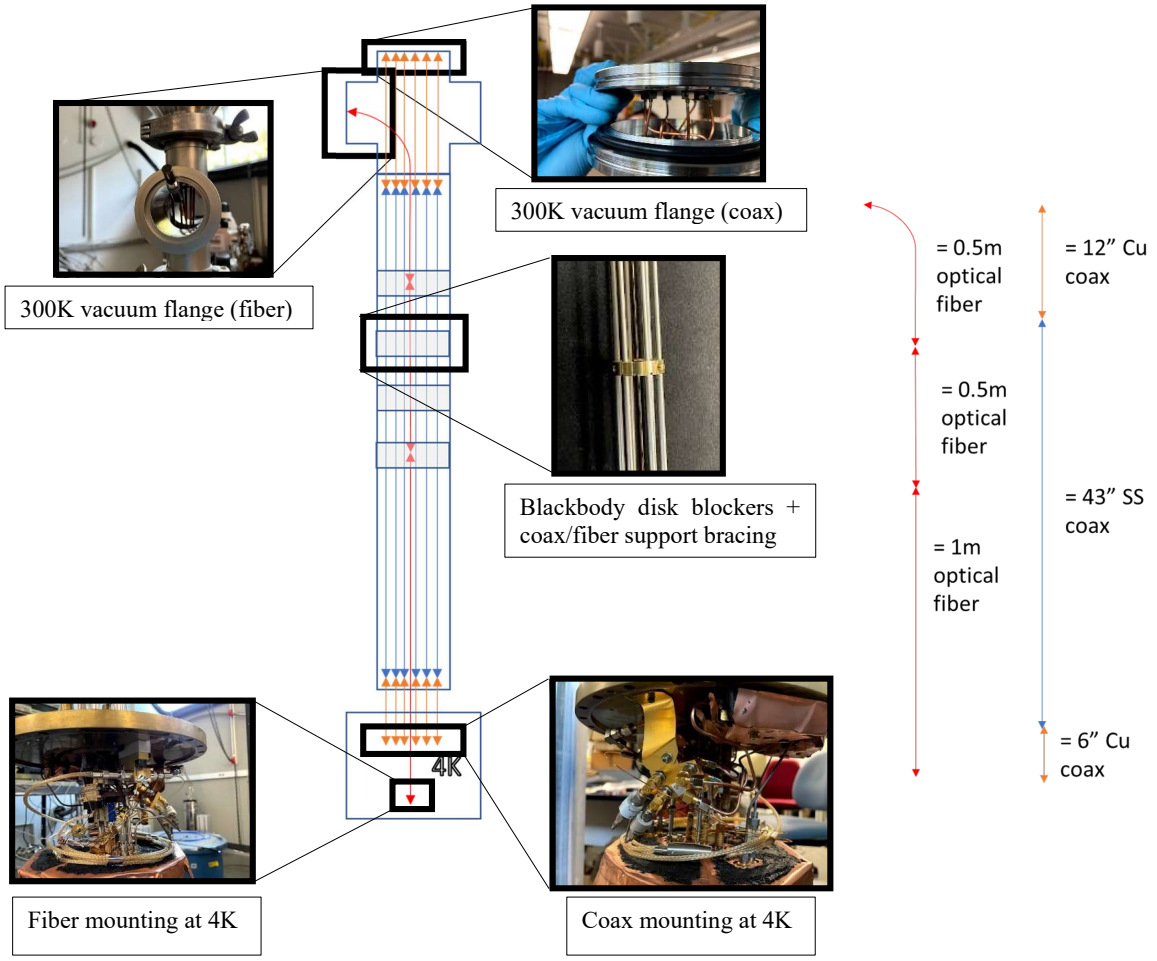

Figure 8.8: **Optical fiber and coax cable setup from 300 K to 4 K in the Cahill B101 Oxford Kelvinox 25 from July 2021 installation.** Left: schematic of six coax cables and one optical fiber routed from room temperature to the 4 K stage, with photos for certain components along the chain. Right: legend that depicts the breakdown of the optical fiber and coax components.

The same noise cleaning and pulsing procedure described in the previous section for the NEXUS measurement was also used for this measurement. Since the time constants of the pulses were much shorter (almost two orders of magnitude!), a faster pulse repetition rate was used for the B240103 measurement at Caltech: 100 Hz instead of 5 Hz as was used at NEXUS.

The coax readout chain includes three stages of intentional attenuation: $20\,\mathrm{dB}_{@4\,\mathrm{K}}$–$10\,\mathrm{dB}_{@1\,\mathrm{K}}$–$10\,\mathrm{dB}_{@100\,\mathrm{mK}}$, in accord with the noise calculation from Section 7.4. We used HEMT 266D, the details for which are provided in Siegel (2016). An unfortunate flaw in the Kelvinox 25 RF readout chain for this particular measurement was that the cables between 300 K and 4 K included 43" coaxes with stainless steel center conductors, adding an unnecessary ∼10 dB additional loss after the HEMT (and on the input side). A beryllium-copper center conductor would have been a better choice.[4]

[4]Stainless steel shields for such long cryogenic cables are usually fine because of the larger cross-sectional area.

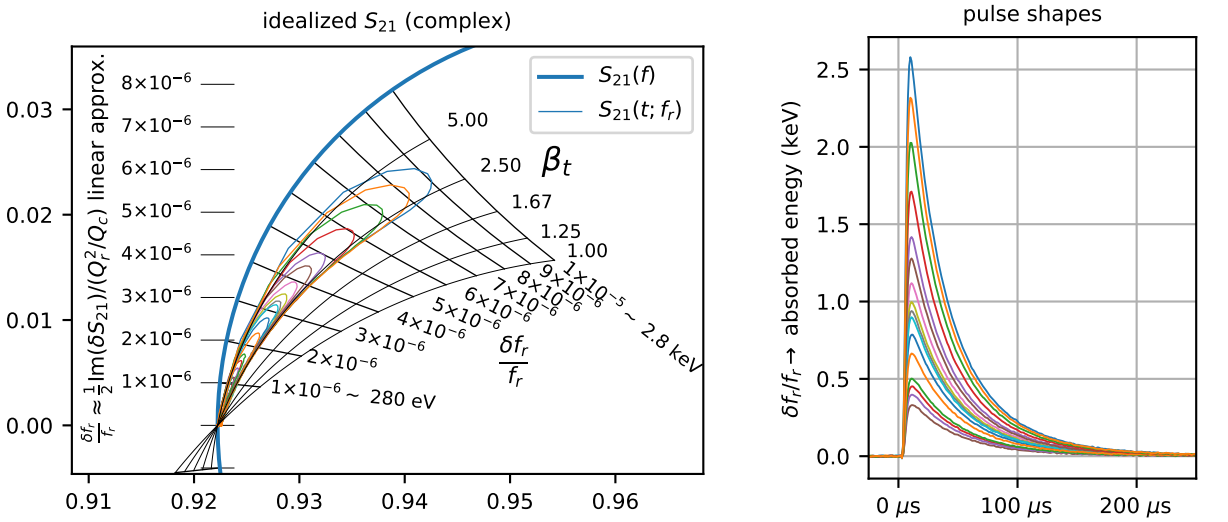

Figure 8.9: **LED pulses in B240103.** Left: B240103 averaged pulse shapes $S_{21}(t;\, f_r)$ with the $S_{21}(f)$ resonator VNA response and $\delta f_r/f_r$–β_t coordinates overlaid. The pulse first rises on the curve of higher β_t then falls along the curve of lower β_t. The rise of the pulse does not follow the expected trajectory of fixed β_t because the rise occurs during the resonator ring-up time. Right: B240103 averaged pulse timestreams.

RF readout was performed with an Ettus X310 USRP. The data-taking procedure was nearly identical to that used for OW200127 in NEXUS (see Section 8.3).

Pulse shapes

The response to 1 μs LED pulses is shown in Figure 8.9. Each orange curve is composed of 4000 pulses averaged together. In the complex S_{21} plot on the left, the pulses have been overlaid on the $\delta f_r/f_r$–β_t coordinate system and converted into $\delta f_r/f_r$ units following Equation 6.41. Pulses in $\delta f_r/f_r$ are then converted into absorbed energy E_{abs}, using the multiplicative factor $V_{\mathrm{ind}}\Delta/\alpha\gamma\kappa_2$.

The use of the nonlinear pulse model for extracting $\delta f_r/f_r$ was critical for the range of photon energies that was used in this analysis. Figure 8.9 Left shows how discrepant the pulses are from the linear approximation: for the largest energy pulse in this data set, the nonlinear model finds a $\delta f_r/f_r$ value of 9×10^{-6} whereas the linear model would yield a value closer to 6×10^{-6}.

In addition, it is visible that the pulse shapes on the falling edge tend to follow a constant β_t, which means that the ratio between the instantaneous $\delta f_r/f_r$ and $\delta(1/Q_i)$ is roughly constant over the course of the pulse. The edge, which follows a higher value of β_t and then transitions to the nominal value, should be ignored in this plot because the resonator response is modified during the resonator ring-up time.

The pulse fall times in Figure 8.9 Right are much shorter than the ms fall times observed in the OW200127 device at NEXUS, shown in Figure 8.5 Right. It is believed that the colder dilution refrigerator (~10 mK versus ~100 mK) and more stringent blackbody and RF shielding, both in free space and along the coaxial cable, contribute to the higher lifetime in NEXUS.

Building an optimal filter for energy reconstruction

There are two main components to the optimal filter: the signal template and the noise PSD. The peak-normalized signal templates are shown in Figure 8.10 Top Left. The inset of the

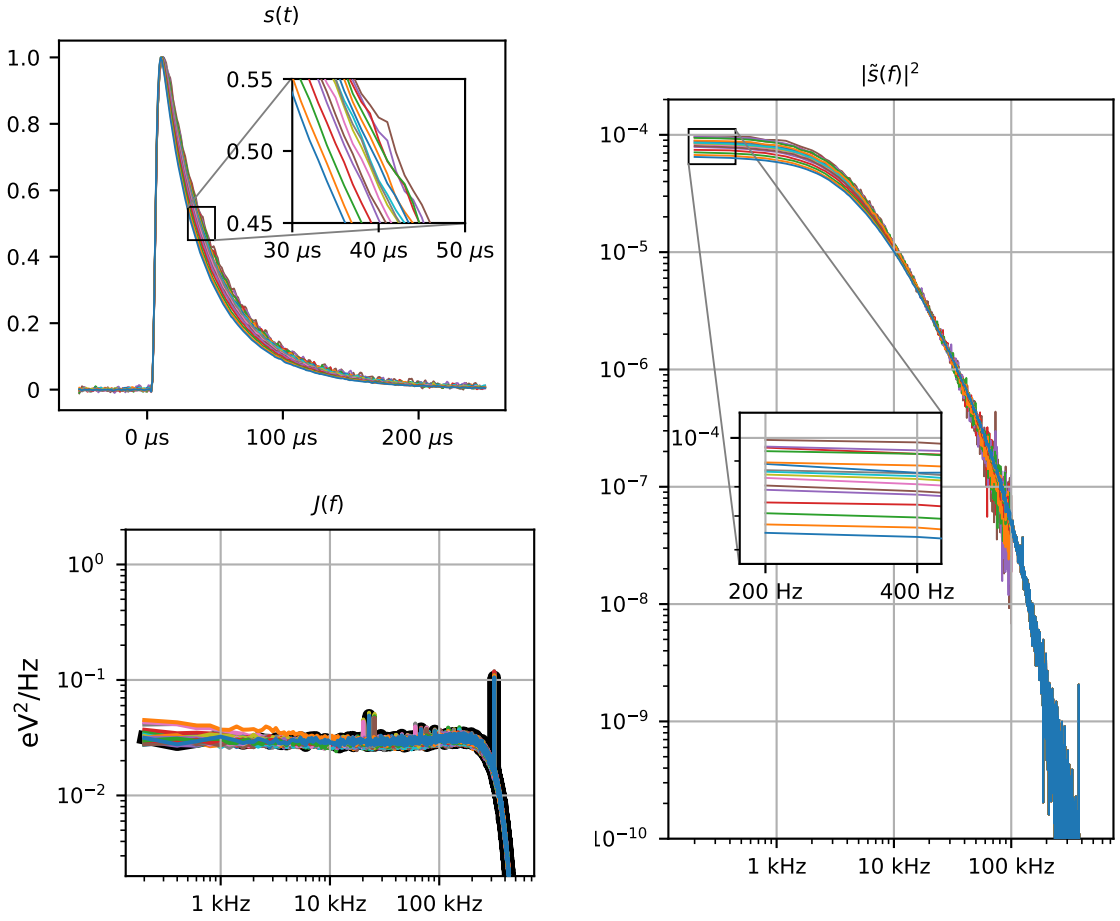

Figure 8.10: **B240103 optimal filter components.** Top Left: peak normalized pulses $s(t)$ with an inset to illustrate the different fall times for the different pulse heights. Right: squared magnitudes of the Fourier transform $|\tilde{s}(f)|^2$ with an inset to illustrate the different levels of low frequency amplitudes, a consequence of the different fall times. Bottom Left: noise power spectral density $J(f)$ for timestreams converted to units of E_{abs} in eV.

plot illustrates how the pulses corresponding to the largest energy depositions (blue, orange, green, etc.) are the pulses with the fastest fall times. All fall times are around 40 μs.

The Fourier coefficients $|\tilde{s}(f)|^2$ of these pulses are shown in Figure 8.10 Right. As was done in the time domain, the inset at low frequencies illustrates the longer pulse fall times for the higher energy pulses. The difference in $|\tilde{s}(f)|^2$ motivates the use of different signal templates for each pulse height. If, say, only the signal template derived from the largest energy pulse is used for amplitude estimation of the smaller energy pulses, then the OF algorithm tends to progressively overestimate the pulse heights as the fall times increase for progressively smaller pulse heights. The overestimation occurs because the individual pulse instances $\tilde{v}(f_n)$ for smaller pulses exhibit larger Fourier coefficients at low frequencies than those of the high energy template.

The noise on the pulse template at high frequencies is significant for the lower energy pulses, so, at high frequencies, we only use the highest energy pulse for all signal templates. The high frequency $\tilde{s}(f_n)$ from the highest energy pulse has been stitched together with the low frequency information for each pulse. The highest energy pulse has the best signal-to-noise ratio at higher

frequencies, and the signal shape is not expected to be different at these higher frequencies, which corresponds to the rise time of the pulse.

The noise PSDs $J(f_n)$ demonstrate good stability over the course of the data-taking as seen in Figure 8.10 Bottom Left. The noise is white due to the much worse Q_r that was observed for this resonator, increasing the relative contribution of amplifier noise compared to OW200127, the noise of which was dominated by TLS noise. An arbitrary PSD is chosen for the construction of the optimal filter, shown in black.

Energy histograms and resolution modeling

In this data set, there are 16 different pulse amplitudes. At each pulse amplitude, 4000 pulses are used to populate a histogram. The histograms display visible broadening with energy in Figure 8.11 Top. There are three expected causes.

- Different OF baseline resolutions as a result of the different pulse fall times and consequent signal templates. We express the energy dependence of the baseline resolution as $\sigma_0(E_{\text{abs}})$. We use a linear model to describe $\sigma_0(E_{\text{abs}})$; see purple line in Figure 8.11 Bottom.
- The reduction of the S_{21} response to $\delta f_r/f_r$ for larger $\delta f_r/f_r$. The resolution degradation takes the following form from Chapter 6:

$$\sigma_{\delta f_r/f_{r,0}} = \left|1 + Q_{r,0}\frac{\delta f_r}{f_{r,0}}\left(\frac{1}{\beta_t} - 2i\right)\right|^2 \sigma_{\delta f_r/f_{r,0}}\Big|_{f_r=f_{r,0}}. \qquad \text{(6.45 revisited)}$$

  Conversion to absorbed energy units produces the following:

$$\sigma^*_{E_{\text{abs}}}(E_{\text{abs}}) = \left|1 + Q_{r,0}\frac{\delta f_r}{f_{r,0}}(E_{\text{abs}})\left(\frac{1}{\beta_t} - 2i\right)\right|^2 \sigma_0(E_{\text{abs}}). \tag{8.6}$$

  $\sigma^*_{E_{\text{abs}}}$ is the energy resolution to a monochromatic energy deposition. The term in $|\cdot|^2$ is shown in brown in Figure 8.11 Bottom. The entire expression is shown in red. A time-averaged value of β_t is used since β_t is roughly constant over the fall time of the pulses, as shown in Figure 8.9 Left.
- Photon shot noise results in different amounts of deposited energy for each pulse (LED pulses are not monochromatic) and is added in quadrature:

$$\sigma_{E_{\text{abs}}}(E_{\text{abs}}) = \sqrt{\sigma^{*2}_{E_{\text{abs}}}(E_{\text{abs}}) + E_{\text{abs}}\eta_{\text{ph}}}. \tag{8.7}$$

  Note that, when we work in units of absorbed energy, the responsivity factor r from Equation 8.3 is equal to η_{ph}, i.e., η_{ph} converts the deposited energy into the absorbed energy.

The baseline energy resolution $\sigma_{E_{\text{abs}}}(0) = \sigma_0(0)$ is 23 eV. It is extracted from the y-intercept of the energy-dependent OF resolutions. The resolution is worse for B240103 than for OW200127 as a result of the lower quality factor and shorter pulse lifetime.

We fit $\sigma_{E_{\text{abs}}}$ versus E_{abs} to Equation 8.7 with one free parameter, η_{ph}. The error bars from Figure 8.11 Bottom are constructed from statistical uncertainties: simply the measured $\sigma_{E_{\text{abs}}}$ of the histogram divided by $\sqrt{4000}$. To estimate the goodness of fit, we compare the χ^2 of the best fit model to the χ^2 distribution for 15 degrees of freedom (the number of data points less the single fit parameter). The computed χ^2 is greater than 98.6% of χ^2 values.

The final reported collection efficiency is 2.7%±0.63%. The uncertainty was calculated by finding the collection efficiencies that would increase χ^2 by 1 from its best-fit value. Because the poor goodness-of-fit calls into question the validity of estimating the uncertainty in this way, one should only think of it as giving a general sense of the uncertainty. The curve for $\eta_{\text{ph}} = 30\%$ is shown for comparison.

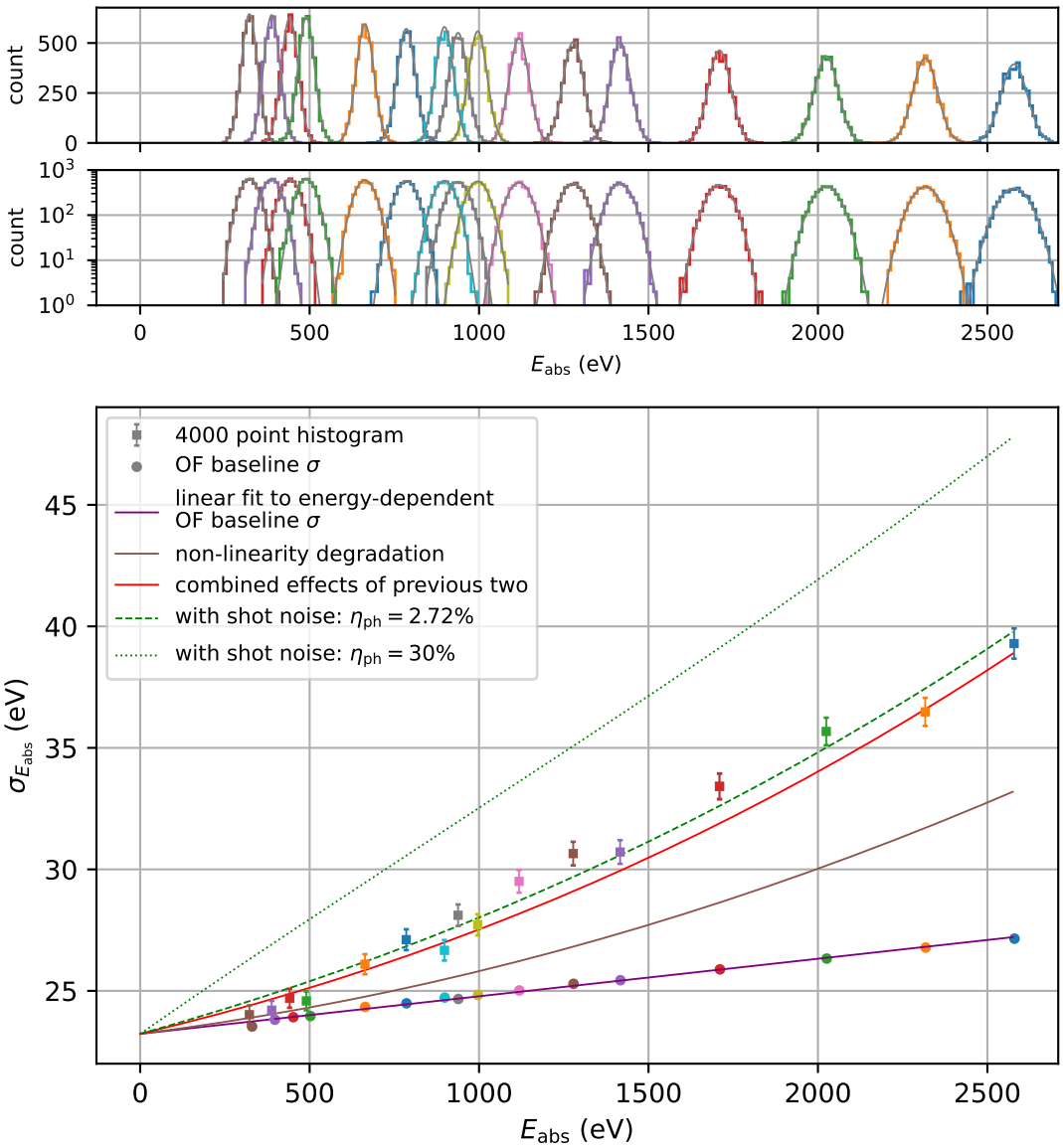

Figure 8.11: **B240103 pulse histograms and resolution modeling.** Top: histograms of the LED pulses in B240103 in both linear and logarithmic y-axes along with fits to each histogram in grey. Bottom: $\sigma_{E_{\mathrm{abs}}}$ versus the mean pulse height as extracted from each histogram. The energy-dependent OF baseline $\sigma_{E_{\mathrm{abs}}}$ is also plotted; the dependence arises from the different low-frequency, long-time behavior of the different energy pulses. The plotted curves are the different contributions to the total energy-dependent resolution and various combinations thereof: the changing baseline resolution, the resolution degradation due to the nonlinearity of the pulse, and photon shot noise.

DMLE2 OW200127 B240103

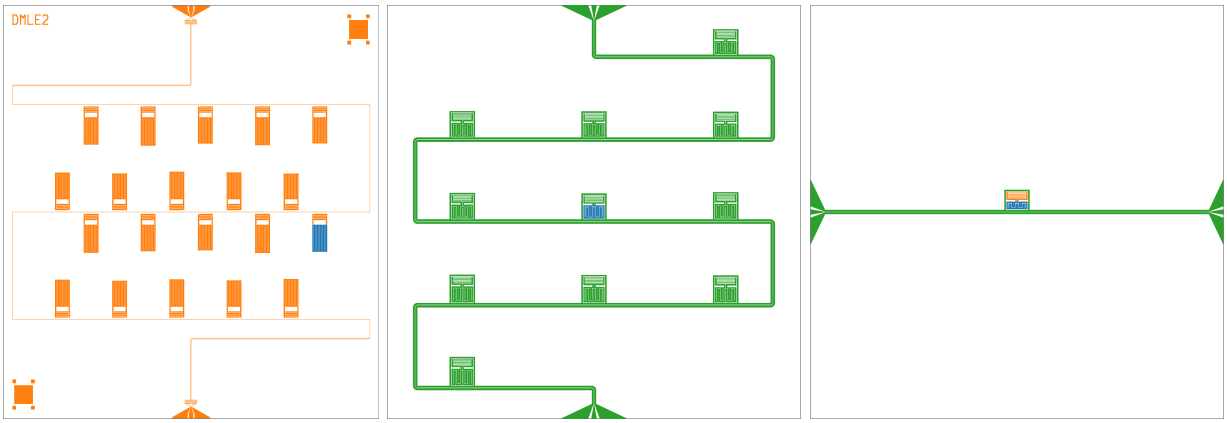

Figure 8.12: **Three KIPM detector designs used to build a model for η_{ph}.**

8.5 Phonon collection efficiency modeling

In this section, we attempt to build a model to describe the phonon collection efficiencies of three different KIPM detectors for which η_{ph} has been measured. They are shown in Figure 8.12. The design of OW200127 attempted to boost the phonon collection efficiency by fashioning all dead metal out of higher-gap material, including the capacitor of the sensing resonator. The collection efficiency, averaged over the two measurements, was 0.78%±0.07%. B240103 removed most of the niobium by having a thinner niobium feedline and no auxiliary resonators and not including a niobium capping layer on the capacitor. The collection efficiency using the nonlinear pulse shape function was measured to be 2.7%±0.63%.

For DMLE2, the phonon collection efficiency of all 20 resonators was 7.0%±1.1%. In order to put this device on the same footing as OW200127 and B240103, only one resonator is examined. What this means is a scaling of the collection efficiency to just the closest resonator to a particular event. Using the nearest resonator is appropriate because, in the OW200127 and B240103 tests, the photons were deposited in the detector right under the single resonator. Using the plots from Figures 1 and 2 from Moore et al. (2012), we estimate that the closest resonator to the event absorbs approximately 1/7 of the total KID-absorbed energy. The estimate of the phonon collection efficiency of just a single resonator is 1.4%±0.25%, the error bar for which comes from taking the error bar of the 20-resonator collection efficiency and dividing by $\sqrt{20}$.

These three detectors are particularly useful for building such a model because their mounting schemes are the same and because the film thicknesses are almost the same: 25 nm for DMLE2 and 30 nm for the other two.

Pair-breaking efficiency η_{pb}

In this model, an initial event's energy is fully converted into phonon energy E. The pair-breaking efficiency η_{pb}, also known as the quasiparticle generation efficiency, puts an upper limit on the fraction of E that is detectable. This factor comes into play during any pair-breaking event, be it from phonons or photons, and is due to the fact that quasiparticles shed phonons until they approach the gap energy.

We estimate that η_{pb} ranges from 0.32 to 0.6. The lower bound comes from a measurement of a 40 nm thick aluminum KID used to detect single 25 μm photons; see Day et al. (2024). These photons were 12.1 THz (50 meV), approximately a factor of 10 higher in frequency and energy

Volume-scaling model inputs					
	V_{Al} (μm^3)	V_{ind} (μm^3)	η_{ph}	$\frac{\eta_{\mathrm{pb}}}{\eta_{\mathrm{ph}}}V_{\mathrm{ind}} - V_{\mathrm{Al}}$ (μm^3)	V_{Nb} (μm^3)
DMLE2	800,000	505,000	7.0% ± 1.1%	6,420,000	0
DMLE2, one resonator	800,000	25,300	1.1%±0.25%	266,000	0
OW200127	21,000	21,000	0.78%±0.07%	2,690,000	1,230,000
B240103	21,900	11,900	2.7%±0.63%	685,000	180,000

Table 8.2: **Volume-scaling model inputs.**

than the ∼1 THz phonons (4 meV) that are absorbed by KIDs in a phonon-mediated detector (see Section 3.2). Lower energy pair-breaking events are expected to more efficiently generate quasiparticles, so the lower bound is set by this measurement even though the 40 nm film is thicker than the films under study in this model.[5]

The upper bound of 0.6 comes from Guruswamy et al. (2014) and is expected to be reached for low energy pair-breaking events. Figure 7 in Guruswamy et al. (2014) estimates $\eta_{pb} \to 0.6$ for 3.6 meV (740 GHz) pair-breaking quanta in the thick film limit. With the Guruswamy et al. (2014) framework, the pair-breaking efficiency in our thin 25 nm to 30 nm aluminum films must be less than 0.6. For the sake of model building, we focus on a benchmark value of $\eta_{\mathrm{pb}} = 0.46$, the average of the upper and lower bounds.

A volume-scaling model

We now model how the $\eta_{\mathrm{pb}}E$ of collectable phonon energy is absorbed in three places: the aluminum, the niobium, or the mounting. By conservation of energy, we write

$$\eta_{\mathrm{pb}}E = \xi_{\mathrm{Al}}V_{\mathrm{Al}} + \xi_{\mathrm{Nb}}V_{\mathrm{Nb}} + E_{\mathrm{loss}} \tag{8.8}$$

where ξ is the efficiency with which either the aluminum or the niobium absorbs the collectable phonon energy. E_{loss} encompasses loss to the mounting or phonon down-conversion at the bare polished surfaces, the bare diced surfaces, and the surface-mount interface. Down-conversion at bare polished surfaces is expected to be subdominant.

We define the phonon collection efficiency as

$$\eta_{\mathrm{ph}} = \frac{\xi_{\mathrm{Al}}V_{\mathrm{ind}}}{E}. \tag{8.9}$$

We care about how much of the initial energy is collected in the inductor. We assume that all of the energy collected in the aluminum is then down-converted and thereafter rendered uncollectible, i.e., phonon recycling from quasiparticle recombination is not taken into account.

Combining Equation 8.8 and 8.9,

$$E = \frac{\xi_{\mathrm{Al}}V_{\mathrm{ind}}}{\eta_{\mathrm{ph}}}$$

$$\frac{\xi_{\mathrm{Al}}V_{\mathrm{Al}} + \xi_{\mathrm{Nb}}V_{\mathrm{Nb}} + E_{\mathrm{loss}}}{\eta_{\mathrm{pb}}} = \frac{\xi_{\mathrm{Al}}V_{\mathrm{ind}}}{\eta_{\mathrm{ph}}}$$

[5] Thicker films are expected to have higher η_{pb} because the time for phonons to escape the film is longer than for a thinner film.

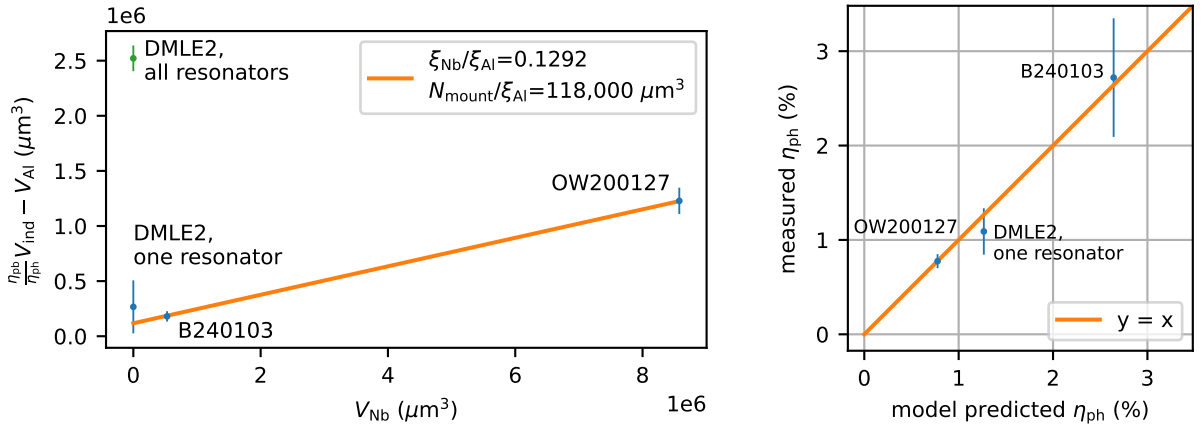

Figure 8.13: $\eta_{\rm ph}$ **volume-scaling modeling.** Left: the results of fitting Equation 8.10 with the three available data points. Right: re-expression of the fit results in terms of measured $\eta_{\rm ph}$ versus modeled $\eta_{\rm ph}$.

$$\frac{\xi_{\rm Nb}}{\xi_{\rm Al}}V_{\rm Nb} + \frac{E_{\rm loss}}{\xi_{\rm Al}} = \frac{\eta_{\rm pb}}{\eta_{\rm ph}}V_{\rm ind} - V_{\rm Al}. \tag{8.10}$$

We now build a linear model where $\xi_{\rm Nb}/\xi_{\rm Al}$ is the slope and $E_{\rm loss}/\xi_{\rm Al}$ is the y-intercept. $\xi_{\rm Nb}/\xi_{\rm Al}$ is the relative efficiency with which niobium and aluminum absorb phonon energy, and $E_{\rm loss}/\xi_{\rm Al}$ is the equivalent amount of dead aluminum that $E_{\rm loss}$ mimics.

The relevant inputs into the model are shown in Table 8.2, including material volume, collection efficiency, and the right-hand side of Equation 8.10. $\eta_{\rm ph}/V_{\rm ind}$ may be interpreted as the efficiency per unit volume of the phonon sensor. This expression dominates the value for $(\eta_{\rm pb}/\eta_{\rm ph})V_{\rm ind} - V_{\rm Al}$. The average DMLE2 resonator is less efficient per unit volume than the B240103 resonator, likely due to position dependence; resonators that are far from the event are less efficient at collecting the event's energy. The "one resonator" data is derived from the closest resonator to the event and shows a more comparable efficiency per unit volume.

The fit of the model to the data is shown in Figure 8.13 Left. The discrepancy between the "DMLE2, all resonators" point and the other two points in the model motivated the investigation of the phonon collection efficiency for a single resonator of DMLE2. The fitted model parameters show that (1) niobium is about 1/8 as efficient in absorbing and down-converting phonons per unit volume than aluminum and (2) the generic $E_{\rm loss}$ is effectively 118,000 μm^3 of dead aluminum, which is 5 to 10 times larger than the phonon-absorbing inductors with volume $V_{\rm ind}$.

The fit parameters $\xi_{\rm Nb}/\xi_{\rm Al}$ and $E_{\rm loss}/\xi_{\rm Al}$ (and by combination $\xi_{\rm Nb}/E_{\rm loss}$) can be used to calculate $\eta_{\rm ph}$, $\eta_{\rm dAl}$, $\eta_{\rm Nb}$, and $\eta_{\rm loss}$, which are the absorption efficiencies in the phonon sensor, the dead aluminum, the niobium, and the unknown losses, for each device.

$$\eta_{\rm ph} = \frac{\xi_{\rm Al}V_{\rm ind}}{N} = \frac{\xi_{\rm Al}V_{\rm ind}}{\xi_{\rm Al}V_{\rm Al} + \xi_{\rm Nb}V_{\rm Nb} + E_{\rm loss}} = \frac{V_{\rm ind}}{V_{\rm Al} + (\xi_{\rm Nb}/\xi_{\rm Al})V_{\rm Nb} + E_{\rm loss}/\xi_{\rm Al}} \tag{8.11}$$

$$\eta_{\rm dAl} = \frac{\xi_{\rm Al}(V_{\rm Al} - V_{\rm ind})}{N} = \frac{\xi_{\rm Al}(V_{\rm Al} - V_{\rm ind})}{\xi_{\rm Al}V_{\rm Al} + \xi_{\rm Nb}V_{\rm Nb} + E_{\rm loss}} = \frac{V_{\rm Al} - V_{\rm ind}}{V_{\rm Al} + (\xi_{\rm Nb}/\xi_{\rm Al})V_{\rm Nb} + E_{\rm loss}/\xi_{\rm Al}} \tag{8.12}$$

$$\eta_{\rm Nb} = \frac{\xi_{\rm Nb}V_{\rm Nb}}{N} = \frac{\xi_{\rm Nb}V_{\rm Nb}}{\xi_{\rm Al}V_{\rm Al} + \xi_{\rm Nb}V_{\rm Nb} + E_{\rm loss}} = \frac{V_{\rm Nb}}{(\xi_{\rm Al}/\xi_{\rm Nb})V_{\rm Al} + V_{\rm Nb} + (E_{\rm loss}/\xi_{\rm Al})(\xi_{\rm Al}/\xi_{\rm Nb})} \tag{8.13}$$

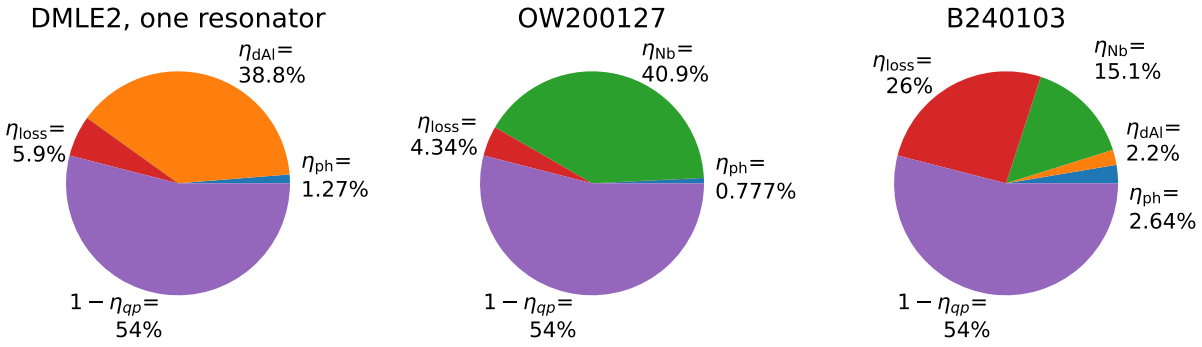

Figure 8.14: **Pie charts of phonon losses.**

$$\eta_{\text{loss}} = \frac{E_{\text{loss}}}{N} = \frac{E_{\text{loss}}}{\xi_{\text{Al}}V_{\text{Al}} + \xi_{\text{Nb}}V_{\text{Nb}} + E_{\text{loss}}} = \frac{1}{(\xi_{\text{Al}}/E_{\text{loss}})V_{\text{Al}} + (\xi_{\text{Nb}}/\xi_{\text{Al}})(\xi_{\text{Al}}/E_{\text{loss}})V_{\text{Nb}} + 1}. \tag{8.14}$$

In Figure 8.13 Right, we show the capacity of our model to predict η_{ph}; the data show good agreement with the model, but there is, of course, a warranted scrutiny for a model with two parameters and three data points. We also provide a breakdown of the model-inferred η_{ph}, η_{dAl}, η_{Nb}, and η_{loss} in each device, shown in Figure 8.14. Here are some observations of the results:

- Phonon losses in DMLE2 were dominated by the other resonators and dead metal on the surface.
- OW200127 was meant to remedy this issue by replacing the dead aluminum with niobium. This remedy did not work; the niobium turned out to be more down-converting than expected.
- B240103 has much less niobium. The result is that the dominant loss mechanism is now the generic loss term, which is currently modeled to be equivalent to approximately 118,000 μm^3 of dead aluminum, about 10 times the volume of active aluminum.

8.6 Summary and outlook toward sub-eV energy thresholds

Table 8.3 provides a summary of the performance of the three KIPM detector designs, reporting σ_E, η_{ph}, $\sigma_{E_{\text{abs}}}$ and the Q conditions and dominant noise source.

For the hypothetical single-resonator version of DMLE2, the intrinsic energy resolution is limited by amplifier noise as a result of being overcoupled ($Q_c \ll Q_i$) but is still respectable at 5.9 eV. The poor phonon collection efficiency results in a resolution on deposited energy much worse than our eV-scale goal.

In OW200127, efforts to improve the phonon collection efficiency did not work, as described in this section, but a better resolution on absorbed energy was achieved due to the long quasiparticle lifetime that was observed. Improvements in the coupling of the resonator to the feedline in combination with a respectable $Q_i > 10^5$ led to a reduced contribution from the amplifier noise but an uncovering of the TLS noise, which should limit $\sigma_{E_{\text{abs}}}$ of this design to 500 meV to 2 eV as calculated in Section 7.3. The TLS noise is particularly harmful to noise performance because it degrades the benefit of improved quasiparticle lifetime: $\sigma_{E_{\text{abs}}} \propto \tau_{\text{qp}}^{-1/4}$ for $f^{-0.5}$ TLS noise instead of $\sigma_{E_{\text{abs}}} \propto \tau_{\text{qp}}^{-1/2}$ for white noise.

Lastly, B240103 showed an improved phonon collection efficiency due to the removal of large amounts of niobium, but the low $R_Q = aQ_r^2/|\hat{Q}_c|$ compared to OW200127 increased the impact of

Summary of KIPM detector performance				
device	σ_E	η_{ph}	$\sigma_{E_{\mathrm{abs}}}$	Q conditions and dominant noise
DMLE2, one resonator*	540 eV	1.1%	5.9 eV	amplifier $Q_c \ll Q_i$, $Q_c = 5 \times 10^3$–5×10^4
OW200127	320 eV	0.78%	2.1 eV	TLS $Q_r = 2.5 \times 10^5$, $Q_i \approx 0.6 Q_c$
B240103	850 eV	2.7%	23 eV	amplifier $Q_i \ll Q_c$, $Q_i = 2.9 \times 10^4$

Table 8.3: **Summary of KIPM detector performance.** The different components of Equation 8.1 are shown, as well as a column for the Q conditions and dominant noise source to show how they may play a role in the intrinsic energy resolution $\sigma_{E_{\mathrm{abs}}}$. The * indicates that all the values in that row are not directly measured and are instead extrapolated from the provided data to a hypothetical single resonator device.

amplifier noise. $Q_r \approx Q_i \ll Q_c$ is much more detrimental to noise performance than $Q_r \approx Q_c \ll Q_i$ because Q_c appears in the denominator of R_Q; low Q_c yields large S_{21} resonator circles.

The next most important step to improve the energy resolution of KIPM detectors is to bring the phonon collection efficiency to the 30% level that has been observed by other phonon-mediated detectors; see Ren et al. (2021), Romani et al. (2024), and TESSERACT Collaboration (2024). There are two techniques that are currently being explored. First, the volume-scaling model described in this section predicts a 26% energy loss to E_{loss}, shown in Figure 8.14. We are not sure whether E_{loss} is dominated by the mounting, but it may be relatively straightforward to reduce E_{loss} by replacing the current mounting with wirebond suspension (Anthony-Petersen et al., 2024).

A second option to improve the phonon collection efficiency is to separate the functions of the phonon absorber and the sensitive element. The 600 nm aluminum fins that have been shown to efficiently absorb phonons could then be used to increase the collection efficiency. The fins would need to be connected to a lower T_c resonator. Hafnium resonators with T_c 140 to 200 mK are being explored as an option (Li, 2024). Reduction of T_c should increase the yield of quasiparticles per unit energy absorbed by decreasing the pair-breaking energy, but materials with a smaller gap may require lower readout powers, increasing the relative contribution of amplifier noise.

Another benefit of incorporating a phonon-absorbing fin into the design is that the resonator can now be made much smaller to increase its sensitivity. The current designs use large resonators to provide sufficient volume and surface area coverage to collect phonons directly. With a separate phonon absorber, the resonator does not need to play this dual role and can be optimized solely as a sensitive element.

A final technique for improving detector energy resolution involves the lengthening of quasiparticle lifetimes, which is believed to be dominantly controlled by environmental blackbody radiation energetic enough to break Cooper pairs in the superconducting sensor. As stated before, energy resolution scales as $\tau_{\mathrm{qp}}^{-1/2}$ to $\tau_{\mathrm{qp}}^{-1/4}$, depending on whether amplifier or TLS noise is dominant. The 50 μs to 100 μs pulses that have been observed at Caltech are much shorter than the 6.5 ms pulses observed in NEXUS at Fermilab, which is one of the main reasons for the much improved $\sigma_{E_{\mathrm{abs}}}$ for OW200127 (in addition to the higher Qs). The colder dilution refrigerator and more aggressive inline filtering at NEXUS are believed to contribute to the improved quasiparticle lifetime. Presently, upgrades are planned for the Caltech dilution refrigerator to improve blackbody radiation shielding.

Part IV

Appendices

Appendix A

Computing dynamic structure factors

In this Appendix, we follow Trickle et al. (2020) and explore how $S(\vec{q},\omega)$ is computed for free nuclear recoils and for single-phonon couplings when dark matter couples only to nucleons. The regimes in which either calculation applies is also explored. As stated in Trickle et al. (2020), the steps to compute $S(\vec{q},\omega)$ are (1) parametrize $|i\rangle$ and $|f\rangle$ according to the excitation type and (2) quantize $\mathcal{F}_T(\vec{q})$ such that it induces transitions $|i\rangle \to |f\rangle$.

Nuclear recoils

For nuclear recoils, $|i\rangle$ and $|f\rangle$ can be expressed as a direct product of momentum eigenstates $|\vec{k}_i\rangle_J$ and $|\vec{k}_f\rangle_J$ where J indexes over each nucleus. Only one nucleus receives momentum in its final state. These eigenstates can be expressed in terms of creation operators:

$$|\vec{k}_i\rangle_J = V^{-\frac{1}{2}}\hat{b}^\dagger_{\vec{k}_i}|0\rangle_J, \quad |\vec{k}_f\rangle_J = V^{-\frac{1}{2}}\hat{b}^\dagger_{\vec{k}_f}|0\rangle_J. \tag{A.1}$$

The creation and annihilation operators $\hat{b}^\dagger$ and $\hat{b}$ follow the canonical commutation relations.

The target form factor, with some manipulation not shared here, can be expressed in terms of a weighted average nucleon form factor $f_N \equiv f_p Z + f_n(A-Z)$, a nucleus form factor $F_N(\vec{q})$, and the nucleus number density $\tilde{n}_N(-\vec{q})$:

$$\mathcal{F}_T(\vec{q}) = \frac{f_N}{f_n} F_N(\vec{q})\tilde{n}_N(-\vec{q}). \tag{A.2}$$

The nucleus form factor $F_N(\vec{q})$ is usually taken to be the Helm form factor. Note that Equation A.2 is referenced against the neutron form factor f_n.

Quantization of $\mathcal{F}_T(\vec{q})$ involves expressing $\tilde{n}_N(-\vec{q})$ in terms of a number operator $\hat{b}^\dagger\hat{b}$. In broad strokes, the matrix element $\langle\vec{k}_f|\tilde{n}_N(-\vec{q})|\vec{k}_i\rangle_J$ takes the form $\langle 0|\hat{b}\hat{b}^\dagger\hat{b}\hat{b}^\dagger|0\rangle$, which can be calculated by sending the annihilation operators to the right so that they act on the ground state. After applying commutation relations, the result is a delta function $\delta^3(\vec{k}_f - \vec{k}_i - \vec{q})$; momentum is conserved. This delta function is evaluated when the sum over the final states f, i.e., the integral over final momentum $\vec{k}_f$, is performed.

Skipping the finer details of the calculation, we reproduce the expression for the dynamic structure factor of nuclear recoils from Trickle et al. (2020):

$$S(\vec{q},\omega) = 2\pi\frac{\rho_T}{m_N}\frac{f_N^2}{f_n^2}F_N^2(\vec{q})\delta\left(\omega - \frac{q^2}{2m_N}\right), \tag{A.3}$$

where a sum over all target nuclei has been performed to compute the rate over the entire detector, making the final result dependent on ρ_T/m_N, the number density of target nuclei. Inserting Equation A.3 into Equation 3.17, we get

$$\frac{dR}{d\omega} = \frac{1}{m_N}\frac{\rho_\chi}{m_\chi}\frac{2\pi^2\overline{\sigma}_n}{\mu_{N\chi}^2}\int d^3v f_\chi(\vec{v})\int\frac{d^3q}{(2\pi)^3}\mathcal{F}_{\text{med}}^2(q)\frac{f_N^2}{f_n^2}F_N^2(\vec{q})\delta\left(\omega-\frac{q^2}{2m_N}\right)\delta(\omega-\omega_{\vec{q}}), \tag{A.4}$$

which is a slightly modified reproduction of Equation 97 from Trickle et al. (2020). Equation A.4 is used for comparison to the phonon production calculation in the following section.

Single phonons

For the production of a single phonon, the initial and final states take the following form:

$$|i\rangle = |0\rangle, \quad |f\rangle = |\nu,\vec{k}\rangle = \hat{a}^\dagger_{\nu,\vec{k}}|0\rangle, \tag{A.5}$$

where phonons are labeled by a branch index ν and a momentum vector $\vec{k}$ within the first Brillouin zone. There are three branch indices for each atom or ion in a primitive cell. The creation and annihilation operators $\hat{a}^\dagger$ and $\hat{a}$ follow the canonical commutation relations.

In the case of single phonon excitations arising from purely nucleon couplings, the target form factor picks up a phase factor that depends on the displacements $\vec{u}_l$ of the lattice sites relative to their equilibrium positions $\vec{x}_l^0$:

$$\mathcal{F}_T(\vec{q}) = \sum_l \frac{f_N}{f_n}F_N(\vec{q})e^{i\vec{q}\cdot(\vec{x}_l^0+\vec{u}_l)} \tag{A.6}$$

where l sums over all the primitive cells. We also assume that there is only one atom in the primitive cell; otherwise, there would be an additional sum over each atom or ion in the primitive cell.

The critical next step in evaluating $\langle f|\mathcal{F}_T|i\rangle$ involves expressing the lattice displacements in terms of phonon creation and annihilation operators:

$$\vec{u}_l = \sum_\nu\sum_{\vec{k}\in 1\text{BZ}}\frac{1}{\sqrt{2Nm_N\omega_{\nu,\vec{k}}}}\left(\hat{a}_{\nu,\vec{k}}\vec{\epsilon}_{\nu,\vec{k}}e^{i\vec{k}\cdot\vec{x}_l^0}+\hat{a}^\dagger_{\nu,\vec{k}}\vec{\epsilon}^*_{\nu,\vec{k}}e^{-i\vec{k}\cdot\vec{x}_l^0}\right), \tag{A.7}$$

where $\vec{\epsilon}$ are unit-normalized polarization vectors, also labeled by ν and $\vec{k}$, and N is the number of primitive unit cells, which in this case is just the number of nucleons. In the limit where $N\to\infty$, $\vec{k}$ becomes continuous. To evaluate $\exp(i\vec{q}\cdot(\vec{x}_l^0+\vec{u}_l))$, we must invoke the Baker-Campbell-Hausdorff formula $e^{X+Y}=e^Xe^Ye^{-\frac{1}{2}[X,Y]}\cdots$ for non-commuting operators X and Y ($\hat{a}$ and $\hat{a}^\dagger$ in this case).

Again skipping the details of the calculation, we reproduce the final result for the dynamic structure factor, Equation 90 from Trickle et al. (2020), simplified in terms of this specific example:

$$S(\vec{q},\omega) = \frac{\pi}{\Omega}e^{-2W(\vec{q})}\frac{f_N^2}{f_n^2}F_N^2(\vec{q})\sum_\nu\frac{\left|\vec{q}\cdot\vec{\epsilon}^*_{\nu,\vec{k}}\right|^2}{m_N\omega_{\nu,\vec{k}}}\delta(\omega-\omega_{\nu,\vec{k}}), \tag{A.8}$$

where Ω is the volume of the primitive cell, analogous to ρ_T/m_N from Equation A.3. The sums over $\vec{k}\in 1\text{BZ}$ and l from the previous equation conspire to enforce $\vec{q}=\vec{k}+\vec{G}$ in Equation A.3, where $\vec{G}$ is the reciprocal lattice vector in the first Brillouin zone. The expression in the exponent $W(\vec{q})$ is known as the Debye-Waller factor and arises from the commutation term of the Baker-Campbell-Hausdorff expansion (the only commutation term that survives the expansion). Without producing the full

expression from Trickle et al. (2020), we report that $W(\vec{q}) \sim q^2/4m_N\omega$. It is also notable that only the conjugated polarization vector $\vec{\epsilon}^*$ survives the calculation because the un-conjugated vector is attached to the annihilation operator. We insert Equation A.8 into Equation 3.17 to get the energy spectrum:

$$\frac{dR}{d\omega} = \frac{1}{m_N}\frac{\rho_\chi}{m_\chi}\frac{\pi^2\overline{\sigma}_n}{\mu_{n\chi}^2}\int d^3v f_\chi(\vec{v})\int \frac{d^3q}{(2\pi)^3}\mathcal{F}_{\rm med}(q)e^{-2W(\vec{q})}\frac{f_N^2}{f_n^2}F_N^2(\vec{q})\sum_\nu \frac{\left|\vec{q}\cdot\vec{\epsilon}^*_{\nu,\vec{k}}\right|^2}{m_N\omega_{\nu,\vec{k}}}\delta(\omega-\omega_{\nu,\vec{k}})\delta(\omega-\omega_{\vec{q}}). \tag{A.9}$$

Appendix B

USRP noise timestream procedure

Procedure for taking noise data with an Ettus USRP using `GPU_SDR`

1. Before using the USRP, use a dedicated VNA to obtain rough values for resonant frequencies and quality factors as well as a sense for the overall transmission of the readout chain.
2. Choose appropriate LOs, cleaning tones, and calibration tones for each resonator under study.
   - Currently, LOs must be rounded to the nearest 50 MHz. There is 160 MHz bandwidth around the LO.
   - Cleaning tones should be separated by at least a few MHz.
   - Calibration tones at a minimum should be taken roughly at $f_r \pm f_r/5Q_r$. More off-resonance tones will improve later conversion of readout noise into the resonator basis.
3. Measure the cable delay for a particular LO. Presently, this step is automated within the USRP code.
4. Measure $S_{21}(f)$ with a VNA scan over how ever many resonators are under study. The USRP will automatically correct for the measured cable delay.
   - It is important to make sure that the VNA scan is over an appropriate frequency range and has appropriate frequency binning. A roughly sufficient frequency range in `python 3` syntax may be `numpy.linspace(fr-5*fr/Qr,fr+5*fr/Qr,1000)`, where `fr` and `Qr` are estimates for the f_r and Q_r, so that there are 100 points between $f_r \pm f_r/2Q_r$, which charts out half of the resonance circle, and so that most of the full resonance circle is visible. For example, if $f_r = 5$ GHz and $Q_r = 10^5$, then the prescribed frequency bins would be spaced by 500 Hz. A larger range beyond [`fr-5*fr/Qr`, `fr+5*fr/Qr`] may be desired if well-measured off-resonance transmission is desired, but appropriate frequency binning between $f_r \pm f_r/2Q_r$ should be maintained.
   - The appropriate amount of averaging or measurement time per frequency point (otherwise known as the IF bandwidth) can be determined empirically but in principle can be estimated by the amount of noise that is anticipated for a given readout chain and a particular resonator's RF responsivity.
5. Fit for the resonant frequencies f_r and quality factors Q_r with Equation 6.30
6. Measure $S_{21}(t;\, f_r)$ at each resonator's f_r. The length of the timestream and the sampling rate are the key parameters to control here. The USRP can stream up to 200 MHz depending on GPU performance. Frequency information beyond 1 MHz is generally not needed for current KIPM detector designs, so decimation down to 1 MHz is somewhat of a default setting. The length of the timestream as well as the desired amount of averaging in Fourier space will

control the lowest frequency bin available to a Fourier transform. If there is no averaging in frequency space, the lowest frequency bin is simply the 1/(length of timestream).

7. It is generally desired to perform steps 3–5 at range of readout powers so as to provide enough context for these measurements. As demonstrated in Section 7.8, amplifier white noise is expected to dominate at low readout powers and would become sub-dominant to TLS noise at some elevated readout power since the white noise contribution to resonator basis noise will go as P_g^{-1} and the TLS noise contribution in the resonator basis will go as $P_g^{-1/2}$.
8. For calibration of $S_{21}(t;\, f_r)$ into quasiparticle units, a temperature sweep of f_r is required. See Section 6.6.

Appendix C

KID debugging

I don't see a resonator! What do I do?

There are a few sanity checks to perform before concluding that the resonator is simply not alive. These checks also apply to resonators whose Q_i are simply lower than expected.

- Is the transmission about what you expect? When operating an RF readout system, it is helpful to have an expectation for what the throughput of the readout chain will be, by accounting for the intentional attenuation placed on the input line, the expected gains of the output amplifiers from HEMTs or warm amplifiers, and whatever components in between, including cables between temperature stages, filters, and warm SMA cables. If the transmission is not what you expect, then there may be a broken component that has developed during or before the cooldown.
- Are you confident that the temperature on the chip is as cold as the mixing chamber thermometer says it is? Before any cooldown, it is of course important to ensure that all elements that are attached to the mixing chamber are thermally well sunk. Some generally good practices for thermal sinking are listed here.
  - Use both lock washers and flat washers between every screw-surface interface. The lock washer is important to maintain tension even when the components undergo thermal contraction. The flat washer is important to spread the pressure from the screw over a larger area.
  - Gold plate or passivate with citric acid all components that are intended to be thermally linked at the mixing chamber to promote thermal conductivity. Beware of nickel under-plating; nickel is magnetic and would be deleterious toward the performance of superconducting resonators.
  - Copper braids can be used as a last-minute fix for thermal linking.
- Could the resonator be loaded in some other way?
  - If there is an optical fiber mounted toward the KIPM detector, it is possible for blackbody radiation to travel down the optical fiber and radiate pair-breaking photons into the resonator. The lower the superconducting gap of the material of the resonator, the more sensitive the resonator is expected to be to these photons.
  - Parasitic loading along the RF transmission line is also possible. HEMTs or other amplifiers may radiate pair-breaking radiation backwards along the line; isolators, in-line IR filters, or generic RF low pass filters are potential ways to mitigate this radiation.

I don't see pulses! What do I do?

Here are some things to consider when you have found a resonator but are struggling to see pulses in a KIPM detector.

- Are you sure that light is entering the fridge? Check that the LED or laser is indeed shining light by flashing it at your palm
- Are the pulses being obscured by noise? If the sampling rate is in great excess of the signal band high frequency white noise, the small signals we seek may not visible. A low-pass filter or some rolling averaging on the timestreams could also remove this noise. There may also be low-frequency phase noise from the USRP that is obscuring these timestreams and cannot be removed by averaging. At lower readout powers, this noise is not so bad. What might not be visible at 1 MHz bandwidth could be visible post-LPF/averaging/cleaning.
- Are you using a wet dilution refrigerator? In the Cahill B101 Oxford Kelvinox 25, we observed a dependence of pulse height on time from the start of the cooldown. Essentially, we discovered that superfluid helium films would form around the surface of the detector and degrade its performance. In principle, the sorb should be able to remove all the helium in the inner vacuum chamber. We decided to add an additional pumping step to the cooling process, ideally between 4 K and 10 K and targeted a leak rate below around 10^{-8} mbar $\cdot$ liter/s to ensure adequate evacuation of helium from the inner vacuum chamber.

Bibliography

Abazajian, Kevork (Mar. 2006). "Linear cosmological structure limits on warm dark matter." In: *Physical Review D* 73 (6), p. 063513. DOI: 10.1103/PhysRevD.73.063513. URL: https://link.aps.org/doi/10.1103/PhysRevD.73.063513.

Abazajian, Kevork, George M. Fuller, and Wallace H. Tucker (Dec. 2001). "Direct detection of warm dark matter in the X-Ray." In: *The Astrophysical Journal* 562.2, p. 593. DOI: 10.1086/323867. URL: https://dx.doi.org/10.1086/323867.

Anczarski, Jadyn et al. (Feb. 2024). "Two-stage cryogenic HEMT-based amplifier for low-temperature detectors." en. In: *Journal of Low Temperature Physics* 214.3-4, pp. 256–262.

Anthony-Petersen, Robin et al. (July 2024). "A stress-induced source of phonon bursts and quasiparticle poisoning." In: *Nature Communications* 15.1, p. 6444.

APS DPF (2021). *Proceedings of the 2021 US community study on the future of particle physics.* URL: https://web.archive.org/web/20240928073619/https://www.slac.stanford.edu/econf/C210711/ (visited on 09/28/2024).

Aralis, Taylor (2024). "SuperCDMS SNOLAB, HVeV Run 3, and development of KIPM detectors." URL: https://resolver.caltech.edu/CaltechTHESIS:11272023-225643701.

Battaglieri, Marco et al. (July 2017). "US Cosmic Visions: New Ideas in Dark Matter 2017: Community Report." In: *arXiv e-prints*, arXiv:1707.04591, arXiv:1707.04591. arXiv: 1707.04591 [hep-ph].

Baumann, Daniel (2022). *Cosmology.* Cambridge University Press.

Campbell-Deem, Brian et al. (Aug. 2022). "Dark matter direct detection from the single phonon to the nuclear recoil regime." In: *Physical Review D* 106 (3), p. 036019. DOI: 10.1103/PhysRevD.106.036019. URL: https://link.aps.org/doi/10.1103/PhysRevD.106.036019.

Cardani, Laura et al. (May 2018). "Al/Ti/Al phonon-mediated KIDs for UV–vis light detection over large areas." In: *Superconductor Science and Technology* 31.7, p. 075002. DOI: 10.1088/1361-6668/aac1d4. URL: https://dx.doi.org/10.1088/1361-6668/aac1d4.

Chandler, Charles E. et al. (1990). "Sub-electron noise charge-coupled devices." In: *Charge-Coupled Devices and Solid State Optical Sensors.* Ed. by Morley M. Blouke. Vol. 1242. International Society for Optics and Photonics. SPIE, pp. 238–251. DOI: 10.1117/12.19457. URL: https://doi.org/10.1117/12.19457.

Chang, Yen-Yung (2023). "SuperCDMS HVeV Run 2 low-mass dark matter search, highly multiplexed phonon-mediated particle detector with kinetic inductance detector, and the blackbody radiation in cryogenic experiments." URL: https://resolver.caltech.edu/CaltechTHESIS:05212023-152144355.

Clowe, Douglas et al. (Aug. 2006). "A direct empirical proof of the existence of dark matter." In: *The Astrophysical Journal* 648.2, p. L109. DOI: 10.1086/508162. URL: https://dx.doi.org/10.1086/508162.

Collaboration, DarkSide-50 (June 2023). "Light dark matter search with DarkSide-50." In: *57th Rencontres de Moriond on Electroweak Interactions and Unified Theories*. arXiv: 2306.12151 [hep-ex].

Cornell, Brett Durcan (2018). "A dark matter search using the final SuperCDMS Soudan dataset and the development of a large-format, highly-multiplexed, athermal-phonon-mediated particle detector." URL: https://resolver.caltech.edu/CaltechTHESIS:06082018-171453167.

Croft, Rupert A. C. et al. (Dec. 2002). "Toward a precise measurement of matter clustering: Lyα forest data at redshifts 2-4." In: *The Astrophysical Journal* 581.1, p. 20. DOI: 10.1086/344099. URL: https://dx.doi.org/10.1086/344099.

DAMIC-M Collaboration (Apr. 2023). "First constraints from DAMIC-M on sub-GeV dark-matter particles interacting with electrons." In: *Physical Review Letters* 130 (17), p. 171003. DOI: 10.1103/PhysRevLett.130.171003. URL: https://link.aps.org/doi/10.1103/PhysRevLett.130.171003.

— (2024). *Detector*. URL: https://web.archive.org/web/20241005121736/https://damic.uchicago.edu/detector.php (visited on 10/05/2024).

DarkSide-50 Collaboration (Mar. 2018). "DarkSide-20k: A 20 tonne two-phase LAr TPC for direct dark matter detection at LNGS." en. In: *European Physical Journal Plus* 133.3.

Day, Peter K. et al. (Oct. 2024). "A 25-micrometer single-photon-sensitive kinetic inductance detector." In: *Physical Review X* 14 (4), p. 041005. DOI: 10.1103/PhysRevX.14.041005. URL: https://link.aps.org/doi/10.1103/PhysRevX.14.041005.

de Visser, Pieter J. (2014). "Quasiparticle dynamics in aluminium superconducting microwave resonators." PhD thesis. Delft University of Technology.

de Visser, Pieter J. et al. (Apr. 2011). "Number fluctuations of sparse quasiparticles in a superconductor." In: *Physical Review Letters* 106 (16), p. 167004. DOI: 10.1103/PhysRevLett.106.167004. URL: https://link.aps.org/doi/10.1103/PhysRevLett.106.167004.

Duan, Ran et al. (July 2010). "An open-source readout for MKIDs." In: *Millimeter, submillimeter, and far-infrared detectors and instrumentation for astronomy V*. Ed. by Wayne S. Holland and Jonas Zmuidzinas. Vol. 7741. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 77411V, p. 77411V. DOI: 10.1117/12.856832.

Essig, Rouven et al. (May 2016). "Direct detection of sub-GeV dark matter with semiconductor targets." In: *Journal of High Energy Physics (Online)* 2016.5. DOI: 10.1007/JHEP05(2016)046.

EXCESS workshop (2022). "EXCESS workshop: descriptions of rising low-energy spectra." In: *SciPost Phys. Proc.*, p. 001. DOI: 10.21468/SciPostPhysProc.9.001. URL: https://scipost.org/10.21468/SciPostPhysProc.9.001.

Faoro, Lara and Lev B. Ioffe (Jan. 2015). "Interacting tunneling model for two-level systems in amorphous materials and its predictions for their dephasing and noise in superconducting microresonators." In: *Physical Review B* 91 (1), p. 014201. DOI: 10.1103/PhysRevB.91.014201. URL: https://link.aps.org/doi/10.1103/PhysRevB.91.014201.

Faramarzi, Farzad et al. (July 2024). "A 4–8 GHz kinetic inductance traveling-wave parametric amplifier using four-wave mixing with near quantum-limited noise performance." In: *APL Quantum* 1.3, p. 036107. ISSN: 2835-0103. DOI: 10.1063/5.0208110. eprint: https://pubs.aip.org/aip/apq/article-pdf/doi/10.1063/5.0208110/20040436/036107_1_5.0208110.pdf. URL: https://doi.org/10.1063/5.0208110.

Fink, Caleb W. et al. (Nov. 2024). "Superconducting quasiparticle-amplifying transmon: a qubit-based sensor for meV-scale phonons and single terahertz photons." In: *Physical Review Applied* 22 (5), p. 054009. DOI: 10.1103/PhysRevApplied.22.054009. URL: https://link.aps.org/doi/10.1103/PhysRevApplied.22.054009.

Gao, Jiansong (2008). "The physics of superconducting microwave resonators." URL: https://resolver.caltech.edu/CaltechETD:etd-06092008-235549.

Golwala, Sunil (2000). "Exclusion limits on the WIMP nucleon elastic scattering cross-section from the Cryogenic Dark Matter Search." PhD thesis. UC, Berkeley. DOI: 10.2172/1421437.

— (Feb. 2016). "SuperCDMS SNOLAB: goals, design, and status." In: *Sources and detection of dark matter and dark energy in the universe*. UCLA Dark Matter. Los Angeles, CA.

Golwala, Sunil and Enectali Figueroa-Feliciano (2022). "Novel quantum sensors for light dark matter and neutrino detection." In: *Annual Review of Nuclear and Particle Science* 72.Volume 72, 2022, pp. 419–446. ISSN: 1545-4134. DOI: https://doi.org/10.1146/annurev-nucl-102020-112133. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-nucl-102020-112133.

Griffin, Sinéad M., Katherine Inzani, et al. (Mar. 2020). "Multichannel direct detection of light dark matter: target comparison." In: *Physical Review D* 101 (5), p. 055004. DOI: 10.1103/PhysRevD.101.055004. URL: https://link.aps.org/doi/10.1103/PhysRevD.101.055004.

Griffin, Sinéad M., Simon Knapen, et al. (Dec. 2018). "Directional detection of light dark matter with polar materials." In: *Physical Review D* 98 (11), p. 115034. DOI: 10.1103/PhysRevD.98.115034. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.115034.

Guruswamy, Tejas, David J. Goldie, and Stafford Withington (Mar. 2014). "Quasiparticle generation efficiency in superconducting thin films." In: *Superconductor Science and Technology* 27.5, p. 055012. DOI: 10.1088/0953-2048/27/5/055012. URL: https://dx.doi.org/10.1088/0953-2048/27/5/055012.

Hall, Lawrence J. et al. (Mar. 2010). "Freeze-in production of FIMP dark matter." In: *Journal of High Energy Physics* 2010.3, p. 80. ISSN: 1029-8479. DOI: 10.1007/JHEP03(2010)080. URL: https://doi.org/10.1007/JHEP03(2010)080.

Helm, Richard H. (Dec. 1956). "Inelastic and elastic scattering of 187-MeV electrons from selected even-even nuclei." In: *Physical Review* 104 (5), pp. 1466–1475. DOI: 10.1103/PhysRev.104.1466. URL: https://link.aps.org/doi/10.1103/PhysRev.104.1466.

Hochberg, Yonit, Yonatan Kahn, Noah Kurinsky, et al. (Oct. 2021). "Determining dark-matter–electron scattering rates from the dielectric function." In: *Physical Review Letters* 127 (15), p. 151802. DOI: 10.1103/PhysRevLett.127.151802. URL: https://link.aps.org/doi/10.1103/PhysRevLett.127.151802.

Hochberg, Yonit, Yonatan Kahn, Mariangela Lisanti, et al. (Jan. 2018a). "Detection of sub-MeV dark matter with three-dimensional Dirac materials." In: *Physical Review D* 97 (1), p. 015004. DOI: 10.1103/PhysRevD.97.015004. URL: https://link.aps.org/doi/10.1103/PhysRevD.97.015004.

— (Jan. 2018b). "Detection of sub-MeV dark matter with three-dimensional Dirac materials." In: *Physical Review D* 97 (1), p. 015004. DOI: 10.1103/PhysRevD.97.015004. URL: https://link.aps.org/doi/10.1103/PhysRevD.97.015004.

Hochberg, Yonit, Matt Pyle, et al. (Aug. 2016). "Detecting superlight dark matter with Fermi-degenerate materials." en. In: *Journal of High Energy Physics* 2016.8.

Hochberg, Yonit, Yue Zhao, and Kathryn M. Zurek (Jan. 2016). "Superconducting detectors for superlight dark matter". In: *Physical Review Letters* 116 (1), p. 011301. DOI: 10.1103/PhysRevLett.116.011301. URL: https://link.aps.org/doi/10.1103/PhysRevLett.116.011301.

Holmström, Eero., Antti Kuronen, and Kai Nordlund (July 2008). "Threshold defect production in silicon determined by density functional theory molecular dynamics simulations." In: *Physical Review B* 78 (4), p. 045202. DOI: 10.1103/PhysRevB.78.045202. URL: https://link.aps.org/doi/10.1103/PhysRevB.78.045202.

Kaplan, Steven B. et al. (Dec. 1976). "Quasiparticle and phonon lifetimes in superconductors." In: *Physical Review B* 14 (11), pp. 4854–4873. DOI: 10.1103/PhysRevB.14.4854. URL: https://link.aps.org/doi/10.1103/PhysRevB.14.4854.

Khalil, Moe S. et al. (Mar. 2012). "An analysis method for asymmetric resonator transmission applied to superconducting devices." In: *Journal of Applied Physics* 111.5, p. 054510. ISSN: 0021-8979. DOI: 10.1063/1.3692073. eprint: https://pubs.aip.org/aip/jap/article-pdf/doi/10.1063/1.3692073/15093308/054510_1_online.pdf. URL: https://doi.org/10.1063/1.3692073.

Knapen, Simon, Tongyan Lin, Matt Pyle, et al. (2018). "Detection of light dark matter with optical phonons in polar materials." In: *Physics Letters B* 785, pp. 386–390. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2018.08.064. URL: https://www.sciencedirect.com/science/article/pii/S0370269318306816.

Knapen, Simon, Tongyan Lin, and Kathryn M. Zurek (Dec. 2017). "Light dark matter: models and constraints." In: *Physical Review D* 96 (11), p. 115021. DOI: 10.1103/PhysRevD.96.115021. URL: https://link.aps.org/doi/10.1103/PhysRevD.96.115021.

Kuflik, Eric et al. (June 2016). "Elastically decoupling dark matter." In: *Physical Review Letters* 116 (22), p. 221302. DOI: 10.1103/PhysRevLett.116.221302. URL: https://link.aps.org/doi/10.1103/PhysRevLett.116.221302.

— (Aug. 2017). "Phenomenology of ELDER dark matter." In: *Journal of High Energy Physics* 2017.8, p. 78. ISSN: 1029-8479. DOI: 10.1007/JHEP08(2017)078. URL: https://doi.org/10.1007/JHEP08(2017)078.

Lewin, J. David and Peter F. Smith (1996). "Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil." In: *Astroparticle Physics* 6.1, 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.

Li, Xinran (Sept. 2024). "Low-Tc phonon-sensitive hafnium microwave kinetic inductance detectors development." Poster presentation at the Applied Superconductivity Conference 2024.

Lisenfeld, Jürgen et al. (Feb. 2015). "Observation of directly interacting coherent two-level systems in an amorphous material." In: *Nature Communications* 6.1. ISSN: 2041-1723. DOI: 10.1038/ncomms7182. URL: http://dx.doi.org/10.1038/ncomms7182.

Luke, Paul N. (Dec. 1988). "Voltage-assisted calorimetric ionization detector." In: *Journal of Applied Physics* 64.12, pp. 6858–6860. ISSN: 0021-8979. DOI: 10.1063/1.341976. eprint: https://pubs.aip.org/aip/jap/article-pdf/64/12/6858/18621786/6858_1_online.pdf. URL: https://doi.org/10.1063/1.341976.

LZ Collaboration (July 2023a). "Background determination for the LUX-ZEPLIN dark matter experiment." In: *Physical Review D* 108 (1), p. 012010. DOI: 10.1103/PhysRevD.108.012010. URL: https://link.aps.org/doi/10.1103/PhysRevD.108.012010.

— (July 2023b). "First dark matter search results from the LUX-ZEPLIN (LZ) experiment." In: *Physical Review Letters* 131 (4), p. 041002. DOI: 10.1103/PhysRevLett.131.041002. URL: https://link.aps.org/doi/10.1103/PhysRevLett.131.041002.

Martinez, Maria et al. (June 2019). "Measurements and simulations of athermal phonon transmission from silicon absorbers to aluminum sensors." In: *Physical Review Appl.* 11 (6), p. 064025. DOI: 10.1103/PhysRevApplied.11.064025. URL: https://link.aps.org/doi/10.1103/PhysRevApplied.11.064025.

Mattis, Daniel C. and John Bardeen (July 1958). "Theory of the anomalous skin effect in normal and superconducting metals." In: *Physical Review* 111 (2), pp. 412–417. DOI: 10.1103/PhysRev.111.412. URL: https://link.aps.org/doi/10.1103/PhysRev.111.412.

McRae, Corey Rae H. et al. (Sept. 2020). "Materials loss measurements using superconducting microwave resonators." In: *Review of Scientific Instruments* 91.9, p. 091101. ISSN: 0034-6748. DOI: 10.1063/5.0017378. eprint: https://pubs.aip.org/aip/rsi/article-pdf/doi/10.1063/5.0017378/19786221/091101_1_online.pdf. URL: https://doi.org/10.1063/5.0017378.

Minutolo, Lorenzo et al. (2019). "A flexible GPU-accelerated radio-frequency readout for superconducting detectors." In: *IEEE Transactions on Applied Superconductivity* 29.5, pp. 1–5. DOI: 10.1109/TASC.2019.2912027.

Moore, David C. (June 2012). "A search for low-mass dark matter with the cryogenic dark matter search and the development of highly multiplexed phonon-mediated particle detectors." URL: https://resolver.caltech.edu/CaltechTHESIS:05172012-142723949.

Moore, David C. et al. (June 2012). "Position and energy-resolved particle detection using phonon-mediated microwave kinetic inductance detectors." In: *Applied Physics Letters* 100.23, p. 232601. ISSN: 0003-6951. DOI: 10.1063/1.4726279. eprint: https://pubs.aip.org/aip/apl/article-pdf/doi/10.1063/1.4726279/14247421/232601_1_online.pdf. URL: https://doi.org/10.1063/1.4726279.

Navarro, Julio F., Carlos S. Frenk, and Simon D. M. White (Dec. 1997). "A universal density profile from hierarchical clustering." In: *Astrophys. J.* 490.2, pp. 493–508. DOI: 10.1086/304888. arXiv: astro-ph/9611107 [astro-ph].

Neganov, Boris S. and V. N. Trofimov (1985). "Colorimetric method measuring ionizing radiation." In: *Otkryt Izobret* 146, p. 215.

Olling, Rob P. and Michael R. Merrifield (Jan. 2000). "Two measures of the shape of the dark halo of the Milky Way." In: *Monthly Notices of the Royal Astronomical Society* 311.2, pp. 361–369. DOI: 10.1046/j.1365-8711.2000.03053.x. arXiv: astro-ph/9907353 [astro-ph].

Particle Data Group (2014). "Review of particle physics." In: *Chinese Physics C* 2014.38, p. 090001. eprint: https://pdg.lbl.gov/2015/download/rpp2014-Chin.Phys.C.38.090001.pdf.

Phillips, William Andrew (Dec. 1987). "Two-level states in glasses." In: *Reports on Progress in Physics* 50.12, p. 1657. DOI: 10.1088/0034-4885/50/12/003. URL: https://dx.doi.org/10.1088/0034-4885/50/12/003.

Planck Collaboration (2020). "Planck 2018 results - VI. Cosmological parameter." In: *Astronomy and Astrophysics* 641, A6. DOI: 10.1051/0004-6361/201833910. URL: https://doi.org/10.1051/0004-6361/201833910.

Pozar, David M. (2005). *Microwave engineering; 3rd ed.* Hoboken, NJ: Wiley. URL: https://cds.cern.ch/record/882338.

Pyle, Matt (2012). "Optimizing the design and analysis of cryogenic semiconductor dark matter detectors for maximum sensitivity." PhD thesis. Stanford University.

Ramanathan, Karthik, Taylor Aralis, et al. (Nov. 2022). "Identifying drivers of energy resolution variation in a multi-KID phonon-mediated detector." In: *Journal of Low Temperature Physics* 209.3, pp. 457–463. DOI: 10.1007/s10909-022-02753-5.

Ramanathan, Karthik, John E. Parker, et al. (2024). *Quantum parity detectors: a qubit based particle detection scheme with meV thresholds for rare-event searches.* arXiv: 2405.17192 [physics.ins-det]. URL: https://arxiv.org/abs/2405.17192.

Ramanathan, Karthik, Osmond Wen, et al. (2024). *Significant noise improvement in a kinetic inductance phonon-mediated detector by use of a wideband parametric amplifier.* DOI: 10.48550/arXiv.2402.05419.

Ren, Runze et al. (Aug. 2021). "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), p. 032010. DOI: 10.1103/PhysRevD.104.032010. URL: https://link.aps.org/doi/10.1103/PhysRevD.104.032010.

Reynolds, Tyler (2020). *Long-term planning sensitivities.* URL: https://confluence.slac.stanford.edu/display/CDMS/Long-Term+Planning+Sensitivities (visited on 12/16/2024).

Ricochet Collaboration (Feb. 2024). "First demonstration of 30 eVee ionization energy resolution with Ricochet germanium cryogenic bolometers." en. In: *European Physical Journal C Particles and Fields* 84.2, p. 186.

Robinson, Alan E. (Jan. 2017). "Coherent photon scattering background in sub-GeV/c^2 direct dark matter searches." In: *Physical Review D* 95 (2), p. 021301. DOI: 10.1103/PhysRevD.95.021301. URL: https://link.aps.org/doi/10.1103/PhysRevD.95.021301.

Romani, Roger K. et al. (2024). *A transition edge sensor operated in coincidence with a high sensitivity phonon veto for photon coupled rare event searches.* arXiv: 2408.11158 [physics.ins-det]. URL: https://arxiv.org/abs/2408.11158.

Rubin, Vera, Norbert Thonnard, and William Kent Ford Jr. (1980). "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/." In: *The Astrophysical Journal* 238, p. 471. DOI: 10.1086/158003.

Sarkis, Y., Alexis Aguilar-Arevalo, and Juan Carlos D'Olivo (May 2020). "Study of the ionization efficiency for nuclear recoils in pure crystals." In: *Physical Review D* 101 (10), p. 102001. DOI: 10.1103/PhysRevD.101.102001. URL: https://link.aps.org/doi/10.1103/PhysRevD.101.102001.

Schottky, Walter (1918). "Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern." In: *Annalen der Physik* 362.23, pp. 541–567. DOI: https://doi.org/10.1002/andp.19183622304. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19183622304. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19183622304.

SENSEI Collaboration (2023). *SENSEI: first direct-detection results on sub-GeV dark matter from SENSEI at SNOLAB.* arXiv: 2312.13342 [astro-ph.CO]. URL: https://arxiv.org/abs/2312.13342.

Siegel, Seth R. (2016). "A multiwavelength study of the intracluster medium and the characterization of the multiwavelength sub/millimeter inductance camera". URL: https://resolver.caltech.edu/CaltechTHESIS:10212015-211417853.

Sikivie, Pierre (2003). "Evidence for ring caustics in the Milky Way." In: *Physics Letters B* 567.1, pp. 1–8. ISSN: 0370-2693. DOI: https://doi.org/10.1016/S0370-2693(03)00863-3. URL: https://www.sciencedirect.com/science/article/pii/S0370269303008633.

Spahn, Gabriel C. (2021). "Improving performance of a multiplexed dark matter detector via infrared isolation." Bachelor's Thesis. University of Minnesota.

SuperCDMS Collaboration (Dec. 2010). "Low-threshold analysis of CDMS shallow-site data." In: *Physical Review D* 82 (12), p. 122004. DOI: 10.1103/PhysRevD.82.122004. URL: https://link.aps.org/doi/10.1103/PhysRevD.82.122004.

— (Apr. 2017). "Projected sensitivity of the SuperCDMS SNOLAB experiment." In: *Physical Review D* 95 (8), p. 082002. DOI: 10.1103/PhysRevD.95.082002. URL: https://link.aps.org/doi/10.1103/PhysRevD.95.082002.

— (Aug. 2018). "Energy loss due to defect formation from 206Pb recoils in SuperCDMS germanium detectors." In: *Applied Physics Letters* 113.9, p. 092101. ISSN: 0003-6951. DOI: 10.1063/1.5041457. eprint: https://pubs.aip.org/aip/apl/article-pdf/doi/10.

1063/1.5041457/13581912/092101_1_online.pdf. URL: https://doi.org/10.1063/1.5041457.

SuperCDMS Collaboration (Nov. 2020). "Constraints on low-mass, relic dark matter candidates from a surface-operated SuperCDMS single-charge sensitive detector." In: *Physical Review D* 102 (9), p. 091101. DOI: 10.1103/PhysRevD.102.091101. URL: https://link.aps.org/doi/10.1103/PhysRevD.102.091101.

— (Aug. 2021). "Light dark matter search with a high-resolution athermal phonon detector operated above ground." In: *Physical Review Letters* 127 (6), p. 061801. DOI: 10.1103/PhysRevLett.127.061801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.127.061801.

— (June 2022a). "Investigating the sources of low-energy events in a SuperCDMS-HVeV detector." In: *Physical Review D* 105 (11), p. 112006. DOI: 10.1103/PhysRevD.105.112006. URL: https://link.aps.org/doi/10.1103/PhysRevD.105.112006.

— (June 2022b). "Ionization yield measurement in a germanium CDMSlite detector using photo-neutron sources." In: *Physical Review D* 105 (12), p. 122002. DOI: 10.1103/PhysRevD.105.122002.

— (2023). *A strategy for low-mass dark matter searches with cryogenic detectors in the SuperCDMS SNOLAB facility*. DOI: 10.48550/arXiv.2203.08463.

SuperCDMS SNOLAB TDR (2018). *SuperCDMS SNOLAB technical design report*. Tech. rep. SuperCDMS Collaboration.

Temples, Dylan J. et al. (Oct. 2024). "Performance of a phonon-mediated kinetic inductance detector at the NEXUS cryogenic facility." In: *Physical Review Applied* 22 (4), p. 044045. DOI: 10.1103/PhysRevApplied.22.044045.

TESSERACT Collaboration (2024). *Low energy backgrounds and excess noise in a two-channel low-threshold calorimeter*. arXiv: 2410.16510 [physics.ins-det]. URL: https://arxiv.org/abs/2410.16510.

Tiffenberg, Javier et al. (Sept. 2017). "Single-electron and single-photon sensitivity with a silicon skipper CCD." In: *Physical Review Letters* 119 (13), p. 131802. DOI: 10.1103/PhysRevLett.119.131802. URL: https://link.aps.org/doi/10.1103/PhysRevLett.119.131802.

Tinkham, Michael (2004). *Introduction to superconductivity*. Dover Books on Physics Series. Dover Publications. ISBN: 9780486134727. URL: https://books.google.com/books?id=VpUk3NfwDIkC.

Trickle, Tanner et al. (Mar. 2020). "Multi-channel direct detection of light dark matter: theoretical framework." en. In: *Journal High Energy Physics* 2020.3.

Viel, Matteo, Martin G. Haehnelt, and Volker Springel (Nov. 2004). "Inferring the dark matter power spectrum from the Lyman α forest in high-resolution QSO absorption spectra." In: *Monthly Notices of the Royal Astronomical Society* 354.3, pp. 684–694. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2004.08224.x. eprint: https://academic.oup.com/mnras/article-pdf/354/3/684/18653214/354-3-684.pdf. URL: https://doi.org/10.1111/j.1365-2966.2004.08224.x.

Wen, Osmond et al. (Nov. 2022). "Performance of a phonon-mediated detector using KIDs optimized for Sub-GeV dark matter." In: *Journal of Low Temperature Physics* 209.3, pp. 510–517. DOI: 10.1007/s10909-022-02764-2.

Wilson, Christopher M. and Daniel E. Prober (Mar. 2004). "Quasiparticle number fluctuations in superconductors." In: *Physical Review B* 69 (9), p. 094524. DOI: 10.1103/PhysRevB.69.094524. URL: https://link.aps.org/doi/10.1103/PhysRevB.69.094524.

XENONnT Collaboration (2024). *Time projection chamber*. URL: https://web.archive.org/web/20240912205053/https://xenonexperiment.org/time-projection-chamber/ (visited on 09/12/2024).

Yellin, Steve (Aug. 2002). "Finding an upper limit in the presence of an unknown background." In: *Physical Review D* 66 (3), p. 032005. DOI: 10.1103/PhysRevD.66.032005. URL: https://link.aps.org/doi/10.1103/PhysRevD.66.032005.

— (2007). *Extending the optimum interval method*. arXiv: 0709.2701 [physics.data-an]. URL: https://arxiv.org/abs/0709.2701.

Zmuidzinas, Jonas (2012). "Superconducting microresonators: physics and applications." In: *Annual Review of Condensed Matter Physics* 3.Volume 3, 2012, pp. 169–214. ISSN: 1947-5462. DOI: https://doi.org/10.1146/annurev-conmatphys-020911-125022. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-conmatphys-020911-125022.

Zobrist, Nicholas et al. (2021). "Improving the dynamic range of single photon counting kinetic inductance detectors." In: *Journal of Astronomical Telescopes, Instruments, and Systems* 7.1, p. 010501. DOI: 10.1117/1.JATIS.7.1.010501. URL: https://doi.org/10.1117/1.JATIS.7.1.010501.

Zurek, Kathryn M. (2014). "Asymmetric dark matter: theories, signatures, and constraints." In: *Physics Reports* 537.3. Asymmetric Dark Matter: Theories, signatures, and constraints, pp. 91–121. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2013.12.001. URL: http://www.sciencedirect.com/science/article/pii/S0370157313004341.

— (2024). "Dark matter candidates of a very low mass." In: *Annual Review of Nuclear and Particle Science* 74.Volume 74, 2024, pp. 287–319. ISSN: 1545-4134. DOI: https://doi.org/10.1146/annurev-nucl-101918-023542. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-nucl-101918-023542.

Zwicky, Fritz (Jan. 1933). "Die rotverschiebung von extragalaktischen nebeln." In: *Helvetica Physica Acta* 6, pp. 110–127.

— (1937). "On the masses of nebulae and of clusters of nebulae." In: *The Astrophysical Journal* 86, p. 217. DOI: 10.1086/143864.