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

Thesis by

Yen-Yung Chang

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

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
Defended May 12th, 2023

Yen-Yung Chang
ORCID: 0000-0002-6441-980X

ACKNOWLEDGEMENTS

This thesis, and all the works done during the last eight years, could not have been achieved without the supports from many people. There are too many people to thank, so I apologize if I fail to mention your name in the following. I thank you for your great help for the completion of this thesis and my PhD.

First and foremost, I thank my PhD advisor, Dr. Sunil Gowala. Your support and trust for students truly made my life at Caltech one of the most memorable and enjoyable times I have ever had. I truly appreciate the freedom I was granted to explore and grow as a scientist. I enjoyed a lot the mentoring relation in the group based on fulfilling every member's potential and passion. I realize this is not always the case for other research groups under the pressure of producing immediate results. For this, I sincerely thank you for such a valuable opportunity to learn and work with you.

For the works introduced in this thesis, and many more that are not mentioned, I thank Taylor Aralis, Bruce Bumble, Karthik Ramanathan, and Osmond Wen, for being absolutely essential to the achievements. Taylor, nothing in this thesis could have been done without you. I am always quick and sloppy. I produce ideas and tools that are always only half working, but I know I can always count on you to make them actually useful. Bruce, it has always been my dream to work at NASA, which could not have come true without your guidance. I feel I was just trying to absorb as much that you know or have achieved in the last thirty years in the unforgettable two fast years. Karthik, what a cool coincidence that we met so many years ago on the UChicago visit day, went on our separate paths, and then the paths met again here at Caltech. Your passion toward novel ideas while pursuing them with care and practicality is truly inspiring. Osmond, I cannot thank you more for bringing the KID work that we started years ago to so much more that it is now. I will never forget seeing the cartoon of phonon spreading in the wafer for the first. It could not have been realized without all your great efforts.

I thank friends in SuperCDMS, in particular Matthew Hollister, Ziqing Hong, Noah Kurinsky, Valentina Novati, Runze "Tom" Ren, and Matthew Wilson, for your valuable intellectual supports as well as fond memories. Matt (H.), I am truly inspired by your spirit of combining science and engineering to "make things work," which is always my self expectation as an experimental physicist. Although we did not overlap long at Caltech, you continued to provide great help for our cryogenic system, which was absolutely essential for everything achieved in this thesis. Noah, you are one of the people I look up to the most as a scientist. Not just for all the close collaboration on projects, your critical thinking and suggestions are always guaranteed to be critical problem solvers. Tom, Valentina, Ziqing, our HVeV R2 results could not have been achieved without you. You made the 1.5 year memory during and after our Northwestern time truly lively. I learned a lot from you in research, but truly the most important thing I learned from you is your diverse approaches and positive attitudes toward the science goal we share. Matt (W.), I would never forget the late night in UMontreal student hotel lobby, where you said you were determined to make the code work to generate limits for our talk tomorrow. In addition to the same inspiration from Ziqing, Valentina, and Tom, I thank you for being the best research partner and friend in the R2 team.

Finally, I would like to thank everyone from Caltech ACT, especially Allen, Bo-Hsuan, Chien-I, Chien-Ying, Hsiao-Yi, MJ, Lucas, Shih-Hao, Yen-Chun, Yu-An, Yun-Ting, Sean, Steve (and my apology if I missed anyone). You are like my families and provided immense joy to make my life at Caltech complete.

I am forever in debt with my families. Without your unconditional love and support, I could not have accomplished anything in life. I thank my wife and our son, for understanding and sharing all the joys and tears in the precious life we share. It has not been an easy journey leaving your old lives to move to the US with me, but in the end your love made every memory full of joy. I thank my brother and sister. You are the reason I fight through difficulties. Mom and dad, even though we cannot FaceTime when I am weak and struggling, you continue giving me strength and love through challenges in life. I cannot wait to share with you all the stories when we meet again. Dad, you had never tried to make me a physicist like you, but I am sure you are proud of me now, not because I still ended up becoming a physicist, but because I am happy. Thank you.

- Yen-Yung, May 18$^{\text{th}}$, 2023.

ABSTRACT

There is ample evidence of dark matter (DM), a phenomenon responsible for ≈85% of the matter content of the Universe that cannot be explained by the Standard Model (SM). One of the most compelling hypotheses is that DM consists of beyond-SM particle(s) that are nonluminous and nonbaryonic. So far, numerous efforts have been made to search for particle DM, and yet none has yielded an unambiguous observation of DM particles.

We present in Chapter 2 the SuperCDMS HVeV Run 2 experiment, where we search for DM in the mass ranges of 0.5–10^4 MeV/c^2 for the electron-recoil DM and 1.2–50 eV/c^2 for the dark photon and the Axion-like particle (ALP). SuperCDMS utilizes cryogenic crystals as detectors to search for DM interaction with the crystal atoms. The interaction is detected in the form of recoil energy mediated by phonons. In the HVeV project, we look for electron recoil, where we enhance the signal by the Neganov-Trofimov-Luke effect under high-voltage biases. The technique enabled us to detect quantized e^-h^+ creation at a 3% ionization energy resolution. Our work is the first DM search analysis considering charge trapping and impact ionization effects for solid-state detectors. We report our results as upper limits for the assumed particle models as functions of DM mass. Our results exclude the DM-electron scattering cross section, the dark photon kinetic mixing parameter, and the ALP axioelectric coupling above 8.4×10^{-34} cm^2, 3.3×10^{-14}, and 1.0×10^{-9}, respectively.

Currently every SuperCDMS detector is equipped with a few phonon sensors based on the transition-edge sensor (TES) technology. In order to improve phonon-mediated particle detectors' background rejection performance, we are developing highly multiplexed detectors utilizing kinetic inductance detectors (KIDs) as phonon sensors. This work is detailed in chapter 3. and chapter 4. We have improved our previous KID and readout line designs, which enabled us to produce our first ⌀3" detector with 80 phonon sensors. The detector yielded a frequency placement accuracy of 0.07%, indicating our capability of implementing hundreds of phonon sensors in a typical SuperCDMS-style detector. We detail our fabrication technique for simultaneously employing Al and Nb for the KID circuit. We explain our signal model that includes extracting the RF signal, calibrating the RF signal into pair-breaking energy, and then the pulse detection. We summarize our noise condition and develop models for different noise sources. We combine the signal and the noise

models to be an energy resolution model for KID-based phonon-mediated detectors. From this model, we propose strategies to further improve future detectors' energy resolution and introduce our ongoing implementations.

Blackbody (BB) radiation is one of the plausible background sources responsible for the low-energy background currently preventing low-threshold DM experiments to search for lower DM mass ranges. In Chapter 5, we present our study for such background for cryogenic experiments. We have developed physical models and, based on the models, simulation tools for BB radiation propagation as photons or waves. We have also developed a theoretical model for BB photons' interaction with semiconductor impurities, which is one of the possible channels for generating the leakage current background in SuperCDMS-style detectors. We have planned for an experiment to calibrate our simulation and leakage current generation model. For the experiment, we have developed a specialized "mesh TES" photon detector inspired by cosmic microwave background experiments. We present its sensitivity model, the radiation source developed for the calibration, and the general plan of the experiment.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

There is ample evidence of dark matter (DM), a phenomenon responsible for ≈85% of the matter content of the Universe that can not be explained by the Standard Model (SM). One of the most compelling hypotheses is DM consists of beyond-SM particle(s) that are nonluminous and nonbaryonic. In this chapter, we first review the evidence of DM and the candidate particle models. We then introduce ongoing or proposed experiments for detecting particle DM and their status.

1.1 Dark matter

1.1.1 Dark matter halo

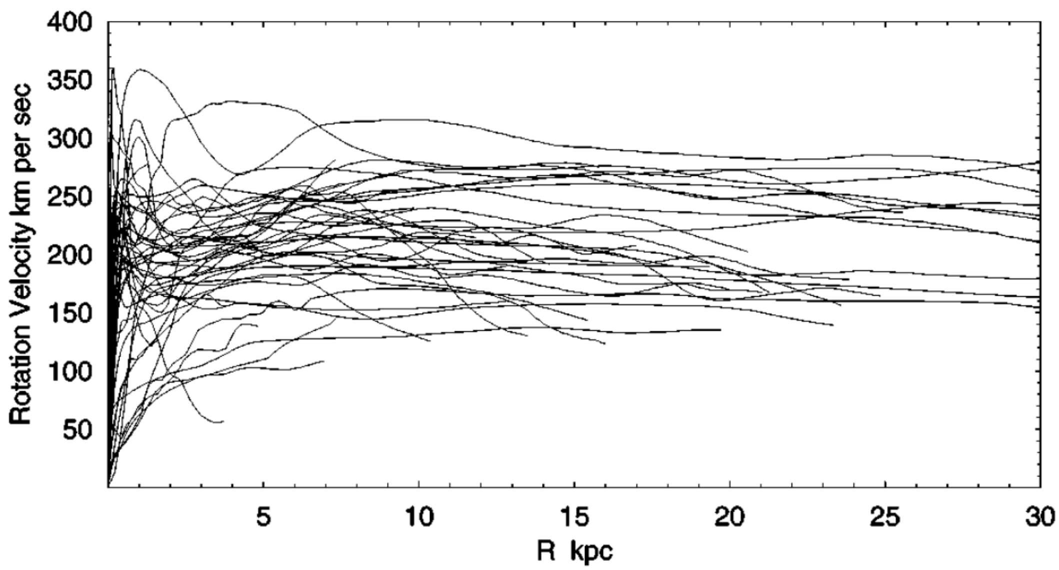


Figure 1.1: The rotational curves of spiral galaxies. The inner-, middle (visible-light outskirt)- and outer (no visible light)-region data were acquired via CO, visible-light, and HI observations, respectively. Figure reproduced from (Sofue and Rubin, 2001) with more detailed information found therein.

Fig. 1.1 presents the classical "evidence" of dark matter (DM), the rotational curves of galaxies. It presents the orbiting velocities of matter versus their distances from the galactic centers. In modern presentations, the rotational curves are typically acquired up to a few times of the visible-light radii of the galaxies, which we now recognize as the full extent of galaxies. These type of data were first acquired in the late 19^{th} to the early 20^{th} century utilizing visible-light telescopes to establish the early recognition that, if assuming Newtonian gravity, the total masses contained in the galaxies must be significantly larger than which inferred from the visible star light in order to support their observed orbiting velocities (Bucklin, 2017). In particular, the non-vanishing high velocities in the visible-light outskirts of galaxies as shown in Fig. 1.1 apparently violate the expectation. The rational hypotheses therefore include, one, large fractions of masses consisting the galaxies were unobservable in the visible light or, two, if the observed mass distributions truly reflected the galaxies' mass distributions, the classical mechanics was then incomplete in describing these systems. While some science historians argue that the term "dark matter" had already been utilized in the literature based on its literal meaning of the "unobservable mass," it is now commonly accepted that F. Zwicky first used the term in its modern meaning that describes distributed "halos" of unobservable masses coexisting with the galaxies (Zwicky, 1933). These massive DM halos resolve the missing mass problem by accompanying with the observable baryonic matter but are inert to electromagnetism (EM) either directly or indirectly through other Standard Model (SM) interactions. In the following decades with the development of the field of radio astronomy, (Rubin et al., 1980) conducted the now-recognized first comprehensive study on 21 spiral galaxies and resolved the possibility that the missing mass problem was due to under-observations outside the visible light. The work established the knowledge of the common existence of the DM halo in every galaxy, representing not just a sufficient mass to modify its rotational curve but the majority of the total mass of the galaxy.

Even though DM does not participate in SM interactions, with DM halo's significant mass contribution to the galaxy construction, it is anticipated to play a critical role in every system's dynamics and evolution via pure gravity. As a consequence, this massive DM gas has left fingerprints in almost every system in the Universe, from the very condensed modern state to the very uniform primordial state right after the big bang. Due to the expansion of the Universe, these modern observations for different systems also correspond to different times in the cosmological evolution and also different length scales from as local as each galaxy to the entire Universe. Among the ample evidences of DM, we select in this chapter iconic examples that span the astronomical length scale spectrum to illustrate the critical nature of DM. For a more comprehensive introduction, we refer the readers to (Tanabashi et al., 2018) and the references therein.

Having introduced DM halo's massive (gravitational) but non-SM galaxy "by-stander" attachment characteristic, we expand the length scale from single galaxies to clusters of galaxies. Unlike the rotational curves, where one can obtain detailed profiles of the DM mass distributions with the help of the baryonic matter, the thinning of the inter-galactic gases and their less well-defined velocity fields make the DM profiling more difficult to be performed similarly to the rotational curves. Nevertheless, thanks to galaxy clusters' large extents and total masses, one can still profile their DM mass distributions with the distorted background light traveling through the massive regions, a technique called weak lensing. Overall, weak lensing observations are consistent with the galactic DM halo picture, in that the galaxy clusters also possess DM halos that are much heavier than their baryonic counterparts. In fact, instead of regarding the DM halo as an attachment to the observable matter, the current best theory that is supported by many-body simulations is that, during the evolution of the Universe, DM condensed gravitationally prior to the baryons due to lacking photon pressure sensitivity like the baryons and then formed the heavy DM halos that in turn gravitationally attracted the baryonic matter into their potential wells. The baryons that were funneled into the condensed DM regions then formed galaxies and compact objects inside. The scenario naturally explains the inevitable pairing of the DM components as the backbones for the formations of the observable astronomical objects. This DM-assisted process has strong consequences for the cosmological history, especially in concentrating materials in the vast empty Universe to form the elements and then the stars and galaxies.

Fig. 1.2 shows a rare incident where two galaxy clusters, both carrying their observable and dark components, collide with each other, a peculiar phenomenon called the Bullet Cluster; the figure is the aftermath picture of the collision. By profiling the baryonic and the DM distributions via optical/X-ray observations and weak lensing, respectively (Markevitch et al., 2002; Clowe et al., 2004), the figure clearly shows that the two baryonic components collided into each other like normal fluid with interaction, leading to two "dragged" splashes toward the outgoing directions. On the contrary, the weak lensing result indicates that the two globular DM halos simply passed each other without any apparent deformation. The result shows, in addition to be non-interacting with SM matter, DM also does not interact with itself at an observable strength (Markevitch et al., 2004). We have summarized with the above examples the cold DM (CDM) nature of DM, where "cold" indicates that DM exhibits an equivalent thermal momentum that is close to the thermalized materials in the Universe, therefore it may condense in the early Universe to form the potential

Figure 1.2: The 1E 0657-56 Bullet Cluster. The blue and the pink regions are artificially colored to represent the mass distributions of DM and baryonic matter, respectively. Their densities are determined by weak lensing and optical/X-ray observations, respectively. The X-ray and the optical/lensing observations are from (Markevitch et al., 2002) and (Clowe et al., 2004), respectively, with the DM self-interaction constraint presented in (Markevitch et al., 2004).

wells that guided the structure formation. CDM is also scatter-less, not just with the SM matter for being "dark" but also with itself, as indicated by the Bullet Cluster example.

1.1.2 Structure formation

We further expand the length scale in discussion to larger structures as seen today, which is, since the Universe is expanding, equivalent to tracing the early physics in the cosmological history further backward in time. Fig. 1.3 shows the recent result by the Dark Energy Survey (DES) (S. Hong et al., 2021), where the fluid CDM distribution is profiled by the modern[1] galaxies in the Local Group. The result shows that DM forms the web-like backbone structure that guides the concentration of baryonic matter into forming galaxies along the so-called Cosmic Web. Due to the expansion and the cooling of the Universe, the structures would still form gravitationally in the absence of the DM. However, in this case the structure formation

[1] To avoid confusion, providing the speed of light and the depth of the observation, the data represent the *modern* universe, not a direct ancient image, but do pose physical implications related to the historical time when the structure was formed/started to form.

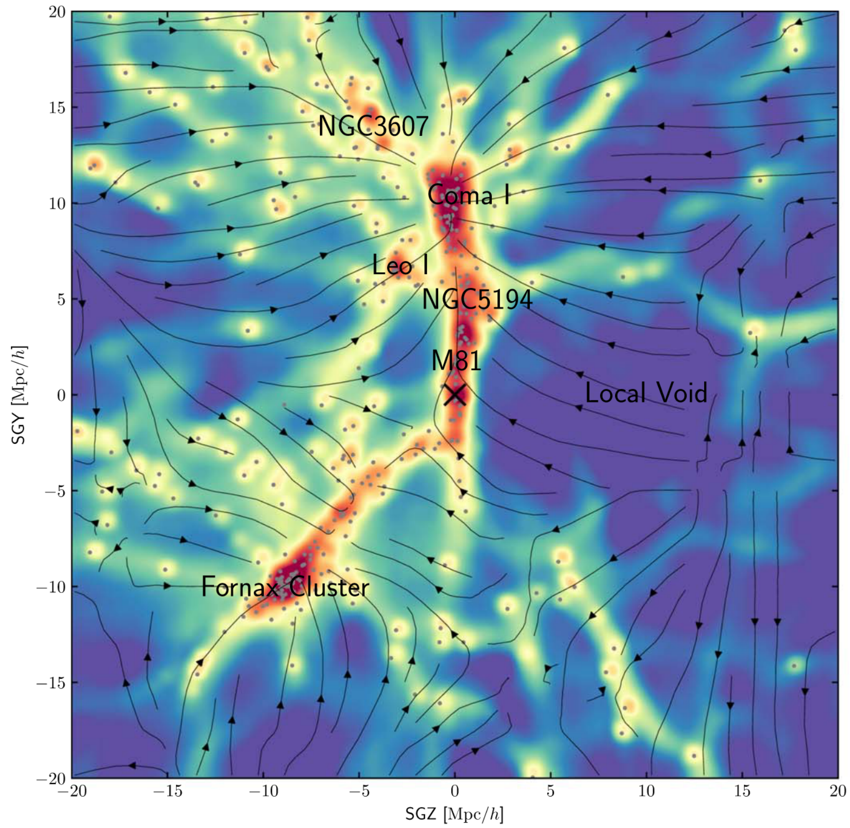


Figure 1.3: The Cosmic Web for the Local Group, with the × symbol at the center marking the Milky Way galaxy. The rainbow colors represent, from red to blue, high to low DM densities, respectively, and the arrows indicate the matter velocity field inferred from the mass distribution. Figure reproduced from (S. Hong et al., 2021), with elaborations found in the corresponding text and the original reference.

is expected to be much slower without the assistance of the DM potential wells. Therefore, by observing the structural concentration as in Fig. 1.3, the data only favor a limited range for the cosmological parameters that participate the evolution of the structures, such as the age and the expansion rate of the Universe, the ratio of the normal and the dark matter, the cooling behavior of DM that depends on the particle physics details such as the particle interactions and the densities. However, since DES or other local structure observations all represent the modern Universe, which requires nontrivial assumptions to be linked to the entire cosmological history, the practical approach for obtaining the constraints is typically by comparing these modern-Universe data to the time-evolving many-body simulations that have all the intermediate steps built-in for phenomenological adjustments.

1.1.3 Cosmic microwave background

First predicted by (Gamow, 1948; Alpher and Herman, 1948) and subsequently confirmed through the discovery of the cosmic microwave background (CMB) by (Penzias and R. W. Wilson, 1965), we now know that everything in the Universe was at one time in an "isotropic" (same in all directions) and "homogeneous" (same from all locations) thermal equilibrium. The thermal radiation field due to this primordial matter distribution gave rise to the CMB when the photons decoupled from constantly scattering with the primordial plasma matter. With the Universe continued expanding and cooling, this ancient thermal radiation field eventually cooled to become the modern CMB we observe today. Aside from all the successes in general cosmology, CMB's major implication for DM is that the CMB represents the distribution of the matter coupled to the CMB photons at the moment they decoupled, which occurred at the known energy scale for ionizing hydrogen atoms and therefore at a calculable time in the cooling history of the Universe. At the moment of the decoupling, free electrons in the primordial plasma ceased receiving sufficient energy from the photon bath to remain in the plasma state, so the electrons quickly attached to the protons in the plasma to form neutral hydrogen atoms. From the perspective of the photons, free charged particles that were in thermal equilibrium with them via constant photon absorption suddenly all disappeared, therefore this density "snapshot" for the suddenly freed photons maps the primordial plasma density (Aghanim et al., 2020).

Besides being similar in constraining the matter distribution through observing luminous objects as for the rotational curve, the weak lensing, and the galaxy survey, CMB is the least uncertain subject in terms of propagating the signal generation to the observation. Thanks to its generation in the much more uniform early Universe, the application of CMB only requires the assumption of a homogeneous, isotropic, Hubble-expanding universe as described by the Friedmann equation (Friedmann, 1999). Fig. 1.4 is the so-called CMB power spectrum that presents the two-point intensity correlation strength of the CMB intensity map. It shows that the CMB field clearly favors certain length scales at a periodicity as indicated by the periodic spectral peaks. The fact that the power spectrum exhibits a periodic structure indicates that the primordial plasma rippled on an acoustic pressure wave like the water in a pond, where the plasma "liquid" is the water and the entire Universe is the pond. This acoustic rippling is the baryon acoustic oscillation (BAO) in the early plasma that embeds the information of the primordial plasma's pressure wave propagation properties. With the CMB power spectrum measured to such a high

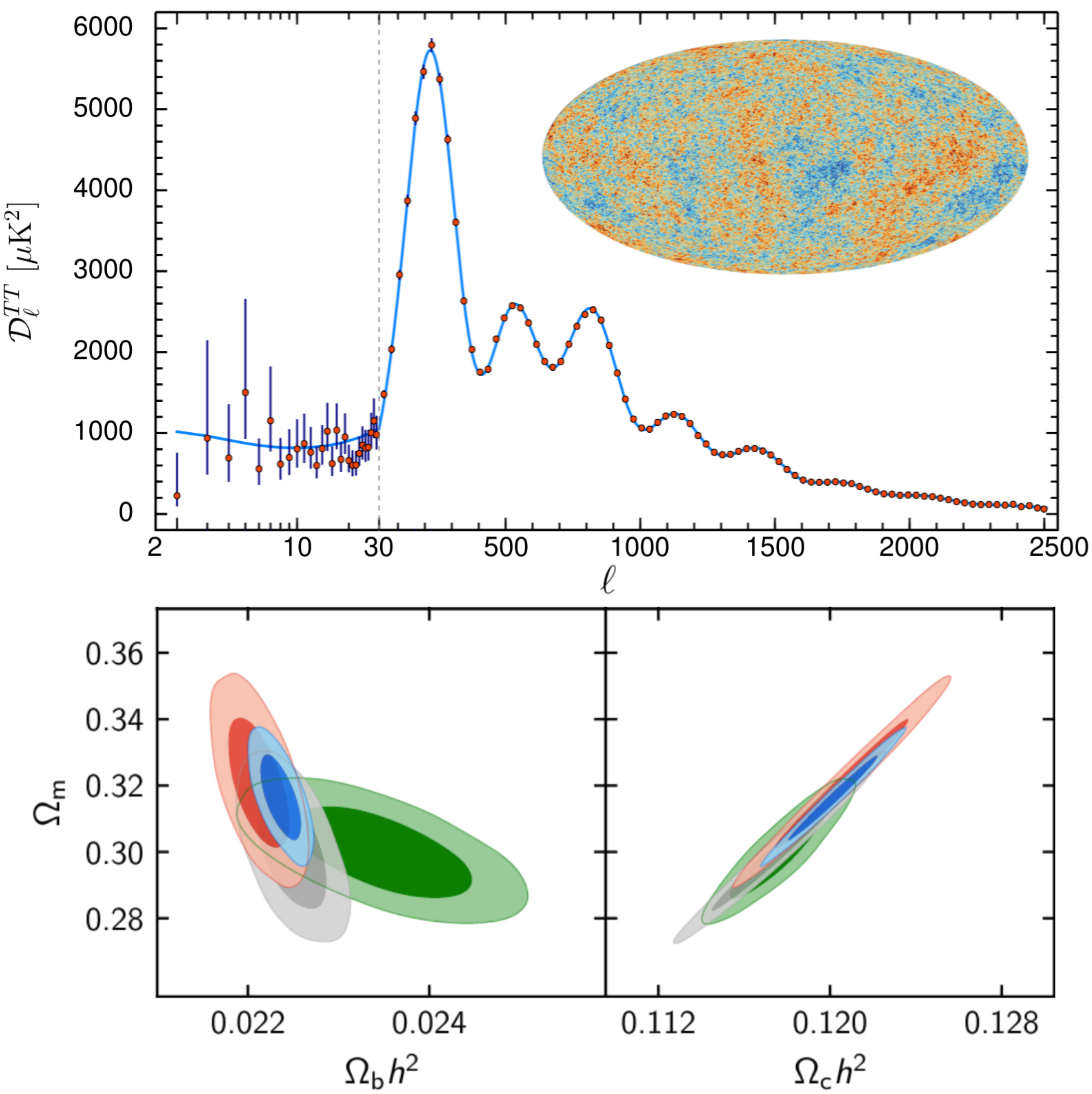


Figure 1.4: Top: The CMB temperature power spectrum and an example data map (inset) reproduced from (Aghanim et al., 2020). Bottom: The Λ-CDM parameter constraints obtained in (Aghanim et al., 2020), presented by the total mass (Ω_m) against the baryonic (Ω_b) and the CDM (Ω_c) components. The ellipses with different colors represent the constraints obtained from the datasets of different observation frequencies, with darker colors representing lower confidence levels. Detailed description for the figure may be found in (Aghanim et al., 2020).

precision, Fig. 1.4's acoustic structure provides an effective tool to constrain the total primordial plasma mass Ω_{m} against the baryon and the DM fractions Ω_{b} and Ω_{c}, respectively; the former is constrained by the absolute strength of the power spectrum, and the latter is most strongly constrained by the relative heights of the acoustic peaks.

The extraction of the above parameters are subjected to the assumptions for how the baryon, the photon, and the DM scatter with each other as well as with themselves. Therefore, one has to constrain these interactions simultaneously based on more detailed particle model assumptions. However, as of today, the simplest perfectly scatter-less CDM assumption remains fully consistent with the latest CMB measurements (Aghanim et al., 2020). Finally, to propagate the CMB from its generation to the modern-Universe observation, one also needs a nontrivial dark energy contribution as described in the Friedmann equation (Friedmann, 1999), which yields the composition shown in Fig. 1.5: The massive energy content of our Universe only represents about 30% of the total energy content of the Universe, while the rest is mostly represented by the dark energy, which is brought out as an allowed integration constant for the Einstein's equation but lacking a consensus physical interpretation. In modern cosmology, it is conventional to follow Einstein's original notation Λ for the dark energy "cosmological constant," so the complete model is named the Λ-CDM model, which is currently the best model of cosmology as of today.

1.1.4 Particle models and constraints

Fig. 1.6 summarizes the currently allowed particle model space for DM. Overall, due to the current non-detection status, the particle mass space for DM is opened to the entire mass range that is only bounded by astronomical observations, spanning from 10^{-22} eV to 10^{14} eV.[2] Nevertheless, depending on the assumed natures of different particle models, any one class of particles cannot satisfy all the existing astronomical constraints across the entire 36 orders-of-magnitude, therefore the models summarized in Fig. 1.6 are seen to roughly group by three different categories:

1) Axion/Axion-like particle (ALP) and ultra-light DM. These DM models typically arise from the misalignment mechanism during symmetry breaking and therefore obtain masses that are inversely correlated with the postulated broken symmetries and their corresponding interaction strengths as pseudo-Goldstone Bosons,

[2] Excluding exotic astronomical compact objects.

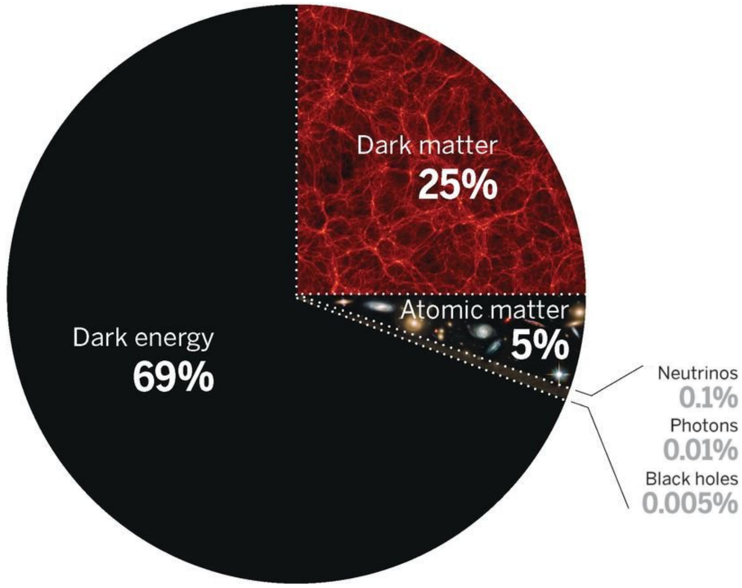


Figure 1.5: The Λ-CDM composition of our Universe. Figure reproduced from (The Daily Galaxy, 2015).

e.g., (Pecci and Quinn, 1977; Weinberg, 1978; Ringwald, 2014). Depending on the size of the misalignment, in principle the Axion/ALP can be unboundedly light and therefore is only bounded in mass by astronomical observations. In particular, currently the most stringent lower bounds are from the extents of the observed DM halos (or any forms of clumps). One may understand these lower bounds as simply forbidding too light a DM particle that exhibits a matter-wave "particle size" larger than the observed DM clumps/halos. Due to this feature, Axion/ALP and similar models are oftentimes loosely referred to as the ultra-light DMs or the wave-like DMs, based on their common characteristic that there must be many particles coexisting within a space of the size of the light particles' large deBroglie wavelength. Under certain conditions, this superposition characteristic can lead to a Bose-Einstein condensate-like coherent DM field similar to a superfluid, e.g., (Sikivie and Yang, 2009).

2) The dark sector, or interchangeably called the hidden/dark sector/valley. This class of models typically describes a standalone structure of particles independent of the SM. Based on the dark sector's parallel existence picture, the DM-built Universe may as well consist a family of different particles or particle bound states analogous to the SM, where the dark particles are allowed to interact strongly within the dark sector yet appear perfectly "dark" to the SM Universe (Spergel and Steinhardt, 2000; D. E. Kaplan et al., 2009). However, knowing that the DM was most likely at one time in a much stronger thermal contact with SM particles than the present condition,

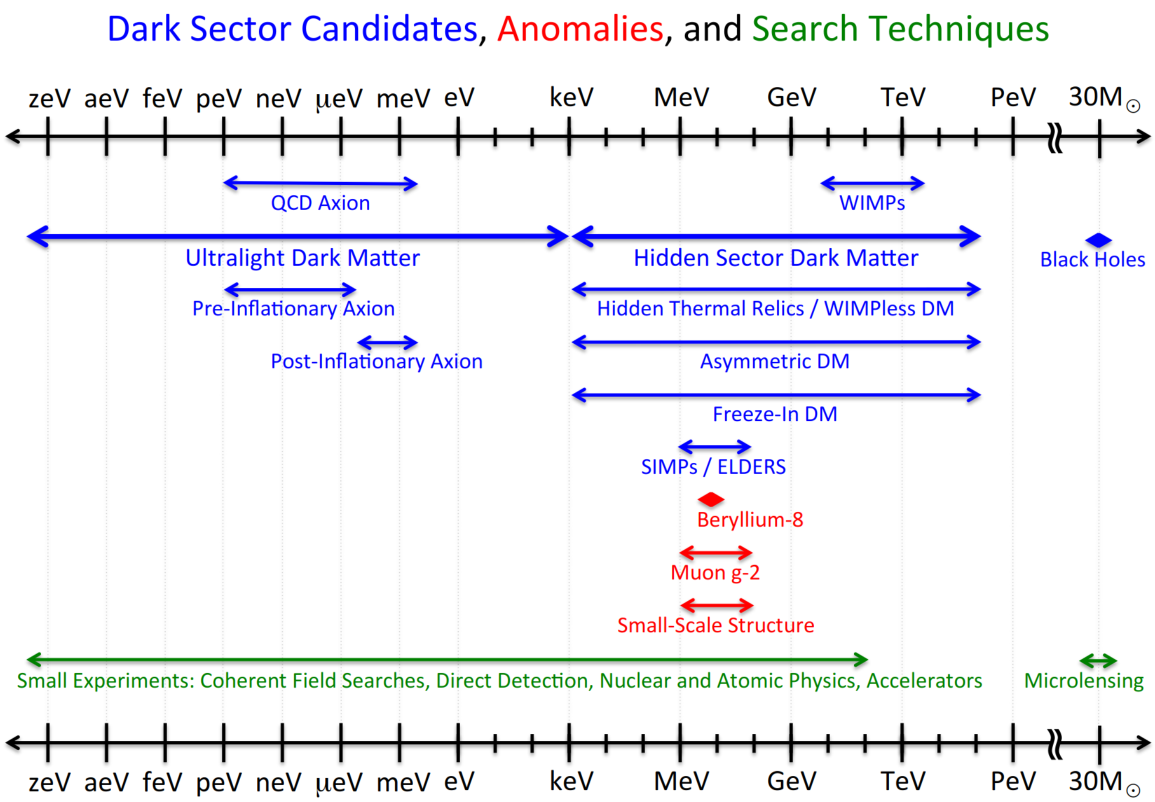


Figure 1.6: The DM model landscape shown by typical mass ranges (blue). Also provided in the figure for comparisons are the persisting experimental anomalies interpreted as the consequences of DM (red), as well as typical detection methods (green) with text locations roughly corresponding to the applicable mass ranges. Figure reproduced from (Battaglieri et al., 2017), with more detailed descriptions found therein.

dark sector models oftentimes associate DM and SM with highly suppressed and yet non-zero interaction strengths, thus still an accessible "valley" in the landscape. Since the interactions are energy dependent, one can derive the thermal history of the Universe based on the expanding Universe's energy density evolution. Among all the possible scenarios, the most iconic one is the so-called thermal freeze-out process, where DM, just like all SM particles that are too heavy to be created at the current-Universe temperature, acquired its frozen relic abundance at the moment the Universe cooled to be unable to annihilate and recreate more of the particle at thermal equilibrium. The exact detail of the thermal decoupling depends on the characteristic mass-energy scales of the models, which in many cases are linked to various observable properties of the Universe, such as the matter-antimatter asymmetry, the bound state energy scales, etc. (Zurek, 2014; Hochberg et al., 2014). It is also possible to postulate mass hierarchies in the models that lead to temporary enhancements of DM creation, i.e., the so-called freeze-in process, or alternating periods of creation, annihilation, and relic abundance determination.

3) Weakly interacting massive particle (WIMP). This class of models generally includes relatively well-formulated extensions of the SM. Historically, the name “weakly” was used to indicate that the extensions were typically applied through the weak portals but with much heavier mediators and thus the suppressed interaction. Since most of such predictions have been stringently constrained, nowadays WIMP has shifted to the meaning of being “weaker than” the weak interaction. The favored types of WIMP are 1) the Super-symmetry (SUSY) and 2) the heavy “Seesaw” neutrinos. The former postulates Boson/Fermion-mirrored particle groups to the SM (Jungman et al., 1996), and the latter attempts to explain the lightness of the SM neutrino by introducing new DM neutrinos that are heavy due to mass-flavor eigenstate rotation, therefore the name “Seesaw” (Weinberg, 1980; Chulia et al., 2018). The Seesaw neutrinos are also called the sterile neutrinos because they are singlets that do not interact with SM particles via the weak force at the tree level. Historically, many iconic WIMP models were proposed to solve the problems other than the DM, such as for the grand unification and the finite Higgs mass by SUSY, or the neutrino mass by the Seesaw mechanism. These models nevertheless predicted the observed DM relic abundance when put in to the perspective of the cosmological thermal history, which added to their plausibility as one of the most believed DM candidates. In contrast with dark sectors, where the model building may be much more creative as long as the astronomical CDM properties are satisfied, WIMP models typically fall in the tighter GeV–TeV range relative to the keV–sub-PeV dark sector range, owing to the heavy mediator coupling and the more strictly formulated and thus constrained theories.

We provide the above classification-style model introduction in light of the persistent non-detection status of particle DM. It is fair to state that currently there is not a single model that is significantly more recognized by the community as the potential true DM model. On the contrary, new DM models are being continuously proposed without violating any empirical constraints due to our limited understanding in the particle nature of DM. Therefore, instead of selecting specific models beyond the grouped introduction above, we discuss in the following well-established and/or -motivated constraints for the particle properties of DM as we know of today, albeit largely derived from astronomical observations. To confine the model landscape as a start, Fig. 1.6 shows that the DM particle mass is bounded from above and below at about $\mathcal{O}(10^2)$ TeV and $\mathcal{O}(10^{-1})$ zeV, respectively. The upper bound was proposed by (Griest and Kamionkowski, 1990) based on the fact that, in order to satisfy the observed relic abundance (Ω_c) and assuming that the DM particles were allowed to efficiently scatter and/or self-annihilate in the early universe, the DM particle mass must not exceed ≈340 TeV (original publication value), otherwise the maximally allowed annihilation partial cross section would become too small and in turn lead to an over-produced relic abundance. Since Griest and Kamionkowski, 1990 calculated the allowed partial cross section from the partial-wave unitarity, this upper bound is now generally recognized as the "unitarity bound" of DM mass. Although the original derivation of (Griest and Kamionkowski, 1990) invoked certain cosmological constant and thermal history assumptions and therefore has been revised by later works, e.g., (Aghanim et al., 2020), the general methodology underlying the unitarity bound is well-accepted, so it is now commonly accepted that the unitarity bound limits the DM mass from above at $\mathcal{O}(10^2)$ TeV. At the opposite end of the mass spectrum are the ultra-light DMs bounded by the sizes of the observed DM halos, where the strongest lower bound for the DM mass is set by young dwarf galaxies with the smallest DM halos ever observed. In these works, the constraints are acquired more delicately by considering the detailed mass functions of the galaxies, which lead to the lower bounds around $\mathcal{O}(10^{-1})$ zeV, e.g., (Calabrese and Spergel, 2016). We note again that, in this mass limit, the particles are light and thus extend to overlap with other DM particles across the entire halo region. This feature prohibits Fermion DM due to not having enough space space in galaxies to accommodate the particles under Fermi exclusion principle. On the contrary, the lightness is natural for the misalignment mechanism that gives rises to Goldstone Bosons, so in Fig. 1.6 the wave-like Axion/ALP DM models occupying this lowest mass region are all ultra-light Bosons.

Between the fairly model-independent upper and lower bounds that are mostly based on observational evidences, there are also intermediate thresholds that are relatively model-dependent but still worth introducing due to their relations with the general features of the above model categories. There are two distinct thresholds in Fig. 1.6, one at ~GeV and the other at ~keV, that distinguish the WIMP from the dark sector models and the dark sector from the wave-like Axion family, respectively. For the WIMP-dark sector threshold, it was first realized by (B. W. Lee and Weinberg, 1977) and thus is generally recognized as the Lee-Weinberg bound that, for any stable DM particles constituting the DM halo that can self-annihilate via the weak portal, specifically the Z Boson in the original publication, the lower mass bound for such DM particles is at ≈2 GeV. Except the model-specific original choice for the Z coupling, the Lee-Weinberg bound is only based on the observed relic abundance and the fact that, for the DM particles to interact via a mediator (Z), the annihilation cross section is proportional to the squared ratio of the DM mass to the mediator (Z) mass. For any DM masses lighter than a few GeV, the suppressed annihilation would result in an over-abundance that violates the observation. The logic underlying the Lee-Weinberg bound is similar to the unitarity bound, both considering simple relations of the masses and the cross sections so to satisfy the relic abundance, so the Lee-Weinberg bound is also generally recognized as a robust threshold for weakly[3] coupled WIMP models.

Contrasting the Lee-Weinberg bound, the ~keV dark sector-ALP threshold is not as generally formulated and therefore typically understood to be in the range shown by Fig. 1.6 but permitting a fuzziness in sub–10s keV. This uncertainty allows detection techniques that are sensitive in the range to simultaneously search for the Axion/ALP and the dark sector DMs, such as the HVeV work of this thesis. Despite the model dependence, the Axion-dark sector threshold typically stems from the assumption that the dark sector particles postulate specific thermal production mechanisms in the early universe, where the DM particles were brought to existence through freeze-out or -in processes from heavier or lighter parents, respectively. The DM relic is then required to be able to cool down during the structure formation as explained and reach the modern Universe condition. The described process strongly depends on model-specific details, especially the freeze-out/in time and the corresponding mass and also the allowed momentum exchange channels. Such dependences lead to model-specific mass lower bounds and uncertainties for different dark sector models. Beyond this transitional mass range, however, the DMs are reliably massive and can cool down for the required structure formation; they are believed to be srtictly "cold" (nonrelativistic). On the other hand, the DMs below the transitional mass range are most likely to be "hot" (highly relativistic) and require mechanisms other than the classical particle-like gravitational condensation to achieve the observed structure formation. The models falling in the sub–10s keV range are therefore frequently referred to as the "warm" DMs. Recalling our earlier introduction to the structure formation studies, currently the best method for constraining the warm DM mass window, hence distinguishing models beyond and below, is by comparing

[3]The original threshold is given by the Z coupling and therefore "weakly" is specifically for weak-interaction coupling. For other "more weakly" coupled WIMP models, the threshold should be adjusted accordingly.

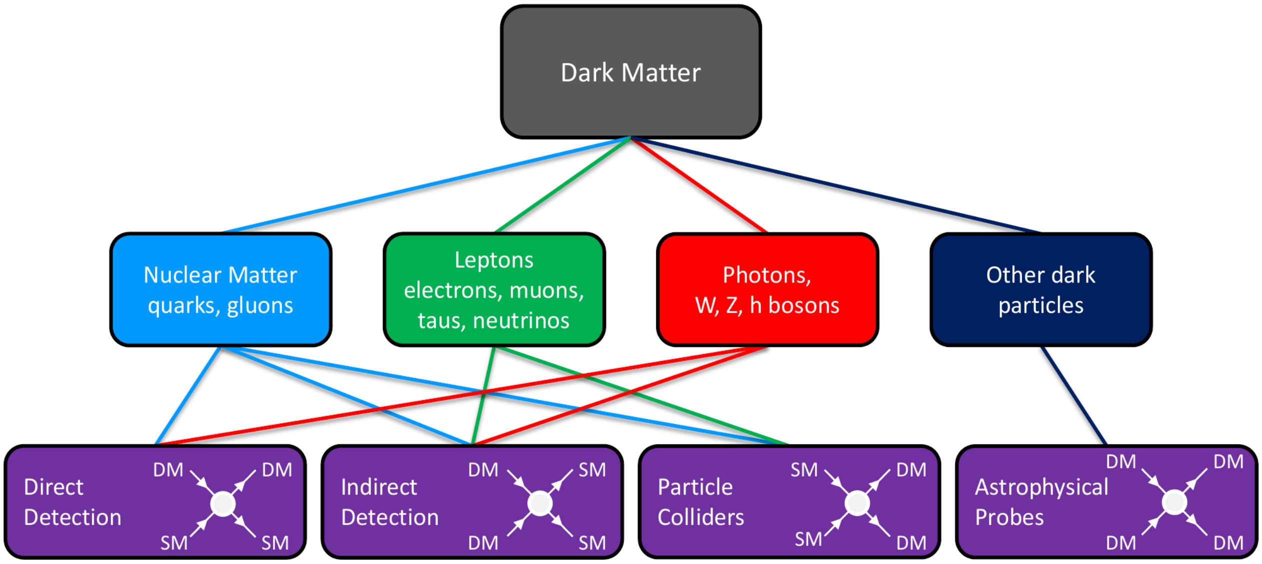


Figure 1.7: A summary of DM detection methods reproduced from (Feng et al., 2014). The middle row shows the theorized interaction channels by colors, and the bottom row are the practical detection techniques with initial- and final-state stable particles shown by the scattering diagrams. The colored lines link the detection methods to the utilizable interactions with the corresponding colors, and the interaction details are loosely represented by the big circles in the scattering diagrams due to diverse model phenomenology. Detailed descriptions may be found in the corresponding text and (Feng et al., 2014).

the observed large structure distributions to the many-body simulations with certain particle model assumptions. In particular, currently the Lyman-α forest observations pose the strongest constraints due to probing the structure formation most deeply into space and time, e.g., Grazilli et. al. (2021), but again the exact placements for the corresponding constraints are nonnegligibly subjected to particle model details.

1.1.5 Experimental detection

Fig. 1.7 summarizes the possible detection channels by their corresponding DM-detector couplings (Feng et al., 2014). Due to model detail variations, the figure specifies the initial- and the final-state SM/DM long-lived particles in the Feynman diagram-style illustrations and represents the intermediate interactions/force mediation by the large circles. Fig. 1.7 categorizes the detection methods into four major types:[4]

1) Astronomical probes, which include the methods that indirectly infer DM's particle properties from DM particles' collective physical consequences in large-scale

[4]To better connect with later contents, we introduce the categories in the figure from right to left.

astronomical objects. All the astronomical evidences for the existence of DM that we have introduced previously generally belong to this category, which are valuable studies in their own rights but nevertheless the least relevant to the particle physics focus of this thesis. Notice that we follow the definition of (Feng et al., 2014) to define only DM-DM self-scattering type of phenomena as shown in Fig. 1.7 as astronomical probes. In practice, other detection techniques that utilize astronomical sources may also be categorized into the other three categories, e.g., cosmic ray-based detection; they will be covered later. In summary, one can understand the detection techniques in this category as observing luminous astronomical objects' behaviors as the consequences of DM particles' momentum exchange properties. The momentum exchange involves the particle mass (gravitational influence), the flux, and the scattering cross section, and thus constraints these properties with the observations. An example is the Bullet Cluster and its conclusion that the DM self-interaction strength is bounded from above. In addition to the large-scale subjects introduced previously, astronomers are also improving the observations to potentially resolve smaller DM structures, targeting yet resolved subjects such as the galactic-center core-cusp and the sub-halo structures, e.g., (deBlok, 2010; Rivero and Dvorkin, 2020).

2) Particle colliders, for which one can understand as attempting to reproduce the early-Universe condition that produces the DM relic and examining the interaction event-by-event. The collider approach is arguably the most informative technique that will immediately provide not just the particle mass but many detailed properties underlying the DM production and interaction. However, such a characteristic also suggests that the collider search is very specific in terms of the theoretical models under investigation, which determine whether the DM is producible in the employed colliding particles/colliders as well as the expected signal reconstruction signatures. If we focus the discussion on only probing the new particles that can constitute the DM halos, in other words long-lived and inert to the particle detectors in realistic colliders, we can exclude those decay-in-detector new physics searches. We then realize that the only practical approach for searching DM in colliders is by collecting all the observable SM particles produced in the collisions so to identify any significant "missing momentum/mass" as the DM signatures, e.g., (Abercrombie et al., 2020; Tenchini et al., 2019). In order to perform this so-called full-reconstruction missing momentum-style searches, the background confusion in different colliders become the key factor for their sensitivities. In particular, the background condition is most strongly determined by the particles utilized for the collisions that produce

DM, which results in the benchmark distinction separating hadron and lepton colliders. Due to hadrons' composite composition and the parton-quark fragmentation, hadron colliders generally yield poorer backgrounds especially toward lower energies, but they are advantageous in raising the collision energy due the hadrons' higher masses; vice versa for lepton colliders. Currently, hadron and lepton collider searches are led by the Large Hadron Collider (LHC, ATLAS/CMS) and the KEK Belle II, respectively, with distinct sensitivity ranges of above $\approx$10 GeV and below $\approx$1 GeV ranges, respectively. Comparing these sensitivity ranges to Fig. 1.6, one may conclude that the LHC and the Belle II missing-momentum searches are close-to-exclusively sensitive to WIMP and dark sector models, respectively. For interested readers, (Abercrombie et al., 2020) and (Campajola et al., 2021) provide comprehensive summaries for the ongoing efforts of LHC and Belle II DM searches.

3) Indirect detection, which encompasses all detection techniques that observe the secondary SM particles produced by DM particles. Comparing the scattering diagrams for the indirect detection and the collider searches, one can loosely regard the indirect detection as DM colliders that annihilate DM particles to produce observable SM daughters in a reversed manner of artificial colliders. However, obviously it is currently impossible to manipulate DM particles as in the colliders to obtain the beamed collisions, so instead the indirect detection is typically achieved by searching for the DM collisions at the locations that one expects DM particles to naturally concentrate. Since there is currently no evidence for DM particles to interact other than gravitationally, we realize that the most reliable practical mechanism to focus DM particles is via gravitational attraction. Therefore, indirect searches mostly aim for observing SM particles that originate from heavy astronomical objects, leveraging their masses for local DM concentrations. The DM signature is therefore the SM particle emissions that are significantly inconsistent with the known radiation mechanisms of the heavy astronomical objects. An inherent limitation for the DM-concentrating technique is that the SM daughter particles, once created inside the astronomical objects, must be able to escape the objects so to be observed. In practice, high-energy γ-ray, proton/anti-proton, light-element nuclei, or neutrinos are the typical candidates for indirect detection. As observatories are built on Earth or its close orbits, local heavy objects including the Sun and the centers of our Milky Way and nearby galaxies have been utilized for such searches (S. K. Lee et al., 2016; Mauro et al., 2019; Aartsen et al., 2013; Abe et al., 2020). Contrasting real colliders whose experimental conditions are carefully understood and controlled, from which so far only limits have been placed for non-observation, astronomical indirect DM

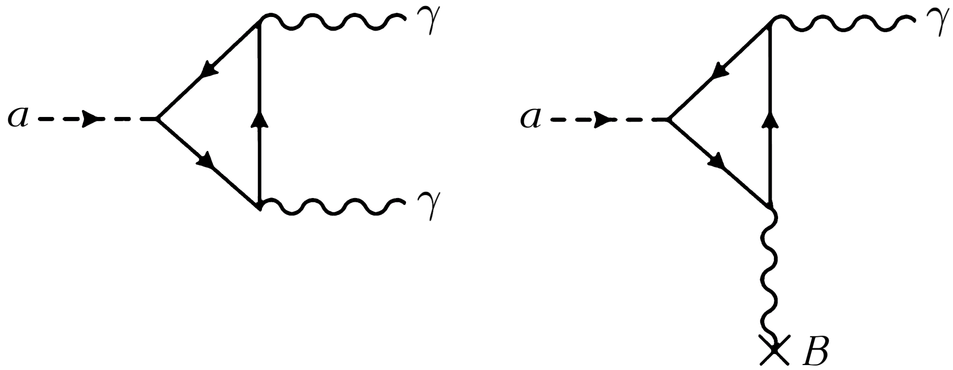


Figure 1.8: The Feynman diagrams for Axion (a) detection. The left- and the right-hand side diagrams, respectively, are the spontaneous di-photon decay and the magnetic field (B)-enhanced mono-photon decay. The applications of these diagrams for Axion/ALP detection is provided in the corresponding text.

searches have reported multiple reproducible anomalies that are still under debate concerning the background modeling, the validity of the DM origins, as well as instrumental calibrations and analysis methodologies, e.g., the 3.5 keV galactic-center γ excess and the ≈500 GeV diffusive $\bar{p}$ bump (Murgia, 2020; Boudaud et al., 2020).

In addition to the above straightforward DM self-annihilation emission searches, there are also detection methods that produce and/or enhance the DM signals on laboratory-built or astronomical object-aided particle focusing mechanisms. Although these methods utilize detection channels other than the simple DM self-annihilation, since we would like to reserve the next category, the direct detection, exclusively for the experimental techniques that have similar features of the main topic of this thesis, we introduce these inspiring but somewhat ill-categorized indirect detection methods here. In particular, the simple SM signal production from DM-DM annihilation in the straightforward indirect detection scheme can be replaced by Axion/ALP models with the iconic photon emission processes shown in Fig. 1.8. Being pseudo-scalars, it is anticipated that Axion/ALPs undertake the spontaneous two-photon decay process shown in the figure hence produce a similar anomalous SM radiation effect as in the simple indirect detection scenario. In practical experimental realizations, however, the two-photon decay is typically exploited by supplying a strong SM photon insertion, illustrated by the magnetic field B in the corresponding diagram in Fig. 1.8, so the anticipated decay signal becomes single-photon while the decay rate is proportionally raised by the applied B-field intensity/photon flux.

Instead of manually building photon detectors that are equipped with strong magnetic fields so as to utilize the processes in Fig. 1.8, one can also search for Axion emissions from local sources that exhibit strong photon activities. The experimental realizations then become much similar to the DM self-annihilation observations, where the local sources help with enhancing the signal by not only providing the photon flux but also a gravitational concentration. An apparent subject near Earth that exhibits a strong photon activity is the Sun. Taking advantage of the high X-ray activity in the Sun, the so-called Axion helioscopes are being built to observe the solar Axion emission, e.g., (CAST Collaboration, 2017). Since the intended Axion production channel in the Sun anticipates the merging of solar X-ray photon pairs, the helioscopes are sensitive in the Axion mass ranges that are roughly equal or below[5] the X-ray range (Fig. 1.11). Another Axion search method exploiting the reversed processes of Fig. 1.8 is that, instead of waiting for two photons to spontaneously annihilate, one can also supply a strong magnetic field to promote the Axion production from an artificial light source. The concept is very realizable in terms of the apparatus requirement, where one can simply construct a strong external B field across a laser beam, and then an "artificial Axion beam" is obtained based Fig. 1.8. Within the coverage of the B field, the Axions produced from the laser can also convert back to photons via the reversed identical process. One proposed experimental realization for examining such a 'photon-Axion-photon" passage is the so-called "light shining through wall" experiment, in which a photon barrier that is only penetrable by DM is placed on the photon/Axion beam path, so the photons must convert into Axions in order to penetrate the "wall" and then convert back into photons to be detected on the other side (Ballou et al., 2015).

Similar to the helioscopes, the photon-Axion production sensitivity falls in the mass range that is comparable or below the original photon energy, therefore the light-through-wall experiments are typically sensitive in the mass ranges that are below the helioscopes due to the available high-intensity laser sources (Fig. 1.11).

4) Direct detection. Direct detection DM search is the main focus of this thesis. Overall, this category includes all the experiments attempting to measure the recoil energies directly deposited by the DM particles when they travel through the detectors. As our solar system steadily plows through the DM halo of the Milky Way, any direct-detection DM detector built on Earth experiences a constant DM flux going through them. If DM indeed exhibits a non-zero interaction cross section with the detector materials, one can then expect recoil energy deposition events due

[5] Since Axions are generated with nontrivial momenta.

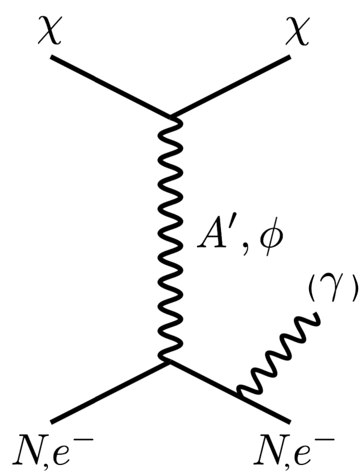


Figure 1.9: The Feynman diagram for elastic-scattering direction-detection DM searches, where χ is the DM, N is the detector neucleon p or n, A' and ϕ are the vector and the scalar mediator fields, respectively. The "N,e^-" labels indicate that the detection utilizes the nucleons and/or the electrons in the detector materials, and the parenthesized photon denotes the final state radiation, which, if occurs, can also be used in parallel with the Ne^- recoils as a detection signature.

to the DM-detector material interaction. From astronomical observations such as Fig. 1.1, we determine the DM mass density at the location of the solar system, ≈300 MeV/cm^3, and we also know the orbiting velocity of our solar system around the Milky Way center, ≈230 km/s, so together we can determine the DM mass flux through the experiments based on the sizes of the detectors. For comparison purposes, it is conventional to further assume that the DM halo is constituted by a single spices of DM particles with an unique mass, therefore the known DM mass flux is translated into the corresponding particle number flux for each examined DM particle mass. Direct-detection experiments also conventionally consider a Maxwell-Boltzmann momentum distribution for the DM particles in the Milky Way halo, based on an unperturbed thermal equilibrium assumption; the simplified constant-velocity picture given above is only for illustration purpose. However, we should note that the assumptions of a single-spices DM halo and the unperturbed thermal momentum distribution are only conventions used for reporting of results. The Maxwell-Boltzmann distribution should also truncate at the escape velocity of the Milky Way galaxy. In reality, these critical input parameters for quantifying direct detection results are actively being examined and refined by astronomical observations and hydrodynamic simulations, e.g., (Posti and Helmi, 2019). For further information, (Lewin and Smith, 1996) provides a general model with a comprehensive introduction for modeling direct detection DM experiments, with more up-to-date numerical data found in (Tanabashi et al., 2018).

Despite the ongoing refinement for the DM influx model, we can already conclude that, if created thermally, the $O(10^2)$-km/s DM particles are highly non-relativistic, therefore the recoil interaction between DM and the SM particles that the detectors utilize for the detection is fully governed by classical mechanics. The recoil interaction is therefore fully analogous to the simple elastic scattering of solid billiard balls. Note that the classical billiard ball analogy is only for the primary DM interaction that generates recoils; the underlying principles for detecting the recoils in different detectors still mostly depend on non-classical physics. Fig. 1.9 presents the Feynman diagram for the direct detection principle. It shows that, considering experimental practicality, the "target" particles that the direct-detection experiments can utilize are very limited, namely only proton, neutron (the nucleon N in the figure), and electron. The diagram also suggests with the vector/scalar mediator labels A'/ϕ that, as long as the target particles anticipate the interactions, all detectable models postulating different interactions are examined simultaneously. This feature is sometimes referred to as "model-independent searches" for direct detection experiments, contrasting the model-specific collider experiments. However, one still needs to interpret the direct-detection data that are collected model-independently for different DM models and therefore generates independent results for different model assumptions.

Assuming the elastic nucleon or electron scattering detection discussed above, Fig. 1.10 presents the recent (2017) constraints on the DM-N/e^- scattering cross sections as functions of the assumed DM masses. Due to null observations, currently all[6] the results report cross section limit curves that rule out larger cross sections from below as in the figure. Similarly for both the nucleon- and the electron-recoil (NR and ER) searches, the limit curves all exhibit linear rises toward higher masses, which reflects the expected sensitivity suppression in cross section due to a reduced "exposure" to the DM flux by the inverse of the assumed mass. The linear sensitivity dependence in the exposure only holds when the instruments can reliably identify DM recoil events, which is not true toward weaker signals. As shown in Fig. 1.10, at lower DM masses that are expected to create weaker recoils in the detectors, every experiment loses its sensitivity over background confusion or its detection threshold and yields a steep rise in the limit curve, resulting in a "mass threshold" effect that depends on the technical details of the experimental construction as well as the

[6] Ignoring detection claims that later contradicted other non-observation results.

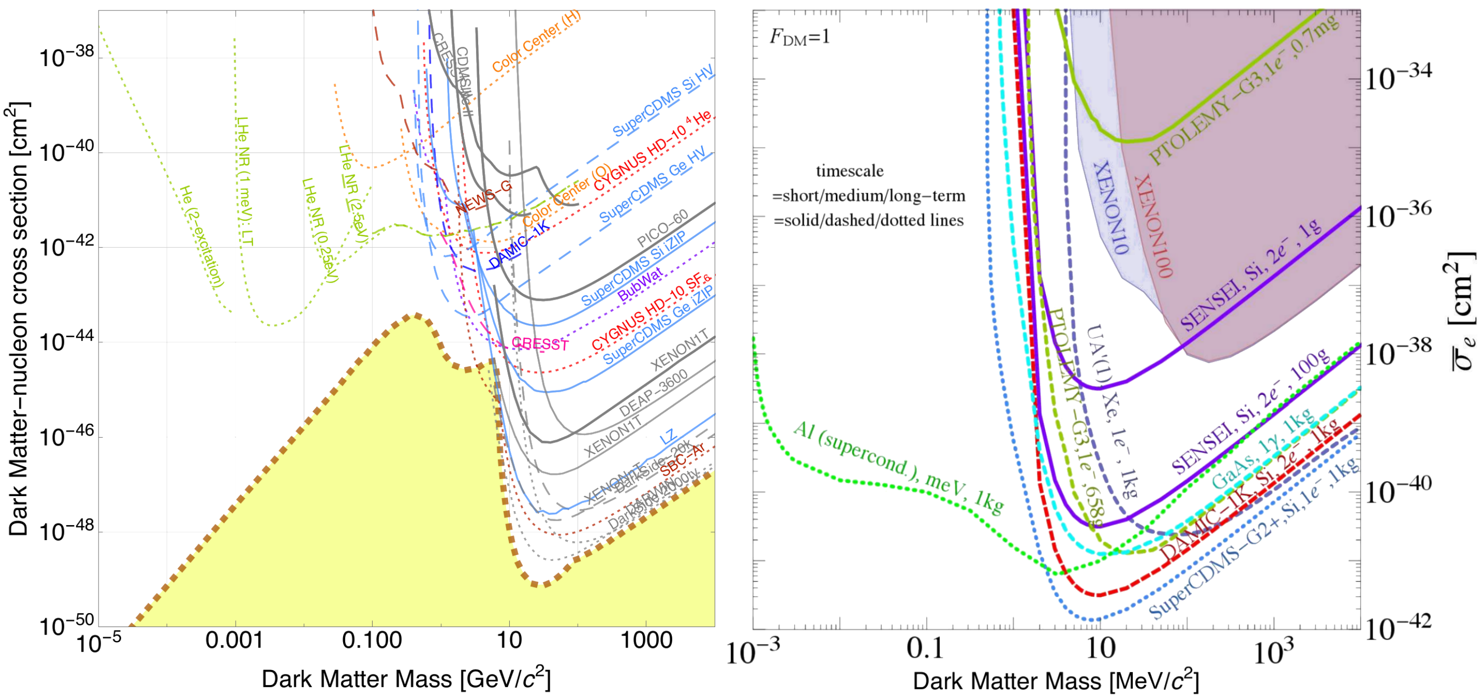


Figure 1.10: Example direct-detection NR (left) and ER (right) constraints reproduced from (Battaglieri et al., 2017), plotting the DM-target particle scattering cross sections against the assumed masses. In both panels, solid, dashed, and dotted curves represent, up to 2017, completed, projected or ongoing near-term, and proposed or planned long-term limits, respectively; a precise classification and the descriptions are found in (Battaglieri et al., 2017). As examples, the NR figure is for the spin-independent scattering, and the ER is for mediators much heavier than the electron. Similar figures assuming different conditions may also be found in the reference. More explanation is provided in the corresponding text.

obtainment of the limits. The background-limited sensitivity is also manifested by the discrete kinks in the limit curves, which reflect the detectors' discrete responses to the background sources that are translated to the corresponding DM masses. In general, the mass thresholds for the limit curves group distinctively at ~GeV and ~MeV for NR and ER searches, respectively. It is because of the classical-scattering momentum transfer efficiency

$$\frac{4Mm}{(M+m)^2}, \tag{1.1}$$

where M and m are the masses of the two colliding objects. The momentum transfer efficiency maximizes when M is equal to m and is monotonically suppressed for larger mass differences. Considering the ≈1 GeV and ≈0.5 MeV masses of nucleon and electron, the momentum transfer efficiency indicates that, given a comparable instrumental noise condition, it is roughly 10^3 times more advantageous in terms of signal-to-noise ratio (SNR) for utilizing electron relative to nucleon to detect DM particles that are much lighter than ~GeV. Likewise, the advantage is with nucleon for searching ~GeV or heavier DMs, therefore the distinct groupings of NR and ER mass thresholds. We compare these characteristic mass ranges to Fig. 1.6 and find that the WIMP models relatively favor NR searches as a combined consequence of the nucleon mass in classical momentum transfer and the Lee-Weinberg bound, while the ER searches are more suitable for lighter non-WIMP dark sector models.

Fig. 1.10 also shows that, currently, the NR searches generally exhibit much stronger cross-section sensitivities than the ER searches. There are both the fundamental physics and the practical reasons for the different ongoing progresses. For the practical reasons, currently the leading NR limits are all from the experiments utilizing scintillating noble liquids as their DM targets, such as Xe and Ar. These experiments utilize the time projection chamber detector design, which has been proven to be a powerful technology in expanding the total target mass capacity, e.g., (Akerib et al., 2020). Nevertheless, as shown in the NR limit plot in Fig. 1.10, noble liquid-based experiments are also not as low in mass thresholds compared to other NR experiments that utilize solid-state targets, e.g., (Agnese et al., 2017). The discrepancy is due to the fact that the solid-state detectors typically exhibit intrinsic background/noise activities at lower energies and at the same time are easier to be integrated with advanced micro-calorimetry technologies. Together, the low-noise characteristic of the solid-state targets and the advanced calorimetry lead to better sensitivities for smaller energy depositions and thus the lower mass sensitivity thresholds. The distinction is also shown in the ER limit plot or Fig. 2.32 for more recent results, where the leading ER limits are all obtained with solid-state detectors due to the requirement for the sensitivity to the much lower energy ER recoils from lighter DM, e.g., (Aguilar-Arevalo et al., 2020). More fundamentally, the ER-NR DM mass sensitivity difference reflects their most advantageous mass ranges for detecting large signals via the classical scattering.

The difference in the ER/NR recoil energy scales pose a strong implication for the signal generation mechanism. For ERs, the expected energy deposition is large enough to ionize single electrons from their bound states to be free particles. For NRs, the nonrelativistic DM exhibits Compton wavelengths much larger than the size of the nucleus and thus scatters all the nucleons in the nucleus coherently. This coherent scattering raises the total scattering amplitude by the number of the nucleons scattered together, which subsequently increases the effective cross section for detection. If the DM-nucleon interaction is independent of isospin, the scattering amplitude multiplication is just the atomic number A, and the cross section or the expected event rate is enhanced by A^2. For interactions that are isospin-sensitive, one still acquires the enhancement by roughly the difference of the numbers of protons and neutrons in the same nucleus. In this case, (Lewin and Smith, 1996) points out that the scattering becomes more sensitive to the inner structure of the nucleus, especially the momentum distribution of the nucleons. In contrast with coherent nuclear scattering, due to the fact that most of the electrons in an atom are in orthogonal orbital states, and the DM-electron interaction is strong enough to probe the orbital transitions, one needs to rigorously calculate the similar coherent-scattering enhancement based on the electronic structures of the target materials, and the enhancement is also expected to be rather close to unity due to the state orthogonality (Essig et al., 2016).

Finally, Fig. 1.10 (c.f. elaboration in (Battaglieri et al., 2017)) shows that, in the coming decade (2021-30), the ongoing experiments project to begin producing NR DM limits that overlap the yellow-shaded region at the bottom of the figure. This shaded area represents the NR DM cross sections corresponding to the scattering rates that can be produced by coherent elastic NRs of neutrinos from various known sources (Billard et al., 2014). On one hand, the prediction presents valuable physics opportunities to examine its underlying assumptions, such as the possible SM interactions or the production mechanisms of the neutrinos On the other hand, the prediction suggests an immediate challenge that, if this neutrino "background" can not be reliably discriminated, the neutrino background confusion will significantly hamper the progress of NR DM experiments. Currently, different approaches for the neutrino background discrimination are being actively investigated, such as through particle flux directionality (Hochberg et al., 2017a) and improving the background model accuracy (Akimov et al., 2017).

In addition to the above relatively model-independent scattering-type searches, there are two more model-dependent detection channels, one for the dark sector models, and the other for the Axion/ALPs. For the Axion, one can utilize the identical strong B field-enhanced photon detectors for the indirect detection but for observing the halo DM flux as for the direct-detection experiments. It is currently the most popular technique for constraining Axion/ALP DM properties, which combines the same source concept for the direct detection and the identical detection technique for the indirection detection, both have been explained. Fig. 1.11 presents the

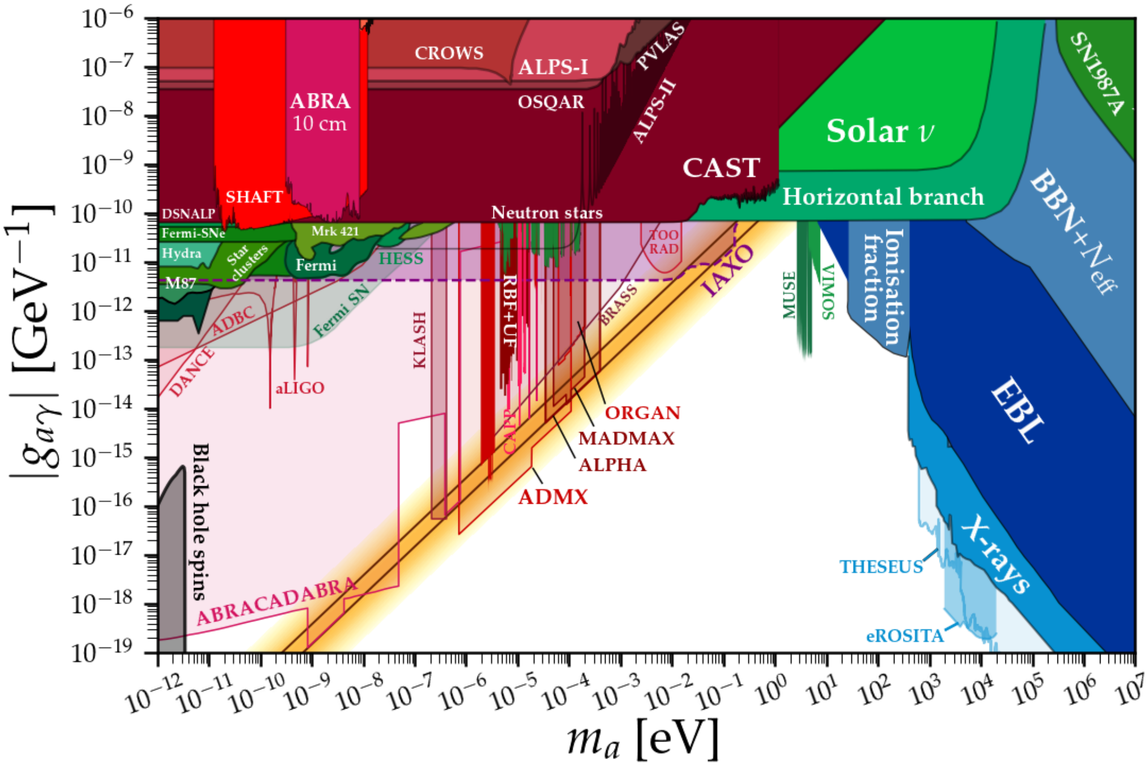


Figure 1.11: Current exclusion limits on the Axion-photon conversion parameter $g_{a\gamma}$ plotted against the assumed Axion masses. The filled and the transparent regions represent the completed and the projected experimental results, respectively. Red, green, and blue indicate the results are from laboratory experiments (including both direct and indirect detection), astrophysical sources, and cosmological observations, respectively. Figure reproduced from (O'Hare, n.d.).

current constraints for the Axion-photon coupling parameter for the diagram in Fig. 1.8. Similar to ER and NR scattering searches, there are currently only exclusion limits bounded from above. Overall, Fig. 1.11 is heavily excluded above ~KeV by cosmological observations (blue in Fig. 1.11). For the warm DM in the (sub-)X-ray mass-energy range, the parameter space is predominately constrained by astrophysical X-ray activities (green in Fig. 1.11) that we have explained can produce Axion/ALP signatures in astronomical indirect-detection observations. Therefore, currently the major direct-detection activities (red in Fig. 1.11) in searching for halo Axion/ALPs mostly focus on the light–ultralight region much below eV. Limited by the relevance to this thesis, we refer interested readers to (O'Hare, n.d.) and the references therein for the production details of Fig. 1.11.

Previously we have explained that the dark sector may in principle consist a zoo of its own diverse particles that exhibits complex evolving phenomenologies similar to

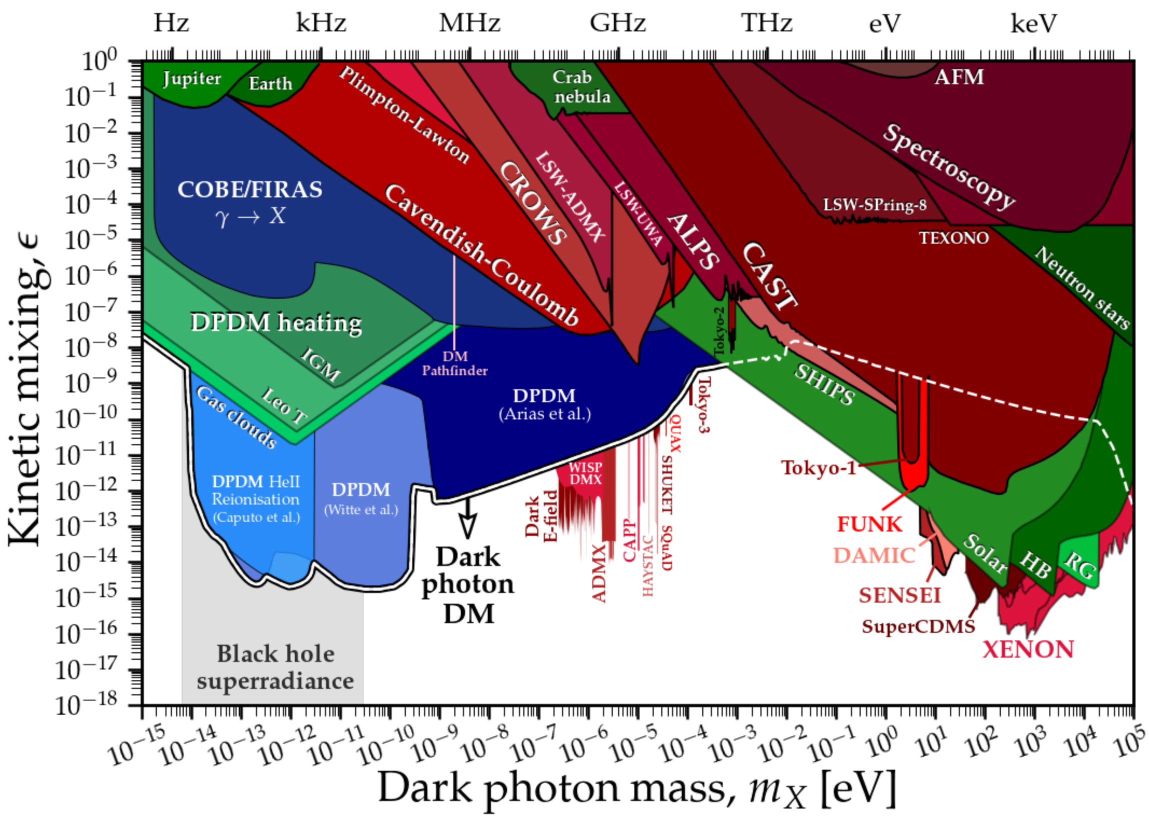


Figure 1.12: Current exclusion limits on the photon-dark photon kinetic mixing parameter ϵ plotted against the assumed dark photon masses. The color scheme is identical to Fig. 1.11. Figure reproduced from (Caputo et al., 2021).

the SM Universe. Although in this case it is relatively difficult to argue for which models are more motivated over others for experimental searches, it is generally agreed that experiments can attempt detection first assuming the simplest dark sector phenomenology. Based on this principle, the simplest possible starting point is a minimal U(1) gauge Boson just like the photon. The difference is that this “dark photon” must be massive in order to consist the DM halos. (Holdom, 1986) then realized that, when combing the SM- and the dark-photon Lagrangians, in addition to the usual (self) kinetic terms $F^{(\prime)\mu\nu}F^{(\prime)}{}_{\mu\nu}$, where we use primed terms to represent the dark photon, a generic “mixing” kinetic term $F^{\mu\nu}F'_{\mu\nu}$ is also allowed in the full Lagrangian. The kinetic mixing mechanism postulates that, at an undetermined mixing size, the usual SM photon interactions can also involve dark photons to complete the interactions. The mechanism is similar to the electroweak mixing that gives rises to Z^0 and the photon but at a much smaller angle than the Weinberg angle. Following the normalization convention

$$\mathcal{L} \supset \frac{\epsilon}{2\cos\theta_{\mathrm{W}}} F^{\mu\nu}_{\mathrm{Y}} F'_{\mu\nu}, \tag{1.2}$$

where Y represents the hypercharge, θ_{W} is the Weinberg angle, and ϵ is the undetermined kinetic mixing parameter, one realizes that every photon detection that are supposedly due to SM photon may be a signature of a dark photon detection at the probability of ϵ^2. In this dark photon detection scenario, the detected photon energy is equal to the total mass-energy of the dark photon (Essig et al., 2016). Therefore, one can search for anomalous photon signals to subsequently constrain the kinetic mixing parameter ϵ as shown in Fig. 1.12. We perform in this thesis such a search via photoelectric effect, and the result is shown in Fig. 2.32 with detailed discussion provided in the next chapter. In addition to the dark photon constraints, we also search for ALP using the same photoelectric, total-inelastic absorption data, assuming that the ALP may trigger the photoelectric effect similarly to a photon via the hypothetical axio-electric effect (Derevianko et al., 2010).

Chapter 2

HVeV RUN-2 ELECTRON-RECOIL DARK MATTER SEARCH

2.1 Overview

The Run 2 (R2) of the HVeV project is a follow-up experiment of (Agnese et al., 2019), which we will call HVeV R1 or simply R1 in the following content. After HVeV R1 demonstrated the first single-charge sensitive phonon-mediated detector and its unprecedented low-mass sensitivity for particle-like electron-recoil (ER) DM, much knowledge has been accumulated in various aspects of the project. These knowledge and improvements, including the DM signal modeling, detector physics and design, calibration techniques, etc., therefore warrant a follow-up Run-2 experiment for implementation. The R2 experiment utilized the upgraded HVeV detector described in (Ren et al., 2020), internally named the NF-C design among the SuperCDMS collaboration. It took place during 04.29–05.16, 2019, in a 44-mK base temperature adiabatic demagnetization refrigerator (ADR) located in a surface laboratory at Northwestern University, IL, USA. The types of data acquired during the experiment include DM search data, ^{57}Co data, and laser data. For the DM search data, we biased the detector at 100 V, 60 V, or 0 V, while the detector (substrate) was held at 50 mK or 52 mK relative to its TES' 65 mK transition temperature. In the end, we collected an 1.2 gram-day total exposure for the 100 V-bias data used by the analysis of this chapter. For each DM search dataset taken between ADR recycles, we took a few laser datasets using the same condition for the DM search data collected immediately after the laser data. We adjusted the laser strength to produce predominately single e^-h^+ events. We took several ^{57}Co datasets at 60 V and 0 V biases after the DM search data collection period. We placed a ^{57}Co radioactive source in front of a beryllium window on the cryostat, so the source's γ emission, up to 136 keV, could illuminate the detector evenly to provide a calibration for position dependence. These data were subsequently analyzed in the following year and were reported in (Amaral et al., 2020a). From HVeV R2, we report constraints on DM-electron scattering cross section σ_e, dark photon kinetic mixing parameter ϵ, and Axion-like particle (ALP) axioelectric coupling g_{ae}, in the mass ranges of 0.5–10^4 MeV/c^2, 1.2–50 eV/c^2, and 1.2–50 eV/c^2, respectively. Due to undetermined background contribution to the observed signal, instead of reporting the corresponding DM-detector interaction strengths, which are subjected to the background confusion, we report our results as upper limits of the interaction strengths. We report the following upper limits at 90% confidence level (CL):

$$\begin{cases} \sigma_e = 8.7 \times 10^{-34} \quad \mathrm{cm}^2 \\ \epsilon = 3.3 \times 10^{-14} \\ g_{ae} = 1.0 \times 10^{-9}. \end{cases} \tag{2.1}$$

In this work, our detector achieves a baseline energy resolution of 2.6 eV, the smallest pulse detector energy resolution ever achieved at the time on phonon-mediated detectors in the X-ray range (Amaral et al., 2020a; Ren et al., 2020), yielding a 2.6% electron recoil-equivalent energy resolution at a 100 V bias. We also perform the first model implementation of charge trapping (CT) and impact ionization (II) effects for DM searches, developed as a R1 follow-up study by (Ponce et al., 2020a), and demonstrate their nonnegligible impacts on the sensitivities of charge-transport type detectors. In the end, we observe a background spectrum that is consistent with R1. Despite the improved energy resolution and exposure, we reported weaker limits with mass thresholds similar to R1 due to the inclusion of CT and II in the presence of a comparable background.

2.2 Detector design and operation

2.2.1 High-voltage ev-resolution (HVeV) detector

SuperCDMS HVeV detectors are designed to collect Neganov-Trofimov-Luke (NTL) phonons generated under a high voltage (HV) bias to obtain the lowest possible threshold. Under the NTL effect, the total phonon energy E_t is given by

$$E_t = E_r + E_{\mathrm{NTL}} = E_r + N_{e/h} e \Delta V, \tag{2.2}$$

where E_r is the recoil energy due to the primary particle interaction, $N_{e/h}$ is the number of electron-hole pairs (e^-h^+) generated, e is the electron charge, and ΔV is the bias voltage across the detector. The equation is reproduced from (Agnese et al., 2018), where the NTL phonon-enhanced detection technique was first demonstrated with more explanation provided therein. We may see from Eq. (2.2) that, given that $\mathcal{O}(1)$ eV[1] is needed to create each e^-h^+ in cryogenic semiconductors, and we assume the recoil energy is efficiently directed into e^-h^+ creation in ER processes, we have

$$\frac{E_r}{\mathrm{eV}} \sim N_{e/h}. \tag{2.3}$$

[1] 3.8 eV for cryogenic intrinsic Si.

Since we aim to lower the DM mass threshold, which in turn suggests the smallest detectable recoil energy, in the limit where only a handful of e^-h^+ are created, we have

$$N_{e/h} \sim \mathcal{O}(1). \tag{2.4}$$

Together with Eq. (2.3) and considering the NTL signal enhancement in Eq. (2.2), we find, in order to substantially leverage the NTL detection mechanism, we need

$$\Delta V \gg \mathcal{O}(1)\ \mathrm{V}. \tag{2.5}$$

If we achieve such a "high" voltage condition, the relative e^-h^+ counting resolution becomes

$$\sigma_{e/h} = \frac{\sigma_{ph}}{(E_t/N_{e/h})} \approx \frac{\sigma_{ph}}{e\Delta V}, \tag{2.6}$$

where σ_{ph} is the intrinsic energy resolution for the phonon energy sensor utilized by the particle detector. For the $\Delta V = 60$ V and 100 V chosen in this work, the above derivation suggests a phonon energy resolution of

$$\sigma_{ph} \lesssim 10\ \mathrm{eV}, \tag{2.7}$$

which allows resolving the quantized generation of e^-h^+ in practice, i.e., unambiguously determining the number of e^-h^+ created in each primary interaction. If an effect that may create non-quantized e^-h^+ or may significantly alter Eq. (2.2) does not exist, we can expect a fully quantized spectrum starting from the single-e^-h^+ peak with negligible between-peak components.

2.2.2 HVeV-R2 detector design

The DM detector employed in this work is a $1 \times 1 \times 0.4$ cm^3, 0.93 g high-resistivity/purity intrinsic Si crystal; Fig. 2.1 shows the detector. The QET-based phonon sensors are fabricated on one of the 1-cm^2 surface of the Si DM target, shaped into a central square and a square annulus around it as in the photo, each covering half of the surface area to achieve an equal phonon sensitivity. To collect the phonons emitted in the substrate when arriving at the phonon sensors, large-area 600 nm thick Al phonon collection "fins" are fabricated covering ≈50% of the instrumented surface. Also shown in Fig. 2.1 is a zoom-in layout for a single QET unit, spanning roughly 140×300 μm^2, that is connected from the Al fins to other identical units by Al traces to construct the two distributed phonon sensors. The distributed sensor design is for suppressing position dependence in the phonon energy distribution and thus the resulting pulse shape. In terms of the electronic

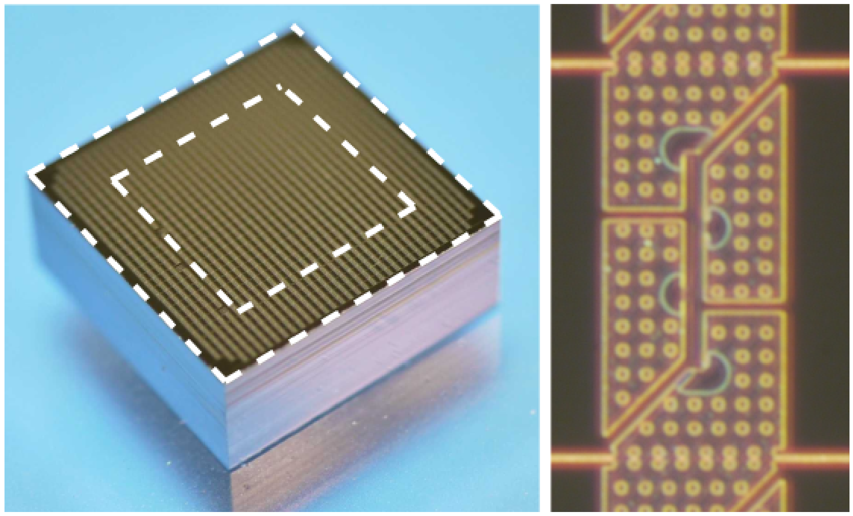


Figure 2.1: Left: The HVeV R2 NF-C detector, consisting of a $1\times1\times0.4$ cm^3 intrinsic Si substrate as the main DM target and dense gold-color QET phonon sensors on the 1 cm^2 surface shown in the photo. The distributed QETs are connectorized into two equal-area phonon sensors as marked by the white dashed lines. Right: A Zoom-in photo of a single QET unit.

circuit connectivity, the Al traces connect the QET units in each phonon sensor in parallel. Multiple fins are connected to a shared narrow rectangular W TES via the sector-shaped “QP traps” to direct the QPs generated by the phonons in the Al fins into the W TES. Due to the much lower T_c = 65 mK chosen for the W relative to Al’s T_c = 1.2 K, the pair-breaking energy that scales roughly linearly with the transition temperature is lowered from the nominal 350–400 μeV for pure Al to a much lower value for the QP traps by the Al-W proximity effect. Because QPs generally immediately relax to the bottom of the band, i.e., possessing half of the pair-breaking energy per QP, once the QPs enter an Al-W QP trap from pure Al, they relax from the higher Al-gap energy and create more QPs that also relax to the bottom of the proximitized Al/W-gap energy. Therefore, these QPs in the QP trap do not possess sufficient energy to create QPs in pure Al; they can only do so in the W TES that has an even lower gap energy. The macroscopic outcome is that the phonon energy absorbed by the QP creation in the large-area Al fins is efficiently collected into the shared W TES through the QP traps, which eventually changes the resistance of the W in its superconducting transition as the signal for phonon detection.

In addition to the main QET architecture, we may see the phonon collection fins are designed with arrays of small square cutouts distributed evenly at a pitch of ≈20

μm. The purpose of these tiny cutouts is to prevent the formation of the magnetic flux-pinning Abrikosov vortices that may interfere the expected Al QP dynamics (Brink, 1995). Finally, in order to apply the bias voltage across the detector, a parquet-pattern grid is fabricated from a 30 nm thick Al film on the opposite surface of the phonon sensors, for which the much thinner Al compared to the photon collection fins is to reduce the phonon loss to the bias grid relative to the fins. For interested readers, much more detailed discussion concerning the detector design, optimization, and characterization may be found in (Ren et al., 2020).

2.2.3 Operation and readout

We operate the QET-based phonon-mediated detector in an ADR in a surface laboratory at Northwestern University. Each time the ADR raises above the Al T_c and recools to below the W T_c for a new data-taking run, we always first scan the current applied to each phonon sensor and record the corresponding TES resistance due to the current's Joule heating effect, while the substrate temperature (base temperature of the ADR) is controlled at a temperature significantly below the W T_c. Specifically for the results reported in this thesis, we control the substrate temperature at 50 mK or 52 mK. After scanning the current, we fix the current individually for each phonon sensor to hold the TES resistance at ≈45% of its normal-state resistance. To make the thermal-electric system a negative feedback system that brings the TES back to the bias condition after pulses, a shunt resistor is connected in parallel with the phonon sensor. When the TES resistance increases, the bias current controlling the TES' in-transition state is temporarily directed to the shunt resistor based on the shunt and the TES resistance ratio, which suppresses the Joule heating for the TES. For the signal readout, a DC superconducting quantum interference device (SQUID) amplifier is installed in each phonon sensor's bias current path, so the fluctuating current can create magnetic flux in the SQUID amplifier's input coil, which becomes a voltage fluctuation across the SQUID that is wired to the outside of the refrigerator. At room temperature, an amplifier for this SQUID output is installed in a typical feedback amplification configuration, whose output voltage is redirected back to the cryogenic environment to be coupled to the SQUID from a feedback coil and subsequently nulls the pickup coil flux. The design achieves the so-called "flux-locked" negative-feedback amplification (Hines et al., 2011; Ren et al., 2020).

2.3 Data acquisition and datasets

2.3.1 Physical configuration and general methodology

Due to the limited ≈12 hours base temperature hold time of the ADR, we need to thermal-cycle the refrigerator hence also the detector between the base temperature and 4 K approximately daily. The system returns to 4 K because it is constantly sunk to a liquid helium (LHe) bath. The daily data taking routine typically involves 1) first confirming the cryogenic system is at the LHe temperature, so to ensure the Nb SQUIDs continue to hold a reliable SC state, then 2) cooling the ADR from 4 K to base temperature by, first, refilling and pumping on the LHe bath to reach about 1.4 K and, second, ramping up the magnetization in the ADR salt pill and, third, (de-)coupling the salt pill (from) to the (1.4-K) base-temperature structure and demagnetizing the salt to achieve refrigeration. During the cooling process, which usually takes a few hours, we regularly 4) perform readout electronics checks, such as the SQUID states and readout noise. We also make use of the long waiting time to 5) pre-bias the detector for about two hours to deplete trapped charge carriers. The depletion of the charge carriers helps reducing the leakage current rate during the following data taking. The pre-bias is nominally done by raising the voltage of the back-side bias grid to 1–2 times of the bias voltage planned for the following data acquisition with respect to the QETs, which are grounded to chassis. Once the detector reaches the base temperature, we then 6) perform the QET-bias current scan introduced previously to fix the phonon sensors into transition, enable the SQUID flux-locking, and from here formally begin data acquisition. We also adjust other conditions such as lowering the HV to the planned values. Before settling the detector into the designated configuration for the dataset, there are a few 7) calibration/diagnostic data to be acquired in advance, in particular the in-transition detector noise and the laser calibration data, which will be discussed in more detail in the next section. We typically also acquire a noise timestream before the detector is biased into transition, i.e., electronics-dominated SC noise, for future diagnosis purposes, but luckily as it turned out these data were never needed for this analysis. Finally, we 8) configure the experiment into the planned data-taking condition.

For the datasets involved in this work, we always record our data in a continuous mode without any real-time triggering. For the DM search data, it means letting the detector observe until the ADR loses base temperature. Depending on the time spent on the above preparation, the ADR typically allows less than 10 hours of DM data exposure, where the ADR could already be at or toward the end of warming up to 4 K. For the ^{57}Co calibration, we readjust daily the relative location of the source to a Beryllium window on the ADR, the calibration photon's entrance, to acquire real-time event rates of about 1–2 Hz. We originally chose the 1-2 Hz ^{57}Co event rate to avoid pulse pile-up but later realized that, after data selection, the data collected at such a low rate did not yield a sufficient data size in our energy range of interest to enable a meaningful calibration for the relative energy scale of the phonon sensors. The readers may find detailed discussions for the data selection and the relative energy scale calibration later in this thesis.

2.3.2 Blind-analysis data

For the data digitization, storage, and reduced quantity generation, we have built an automatic pipeline that processes the accumulated data in parallel with the continuous data acquisition and splits the stored data according to our blinding scheme for the analysis. We use a staged (un-)blinding scheme for the DM search data, where we bin the data timestream by every 10 seconds and, for analysis strategy development, immediately unblind the 1$^{\text{st}}$-second raw timestream with all reduced quantities computed. We call these data the stage-0 data and store them in an open folder for the analyzers' open access. The rest of the data, namely the *blinded* 2$^{\text{nd}}$–10$^{\text{th}}$ seconds, are divided into 2$^{\text{nd}}$–3$^{\text{rd}}$-second (stage-1) and 4$^{\text{th}}$–10$^{\text{th}}$-second (stage-2) and are stored separately for later dark matter searches when the analysis strategy is fully determined with the stage-0 data. In contrast with the blinded DM search data, the laser and ^{57}Co calibration data are not blinded due to their necessity for the analysis methodology development. It was determined prior to the start of the HVeV R2 data-taking/analysis campaign that, instead of having the 1$^{\text{st}}$-second data unblinded with all the rest saved for DM search, we would use part of the blinded data, the stage-1 portion, to verify that the analysis methodology developed from the stage-0 portion is not overtuned to data. More details on the (un-)blinding may be found in Sec. 2.7. If the analysis methodology established passes the verification on the intermediate unblinded stage-1 data, we proceed to fully unblind the data and use the *stage-1 and -2 data together* for the DM search. If the verification had failed, the treatment *determined prior to the start* of HVeV R2 was that we should combine stage-0 and -1 data, resolve the cause to the discrepancy, and then redevelop the analysis strategy based on the combined data. In this case the stage-2 data alone would be the DM search data. We acknowledge the possibility that our stage unblinding scheme may still introduce human bias, as there is one opportunity reserved for revising the provisional analysis methodology. However, since the entire procedure was settled in advance, and more importantly the data used for the analysis strategy development are strictly excluded from the limit setting, we believe the bias introduced, if any, should be modest and would not invalidate the robustness of the reported results. Since HVeV R2 was the first blind analysis attempted in the HVeV project, we reasoned such a stage unblinding scheme serves the best interest not only for the R2 results but also for the development of general blinding techniques for future HVeV analyses.

2.3.3 Data acquisition and signal reconstruction

We digitize the analog output signal from the room-temperature readout electronics with a 16-bits National Instrument PCIe-6374 DAQ at 1.51 MHz. While the digitized timestream from each phonon sensor is continuously being stored, the data acquisition pipeline also searches for pulse triggers and generates reduced quantities such as the pulse height and time for each trigger window,[2] which is chosen to be from 1024 samples before to 3072 samples after the trigger point, corresponding to a window of -675∼+2000 μsec relative to the trigger. To issue triggers, the timestream is continuously convolved with an approximated pulse template and then the convolved timestream is compared to an empirically determined trigger threshold to define rough signal duration(s) that exceed the threshold. The approximated pulse template, consisting of a 20-μs exponential rise time and an 80-μs exponential fall time, was empirically obtained from the same detector prior to HVeV R2 and is only used for triggering purpose in this analysis. We will show in the following sections that the chosen trigger threshold corresponds to ≈30 eV at a 100-V HV bias, i.e., 0.3 e^-h^+. The algorithm then identifies the convolved maximum in each trigger duration as the trigger point.

Although in principle the trigger algorithm we use already provides pulse parameter estimates, in order to be accurate on these critical reduced quantities for the analysis, we adapt the optimal filter (OF) formalism as the primary pulse size and time estimator for this work (Zadeh and Ragazzini, 1952; S. Golwala, 2000), while reserving the trigger information for data selection and future reference. Being one of the most critical algorithms in this thesis, not only for the HVeV R2 work but also the KID-based detector R&D, we will provide an in-depth introduction to the OF formalism when we use it in Sec. 4.6 for KID energy resolution modeling. So here we simply state as given that, to optimally estimate the pulse time and height from a noisy timestream, the OF formalism requires a height-normalized pulse template and

[2] To be more precise, the timestreams are stored by small chunks, so the algorithm may achieve a close-to-simultaneous new data writing and triggering that do not interfere with each other.

an expected noise power spectrum to construct the filter that maximizes the signal-to-noise (SNR) ratio for the detection. Through the OF formalism calculation, one may obtain the estimated pulse height and time offset relative to the template, as well as a χ^2 statistic indicating the resemblance between the observation and the pulse template. Since we always record a stream(s) of pure in-transition noise before each dataset, the noise provides not only a data quality check but also the PSD input for the OF formalism. We always record 5–10 sec of the noise and visually confirm that there are no apparent pulses in the timestream, e.g., due to muons.

We proceed to the laser calibration after the noise data acquisition, where we obtain the pulse template input for the OF after visually removing pulses that are obviously distorted, e.g., overlapping pulses due to background particles/muons, anomalously large pulses, etc., and then average the accepted ones for noise suppression. Similar to the noise input, the laser pulse template is separately determined for each DM search dataset with its own laser calibration data. To best represent the low-mass DM recoil signal, we adjust the laser strength to create predominately single e^-h^+, and for each dataset, we accumulate a data size that allows calibrating up to the 4^{th} e^-h^+ peak given by the Poisson fluctuation. We collect for each dataset 15–30 minutes of laser pulses at 100–300 Hz, with a synchronized TTL signal sent to the DAQ from the laser driver for laser event tagging.

2.4 Calibration

2.4.1 e^-h^+ quantization

We use a 635-nm (1.95 eV) diode laser as the main energy scale calibration source of this work. The photons are transmitted from room temperature via a single-mode optical fiber, through two KG-3 glass filters mounted at 1.4 K, to illuminate at the center of the instrumented detector surface. We use the KG-3 filters to suppress the infra-red (IR) and sub/mm blackbody radiation transmitted with the calibration photons through the optical fiber. The photon-emitting end of the optical fiber is directly mounted on the lid of the hermetic copper detector housing, pointing vertically at the center of the detector that forms a laser spot size of ≈4 mm in diameter on the instrumented surface. By absorbing the photons with electrons in the Si substrate to generate quantized e^-h^+ and assuming a perfect efficiency for the photoelectric effect-to-phonon energy conversion, the total phonon energy associated with the ER events in the presence of the NTL-effect HV bias is

$$E_N = N_{e/h}(E_\gamma + e\Delta V), \tag{2.8}$$

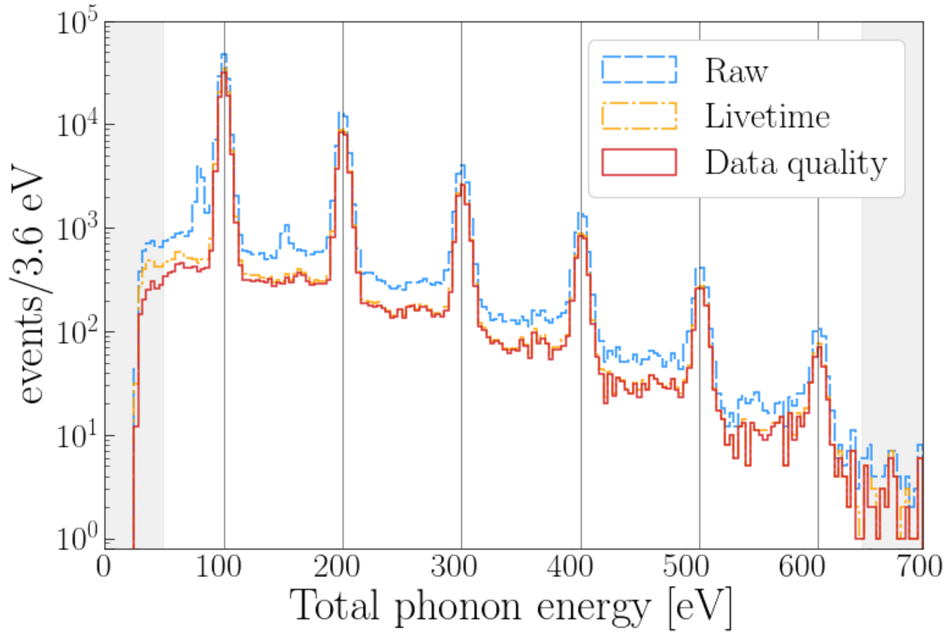


Figure 2.2: 100-V laser calibration data combining all laser datasets after energy scale calibration. The blue dashed, yellow dot-dashed, and red-solid curves represent the raw data, the livetime-selected, and the event-by-event selected spectra, respectively. The vertical gray lines mark the expected e^-h^+ peak locations according to Eq. (2.8), and the gray-shaded regions indicate the regions outside the region of inters (ROI) of this analysis. The spectra are binned by 3.6 eV, the post-selection peak-averaged nominal energy resolution used for the limit-setting. Further information about the calibration, selection, ROI, and limit-setting may be found later in the corresponding sections.

where $N_{e/h}$ is the number of e^-h^+ generated, E_N is the total phonon energy corresponding to $N_{e/h}$ e^-h^+, and $E_\gamma = 1.95$ eV is the primary ER energy contributed by each 635-nm laser photon (Amaral et al., 2020a). Fig. 2.2 shows the laser spectrum. Benefiting from our detector's e^-h^+-counting resolution, we may achieve an energy scale calibration with Eq. (2.8) by straightforwardly identifying the $N_{e/h}$-to-E_N correspondence for the clearly quantized laser energy spectrum and assigning to each peak the E_N according to Eq. (2.8).

2.4.2 Calibration strategy

We calibrate for the energy scale dependences discussed in this section that cause the relative energy scale to deviate dataset-to-dataset. The goal is to correct for these dependences by aligning the quantized e^-h^+ peaks across different laser calibration datasets, so we can apply the same treatment for the DM search data and then combine them for the DM search. To begin, we anticipate ΔV to be a major variable to be calibrated between datasets due to the reapplication of the HV bias. While this reapplied value is known to vary due to analog knob adjustment, the set value is nonetheless stable for each dataset once set. In addition to the external HV that may alter the calibration, internally the TES' responsivity also depends on the thermal link between the heat bath (substrate) and the Joule-heated TES film. At thermal equilibrium, one has

$$\frac{V_b^2}{R_0} = K(T_c^x - T_b^x), \qquad x \approx 5, \tag{2.9}$$

where the left-hand side is the Joule-heating power given by the TES DC bias-voltage V_b and its resistance R_0, the right-hand side represents the electron-phonon coupling for the dissipation of the applied bias power, T_b is the substrate/bath temperature, and empirically x is found to be ≈ 5. (Pyle, 2012) Eq. (2.9) suggests that not only does the energy scale of the detector depend on the bath temperature, hence needed to be calibrated separately for the 50 mK- and 52 mK-T_b datasets, but it can also fluctuate for a chosen T_b due to the temperature fluctuation of the ADR. (Pyle, 2012; Kurinsky, 2018) The latter is mostly because our refrigerator's PID temperature control only holds the temperature within about $\pm 2\%$ fluctuation and at the same time regularly causes temperature divergences roughly hourly. Eq. (2.9) also suggests that the energy calibration's temperature dependence is to first order independent of the HV applied and therefore may be calibrated independent to the HV. Based on this argument, we decide to calibrate the energy scale by first correcting and subsequently removing the temperature dependence of the data, and then we rescale the temperature dependence-removed data by the recorded HV offset relative to the nominal value. Note that we perform the above steps with the pulse size data in the relative electronics unit but not a physical energy unit such as eV. After the two dependences are calibrated based on known effects, we empirically adjust the energy scale so that the e^-h^+ peaks are evenly spaced as expected by Eq. (2.8). Finally, we may convert the energy scale from the DAQ unit to the physical energy in eV by assigning the fully calibrated e^-h^+ peaks in the laser data with their corresponding quantized energies, and then the energy scale defined by the laser data may provide the standard energy scale for the physics data.

2.4.3 Temperature calibration

We anticipated the relative calibration would depend on the fluctuating temperature, so after we concluded the full DM search data acquisition campaign, a series of dedicated laser datasets were taken for developing the temperature calibration. As discussed previously, we expect the temperature calibration to be independent of the

HV applied but strongly depends on the designated base temperature, so to calibrate the 50-mK and 52-mK data, the temperature calibration data were taken at 60 V, scanning 49.9–51.0 mK and 51.9–53.0 mK at 0.1 mK increments. Fig. 2.3 shows

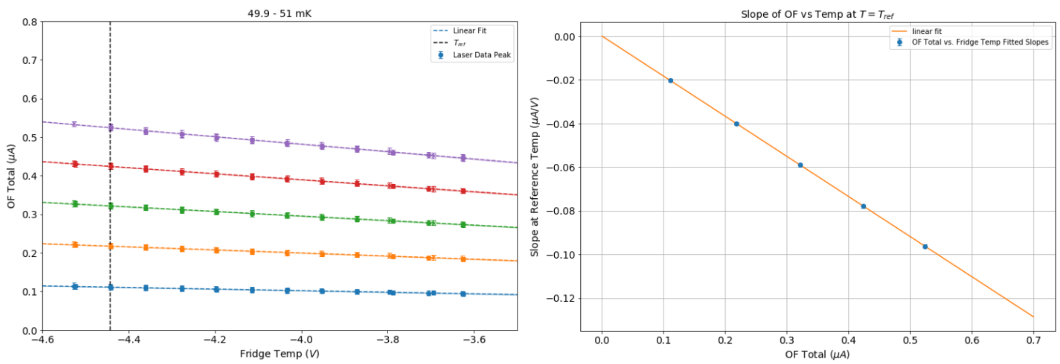


Figure 2.3: Left: Laser peak locations, from bottom to top for $N_{e/h} = 1$ to 5, plotted against the ADR temperature, both in raw electronic units. The vertical dashed line marks the nominal operation temperature, in this figure 50 mK as an example. Right: The laser peaks' temperature dependences, i.e., the slopes of the colored solid lines in the left-hand side figure, plotted against the "correct" corresponding peak locations, i.e., the dashed-solid line intersections in the left-hand side figure. Figures produced by M. Wilson.

the data for the calibration study, where we use a Gaussian function to extract the peak position for each peak in every dataset. We may see, due to the controlled temperature variation, the observed OF signal size ($\texttt{OF}_{\texttt{tot}}$) is indeed continuously suppressed toward higher temperatures as expected. Owing to the small variation in temperature (1.1 mK) relative to the absolute temperatures of 50 mK and 52 mK, we linearize the energy scale's dependence in base temperature to

$$\texttt{OF}_{\texttt{tot}} = m(x - x_0) + \texttt{OF}_{\texttt{cor}}, \tag{2.10}$$

where x is the base temperature thermometer reading as shown by the x axis in Fig. 2.3, x_0 is the targeted 50-mK or 52-mK thermometer reading, m represents the linear temperature dependence to be determined, and $\texttt{OF}_{\texttt{cor}}$ is the "correct" OF signal size to calibrate to, i.e.,

$$\texttt{OF}_{\texttt{cor}} = \texttt{OF}_{\texttt{tot}}(x = x_0). \tag{2.11}$$

Note that different numbers of $e^- h^+$, corresponding to different signal sizes, should yield different m and $\texttt{OF}_{\texttt{cor}}$, and the parametrization needs to be linearized/extracted locally for 50 mK and 52 mK. By fitting Eq. (2.10) to the data separately for every laser peak, we may further extract the model slope (m)'s dependence in $N_{e/h}$, also shown in Fig. 2.3, and subsequently promote the Eq.-(2.10) relation extracted discretely for integer $N_{e/h}$ to a linear continuous relation that describes the responsivity suppression due to raised base temperatures,

$$m = m' \, \mathtt{OF}_{\mathtt{tot}}|_{x=x_0} + m_0 = m'\mathtt{OF}_{\mathtt{cor}} + \cancel{m_0}, \tag{2.12}$$

where m' is a negative constant representing the increasing suppression on larger pulses, and m_0 is phenomenologically used by the linear equation that is expected and indeed found to be consistent with 0. With the sets of (x_0, m', m_0) separately determined with the 50 mK and 52 mK temperature calibration data series, we may combine Eq. (2.10) and Eq. (2.12) to correct every event based on its observed OF signal and recorded instantaneous temperature by

$$\mathtt{OF}_{\mathtt{cor}} = \frac{\mathtt{OF}_{\mathtt{tot}}}{1 + m'(x - x_0)} \tag{2.13}$$

on an event-by-event basis. Fig. 2.4 shows the result of the temperature calibration, where we may see, through this method, the amount of correction is generally within ±0.3% during normal PID control periods and periodically reaches 2–3% during PID divergences.

2.4.4 High voltage calibration

We proceed to correct for the HV fluctuation for each dataset after the temperature dependence is calibrated. We expect this HV fluctuation to be dominated by the limited practical accuracy with resetting the HV bias dataset-to-dataset using the analog knob on the HV power supply. Motivated by the poor precision (≈V) given by the power supply panel, a voltmeter accurate to 10^{-2} V is installed to record the real-time HV for every event during data-taking. From the voltmeter data, we find the HV is stable once the knob is set, but the knob-set value may vary dataset-to-dataset by as much as ±2 V, even though shown exactly 100 (60) V by the power supply. Fig. 2.5 shows the energy scale versus the voltmeter-recorded HV. To extract the energy calibration's dependence in HV, we fit a linear relation to the temperature-corrected OF signal size versus HV independently for each dataset. In order to be representative for our interest in small signals yet avoiding high background bias, instead of using the single e^-h^+ data, we extract the second e^-h^+ peak position using $\pm 3\sigma$ $\mathtt{OF}_{\mathtt{cor}}$ data around the peak. We then fit a line

$$\mathtt{OF}_{\mathtt{cor}} = a \times \mathtt{HV} + b \tag{2.14}$$

to the peak position values, where $\mathtt{HV}$ is the recorded HV, and a, b are the linear parametrization. We use this linear model to further calibrate the temperature-

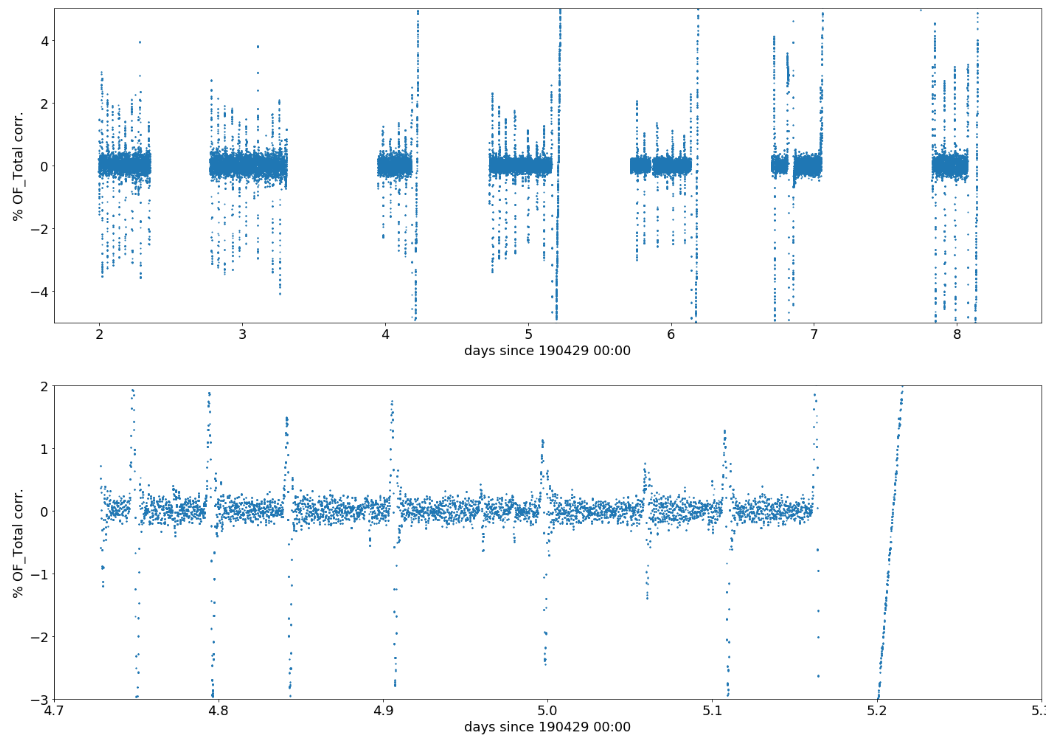


Figure 2.4: Percentile temperature correction for every event in the stage-0 data. Top: All the datasets plotted by their absolute time, where we may visualize the daily data-taking schedule, the hourly PID divergences, and the loss of ADR base temperature at the end of each dataset. Bottom: A zoom-in to the 4$^{\text{th}}$-day data, showing the 0.3% correction for the nominal temperature reading fluctuation, the 2–3% correction for the hourly base temperature divergence, and the uncalibratable loss of base temperature toward the end of data-taking. We acknowledge that the 0.3% correction under nominal data-taking condition is likely due to instrument stability or noise, but we apply the temperature correction regardlessly for data processing consistency and simplicity.

corrected OF signal size by

$$\mathtt{OF}'_{\mathtt{cor}} = (\mathtt{OF}_{\mathtt{cor}} + (\mathtt{HV}_{\mathtt{avg}} - \mathtt{HV}) \times a) \times \frac{\mathtt{HV}_{\mathtt{0}}}{\mathtt{HV}_{\mathtt{avg}}}, \tag{2.15}$$

where $\mathtt{HV}_{\mathtt{avg}}$ is the average HV of the dataset, and $\mathtt{HV}_{\mathtt{0}}$ is the intended nominal HV (100 V or 60 V). The parenthesized term in Eq. (2.15) first corrects the event-by-event HV fluctuation to the average value of the dataset, and then the following HV ratio term rescales the knob-set HV to the nominal HV. We find, as expected by the 10^{-2}-level HV instability, the parenthesized relative correction in Eq. (2.15) presents a $\lesssim 10^{-4}$-level modification to the temperature-corrected $\mathtt{OF}_{\mathtt{cor}}$, which is negligible compared to the %-level detector resolution. However, due to the daily HV reset with a mechanical knob, the rescaling between the target HV and the obtained average

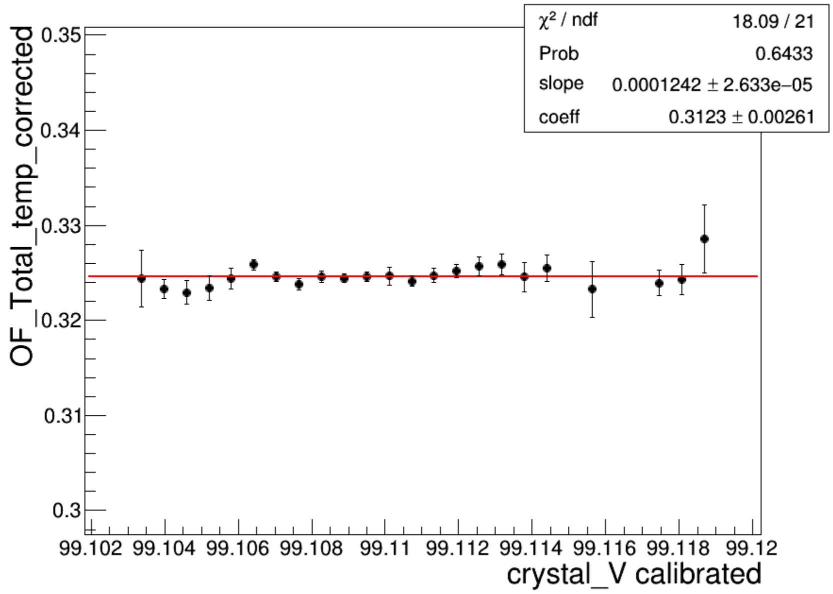


Figure 2.5: Temperature-calibrated 2nd e^-h^+ peak position plotted against the voltmeter-recorded HV. Due to the digitization of the voltmeter, only a range of certain values are assigned to the events as shown by the few data points in this figure. It therefore allows us to bin the date with the same HV to separately locate their corresponding 2nd e^-h^+ peaks. Figure produced by V. Novati.

does correct the energy scale at a comparable level to the detector's intrinsic 3–4% energy resolution and therefore is necessary to be considered.

After calibrating for the anticipated energy scale dependences discussed so far, we pause and examine the result before applying further empirical treatments to improve the data quality. As shown in Fig. 2.6, when we examine the potential impact on the calibration due to the choice of different laser photon fluxes, we find the "0th peaks," i.e., the e^-h^+-less TTL-tagged events centering at $\mathtt{OF}'_{\mathtt{cor}} \gtrsim 0$, present an intriguing positive correlation to the Poissonian mean numbers (λ) of e^-h^+ in datasets with different levels of laser illumination. Fig. 2.6 visualizes the 0th-peak offset against the applied laser strength. The fitted linear trend line suggests that the "e^-h^+-less detection/offset" is not just proportional to the selected laser strength, but it also suggests a constant "background light" by the positive intersection to the y axis; in the case of Fig. 2.6, about 0.3 eV-equivalent. We therefore hypothesize that there are at least two different mechanisms causing such a phenomenon, where one should correlate to the generation, emission, or detection of the calibration laser photons, while the other is independent and generates a constant signal that offsets the 0th peak regardless of λ. For the former, currently the hypotheses include 1) laser

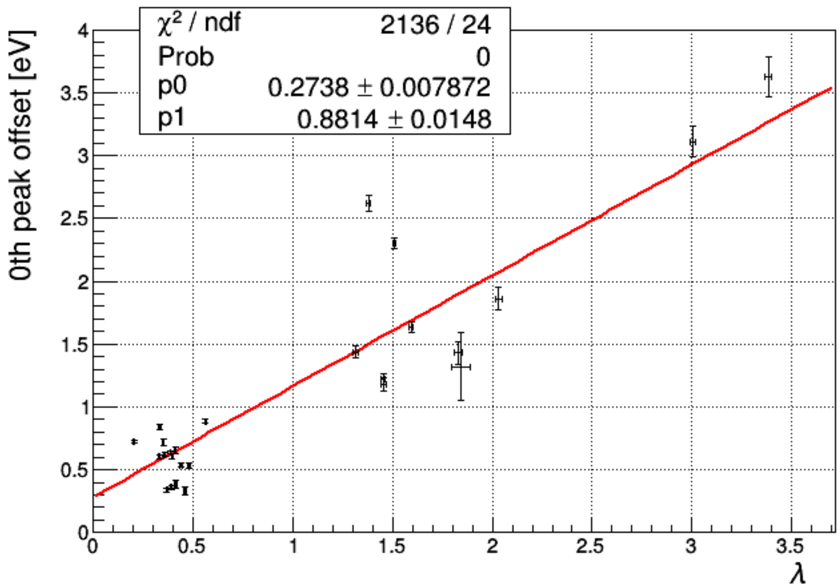


Figure 2.6: e^-h^+-less laser peak position in each laser dataset plotted against the average number of e^-h^+ generated.

driver-caused, e.g., electrical crosstalk, or the laser/laser-driver may generate non-635-nm radiation that is sent down the optical fiber with the calibration photons and subsequently detected. It could also be 2) a clue to an incomplete modeling for the actual calibration signal. For example, due to a finite photon detection efficiency, it is natural that multiple photons need to be generated in order to excite each e^-h^+ for the calibration. In this scenario, those nonabsorbed photons may still generate signal via, for example, direct absorption by the QET film or creating temperature instability in the detector. For the latter, it is in fact easier to understand and, as a nondiminishing background, requires simpler conjectures to satisfy the observation. Currently the hypothesized sources for the constant-energy background, i.e., the y offset at $\lambda = 0$, include 3) possible scintillation or florescence due to a range of materials surrounding the detector, such as the printed circuit boards (PCBs). These materials may scintillate due to cosmic ray or background radioactivity bombardments and therefore lead to the observed small background energy. In terms of the physical size of the λ dependence, which we will calibrate later but choose to show already in Fig. 2.6 for a better illustration, the correlated generation of extra signal is $\approx$1 eV/λ, roughly a 1% excess for a HV bias of 100 V, and the uncorrelated component is ≈ 0.3 eV, which, due to sparse data, may still be consistent with zero. Many of the aforementioned mechanisms for generating the 0^{th}-peak offset may in fact be tested by methods such as correlation detection with multiple detectors arranged to be sensitive to the conjectured sources, or simply suppressing the overall background radioactivity[3] by means of lead shields or moving to underground locations. It is exactly one of the main missions for the ongoing HVeV Run 3 effort, where four HVeV detectors are deployed simultaneously in a better-shielded environment at the NEXUS underground laboratory (Z. Hong et al., 2019). Meanwhile, since the phenomenon is unexpected and first discovered in HVeV R2 without much handle for correction, we decide to only calibrate for the factors that we may physically justify and preserve this empirical observation in the following empirical calibration steps.

2.4.5 Energy scale assignment and nonlinearity

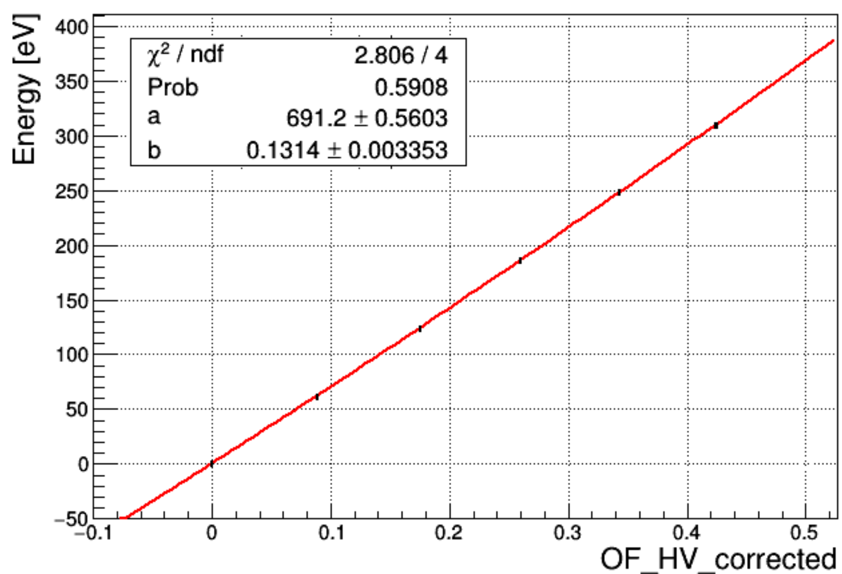


Figure 2.7: An example 2nd-order polynomial fit to the physical energy-OF pulse height relation for a laser dataset taken at 100 V.

We now assign the e^-h^+ peaks in the laser data with the energies according to Eq. (2.8), where we also consider empirically a mild non-constant conversion between the calibrated DAQ-unit $\mathtt{OF}'_{\mathtt{cor}}$ and the physical energy in eV. Based on Eq. (2.8), we assign the laser peaks with

$$E_N = N_{e/h}(1.95 + \mathtt{HV_0})\ \text{eV}, \text{ where } \mathtt{HV_0} = 100 \text{ or } 60. \tag{2.16}$$

To empirically accommodate for the nonlinearity, we model the conversion from the DAQ-unit temperature- and HV-calibrated $\mathtt{OF}'_{\mathtt{cor}}$ to the eV-unit final OF energy

[3] Effective for the $\lambda = 0$ offset.

$\texttt{OF}''_{\texttt{cor}}$ by

$$\texttt{OF}''_{\texttt{cor}} = E_N = a \cdot \texttt{OF}'_{\texttt{cor}} \times (1 + b \cdot \texttt{OF}'_{\texttt{cor}}), \tag{2.17}$$

where a and b represent the parabolic parametrization fitted from $E_N(\texttt{OF}'_{\texttt{cor}})$ assigned by Eq. (2.16). The parabolic fit is shown in Fig. 2.7. The $\texttt{OF}'_{\texttt{cor}}$ data points corresponding to the quantized laser peaks are extracted using Gaussian distributions on the $\pm 2\sigma$ regions around the peaks, and the error assigned to each peak position is the uncertainty given by the Gaussian fit. Note that with Eq. (2.17), not only do we perform an empirical modeling between E_r and $\texttt{OF}'_{\texttt{cor}}$ that effectively corrects the nonlinearity, but we also implicitly convert the unit for the detected signal size from the DAQ unit of μA for $\texttt{OF}'_{\texttt{cor}}$ to the physical energy unit of eV for E_r. Therefore for future reference, to ensure the calibration is applicable for the corresponding physics dataset, in particular the unit-full a and b fitted from the laser data, the same unit for the input $\texttt{OF}'_{\texttt{cor}}$ needs to be used. Similar to previous steps, the (a, b) calibration is independently obtained from each laser calibration dataset and then applied to its associated physics dataset.

According to the extracted (a, b), we find that the nonlinearity yields a signal size suppression of $\approx 8\%$ at $N_{e/h} = 6$, roughly the energy upper bound[4] for this analysis. After all the above calibration procedures, we have now corrected for the temperature and HV fluctuations as well as empirically calibrated the raw DAQ unit into eV, so we combine all the datasets, for both the laser calibration and the physics data[5], to yield the final spectra for subsequent DM analyses. Fig. 2.8 shows the fully calibrated combined laser data from all datasets, where we obtain a final Gaussian width of 3.52 ± 0.01 eV for the single e^-h^+ peak, with increasing values up to 5.13 ± 0.06 eV for higher-$N_{e/h}$ peaks up to $N_{e/h} = 6$. We conclude the e^-h^+-counting resolution obtained in this analysis is

$$\sigma_{e/h} = \frac{3.52}{101.95} = 3.45 \times 10^{-2} \quad [N_{e/h}], \tag{2.18}$$

achieving the planned e^-h^+ counting sensitivity for a quantized ER DM search.

2.4.6 Partition

We discuss below a study of the relative calibration of the two phonon sensors on the detector. We find that the sensors exhibit consistent sensitivities and thus can be treated as one sensor by combining their signals. We had hoped to illuminate the

[4]Introduced in Sec. 2.5.
[5]Only the unblinded data.

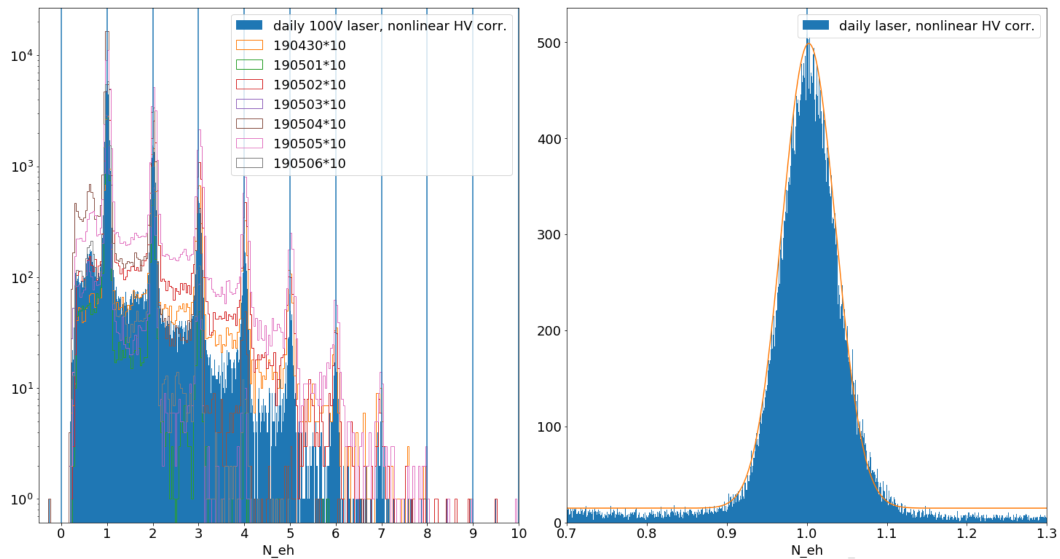


Figure 2.8: Left: The fully calibrated laser data in the unit of $N_{e/h}$ combining all datasets. Right: A zoom-in to the first e^-h^+ peak.

detector uniformly with the ^{57}Co source to obtain a dedicated position dependence calibration, but the collected dataset was too small to yield a more precise result than that described below.

Fig. 2.9 presents the raw signal sizes in the central and the annular sensors. We select the events plotted in Fig. 2.9 identically to that done to $\mathtt{OF}_{\mathtt{tot}}$. As shown in the figure, the quantized e^-h^+ peaks follow a line of a slope of 1, indicating the sensors' responsivities are equal. Based on the result, we calibrate the energy scale of the detector using the combined signal from the two phonon sensors, which are added without relative weighting. The calibrated signal in the unit of eV is called $\mathtt{OF}''_{\mathtt{cor}}$. We then compute the contributions from the inner and outer sensors by

$$\mathtt{OF}''_{\mathtt{cor,in(out)}} = \mathtt{OF}''_{\mathtt{cor}} \left(\frac{\mathtt{OF}_{\mathtt{tot,in(out)}}}{\mathtt{OF}_{\mathtt{tot,in}} + \mathtt{OF}_{\mathtt{tot,out}}} \right), \tag{2.19}$$

where the subscripts `in` and `out` denote the inner square and the outer annulus.

We define the partition quantity η, an inner-outer detection asymmetry parameter that indicates the "off-center-ness" of an event. We follow SuperCDMS' sign convention to define

$$\eta = \frac{\mathtt{OF}''_{\mathtt{cor,out}} - \mathtt{OF}''_{\mathtt{cor,in}}}{\mathtt{OF}''_{\mathtt{cor,in}} + \mathtt{OF}''_{\mathtt{cor,out}}}, \tag{2.20}$$

which is zero for equal inner- and outer-sensor signals and is positive for events with higher outer-sensor signals. Notice that the partition quantity may in principle

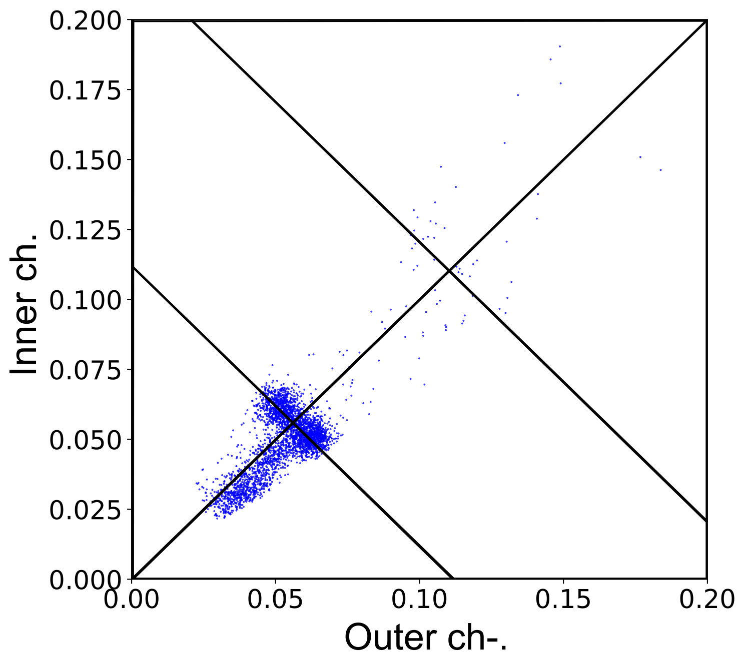


Figure 2.9: The uncalibrated stage-0 OF signal sizes of events detected by the inner and outer phonon sensors. Figure produced by V. Novati.

be nonlinear with respect to the physical radius depending on the design and the geometry of the detector. Although we are unable to quantitatively implement the partitioning for the data selection and therefore did not report the information in (Amaral et al., 2020a), we will report interesting qualitative findings later in this chapter based on the partition quantity. The result may potentially reveal critical implications for future solid-state target DM searches thus are worth further studies in the future.

2.5 Data selection

2.5.1 Selection strategy

In order to improve our data's reliability for the later DM signal identification, we apply the following data selections to remove the data that are more likely to be caused by non-DM sources or incorrectly reconstructed. The first stage of the data selection is the discarding of anomalous time periods, dubbed livetime selection in the following, where we identify periods of time that the detector and/or the experimental construction as a whole perform abnormally or were relatively prone to generating low-quality data. Based on well understood physical causes, we regard these periods of time as inactive times for the experiment due to overly compromised data quality or data-taking condition. We discard all the data acquired during these times and adjust the total DM exposure accordingly for the DM search. The second stage of the data selection is the event-based selection, where we further discard events on an event-by-event basis according to certain characteristics of the events. These characteristics can either indicate that the events are unreliably reconstructed or are more likely to be non-DM events. For the event-based selection, we estimate the probabilistic loss of DM events (efficiency) due to the selection and account for it when calculating the corresponding DM interaction strengths.

In general, the livetime selection is more effective in removing unreliably reconstructed events, which in principle may still be DM events but nonetheless introduce a high uncertainty to the DM interaction strength calculation and thus can reduce the data's DM sensitivity in practice. On the contrary, the event-based selection is designed based on our knowledge and assumptions in the detector and its DM interaction, therefore the selection may be regarded as a model/assumption-driven veto that primarily discriminates events that do not resemble our expectation for DM events. While preferentially removing the background events with respect to good DM candidates is beneficial to the "purity" of the data, the event-based selection may randomly discard a fraction of potential DM candidate events. Assuming the ER DM models that we target in this work produce identical signals to the calibration laser photons, we apply the same event-by-event selection to the laser data to measure the selection efficiency. The efficiency is accounted for when calculating DM limits.

2.5.2 Trigger burst livetime selection

We consider for the livetime selection 1) significantly raised temporary trigger rate, 2) the height of the pre-trigger signal trace, and 3) the temperature of the detector, applied in the written order. Based on existing DM interaction cross section limits, 1) suggests the data acquisition is compromised by non-DM sources, such as sudden electronic noise pick-ups. We use 2) as an indication for the events overlapping with preceding events, the so-called "pile-up" events, where the energy reconstruction may be biased due to distorted pulse shapes. We have explained that the detector temperature is a major dependence for the energy scale, which in practice is difficult to be reliably corrected outside a limited range due to uncorrelated fluctuations. Therefore, we select on 3) to limit the temperature range for the correction in our energy reconstruction in order to prevent such an energy resolution degradation.

The first selection in the livetime selection, the trigger rate-based selection, or internally called the "trigger burst" selection, is performed mostly identically to HVeV R1 (Agnese et al., 2019). However, since the selection methodology depends on the data's time-varying trigger rate, the automated data processing tool developed for the non-blind R1 analysis is not immediately applicable for this work due to our time-domain blinding scheme. In order to explain the difference, we should briefly introduce the trigger burst selection algorithm, and we refer the readers to (M. Wilson, 2021) for more detailed information by the developer of the R2 selection tool. In summary, we would like to remove time bins where "significantly" high trigger rates appear temporarily, which are obviously inconsistent with our expectation for DM. However, the determination for such a significance requires referencing a "normal" trigger rate of the data. As it turns out, such a normal trigger rate may slowly drift over time and also correlate with long time-scale effects such as the daily cycle of human activities, e.g., cellphone activities. So instead of calculating the level of trigger rate fluctuation for each time bin with respect to the average of the entire dataset, which spans much longer than the drifting time scale, we designate another bin size that is much smaller than the drift time scale, yet much wider than the time bin to be discarded, so to calculate a "local normal" trigger rate for reference. The locally averaged rate is then compared to the "instantaneous" trigger rate represented by the narrowest time bin to determine if it should be rejected from the analysis. Therefore, it is understandable that the binning treatment, namely the need for the rolling calculation of the reference local trigger rate, was not an issue when the data were fully revealed in R1, wheres in the blinded R2 it could be biased due the choice of representing the data with the 1$^{\mathrm{st}}$ second in every 10 seconds (Sec. 2.3). For example, a periodic variation at the 10-sec. (or the multiples of 10 sec.) periodicity can be systematically missed by the unblinded portion and lead to a bias. To take advantage of a modified data processing script from R1, we compensate for the artificial effect by the assumption that the data in every 10 seconds are represented by the 1$^{\mathrm{st}}$ second, so the trigger burst identification algorithm described in (Agnese et al., 2019) may be applied identically. We choose to reject time bins which exhibit $+3\sigma$ deviations from the locally determined normal trigger rates. Notice the positive sign since only a raised trigger rate is undesirable.

We apply the trigger burst selection in two steps, first the overall trigger burst selection and then the leakage trigger burst selection. The overall trigger burst is already described above, considering all triggers independently of energy, while the leakage trigger is further defined by only determining the event rate with

$$\mathrm{OF}''_{\mathrm{cor}} > 0.8 \cdot N_{e/h}, \tag{2.21}$$

where $N_{e/h}$ is calibrated to the corresponding single-e^-h^+ phonon energy depending on the HV of each dataset. The motivation for the two-steps selection is due to the fact that, because of our rather liberal choice of the $\lesssim 0.3 N_{e/h}$ (100 V) trigger threshold for preserving more data, the data collected are always overwhelmed by very low-energy triggers. These triggers are typically much below single e^-h^+ and are mostly due to digital electronics sources such as cellphones and computers. Contrasting with these electronic noise triggers, when approaching the e^-h^+ quantizatioin threshold and above, the electronics background is largely suppressed, leaving the data dominated by quantized events that really indicate particles creating phonon pulses in the detector. We would like to reject transient pick-ups causing the very low-energy triggers, hence the first-layer overall trigger burst rejection. However, with the data dominated by these triggers, the selection algorithm would effectively ignore the contribution from quantized leakage current events, where series of consecutive triggers are generated by γ or muons that destabilize metastable atomic states. We therefore apply the second-layer selection in the leakage current-dominated regime after the overall trigger burst selection. According to the stage-0 analysis development data, we expect a 10–11% live time loss due to the trigger burst rejection, and we will present the final value in the unblinding discussion (Sec. 2.7).

2.5.3 Baseline (pile-up) livetime selection

Next we remove extended periods of time when the signal traces are distorted by the trailing preceding pulses. Due to pulse shape distortion when occurring on a tilted trace baseline (tail of the preceding pulse), these events' OF-reconstructed energies are expected to be systematically biased by the nonrandom pulse shape perturbation relative to the input templates (Sec. 2.3). To quantify and select the pre-pulse baseline flatness, we first calculate the "mean baseline" reduced quantity `MB` that is the averaged trace height before the trigger, i.e., from 675 μs before to the trigger point. Fig. 2.10 shows an example `MB` timestream from an unblinded continuous laser dataset. With the dense laser data points profiling `MB` at 100 Hz, we see a series of large background events that last for a few to ten seconds. We therefore decide to bin the livetime to be selected by 1 second, which is not only suitable for identifying these large pileup-generating events but also compatible with the 1s-unblinded DM data. We calculate the mean of `MB` for the events contained in each 1s time bin,

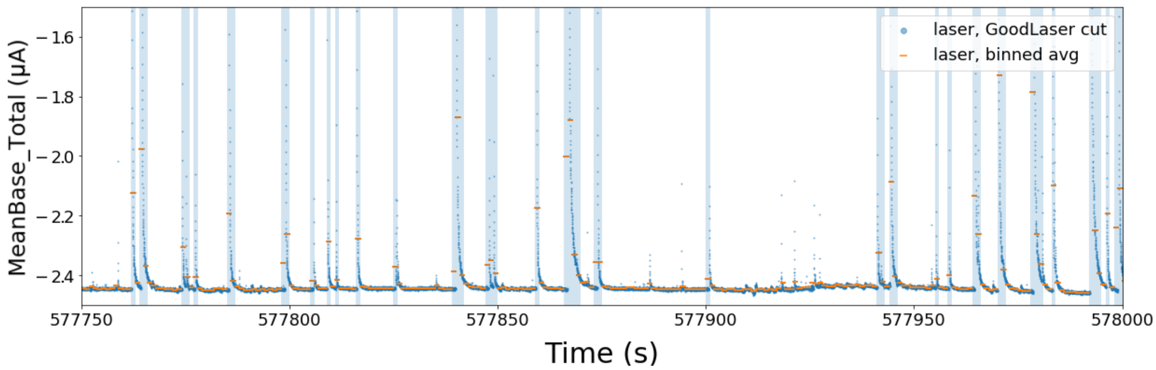


Figure 2.10: An example of pre-pulse baseline height timestream of events taken from a laser dataset, which samples the value at 100 Hz in time. The blue dots and orange bars represent, respectively, the values of individual events (`MB`) and their averages in the 1-sec. time bins (`MMB`) used for the selection. The shaded time periods mark the livetime exhibiting high `MMB` that are subsequently rejected. The "GoodLaser cut" noted in the legend indicates that, to improve data quality, we use only the laser events correlated with the TTL signal generated by the laser pulse generator in this figure.

called `MMB`, to represent the local baseline height in each second. Due to relocking the SQUID feedback loop dataset to dataset, the absolute signal baseline level as read out by the current signal naturally varies. So, in order to make the datasets more comparable for a simpler data selection later, we subtract the nominal `MB` value due to relocking the SQUIDs for each dataset from `MMB`, which we call Δ`MMB`, and use it for the following analysis. Fig. 2.11 shows the results, where the distributions indeed exhibit tails toward higher values as expected by the heightened baseline. However, Fig. 2.11 also shows the laser datasets exhibit significantly different Δ`MMB` values relative to their DM search data, as quantified by the Gaussian fits for the laser Δ`MMB` distributions. It is currently inconclusive on the cause for the laser and their corresponding DM search datasets to exhibit such differences in Δ`MMB`, especially when the sign and the size of the difference are not consistent over time. We note the laser/DM dataset pairs were acquired back-to-back within minutes when the readout system, SQUID locking included, should have already been fully fixed in advance.

The phenomenon causes a major concern for the applicability of our laser-based energy scale calibration. To examine its effect, we explicitly compare each laser-DM data calibration pair and confirm the quantized $e^{-}h^{+}$ peak widths and positions are indeed consistent, hence still justifying the application of the quantized laser spectrum to calibrate the DM search. However, we do find from the laser-DM comparison work that there is one dataset, acquired on May 5$^{\text{th}}$, 2019, that shows

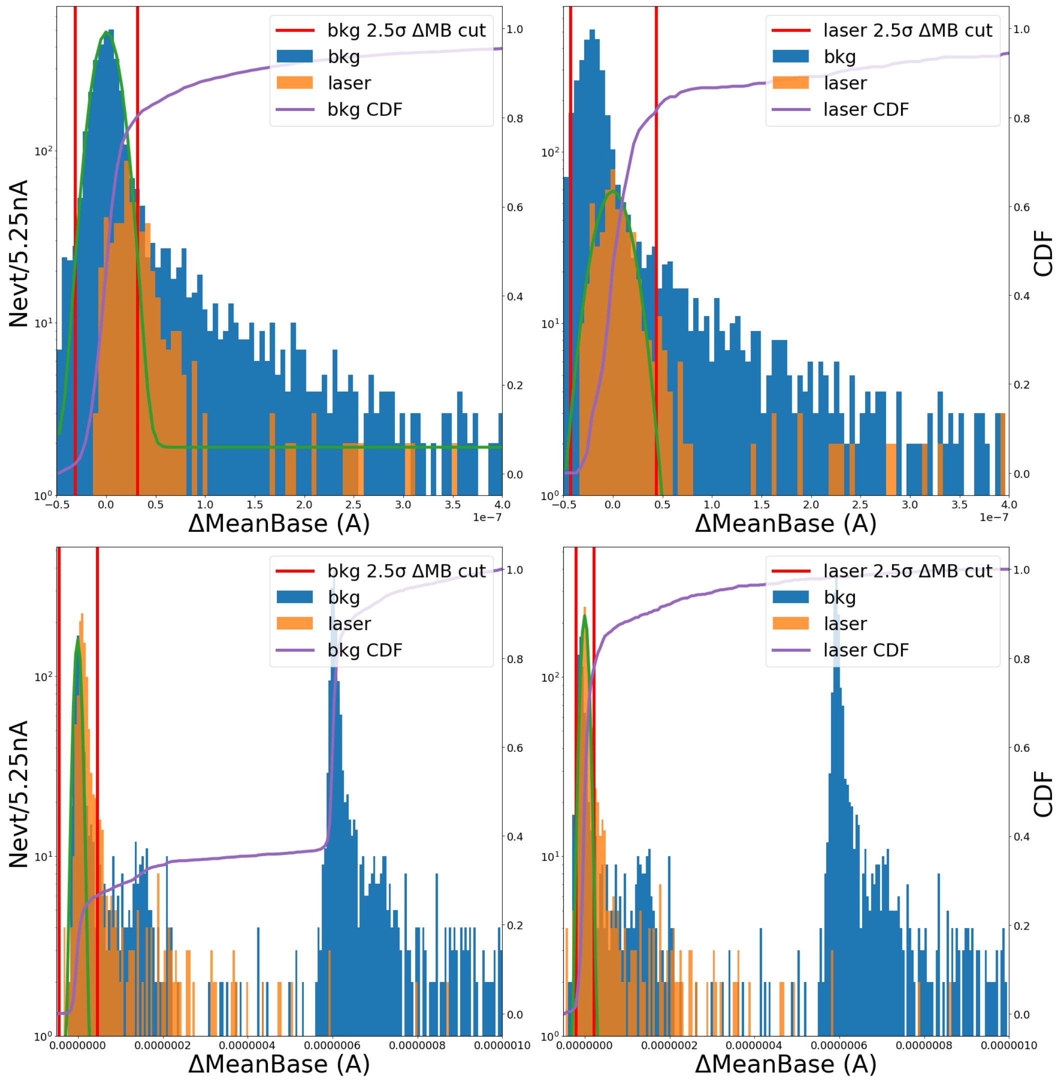


Figure 2.11: The Δ`MMB` distributions taken from a "normal-day" dataset, shown in the top row, and the SQUID-jumped day, shown in bottom row. The blue and orange histograms represent the DM search data and the laser data, respectively, and are plotted identically in the same rows for comparison. The green curves show the Gaussian fits to the data, for the DM search data on the left and the laser on the right, with the vertical red lines marking the $\pm 2.5\sigma$ selection windows corresponding to the Gaussian fits in the same panels. In each panel, there is a purple curve (right axis) showing the CDF for the readers' comparison to the selection window.

an apparent "splitting" to its quantized $e^- h^+$ peaks in the DM search spectrum. The splitting is shown in Fig. 2.12. When we compare the laser spectrum to the split DM search spectrum acquired on this day, we find the laser peaks align to one of the two split copies in the DM search spectrum, leaving an apparent misalignment to the other copy. When we examine the system temperature record, it becomes clear

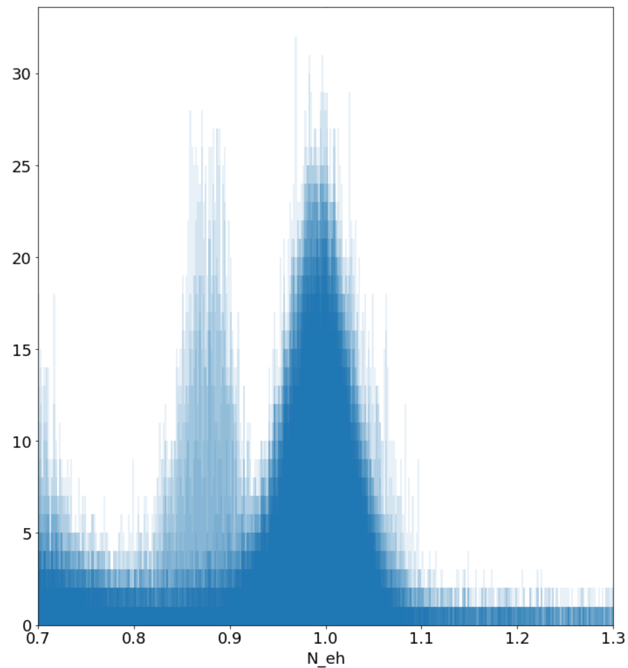


Figure 2.12: A dataset-to-dataset calibration consistency visualization by stacking all the semi-transparent post-calibration stage-0 spectra on to each other. The deep-color single e^-h^+ peak indicates the calibrated quantizations across datasets are nominally consistent therefore align the single e^-h^+ peak. The misaligned light-color peak centering at $\approx 0.9 N_{e/h}$ suggests it is contributed by much less data. The same splitting applies to multi-e^-h^+ peaks as well but not shown here for visual clarity. Detailed explanation is provided in the corresponding text.

that a large `MB` (Δ`MMB`) jump occurred at about 3 hours into the DM data acquisition, when an atypically large temperature runaway occurred simultaneously. Although the temperature was later recovered by the PID control, the large `MB` jump never recovered. Fig. 2.12 unambiguously proves that the incident permanently altered the energy calibration until the end of the dataset, which forces us to discard the entire post-jump dataset from the analysis due to its uncalibratibility.

In fact, we will show later when introducing the temperature livetime selection that, since we allow the ADR to run until losing base temperature at the end of each dataset, it is typical that hours-long of data without proper detector condition were always recorded toward the end of each dataset. With the temperature quickly rises, `MB` always drifts in accordance due to TES bias condition change, and at some point the drifting signal baseline would always jump to another value discontinuously. It is the well understood "SQUID jump" phenomenon due to injecting an excessive current that drives the SQUID to the outside of the local flux quantization, resulting in the SQUID recovering to the new local minimum like a baseline "jump." If such a phenomenon occurs and the SQUID is immediately re-locked by the feedback control, the SQUID transfer function is randomly readjusted, so is the resulting gain and therefore the energy calibration. This scenario is exactly the cause to the May 5th incident. We define the selection criterion for the proper-baseline livetime after the data are inspected and understood. Shown in Fig. 2.11, we fit to each DM search dataset a Gaussian function to identify the $\pm 2.5\sigma$ region around the nominal Δ`MMB` value and discard the 1-sec. time bins falling to the outside of the selection window. While it is not straightforward to visualize the effect of the selection with the largely blinded DM search data, one can see clearly from the laser `MB` timestream (Fig. 2.10) that the selection effectively rejects the time periods contaminated by trailing large events.

We hypothesize that the the large events rejected by the baseline selection are predominantly muons since we are in a surface laboratory. To verify this intuitive hypothesis, we estimate the muon flux for the HVeV R2 detector. We use the muon flux measurement provided in (Shukla et al., 2016), where the flux for muon at energies greater than 0.5 GeV is given by

$$\Phi(\theta) = I_0(\cos\theta)^{2.01\pm0.03}, \tag{2.22}$$

where θ is the zenith angle, and

$$I_0 \approx 85.6 \pm 2.4\ [\mathrm{m}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}]. \tag{2.23}$$

We estimate the muon "in-"flux through each vertical or horizontal surface by

$$\begin{aligned}
\Phi_{\text{hori.}} &= \int_0^{2\pi}\int_0^{\pi/2} \phi_0 \cdot \cos\theta \cdot \sin\theta d\theta d\phi \quad \approx 134 \quad [\mathrm{Hz/m^2}] \\
\Phi_{\text{vert.}} &= \int_0^{\pi}\int_0^{\pi/2} \phi_0 \cdot \sin\theta \cdot \sin\theta d\theta d\phi \quad \approx 53 \quad [\mathrm{Hz/m^2}].
\end{aligned} \tag{2.24}$$

Note the range of integration for the vertical surface is for each surface, for which we multiply the result by 4 to account for the total muon influx for the detector. With Eq. (2.24) and given the $1 \times 1 \times 0.4\ \mathrm{cm}^3$ geometry of our detector, when the normal direction of the instrumented $1 \times 1\ \mathrm{cm}^2$ surface is oriented parallel to the ground as in the experiment, we estimate the muon in-flux for the detector to be

$$\Phi_{\text{HVeV R2}} \approx 0.02\ \ \text{Hz (estimate)}, \tag{2.25}$$

while we obtain with the `MB` livetime selection algorithm a $\approx$0.12-Hz occurrence of extended `MB`-elevated periods (Fig. 2.10). The discrepancy suggests the large events

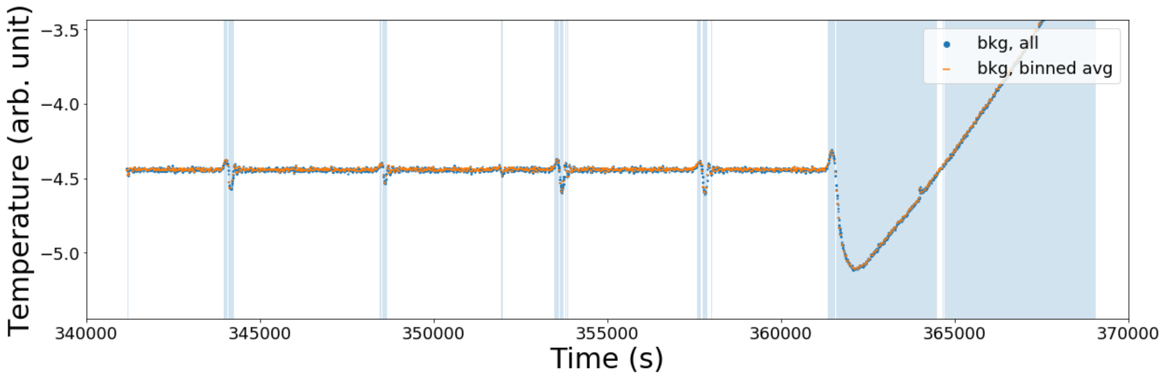


Figure 2.13: A typical temperature timestream data taken from a stage-0 dataset. The blue dots and orange bars represent, respectively, the values of individual events and their averages in the 30-sec. time bins used for the selection. The blue-shaded time periods mark the livetime exhibiting high temperature deviations that are subsequently rejected.

are not dominated by just direct muon incident but possibly also muon-generated secondaries and/or other environmental radioactivities. The result may serve as a reference for future HVeV runs when the experiments move to underground sites.

2.5.4 Temperature livetime selection

For the last parameter in the livetime selection series, we consider the temperature of the detector, a critical factor that defines the electro-thermal bias condition for the TESs. We first take a general inspection to the typical temperature profile of our experiment, shown in Fig. 2.13. We find the temperature is generally controlled to within sub-% fluctuations, a fact we have already invoked in the energy calibration (Sec. 2.4). However, aside from this nominal condition with a small temperature variation, we may also see the temperature profile exhibits periodic spikes at roughly every 1.5 hours, giving up to 5% variations relative to the nominal temperature and may persist for 5–10 min. each time. Toward the end of each dataset, the structure where the temperature first decreases by more than 10% and then starts to rise indefinitely always presents, marking the exhaustion of the ADR cooling power.

Based on the above general pattern, we develop a two-steps temperature selection similar to the trigger burst, each targeting an undesirable quality of the data at its unique time scale. We first identify and remove the long-time structures where the detector is apparently under a close-to-malfunction condition, including the PID temperature control reset occurring roughly every hour and the end-of-dataset loss of base temperature. After these unusable time periods are excluded, we then try to

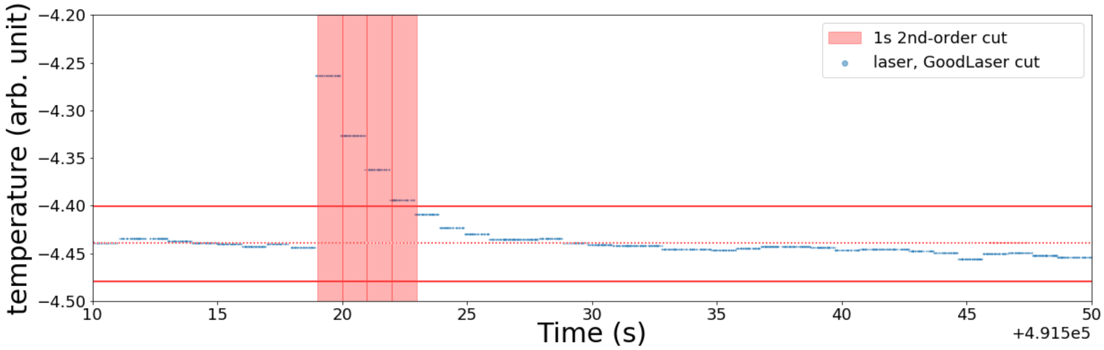


Figure 2.14: A typical temperature fluctuation rejected by the fine-binned temperature livetime selection. Unlike Fig. 2.13, this figure is produced from the laser data, because at stage-0 the DM search data are too sparse to visualize the fine-binned effect. The blue dots represent the values of individual events, which show the stair-like 1-Hz temperature reading refresh, while the red vertical bands mark the time periods exhibiting high temperature deviations that are subsequently rejected. The dotted and solid red horizontal lines represent the nominal temperature and the boundaries of the selection window, respectively.

improve the data quality by finely select on the random fluctuation of the temperature and reject high deviations with respect to nominal values. In terms of a practical implementation, we adapt the same methodology for the baseline selection, where we first bin the data in time, calculate the average temperature in each time bin, and from the distribution of these binned average values, we extract the standard deviation and reject the bins falling to the outside of the $\pm 3\sigma$ window.

We first bin the data by 30 seconds, a time scale that is comparable to the detector's characteristic thermal response time but is much shorter than the PID control resets. The binning is expected to be able to identify the ADR-related detector temperature drifting but otherwise insensitive to random temperature fluctuations. Indeed as shown in Fig. 2.13, the algorithm effectively identifies the targets, namely the hourly shaded temperature-resetting times and a large portion at the end of the dataset. For the fine selection, in principle, it would be the most precise if we could perform the selection on an event-by-event basis. Unfortunately, our data acquisition system was designed to update the temperature reading only at every second. Since all the events occurring during every 1-sec. time period are assigned with the same temperature due to the 1-Hz temperature update, the selection provide no further precision whether it is performed event-by-event or as a livetime selection with a bin size equal to or narrower than 1 sec. On the contrary, if in this case we still choose to select the temperature event-by-event and estimate the selection efficiency with

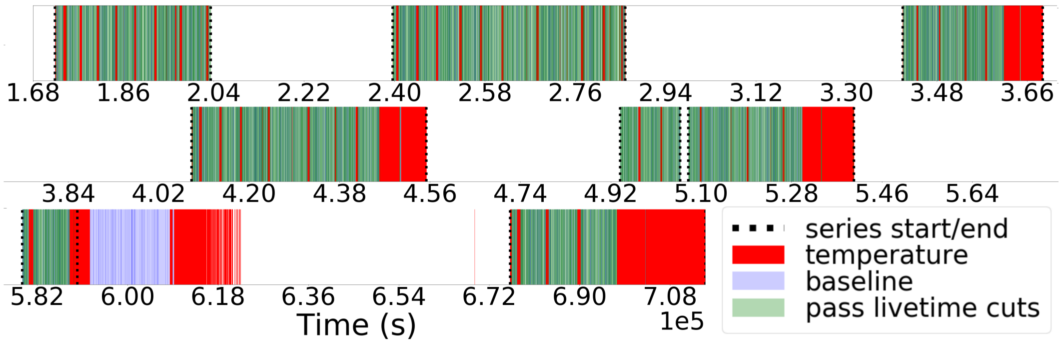


Figure 2.15: To be read from top-left to bottom-right, the raw DM search data livetime marked by the selected qualities. Based on the selection sequence introduced in the text, red, blue, and green indicate the time is first rejected by the temperature selection, the baseline selection, or passes the selection as the final livetime-selected data for limit-setting, respectively. Note that, due to the narrowness of the blue stripes, they might appear to some readers as darker green stripes on the nominal green regions.

the laser data, which were acquired at very different event rates compared to the DM search data yet both assigned with Hz-updated temperatures, combining with the complexity due to then 1 second-binned unblinded data, we worry that the selection may introduce an unquantifiable discrepancy thus an unnecessary extra uncertainty. We therefore choose to apply the fine temperature selection as a livetime selection binned synchronously to the 1-Hz temperature update, i.e., not only at 1 Hz but also synchronized to the exact seconds according to the DAQ design, which is the most accurate option permitted by the data structure. Despite the somewhat misplanned temperature data acquisition, Fig. 2.14 shows that the temperature fluctuations that may practically exceed the 3σ rejection threshold are typically at multi-seconds time scales hence identifiable by the 1-sec. binning. We may also see in Fig. 2.14 that the really trend-less rippling of the temperature profile always remains within the 3σ-allowed range, where our temperature-based energy calibration is effective and reliable.

2.5.5 Livetime-selected data

It is difficult to quantify and present the livetime loss due to each livetime selection in a simple independent fashion because, first, it is common for multiple selection criteria to identify and subsequently discard the same time periods due to shared causes/symptoms, such as the loss of temperature control and the consequential mean baseline shift. Second, the livetime loss estimated at this stage by the heavily blinded data, especially with the choice of a time-domain (un-)blinding scheme, may be substantially biased depending on non-random temporal patterns of the physics origins for the vetoed qualities. Nevertheless, we still provide a rough illustration for the relative "additional" effects for the selections using a specific order of the selections, from which we may understand the extra differences based on the principles of the selections. As an example, we first apply the wide-bin temperature selection to identify the periods with uncontrolled temperatures, and then by adding the baseline selection, we further mark the muon/background particle pile-ups in addition to the temperature loss. We then apply the fine-bin temperature selection to further reject random temperature fluctuations for an improved calibration quality. Finally, we apply the trigger burst selection that is expected to be largely uncorrelated with previous criteria.[6] Based on this particular sequence, Fig. 2.15 visualizes the livetime selection with the sequential livetime reduction summarized in Tab. 2.1. Fig. 2.15 shows that the wide-bin temperature livetime selection presents apparent red stripes that mark the hourly temperature resets and the large end-of-dataset removals, effectively defining the active time of the experiment. Numerous narrow blue stripes blend in the green active-time regions to denote the pile-ups, with an exception that, for the May 5$^{\text{th}}$ dataset (6$^{\text{th}}$ from the start), we may actually see a much wider hourly temperature reset, and then a following `MB`-rejected blue region appears to indicate the data exhibit a properly recovered temperature but nonetheless are rejected by the offset high `MB`. Such a combination is not seen in any other red regions because of simultaneous temperature and `MB` rejections, but it should be easily understandable based on our previous explanation.

Tab. 2.1 summarizes Fig. 2.15 in numbers, where we find that, due to the ADR temperature control pattern, about 1/3 of the full data collected are unusable. And by comparing the data before and after the baseline selection, we realize the pile-up contamination roughly occupies 22–24% of the proper detector working time. Finally, after the fine temperature selection rejects unnoticeable amounts of livetime as calculated on the small stage-0 data, we acquire an 8% further reduction due to trigger burst rejection. We should emphasize that, while the values quoted in Tab. 2.1 may vary depending on the sequential application of the inter-correlated selections, the final livetimes are correct regardless of the order of the selections. We obtain 3.6 hours and 1.2 hours of livetime-selected exposures at 100 V and 60 V HV biases, respectively, for the 10%-unblinded stage-0 datasets. We will continue using these datasets for the development of the event-based selections to further

[6] Probably still correlates with the pile-up removal, especially the leakage trigger bursts. But again it only indicates the livetime selections are correlated.

	100 V		60 V	
	hour	%	hour	%
Raw livetime (stage-0)	7.8	100.0	2.5	100.0
Wide-bin temp.	5.2	66.8	1.7	68.4
Baseline	3.9	50.8	1.3	53.4
Fine-bin temp.	3.9	50.8	1.3	53.4
Trigger burst	3.6	46.6	1.2	49.9
Stage-0 final	3.6	46.6	1.2	49.9

Table 2.1: The sequential livetime reductions for the 100-V and the 60-V stage-0 DM search data.

reject background-like events and discuss the unblinded results afterward.

2.5.6 OF pulse time selection

We continue to develop the event-based selections with the livetime-selected data. We use the laser data as the proxy for the anticipated ER DM signal, defining with them the criteria for a reliable signal reconstruction and correspondingly the selection criteria. To ensure consistency, identical livetime selections are applied to the laser data prior to using them for developing the event-based selections. In addition to applying the identical trigger scheme, the OF reconstruction (Sec. 2.3), and the livetime selections, we further select and use only the laser data that are synchronized with the TTL signals sent from the laser driver directly to the DAQ system. The TTL tagging provides an effective background trigger rejection hence improves the reliability of representing the DM events with the laser data (Fig. 2.16). With the laser proxy data defined, we straightforwardly determine the ranges of parameters that distinguish DM candidates from background-like events using the laser data distributions. We accept only those events in the DM search data satisfying the criteria for the final limit-setting. The parameters we utilize include, applied in the written order: 1) The pulse time calculated by OF, 2) `MB`[7], 3) the χ^2 statistic for the OF calculation, and 4) correlated telegraph noise detection.

In this analysis, we use a specific version of the OF algorithm to obtain the pulse height (energy) estimates, where we allow the algorithm to not only vary the pulse amplitude to minimize the χ^2 statistic but also to very the relative offset in time between the trigger point given by the aforementioned trigger algorithm and the laser-based pulse template. The readers may find in-depth discussions for the mathematical basis of our OF calculation in (S. Golwala, 2000) as well as later

[7] Defined in 2.5.3.

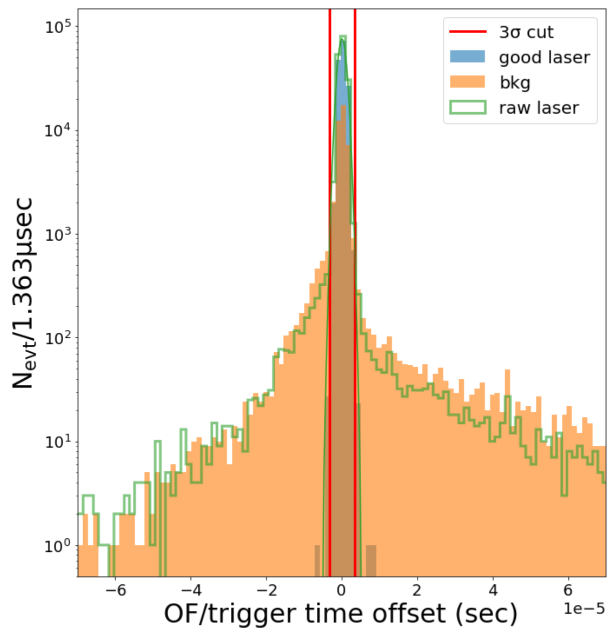


Figure 2.16: The OF-determined pulse time relative to the trigger time. The blue shaded histogram and the green curve are the TTL-tagged and the raw laser data, respectively. The orange shaded histogram is the stage-0 DM search data. Without TTL tagging, the raw laser and the stage-0 DM search data are triggered and selected identically, and they show consistent shapes except the rate for the centered peak is higher for the laser data due to a higher event rate. The vertical red lines mark the $\pm 3\sigma$ selection window determined with the TTL-tagged laser data distribution.

in this thesis (Sec. 4.6). Although we will explain in Sec. 4.6 that the temporal cross correlation in our trigger algorithm is mathematically equivalent to OF, the simplified single-pole pulse shape that we use for the triggering is only precise enough for identifying pulses but not enough for a satisfying pulse height/energy calculation. Due to the modest but well-identified pulse shape variation dataset-to-dataset and also in an energy dependent fashion, especially for smaller signals, the most precise pulse time may vary noticeably with respect to that determined by the approximate triggering pulse shape. We find, if we allow a slight variable time offset in our OF calculation to accommodate the difference, the OF may further improve the χ^2, indicating a better representation for the detected signal. In this case the OF yields more precise pulse height and time estimates. In fact, due to small pulses' higher tendency to be distorted by preceding events, noise, or other not yet understood detector physics, etc., allowing the time offset has been proven to be especially beneficial for the energy resolution of small signals, which is our main interest and therefore is critical that we adopt the technique.

However, the additional degree of freedom may also introduce a bias when the pulse time and subsequently the amplitude are erroneously estimated in the presence of a pulse shape distortion, e.g., tilted baseline due to a pile-up event or simply identifying a separate event that happens to occur in the same trace window. Fig. 2.16 shows a comparison of the OF-determined pulse time obtained from the laser data and the stage-0 DM search data, presented with respect to the trigger time as offsets. The "raw laser" data in the figure represent the full laser dataset, and the "good laser" data represent the laser data that correlate with the TTL tagging. The well-defined good laser data exhibit an anticipated Gaussian distribution, with an 1.1 μsec. standard deviation (σ) that is consistent with being dominated by the DAQ timing resolution, i.e., the 1.5 MHz sampling rate. Without enforcing the TTL tagging, the raw laser data exhibit a wide decaying distribution, indicating the trigger algorithm alone is not robust enough to reject erroneous triggers at higher trigger time offsets. The DM search data ("bkg" in Fig. 2.16) show a similar spectrum to the raw laser data, which includes a centered peak consistent with the good laser data and a broad decaying continuum due to erroneous triggers. We therefore choose to reject data exceeding $\pm 3\sigma$ in the trigger-OF time offset defined by the good laser data as being misreconstructed due to the temporal degree of freedom allowed in our OF.

We should state that, although we have argued that including this degree of freedom is helpful in terms of small-signal energy resolution, due to the degrading OF timing resolution at smaller signal-to-noise ratios, imposing such a selection in time is expected to also introduce a soft energy-dependent detection threshold. We will rigorously quantify the effect when we introduce the efficiency for the event-based selections, in which we obtain in the end a 30–50 eV soft data threshold, which is raised from the 30 eV DAQ triggering threshold due to the combined effect from all event-based selections. Since we aim to utilize the $e^{-}h^{+}$ quantization for the DM search, which begins at 100 V owing to the applied HV, we conclude the selection threshold is effectively negligible and post-justifies the choice of the $\pm 3\sigma$ selection window for this OF-trigger timing selection.

2.5.7 Baseline selection

We proceed to select the data on an event-by-event basis for rejecting distorted pulses, again quantified by the `MB` reduced quantity relative to the nominal SQUID-locked values dataset-to-dataset. As the majority of the events exhibiting unacceptable `MB` has already been rejected in the baseline livetime selection, we anticipate the remaining data to show distributions concentrating at their nominal values, as opposed

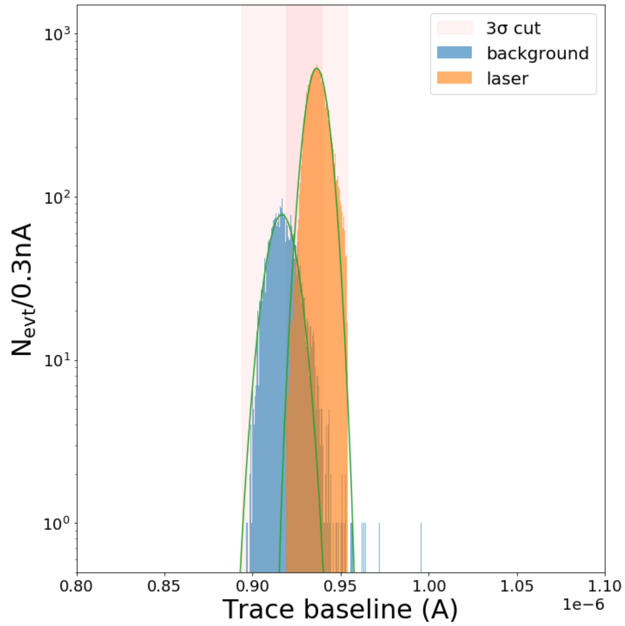


Figure 2.17: The pre-pulse mean baseline height (MB) of a laser/DM dataset pair. The orange and blue shaded histograms are the TTL-tagged laser and the stage-0 DM search data, respectively, with the green curves showing their corresponding Gaussian fits. The red vertical bands mark the selection windows separately for the laser and the DM search data.

to particular structures that are inconsistent with a random fluctuation as seen in the baseline livetime selection. Still, we choose to apply this event-by-event baseline selection to safeguard the data from unexpected potential baseline outliers in the not yet unblinded data. As expected, Fig. 2.17 presents the well-modeled Gaussian distributions and their $\pm 3\sigma$ selection windows for MB. As it turns out, indicated by both the laser calibration data already and later the fully unblinded DM search data, this precautionary selection indeed imposes little impact to the data after the high-MB (MMB) livetime is already excluded from the analysis.

For future reference, there are several reasons to carry out the baseline selection primarily as a livetime selection but not event-by-event. First, using the SQUID jump incident on May 5$^{\text{th}}$, 2019 as an example, we may easily understand that the anomalous baselines may occur as extended special events but not just stochastically due to random pile-up events, even though they are originally the targeted phenomenon to be removed. If the MB selection had been developed only event-by-event without a prior livetime selection while the data are largely blinded, this type of extended occasional occurrences would be difficult to foresee and accounted by the selection efficiency with certainty. Second, and in fact an intriguing new lesson learned from this work, based on the working principle of our OF calculation, it is understandable that the energy reconstruction by the OF is not only randomly smeared when an event correlates with pile-ups or simply distorted, but the reconstructed energy value is expected to be biased in an energy- and pulse shape-dependent fashion. Particularly for our large preceding event rejection scenario, we realize, if the baseline selection is implemented as an event-based selection, the selection would predominantly target a specific class of pulse shape distortion, namely the “tilted” pulse on a sloped baseline. Due to this systematic but not random pulse shape distortion, combining with the correlation between OF energy (mis-)reconstruction and the observed pulse shape, the selection may then correlate with the reconstructed energy that in turn leads to an energy-dependent artificial bias in the selection efficiency. We find, with our detector’s outstanding energy resolution, we may indeed start to resolve a nontrivial dependence in `MB` for the OF-reconstructed energy. Although qualitatively understood, we find this dependence difficult to quantify or predict from first principle and is obviously unphysical; we will revisit this issue with more detail when we formally define the selection efficiency. In summary, we find that, by varying the tightness of the `MMB` livetime selection applied prior to the event-by-event selections, the size of the artificial efficiency curve structure due to the event-by-event `MB` selection responds in agreement with the remaining high-`MB` events that are not removed by the livetime selection. For future HVeV analyses, we propose explicitly simulating the OF formalism’s reconstruction performance under the influences of practical effects, such as specific noise sources or certain pile-up patterns. For HVeV R2, the above arguments motivates us to tighten the livetime selection for anomalous high `MB` to a relatively more stringent level of 2.5σ, which for the event-based selection allows us to choose a common $\pm 3\sigma$ selection window just to select on potential trendless random fluctuations.

2.5.8 χ^2/DOF (degree of freedom) selection

We next apply selections to the degree of freedom (DOF)-normalized OF χ^2/DOF statistics for identifying the events that resemble the pulse templates in shape. We independently define for each DM search dataset its own pulse template using the dataset’s corresponding laser dataset. The selection is illustrated in Fig. 2.18. To begin the discussion, we first acknowledge the apparent fact that our OF χ^2/DOF calculation is not properly normalized and thus yields non-unity χ^2/DOF values for proper minimizations. It is nevertheless still usable in practice for defining DM candidates based on the χ^2/DOF distributions of laser given by the the same

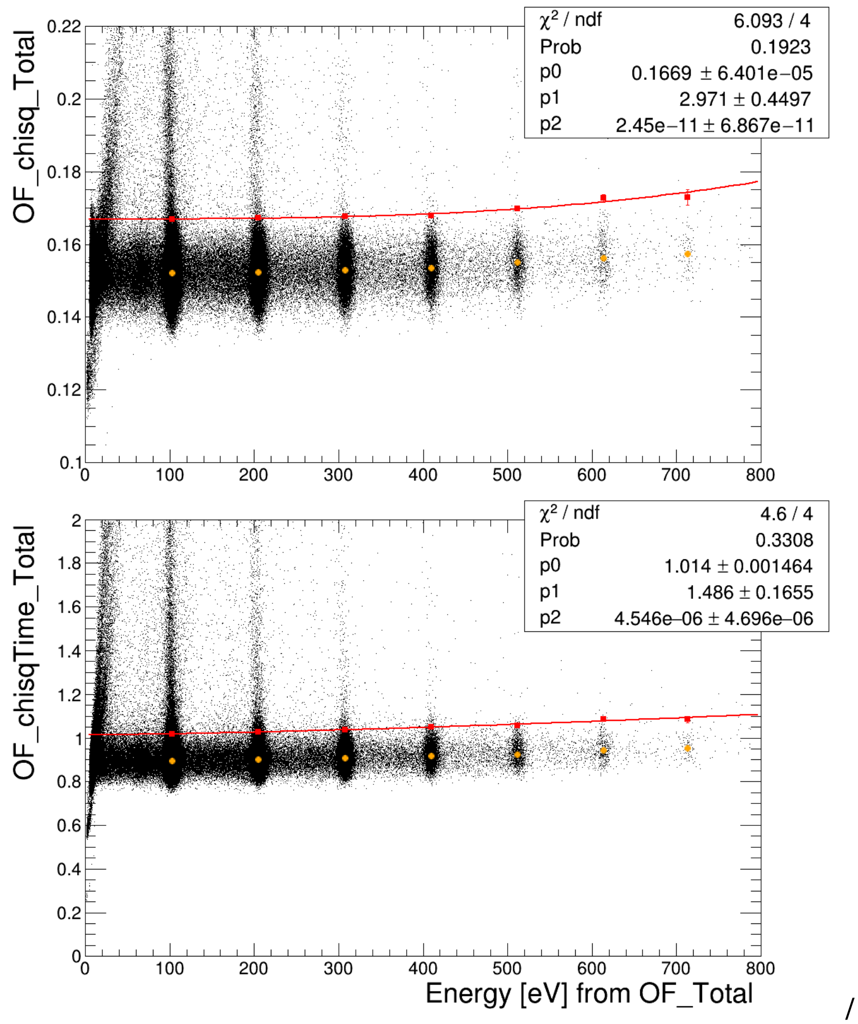


Figure 2.18: The degree of freedom (DOF)-normalized frequency- (top) and time- (bottom) domain OF χ^2 for the laser data plotted against the calibrated energy. The orange and the red dots represent the $e^- h^+$ peaks and their corresponding $+3\sigma$ values in χ^2/DOF, respectively, with the red solid curves showing the 2^{nd}-order polynomial fits to the $+3\sigma$ points. Detailed description for the generation of this figure is provided in the corresponding text. Figure produced by V. Novati.

algorithm. For this selection, we utilize the χ^2/DOF in both the frequency and time domains. The former corresponds to the minimization that provides the major pulse parameters used throughout this analysis, i.e., height and time. We also directly evaluate the trace difference between the signal and the template in the time domain, which, relative to the frequency-domain χ^2/DOF, provides another signal-template resemblance indicator that is commonly recognized to be more sensitive to pulse shape distortions outside of the major frequency band of the template, such as the very-long time-scale tilted baseline. The readers may see in Fig. 2.18, instead of designating constant values for the frequency- and time-domain χ^2/DOF cuts, 2$^{\text{nd}}$-order polynomials are empirically fitted to represent the continuous $+3\sigma$ upper edges, defined by the discrete $+3\sigma$ values of the e^-h^+ peaks, that are used to reject the events located above the curves. In producing the 3σ selection boundaries, we select the $\pm 2\sigma$ data in energy of each e^-h^+ peak and then extract the $+3\sigma$ χ^2/DOF upper value given by the χ^2/DOF distribution of the selected data. The $+3\sigma$ χ^2/DOF of the first seven e^-h^+ peaks are then used to determine the 2$^{\text{nd}}$-order polynomial cut boundaries. The χ^2/ statistics' increase with energy as described by the fitted curves is the consequence of the pulse shape's dependence in energy, while we define the templates for the OF calculation by averaging the laser data predominately consisting single (and double) e^-h^+ pulses.

2.5.9 Anti-coincidence selection

Finally in the event-based selection series, we apply a coincident detection rejection, which is justified by the low likelihood of acquiring independent DM interactions in multiple detectors in the same event time frame, considering the existing DM interaction cross section limits and the exposure of this work (Sec. 2.2). The coincidence detection is achieved using another TES-based detector that is located alongside the main DM detector and is biased and read out identically. This coincidence veto detector is designed to be a R&D variant for the low-threshold SuperCDMS detectors. It is fabricated using the same Si substrate geometry, but the phonon sensor is equipped with a mm-size coiled bias line that is identical to the common square-spiral RFID (radio-frequency identification) antenna. The design makes the TES particularly sensitive to environmental RF/microwave pick-up, thus proving to be useful for monitoring the telegraph noise for other detectors but not an improved DM detector by itself. Using artificial telegraph noise sources such as cellphones close to the ADR, we demonstrate that the veto detector can sense the RF pick-up by outputting square waves that reflect the artificial interference's signal encoding, while during these test periods the DM search detector also exhibits elevated trigger rates in a perfect coincidence. Originally aiming for a lower T_c to improve the TES energy resolution, the W film recipe for the veto detector is also adjusted to give $T_c \approx 52$ mK, which is too close to the designated operation temperature for the main DM search, so while the detector is able to be properly electro-thermally biased into a stable transition, it exhibits a very limited dynamic range that only permits detection but not in a practical energy-resolving manner. Fig. 2.19 shows the histogram for the $\mathtt{OF}_{\mathtt{tot}}$ pulse sizes observed in coincidence with the laser and DM candidate

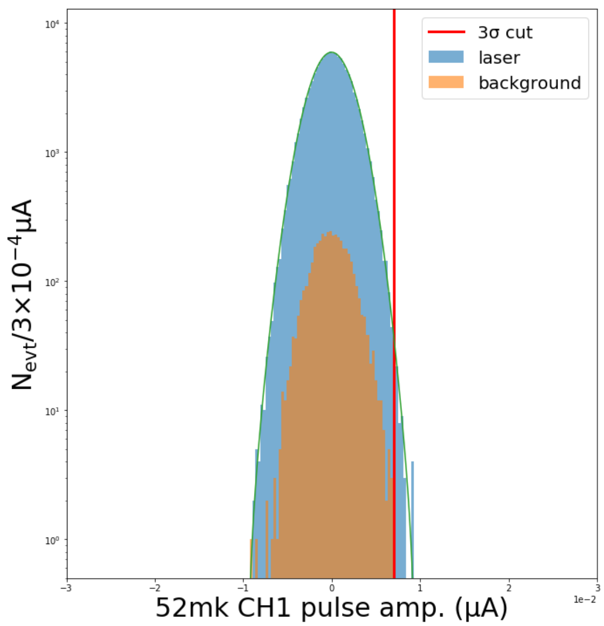


Figure 2.19: The uncalibrated raw OF signal size ($\mathtt{OF_{tot}}$) in the telegraph noise monitoring detector observed in coincidence with the main detector triggers. The orange and blue shaded histograms are for the TTL-tagged laser and the stage-0 DM search data triggers, respectively, with the green curve showing the corresponding Gaussian fit, and the red vertical line marking the selection threshold.

data. The values are presented in the raw DAQ unit without calibration since, as explained above, we are only interested in their ability to signal coincidences but not the physical size of them. Fig. 2.19 shows, after all the preceding selections, the coincident signal in the telegraph noise veto detector is nothing but a Gaussain random fluctuation around non-detection, as anticipated by a pure noise condition. We therefore choose to reject any main-detector data that coincide with greater than $+3\sigma$ detections ($\mathtt{OF_{tot}}$) in the telegraph noise veto detector. We find the rejection is consistent with randomly removing events from the DM search data without an obvious temporal pattern.

2.6 Signal model

2.6.1 Preparation

There are two main components in the signal modeling: The first is to properly estimate the DM signal loss due to all the data selections detailed previously, which is represented by a selection efficiency that is necessary for connecting the expected DM signal to the post-selection DM search data. The second is to simulate a series of the anticipated DM signal spectra based on the assumed DM properties and our knowledge in the detector and the experiment, such as the exposure and the selection efficiency, so we can compare the expected and the obtained spectra post-selection to constrain the DM properties as shown by the data. Besides the selection efficiency, there are a few more input parameters and assumptions needed for generating the anticipated signal spectra. The first and foremost, different DM model assumptions that govern the DM-e^- interaction for depositing observable ER energy into our detector. This part of the modeling also includes assuming a range of different DM particle masses for each model considered, which effectively transfer different amounts of ER energies to the electron via the assumed interaction channel, hence yielding different ER energy spectra. To compare with the data, these ER energy spectra need to be further modified by the realistic detection technique utilized by our detector, namely the quantized e^-h^+ and the subsequent NTL phonon generations. The readers will find in later discussions, in addition to the canonical NTL phonon generation process assumed in the HVeV R1 analysis, we further consider in this work a more realistic e^-h^+ transport model that includes the charge trapping (CT) and the impact ionization (II) effects. The improvement leads to a substantial difference in the expected signal spectra as well as the final DM limits. After the detected and all the expected signal spectra are prepared, we use the detected spectrum to constrain the likelihood of the data being produced by DM for each assumed model, therefore subsequently constraining the DM parameter assumed for producing each model's expected spectrum. Finally, all these constrained DM parameters are collected to be the final DM ER interaction limits reported in this work.

2.6.2 Selection efficiency and data spectrum

Since we assume the laser photon signal stands for a good proxy of the expected DM ER signal and have developed all the selections based on its properties, we continue to use the laser data to quantify the DM candidate loss due to our selections. It is represented by the efficiency ϵ for preserving good DM candidates or laser events by

$$\epsilon_i(E) = \frac{N_i(E)}{N_{i-1}(E)}, \tag{2.26}$$

where the dependence E denotes that the quantities are calculated for the energy bin $E \sim E+dE$, i labels the i^{th} selection, and N_i is the number of the remaning laser events after the i^{th} selection. Assuming the finite-size sampling of the events falling in each energy bin follows the binomial distribution, we may estimate the statistical uncertainty for the binned efficiency in Eq. (2.26) by

$$\sigma_{\epsilon_i} = \sqrt{\frac{\epsilon_i(1-\epsilon_i)}{N_{i-1}}}, \tag{2.27}$$

where σ_{ϵ_i} is the standard deviation for a near-Gaussian ϵ_i distribution if the sampling is performed $\gg$1 times (Thomason, 2015); we suppress the (E) dependence in the above equation for readability. Fig. 2.20 summarizes the progressive impact of the livetime and event-based selections, and Fig. 2.21 shows the corresponding efficiency curves given by the laser spectra in Fig. 2.20 and Eq. (2.26 / 2.27).

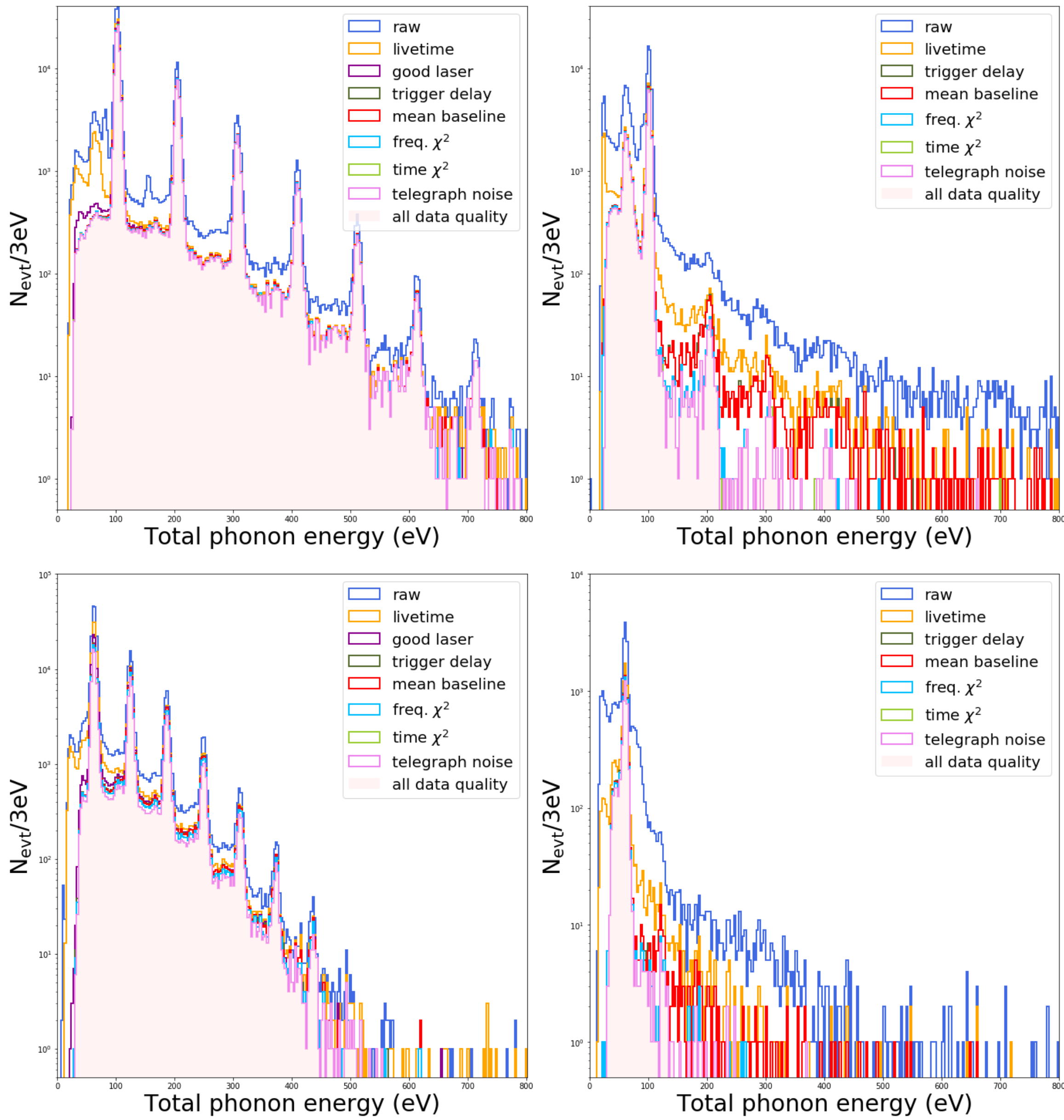


Figure 2.20: Selection-by-selection spectra for the laser (left) and stage-0 DM search (right) data, taken at 100 V (top) and 60 V (bottom) HV biases. For every panel, the histograms from top to bottom correspond to all raw data, livetime-selected, TTL-tagged (laser only), event-by-event selected up to OF time, `MB`, frequency-domain χ^2, time-domain χ^2, telegraph noise coincidence, and finally the pink-shaded fully selected DM search data.

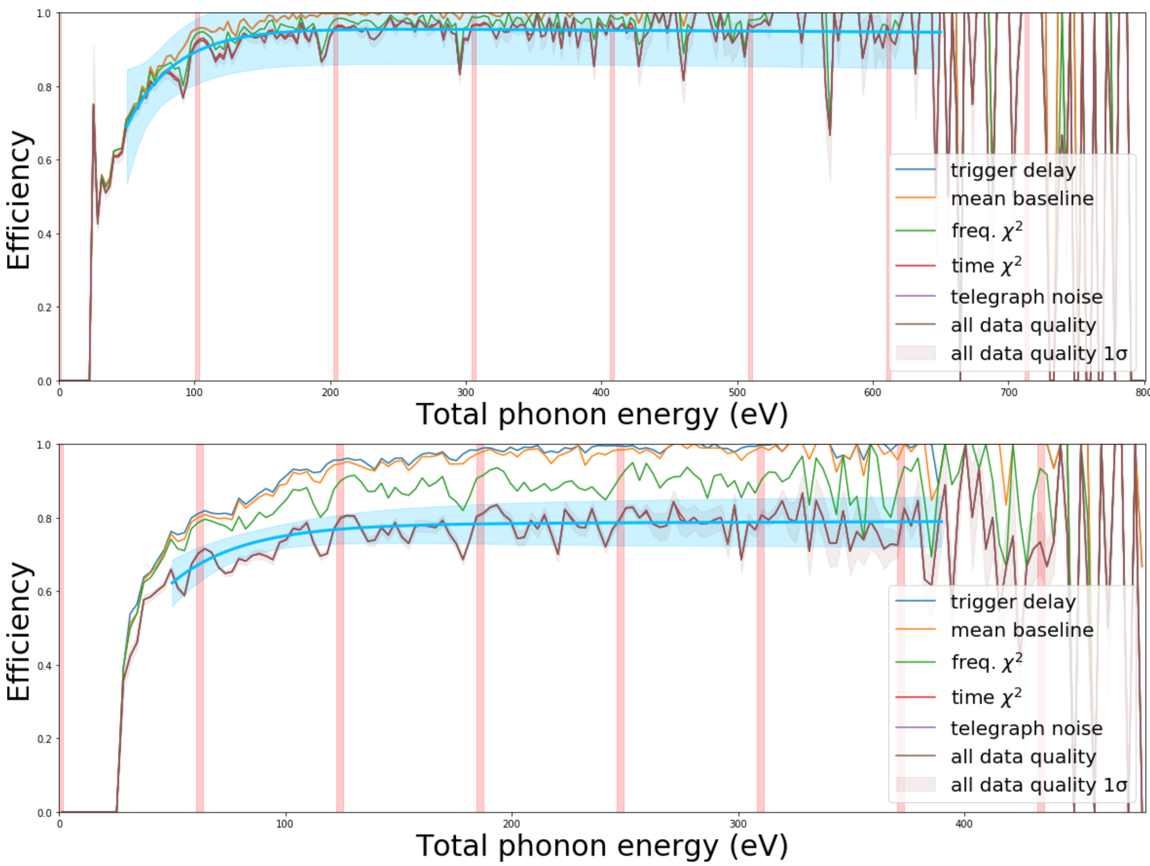


Figure 2.21: Selection-by-selection efficiency curves for the 100 V (top) and 60 V (bottom) data. For every panel, the curves from top to bottom correspond to being selected up to OF time, `MB`, frequency-domain χ^2, time-domain χ^2, telegraph noise coincidence, and finally the fully selected limit-setting efficiency with its shaded 1σ uncertainty range based on Eq. (2.27). The blue smooth curves represent the efficiency model fits using Eq. (2.29), with the attached blue shaded regions representing the uncertainty ranges for the fits as described in the corresponding text. To help with interpreting the efficiency curves, we mark the expected quantized e^-h^+ energies by the evenly spaced vertical red lines, which have the width of the detector's rough phonon energy resolution of 3 eV.

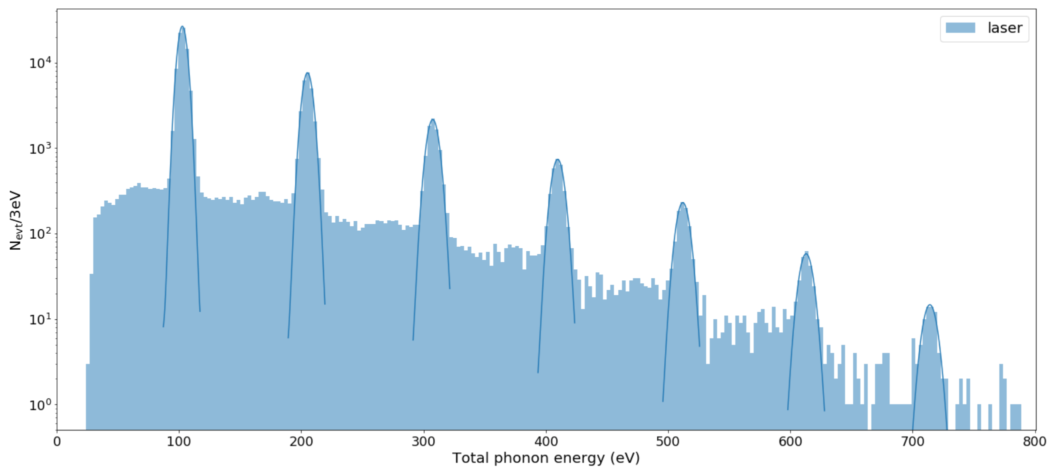
2.6.3 Final energy resolution

From the post-selection laser spectrum, we obtain by Gaussian fits

(eV)	1st	2nd	3rd	4th
peak	102.69 ± 0.01	205.07 ± 0.02	307.53 ± 0.03	409.69 ± 0.03
σ	3.52 ± 0.01	3.95 ± 0.02	4.59 ± 0.03	4.66 ± 0.03
	5th	6th	7th	
peak	512.03 ± 0.04	612.88 ± 0.06	714.08 ± 0.06	
σ	4.91 ± 0.04	5.13 ± 0.06	5.39 ± 0.06	

Table 2.2: Post-selection data resolution, i.e., spectral widths for the quantized e^-h^+ peaks.

for the e^-h^+ peaks at a 100 V HV bias. These "data resolutions" are later used as input parameters to determine the quantized windows in energy for the DM searches. We also obtain the data resolutions for the 60 V data as well as the resolutions given by the distinguishable peaks in the DM search data. Since we confirm all the resolution values are consistent between different data types given their fitted peak widths (σ), and in practice only the 100-V laser-data resolutions are used for later DM searches, we summarize the resolutions with Fig. 2.22 and Fig. 2.23 and refer the interested readers to the complete resolution data tables similar to Tab. 2.2 in (Chang, 2019a) (SuperCDMS internal).



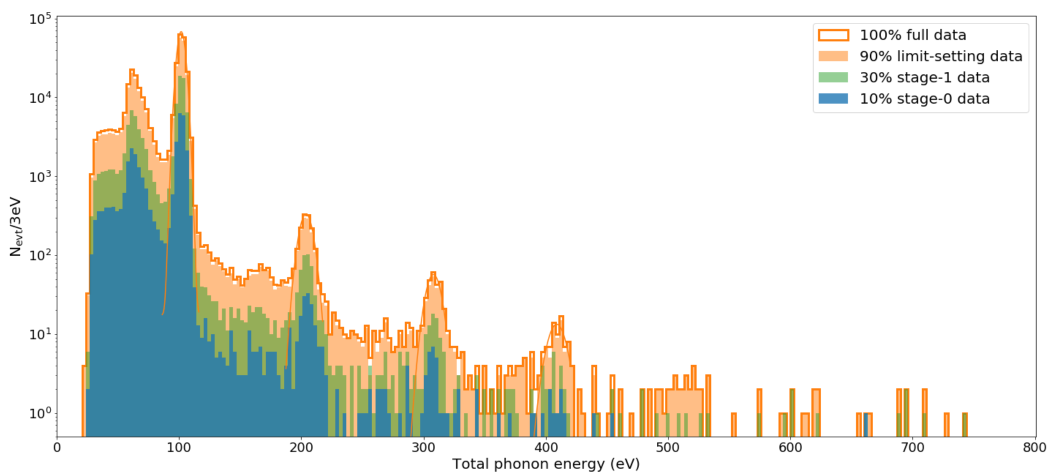


Figure 2.22: The top and the bottom figures are the 100 V post-selection laser and DM search data spectra, respective. For each distinguishable peak, we fit a Gaussian distribution to the data as shown by the fitted curve to extract the peak position and the standard deviation. For the DM search data, only the fits to the fully unblinded data are shown for clarity; other fit results for intermediately unblinded data can be found in (Chang, 2019a) (SuperCDMS internal).

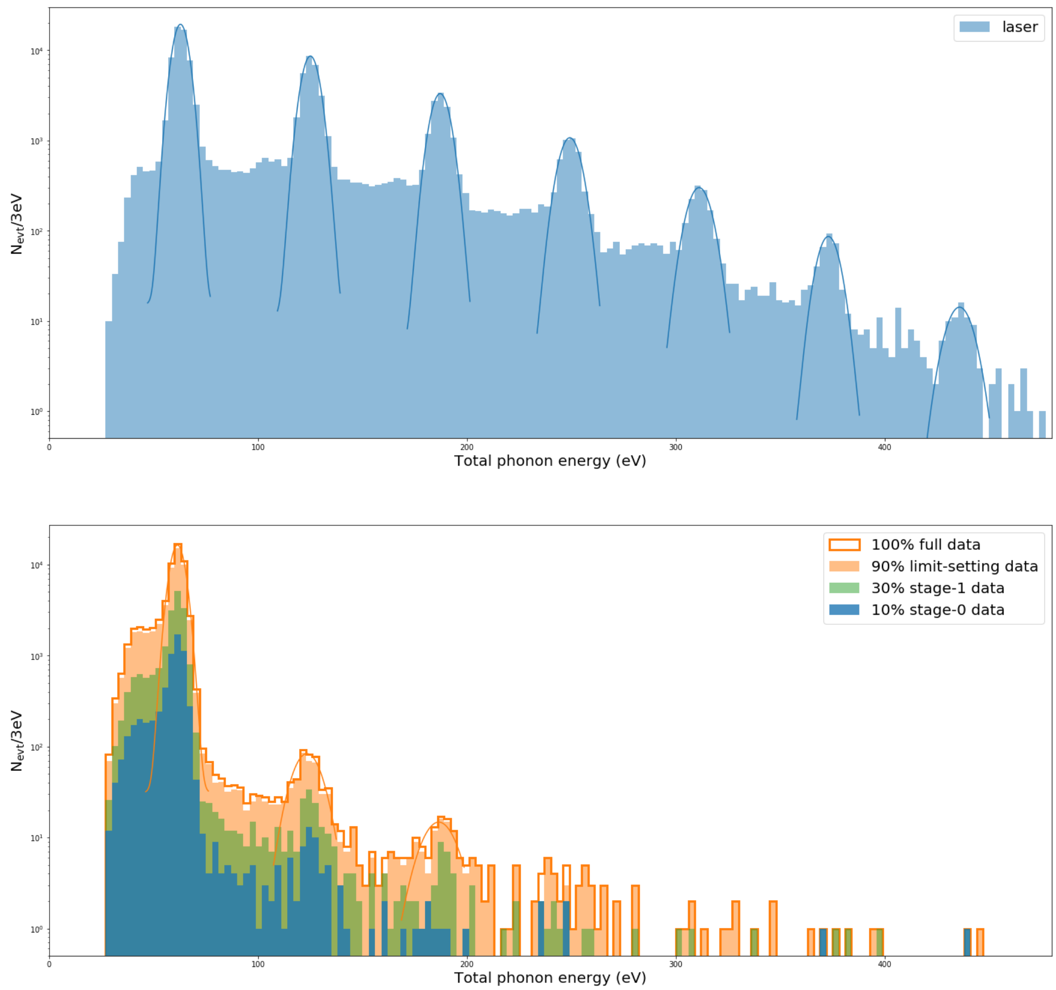


Figure 2.23: The top and the bottom figures are the 60 V post-selection laser and DM search data spectra, respective. For each distinguishable peak, we fit a Gaussian distribution to the data as shown by the fitted curve to extract the peak position and the standard deviation. For the DM search data, only the fits to the fully unblinded data are shown for clarity; other fit results for intermediately unblinded data can be found in (Chang, 2019a) (SuperCDMS internal).

2.6.4 Charge trapping and impact ionization

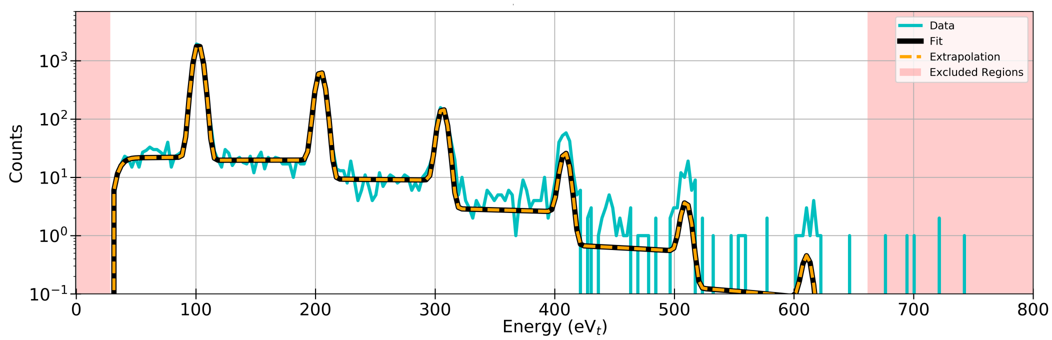


Figure 2.24: A laser spectrum fit to our 100 V laser data using (Ponce et al., 2020a)'s quantization model considering the CT and II effects. The fluctuating green curve and the smooth black curve represent the data and the fit, respectively. The fit shows the between-peak regions are indeed well-described by the CT/II-generated continuum. Credit: F. Ponce.

While the laser spectra we obtain in Fig. 2.20 are indeed highly quantized as expected by the dominant photoelectric e^-h^+ generation, there are some peculiar structures indicating the quantization is not the only observable mechanism leading to the result. In particular, we observe a high rate of the non-quantized events, i.e., between peaks, that appear to be irreducible after excluding the understood detector-compromised livetime and also tagging the true laser events with the TTL coincidence. If the non-quantized continuum were generated by random background coincidence, it should be greatly suppressed once the TTL tagging in enforced. After TTL tagging, it is expected that the highly suppressed background coincidence may contaminate the laser spectrum predominately through biasing the quantized pulse height as pile-ups rather than an uncorrelated addition to the spectrum. This type of pulse shape bias should respond to selections such as `MB` or χ^2, which is inconsistent with Fig. 2.20. In other words, the result indicates the continuum is produced as a consequence of the quantized e^-h^+ generation.

In order to accommodate the continuum, (Ponce et al., 2020a)[8] proposes including the charge trapping and impact ionization mechanisms into our crystal phonon emission model, where the originally generated e^-h^+, while traversing through the substrate under HV, may be trapped (CT) or liberate loosely bound charge carriers (II) at charged defect sites. In CT, since the original e^-h^+ does not traverse the nominal distance in the crystal, the amount of the NTL phonons it emits is proportionally reduced by the shortened path. Contrasting with CT, II is where

[8]Detailed derivation in (Ponce et al., 2020b).

the original e^-h^+ liberates extra charge carriers at charged sites by impact, hence subsequently providing more charge carriers that emit NTL phonons. However, depending on the locations of II, these charge carriers also do not traverse the full crystal size therefore result in a non-quantized extra NTL phonon generation. The continuum is therefore a consequence of the removal or addition of phonon-emitting charge carriers at random locations. Assuming constant CT and II probabilities for each charge carrier traveling an unit length along the HV, the total phonon energy produced by each e^-h^+ hence the modified quantized spectrum may be rigorously predicted. The modification is parameterized by the integrated CT and II probabilities for each e^-h^+ traversing the entire substrate, where these (integrated, not unit-length) CT and II probabilities may be extracted by fitting the expected spectral model to the quantized laser spectrum *with continuum*. In particular, the CI and II probabilities are most strongly constrained by the relative heights of the quantized peaks and their surrounding continuum. Fig. 2.24 presents such a fit to our 100 V post-selection laser data, where we extract $13 \pm 2\%$ and $1 \pm 1\%$ CT and II probabilities, respectively. These values will be used later as input parameters for producing more realistic DM spectra.

2.6.5 High-radius background

The readers might have noticed there is an unnatural "half" e^-h^+ peak in the pre-selection laser spectrum as well as the DM search spectra pre- and post-selection (Fig. 2.20). It is regarded "unnatural" because, if we assume all the particle detection by our detector is through quantized e^-h^+ generation, even considering CT and II, the expected signal should only be strictly quantized to the e^-h^+ peaks or randomly scatted to be the continuum due to probabilistic CT/II, neither permitting the prominent single-value half e^-h^+ energy. If we further compare its progressive removal in the laser and DM search data, we may conclude it is caused by an uncorrelated non-laser source, and the emission rate is rather high, i.e., $\mathcal{O}(10)$ Hz, considering the laser data were taken at 100-300 Hz yet the half e^-h^+ contribution is still observed at the relative rate shown by the peaks. In fact, due to such a high event rate, the half e^-h^+ contribution is even randomly added to quantized signals and is observed as a 1.5 e^-h^+ peak before TTL tagging. While this non-integer e^-h^+ emission is by-default nonexistent in our quantization-only ER DM search, therefore practically unlikely to affect the experiment's DM sensitivity,[9] such a well-established yet completely unexpected phenomenon does signal our ignorance in the detector physics

[9]The readers will find later that we exclusively use the quantized data for the DM search.

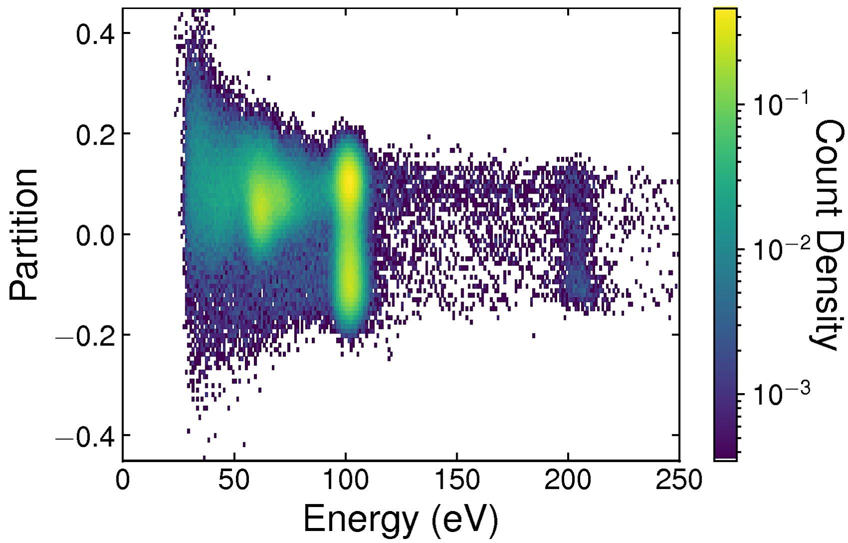


Figure 2.25: The partition variable η plotted against event energy for the post-selection 100 V stage-0 DM search data. Detailed explanation is provided in the corresponding text.

and motivate a dedicated study to resolve and mitigate its potential impacts for future experiments.

After reexamining the detector, currently we attribute the half e^-h^+ emission to a visible surface trench on the middle line of the diced detector sidewalls, i.e., the four uninstrumented 1×0.4 cm^2 surfaces. It is due to the constraint for dicing half of the thickness of our atypically thick substrate each time and then flipping to dice from the other side, which inevitably results in crudely crushing the Si crystal along the middle line due to slight misalignment for the front- and back-side dicing. The hypothesis underlying the middle-line half-e^-h^+ emission is that, if the rough crystal lattice in the trench hosts a high population of loosely bound charge carriers, a small external energy deposition may easily liberate these charges not as pairs but singularly. Coincidentally, this trench is produced transverse to the HV direction defined by the instrumented surfaces, so these charge carriers, when freed, are expected to almost always traverse half the depth of the detector and generate half the NTL phonons compared to the paired e^-h^+. We further verify the hypothesis by investigating the spatial distribution of the half e^-h^+ events quantified by the partition variable η (Fig. 2.25), which gives a strong evidence that these events are exclusively "high radius," i.e., closer to the sidewalls. Unfortunately, recalling the discussion when we defined η (Sec. 2.4), we did not collect sufficient ^{57}Co data for calculating the partition selection efficiency and therefore can not apply the select to reject these half e^-h^+ events. Motivated by the result, SuperCDMS collaboration is currently attempting to implement a robust partition selection efficiency calibration as well as developing the technique for polishing the diced sidewalls post-fabrication for future detectors. We are expecting the next batch of HVeV detectors, hopefully much improved by the knowledge gained in R2, to be deployed in winter, 2021.

In addition to the immediate impact of the leakage current background, the hypothesized mechanism for the half e^-h^+ generation poses a deeper long-term implication for future low-threshold DM searches. Also shown in Fig. 2.25, we may see, not just the half e^-h^+ peak, but in fact the broad continuum below 2 e^-h^+ is asymmetrically high radius. Comparing to the quantized events that are symmetric in η, we may then conclude, considering the e^-h^+ origin that causes the CT/II-generated continuum, the data indicate there exists another component of background that is not associated to e^-h^+ creation and is predominately high radius. It could be caused by a similar mechanism where external energy liberates loose charge carriers. Since currently all low-threshold DM experiments are limited by unidentified low-energy backgrounds, understanding the exact physics origin leading to the half e^-h^+ in our data, such as the precise channel(s) for the energy deposition, the condition for accumulating/liberating loose charges, etc., would be valuable for future DM experiments in general.

2.6.6 Efficiency model

We now examine the efficiency curves in Fig. 2.21 now that we have fully inspected the spectra used for the calculation. In general, the obtained trend is much similar to typical selection efficiencies, consisting of a rising edge manifesting the inability for the combined experimental design to detect smaller signals, whether it is due to hardware level limitations, e.g., triggering, detector physics, etc., or software/artificial level effects, e.g., reconstruction algorithm, data selection, etc. Above the experimental sensitivity threshold, roughly 20–40 eV in our case (100 V), the selection efficiency smoothly “turns on” and asymptotes to a constant in energy at 95–96%. It shows the experiment is roughly equally sensitive in the region therefore not particularly prone to systematic uncertainties if the energy spectra are to be utilized for physics searches, and the efficiency we acquire is also satisfyingly high. However, if we take a closer inspection, there are some interesting details that require further investigations.

First, our laser data are too sparse and lead to an uncalculable efficiency beyond

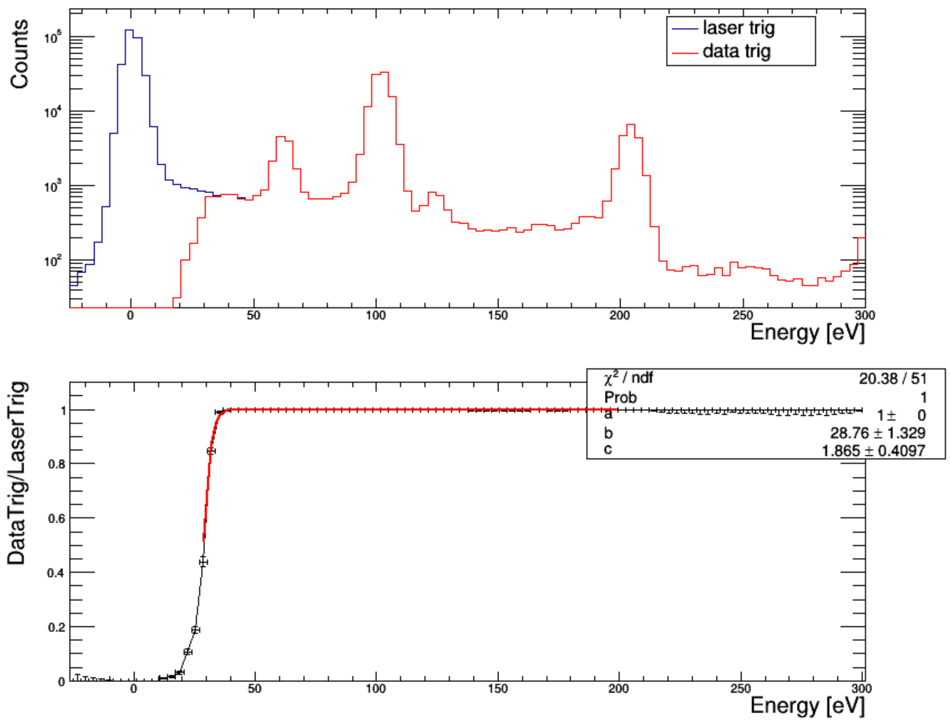


Figure 2.26: Top: 100 V laser event spectra identified by TTL-only tagging (blue) and also by DAQ triggering (red). Bottom: The trigger efficiency curve calculated taking i to be the DAQ trigger selection according to Eq. (2.26). The fitted Eq. (2.28) curve and corresponding fitted parameters are shown by the red curve and in the top-right inset. Figure produced by Z. Horng.

$\lesssim$700 eV, which can also be seen from the lack of the laser continuum data above this energy in Fig. 2.20. We therefore define the upper bound of our DM search region of interest (ROI) to be 650 eV. Since we plan to mainly use the quantized peaks for the DM search, spectral region far below the first e^-h^+ peak by multiple times of the $\sigma \approx 3\%$ $N_{e/h}$ resolution is essentially irrelevant, e.g., below 90% (roughly $\approx 3\sigma$) of the single e^-h^+ energy. We therefore choose 50 eV to be the lower bound of our ROI, based on its sufficient distance from the first e^-h^+ peak with a well-calculated efficiency on the abundant data, and we also choose to round the upper bound of the ROI by 50 eV to be 650 eV. Based on similar arguments, we also define the ROI for the 60 V DM search to be 50–390 eV.

Second, although we have briefly described our 20–40-eV efficiency threshold as a typical smooth turn-on, with a closer inspection, we may notice a much sharper rise at <30 eV that connects to a more rounded curve with a kink, which then smoothly transitions to the full efficiency. Our interpretation for the noticeably piece-wise efficiency curve is that, due to the small energy resolution ($\sigma_{N_{e/h}}$) of our data, we may in fact resolve the transition of the effects dominating the efficiency turn-on in different energy ranges. To rigorously model the efficiency curve, we first isolate the trigger efficiency by comparing all the TTL-identified laser events, even when no phonon pulses are detected, to those also triggered by our nominal DAQ trigger algorithm, i.e., taking i in Eg. (2.26) to be the DAQ trigger "selection" on the entire data stream. The trigger efficiency curve is shown in Fig. 2.26. To quantify the energy and the transition width of the trigger efficiency threshold, we empirically model the *trigger-only* efficiency curve by

$$\epsilon_{\text{trig.}} = \frac{a}{1 + e^{-(E-b)/c}}, \tag{2.28}$$

where we find a is consistent with unity, $b = 28.8 \pm 1.3$ eV, and $c = 1.8 \pm 0.4$ eV, indicating the hardware trigger threshold is at 29 eV with a transition as sharp as 4 eV. The isolated trigger-only threshold is consistent with the steep rise observed in the combined efficiency curve. We then compare the trigger efficiency curve to the event-based selection efficiency curves (Fig. 2.21), in particular in the 40–150 eV region that exhibits a perfect trigger efficiency, which indicates the difference between the gradually increasing combined efficiency and the trigger efficiency is dominated by event-based selections beyond 30 eV. Judging by the compact packing of the efficiency curves from the first selection to the last, we may conclude the smooth turn-on of the combined efficiency beyond 30 eV is due to the rejection to the misassociated trigger/OF-identified pulses.

Third, while at large the efficiency curve we obtain is smooth and only fluctuates within a reasonable range given by the finite data statistics, which is anticipated based on the detector's continuous response to the phonon energy, we notice sharp dips in the 100-V efficiency curve that seem to consistently appear at energies 10–20 eV below the expected e^-h^+ energies. These dips are noticeable at $N_{e/h} = 1, 2, 3,$ and 4 in Fig. 2.21. After inspecting the cut-by-cut evolution of the efficiency curves, we identify the χ^2 selection to be the cause of the dips. We examine the correlation between the χ^2 distribution and the energy of the quantized events, shown in Fig. 2.27, where we find the two parameters plotted, χ^2 and $\texttt{OF}''_{\text{cor}}$, exhibit a slight yet resolvable anti-correlation at every quantization. Higher-χ^2 events are preferentially lower-energy, which would result in rejecting by the χ^2 selection slightly more events at lower energies, in turn yielding reduced efficiencies at the low-energy sides of the quantizations that are manifested by the dips. However, according to the mathematical formulation of OF, the minimum χ^2 value determined by the OF should be independent of the size of the pulse, assuming the pulse shape matches

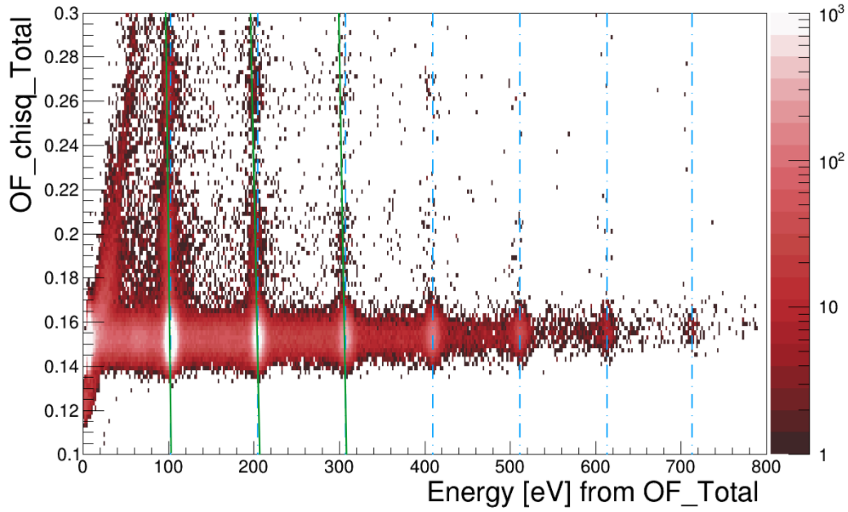


Figure 2.27: Laser frequency-domain χ^2 plotted against the calibrated event energy. The dashed vertical lines denote the e^-h^+ quantization energies per Eq. (2.8), and the solid green lines represent the peak energies as functions of χ^2 for the first three e^-h^+ peaks.

the template. In other words, the observed anti-correlation indicates the pulses are distorted in a systematic way, as opposed to randomly due to noise, that consistently biases the OF energy estimates toward smaller values.

Fig. 2.28 shows a few example pulses in the biased $\chi^2(\mathtt{OF}''_{\mathrm{cor}})$ clusters. In all these events, even though the pulses are correctly identified, they are assigned with erroneously suppressed heights, so the baseline of the template may compensate the existence of non-flat baseline in the signal traces, which, as shown by the examples, may be caused by close packs of pulses, non-phonon pick-ups, or the tails of large preceding pulses. The result unambiguously proves that the dips are unphysical for the detector's true DM detection efficiency but are generated due to the imperfect rejection of pile-ups and distorted/contaminated signal traces. Despite being due to physical origins, these dips would not exist for DM events therefore should not be considered in our DM limit-setting. In fact, this observation led us to revise the `MMB` acceptance window from an original 3σ selection. By incrementally tightening the `MMB` livetime selection, we may consistently suppress the efficiency dips hence further prove the phenomenon is truly artificial. We eventually choose the current 2.5σ `MMB` livetime selection, which we find best balances the efficiency dip size and the preservation of exposure.

Motivated by the result, we decide to provide the efficiency description not as a binned raw calculation from data but a smooth empirical model constrained by the

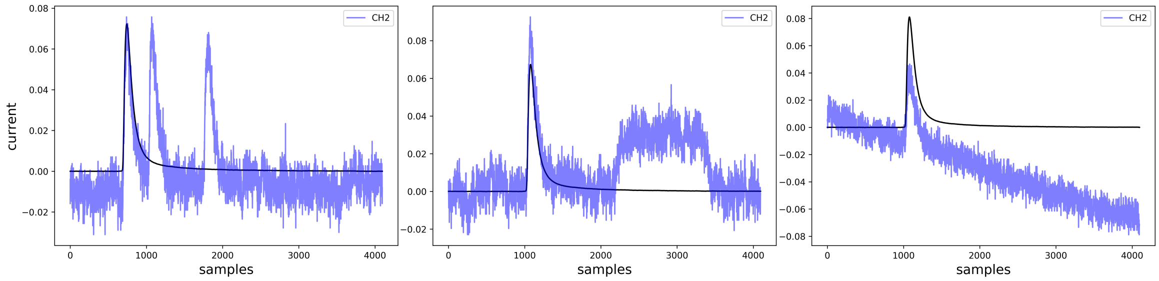


Figure 2.28: Example pulses belonging to the group exhibiting anti-correlated χ^2-`OF"`$_{\texttt{cor}}$ that generate the efficiency dips, shown in arbitrary DAQ unit. OF-reconstructed pulses based on the pulse template are shown by the smooth black curves. Figure produced by V. Novati

binned data. We model the efficiency by

$$\epsilon(E) = \epsilon_{\text{trig.}} \times (aE + b) \times \left(1 - e^{-(E-c)/d}\right), \tag{2.29}$$

where the first term is just the trigger efficiency curve obtained previously, the second term is a near-constant line describing the full efficiency at high energies, and the third is an exponential turn-on for the selection efficiency. We perform independent fits to the binned efficiency data by randomly generating 5000 sets of initial (a, b, c, d) values. The resulting peak efficiency at every energy and its $\pm 1\sigma$ range around the fitted peak value form the efficiency curve and its corresponding uncertainty range. The fitting results are shown in Fig. 2.21.

2.6.7 Dark matter signal model

For generating the expected DM signal spectra, we consider in this work three classes of DM models: DM-e^- scattering, dark photon absorption, and Axion-like particle (ALP) absorption. Owing to the fact that the general methodology for generating signal models are mostly identical to HVeV R1 (Agnese et al., 2019), and in fact all necessary ingredients, including the livetime- and event-selected datasets, efficiency files, our final DM limits, and a carefully written instruction for the usage of the above, are all provided publicly on `arXiv.org` (Amaral et al., 2020b), we refer the readers to these references, as well as M. Wilson's dissertation (M. Wilson, 2021) for shared technical details. Here we focus on the contents that are physically meaningful or specialized in the HVeV R2 analysis.

We start from the expected ER spectrum for the DM-e^- scattering, where extreme cases that the interaction mediator particle for the halo DM is much heavier or lighter than the ER target particle, i.e., Si nucleus in our case, are considered. According to (Essig et al., 2016), the differential scattering rate R in the recoil energy E_R for the process is given by

$$\frac{dR}{d\ln E_R} = V \cdot \frac{\rho_{\mathrm{DM}}}{m_{\mathrm{DM}}} \cdot \frac{\rho_{\mathrm{Si}}}{2m_{\mathrm{Si}}} \cdot \sigma_e \alpha \cdot \frac{m_e^2}{\mu_{\mathrm{DM}}^2} I_{\mathrm{crys.}}, \tag{2.30}$$

where V is the volume of the detector, $\rho_{\mathrm{DM}} = 0.3\ \mathrm{GeV/cm}^3$ (notice assumed value) and ρ_{Si} are the DM halo and Si crystal densities, m_{DM}, m_{Si}, m_{e} are the particle masses for DM, Si atom (nucleus), and e^-, μ_{DM} is the reduced mass for the DM-e^- system, α is the fine-structure constant, $I_{\mathrm{crys.}}$ is the scattering integral over the phase space for the detector crystal as defined in (Essig et al., 2016), and the interaction strength to be constrained, or equivalently the expected spectrum, is encoded by the unknown DM ER cross section σ_e. The heavy and ultra-light mediator assumptions are built in to Eq. (2.30) through their respective form factors for the Si nucleus in the interaction, $F_{\mathrm{DM}} = 1$ and $F_{\mathrm{DM}} = 1/q^2$, respectively, which are subsequently taken as dependence for $I_{\mathrm{crys.}}$, and q is the momentum transfer in the event. Instead of manually constructing the above equation into our limit-setting program, we adopt the computational tool `QEdark` to generate the expected ER spectrum similar to HVeV R1. (Essig et al., n.d.) Similarly, the dark photon spectrum is given by

$$R = \frac{1}{\rho_{\mathrm{Si}}} \cdot \frac{\rho_{\mathrm{DM}}}{m_V} \cdot \epsilon_{\mathrm{eff}}^2\, \sigma_1|_{\omega=m_V}, \tag{2.31}$$

where m_V is the mass for the dark U(1) vector Boson, the so-called dark photon, σ_1 is the real part of the Si conductivity, and ϵ_{eff} is the effective dark-SM photon mixing parameter

$$\epsilon_{\mathrm{eff}}^2 = \frac{\epsilon^2 m_V^2}{m_V^2 - 2m_V\sigma_2 + \sigma_1^2 + \sigma_2^2}, \tag{2.32}$$

where σ_2 is the imaginary counterpart of σ_1, and ϵ is then the unmodified (by detector material) dark-SM photon mixing parameter to be constrained. In this analysis, the Si conductivity data are taken from (Hochberg et al., 2017b) and (Bloch et al., 2017a) in a recoil energy-dependent fashion. Finally for ALP absorption,

$$R = \rho_{\mathrm{DM}} \cdot \frac{3g_{ae}^2 c m_a}{16\pi\alpha m_e^2} \cdot \sigma_{\mathrm{p.e.}}, \tag{2.33}$$

which describes the ALP being absorbed via the Axio-electric effect as a modification of the photoelectric effect, where m_a is the Axion mass, $\sigma_{\mathrm{p.e.}}$ is the photoelectric cross section that depends on the energy absorbed, thus also m_a, and g_{ae} is the unknown Axion-electron coupling that governs the process. We note that, due to the

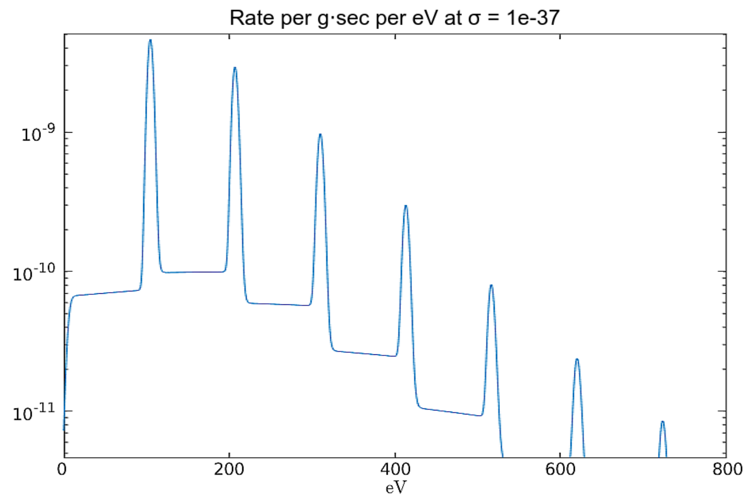


Figure 2.29: An example signal model assuming DM-e^- scattering, $m_{\mathrm{DM}} = 1$ GeV/c^2, $\sigma_e = 10^{-37}$ cm^2, $F_{\mathrm{DM}} = 1$ (heavy mediator), $F = 0.155$, a 100-V HV bias, and the CT and II fractions of 15% and 1% extracted previously. Figure produced by M. Wilson.

totally inelastic absorption nature of the dark photon and Axion ERs,

$$E_R = m_V \quad \text{or} \quad m_a \tag{2.34}$$

applies to to Eq. (2.31) and Eq. (2.33).

After the expected recoil spectra for the above DM models are constructed, we modify them based on our detector response, namely the quantized e^-h^+ detection with CT and II. Using the same detector model of HVeV R1 (Agnese et al., 2019), we convert the continues recoil energy spectra given by the theoretical equations above into number of e^-h^+ by

$$N_{e/h} = \begin{cases} 0 & , E_R < E_{\mathrm{gap}} \\ 1 & , E_{\mathrm{gap}} < E_R < \epsilon_{eh} \\ E_R/\epsilon_{eh} & , \text{otherwise,} \end{cases} \tag{2.35}$$

where $E_{\mathrm{gap}} = 1.2$ eV is the Si ionization bandgap, and $\epsilon_{eh} = 3.8$ eV is the average energy for creating each e^-h^+. Instead of simply converting the recoil energy to the exact $N_{e/h}$ according to Eq (2.35), we adopt an identical approach of HVeV R1 to account for the statistical fluctuation using the effective Fano factor

$$F_{\mathrm{eff}} = F \times \frac{E_R/\epsilon_{eh}}{E_R/\epsilon_{eh} - 1}, \tag{2.36}$$

where we assume an unmodified Fano factor $F = 0.155$, for which the modification practically prevents $N_{e/h} < 1$ as proposed in (Agnese et al., 2019). The readers may find a much more detailed prescription for our practical implementation of the Fano factor in the public data release package of (Agnese et al., 2019), which may be obtained from the corresponding arXiv page of (Agnese et al., 2019). Finally, we may calculate the corresponding total phonon energy using Eq. (2.8), convolve the resolution-less prediction with the data resolution tables introduced previously, and then further adjust the quantization based on (Ponce et al., 2020a)'s CT/II prescription to generate the between-peak continuum. Fig. 2.29 shows an example for the simulated spectrum assuming heavy mediator DM-e^- scattering at $m_{\mathrm{DM}} = 1$ GeV/c^2, $\sigma_e = 10^{-37}$ cm^2, and other input parameters listed in the figure caption.

2.7 Dark matter constraints

2.7.1 Unblinding

Based on the blind analysis principle, before we reveal the blinded data for the official DM limit-setting, i.e., stage-1 and -2 data, we should complete two necessary procedures in advance and, importantly, fix thereafter: 1) The criteria for determining the successfulness of the (un-)blinding, 2) the intended limit-setting methodology, which should be determined based on the already unblinded calibration and the stage-0 DM search "background proxy" data. In this section, we will first discuss the blinding verification criteria and report immediately afterward the unblinding verification results. We then explain the process of determining our limit-setting algorithm, which is one of the major differences in methodology relative to HVeV R1 and most modern[10] SuperCDMS analyses. We should emphasize the determination process for the limit-setting methodology took place before any previously blinded data were revealed, meaning it was performed on the stage-0 data before we validated the successfulness of the blinding. We carefully set up the rules to prevent the generation of any kinds of DM parameter limits or indicative projections during the intermediate unblinding steps, so to ensure the unblinding verification introduces the least bias to the analysis while still allowing us to examine the consistency between stages. Based on the criteria we choose, we achieve in the end a successful blind analysis. With the fully unblinded stage-1 and -2 data, we generate the final DM parameter limits for σ_e, ϵ, and g_{ae} through the data processing and limit-setting methodologies that are constrained prior to the final unblinding. Finally, we discuss the implication of the physics results as the conclusion of the chapter.

In general, we want to verify whether the analysis methodology developed on the

[10] i.e., 2010–.

stage-0 data is equally applicable for the full data, in the sense that there should be no unexpected data features that would otherwise degrade the expected DM search sensitivity. The current analysis is developed using the 1^{st}-second data in every 10 seconds, while the other 9 seconds are divided into stage-1, the 2^{nd}–3^{rd}, and stage-2, the 4^{th}–10^{th} seconds. This division scheme is prepared so that we may apply the stage-1 data for an intermediate verification for the unblinding, before the full data are unblinded for the final limit-setting. We acknowledge that the (un-)blinding scheme is not the simplest, most bias-robust option, but since HVeV R2 is the first analysis in the HVeV series adapting a blind analysis, the blinding scheme is selected so we may gain more insights for future blind analyses. If stage-1 data meet the criteria for a robust, unbiased analysis developed on the stage-0 data, we then proceed to unblind the stage-2 data and use the stage-1 and -2 data together, i.e., excluding the data employed in analysis development, to generate the final limits. If unfortunately stage-1 data fail to satisfy the criteria, we decided prior to the data-taking campaign that the stage-0 and -1 data will be combined for a subsequent analysis re-development, and the revised analysis methodology will be directly applied to the stage-2 data for limit-setting, with bias/uncertainty quantified accordingly, but without further partial unblinding steps for verification or adjustment purposes. If the unblinding had proceeded in this scenario, we would have reported the full sequence in (Amaral et al., 2020a).

In terms of the criteria for a successful unblinding, we decide to examine the stage-0 and -1 consistency of 1) the selected livetime, 2) the parameter distributions for event-based selections, and 3) the data resolution indicated by the quantized $e^{-}h^{+}$ peaks in the DM search data. Note that the laser data have been fully unblinded since the beginning and are irrelevant to the blinding for the DM search. In principle, the only category among the three that, if systematically biased, may affect the conclusion of the analysis is the parameter distributions, which subsequently determine the selection efficiency used in the limit calculation. On the other hand, the livetime and the resolution may simply be updated to those yielded by the unblinded data. However, if these qualities that the analysis directly depends on are indeed consistent between stage-0 and -1, we should expect the post-selection final livetimes estimated by the partially unblinded data to be consistent after rescaling by the unblinded fractions. It would demonstrate that the parameters utilized by the livetime selection are distributed independent of our time-domain division, i.e., evenly at a Hz-level time scale. Similarly, the quantization resolution from the DM search data, specifically the central values and widths for the $e^{-}h^{+}$ peaks, should also be consistent between the two stages, if the background is truly independent of our blinding scheme. We note again, if the livetime and the resolution are somehow correlated or biased by the blinding, we should apply the values indicated by the actual limit-setting data, so the obtained results may still be regarded as properly reflecting the physics given by the data. It is nevertheless less optimized in sensitivity given the analysis is developed on a noticeably incomparable dataset. On the contrary, if the event-by-event selected parameters are biased, we would acquire systematic biases in the resulting limits that should either be included or trigger an analysis revision.

For validating the consistency in livetime selection, due to the already elaborated correlation between selections, we decide to compare the combined remaining livetimes after all livetime selections are applied. The selected stage-0 and stage-1 livetimes should differ by a factor of 2 according to their sizes, and we set the criterion that the values should be consistent within a difference of 5%. Otherwise we should perform a further investigation to identify the causes to determine if the analysis is biased and requires a revision. For the post-selection data resolution, we require the first four quantization peaks (higher peaks are unresolvable due to low statistics of the stage-0 data), defined by the $\pm 3\sigma$ regions around the spectral peaks, should yield consistent peak energies and widths based on Gaussian fits. Finally for the most critical event-based selection parameters, we adapt the two-sample Kolmogorov–Smirnov (KS) test (Hodges, 1958) to assess the consistency of the parameter distributions taken separately from the stage-0 and -1 data. We choose the criterion that, if the null hypothesis is not rejected at $\alpha < 0.01$, i.e., KS test does not support that the two distributions are sampled from the same mother distribution at more than 99% C.L., we should then investigate the causes, since the null hypothesis test does not immediately signal a resolvable difference but rather can not support the consistency statement at the designate C.L. Through the investigation, we should either conclude the blinding is bias-less, quantify additional uncertainties made into the limits, or revise the analysis.

As a result, we find the selected stage-1 livetime is 3.5% lower than that expected by stage-0, and the resulting post-selection data resolutions are fully consistent (Fig. 2.22/2.23). Detailed numerical values may be found in (Chang, 2019b; Chang, 2019c) and (Chang, 2019a) (SuperCDMS internal). For the KS tests, we explicitly examine the temperature, `MB`, frequency- and time-domain χ^2, and $\mathtt{OF}_{\mathtt{tot}}$ for the telegraph noise monitor. When testing the χ^2 distributions, we use the 1-D χ^2 data

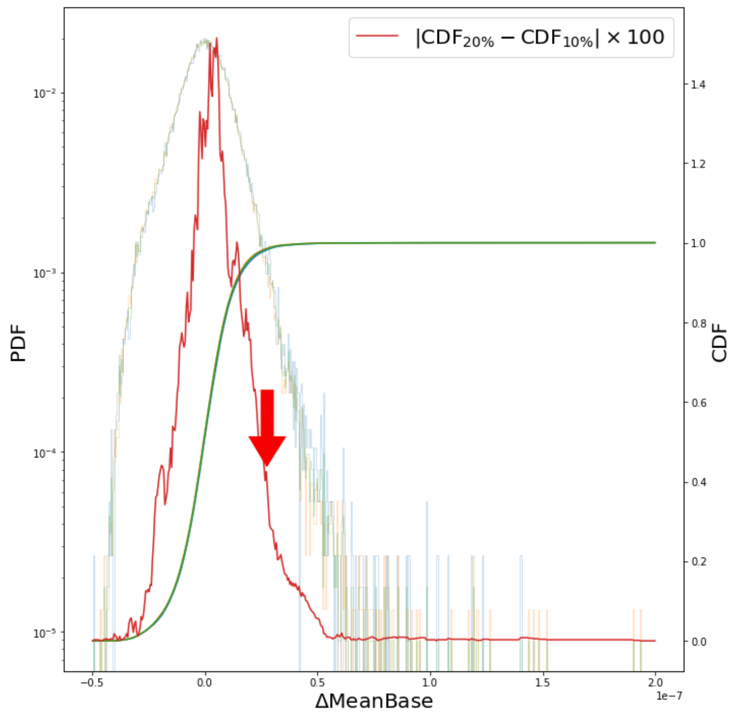


Figure 2.30: The ΔMB distributions combining all DM search datasets, i.e., combined after the SQUID-locked baselines are subtracted. The lightly colored background histograms are the normalized probability distribution functions (PDFs) for the stage-0 (blue), stage-1 (yellow) and stage-0 and -1-combined (green) data, and the foreground curves of the same colors represent the CDFs of the corresponding PDFs (all overlapping). The red curve shows the absolute CDF difference between stage-0 and -1, inflated by a factor of 100 for visibility. The peak value for the red curve determines the KS statistic, while the location highlighted by the arrow indicates our MB selection threshold. Detailed explanation is provided in the corresponding text.

histograms to simplify the KS test computations, although the χ^2 selections are performed in the 2-D χ^2-energy space. To our satisfaction, all the KS tests satisfy our criterion except for the MB distribution, which is shown in Fig 2.30. Based on the KS test calculation, where the KS "*D*" statistic is determined from the maximal cumulative distribution function (CDF) difference between the input distributions, which is 1.4% in our case as shown in the figure. However, the difference at the designated 3σ selection point is about 0.4%. We therefore conclude from the explicit CDF comparison that the MB sample discrepancy suggests a trivially small uncertainty (0.4%) to the efficiency. In fact, even if we had unreasonably selected at the CDF difference maximum that determines the *D* statistic, the uncertainty made into the efficiency due to blinding is still a negligible 1.4% compared to the efficiency modeling uncertainty (Fig. 2.21). We conclude that the staged unblinding validation result indicates that we may proceed to unblind the full data for limit-setting without revising the stage-0-determined analysis methodology. It is worth noting for future reference, from a follow-up study after the full unblinding, we later find, by validating the stage-2 data against the stage-0 and -1-combined data as if the first 3-seconds bin was initially unblinded for the analysis development, the livetime difference is $\lesssim 1\%$, the data resolution remains fully consistent, and all the the KS tests, including the `MB` selection, pass the 99%-C.L. criterion.

2.7.2 Limit calculation

Before formally introducing the main limit calculation, we first point out that, while we do obtain a good fit to our laser data with (Ponce et al., 2020a)'s CT/II model (Fig. 2.24), in which we extract the CT/II contributions with the continuum data that are in turn used to generate the expected DM signal model, the high non-quantized low-energy background shown in the stage-0 DM search spectrum (Fig. 2.20) is clearly in disagreement with any of the signal models we can generate based on current assumptions. The disagreement may be easily seen by comparing Fig. 2.29 to Fig. 2.20, in particular the between peak regions, and realize our model expect a much lower between-peak continuum around the quantizations, while in the obtained data the continuum exponentially rises toward lower energies. In fact, we will explain later that the model-data disagreement also applies to the quantized signal. In summary, at this moment we do not have a complete model to objectively distinguish the DM contribution from unidentified background sources, whether or not the between-peak data are included for the limit calculation.

Based on this critical fact, we report in this work only DM parameter constraints, i.e., upper limits for the interaction strengths, due to our inability to distinguish potential DM events from non-DM *background* events in the post-selection DM *candidate* events. This condition leads to a limit-setting power that is fully limited by the observed event rate, up to the highest possible DM interaction strength that would otherwise yield a rate that can not be explained by the observed event rate. In this observation-limited limit-setting, we take into account the event rate's energy dependence, i.e., the spectral distribution, and the statistical fluctuation at the chosen CL. Motivated by the same reason, we decide to exclude the continuum from the limit-setting process as it is deemed to be, relative to the quantized spectral regions, dominated by unidentified background on top of the modest contribution from the CT/II effect-smeared quantized events. We exclude the data outside the

$\pm 3\sigma$ regions of the 1st–4th $e^- h^+$-quantization spectral regions and use only these windows for setting the limits. Note that these windows are defined by the observed four quantization peaks in the stage-0 DM search data but not from the laser data, because depending on the mass of DM, the observed DM-candidate peaks may be offset differently from the NTL-only energy multiples, i.e., the $(N_{e/h} \cdot \mathtt{HV_0})$ term in Eq. (2.16). It is analogous to the offset due to calibrating the detector with different photon energies. If the quantization peaks are due to non-DM sources in the DM search data, their offsets also similarly encode the characteristic primary ER depositions of the background sources. Due to the improved resolution of our detector combining with the peak-only limit-setting scheme, we find this subtle detail in peak position may introduce a significant uncertainty to the limit calculation, which was not the case for HVeV R1.

We realize such an effect is significant and needs to be taken into account when we attempt to generate trial limits on the stage-0 data as part of the limit-setting algorithm development process. We find, due to the 3 eV energy resolution, any eV-level offset in the data-model spectral comparison process is resolvable and would lead to final limits that are sensitive to the difference. We previously hypothesized possible scintillation/fluorescence emissions as the cause to the deviation of the 0th peak in our laser calibration data (Sec. 2.4). We notice, in the presence of these unidentified energy sources, if the limits are to be calculated by the optimum interval (OI) algorithm (Yellin, 2002; Yellin, 2007) that is commonly adopted by SuperCDMS analyses for providing generally stronger limits, such a sensitivity to the exact peak positions would introduce high uncertainties in the resulting limits. When we compare the OI-determined most possible DM spectra to our windowed data spectrum, shown in Fig. 2.31, it becomes clear that the misalignments at the edges of the quantization windows are unphysically emphasized by the OI algorithm, as they provide the strongest constraints to the allowed DM interaction strength. The OI in turn unphysically sets limits with these edge intervals that are strong yet susceptible to the systematic uncertainty. The phenomenon was not anticipated but immediately understood, since the OI algorithm is designed to "optimally" compare the obtained event distribution intervals to the expected distribution, hence not only considering the number of events observed but in our case aggressively reweighting each event by its discrepancy to the expected signal model. We conclude, as our detector resolution has improved to the level that resolves the aforementioned spectral features, it is overly uncertain to provide DM constraints based on the OI algorithm.

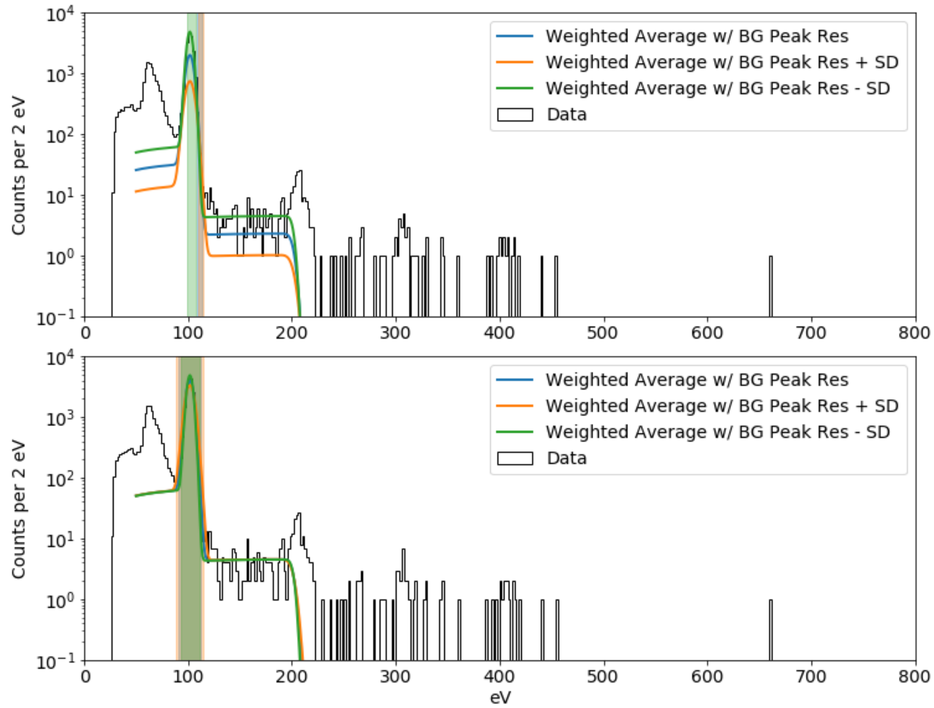


Figure 2.31: The stage-0 data (black curve) overlaid with the 90%-CL signal models determined by the IO (top) and Poisson (bottom) limit-setting algorithms. For this figure, we use a dark photon absorption model of $m_V = 1.8$ eV as an example to emphasize the single e^-h^+ peak. By varying the data resolution by $+1\sigma$, using nominal value, and -1σ, shown, respectively, in yellow, blue and green, the data windows dominate the limit-setting are highlighted by the vertical bands, and the resulting signal models are shown by the corresponding colors. We may see, as a pure counting method, the Poisson limit-setting faithfully utilizes the actual $\pm 1\sigma$ data ranges and provides stables predictions, while the OI algorithm may identify the edge of the peak as the optimum interval, de-weighting most of the data, and set unstable limits based on the biased interval. Figure produced by M. Wilson.

Motivated by the finding, also as a conservative choice for the first blind HVeV analysis, we eventually take the opposite approach to completely ignore the in-peak spectral shape information and calculate the HVeV R2 limits based only on the total numbers of events obtained in the e^-h^+ peaks. We define the peak regions by their $\pm 3\sigma$ windows around the peaks and count the numbers of events within the windows. We weight the limits given independently by different peaks as they may be separately compared to the signal model and yield different limits. Already mentioned previously, we find the relative heights of the observed quantization peaks do not agree with any of the signal models we may generate, especially the first and second e^-h^+ peaks. It indicates that the sources dominating some or potentially all of the quantized peaks are likely different from our signal model, which suggests different peaks may yield significantly different limits. Therefore, it is necessary to combine the limits determined by different peaks with a proper statistical reweighting to be the final result given by the data. For our pure counting-based limit-setting, we calculate the limits assuming simple Poisson fluctuation for the obtained number of events in each peak. By comparing the number of events allowed by Poisson fluctuation to the expected peak height for a given DM mass and interaction strength, the allowed number of events may be translated into a 90% C.L.-allowed σ_e, ϵ, or g_{ae}. The choice of providing 90%-C.L. exclusion limits is to be comparable with HVeV R1 and other recent low-mass DM searches. To obtain the relative weighting for each peak to be used for the limit-setting, for each DM mass, we vary the signal model by randomly sampling the nominal distributions of the input parameters with systematic uncertainties, listed in Tab. 2.3, and take the relative occurrence of each peak dominating the limit-setting by yielding the most stringent limit as its weight. For a peak that dominates the limit-setting n times in the random sampling process, we then calculate for the peak a limit that exhibits a C.L. of

$$C_n = 0.9^{1/n}, \tag{2.37}$$

so when the peak is used n times, the combined C.L. is reduced to 90%. One can then calculate C_n separately for each peak and compare it to other C_n to be the relative weighting. We should note, for every mass and every model, this relative weighting is determined using the Poisson algorithm described above on the stage-0 data *before unblinding the stage-1 data* and fixed thereafter for the final limit-setting. Tab. 2.3 also summarizes other more detailed uncertainty items assumed for this work, and we again refer interested readers to (M. Wilson, 2021; Amaral et al., 2020b) for technical instructions on the Poisson limit-setting and other technicalities for our particular calculation.

2.7.3 Physics result

Fig. 2.32 presents the final results for the HVeV Run 2 DM search, based on the 1.2 (0.4) gram-days of the stage-1 and -2 data collected at the bias HV of 100 (60) V. In general, we obtain consistent DM mass sensitivity thresholds with HVeV R1 at 0.5 MeV/c^2 and 1.2 eV/c^2 for the scattering and absorption types of ER DM models, respectively. As an improvement relative to HVeV R1, for the first time by a direct detection experiment, we provide the ALP absorption limit in a similar mass range to our dark photon result. The limits obtained in the work are generally weaker than HVeV R1 by a factor of a few depending on the mass ranges, due to various

Sys. uncertainty	Distribution	100 V		60 V		Unit
		Mean	Range (σ)	Mean	Range (σ)	
Energy resolution	Gaussian	3.6	0.3	3.5	0.3	eV
Energy calibration	Gaussian	0	0.5	0	0.5	eV
CT fraction	Gaussian	11	3	16.0	0.7	%
II fraction	Gaussian	2	3	3×10^{-13}	4×10^{-13}	%
Efficiency	Gaussian	Fig. 2.21		Fig. 2.21		
P.E. absorption ($\sigma_{\text{p.e.}}$)	Uniform	(Amaral et al., 2020b)		(Amaral et al., 2020b)		
Fano factor	Uniform	0.155	$10^{-4} \sim 0.3$	0.155	$10^{-4} \sim 0.3$	

Table 2.3: The nominal input parameters for the HVeV Run 2 limit-setting calculation. For items noted as Gaussian distributions, the range values represent the Gaussian standard deviations, while for uniformly distributed items, the assumed nominal values are noted under the means, which are not necessarily the centers of the assumed uncertainty ranges. For the efficiency and the photoelectric photon absorption cross section, since the actual inputs utilized are energy dependent, we refer the readers to Fig. 2.21 and (Amaral et al., 2020b) for detailed data. This table is reproduced from (Amaral et al., 2020a; team, 2020a) with less critical technical information removed.

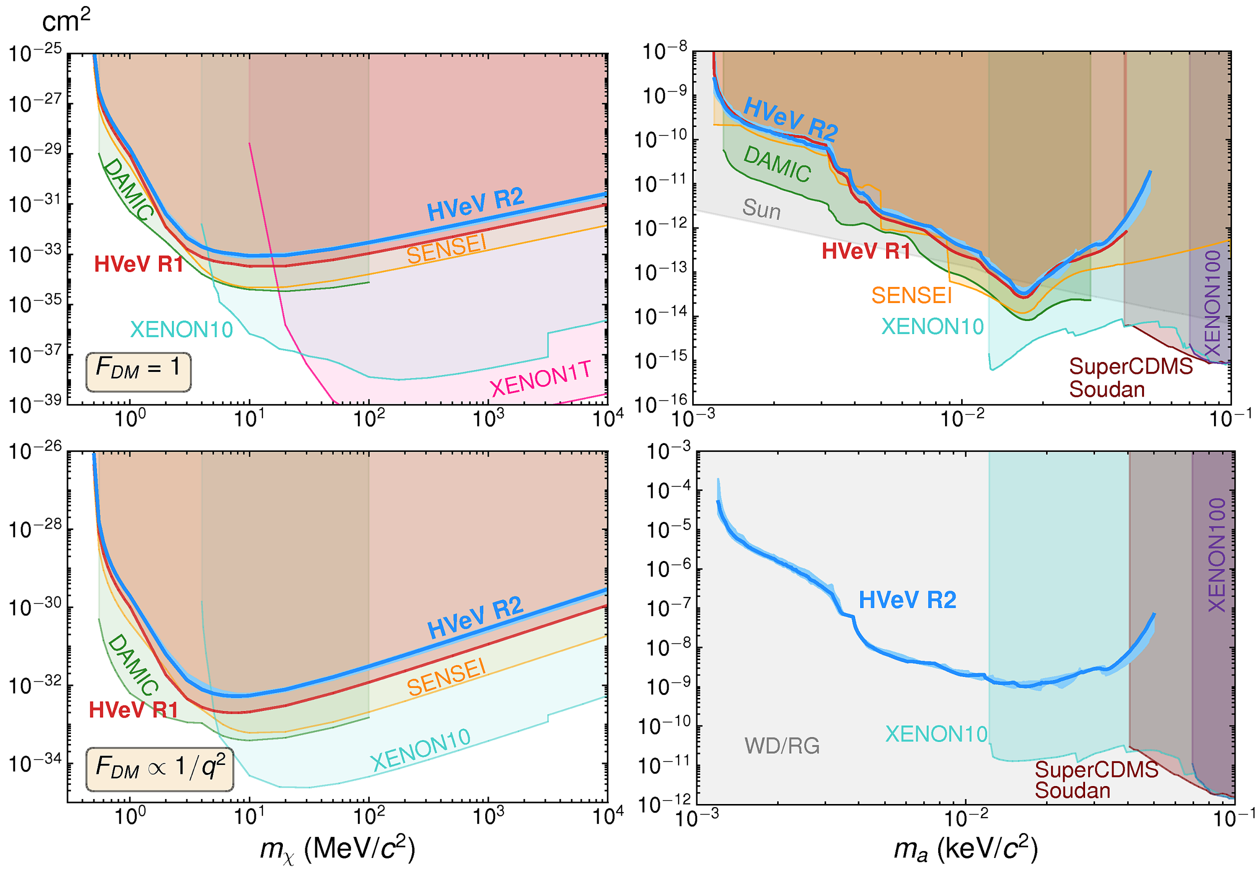


Figure 2.32: The final limits obtained by HVeV Run 2 analysis. The left-hand side panels show the results for DM-e^- scattering, heavy and ultra-light mediator cases in the top and bottom panels, respectively. The right-hand side panels show the results for absorption type of ER DM models, dark photon and ALP cases in the top and bottom panels, respectively. The results presented in this figure are based on our 100-V DM search data, where we choose to suppress the 60-V results for visual clarity, since they are generally consistent thus overlapping with the 100-V curves; interested readers may find the 60-V results in (team, 2020b). Other constraints shown here for the DM-e^- scattering include SuperCDMS HVeV R1 (Agnese et al., 2019), DAMIC (Aguilar-Arevalo et al., 2020), SENSEI (Abramoff et al., 2020), XENON10 (Essig et al., 2017; Essig et al., 2012), XENON1T (Aprile et al., 2019), and for dark photon and ALP absorption, SuperCDMS Soudan (Aralis et al., 2020), XENON10, XENON100 (Bloch et al., 2017b), HVeV R1, DAMIC, SENSEI, the sun (An et al., 2020), red giant and white dwarf stars (Viaux et al., 2013; Tanabashi et al., 2018; Bertolami et al., 2014).

factors combined. First, due to the now observable quantization modeling details that were not relevant for R1, we are limited to adopting the less aggressive Poisson limit-setting rather than the OI as in R1. Second, as we have pointed out previously, we are the first experiment to consider the charge transport signal modification, namely the CT and II effects, which in our case results in a ≈16% probability of charge carrier loss *for each carrier*. The total loss of signal exponentially increases when multiple e^- and/or h^+ are generated by DM ER, thus weakening the resulting limits by a factor of a few. If we focus on the lower-mass regions, e.g., the lower half of each of the limit graphs, we may notice currently the leading limits in these regions are all provided by charge transport type detectors, such as the SENSEI and DAMIC CCD detectors as well as HVeV detectors. For these experiments, we have demonstrated that the impact due to CT/II may be significant and thus suggest future solid-state low-mass DM experiments to understand. Motivated by this conclusion, we apply the same model utilized in this work to retrospectively study the CT/II effect in the legacy R1 laser calibration data. We find the CT and II fractions in the HVeV R1 detector are intriguingly consistent with zero, suggesting the published limits are still reliable. The finding nevertheless proves such a significant variation, in our particular case we contribute to fabricating the R1 and R2 detectors from different Si wafers, undoubtedly exists and is critical to be considered.

Finally, a direct comparison of the data spectra obtained in HVeV R1 and R2 is provided in Fig. 2.33. We see, aside from the much improved resolution, the quantized event rate in every e^-h^+ peak is surprisingly consistent, despite R1 and R2's very different experiment constructions, exposures, and also the detector themselves. The result further justifies the R1/R2 limit difference is predominately due to the deliberate choice of analysis methodology but not the physically detected data. Since currently all the low-mass DM experiments are limited by poorly modeled low-energy backgrounds, our finding provides a strongly indicative information to the physics origin(s) of the low-energy background, namely it is likely independent of the differences of HVeV R1 and R2, such as the location, the crystal condition, the cryo-mechanical constriction, etc. The finding is of great interest to the community and subsequently helps with various studies hypothesizing different possible sources of the background. We recommend the readers to (Du et al., 2020) for a comprehensive summary to the current (2020–2021) understandings, theories, and future implications for a spectrum of different background hypotheses. It is also worth mentioning that, it is further noticed by (Kurinsky et al., 2020), not only between HVeV R1 and R2, with the proposed rescaling, the authors even bring in other experiments utilizing noble gases and scintillators as their detectors and conclude the observed low-energy background across all the current experiments might be a consequence of a shared origin. If one is convinced by the methodology of Kurinsky et. at. (2020), then the finding becomes more indicative to a real DM detection but somehow modified by poorly understood detector physics.

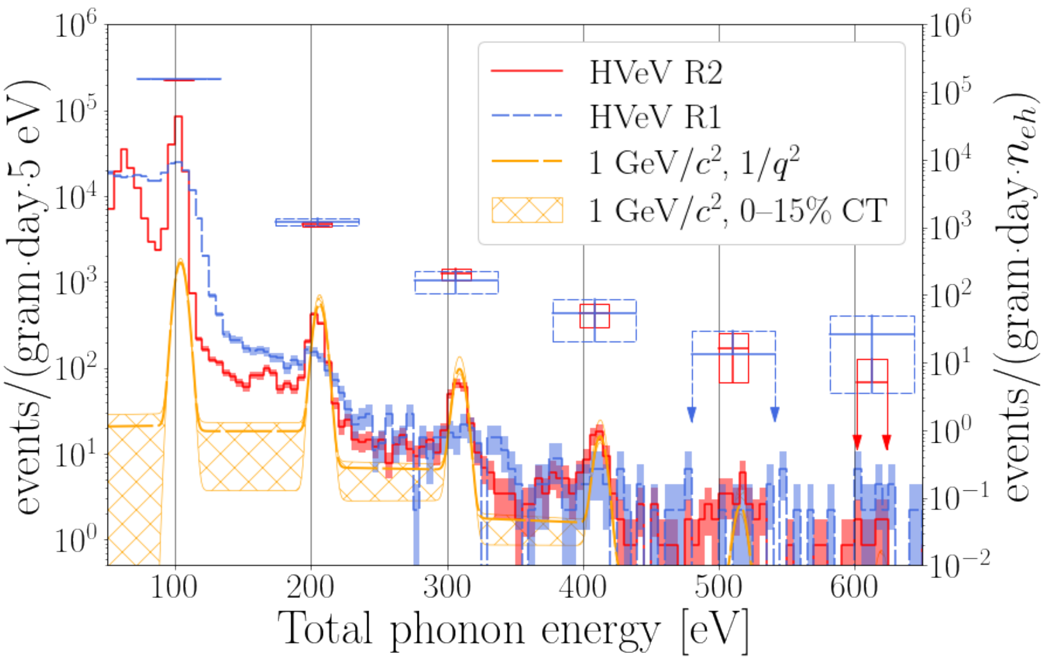


Figure 2.33: The HVeV R1 (blue) and R2 (red) DM search data spectra, normalized to the units of events/gram-day per energy bin (left axis) or per e^-h^+ generation (right axis). The full spectra are plotted using the left-axis unit, while the total number of quantized events in each e^-h^+ peak, defined by the $\pm 3\sigma$ range of each peak, is presented by the box above the peak using the right-axis unit. The shaded regions attached to the full spectra represent their uncertainty ranges. The horizontal middle line, the vertical middle line, the width, and height of each box, respectively, represent the nominal e^-h^+-generation event rate, the peak position, the peak's $\pm 3\sigma$ energy range, and the event rate uncertainty. Also shown in this figure is an example signal model (yellow) constrained by the R2 data using our Poisson limit-setting algorithm, assuming $m_{\rm DM} = 1$ GeV/c^2 for heavy-mediator DM-e^- scattering, and the attached hatch region marks the uncertainty range due to varying the CT fraction in 0–15%.

2.8 Future work

Since the completion of HVeV R2, the HVeV project has moved from above-ground sites to the NEXUS facility (Z. Hong et al., 2019), an underground cryogenic facility in the 255-meter water-equivalent MINOS near-detector tunnel at FNAL, IL, USA. At NEXUS, in addition to the overburden shielding, multiple HVeV detectors are simultaneously deployed to achieve correlated background rejection as well as accumulate greater exposures. To further understand the critical background that limits HVeV R1 and R2 searches, and in fact all current low-mass DM experiments, in parallel to the overburden impact and multi-detector correlation studies, NEXUS will host a dedicated experiment developed by this thesis to characterize the potential background contribution by the blackbody radiation in the cryogenic system. The blackbody radiation background study is detailed in Ch. 5. NEXUS has also adapted a kinetic inductance detector (KID)-based R&D particle detector, developed based on the works of this thesis in Ch. 3 and Ch. 4, for a collaborative development for future highly-multiplexed RF-based phonon-mediated particle detectors.

Chapter 3

KINETIC INDUCTANCE DETECTOR-BASED DARK MATTER DETECTOR—DESIGN

3.1 Kinetic inductance detector-based dark matter detector

3.1.1 Kinetic inductance detector

Kinetic inductance detector (KID) is a type of thin-film superconducting (SC) pair-breaking detectors similar to the QET, which SuperCDMS DM detectors currently employ for phonon detection. Unlike the QET, in a KID a pair-breaking energy deposition that creates QPs is sensed with the inductance increase caused by the excessive QPs, as opposed to the resistance increase in the TES. First proposed and demonstrated by (Day et al., 2003) at NASA JPL and Caltech, such a detection principle was realized by constructing a μm-scale quarter-wavelength ($\lambda/4$) transmission line resonator in the microwave range. When the KID was cooled to much below T_c and experienced a transient QP generation due to X-ray illumination, the transient sheet inductance change led to a corresponding resonance modulation. In the complete construction of (Day et al., 2003), the $\lambda/4$ resonator was capacitively coupled to a 50 Ω-matched microwave transmission line for a transmission (S_{21})-style measurement. One continuously passes a near-resonance monotone readout microwave signal through the transmission line and monitors the resonant transmission at the output as the signal. Owing to the 50 Ω transmission line's microwave feeding function, it is commonly called a "feedline." When the resonance is shifted by the excessive QPs, one then observes a output signal (S_{21}) change in both the amplitude and phase.[1] In addition to the pulse detection, (Day et al., 2003) also showed, by steadily controlling the temperature of the KID to reach different equilibrated QP densities and their corresponding steady-state S_{21}, one may obtain a S_{21}-QP density correspondence that calibrates the observed pulse amplitudes to the energy deposition. This procedure was only qualitatively demonstrated in (Day et al., 2003). We reproduce in Fig. 3.1 the cartoon chart originally provided in (Day et al., 2003) for illustrating the working principle of KID, with a step-by-step explanation in the figure caption.

[1] The signal is now commonly measurable in both quadratures, but it was only reported in phase in Day et. al. (2003).

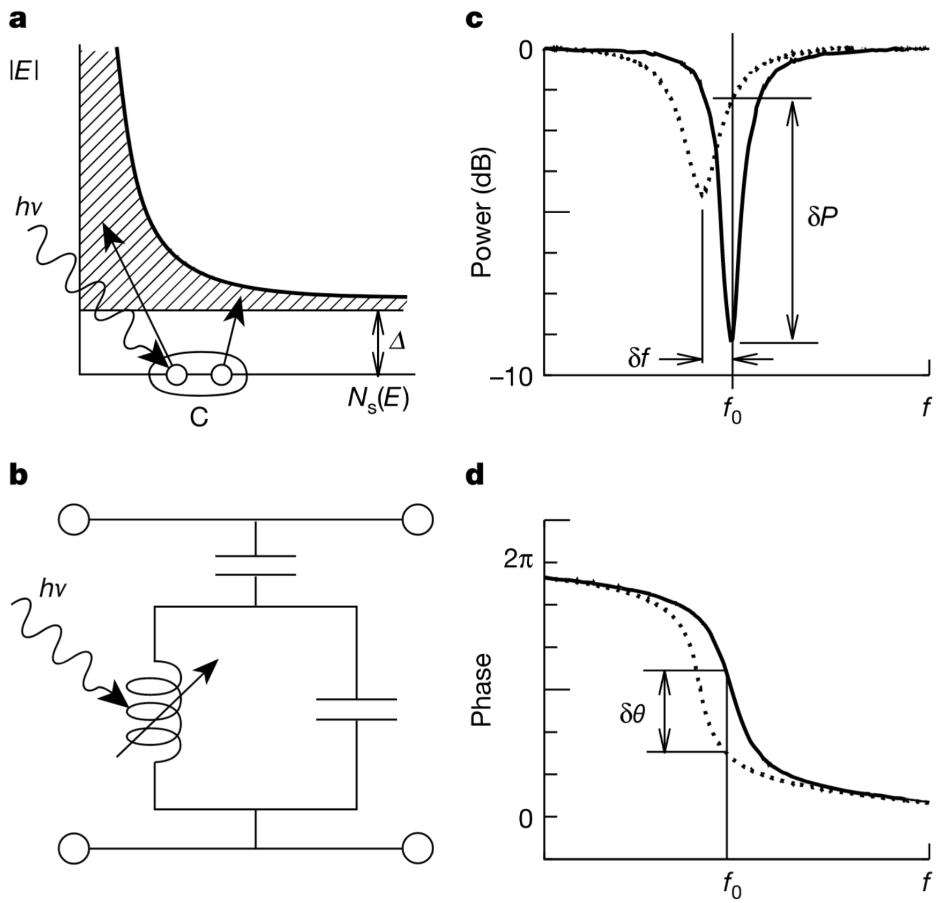


Figure 3.1: The KID working principle illustration reproduced from (Day et al., 2003). **a)** QP energy $|E|$ plotted against the QP density of states N_s (shaded). The elliptical object in the figure represents a Cooper pair on the Fermi surface formed by two electrons. In this example, a photon with an energy of $h\nu \geq 2\Delta$ excites the two electrons across the QP generation bandgap Δ into the shaded QP population. **b)** The equivalent circuit model for a KID. The top and bottom lines represent the two polarities of the feedline. The capacitor between the LC resonator and the feedline represents the capacitive coupling. The LC resonator is the KID with a variable inductance due to external energy deposition. The external energy is again shown by the photon, but we will elaborate later the working principle applies identically to phonon detection. **c)** The transmission (S_{21}) in amplitude versus frequency. f_0 is the resonant frequency. When the resonance shifts from the solid line to the dotted line due to excessive QPs, the readout signal at f_0 experiences a fractional transmitted power change of δP as marked in the figure. **d)** A similar diagram to c) but for the S_{21} in phase, so in this case the signal is given by the phase change $\delta\theta$. In c) and d), the solid Lorentzian curves represent the steady-state resonance before detection, and the dotted curves illustrates a detection with an elevated QP density relative to the pre-detection state.

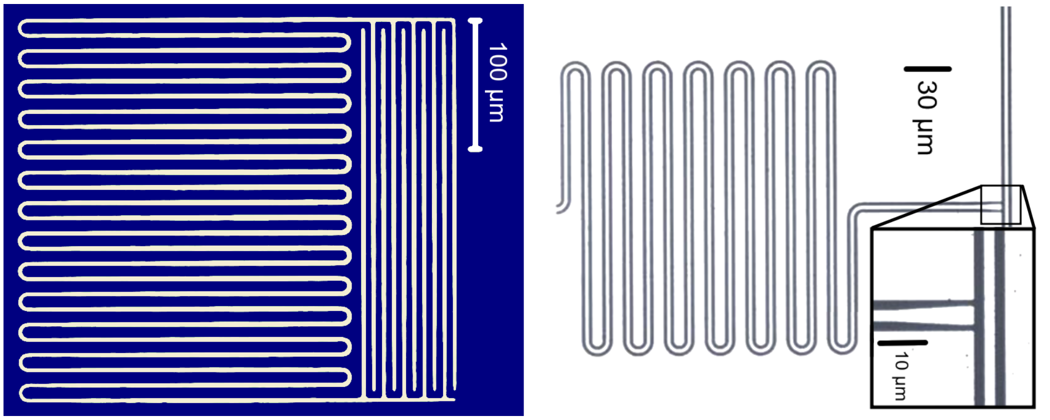


Figure 3.2: Left: The original (Doyle, 2008) LE KID. The single-trace meander on the left is the inductor, and the interdigitated parallel traces on the right combine to be the capacitor. Right: The original (Day et al., 2003) $\lambda/4$ transmission line KID. The inset shows the capacitive coupling design. More information may be found in the corresponding references. Both photos are reproduced with color adjustments for better visibility.

It is easier to understand the kinetic inductance in the context of lumped-element (LE) LC resonators, where one can typically identify isolated inductor(s) or certain portion(s) of the circuit layout as the main contributor(s) of the inductance. Without going into the detail provided later in Sec. 3.2, within a few years after (Day et al., 2003)'s first $\lambda/4$-transmission line resonator realization, the concept was also realized on LE LC resonators independently by multiple works, e.g., (Doyle, 2008; Moore et al., 2012). In the LE resonator design, the inductor and capacitor components within the resonator are distinctly defined and are visible in the forms of, taking the typical design in Fig. 3.2 as example (Doyle, 2008), interdigitated parallel traces and a long meandering trace. The former serves as a capacitor whose interelectrode gap is filled by the substrate dielectric material and vacuum, and the latter creates a finite-area current loop for its own magnetic field emission and thus becomes the inductor. It is clearer to explain with this picture that the meandering current contributes an inductance even in the absence of the SC film's sheet impedance. The total capacitance of the LE is nominally dominated by the interdigitated capacitor by design.

The *kinetic* inductance refers to the additional contribution to the total inductance that arises from the surface impedance of the SC material, which stems from the kinetics of the Cooper pairs and typically has the surface density unit of nano-Henry per square (nH/□). Due to the effective mass of Cooper pairs when being oscillated by the penetrating microwave readout field, which depends on the QP

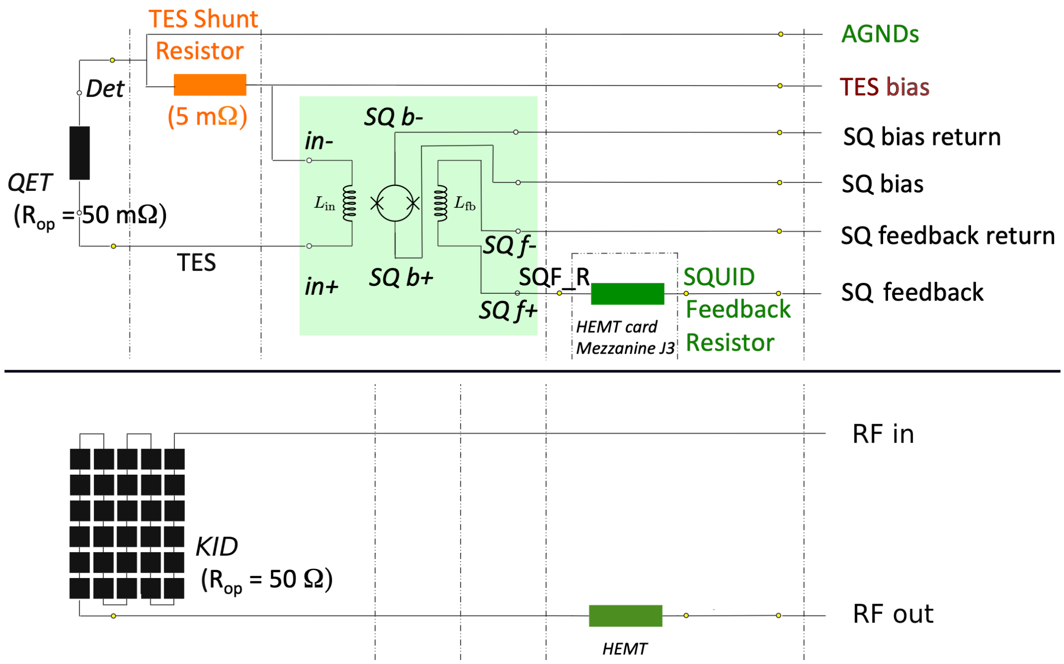


Figure 3.3: The control and readout electronics diagrams for TES (top) and KID (bottom). The orange- and green-shaded subsystems in the TES diagram represent the electro-thermal biasing and the SQUID flux-locked feedback amplification, respectively. In both diagrams, HEMT amplifiers are assumed for the subsequent readout amplification. Each black rectangle in the diagrams represent one TES or KID sensor. From right to left, the regions separated by the vertical dashed lines are at room temperature, 4 K, 0.7 K, 0.2 K, and 30 mK. More information about the TES readout may be found in (SuperCDMS Collaboration, 2018). Top figure produced by P. Brink.

density, the name therefore emphasizes that the *sensitive* inductance underlying the detection principle is the *kinetic* inductance that varies with the QP density under external energy injections. The kinetic mechanism is analogous to the Drude-model AC inductance for classical conductors due to the effective mass of their charge carriers, i.e., the damping frequency (Jackson, 1998), but for classical conductors, the practical kinetic inductance are generally to small to realize a LC resonator below THz.

3.1.2 Motivation

The advantages of using KIDs for phonon sensing include: First, the readout is in principle straightforwardly achievable with the by-design 50 Ω feedline, which is immediately compatible with a variety of commercial microwave electronics. However, we should note that, as we will discuss in more detail in the following and the next chapters, at the moment designing a satisfyingly 50 Ω-matched feedline remains a challenging RF engineering task, especially when the feedline extends comparably or more widely than the resonant wavelengths of the KIDs. Nevertheless, we can still take the TES technology currently utilized by SuperCDMS detectors to illustrate KID's potential advantage. Due to TES' small-impedance nature, the TES technology requires dedicated treatments for the readout. In particular, SuperCDMS currently adopts the SQUID amplification readout scheme (Hines et al., 2011; Ren et al., 2020; SuperCDMS Collaboration, 2018), where the small TES current signal is magnetically coupled between SC SQUID coils for a flux-locked room-to-cryogenic temperature feedback readout (Sec. 2.2). Compared to the well-tested commercial RF amplifiers, the SQUID flux-locked readout scheme is relatively prone to interference, such as flux trapping or thermal instability, etc., if not hosted by a carefully designed and constructed cryogenic system. As shown in Fig. 3.4, TES requires much more complex cold and warm electronics to be controlled and read out, which reduces its human- and infrastructure-error robustness as well as cost effectiveness relative to KID. As a result, the TES-based phonon sensors in the previous-generation (Soudan) SuperCDMS experiment yielded ≈3/4 of the readout channels (Doughty, 2016), and as elaborated previously, HVeV Run 2 also lost ≈10% of the exposure due to the loss of SQUID control (Sec. 2.5). As introduced later in this chapter, we have demonstrated with an 80-KIDs prototype detector that we project to achieve more than 90% of a KID sensor (channel) yield for 250 KIDs in one detector, while for the prototype detector containing less than a hundred KIDs, the resonance identification limit leading to the 90% projected yield did not cause a loss of any KID in the prototype detector.

Second, because the KID senses the QPs created from the quiescent Cooper pair bath strictly in the SC phase, it is possible to achieve a sensitivity that is independent of the device temperature, as long as the device remains much below T_c. However, we should point out that this feature is only beneficial in avoiding the need for a dedicated temperature control but does not suggest the energy sensitivity of a KID-based detector is commonly insensitive to the device temperature. Effects that remain sensitive to the temperature at much below T_c, such as the generation-recombination noise analyzed in the next chapter, can still be the sensitivity limiting factors and lead to a temperature-dependent detector sensitivity. For the prototype detectors developed in this thesis, we fabricate our devices with Al films for the KIDs, which exhibit $T_c \approx 1.2$ K, and we find the operating temperature becomes irrelevant in practice at below ≈100 mK. The result is consistent with our expectation that the current sensitivity is limited by the noise of the readout electronics, which are mounted at higher temperatures and therefore do not affect the detector sensitivity under device temperature fluctuation. So, under such a condition, we may mount the devices to the base-temperature plates of any properly functioning dilution refrigerators (DRs), and the KIDs should work without a dedicated temperature control mechanism and provide a consistent sensitivity. For a comparison, we have introduced the relatively complex temperature control scheme for driving the TES into the transition state (Sec. 2.2), where the bias control needs to be applied individually to each TES due to practical T_c and parasitic power variations. Similar to the need for the SQUID readout, the TES bias scheme leads to practical cryogenic electronic and thermal engineering challenges and increases the associated cost. Together they have gradually become the limiting factor for the total readout capacities for large TES-based instruments, such as CMB-S4 and SuperCDMS SNOLAB (Abitbol et al., 2017; SuperCDMS Collaboration, 2018).

Third, not only does the KID operate on its own without complementary cryogenic readout or control electronics, the signals from multiple KIDs can also be combined and transmitted using one shared feedline for later decomposition at room temperature. It is achieved by KID's frequency-domain narrow-resonance detection scheme relative to the traditional time-domain discharging pulse detectors, e.g., PMT (photon multiplier tube), diode, TES, etc. The S_{21} signal for each KID is labeled by a distinct resonant (carrier) frequency that is not lost when combined, which is not the case for a current signal if transmitted in a shared line. Therefore, in addition to be free from supporting cryogenic electronics, the number of readout cables per DM detector does not increase with the phonon sensors instrumented on the detector. In practice, one achieves this frequency-domain multiplexing by, instead of sending one monotone readout signal through the feedline as illustrated previously, all the KIDs are coupled to a shared feedline, so a readout signal combining all their resonant frequencies is used instead. As long as the KID resonances are sufficiently isolated, each frequency in this "frequency comb" readout signal only resonates with its corresponding KID and therefore is distinguishable in the frequency domain. By knowing the mapping of the *by-design* physical KID-to-resonant frequency correspondence, the composite output comb signal carries all the signals from the KIDs with distinct contributions that are demodulated individually at room temperature with commercial RF equipment. Given that off-the-shelf modern commercial RF electronics provide 10^2 MHz bandwidths, and thin-film SC resonators have been shown to exhibit quality factors of $10^{5\text{–}6}$ in GHz ranges, with one set of commer-

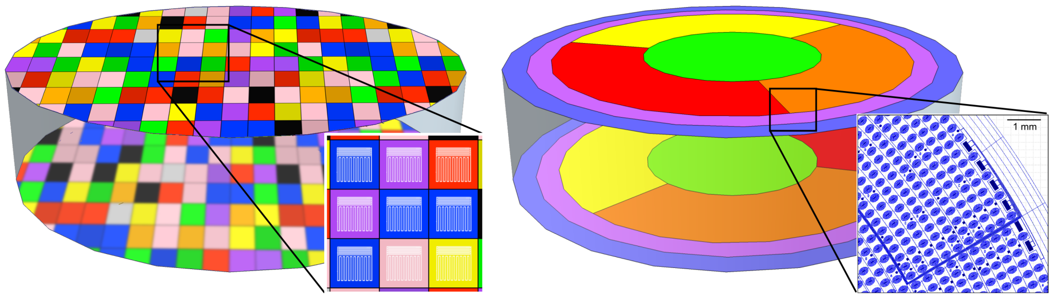


Figure 3.4: Phonon sensor mappings for the TES-based SuperCDMS SNOLAB HV-type detector (right) and a projected KID-based highly multiplexed detector (left). The zoom-in figure for the SNOLAB HV detector shows the QET phonon sensors by the blue ellipses, and the zoom-in for the KID-based detector is intended to illustrate that each KID serves as a separated phonon sensor. The HV detector and the QET pattern figures are reproduced from (Agnese et al., 2017) and(SuperCDMS Collaboration, 2018).

cial RF readout equipment on a single readout line and no additional multiplexing electronics, we may conceivably carry 10^3 KIDs with one readout cable, which is otherwise a very challenging number for technologies that scale the supporting electronics with sensor count.

Fourth, now supported by the dramatically improved experimental scalability and detector multiplexibility, we present in Fig. 3.4 a comparison of phonon sensor layouts for a TES-based SuperCDMS detector to a conceptual design utilizing KIDs as phonon sensors. The readers will find in later chapters a highly multiplexed detector has indeed been realized as a major contribution of this thesis. Fig. 3.4 shows, with the KID's inherent multiplexibility, we picture covering the phonon sensor-instrumented surfaces with hundreds of $\lesssim$mm-size LE KIDs similar to which shown in Fig. 3.2. Each KID operates as an individual phonon sensor yet shares a single feedline for the readout, contrasting with SuperCDMS' current distributed-QET phonon sensor design, where the detectors are similarly covered by numerous $\lesssim$mm-size individual QET phonon sensors, but the sensors are connected to be a few large-area joint phonon sensors as in Fig. 3.4 or Fig. 2.1. We explained in Ch. 2 that the motivation for the distributed-QET sensor design is that, aside from the aforementioned challenge to read out so many QETs separately, it is well-established that the phonon pulses observed at different locations on the surfaces exhibit appreciably different pulse shapes due to incompletely understood or modeled detector physics (Pyle, 2012; Kurinsky, 2018). Since such a position dependence may interfere with the detector performance through, for example, a pulse shape variation that degrades the optimal filter's energy reconstruction ability and subsequent its resolution, the current solution taken by SuperCDMS detectors is to expend the distributed-QET sensor coverage to suppress the position dependence, which leads to the sensor mapping in Fig. 3.4. The three wedge-shape sensors "triangulate" the event position based on their relative signal sizes, with the central sensor providing the least biased energy estimate. There are two additional outer rings designed for this particular detector. Recalling the high-radius background observed in HVeV Run 2 (Fig. 2.25), the purpose of these rings is to identify this type of near-sidewall events that are commonly distorted and are otherwise difficult to reject under the HV detection scheme. However, the current approach developed for the QET layout has also been shown to only partially remove the position dependence, and the "guarding ring" technique also does not fully distinguish the sidewall background due to surface event distortion (Kurinsky et al., 2020). With only the few large-coverage sensors, very shallow events that are heavily distorted near the surfaces may potentially lead to misreconstructed events and present themselves as an irreducible background in more sensitive future DM detectors. We therefore propose KID as a promising alternative of the TES-based phonon sensing with orders-of-magnitude finer pixelization for future phonon-mediated rare-event particle detectors.

Fifth and finally, KID-based phonon-mediated detectors may present other application-dependent features that are relatively preferable in certain aspects. For example, we currently fabricate our KIDs with aluminum, the material SuperCDMS and other similar applications identify as a suitable material for a simultaneous QET phonon absorber and QP trap (Sec. 2.2). Being able to fabricate KIDs with Al means the Al film not only provides a high kinetic inductance suitable for sensitive KIDs but also allows the KIDs to become their own phonon absorbers. In other words, we simplify the fabrication and the associated hetero-material QP transport complexity by completely removing the dedicated phonon-absorbing structure. Such a difference naturally provides benefits in fabrication yield, reproducibility, and also cost, as well as permits more design flexibility in the phonon sensor layout. Nevertheless, being utilized simultaneously as the KID and also the phonon absorber, the constraints applying to one role may also limit the other. In order to leverage kinetic inductance's inverse quadratic dependence in thickness in the thin film regime, it is usually preferable to fabricate the KIDs with films that are much thinner than the material's EM penetration depth. Specifically for Al, the requirement implies $\lesssim$40 nm for the film thickness, which contradicts the optimization for efficient phonon absorbers, where 100s nm is usually chosen to ensure the energy carried by the pair-breaking phonons is efficiently exhausted before returning to the substrate. Meanwhile, concerning future upgrades from Al KID, by the end of Ch. 4, we recommend replacing Al with lower SC bandgap materials as one potential path for producing KIDs with lower energy resolutions and detector thresholds. The approach might as well invalidate the statement that KID-based phonon-mediated detectors do not require a dedicated phonon-absorbing mechanism. The possibility remains to be evaluated in the future.

Another advantage of KID is that it is in principle more sensitive when operates arbitrarily colder, which is not always the case for different sensor technologies. Taking the TES for comparison, it simply can not operate elsewhere except at the transition temperature. As it turns out, the in-transition operation results in a phonon noise-dominated condition that determines TES' sensitivity. It has been demonstrated that, by fabricating detectors with different T_c, the TES energy resolution scales as T_c^4, which agrees with the theory for a phonon noise-limited condition (Pyle, 2012). Therefore, to improve the energy resolution under this condition, the path forward for future TES is lowering T_c for lower bias power devices. The approach requires perfecting the low-T_c SC film growth technique, and it has been demonstrated on several test devices that it is possible to fabricate W TESs with T_c as low as 20 mK. On the other hand, under the nominal "dark" operation condition with minimal external energy deposition for DM experiments, it has been demonstrated that one can suppress the GR noise to be subdominant with respect to the readout electronics, therefore removing the KID-intrinsic noise contribution from the practical factors affecting the sensitivity. The KID sensitivity under the low GR noise condition generally becomes dominated by the first-stage low-noise amplifier (LNA), and an immediate advantage of the LNA-dominated operation principle is that it is straightforward to achieve a resolution improvement by swapping LNAs to lower-noise products. We discuss later in this thesis that, based on already demonstrated dark environment techniques, one can suppress the GR noise to be subdominant to the quantum-limited noise, in other words allowing any LNAs that do not invoke quantum squeezing techniques to be the dominant noise sources. Meanwhile, works have also been done to create LNAs that exhibit quantum-limited performances. So, combining the two, it is promising for future KID devices to offer a quantum-limited sensitivity based on already demonstrated solutions.

3.2 Previous works

3.2.1 From $\lambda/4$ transmission line to lumped element

KID was first realized by (Day et al., 2003) with Al $\lambda/4$-transmission line resonators using the coplanar waveguide (CPW) design for both the feedline and the KIDs (Fig. 3.2). In the following few years, works achieved the SC resonator-based frequency-domain multiplexing architecture utilizing physically distinguishable inductors and capacitors as LE resonators, e.g., (Doyle, 2008) (Fig. 3.2). The motivation for adopting the LE design was that, in order to sense a stronger KID inductance change due to a pair-breaking event in the KID film, the event needs to occur at a location with preferably a stronger current density. One therefore prefers a stronger readout signal that produces a higher absolute current density in the KID. However, while one can always increase the readout signal for the absolute current density, the relative current density variation in the KID does not change for the same KID design. So, with a high current density variation, the consequent sensitivity to QP density fluctuation becomes equally nonuniform in the KID and is not improved by adjusting the readout signal level. Considering the $\lambda/4$ standing-wave nature of the original transmission line resonator design, we may understand the design is fundamentally impossible to avoid large current density variation along the transmission line. On the contrary, a LE KID is by definition much smaller than the wavelength, therefore the readout RF signal experienced by the KID is approximately uniform across the region of the KID, resulting in a current density that is much more uniform compared to the $\lambda/4$ design. In practice, to realize a microwave resonator that is small enough to be a LE, (Doyle, 2008) adopted the interdigitated capacitor (IDC) design, which provided a sufficiently large capacitance with the field running between neighboring IDC parallel traces, or commonly called the "fingers," through the substrate and the vacuum above. (Doyle, 2008) then chose a microsctrip meander design for the inductor, which provided the necessary inductance for the IDC at a small enough size that sufficed the criterion of a LE. To simplify the circuit layout, the meander used in (Doyle, 2008) was the simplest single trace but not the CPW utilized by (Day et al., 2003). (Doyle, 2008) also showed that, with the IDC design, one may easily adjust the capacitance to achieve a systematic resonant frequency placement, by incrementally tuning the length and/or number of the IDC fingers. Owing to its reproducibility and versatility, as of today, the "IDC-single trace meander" LE KID design continues to be one of the widely employed KID layouts.

3.2.2 Theoretical development

During the same era, multiple works established the fundamental physics of KID, and in many cases the general physics of thin-film SC resonators. Among these theoretical developments, the most noticeable works are by Gao and collaborators, published collectively as Gao's PhD dissertation (Gao, 2008). Starting from Mattis-Bardeen theory (Mattis and Bardeen, 1958), Gao et. al. established an analytical model for the surface impedance fluctuation for SC films due to generic pair-breaking mechanisms, such as temperature fluctuation or external energy deposition. They continued to apply the result to a $\lambda/4$ transmission line-based resonator circuit model, i.e., assuming (Day et al., 2003), which then yielded the theoretical model describing KID resonances' steady-state as well as transient detection responses. Until today, this model is still generally recognized as the most appropriate baseline model for almost all practical thin-film SC resonators. In most cases, it describes the data fairly well without invoking design-specific modifications, especially for elemental superconductors. In addition to the model-building, since the model was based on Mattis-Bardeen thoery and therefore naturally related thermodynamic properties to the QP density, (Gao, 2008) was able to prescribe a rigorous procedure for calibrating the KID as postulated in (Day et al., 2003). By steadily controlling the KID into different temperature equilibria, (Gao, 2008) demonstrated it is possible to simulate the QP creation due to external energy deposition by the calculable thermal QP creation at different temperatures. The procedure therefore calibrates the pair-breaking energy scale along with many characteristic intrinsic properties of a KID. Just like the model itself, this technique has also been widely adapted to different KID applications nowadays.

Following (Gao, 2008), the theoretical development for the physics of KID naturally proceeded to 1) more realistic QP dynamics different from (Gao, 2008)'s homogeneous assumption, 2) the physics extrinsic to the KID or its QP system that causes the realistic QP dynamics, and 3) application-specific modifications or further derivations for the (Gao, 2008) model, including both SC physics that applies to individual KIDs and also the microwave physics that addresses the overall construction of the KID-feedline-environment integration. However, currently many of the above works are derived from the R&D efforts under particular projects for KID-based instrument developments. We find the knowledge is not as easily accessible or systematically organized as in (Gao, 2008) and often attached to the literature focusing on obtaining physics measurements or as internal documents. For this reason, we try to systematically organize as many materials involving the fundamental physics of our application in the next chapter, while some of them are not yet published or not in easily accessible sources. Concerning the QP dynamics, the references we most frequently refer to are the internal notes of Caltech Observational Cosmology Group written by Zmuidzinas in 2008–2012, which are largely summarized in (Zmuidzinas, 2012) as a comprehensive review article. In addition to the QP dynamics, (Zmuidzinas, 2012) also summarizes various empirical or semi-first-principle models for the energy transport and pair-breaking mechanisms, such as the readout power, the thermal or non-thermal radiation, and the losses to the substrate's two-level system (TLS) or other dissipative channels. Particularly for the TLS, (Gao et al., 2008) presented based on (Phillips, 1972)'s general amorphous material state tunneling model the first detailed theoretical study, which also provided an experimental evidence supporting that the $1/f$ noise frequently dominating KIDs in the resonant frequency signal (df)[2] is likely due to the TLS in the substrate.

3.2.3 Low radiation interference operation

There are two major types of unintended extrinsic "background" energy sources that contribute to the QP generation in KIDs: One, large-area direct radiation exposure and, two, the readout power QP generation. Especially for astronomical instruments that are designed to be exposed to external radiation, a direct radiation exposure is inevitable and therefore generally dominates their performance, i.e., the photon noise-limited sensitivity. The fundamental quantum theory for the photon shot noise was first presented by (Zmuidzinus, 2003) and has been achieved in modern KID-based instruments, e.g., on Al LE KID (Mauskopf et al., 2014) and more recently the thermal KID (TKID) (Wandui et al., 2020). However, the photon noise-limited condition is not immediately relevant to our DM search application, since for most DM direct detection experiments, not only do the detectors operate inside light-tight enclosures, but any types of known radiation are also carefully shielded to the lowest achievable level. While one might then be convinced that, under such a condition, the concern of large-area radiation exposure is then alleviated from practical relevance, but on the contrary, (de Visser et al., 2011; Baselmans et al., 2012) found that the QP dynamics in the SC films may still be significantly affected solely by the blackbody (BB) radiation emitted by the surrounding cryogenic mechanics hosting the KIDs. The authors further demonstrated in (Baselmans et al., 2012) and (de Visser et al., 2014) that, by adopting a specialized detector housing scheme, they were able to fully suppress the BB radiation interference

[2] See Ch. 4 for detailed derivations and nomenclatures.

and subsequently achieve the highest pair-breaking sensitivity that was limited by other non-QP-origin losses. The special housing scheme, dubbed the "box-in-box photon labyrinth" by the authors, consisted of multiple light-tight detector enclosures with customized mm-wave radiation absorber coated on the surfaces. Fig. 3.5 reproduces the original results by de Visser et. al. Enabled by the the ultra-low radiation housing, the authors then observed that, in such a low thermal radiation environment, the radiation traveling with the manually injected KID readout microwave in the feedline may couple to the KIDs and together become the dominant pair-breaking energy source that limits the device sensitivity. This is the latter type of the undesirable extrinsic energy that increases the QP density, which leads to a suppressed QP lifetime and thus a reduced responsivity. The increased QP density also results in a raised QP generation-recombination noise. To reduce unnecessary pair-breaking energy injection via the feedline, the authors installed customized inline IR filters to block the Johnson-Nyquist thermal radiation in the coaxial cables that feeds the readout microwave through the ultra-low radiation housing. Finally, with both pair-breaking radiation mitigation strategies, the authors needed to reduce the original readout power to achieve the least necessary QP generation with the readout signal. De Visser et. al. (2014) demonstrated an empirical procedure for the readout-power optimization, and we will present a rigorous first-principle model for the optimization procedure in the next chapter.

3.2.4 Microwave performance

Various more advanced applications strengthened by the refined model began to emerge after 2010. Evolving beyond the original isolated single-KID model, (Noroozian et al., 2012) presented a model for the inter-KID microwave coupling utilizing the (Doyle, 2008) LE KID design. (Noroozian et al., 2012) successfully describes the data of a tightly-packed KID array, where the KIDs behave as weakly coupled oscillators perturbed from their original frequencies by the observable capacitive interconnections. The result shows that, when arranged sufficiently close in physical space, the KIDs are not only capacitively coupled to the feedline but also pairwise linked to one another via nonnegligible capacitances, which results in observable resonant frequency offsets relative to the supposedly perfectly isolated situation. The work was among the early successes in realistically implementing KIDs for highly-multiplexed tight arrays, which was anticipated by the frequency-domain multiplexing scheme since (Day et al., 2003) but had not been realized due to the practical inability to reproduce the designed resonant frequencies in fab-

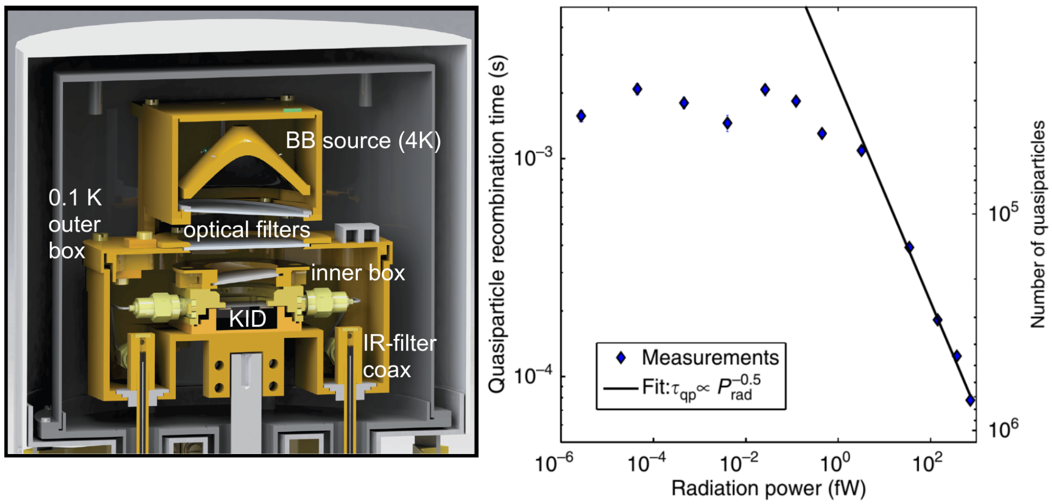


Figure 3.5: Left: The original (de Visser et al., 2014) experimental construction with components labeled, consisting of a BB radiation source at 4 K, a 0.1 K outer enclosure, a base-temperature inner enclosure, IR absorber-filled coaxial readout cables, the KID device, and a series of radiation filters on the photon path between the BB source and the device. The inner walls of the enclosures are all coated with mm wave-absorbing materials. Right: The observed QP lifetime as a function of the impinging BB radiation power calculated from the controlled source temperature. The data deviate from the BB photon-dominated regime, represented by the fitted theoretical curve, when the radiation power is lowed to 1 fW and below. Figures reproduced from (de Visser et al., 2014)

ricated devices. For previous implementations, the practical limits for the tightest confusion-less resonance packing in their readout bands were not the intrinsic widths of the resonances, but the much wider frequency ranges needed for the uncertain resonant frequency placement. Based on their microwave crosstalk model, (Noroozian et al., 2012) was able to contribute a dominant fraction of such frequency scattering to the coupling of the resonances. In addition to quantifying the frequency displacement, the authors further demonstrated one may suppress the inter-KID coupling by, one, replacing the single-trace meander in (Doyle, 2008) LE KID with a coplanar stripline (CPS) and, two, enclosing each LE KID with the so-called "ground shield." In the first technique, a CPS contracts the guided RF field that otherwise radiates to infinity for a single trace to predominately between the parallel traces. For the second, the ground shield isolates the entire KID from its neighbors and is shorted to one side of the CPW feedline's ground trace. The upgraded LE KID layout by (Noroozian et al., 2012) is presented in Fig. 3.6.

Large/dense KID array projects started to realize that, due to better readout signal

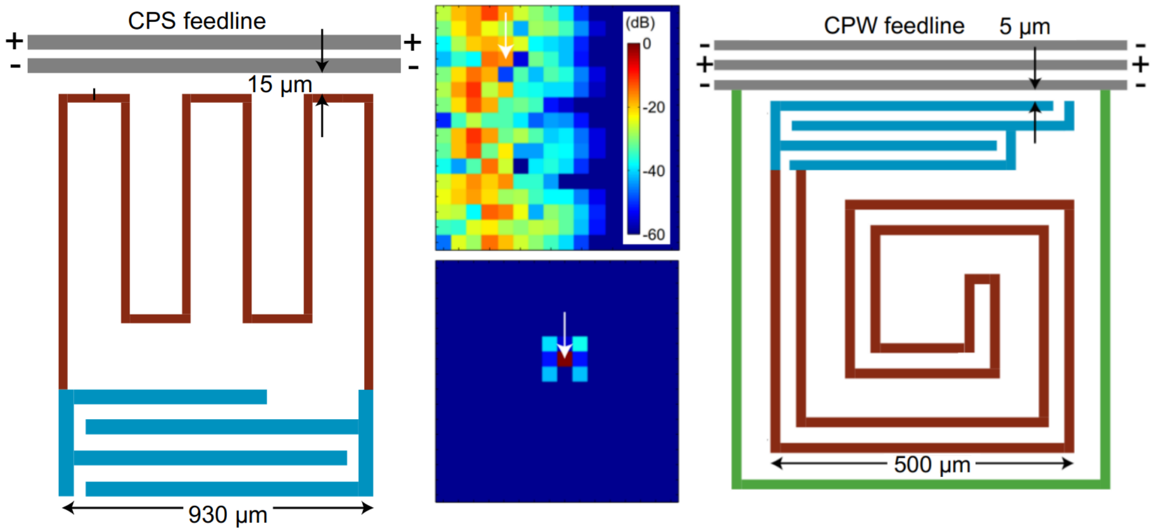


Figure 3.6: The Noroozian et. al. (2012) LE KID designs before (left) and after (right) the modifications for microwave crosstalk mitigation. Note that the two layouts are in different scales as noted by the feedline-KID distances. The inductors, the capacitors, the feedlines, and the ground shield are in brown, blue, gray, and green, respectively. The two intensity maps at the center show the calibration images taken by the KID arrays before (top) and after (bottom) the upgrade, where the arrow indicates the illuminated pixel. Figures reproduced from Noroozian et. al. (2012).

concentration from the feedline sections to their corresponding KIDs, waveguide designs with less out-of-plane radiation typically exhibit better overall transmissions and reproducibilities. The underlying mechanism is much similar to the inter-KID coupling found in Noroozian et. al. (2012), and for the feedline, one major way to form a stronger unintended long-distance coupling is with a dipole far field, which decays much more slowly than a far field dominated by higher multi-poles. In practice, such a net dipole asymmetry always exists in unbalanced type waveguide designs, such as CPS, while it can also occur locally in balanced type waveguides like CPW. To correct this practical imperfection, in addition to avoiding unbalanced waveguide designs, techniques such as wirebonding ground planes across CPW central trace or even dedicated micro-fabrications for intricate ground bridges began to gain popularity for improving the transmissions and the resonant frequency placement, e.g., Mazin et. al. (2012). In the meantime, with more KIDs fabricated and characterized with (Gao, 2008)'s resonator model, it was noticed that realistic KID resonances occasionally showed appreciable near-resonance phase offsets relative to the feedline transmission. This extra phase is strictly speaking not supposed to occur in the (Gao, 2008) model assuming a perfectly matched feedline but nevertheless empirically accommodated by an artificial phase term in the formula. Since the empirical phase parameter in the (Gao, 2008) model was only used to stabilize/improve the fit, it was not expected to be large, otherwise other more physical model-motivated parameters should be modified accordingly. Adopting a similar approach of Noroozian et. al. (2012) in modifying (Gao, 2008)'s circuit model with physical elements to better describe the data, (Khalil et al., 2012) noticed that the near-resonance extra phase may be phenomenologically modeled and therefore postulated the existence of an extra input capacitance in the feedline, which is only significant at near-resonance frequencies. This "capacitor" provides the impedance mismatch giving rise to the phase offset. In practice, the exact inline capacitance is difficult to predict in first-principle RF simulations due to case-by-case design, fabrication, and/or microwave environment differences, while it remains possible that the hypothesized capacitor is only a misinterpreted mathematical equivalence. Nevertheless, the consequent resonance formulae for the phase and the internal loss due to (Khalil et al., 2012), compared to which originally given by (Gao, 2008), do in many cases appropriately recover better consistencies between physical expectations and observations. Since the (Khalil et al., 2012) model continuously reduces to (Gao, 2008)'s formula in the absence of the postulated feedline input capacitance, it is now generally accepted as a better mathematical description for the resonance data, especially when the internal loss of the resonator is of interest.

3.2.5 KID-based phonon-mediated detector

The application of KID has been rapidly growing Since 2010 for not just proof-of-principle demonstrations but installations projecting substantial physics impacts. While the majority of the effort is in camera- or telescope-style applications, where the originally proposed direct photon detection scheme and the immediate multiplexibility advantage align well with the field's interest, it was first proposed by Golwala and demonstrated in (Moore et al., 2012) that the KID serves the same functionality of the TES in QET phonon sensors for SuperCDMS-like phonon-mediated particle detectors. Using a CPS feedline feeding twenty (Doyle, 2008)-style Al LE KIDs, distributed roughly evenly on a $2 \times 2 \times 0.1$ cm^3 Si substrate, (Moore et al., 2012) successfully demonstrated that the Al KIDs on this detector simultaneously served as SC pair-breaking sensors and their own phonon absorbers. Although the energy and position resolutions obtained did not immediately exceed their then QET-based SuperCDMS counterpart, the result nicely verified the energy resolution model developed by Golwala, which encouragingly forecasts a competitive future

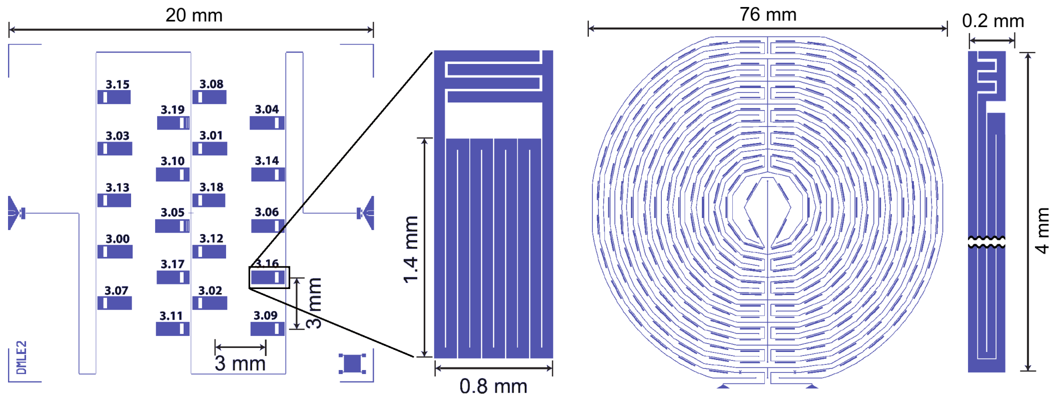


Figure 3.7: The detector layouts for the previous-generation KID-based phonon-mediated particle detectors for the DMKID project; left: (Moore et al., 2012), right: (Cornell, 2018). Note that the two layouts are of distinct scales as noted. Also notice that the zoom-in KID design panel for the (Cornell, 2018) layout is shortened to visualize details. The numbers attached to the KIDs in the (Moore et al., 2012) design denote their resonant frequencies in GHz. More details are found in the references. Figures reproduced from (Moore, 2012; Cornell, 2018).

resolution with clear paths for improvements. On the other hand, the obtained position resolution with the modest multiplexing design was already comparable with the SuperCDMS detector. In addition to the detector performance, (Moore et al., 2012) also showed that the much simplified readout apparatus promised by the frequency-domain multiplexing scheme did permit the experiment to perform on much simpler cryogenic and room-temperature infrastructures. Following (Moore et al., 2012), work had been done by (Cornell, 2018) in the hope to scale up the detector to standard 3" semiconductor substrates, who began by adapting (Moore et al., 2012)'s general architecture, namely the CPS-fed (Doyle, 2008)-style Al LE KIDs. However, (Cornell, 2018) proved that the up-scaling is not as straightforward as copying and expending (Moore et al., 2012)'s layout, since maintaining a comparable KID density on the substantially larger substrate requires more KIDs, which in turn leads to a much longer and more complex feedline routing that introduces unforeseen RF engineering challenges. The designs of Moore et. al. (2012) and Cornell (2018) are shown in Fig. 3.7.

There were new modifications implemented for the 3" devices in addition to just replicating the Moore et. al. (2012) layout. At the time, (Cornell, 2018) was able to take advantage of (Noroozian et al., 2012)'s revised LE KID design with the spiral CPS inductor for reducing crosstalk (Fig. 3.6). Cornell also decided to minimize the IDC for the KIDs and keep them all identical, so the phonon-absorbing material due to these current-less insensitive capacitors was kept minimal. Therefore, the choice led to a relatively uncommon frequency placement scheme by incrementally adjusting the dimensions of the inductors, as opposed to with capacitors in most other works. However, despite the hopefully improved KID design or the newly introduced inductor-based frequency tuning, Cornell's effort to scale up the detector was not a simple success. As it turned out, the CPS feedline option following (Moore et al., 2012) requires a rather narrow separation between the parallel traces, which did not present an issue for (Moore et al., 2012), since multiple identical small dies were fabricated simultaneously on a same wafer, permitting the freedom to select the highest-quality die for the experiment. On the contrary, the fabrication of the much extended CPS feedline for the 3" devices became a challenging practical constraint because of the narrow gap's vulnerability to micro-fabrication defects, e.g., dust particles or photolithographic imperfections.

Aside from the main fabrication difficulty that hampers the progress, the CPS designs at increasing lengths and routing complexities yielded a peculiar feature that did not appear in the previous generation devices. Cornell found that, roughly in accordance with the feedline complexity, resonance-like wide-band structures appeared in the feedline transmission (S_{21}) for KID-less devices. One hypothesis for the phenomenon was that, although the transmitted field was confined to a near-waveguide space relative to the single trace, due to CPS' inherent oscillating polarization formed by the opposite traces, the waveguide may appear as a weak yet non-vanishing radiating dipole at distance and therefore promote the coupling between the feedline and the environment. Considering the GHz wavelengths in discussion and the dimension of the copper device holder, it is plausible that the detector holder may become an EM cavity that resonates with the feedline radiation, resulting in the observed S_{21} structure. Since the 3" devices generally included feedlines that were much longer than the GHz wavelengths, while the feedline on the 2-cm chips should behave much more like a LE, it is also possible for the long feedlines to exhibit not just local dipoles between the CPS traces but also between unbalanced spots along the feedline due to traveling wave effects. In this case, the distances of these nodes, the transmitted wavelengths, and the size of the cavity/device holder all share the comparable length scale, making them more easily to couple to each other. Nevertheless, due to the already low fabrication yield and the obvious necessity to replace the CPS, Cornell did not proceed to resolve the hypothesis but rather started an effort of redesigning new feedline layouts with CPW, which we continue as part of this thesis. After the feedline design, we continue to refine the theory for KID-based phonon-mediated detectors which Golwala et. al. originally developed (Moore, 2012), improve the fabrication technique with our new detector design, and finally present an 80-KID 3" prototype detector with a much improved performance. Following T. Aralis' suggestion for combining "DM" and "M(icrowave)-KID" to be "DM^{2}KID," which was later simplified to "DMKID," we will frequently refer our KID-based DM detector development work as the DMKID project in the following chapters.

3.2.6 Ongoing and complementary works

Before introducing our research, we highlight a few ongoing efforts continuing in parallel with our works presented in this thesis (2015–2021) to conclude the section. We also summarize interesting recent progresses in complementary topics that are readily or potentially applicable for our application in the future. (Battistelli et al., 2015) (later became the CALDER collaboration) proposed adopting a detector design based on (Moore et al., 2012) onto TeO_2 substrates for a neutrino-less double beta decay ($0\nu\beta\beta$) experiment. The authors argued that, due to almost identical requirements to direct detection DM experiments except the energy range, KID's advantages in multiplexibility, scalability, as well as the anticipated energy resolution that motivate our effort also apply to $0\nu\beta\beta$ experiments. The authors presented that, by virtually replicating (Moore et al., 2012)'s original detector layout onto $0\nu\beta\beta$-emitting materials, the detector architecture naturally became a valid realization for $0\nu\beta\beta$ searches. Owing to the shared heritage for this thesis from (Moore, 2012) at a comparable timeline, CALDER's onward R&D progresses have always been inspiring and usually immediately applicable for our works. Following the original demonstrator, (Cardani et al., 2018) presented a detector that used an Al/Ti/Al-multiplayer film for its KID fabrication modified from the original single-layer Al. By proximitizing the Al with the sandwiched Ti, the detector showed that one may indeed adjust the SC pair-breaking energy to achieve a resolution enhancement, which is one of the main optimization path proposed by (Moore et al., 2012; Chang et al., 2018). Since then, the approach of replacing high kinetic inductance materials that are commonly adopted by current-generation KIDs (Al, TiN, etc.) with low-T_c, low-gap materials, such as PtSi, Hf, AlMn, etc., has been actively pursued by astronomical/cosmological instruments, e.g., (G. Jones et al., 2017; Walter et al., 2020; Mazin et al., 2020). Coincidentally driven by the search for the "super inductor" materials for SC qubits, (Grunhaupt et al., 2018) found that, by intentionally disturbing the Al film sputtering to prevent large-scale crystallization, it is possible to obtain highly granularized Al films that are at the border of becoming insulators. This kind of fine-grained Al film exhibits intrinsic kinetic inductances that are typically two orders-of-magnitude higher than traditional sputtered Al films. (Grunhaupt et al., 2018; Valenti et al., 2018) further identified the mechanism causing such a dramatic enhancement to be the formation of a meta-stable state(s) for the QP recombination process. Not only do the meta-stable state(s) change the kinetic inductance macroscopically, microscopically, because the meta-stable state(s) can trap QPs and thus prevent them from directly recombining with each other, the observed QP lifetime was extended by also orders-of-magnitude to a 0.1–1 sec. range. While the effects on KID-based phonon-mediated detectors if adopting such a novel material remains to be understood, a recent study by (Cardani et al., 2021) announced that a CALDER detector has adopted the granular Al for its KIDs and has not yet observed a significant improvement in the phonon energy resolution, where the authors cited acoustic mismatch between the film and the substrate that impedes the phonon transport as a potential explanation.

Surrounding the main KID R&D, there are also ingenious engineering developments driven by the growing interest for large-scale implementations of KID-based instruments that enhance the practicality and frequently the sensitivity as well. Among these achievements, the most important work that we must highlight due to our direct adoption is the flexible commercial GPU-based resonator array readout system (Minutolo et al., 2019). Before this system, the most commonly utilized platform for the microwave readout was the so-called ROACH system, which included a Xilinx Virtex-series field-programmable gate array (FPGA) that controlled external on-board RF electronics to coordinate the data acquisition, e.g., (McHugh et al., 2012; Siegel, 2016). Being FPGA-based, the ROACH system was in principle programmable/modifiable on the firmware level but required specialized skills in FPGA programming, certainly not as flexible and friendly as directly through commercially available softwares. The major contribution by Milutolo et. al. was the demonstration of a full migration from the FPGA platform to a completely software-defined radio (SDR) system, which has been proven much simpler to debug, upgrade, and integrate with analysis scripts written in widely used languages such as Python or C++. In addition to the major firmware-to-software-level improvement, the authors were able to further enhance the digital control and data acquisition performance for the commercial SDR hardware with off-the-shelf GPU processors, making the full system straightforward to be (re-)constructed from Git repository at a relatively low cost.

Another decisive engineering science development for KID and other sensitive microwave-range quantum devices is the recent realization of the (near-)quantum noise-limited amplifiers, which includes the Josephson junction parametric amplifier (JPA) (Castellanos-Beltran, 2010), the traveling-wave parametric amplifier (TWPA) (Macklin et al., 2016), and the kinetic inductance parametric amplifier (KIPA) (Klimovich et al., 2019). These amplifiers allow quantum devices to be read out at the lowest theoretically permitted noise and therefore achieve either the minimal readout noise or the shortest noise-suppressing averaging times. Of these, JPA is inherently narrow-band due to its resonance-based operation principle and thus not as useful for highly multiplexed KID devices. Comparing TWPA and KIPA, while both provide wide bandwidths that are suitable for practical frequency-domain multiplexing, KIPA has shown an unambiguous quantum-limited noise with a higher gain, while for TWPA only close to quantum-limited performances were demonstrated on lower gains. In fact, we are fortunate enough to collaborate locally with one of the leading groups in the KIPA research, the P. K. Day group at Caltech/NASA JPL. We dedicate Sec. 4.7 for a more detailed discussion on its working principle and our plan for implementation.

Finally, concerning more creative LE KID designs, it has been demonstrated that it is possible to perform an empirical post-fabrication resonant frequency adjustment by designing the IDC fingers into a "trimmable chain" shape (Liu et al., 2017). Such an approach, however, requires first mapping the resonance-resonator correspondence with a dedicated cryogenic LED system, whose complexity increases quickly with the device's multiplexing, and one then empirically trims each IDC finger to achieve the desired frequency placement based on the mapping. Although the procedure is fairly time-consuming, works have been done to show the method does lead to better frequency placement accuracies and subsequently higher resonance/resonator packings in the frequency/physical space. There are also designs replacing the IDC with a parallel-plate patch capacitor, which provides much larger capacitance for reducing the inductor size, hence allowing higher focal-plane[3] KID densities due to the reduced KID form factors (Coiffard et al., 2017; Beldi et al., 2019). For these parallel-plate KIDs to achieve high quality factors with as little in-substrate losses as possible, at the same time also proving more practical flexibility in fabrication, works such as (S. R. Golwala et al., 2019) have also be done in pursuing very low-loss amorphous dielectric material growth techniques.

[3]Currently this technique is mostly adopted by astronomical instruments.

3.3 Feedline design

3.3.1 CPW geometry

As explained in Sec. 3.2, we replace the CPS feedline used in (Moore et al., 2012; Cornell, 2018) with a CPW design to alleviate the fabrication yield constraint due to CPS' narrow gap. Being a widely adopted waveguide option, the (approximated) analytical model for the CPW impedance is commonly available in microwave electronics textbooks/handbooks such as (Simons, 2001; Wadell, 1991). In one form of the impedance formula,

$$Z_0 = \frac{30\pi}{\sqrt{\epsilon_{\text{eff}}}} \frac{K'(k)}{K(k)}, \tag{3.1}$$

where ϵ_{eff} is the effective permittivity of the waveguide. According to (Ghione et al., 1984; Ghione et al., 1987), for a CPW, ϵ_{eff} may be approximated by the average permittivity of the substrate and the environment hosting the CPW. In our case, that is

$$\epsilon_{\text{eff}} = \frac{\epsilon_{\text{Si}} + \epsilon_{\text{vac}}}{2} \approx 6.5, \tag{3.2}$$

where ϵ_{Si} and ϵ_{vac} are the Silicon and vacuum permittivities, respectively, We note that the simple average for ϵ_{eff} does not generally apply to other waveguide designs.

$$\begin{cases} K(k) &= \displaystyle\int_0^1 \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}} \\ K'(k) &= \displaystyle\int_1^{1/k} \frac{dx}{\sqrt{(x^2-1)(1-k^2x^2)}} = \sqrt{1-K^2(k)} \end{cases} \tag{3.3}$$

are the complete elliptic integral of the first kind (K) and its complement (K'), which take the argument

$$k = \frac{a}{a+2b}, \tag{3.4}$$

where a and b are, respectively, the widths of the CPW central trace and its separation from the ground planes. We will loosely call this separation "the (CPW) gap" in the following content if the context permits without confusion.

Aiming for a $Z_0 = 50\ \Omega$ and taking $\epsilon_{\text{Si}} = 11.5 \sim 12.0$, Eq. (3.1) yields

$$\frac{a}{b} = 1.7 \sim 1.8 \text{ (revised to 1.9, see later).} \tag{3.5}$$

In fact, the above computation is so standardized that, instead of explicitly performing the elliptic integrals numerically, we use online tools such as (Microwaves101.com, 2001) for a convenient scanning between the targeted characteristic impedance, the substrate permittivity assumption, and the required central trace-to-gap width ratio a/b. The critical conclusion from the calculation that benefits our application in practice is that, given typical semiconductor permittivities with ideally a 50-Ω match, CPW allows the central trace to be at a comparable width of the gap, i.e., $a/b \approx 1$–2 as in Eq. 3.5, while the ground traces/planes on the sides may in principle be arbitrary. The result effectively allows a CPW to be designed with a gap width that is at the same order-of-magnitude of its full width. On the contrary, analytical formulae or online calculators yield trace-to-gap width ratios of $12.5 \sim 13.8$ for a CPS of the same material and impedance target (Darwis et al., 2018; I-Laboratory, 2000). In other words, the CPS gap width needs to be order-of-magnitude narrower than the allowed total width of the waveguide.

Unfortunately, constraints for an applicable waveguide width do exist in reality. Above all, we will explain in the next chapter that we aim for certain total resonator quality factors for optimizing the detector sensitivity. Since we expect the feedline-KID coupling to dominate the total quality factor, the targeted quality factors are obtained through adjusting the separation between the KIDs and the waveguide. For a CPW, this distance adjustment is to-first-order unconstrained, since an ideal CPW has no design requirement on the ground trace width.[4] As long as the coupling distance needed is a few times larger than the minimal gap width allowed by the fabrication, Eq. (3.5) shows that there is sufficient space to fit the ground plane between the KID and the trace gap to construct the CPW geometry. Contrasting with a CPW, a CPS' impedance depends strongly on its trace width and therefore poses a corresponding lower limit for the coupling distance. Combining with the fact that the realistic coupling distances needed are typically around $\approx$100 μm, and our previous statement that CPS requires $a/b \gtrsim 10$, the coupling tuning for CPS effectively requires gap widths no more than a few μm. This limit eventually led to the fabrication difficulty in (Cornell, 2018), except that in fact (Cornell, 2018; Moore, 2012) had already attempted to ease the constraint by introducing other more advanced techniques allowing non-50 Ω feedline sections.

In addition to the constraint due to the coupling adjustment, we would also like to reduce the feedline's footprint on the detector in general, so less phonons are absorbed by these structures that do not produce signals. Based on the same total width argument, one may also argue that a CPW is potentially more flexible for

[4]"Ideal" in that one usually assumes the ground traces are sufficiently wide and thus contribute a negligible inductance per unit length compared to the central trace. Such an assumption could be invalid if the ground traces are of a comparable width to the trace gaps or the central trace.

	Parameter	Value	Unit
Si	Dielectric const.	11.9	
	Elec. loss tangent	0	
	Permeability	1.0	
	Mag. loss tangent	0	
Nb	DC resistance	0	$\Omega/\square$
	RF resistance	0	$\Omega/\square\sqrt{\text{Hz}}$
	DC reactance	0	$\Omega/\square$
	Sheet inductance	0.13 (0.05)	pH/$\square$
Al	DC resistance	0	$\Omega/\square$
	RF resistance	0	$\Omega/\square\sqrt{\text{Hz}}$
	DC reactance	0	$\Omega/\square$
	Sheet inductance	0.8	pH/$\square$

Table 3.1: The material parameters used for SONNET simulations. For most items in the table, the zero values are given by the superconductivity, and the units with "□" represent surface density type quantities. The parenthesized Nb sheet inductance value is the data value later obtained from our 300-nm Nb film, while the non-parenthesized one is the nominal value used in the simulation.

reducing the overall feedline footprint than a CPS. We should note that, such an argument is valid only when the CPW ground trace is sufficiently wide to not affect the CPW performance but is still narrow enough to be neglected. Otherwise, the ground trace width should be included, which then reduces the size difference of CPW and CPS. As a balanced-type waveguide, CPW also allows the adoption of the ground shield design for crosstalk suppression (Sec. 3.2), which would otherwise introduce a significant RF performance impact for unbalanced waveguides such as the CPS. We will also introduce in the next chapter that we have developed the technique for fabricating the feedline of higher bandgap materials to avoid phonon absorption, which helps with alleviating the feedline footprint constraint.

We continue to fine-tune the CPW design with more realistic geometries and parameters in the finite-element EM simulation program (SONNET, n.d.) following Eq. (3.5). We first choose $a = 20$ μm, which gives

$$(a, b, c) = (20.0, 11.5, 100.0)\ \mu\text{m} \quad \text{(revised, see later)} \tag{3.6}$$

according to Eq. (3.5), where we assume that we will adopt finite-width ground planes, and c is the ground trace width chosen ad hoc to begin the simulation. We also round the dimensions, in particular the gap width b, to 0.5 μm based on the finest photo-mask resolution offered by (Photo Science Inc., n.d.), the vendor we have been

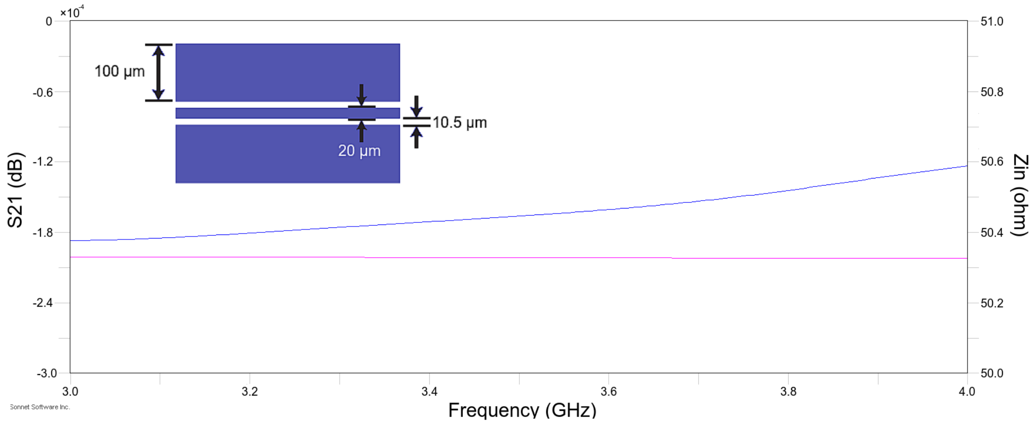


Figure 3.8: DMKID CPW design simulation. The blue and pink curves in the figure represents, respectively, the S_{21} in dB by the left axis and the characteristic impedance in Ω by the right axis; both plotted against the transmitted frequency in 3–4 GHz. Also shown in the top left is a schematic for the simulated CPW geometry.

working with for contact mask productions for the DMKID project. We first lay down a straight CPW using the parameters in Eq. (3.6), edge-to-edge on a 1 mm thick, 1 cm^2 area, square Si substrate. We set the CPW film thickness to be a variable in a practical range of 50–300 nm, knowing that such thicknesses (and variation) are negligible with respect to the CPW size according to analytical models that consider the film thickness. The simulation results indeed support the conclusion that we may freely choose the film thickness in the assumed range without a noticeable difference. We assume the feedline film to be SC Nb, which offers a higher pair-breaking threshold (2Δ) than the phonon energy in Si. We expect the QPs created in the Nb film by high-energy phonons to recombine and release the energy back into phonon modes, first in the Nb and then back into the Si substrate. On the contrary, SC Al with a bandgap lower than the Si phonon energy can not regenerate phonons and therefore is much more likely to permanently absorb the energy transported by the original incoming phonons. We use the material parameters summarized in Tab. 3.1 for the full simulation, where we inherit the Al and Si parameters from (Cornell, 2018) but in our case reserve Al only for the KIDs.

The rounded gap dimension b results in an impedance that is closer to 51 Ω than to 50 Ω, whose frequency dependence in 2–6 GHz is negligible compared to practical microwave electronics. To further minimize the impedance offset, we test b = 11.0, 10.5, 10.0, 9.5 μm and settle with

$$(a, b, c) = (20.0, 10.5, 100.0)\ \mu\text{m}, \tag{3.7}$$

which has been proven to be a robust design that we continue to apply for DMKID devices until today (2021). The simulation results are shown in Fig. 3.8. This 1 cm straight CPW simulated in SONNET exhibits a ≈0.3 Ω offset from 50 Ω in 2–6 GHz with a negligible variation. Its simulated end-to-end loss of 10^{-4} dB is negligible in practice and is likely a result of numerical simulation uncertainty.

Before proceeding to more complex detector designs with Eq. (3.7), we construct a KID identical to (Moore et al., 2012) beside this CPW feedline for a validation. By adjusting the distance of the KID and the CPW from 100 μm to barely touching the ground trace, we obtain the total quality factors for the resonance ranging from unidentifiably high in SONNET ($>10^7$) to absurdly low ($<10^2$). The result ensures that the ground plane width chosen ad hoc in Eq. (3.7), mostly just to avoid future layout and fabrication complexities, allows sufficient design flexibility for the target total quality factors.

3.3.2 Corner design

In order to realistically implement the CPW feedline, there is one more necessary ingredient in addition to Eq. 3.7, the turning/cornering of the waveguide. Fig. 3.9 shows the four designs that we simulate: 1) No modification (not in Fig. 3.9); 2) the traditional one- and two-section chamfers, which are the most commonly utilized designs in commercial RF electronics; 3) the minimal rounding, whose radius is determined by the width of the CPW; 4) the maximal rounding, whose radius is determined by the distance of parallel CPW lines. To include these corner schemes into the simulations, we modify the previous straight CPW on an 1×1-cm^2 substrate simulation to have a 90° corner at the center of the square substrate, i.e., connecting two 0.5-cm CPW sections that terminate at the centers of two neighboring edges.

For both 1) and 2), the simulations show clear oscillating transmissions (S_{21}) corresponding to the 0.5 cm sections, which indicates that standing waves are formed between the corners due to impedance perturbation. The sizes of the S_{21} oscillations for 1) and 2) are 0.1 dB and 0.06 dB, respectively. We attribute the oscillations to the introduction of localized impedance perturbations, because the periodicity of the oscillations is precisely as expected by the length of the waveguide section between the corners and the guided wave's group velocity $c/\sqrt{\epsilon_{\text{eff}}}$. The small but nonetheless unambiguous oscillation suppression comparing with and without a corner design also suggests that the chamfered corner provides certain level of reflection suppression. However, based on our past experience that each corner usually introduces

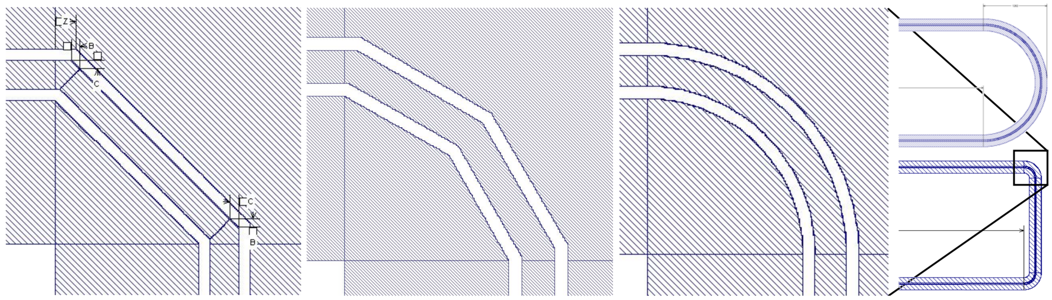


Figure 3.9: Different CPW corner schemes simulated in this work, from left to right, the one-section chamfer, the two-section chamfer, the minimal rounding, and in the rightmost two small panels the maximal (top) and minimal (bottom) roundings. The two small panels show that, in the top panel, the rounding radius is maximally limited by the distance of the preceding and following straight sections, while shown in the bottom panel for comparison, the minimal rounding is limited by the width of the CPW regardless of the straight sections it connects.

a sub-to-1 dB loss to the transmission, we suspect that the simulated $\lesssim$0.1-dB absolute losses are optimistic. As an idealized simulation in both material property and model geometry, it is most likely that SONNET underestimates the scattering of energy from one frequency to others at reflections. This effect is expected to lead to an underestimated oscillation, since the simulated result is only contributed by the exact frequency being simulated. In addition to this generic difference due to idealization, the model that we simulate is indeed physically different from that we fabricate. The 1-cm^2 simulated models are small enough compared to the GHz wavelengths and thus can behave as LEs, while in reality we frequently produce devices that are larger than the wavelength scale. It is possible that the impact of the corners are really suppressed in LEs, as localized features can be neglected, so the simulated effect may be realistic but under-represent the large devices we fabricate. However, when we initially obtained the result, we were limited by our computation resource from expanding the simulations to larger layouts to test the hypothesis. For future readers who would like to revisit the CPW study, we have upgraded the computer for the SONNET program as well as reconstructed identical CPW simulations in HFSS (Ansys, n.d.), another finite element simulation program known to be capable of simulating larger structures more efficiently than SONNET. While HFSS is also known to be generally not as accurate as SONNET in reproducing narrow-frequency structures, e.g., high-quality resonances, we have preliminarily confirmed that HFSS is capable of simulating the much more complex full CPW layouts for our 3" DMKID devices and reproduces S_{21} at an expected consistency.

For the smoothly rounded corners 3) and 4), SONNET does not show any noticeable structures in their transmissions, which is not unexpected given that they do not possess any geometric discontinuities as localized nodes. For the maximal rounding case, we round the CPW to a radius of 0.5 cm, effectively forming a quarter circle on the 1-cm^2 chip. This particular radius is suitable for connecting parallel CPW lines that are separated by 1 cm, which is roughly the scale we anticipate for layouts on 3" wafers. This design shows a very consistent transmission as if it is a straight line, therefore, despite occupying a relatively large area that limits the total number of resonators, it always serves as a reliable option for the CPW routing. In fact, all devices from the Mazin group at UCSB adopt the maximal rounding for their CPWs. At the opposite extreme, the minimal rounding corner scheme may be regarded as directly turning the CPW line while smoothly preserving the (a, b, c) values with respect to the center of the curve. Interestingly, instead of generating a periodic S_{21} structure as in the discrete designs, the minimal rounding corner exhibits a near-constant loss of 0.01–0.02 dB, with a mild hint for increase toward higher frequencies. While the result is, again, qualitatively understandable from the perspective of a constant loss per wavelength in the corner geometry, thus increased at higher frequencies, the simulated absolute loss is unrealistically low. Nevertheless, one advantage compared to the designs with discrete kinks is that the rounded designs do not impose a significant frequency dependence in the transmission. So, considering the simulated frequency-dependent structures are likely to be more pronounced in real devices, we decide to adopt the rounded corner scheme for our CPW feedlines.

3.3.3 Layout automation

We have created a Python library for the CPW feedline design in order to standardize the design process for the 3"-wafer devices of this thesis as well as more complex future layouts. The library may be downloaded from (Chang, n.d.). Using (gdsCAD, n.d.) as base library, our library provides tools that enable a fast construction of CPW layouts in the `.gds` file format. It allows the users to specify an array of 2D coordinates, and then the coordinates are connected with straight-line segments in the specified order as the *central line* of a CPW. At the intersections of the line segments, the CPW is reshaped to the minimally rounded corners discussed previously. To produce the corner layout, the algorithm first determines the arc center based on the angle of the straight sections the corner connects. The algorithm then determines the radii with respect to the arc center for the edges of the central

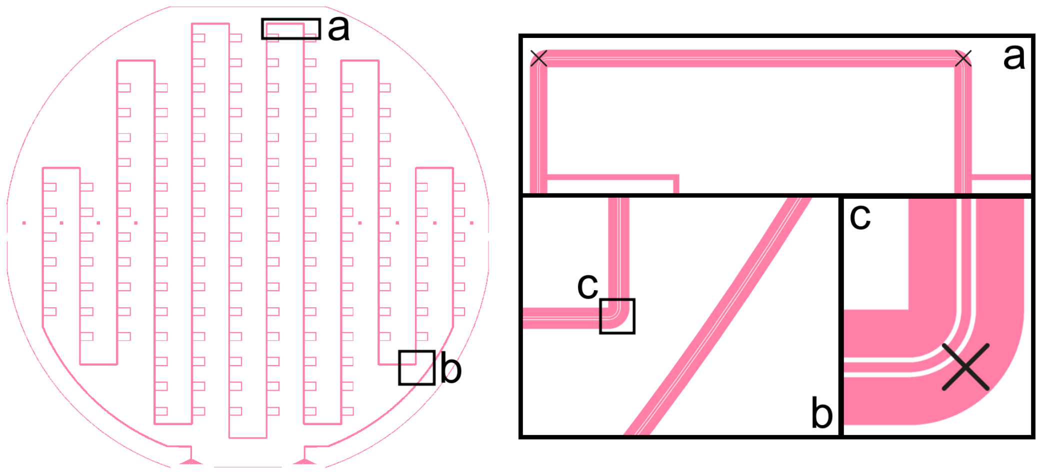


Figure 3.10: Left: The complete view of a 140-KID 3"-detector CPW design with ground shields and mask alignment marks; further description is provided in the following sections. Right: Zoom-in panels for the labeled corresponding locations. a) The meander layout defined by the ×-marked sparse coordinates. b) The smooth large arc defined by a dense array of coordinates automatically generated from a circular function. We manually adjust the horizontal distance of the arc and the c-labeled corner in this figure, so the details of the CPW may be visualized. c) A zoom-in to the minimal-rounding corner generated by the ×-marked coordinate. Note that the mark is always on the central lines of the linked CPW segments, while the relative location of the center of the curved corner varies depending on the angle and dimension of the linked CPWs.

and the two ground traces based on the chosen trace widths. The result is that every trace edge in the rounded corner is tangent to the connected straight CPW trace edges at the minimally required radius, while the (a, b, c) values are maintained throughout the corner. We then realize that, not only does the algorithm allow specifying sparse coordinates for piece-wise straight-line style CPW layouts, e.g., the meandering CPW in Fig. 3.10, when giving an array of coordinates that are dense enough to prevent forming straight CPW sections between corners, each coordinate in the dense array becomes the center of a partially formed rounded corner. Together this series of sector-shaped partial corners then forms an approximated long smooth curve that is tangent to the curve at every given point. This is the technique that we utilize to generate the large curves connecting the bonding pads to the ends of the meandering CPW in Fig. 3.10. With this technique, we are able to design abstract yet fully reconstructable free-formed CPW trajectories from either manually specified coordinate series or points calculated from close-form expressions. Since the algorithm assumes minimal rounding for the calculation/layout generation, which is only possible when the specified points are separated further than the width of the feedline for forming turns, the technique of approximating smooth curves with dense points also allows us to generate a rounded corner at any given radius with a few more points given before and after each sparse corner coordinate.

3.4 KID phonon sensor design

3.4.1 Overview and general architecture

We first reproduce an identical LE KID that is known to work in (Moore et al., 2012; Battistelli et al., 2015) in SONNET to test the coupling of the new CPW feedline and the KID. After confirming the targeted coupling quality factor is achieved with an order-of-magnitude flexibility at reasonable KID-feedline distances and relative orientations, we proceed to replace (Moore et al., 2012)'s original (Doyle, 2008)-style LE KID with the crosstalk-suppressed design by (Noroozian et al., 2012). Since the crosstalk-suppressed "CPS KID-with-ground shield" design is expected to reduce not just the inter-KID microwave coupling but also to the feedline, we readjust the feedline-KID coupling and confirm that it is still within practical reach. After we establish the single-KID architecture, including the KID, the coupled feedline section, and the ground shield, we construct more KIDs in the simulation to study multiplexing properties, such as the level of inter-KID crosstalk and the viable resonant frequency tuning strategies. In particular, we first attempt the inductor-dimension frequency tuning scheme proposed by (Cornell, 2018), motivated by the same rational of maintaining the TLS noise that is found to vary with the capacitor design (Zmuidzinas, 2012; Siegel, 2016). We notice that, in order to maintain the resonant frequencies in the targeted range using specifically the aggressively shrunk IDC by (Cornell, 2018), the small capacitance offered by the IDC requires a substantial enlargement to the inductors. (Cornell, 2018) assumed practical photolithographic limits as the only IDC design limitation, and for further discussion on the ideal resonant frequency range considering TLS noise, see Sec. 4.6 of this thesis or (Moore et al., 2012) for previous DMKID works.

We then realize, even though the enlarged inductors may be spiraled inside a LE-sized area, their uncoiled total CPS lengths are comparable or exceed the resonant wavelengths. The consequence is that these inductors made of long CPSs also exhibit appreciably high current density variations along the spiraling CPS routing, similar to which (Doyle, 2008) pointed out for (Day et al., 2003)'s traditional $\lambda/4$ resonators, which results in non-uniformity of response $\propto J^2$, where J is the current density. We therefore propose the novel symmetric-CPS (sCPS) LE KID design to improve the current density uniformity while simultaneously preserving all the features included so far (Chang et al., 2018). Our KID adopts the TLS noise-robust IDC by (Moore et al., 2012) at a minimal size with a CPS-formed inductor, but instead of adopting simple meander or spiral as in (Doyle, 2008) or (Noroozian et al., 2012), respectively, it uses a novel symmetric double-sided geometry to reduce the effective CPS length. As a result, the sCPS LE design achieves an unprecedented >95% high current density uniformity with a simultaneously mitigated inter-KID crosstalk. It not only alleviates the sensitivity uniformity concern but at the same time permits a straightforward and highly accurate frequency placement (σ<0.07%). The resonant frequencies given by sCPS KIDs follow a simple relation of $f_r \propto 1/\sqrt{l}$, where l is the variable frequency tuning length for the sCPS inductor that is proven to be proportional to the inductance it provides. The applicability of such a frequency tuning relation indicates that our KID design is governed by the ideal LE LC resonator relation $f_r \propto 1/\sqrt{LC}$, which is rarely the case for other SC thin-film LC resonators, e.g., the elongated (Cornell, 2018) design. With these ideal LC resonators, we subsequently realize the inductor-based frequency placement scheme (Cornell, 2018) proposed with a much improved precision for future highly multiplexed detectors.

As shown in Fig. 3.11, we place LE KIDs reproduced from (Moore et al., 2012; Battistelli et al., 2015) to the side of the previously established straight CPW. The KIDs assume a 30 nm thick Al film of material parameters in Tab. 3.1. Following (Cornell, 2018)'s argument, we first orient the KIDs relative to the feedline as shown in the figure, so that the KIDs' elongated direction is in parallel with the CPW for a smaller feedline-KID distance variation. (Cornell, 2018) argued that the choice likely provides a more uniform coupling across the inductor area thus a more predictable/controllable quality factor relation in the feedline-KID distance. We note that, since SONNET does not consider materials' internal quality factors, in the LE regime, the simulated quality factor is dominated by the coupling.[5] Without going into the details that we will provide in Ch. 4, we simply state here that we would like the KIDs to exhibit quality factors that match the anticipated rise time of the phonon energy injection, so the detector's position resolution is optimized. Given the 3" detector geometry, the anticipated multiplexing/phonon absorber coverage, and other assumptions referencing SuperCDMS SNOLAB detectors, the quality factor matching requirement effectively translates to a total quality factor of $(1 \sim 2) \times 10^5$. Further considering the much higher typical internal quality factors for Al KIDs,

[5] For non-LE, the specific geometry of the resonator also contributes.

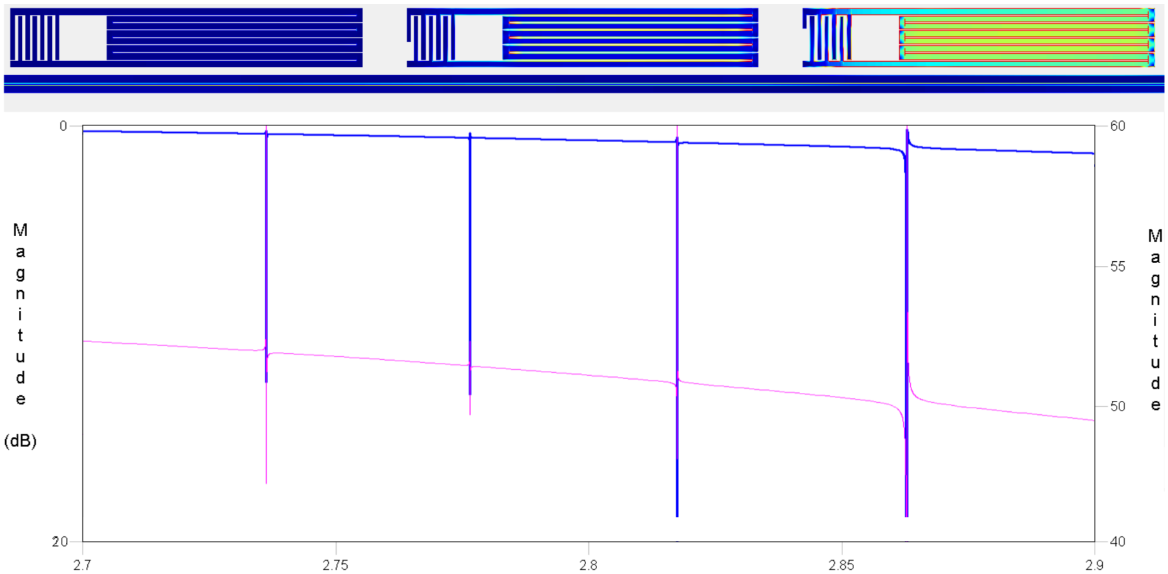


Figure 3.11: Top: The current density distribution for a 4-KID simulation based on (Moore et al., 2012; Battistelli et al., 2015) (one not shown due to page size). To achieve a simulation at this size and complexity, we reduced the finite element meshing to as coarse as allowed by the trace widths. We will explain in later text that, as we later realized, it was not a completely appropriate choice. Nevertheless, (Moore et al., 2012; Battistelli et al., 2015) took such an approach, so we followed for a better reproduction. The frequency placement was achieved via modifying the lengths of the last IDC fingers identical to the references, as may be seen in the figure. The current density is plotted at the resonant frequency of the rightmost KID, which is manifested by its much higher current/RF field density. Bottom: The corresponding S_{21} (blue) and impedance (pink) curves in 2.7–2.9 GHz. In this particular example, which we select for a better visualization, the feedline utilized was still in the process of optimization, slightly different from that introduced previously, and therefore exhibits a steeper frequency dependence as shown by the continuum.

we conclude our KIDs' total quality factors will be dominated by the coupling, which is then tunable by adjusting the feedline-KID distance (Moore, 2012). This coupling-dominated feature enables a convenient design process by the simulation, in particular scanning the feedline-KID distance to obtain the targeted (coupling) quality factor, which will be the desired total quality factor according to the argument.

We determine based on past experience in contact-mask microfabrication that the smallest practical distance between the KIDs and the CPW is 10 μm. For a better reproducibility between the simulation and fabricated devices, we choose 100 μm as a design upper bound for the same distance. Due to the fact that we anticipate populating KIDs at ~mm pitches in future highly multiplexed layouts, we require the KID-feedline distance to be much smaller than between adjacent KIDs, so to achieve an "exclusive/localized" coupling between each KID and its closest CPW section, which is better represented by the small-scale simulation without distant elements. In the 10–100-μm trial CPW-KID distance range, we obtain simulated corresponding quality factors from as strongly coupled as $10^{\lesssim 3}$ to $10^{>6}$, where typical Al-KID internal quality factors start to dominate. We replace the (Moore et al., 2012; Battistelli et al., 2015) KIDs with the CPS-inductor (Noroozian et al., 2012; Cornell, 2018) KIDs and obtain a comparable flexibility in the quality factor tuning. The result demonstrates that the feedline is robustly designed to support any practical quality factor requirements for reasonably designed KIDs in future layouts.

3.4.2 Symmetric-CPS LE KID

We compare the current density distributions of different KID designs as illustrated in Fig. 3.12. It becomes clear to us that all of the KID designs we have considered so far exhibit appreciable levels of current density variation. In particular, Fig. 3.12 clearly shows the most elongated (Cornell, 2018) layout exhibits a current density distribution following the CPS winding that is asymmetric in the two-dimensional area of the inductor. Contrastig with the long (Cornell, 2018) design, the current density distributions for the other designs in Fig. 3.12, if neglecting the edge current concentration for (Battistelli et al., 2015) that is not visualized in coarse simulations,[6] all distribute roughly symmetrically in the overall inductor areas as if the inductors are undivided patches but not formed by narrow winding traces. One would expect an ideal LE to exhibit such a symmetric result rather than that shown in (Cornell, 2018). We note that this observation does not exclude the possibility that the current densities for (Battistelli et al., 2015) and (Moore, 2012) could also be varying primarily along the trace winding while disregarding the inductor orientation, but the difference can not be distinguished on these shorter and more folded layouts.

We are inspired by the similarity of this observation to the typical field modulation in $\lambda/4$ transmission line resonators due to standing wave formation. We interpret the result as, although the entire KID appears to be a LE resonator formed by distinguishable inductor and capacitor, the (Cornell, 2018) design in fact has a nontrivial contribution from a non-LE traveling-wave effect in the CPS-formed inductor. By treating the shorted end of the CPS and the opposite end broken by the IDC as reflection nodes, the inductor behaves similarly to a $\lambda/4$ transmission line resonator. The hypothesis motivated us to design a series of many different LC resonators,

[6] Or integrating the current transversely.

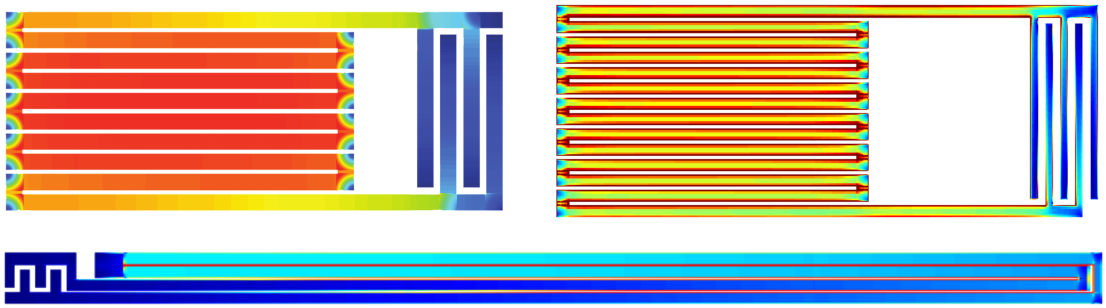


Figure 3.12: The current density distributions of the (Moore, 2012) (top-left), the (Battistelli et al., 2015) (top-right), and the (Cornell, 2018) (bottom) designs; KID geometries reproduced from combining reference text descriptions and example figures. A much larger variety of designs was studied for this part of our work, but we choose to highlight the representative ones shown in this figure to represent the characteristics shared by all the KID designs we studied. Note that, for a clearer visualization, the length scales for different KIDs in this figure are roughly comparable but not precisely the same. The trace widths for the Moore design, the Battistelli design, and the Cornell designs are 70 μm, 80 μm, and 45 μm, respectively. Due to different excitation powers that are subject to feedline-KID couplings, the color scales are only relative and meaningful within the same KIDs but should not be compared between different KIDs.

all utilizing the same IDC but different sizes of CPS inductors coiled in different creative ways. Then, by comparing their resulting current density distributions, we conclude, while different CPS winding schemes unsurprisingly provide different levels of current nonuniformity, once the total length of the CPS becomes longer than a significant fraction of its operating wavelength, pronounced low-current density regions always appear at the capacitor-inductor joints to subsequently increas the nonuniformity from tens of percent to a factor of a few. Given that we primarily aim for a few GHz resonant frequencies using semiconductor substrates, the operating wavelengths fall in 2–4 cm and therefore place a CPS length limit of $\lesssim$mm. The result is qualitatively consistent with our observation that a localized non-LE field indeed forms in the CPS inductor even when the CPS is coiled within an area satisfying the LE condition. Therefore, in order to have the entire KID resonate as a LE, not only do we require the overall 2D dimension of our new KID layout to be much smaller than the wavelength, but if to adopt CPS for crosstalk mitigation, we apply a stronger constraint that the uncoiled CPS must also be much shorter than the wavelength. Our model also explains why the current density distributions in single-trace meanders vary less along the trace winding compared to CPS-formed inductors: It is because the guided field is less confined to the single trace compared to the CPS, therefore the traveling wave effect is also relatively suppressed in the single-trace meanders.

We now reexamine the Cornell (2018)/Noroozian et. al. (2012) design with the new constraint. We first state that, according to the detailed theory in Ch. 4, we anticipate the energy resolution of the detector to degrade with the resonant frequency. Therefore, it is preferred to design for lower resonant frequencies, which in turn implies a larger form factor for the KID. On the other hand, the TLS noise increases at low frequencies by $1/f$ (Gao et al., 2008) and, as observed in (Moore et al., 2012), is negligible at $\gtrsim$3 GHz in the presence of the electronic noise. We explain in the next chapter that we want to avoid TLS noise dominated operation, so the energy resolution benefits from the lower and engineerable white amplifier noise, which is also simpler to model. Combining the two factors, namely lowering the resonant frequency for energy resolution and raising it to avoid TLS noise, we choose $\approx$3 GHz as the design baseline, which corresponds to a guided wavelength of $\approx$40 mm. Following the CPS length constraint discussed previously, we therefore choose to limit the uncoiled CPS in the new KID design to 4 mm or less. Notice that this limit value is consistent with the fact that the (Cornell, 2018) design in Fig. 3.7, which utilizes a 7–10 mm CPS depending on the frequency tuning, always exhibits a strong guided-wave-like current density variation. Owing to the result, when we apply the 4-mm constraint to our design process, we need to abandon the aggressively minimized IDC by (Cornell, 2018) to maintain the resonances at around 3 GHz. This requirement has the unfortunate consequence of potentially increasing the phonon lost in the insensitive capacitor.

Motivated by the CPS length study, we search for a new KID design that utilizes a reasonably sized IDC with a sufficiently large inductor that compensates the IDC phonon loss. At the same time, the inductor should always satisfy the CPS length constraint. After carefully reconsidering all the existing LE KID designs, it occurred to us that, if one uses the (Moore et al., 2012) design as a template and, instead of adopting a spiral-CPS inductor as in (Noroozian et al., 2012), rebuilds the single-trace inductor with CPS, the new KID design will hopefully provide the crosstalk reduction (Noroozian et al., 2012) proposed while preserving other readily usable characteristics in the (Moore et al., 2012) design. The concept is supported by the anticipation that the CPS-formed meander provides a comparable inductance to its traditional single-trace counterpart due to a mirrored waveguide routing, and then the remaining task is to ensure its total length satisfies the 4 mm constraint. The

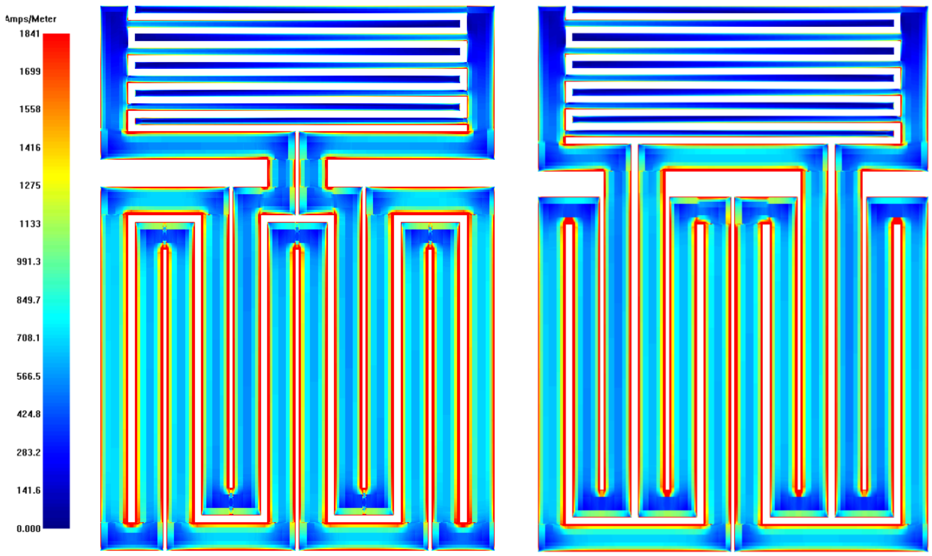


Figure 3.13: The current density distributions for the basic (left) and the doubly-folded (right) sCPS LE KIDs; detailed layout geometries are listed in the corresponding text. The color scale for the current density is attached to provide visual guidance but only a crude numerical reference. In reality as well as in the simulation, the current density highly depends on the readout power, feedline transmission, and the KID-feedline coupling, while the simulated value is also sensitive to the selected finite element mesh. We set the meshing option in SONNET to the finest to appropriately account for the edge current concentration visualized in the figure, which should not be confused as an indication of a highly nonuniform sensitivity distribution. In reality, the sensitivity distribution as suggested by the current density is smeared by the QP diffusion length, which we purposely choose to be the width of the inductor traces. More discussion on the sensitivity distribution may be found in the corresponding text.

unpacked trace length in (Moore et al., 2012)'s inductor is 2–3 times of the limit. Instead of declaring the design concept is unviable, we seek novel layouts that divide the simple meander into multiple parts, each forming an individual inductor that is compatible with a more complex practical CPS connection. Our answer to this inductor division problem is the novel symmetric-CPS (sCPS) LE KID shown in Fig. 3.13.

At first glance, the sCPS LE KID design may appear highly similar to the classical (Moore et al., 2012) layout, with an IDC on top and a meandering pattern for the inductor at the bottom. One apparent difference is that, as we have already elaborated as one of the main design goals, we replace the single-trace inductor with the double-trace CPS to confine stray EM wave for crosstalk suppression. To realize

CPS' double-trace connection, the outer frame of the IDC needs to be split at a shared point as opposed to connecting the two most distant points; the comparison may be seen in Fig. 3.6. However, unlike in the (Noroozian et al., 2012) and the (Cornell, 2018) designs (Fig. 3.6 and Fig. 3.12, respectively), where the IDC frames split on one side for connecting the asymmetric spirals, in the sCPS design we divide the meander into multiple mirrored sections that connect symmetrically to the IDC frame. For each of these divided inductor sections, one CPS trace connects to the IDC frame, and the other connects to one of the CPS traces in the mirrored inductor section. By bringing the middle point(s) of the CPS meander to be close to the IDC frame, we effectively create another "node" in addition to the shorted and the IDC-broken waveguide ends. As shown by the symmetric current density distribution in Fig. 3.13, each inductor section forms its own in-CPS traveling wave from the shorted end to where the inductor sections meet. The design therefore achieves the sought division for the long traditional meander/spiral inductors. In terms of the practical dimensions for the basic-version sCPS LE KID shown on the left of Fig. 3.13, which we utilize in all DMKID devices since its creation, the meander trace width is 80 μm; the trace gaps inside and outside of the CPS are 20 μm and 10 μm, respectively;[7] the IDC frame width is 80 μm; and the IDC finger width and the interdigitating gap are both 20 μm.

Our IDC design is largely copied from (Moore, 2012), where we choose identical finger and interdigitating gap widths to maintain the same field strength between the fingers. We hope to obtain a similar TLS noise-free performance as (Moore et al., 2012) in doing so, based on the widely accepted current theory that the TLS noise in KIDs are likely due to the excitation of substrate TLS with the EM field in the capacitors. Since the capacitance of an IDC is predominately contributed by the dielectric filling in between of the IDC fingers but not the SC materials, we originally attempted to reduce the width of the fingers so to minimize the capacitor area. However, we quickly noticed such a modification concentrates the current in the traces. According to (Gao et al., 2008)[8], this elevated field in the capacitor senses more strongly the tunneling of the dielectric constants of the Si and the SiO_2 on the substrate surface, which leads to an elevated TLS noise, thus strongly disfavoring the narrowing of the IDC traces. We nevertheless further widen (Moore et al., 2012)'s IDC frame to be as wide as our CPS trace for 1) an impedance-matched smooth

[7] The readers should be able to distinguish the 20- and 10-μm gaps in the figure if it is unclear which gaps are inside or outside of the CPS.

[8] c.f. Ch. 4 for a more detailed discussion.

connection to the new inductor design and 2) a reduced current density in the IDC frame that is not elevated relative to the fingers.

We then take an opposite approach of enlarging the inductor area to compensate for the phonon loss in the capacitor. We first identify the QP diffusion length as the length scale for the oscillating current in the inductor to respond to localized QP creation events. In other words, we anticipate the sensitivity to be subject to the current density fluctuation, if the QP creation events can take place further than the QP diffusion length, and vice versa. Therefore, in order to suppress such a length scale dependence that is not indicated by the simulated current uniformity, we limit the inductor trace width by the QP diffusion length. We choose 80 μm, a value appreciably reduced from the typical QP diffusion lengths found for thin-film Al, as the width of our CPS traces. This design approach ensures that the QPs can completely diffuse transversely in the inductor traces and thus suppresses the KID's sensitivity to the chosen inductor trace width. We then choose 20 μm and 10 μm for the gaps inside and outside of the CPS, respectively, based on the smallest dimensions that we can reliably fabricate. We choose to minimize these dimensions so to produce the smallest KID that better behaves as a LE. We nominally take 10 μm as the reliable trace-gap limit for our photolithography due to our $\lesssim \mu$m feature resolution, and we double the limit for the inside gap of the CPS to prevent it from being shorted by defects. Based on past experience, when the inductor traces are shorted by defects on the CPS' outside edges, the impact is at most a degradation to the resonator's quality factor if not negligible. On the contrary, if the inductor waveguide is shorted inside, the resonant frequency can be shifted significantly due to the obvious change of the effective meander length.

3.4.3 Simulation technique

Before presenting the characterization for our new design, we first introduce the simulation techniques that distinguish our design process and the following characterization from previous works. Being a widely adopted simulation tool for most KID development works, SONNET claims it is a fully 3D simulation operating on a predominately 2D user interface,[9] and it simulates the surface- and edge (film sidewall)-concentrated current in conductive materials with a constantly improving accuracy. Since 2015, SONNET further claims to have improved its predictability specifically for high-quality SC micro-resonators. While it remains to be clarified as whether the improved surface current modeling for SC materials now considers

[9] It is possible to visualize the model in a 3D view, but the functionalities are much more limited.

the $1/\lambda$ EM field attenuation according to the Mattis-Bardeen theory but not the classical $1/\lambda^2$, λ being the effective EM penetration depth, it is a fact that one can not explicitly designate a material in SONNET to be a SC aside from setting a zero resistivity. We therefore question whether SONNET really distinguishes a SC from a perfect classical conductor and simulates the currently density distribution accordingly. Furthermore, as seen in all the KID design works we have cited so far, they all chose to mesh their simulations to be as coarse as their metal traces, e.g., see the coloring of the (Moore et al., 2012) design in Fig. 3.12. In other words, all the previous simulation works we have mentioned used single lines of 10s-μm wide finite-element cells without the resolution for the edge-concentrated current. As a comparison, when we enable the finest meshing option in SONNET, we always obtain more than 50% of the total current concentrated within a few μm from the edges in our 80-μm SC Al traces, such as in Fig. 3.13.

Although due to lack of observational confirmation, we can not comment on whether the current uniformity provided by the coarsely meshed simulations is correct, we do find, by adjusting the meshing option from the finest to the coarsest, the simulated resonant frequencies can vary significantly with respect to the narrowness of the SC resonators at high quality factors. We therefore decide to perform all the following simulations at the “finest” meshing setting in SONNET and present the results as such. Under this setting, SONNET meshes the models to be much finer than their local current concentration length scales in a frequency-dependent fashion, which ensures the current concentration effect is included. When we compare these finely meshed results to previous works on coarse meshes, we then average each local current destiny value with its surrounding 80×80 μm^2 pixels within the conductive region, i.e., excluding the outside of the SC film if the pixel is close (<80 μm) to the edge. We believe our averaging approach more appropriately represents the distribution of the sensitivity to QP density by considering simultaneously the sensing current density and the intrinsic smearing due to QP diffusion. It is worth noting that, when testing this methodology on the (Moore, 2012) design, we obtain a significantly lower “sensitivity uniformity”[10] value than the simple current density uniformity reported in (Moore, 2012). The mesh therein was set to the coarsest possible but coincidentally also 80 μm, because it was limited by the KID design’s 80 μm trace width. The difference of (Moore, 2012) and our results shows that, in terms of inferring the QP sensitivity uniformity with SONNET simulation, the simple current density uniformity is likely not a good metric even if the simulation

[10] i.e., the above current density uniformity with QP diffusion averaging.

is meshed to the QP diffusion length.

The second technical remark, which is also a tip for future DMKID group members, is for a quick resonant frequency identification for high-quality resonances. As pointed out in (Wisbey et al., 2014), simulation tools like SONNET tend to assume that the frequency-domain structure of interest occupies a good fraction of the designated bandwidth of the simulation. The algorithms would sample the frequency space accordingly to efficiently profile such a structure. If any structure is identified by the initial sparse sampling, the algorithms continue refining the simulation around the frequencies until the result is deemed appropriately determined. However, due to the narrowness of KID resonances relative to the frequency range they are placed, it is common for the simulation to miss the resonances and conclude the spectrum contains nothing but the feedline transmission structure. Before (Wisbey et al., 2014), one technique to circumvent the issue was by first placing the resonators exaggeratedly close to the feedline, so the total quality factors are heavily diluted by the strong coupling. One is then able to identify the broad resonances even though the precise resonant frequencies are likely also biased. And then by incrementally moving each resonator away from the feedline, one may hopefully track the resonance with the increasing quality factor while narrowing the simulated frequency window. The method is obviously very time- and labor-consuming, especially when the computers were not as powerful.

The ingenious technique (Wisbey et al., 2014) proposed, dubbed the "three port method" in Caltech ObsCos group, solves the problem by inserting a wave port in addition to the two existing feedline excitation ports. For traditional (Doyle, 2008)-style LE KID consisted of an IDC and a single-trace meander, (Wisbey et al., 2014) suggests placing the third wave port on one of the linking traces between the capacitor and the inductor. For this third port, the equivalent circuit model is a series R-L-C resonator, whose resistance represents the loss of the SC and thus is close to zero, the inductance and the capacitance are that designed for the KID resonator, and the two feedline excitation ports are removed from the model by matching loads. (Wisbey et al., 2014) shows that the input impedance formula for the third port does not contain a singularity as for the feedline excitation ports and thus is slow-varying in frequency. It is then easier and more computation-efficient for the simulation to profile a wideband spectrum for the circuit and accurately determine the frequency that the imaginary part of the third-port input impedance crosses zero, which is the resonant frequency. One can also extract the quality factor for the pre-modified resonator from the slope of the simulated third-port input impedance (Wisbey et al., 2014). While we sometimes notice that the parameters extracted from the third-port simulation can disagree noticeably with that from a direct 2-port simulation, the inferred quality factor and the resonant frequency by the (Wisbey et al., 2014) method are precise enough to guarantee a frequency window that is mostly occupied by the resonance of the original model. With the help of these calculated values, the iterative narrowing process becomes much more efficient for one to simulate the pre-modified layout and directly determine the resonance frequency and the quality factor. However, it is unclear whether (Wisbey et al., 2014)'s 3-port circuit model and therefore the formulae apply to our sCPS LE KID. Following the spirit of (Wisbey et al., 2014) using layout similarities, we test inserting the third port at one of the IDC-inductor connections, the connection of the two symmetric CPS meanders, and the shorted end of one of the CPS meanders. We find the last option yields a similar frequency-domain response to a traditional LE KID with the third port. We did not determine if the circuit model by (Wisbey et al., 2014) strictly apply to our 3-port modification, but our modification does yield comparable peak position and width for the resulting broadband structure as in (Wisbey et al., 2014). The information is sufficiently accurate to be substituted into (Wisbey et al., 2014)'s formulae to predict sCPS KID's resonant frequency and quality factor to order-of-magnitude, so one can narrow the simulation window for the two-port structure.

3.4.4 DMKID sCPS KID design

Fig. 3.14 presents the QP diffusion-averaged current density distribution. The particular example shown is one of our sCPS LE KIDs at 3.353 GHz. Unlike in (Moore, 2012), where the single-trace meander exhibits a 2D-symmetric current density distribution that disregards trace discontinuities and gaps (Fig. 3.12), the distribution in Fig. 3.14 clearly responds to edges and corners and continuously modulates along the CPS winding as expected by our traveling wave interpretation.[11] We define the relative current density nonuniformity by

$$\frac{|\sigma_{\mathrm{J}}|}{\overline{\mathrm{J}}}, \tag{3.8}$$

where $\overline{\mathrm{J}}$ and σ_{J} are the mean and the standard deviation for the simulated current density after the QP diffusion averaging in the inductor area, respectively. Notice that we alternately use "uniformity" and "non-uniformity" in the following, where

[11] Due to the much improved uniformity, the readers might need to significantly expand Fig. 3.14 to distinguish the color variation.

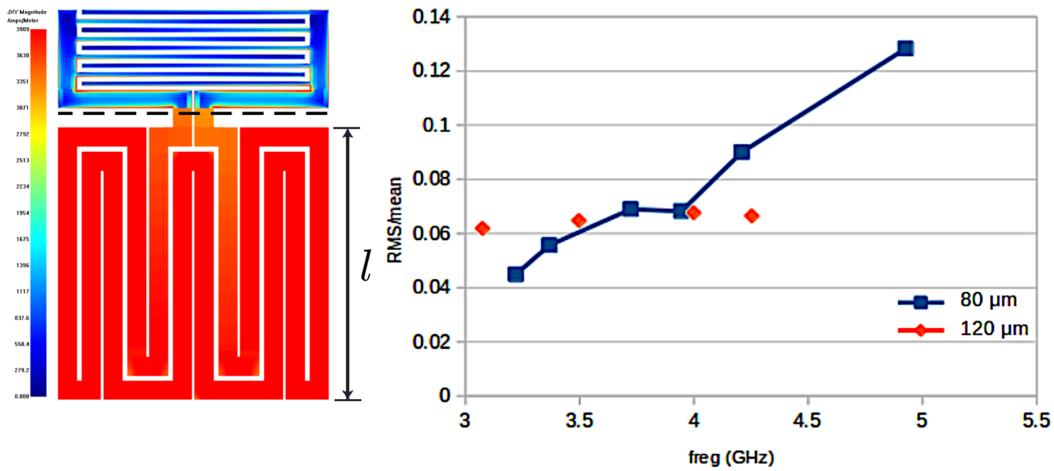


Figure 3.14: Left: The sensitivity/current density distribution for a 3.353-GHz sCPS LE KID, obtained by the QP diffusion-averaged technique discussed previously. For comparison, we leave the IDC area above the dashed line unprocessed by the averaging technique. Right: The simulated sensitivity/current density nonuniformity per Eq. (3.8) for different KIDs, plotted against their resonant frequencies. In addition to the nominal 80 μm-trace design represented by the blue squares with a solid curve, we attach in the figure another series of data represented by the curve-less orange diamonds for results utilizing 120-μm CPS traces.

the uniformity simply means the difference of unity and the nonuniformity defined above, i.e.,

$$1 - \frac{|\sigma_{\mathrm{J}}|}{\overline{\mathrm{J}}}. \tag{3.9}$$

Fig. 3.14 summarizes the nonuniformity values for different resonator sizes/frequencies that we simulate. Although strictly speaking the traveling wave-like distribution manifests the non-LE quality of the design, the result shows the sCPS LE KID generally exhibits sensitivity uniformities greater than 93% below 4 GHz, the primary range we target for our DM detectors. This uniformity is to our knowledge the highest ever achieved if directly compared to previously reported current density uniformities obtained with coarse meshing, which are likely optimistic according to the analysis detailed previously. Despite the observable percent-level variation, the obtained high uniformity indicates that the sCPS LE KID is by far the design that most resembles a perfect LE.

We adjust the dimension of the inductor, specifically the length l marked in Fig. 3.14, to achieve the multiplexing frequency placement following (Cornell, 2018)'s proposal. As shown in Fig. 3.15, we find the resonant frequency f_r and the dimension of the inductor l follow

$$f_r \propto \frac{1}{\sqrt{l}}. \tag{3.10}$$

Since we expect the resonant frequency to follow

$$f_r \propto \frac{1}{\sqrt{LC}}, \tag{3.11}$$

and we control l while keeping the capacitance fixed, Eq. (3.10) suggests

$$L \propto l \tag{3.12}$$

in the range of l (or equivalently f_r) that we test. We find the improved sCPS LE KID design is governed by the ideal LC resonator relation as in Eq. (3.11), contrasting with the variable $f_r \propto 1/l^{0.5-1}$ frequency placement relation (Cornell, 2018) obtained, which we have argued is likely due to the long meander's traveling wave behavior. The result supports our design concept that, by carefully improving the LE quality of the KID design, especially the current uniformity in the inductor, we recover the ideal LC resonator property for the KID. Later we will also present in more detail that, as it turns out, not only does the sCPS LE KID exhibits a well-modeled LE response in the simulation, but more importantly, the practical frequency placement for a fabricated RF assessment device also achieves a much greater accuracy than previous works.

3.5 Detector design

3.5.1 KID-feedline and inter-KID coupling

We combine our CPW feedline design and the basic-version sCPS LE KID to construct the full detector layout. We choose to orient the KIDs relative to the feedline symmetrically, i.e., the KIDs' symmetric axes are perpendicular to the CPW's, contrasting with (Cornell, 2018)'s asymmetric orientation. The symmetric and the asymmetric orientations are shown by the right- and the left-hand side panels in Fig. 3.16, respectively. (Cornell, 2018) chose the asymmetric orientation to acquire a manageable feedline-KID coupling variation for the overly elongated KID design (Fig. 3.7), which is now alleviated by our more square KID layout. We plan to distribute the resonances in 3–5 GHz, which corresponds to

$$l = 1400\text{–}500\ \mu\text{m} \qquad (\text{for } f_r = 3\text{–}5\ \text{GHz}) \tag{3.13}$$

according to the fitting result in Fig. 3.15.

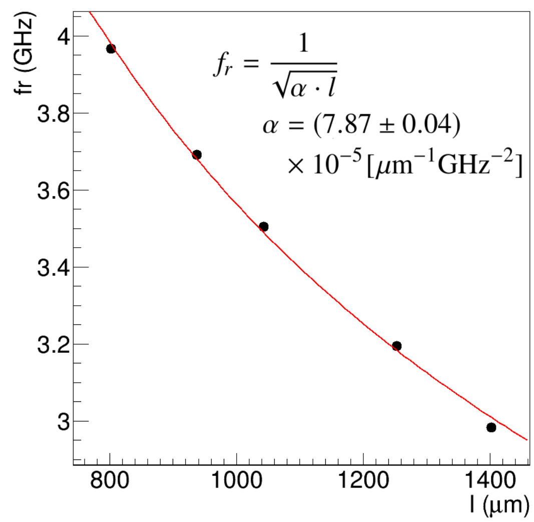


Figure 3.15: The simulated resonant frequencies for sCPS LE KIDs with different inductor sizes (l) in the 3–4 GHz range. The data are fitted by the curve shown in the top-right corner. For future applications, notice that α is not unitless and therefore depends on the unit of the input l. In principle, α also depends on the physical properties of the device, such as the material parameters, KID design, fabrication quality (for real devices), etc.

Fig. 3.16 illustrates the coupling field comparison. Although we place the variable dimension l transverse to the feedline, which leads to a variable distance from the feedline to the center of the KID, the practical value and the range of variation in Eq. (3.13) are both negligible compared to the much larger resonant wavelengths. Therefore, we expect a consistent CPW-KID coupling insensitive to l that is predominately determined by the CPW-KID distance. The argument motivates choosing the symmetric orientation over the asymmetric orientation for a more balanced forward/backward-wave coupling, despite the possible small coupling strength variation due to l difference. We suspect the asymmetric coupling is responsible for the unexplained large coupling phase and want to avoid the associated quality factor/readout power complexities proposed by (Khalil et al., 2012). One can also argue that, intuitively, the asymmetry of the coupling is likely to depend on l when placing l parallel to the feedline, and we expect the coupling between the feedline and the KID to be sensitive to the asymmetry, therefore the asymmetric orientation is intrinsically more sensitive to l even though the CPW-KID distance can be fixed.

Fig. 3.17 shows our adaptation of the ground shield concept by (Noroozian et al., 2012) following the CALDER detector design (Battistelli et al., 2015), which reduces the crosstalk coupling between neighboring KIDs by attracting stray field

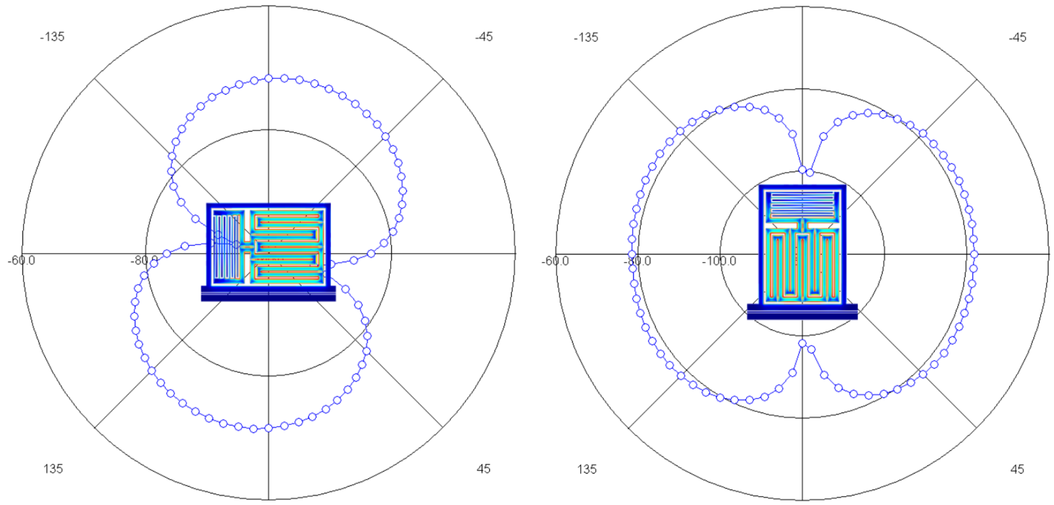


Figure 3.16: The simulated field strengths for the complete "KID-feedline-ground shield" constructions shown at the corresponding centers of the charts. We choose a physical radius in the transitional near-to-far-field regime for the field calculation, and the values are shown in dB for a relative comparison. We insert the layouts following the radar chart orientation for a field directionality reference, and the layouts are colored by their on-resonance current densities.

to the box-shaped extension of the CPW ground trace. Since crosstalk leads to resonant frequency and quality factor modifications, the ground shield design is expected to maintain our frequency placement accuracy. Currently, we choose 80 μm for the width of the ground shield, identical to the CPS meander. Although we plan to fabricate the ground shield of Nb to suppress its phonon absorption relative to the Al KID (Ch. 4), we still prefer it to be not significantly larger than the KID and potentially interfere with the phonon propagation in the detector. Nor do we want the ground shield to be small compared to the KID and provide an insufficient crosstalk suppression, so we choose the same trace width as the KID and also for aesthetic reasons. With the integrated feedline, KID, and the ground shield layout, we scan the distance d between the feedline and the KID in the simulation (Fig. 3.17) to determine the dependence of the resonant quality factor versus the distance. We subsequently select

$$d = 30\ \mu\text{m}, \tag{3.14}$$

which corresponds to

$$Q \approx Q_c \approx 1 \times 10^5, \tag{3.15}$$

where Q_c is the coupling quality factor that we expect to dominate the total quality factor Q. We explain our choice for the specific quality factor in the next chapter.

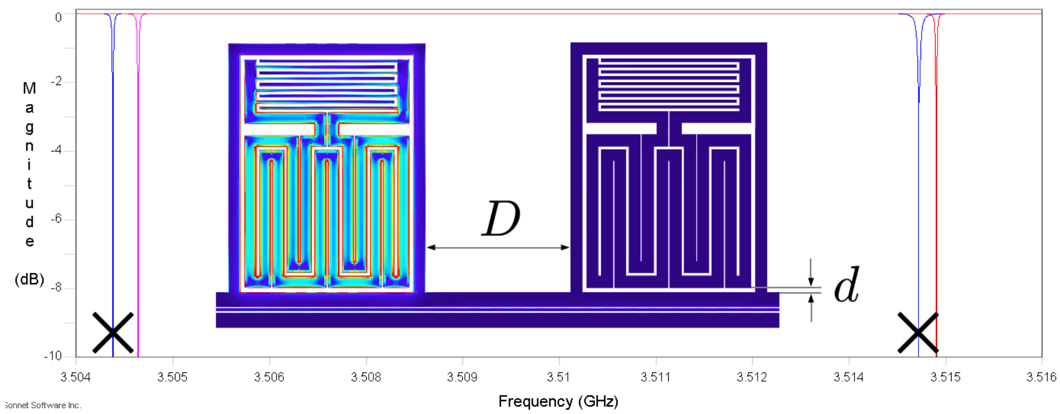


Figure 3.17: A comparison of the S_{21} without (red/pink) and with (blue, $\times$ mark) crosstalk bias. We obtain the nominal transmission curves (red/pink) from the single-KID simulations and the crosstalk (blue) curve by placing the two KIDs at $D = 50\ \mu$m as shown by the inset. d is the variable distance for adjusting the coupling quality factor; more discussion is found in the corresponding text. The coloring of the inset represents the current density distribution at the resonant frequency of the left-hand-side KID, which shows a crosstalk-free condition when $D = 1$ mm and $d = 30\ \mu$m.

After determining d, we place two KIDs that are nominally 10 MHz apart in frequency space side by side as shown in Fig. 3.17. We probe their microwave crosstalk as a function of D, utilizing their resonant frequency offsets according to (Noroozian et al., 2012). Assuming capacitive coupling, in the weakly coupled regime ($C_{12} \ll C$, see Eq. (3.16)), the frequency offset δf for a resonator under crosstalk is given by

$$\delta f \approx \frac{C_{12}}{C_1 + C_2} \times \bar{f}, \tag{3.16}$$

where C_1 and C_2 are the designed capacitances for the resonators, which should be the same in our design, C_{12} is the coupling capacitance, and $\bar{f}$ is the mean of the crosstalk-free resonant frequencies. For the rather extreme $D = 50\ \mu$m example in Fig. 3.17, we obtain

$$\frac{C_{12}}{C} < 1.7 \times 10^{-4}, \tag{3.17}$$

where we take $C = C_1 = C_2$ based on our design. However, we notice the simulation does not yield equal-magnitude and opposite-sign deviations for the resonances, which is expected by the weak coupling assumption. Instead, the mean of the resonant frequencies shifts significantly, which is apparent in Fig. 3.17 especially for the narrow resonances. The result suggests that the added capacitance C_{12} is likely not as small relative to the designed capacitances for the resonators and therefore should not be omitted in $\bar{f}$. We incrementally increase D and find the resonances are

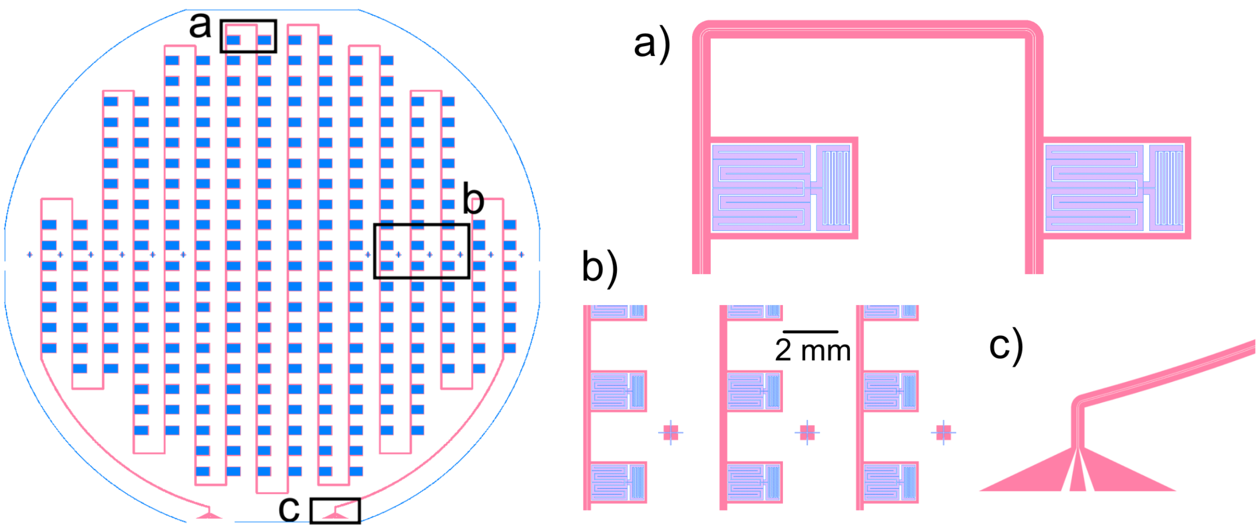


Figure 3.18: Left: The complete view of the 260-KID 3"-detector design. Right: The corresponding detailed features as labeled in the left-hand side schematic. a) The CPW feedline with ground shields (pink filled) and the sCPS LE KIDs (purple hatched). b) A zoom-in view to show the detector multiplexing, sensor density, and photo-mask alignment marks. c) The tapered wirebonding pad and its connection to the CPW feedline.

offset comparably to that in Fig. 3.17 but without an apparent trend against D. After $D \gtrsim 500$ μm, the randomly displaced resonances begin to approach their nominal frequencies monotonically and become indistinguishable from the crosstalk-free condition at $D \approx 1$ mm. Despite the result apparently requires a more sophisticated model, we conclude that the expected fractional frequency displacement due to crosstalk is at most in the 10^{-4} range and can be fully mitigated. The highest displacement corresponds to placing adjacent KIDs with ground shields almost in contact, while the crosstalk can be fully mitigated at a ≈mm separation.

3.5.2 Feedline routing

We proceed to the full design for 3" circular wafers. In order to take advantage of the existing infrastructures from (Moore, 2012; Cornell, 2018), such as detector housing, readout cabling, and cryogenic mechanics, we set the design rule that the input and the output terminals for the CPW feedline, at least for near-term designs, should be located 18 mm apart from each other along the wafer flat, as shown by the (Cornell, 2018) design in Fig. 3.7 or our design in Fig. 3.18. We then restrict the use of the 500 μm-wide peripheral region around the standard 3" single-flat (100) wafer edge to avoid practical fabrication issues, such as edge rounding issues or edge bead. More detail about our edge treatment is provided in Sec. 4.1. Under the two general constraints, we design the tapered wire-bonding pads at the terminals of the feedline as shown in Fig. 3.18. The design adopts the commonly utilized taper technique that preserves the width ratio of the CPW, namely our (a, b, c) parameters in Eq. (3.7), hence maintaining the designated impedance based on Eq. (3.5). We widen the nominal CPW geometry by 15 times in a feedline length of 800 μm, i.e., the "height" of the triangular shape, and we place the widened bottom edges of the tapers at 500 μm from the border of the wafer as stated.

Our feedline design utilizes the meandering routing scheme shown in Fig. 3.10/3.18 to evenly cover the 3"-diameter detector surface (minus 500 μm border) at pitches determined based on the multiplexing requirement; details about the pitch calculation is provided in the next section. We link the meander and the bonding pads by two large arcs that are concentric to the wafer. Instead of allowing the arcs to run arbitrarily close to the bonding pads, we use a 2-mm straight vertical section to introduce each bonding pad taper to the CPW in order to reduce the potential impedance mismatch by an aggressive geometry change. The arcs then connect to these vertical sections via the horizontal sections at 90° corners. As introduced in Sec. 3.5, we have developed an efficient automated CPW layout program that generates idealized feedline designs based on specified requirements, so instead of hard-coding the geometries for the routing plan, we simply provide the above descriptions and constraints for the program to generate the layout automatically. We also include all necessary components into the tool library, including the tapered wire-bonding pad, wafer boundary and mask alignment marks, and most importantly, sCPS LE KID with ground shield that are automatically generated by the designated (l, d) inputs. We place the above components using coordinate and orientation inputs similar to the feedline generation.

3.5.3 Phonon sensor population

We have not yet specified the most critical parameter: The total number of KIDs each detector layout contains. In order to allow the feedline to terminate on the same side of the wafer, we are limited to choosing an even number of parallel lines at a specific pitch for the CPW meander, which subsequently limits the reasonable number of KIDs to be populated along the feedline.[12] Therefore, instead of arbitrarily specifying the number of KIDs, we determine it is more practical that we instead specify the number of parallel meander lines with a set of KID-populating criteria

[12] It is easier to reference Fig. 3.18 while reading the following content.

and allow the "algorithm" to generate the layout. We apply the KID-populating rules:

1. For the specified even number of CPW meander lines, the lines are evenly separated within the non-peripheral region. The space distribution includes reserving the left- and right-hand side distances for the outermost meander lines.
2. In each column of KIDs that attaches to the side of the CPW, the vertical separation of KIDs is equal or larger than the horizontal separation between two parallel CPW lines. The KIDs are equally spaced in all KID columns.
3. The top and bottom KIDs in each column are located by more than half of the nominal KID-to-KID distance from their closest CPW corners.
4. We allow the algorithm to minimally adjust the vertical positions of KIDs, so all the KIDs in the layout align to a global Cartesian grid, not just vertically along the feedline but also horizontally to the KIDs in other parallel columns.
5. We arrange the resonant frequencies of KIDs to increase at the same increment in frequency along the feedline routing. Due to the feedline meandering, this choice results in opposite frequency ordering (resonant frequencies increase toward the top or bottom of the wafer) in adjacent KID columns.

We examine 3. in SONNET and confirm that this distance is sufficient to prevent the top and the bottom KIDs from having a higher coupling to the CPW than the middle KIDs. We designate 5. as a simpler baseline design, which may be straightforwardly modified for advanced future layouts with our KID/frequency placement program. We find, simply by varying the number of meander lines, the above design rule produces any target number of KIDs at a %-level discrepancy, as long as the targeted number is higher than 30. Considering that we mainly aim for a high multiplexibility for this work, the 30-KID limit is sufficiently small and does not pose a practical constraint to our goals.

For the detectors we present later in this thesis, we adjust the inductor dimension l based on Fig. 3.15 to achieve the selected resonant frequencies:

$$f_r = 3.05 + \frac{3.45 - 3.05}{N_{\mathrm{KID}} + 1} \times i_{\mathrm{KID}} \quad [\mathrm{GHz}], \quad i_{\mathrm{KID}} = 1 \sim N_{\mathrm{KID}}, \tag{3.18}$$

where N_{KID} and i_{KID} are the total number and the index of KIDs, respectively. Eq. (3.18) places the resonances in 3.05–3.45 GHz, evenly spaced for the designated number of KIDs. The arrangement also separates the first and the last resonances from the edges of our readout band, which we choose based on the near-3 GHz frequency placement decision explained previously assuming the bandwidth given by the readily available ROACH DAQ system described in (Siegel, 2016). Around the same time we designed this multiplexing scheme, we also acquired the GPU-based SDR DAQ system described in (Minutolo et al., 2019), which is more flexible than the ROACH system and allows future DMKID devices to revise Eq. (3.18) based on the installed SDR components. It is worth noting that our current simple increasing arrangement of f_r along the feedline is not the most sophisticated option ever demonstrated. For example, in order to suppress the crosstalk to improve the detector's imaging/positioning ability, (Siegel, 2016) utilized a non-linear mapping between the resonant frequencies and the physical locations of the KIDs, which distances the KIDs closer in the frequency space further on the physical device. We do not immediately adopt such an improvement for the devices introduced in this thesis, because we prefer the early measurements to be easier to understand and be compared to our model for R&D purposes. As introduced later, we find the crosstalk mitigation due to our detector design allows the future DMKID detectors to continue adopting the simple linear f_r arrangement without significant crosstalk confusion.

3.5.4 Design realization

Based on the design procedure detailed above, we generate three different detector layouts: 80 KIDs, 140 KIDs, and 260 KIDs; the designs are shown in Fig. 3.19. We choose the 80-KID design for its roughly 5 mm separation between KIDs, a similar KID density to (Moore et al., 2012)'s original design. The purpose of this layout is to reproduce (Moore et al., 2012) and to study the position and energy resolution differences, which are presumably dominated by the substrate geometry change. On the other extreme, we choose 260 KIDs based on approaching the aforementioned 1-mm crosstalk-free KID distance limit. We then taken the geometric mean of 80 and 260, slightly rounded to 140, as an intermediate step for an up-scaling verification as well as data processing and other R&D practices. In fact, based on our phonon-mediated detection principle, we suspect that the dramatically increased multiplexibility will quickly approach an optimal position resolution permitted by the KID technology and then saturate. Since we do not expect the detector performance, especially the position reconstruction reliability, to improve unlimitedly with the multiplexing factor, the 140-KID detector will

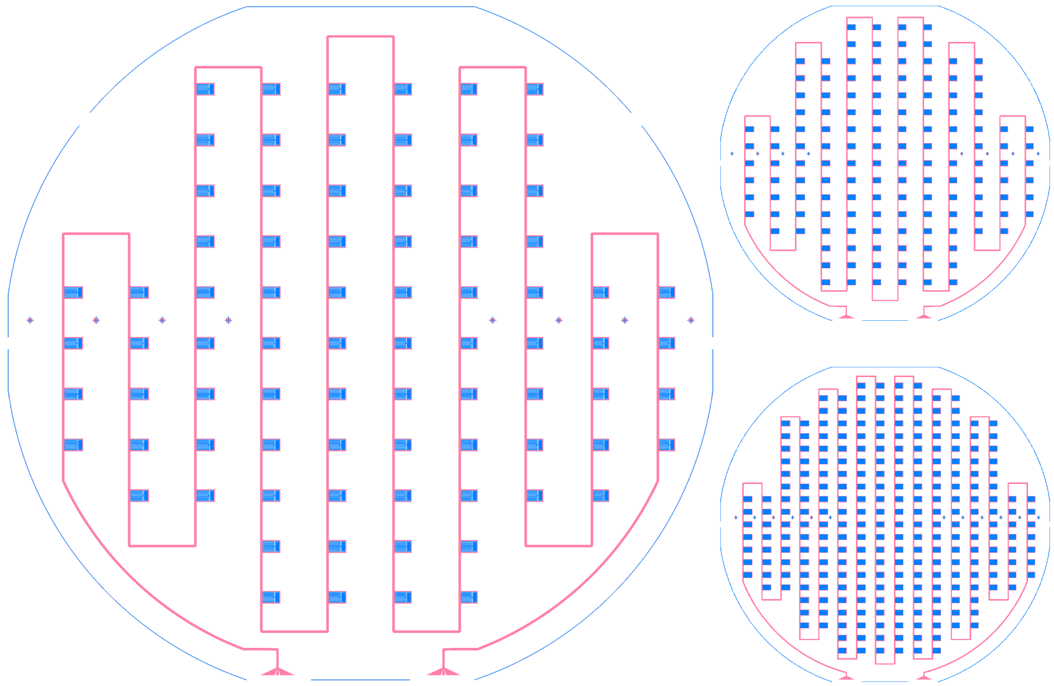


Figure 3.19: The complete views of the 80-KID (left), 140-KID (top-right), and 260-KID (bottom-right) 3" detector designs. The Al and Nb features are in pink and blue, respectively. For detailed features in the schematics, we refer the readers to Fig. 3.10, Fig. 3.13 , Fig. 3.17, and Fig. 3.18 for the details of the CPW, the sCPS KID, the feedline-KID integration, and the general detector design, respectively.

help with constraining the detector performance's scaling in multiplexibility and thus the optimization for future detectors. Based on preliminary signal and noise measurements with certain model assumptions, we provide an analytical model in Ch. 4 for the detector's expected phonon energy resolution $\sigma_{E_{\text{ph}}}$, which indicates

$$\sigma_{E_{\text{ph}}} \propto \sqrt{N_{\text{KID}}}. \tag{3.19}$$

We have fabricated the photo-masks of the 80-, 140-, and 260-KID layouts for future works to continue the N_{KID} dependence study.

Finally, we discuss the material considerations for the detector realization. Previously, we briefly mentioned that we plan to fabricate the KIDs of Al, which exhibits a high kinetic inductance, while other phonon-insensitive features are fabricated of Nb, including the feedline and the attached ground shields. It is an evolution from the material composition strategy of former DMKID projects' single-layer Al fabrications. The readers may find an in-depth discussion in (Moore, 2012) for the motivation of utilizing Al for the KIDs. In summary, Al is an efficient Si phonon absorber that was first proposed by (Irwin, 1995) and has been adopted by Super-

CDMS detectors until now (Pyle, 2012). It is also one of the SC materials with high kinetic inductances that are relatively easy to manipulate in practical thin-film device fabrication. However, being an efficient phonon absorber, the existence of the large-area Al feedlines in (Moore et al., 2012; Cornell, 2018) also implies significant phonon losses to the feedlines. Given that the detector designs (geometries, phonon absorption times, etc.) in these previous works likely led to nearly homogenized phonon distribution and long timescales, we expect these all-Al designs to suppress the signal size roughly proportionally to the feedline-to-KID coverage ratio. One can easily see that it is a significant suppression just by comparing the pink and the blue areas in Fig. 3.19. Therefore, we propose replacing the feedline material with a SC material that exhibits a larger bandgap, ideally much larger than the ≈meV Si phonon energy, so we can prevent the phonon loss in the feedline. According to Bardeen-Cooper-Schrieffer (BCS) theory, well-modeled[13] SC materials should exhibit bandgap energies that are proportional to their T_c, which are more widely and accurately determined (Ashcroft and Mermin, 2021). We then identify Nb, which has a T_c about 8 times of Al, as an ideal high bandgap option for fabricating the "dead materials" in our detectors. In addition to the high bandgap, there are also practical benefits of adopting Nb. First, it is one of the highest-T_c materials immediately available in our JPL fabrication facility. Second, among these available high-T_c options, Nb is elemental and therefore simpler to incorporate into our fabrication. Lastly, due to the above reasons, our colleagues at JPL MDL have accumulated valuable experiences specifically in fabricating Al/Nb bi-material devices with the facility, which has greatly helped with the development of the fabrication process.

For a final remark to close the detector material discussion, even though Nb typically exhibits a pair-breaking energy as high as 2.7–3.0 meV (Ashcroft and Mermin, 2021; Gao, 2008), making it suitable for preventing the phonon loss, our recent work has tentatively shown it might not be as effective as we previously expected. E. Lindeman has constructed a simulation based on the (Geant4, n.d.) package with customized phonon physics based on (S. B. Kaplan et al., 1976). We notice that, due to the wide phonon energy range, the simulation yields a possible phonon absorption that is not as minimal as we had hoped based on the difference of the ≈1 meV peak phonon energy and the ≈3 meV Nb pair-breaking energy. We are currently investigating the result, which we suspect involves the early-phase physics of the phonons before they down-convert to lower energies. It is also worth mentioning that, recently, O. Wen further adapts the Nb-feedline concept to a new fabrication effort, where he also

[13] There are many SC materials that cannot be modeled by the BCS theory, but Nb is not included.

changes the material of the capacitors in KIDs, leaving only the phonon-sensitive inductors pure Al. However, instead of simply fabricating the capacitors from the same single-layer Nb with the feedline, Wen fabricates the all-Al KIDs identical to the devices of this thesis. He then covers the Al capacitors with Nb to proximitize the Al with the Nb bandgap. The improvement due to the proximitization technique is being evaluated in parallel with the ongoing characterization of ours and Wen's devices. These continuing efforts will provide valuable knowledge for our future R&D, and we look forward for detailed discussions in O. Wen's PhD dissertation that is to be completed in 2024.

3.6 Nb RF assessment device

3.6.1 Feedline test device

In order to assess our detector design methodology, before fabricating the planned Nb-feedline Al-KID bi-material detector(s), we created a series of single-layer Nb devices with gradually increasing complexities. These Nb devices offered us the opportunity to adjust our Nb fabrication parameters, develop the analysis methodology and software, and, taking advantage of Nb's negligibly low kinetic inductance, assess the RF performance dominated by the circuit layout for each of the components comprising the full detector. The layouts and transmissions of these Nb test devices are shown in Fig. 3.20. Our CPW feedline development campaign started in the later part of (Cornell, 2018), where we fabricated cm-long straight CPW feedlines to verify the CPW impedance determined by Eq. (3.7) in this thesis. Due to the practical requirement of connectorizing the devices, the tapered wire-bonding terminal design was included and therefore tested simultaneously. The result was a ≈dB transmission loss that, considering our readout system calibration, can be regarded consistent with the simulation. It shows our (a, b, c) dimension design for the new CPW is sufficiently 50 Ω-matched. Shown in Fig. 3.20, we continued (Cornell, 2018) to expand the CPW design to a minimal round trip, a single round trip across the 3" wafer, and then double round trips with arcs. The double round trip layout includes all the design elements that are used in our standardized feedline layout scheme.

Following the increasing complexity of the CPW routing, broad resonance-like structures observed in (Cornell, 2018)'s CPS feedlines began to appear also in our CPW feedlines. Nevertheless, contrasting with the 10s-dB deep structures that could obscure the KID resonances in (Cornell, 2018), our CPW transmission generally fluctuated within 5 dB, which we deemed manageable comparing to LE KIDs'

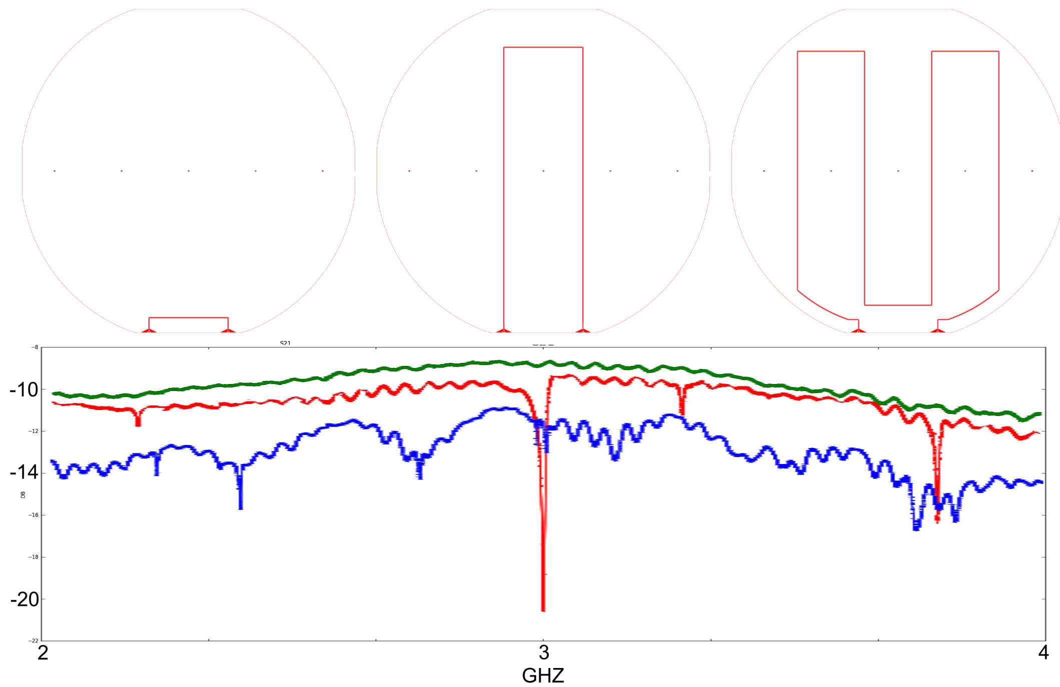


Figure 3.20: Top: Layouts of the Nb CPW feedline test devices. Bottom: The S_{21} for the top-panel devices. The layouts from left to right in the top panel correspond to the green, the red, and the blue curves in the bottom panel, respective. Due to missing readout attenuation records, the overall levels of the curves are not meaningful.

typical 10s-dB resonance size. As shown in Fig. 3.20, we did obtain several peaking structures in the feedline transmissions when the routing became more complex, but the structures were much wider than the expected narrow KID resonances, so one can distinguish the feedline structure from the KID resonance without confusion. The number of such peaking structures was also low enough to preserve most of the bandwidth for populating KID resonances. Therefore, we conclude our CPW-meander feedline routing scheme permits a clear identification of KID resonances and thus meets our goal. During the RF characterization, we frequently fabricated multiple devices of the same layouts with the then evolving fabrication techniques. Based on the high reproducibility of the transmissions given by the same feedline designs, we conclude the resonance-like structures are due to the routing designs of the feedlines, indicating it is RF design-related, but not a consequence of material or fabrication quality variation. More information about our early study for the CPW feedline is also provided in (Cornell, 2018).

In order to further resolve the undesirable peaking structure in our CPW transmission, we performed a series of measurements with a device that exhibited a

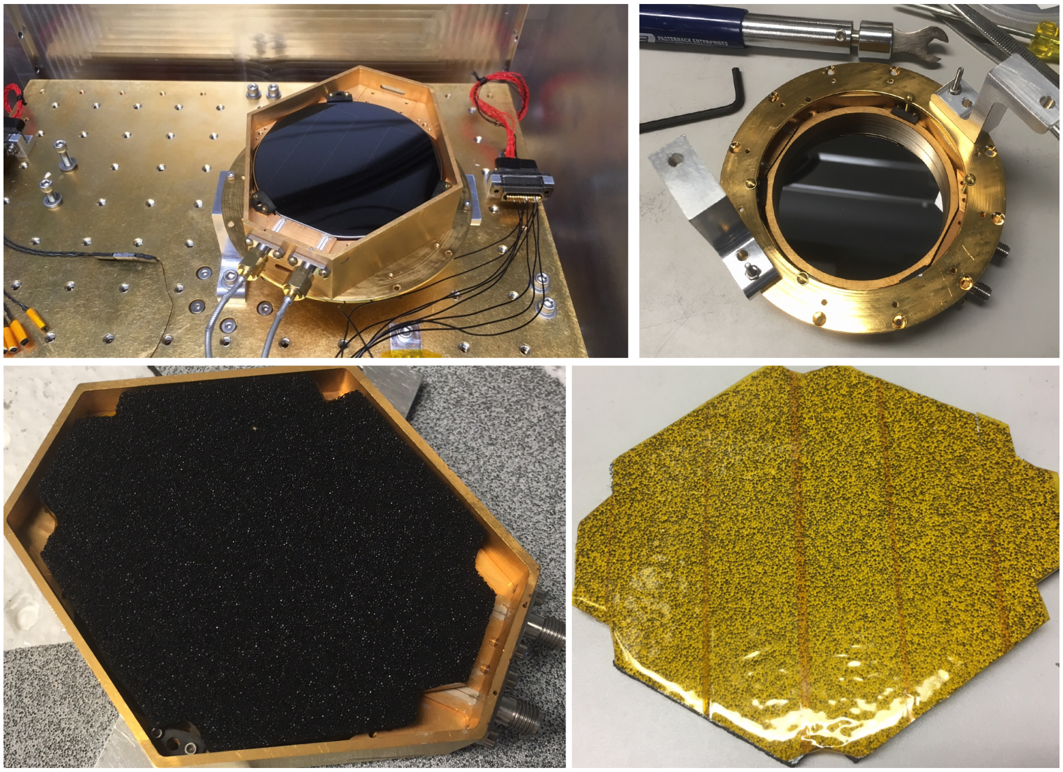


Figure 3.21: Top: The experimental construction for the lid-removed CPW S_{21} test. The left-hand side panel shows the 50-K Al enclosure surrounding the lid-removed device (top of Al enclosure removed for visibility). The right-hand side panel presents the bottom view of the device and its fixture, where we utilized the leg-shaped Al stand to raise the device from the refrigerator's bottom plate. The bottom lid of the device holder was also removed in this test. Bottom: The Eccosorb foam pad in the device holder (lid removed for visibility). The right-hand side panel shows the Eccosorb surface facing the device, which we covered with Kapton tapes to prevent the foam material from flaking onto the wafer.

particularly problematic S_{21}, in which we modified the device's surrounding RF environment by 1) removing the copper lid of the device holder that faced the CPW at a distance of ≈1 cm, so the CPW was exposed to a ≈30 cm wide, 50 K Al large enclosure. Alternatively, we 2) installed a piece of ⌀3" × 0.25" (Eccosorb, n.d.) foam pad, a strong GHz-range microwave absorber, in the space between the lid of the device holder and the CPW. The foam pad filled the space with a ≈mm tolerance. With Eccosorb's near-free space impedance, we expect it to suppress the RF reflection between the CPW and the lid of the device holder, as if the device was radiating into an open space. Fig. 3.21 shows the the experimental setups, and Fig. 3.22 shows the resulting transmissions for the different configurations.

We noticed that the narrow structures in the transmission were significantly reduced

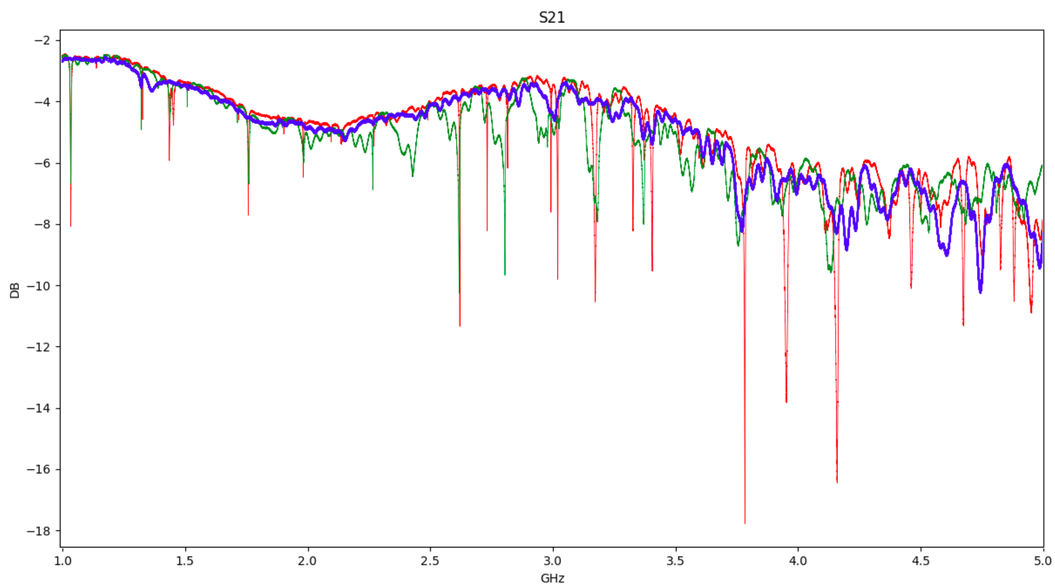


Figure 3.22: The CPW feedline transmissions for the unmodified (red), the lid-removed (green), and the Eccorsorb-filled (blue) configurations. The three curves were taken under an identical readout electronics configuration and therefore comparable, but we believe the overall S_{21} level of the data is uncertain by a few dB due to uncalibrated cable attenuation.

by removing the closely located lid that formed the hermetic copper enclosure. However, with a careful inspection, Fig. 3.22 shows that the narrow resonances in the original transmission were in fact "broadened" to be wider and shallower but mostly still remained. Despite such imperfection, the fluctuation in this case is generally shallow enough to allow a reliable identification of typical KID resonances. We proceeded to simulate an open space situation by inserting the Eccosorb pad, which we assumed created an equivalent configuration of completely removing the copper lid from the CPW-copper system. Note that we only installed the Eccosorb material at the instrumented side of the substrate to separate the closer copper lid, while there was another lid 2.5 cm away from the uninstrumented side of the substrate. Fig. 3.22 shows that the fluctuation was strongly mitigated by the modification, even compared to the physically opened scenario. The result demonstrates that a closely located copper structure can couple with the large CPW-only device to create the resonance-like structures in its transmission, and in this demonstration, the coupling is near-field due to the lid and the CPW's short distance compared to the wavelength. Note that, while the result intuitively suggests one can utilize the Eccosorb materiel to decouple the CPW and the surrounding mechanical structures, our experiment did not distinguish a CPW-copper lid coupling from a possible CPW self coupling, where different parts of the CPW separated further than the wavelength coupled to each other through the Eccosorb space and thus suppressed. It is also possible that, instead of a true decoupling, the Eccosorb material just created a much more lossy condition for the coupling and thus led to much widened resonances that were too wide and shallow to be identified from the continuum.

Motivated by the test result, we decided to utilize the Eccosorb shown in Fig. 3.21 for all the following measurements of this chapter. However, we have to report that, while inserting a piece of Eccosorb seems to be a straightforward solution for future applications, it is only a temporary solution to continue the R&D for the following reasons. First, we find the material difficult to be cooled to much below 1 K. While we were able to perform the experiment described above in a 4 K cryostat, when we repeated a similar experiment in a dilution refrigerator (DR) at 60 mK, we obtained a result that is consistent with the Eccosorb remaining much warmer than the base temperature. In the DR experiment, we replaced the KID-less Nb feedline-only device with a device that contained an identical feedline with Al KIDs and Nb ground shields. Comparing the data with and without the Eccosorb for this device, we find the internal quality factors of the KIDs became insensitive to the base temperature at higher temperatures when there was an Eccorsorb pad. We attribute the result to having a warm Eccorsorb constantly thermally radiating toward the KIDs, which held the KIDs' QP systems at higher equilibrium temperatures than without the Eccosorb. So, in the presence of these radiation-created QPs, the difference of the thermally created QP population became negligible at higher temperatures, and so as the corresponding internal quality factor change. It is still worth emphasizing that we did acquire the intended transmission improvement by comparing to the data with and without the Eccosorb. For future reference, we had foreseen the cooling issue with the loosely mounted foam type Eccosorb and changed it to a solid Epoxy type, but the difference was insufficient for our refrigerator to cool the large quantity of Epoxy. Second, for DM experiments to eventually adopt the KID technology, any associated apparatus locating closely to the main detectors must be either intrinsically low-radioactivity or can be purified. However, the underlying principle for the Eccosorb or similar type compounds to be lossy in RF is that they contain high concentrations of ferrite and usually also other carbide and metal powders. Considering modern rare-event experiments' sensitivities, this composition is unacceptably high-radioactivity, while a straightforward method for purifying such a complex mixture is unavailable at the moment. Regardlessly, to focus on the detector development at this stage, we utilized

Eccosorb in the following 4 K measurements and assume cryogenic-compatible, radio-pure substitutes will be available in the future. It remains for future DMKID group members or interested readers to identify alternative materials that exhibits the same microwave absorption as Eccosorb, or fundamentally understand the RF coupling so to mitigate the phenomenon at the circuit design level.

3.6.2 Complete detector

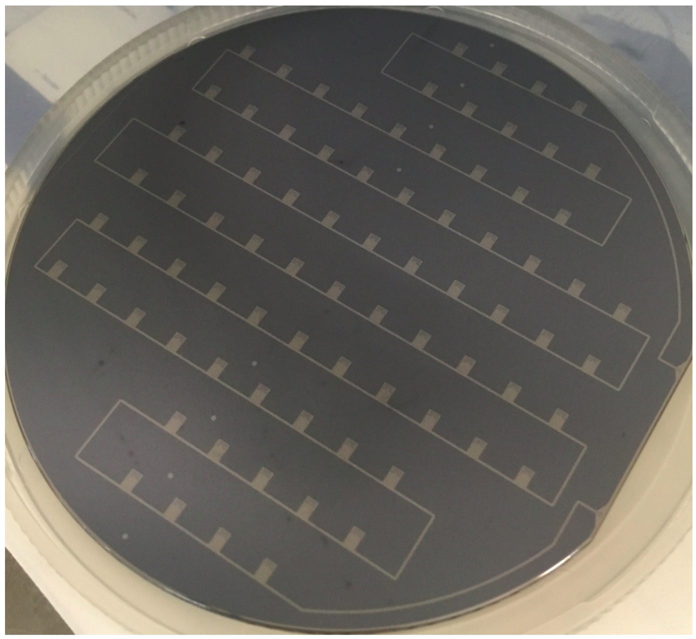


Figure 3.23: The 80-KID 300 nm-Nb RF assessment device.

We fabricated a complete 80-KID device, with the feedline, the KIDs, and the ground shields as in Fig. 3.19. We fabricated all the layout elements together in a single-etch fabrication from a single-layer 300 nm thick Nb film. Fig. 3.23 is the photo of the device. The fabrication technique is identical to the feedline fabrication part of our complete fabrication process for the Al-KID Nb-CPW devices. The readers may find a detailed discussion for the fabrication in Sec. 4.1. For this device to offer a more accurate RF performance assessment, we chose the Nb film thickness to be much larger than its SC penetration depth,[14] so the Nb film exhibited a minimal kinetic inductance that allowed the geometric inductance to dominate. Note that, since we directly adopted the layout in Fig. 3.19, in particular the identical inductor sizes (l) for placing Al KIDs in 3.0–3.5 GHz, these Nb resonators were expected to shift to higher resonant frequencies due to lower inductances ($f_r \propto 1/\sqrt{L}$).

[14]E.g., see (Gao, 2008) for a numerical calculation and reference values.

For our KID design, we expect the KIDs made of the same material to exhibit a similar kinetic inductance fraction, which is defined by

$$\alpha = \frac{L_k}{L_g + L_k}, \tag{3.20}$$

where L_k and L_g are the kinetic and the geometric inductances, respectively. The expectation is based on sCPS LE KID's high current density uniformity that yields $L \propto l$ (Eq. (3.12)), where we now write

$$L = L_g + L_k. \tag{3.21}$$

Knowing that the kinetic inductance L_k scales as the "square" count of the inductor geometry, which should be proportional to l since we do not vary the inductor trace width, Eq. (3.12) and Eq. (3.21) together suggest that L_g also scales as l. This inferred $L_g \propto l$ relation is expected for an inductor with uniform current density and thus is consistent with previous results. Since both L_g and L_k are proportional to l, we therefore anticipate the kinetic inductance fraction α as defined by Eq. (3.20) to be insensitive to l when we vary l for the frequency placement.

Using the above equations to rearrange Eq. (3.11), one can show

$$\frac{f_{r,\mathrm{Nb}}}{f_{r,\mathrm{Al}}} = \sqrt{\frac{1-\alpha_{\mathrm{Nb}}}{1-\alpha_{\mathrm{Al}}}}, \tag{3.22}$$

where the subscripts $_{,\mathrm{Al}}$ and $_{,\mathrm{Nb}}$, respectively, denote the designed resonant frequency for an Al KID and its corresponding value when the identical KID is made of Nb. Following our assumption that the Nb film possesses a negligibly low kinetic inductance compared to its own geometric inductance, we simply Eq. (3.22) to be

$$f_{r,\mathrm{Nb}} = \frac{f_{r,\mathrm{Al}}}{\sqrt{1-\alpha}} = \frac{1}{2\pi\sqrt{L_{\mathrm{Nb}(g)}C}} = \frac{1}{2\pi\sqrt{(1-\alpha)L_{\mathrm{Al}}C}}, \tag{3.23}$$

where we use $\alpha = \alpha_{\mathrm{Al}}$ since the assumption suggests $\alpha_{\mathrm{Nb}} \approx 1$. Considering a common $\alpha = 5$–10% for Al LE KIDs, Eq. (3.23) suggests that the Nb KIDs should offset by 3–5% upward in frequency. Since $f_{r,\mathrm{Nb}}$ and $f_{r,\mathrm{Al}}$ only differ by a factor of $\sqrt{1-\alpha}$, which we have argued can be treated as a constant, and also due to the fact that we designed for evenly spaced resonant frequencies assuming Al inductors (Eq. (3.10)), we expect the same set of the selected l to yield evenly spaced resonances also for these Nb KIDs.

Fig. 3.24 presents the S_{21} data we obtained at 4.6 K for the 80-KID 300-nm Nb RF assessment device. We observed 80 resonances unambiguously associated to the

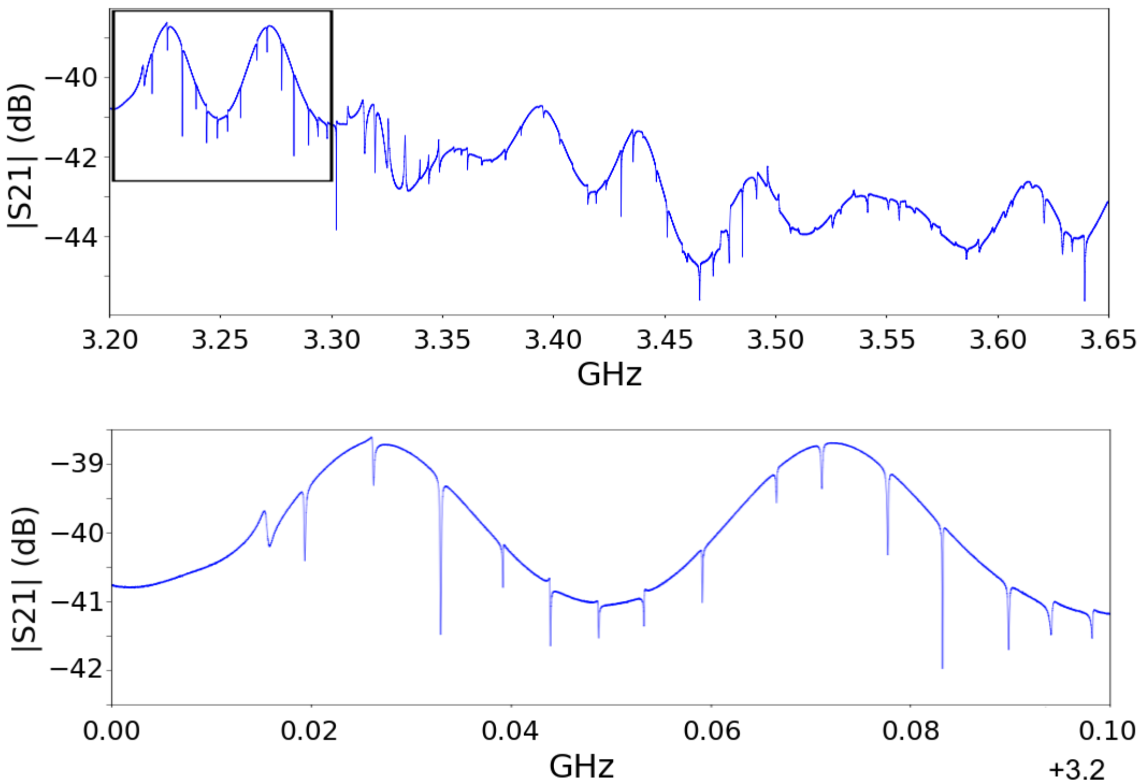


Figure 3.24: Top: The 4.6-K transmission of the 80-KID 300 nm-Nb RF assessment device. Bottom: A zoom-in of the the top figure to the marked region.

KIDs using the resonance formula developed in (Gao, 2008). All the resonances are distinguishable even though their absolute sizes are significantly suppressed by Nb KIDs' low internal quality factors.[15] Fig. 3.25 summarizes the data and shows that the resulting Nb resonances ranged from 3.216 GHz to 3.639 GHz. Recalling that we designed these Nb KIDs' Al counterparts to be in 3.055 GHz to 3.445 GHz, Eq. (3.23) and the assumed Al kinetic inductance value for the design/simulation (Tab. 3.1) suggest

$$\alpha_{\mathrm{Al},\Delta f_r} \approx 15.0\%, \tag{3.24}$$

which is possible but rather high compared to common Al KIDs. The subscript $_{\Delta f_r}$ denotes that the result was inferred from the range of spreading of the resonant frequencies but not from an individual absolute frequency(s). We also considered independently the absolute frequency placement result. We compared the average (central) frequencies of the fabricated Nb and the designed Al resonances, which

[15]The readers may find an introduction to the resonance characterization methodology developed by (Gao, 2008) in Ch. 4, where we also elaborate our particular mathematical interpretation for Gao's model and the derived code implementation based on the interpretation. The statement for the shallow resonances due to low internal quality factors is also justified therein.

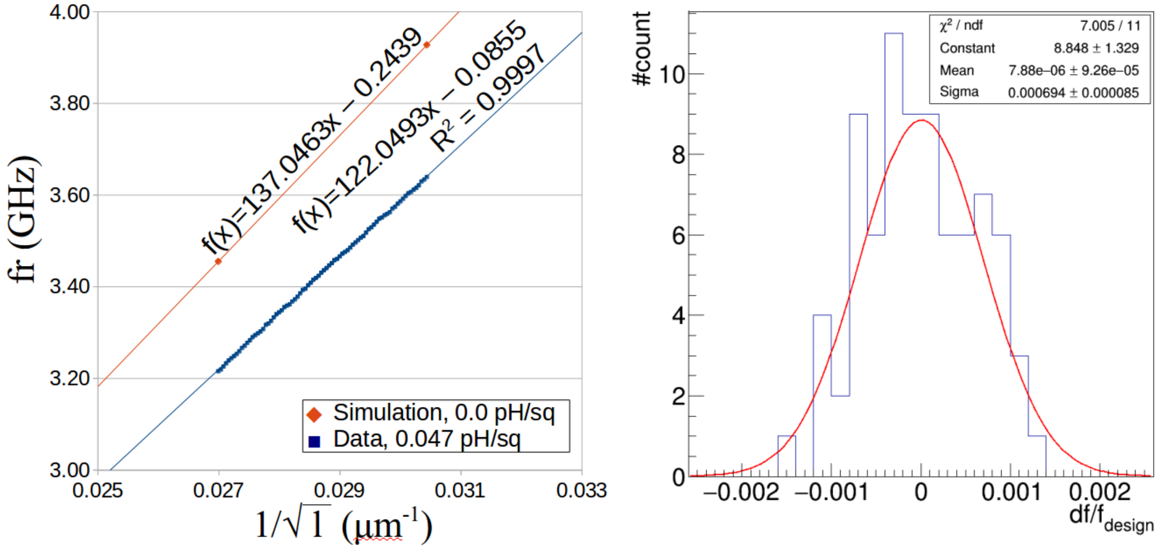


Figure 3.25: Left: The frequency placement result for the 80-KID Nb RF assessment device, presented as the obtained resonant frequencies (blue square) as a function of the designed $1/\sqrt{l}$. The blue curve is a linear fit to the data, with fitting result shown on the plot. The orange diamonds and their associated fit curve represent the L_g-only simulation prediction by setting $L_k = 0$ pH/□ in SONNET. For visual clarity, we choose to show only the first and the last simulated resonances, because by design the simulated data in between are perfectly evenly spaced and follow the linear design curve without error. By comparing the obtained Nb resonator data to the $L_k = 0$ simulation, we extract the Nb kinetic inductance of 0.05 pH/□ using the identical method of comparing the Al and Nb KIDs. Right: The statistics for the frequency placement accuracy, presented by the frequency difference of the observation and the linear model shown in the left-hand side figure. The difference is normalized by f_r to be an fractional accuracy. We use a Gaussian fit shown by the red curve to quantify the fractional accuracy distribution. More discussion may be found in the corresponding text.

yielded

$$\alpha_{\mathrm{Al},|f_r|} = 10.1\%, \tag{3.25}$$

where we use the subscript $_{|f_r|}$ to denote that the result was inferred from the absolute frequencies. Despite the nontrivial difference of Eq. (3.24) and Eq. (3.25), the increased 10–15% kinetic inductance fraction compared to the typical 5–10% is encouraging, because according to our theoretical model (Ch. 4), we expect the detector energy resolution (σ_E) for a KID-based phonon-mediated particle detector to scale as α^{-1}. We also think the difference of Eq. (3.24) and Eq. (3.25) is within a reasonable range, considering that the inductor possesses part of its inductance that does not scale by l, e.g., by the unmodified inductor traces transverse to the l direction (horizontal traces in Fig. 3.14).

There is also the practical frequency placement aspect of the RF performance. By selecting l according to Eq. (3.10) for the frequency placement, Fig. 3.25 shows that the fabricated Nb-KID resonant frequencies exhibited a well-modeled Gaussian distribution, centering at zero misplacement with a fractional uncertainty of

$$\sigma_{df_r/f_r} = 0.07\%, \tag{3.26}$$

i.e., ≈2.3 MHz depending on the exact f_r. To our knowledge, such a frequency placement accuracy was the highest ever achieved without post-fabrication treatments.[16] This result gives us confidence in our design and analyses. The highly predictive $f_r(l)$ relation also supports our earlier conclusion that the sCPS LE KID is dominated by the tunable inductance that is proportional to l, which allows a precise practical frequency placement for the multiplexing.

Fig. 3.26 shows the obtained Gaussian-distributed internal quality factor Q_i and coupling quality factor Q_c, which center at

$$\begin{cases} Q_{c,\mathrm{Nb}} = 10^{5.4\pm0.4} \\ Q_{i,\mathrm{Nb}} = 10^{4.0\pm0.3}. \end{cases} \tag{3.27}$$

In addition to the inherently low kinetic inductance of Nb, we performed the measurement at a nonnegligibly high temperature (4.6 K) relative to the Nb T_c (9.0–9.5 K). So, we anticipated the Nb KIDs to exhibit low internal quality factors as found above, which to the most part is irrelevant to this study. However, we were intrigued by the obtained coupling quality factors, which were mostly higher than the targeted $Q_c \approx 1 \times 10^5$ and yielded a distribution that was only marginally consistent with the design at the 1σ lower bound. The alarming implication of this result is that, as Fig. 3.26 presents the data in logarithmic scales, the observed 7% uncertainty in $\log(Q_c)$ in fact indicates almost an order-of-magnitude spreading from 1×10^5 to 6×10^5. Since we plan to control the coupling quality factor (dominating total Q) to achieve the optimal signal bandwidth, which allows us to fully reconstruct the phonon pulse shape and subsequently the event position, such high Q_c can potentially distort the pulses and reduce the detector's position reconstruction performance. At the same time, it is possible for the coupling to depend on the properties internal to the resonator but not just the KID-feedline distance. We repeated the simulation in SONNET using the Nb film thickness with a sheet inductance of zero. Although it was counter-intuitive to us that the coupling was not fully dominated by the KID-feedline distance in such a weakly coupled regime, we found the simulated

[16] As of 2018.

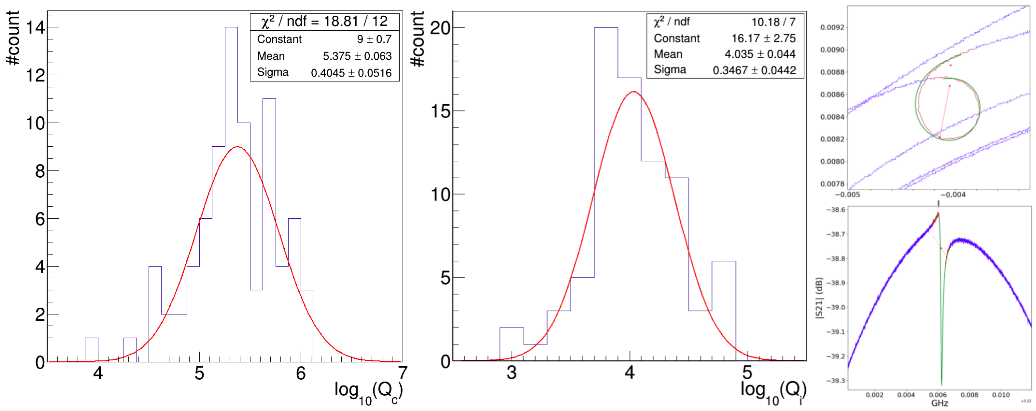


Figure 3.26: The coupling (left) and the internal (middle) quality factor distributions for the KIDs in the RF assessment device. The red curves are the Gaussian fits to the data, with fitting results shown in the top-right insets. We also attach on the right-hand side an example resonance fit for the parameter extraction; top: the fitted complex S_{21} plane, bottom: the fit result converted into absolute S_{21}.

Nb coupling quality factors were indeed increased by a factor of ≈2 from that we originally designed for the Al KIDs. If taking the factor of 2 difference to resclae the Nb-KID data, the observed $Q_{c,\mathrm{Nb}}$ distribution suggests

$$Q_{c,\mathrm{Al}} = 10^{5.1\pm0.4}, \tag{3.28}$$

which is now consistent with the design target.

We should note the simple rescaling to the central value of the Q_c distribution does not explain the undesirable wide spreading of the fabricated values. As we designed the CPW-KID distances to be identical, the large variation was not anticipated by the simulation nor by our physical intuition. The potential causes may be: 1) Fabrication-related nonuniformity, such as the film thickness, T_c, or L_k.[17] 2) KID-related, based on that we did not independently simulate the CPW-KID distance dependence for all 80 KIDs. 3) Integrated detector design-related. As mentioned previously, we were unable to simulate large-area multi-KID models in SONNET when we generated the current full-wafer design. Among these potential causes, we believe we can exclude 1) based on our ability to control the film deposition. Especially for such a thick and low-L_k film, the film property variation for a 3" device should be negligible. For 2), it is again not impossible but counter-intuitive to us that, given the LE nature, the coupling depends on the exact KID geometry (l) at such a significance. We suspect 3) to be the most possible cause to the Q_c

[17] These parameters are strongly correlated in reality.

variation. An example for the unexpected large-area effect due to integration is the feedline transmission. While the dimensions and the corners of the CPW were separately designed to be nearly perfect, we still found that the CPW can couple to the surrounding structures in the Eccosorb experiment. When we later compared the Eccosorb-equipped and the Eccosorb-free measurements for an Al-KID Nb-CPW device of an identical 80-KID layout, we found the coupling quality factors were drastically different. In the absence of the Eccosorb, KIDs consistently exhibited larger coupling phases similar to that described in (Khalil et al., 2012). Some of the Al KIDs were even fitted with unphysical negative Q_c by the (Khalil et al., 2012) model, suggesting the model is also insufficient in explaining our data. The result demonstrates that the coupling between the CPW and the KIDs in our design was subject to currently unidentified RF effects and therefore supports that the large Q_c variation as a derived consequence. A better model for our CPW-KID coupling, as well as for the transmission structure due to long-range effects, is currently being pursued by T. Aralis and K. Ramanathan.

To conclude the chapter, except for the still nonideal fluctuating S_{21} continuum and the Q_c variation, both pointing to unmodeled long-range RF effects, we have fully understood the Nb RF assessment device. We conclude that, aside from the two nonideal features, we have successfully reproduced the specified design target. We therefore conclude the work and proceed to the next stage, where we realize the Al-KID, Nb-feedline, highly multiplexed, KID-based, phonon-mediated particle detector.

Chapter 4

KINETIC INDUCTANCE DETECTOR-BASED DARK MATTER DETECTOR—REALIZATION

4.1 Fabrication

4.1.1 Overview

Fabrication yield had been a great challenge since we scaled up from (Moore et al., 2012)'s 2×2 cm^2 square layout to the full 3"-wafer circular layouts (Fig. 3.7). While KID's single-feedline readout enables its high multiplexibility, the fact that all the KIDs are read out by a shared feedline also makes the architecture vulnerable to feedline issues. In particular, as we began utilizing increasingly long feedlines, it became very difficult to avoid obtaining defects in the fabricated feedlines. As a result, our fabrication yield for working devices decreased significantly and became the dominant factor that limited the progress of the DMKID program.

In Ch. 3, we address this issue by replacing (Moore, 2012; Cornell, 2018)'s CPS feedlines with our CPW design. The change increased the feedline trace separation from a few μm to 11 μm (Sec. 3.5). We expected the improvement to be qualitative because the particles we typically obtained in the feedline gaps, which can lead to shorts between traces, were also of the size of a few μm. In addition, the fabrication technique (Cornell, 2018) adopted had a photolithographic resolution of $\approx$ μm, thus making fabricating the narrow-gap CPS inherently more challenging.

However, we only improved our fabrication yield to $\approx$50% with the design improvement. We would occasionally obtain large particles that created shorts between the CPW traces, which, as it turned out, still prevented us from fabricating working feedlines longer than 15 cm like that in Fig. 3.20 at a practical yield. We fabricated all our devices in JPL Micro-device Laboratory (MDL) H6 class-10 cleanroom. Given the geometry of our CPW and the supposed class-10 particle density in the laboratory, one can easily conclude the probability for one particle to randomly fall on the feedline region should be much lower than that we obtained. Therefore, we concluded there must be a mechanism(s) other than random occurrence for our this low yield.

The above particle acquisition estimate is based on the standard cleanroom classification, which is defined for midair space in an unoccupied cleanroom. Therefore, knowing human fabricators are the primary source of dust emission in a cleanroom, our devices could be experiencing a higher particle density due to human handling in close proximity. A quick survey also yielded that the particle density in a cleanroom can be orders-of-magnitude higher close to still objects and walls than in midair, because particles can naturally or be forced to plate onto these locations by the designed airflow for dust filtration.

We carefully revisited each step in our fabrication based on the above findings, which led to our current standard fabrication technique that has not failed a device since the improvement. Fig. 4.1 depicts the step-by-step fabrication for our Al/Nb-hybrid detectors. In summary, the fabrication comprises:

1. Sputter a 30 nm thick Al film. The thickness is chosen to be thinner than SC Al EM penetration depth for high kinetic inductance.
2. In the same chamber without breaking vacuum, sputter a thin layer (3–6 nm) of TiN on Al. TiN is hard and thus is chosen as the protection layer for the fragile Al film underneath.
3. Spin-coat the wafer with PR. We use an under-coat of polymethyl methacrylate (PMMA) to assist with PR adhesion, development, and removal.
4. Align the contact mask for the KID (Al) pattern to the wafer and expose the PR with 300 nm UV-B.
5. Post-bake the PR to harden the UV-dosed KID (Al) pattern and then develop the pattern in a metal-ion free (MIF) developer bath.
6. Use chlorine-based reactive ion-etch (RIE) to define the Al KID pattern.
7. *Immediately* after the Cl RIE, perform O_2 RIE and DI water cleaning to purge the residual Cl atoms that can damage the Al resonators. This step is critical for producing high quality Al KIDs.
8. Sputter a 300 nm thick Nb film. The large thickness is chosen for film homogeneity and the robustness against feedline defects.
9. Spin-coat the wafer with a designated thick PR, which also requires a PMMA under-coat as above.
10. UV-expose, post-bake, and develop the feedline (Nb) pattern similar to the KID patterning.
11. Use *buffered* Fluorine-based RIE to define the Nb feedline and ground shield patterns. The thick PR and the buffered RIE recipe, together with the control

techniques detailed later, permit an optimal timing for this step. Precise timing is critical for minimizing the Al KIDs' exposure to the F RIE, which can seriously damage the Al film. The thin TiN protection layer is immediately removed by the F RIE once exposed and thus does not give a detectable signal.

12. Final polymer cleaning process. Due to the hardness and the large quantity of the PR/PMMA chosen for our slow Nb etch, multiple steps of strong oxygen plasma etching and possibly stronger PR strippers like toluene are needed in addition to typical solvent cleaning.
13. If needed, dice and/or wirebond the water.

Most of the techniques described in the following are not profound, but they were only developed because we thoroughly reexamined every detail in our process, from understanding the airflow in the space to the fabricator's natural reflexes. For future DMKID fabricators, we advise to first understand the entire instruction, visualize and simulate the process in their minds, and then practice with mock-ups as much as possible before working with real devices. Most importantly, one should be very mindful about every small movement and decision during fabrication, and be sure to fully understand and evaluate the consequence before execution. While the advice may read generic, it has been proven decisive for our fabrication and micro fabrication in general.

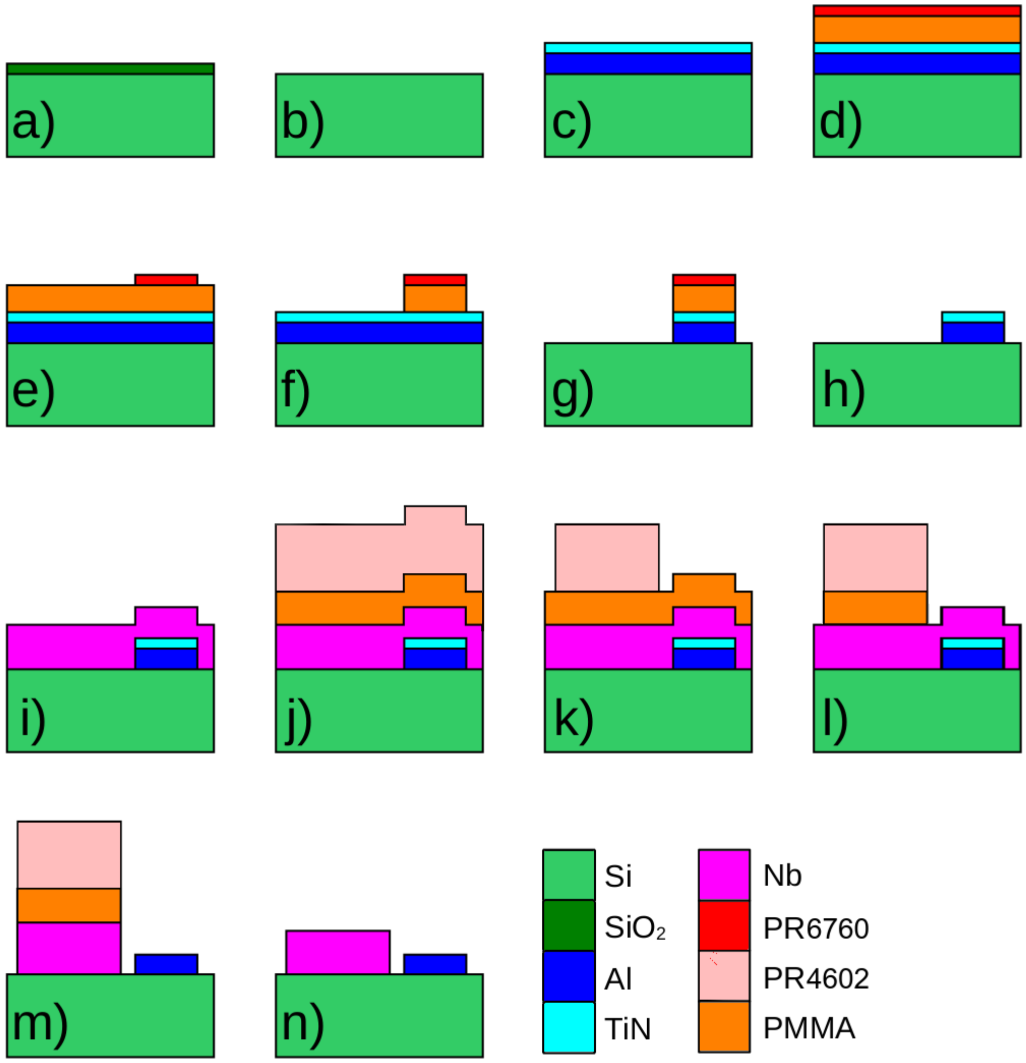


Figure 4.1: A cartoon summary for the DMKID Al/Nb-hybrid fabrication (not to scale). a) Si substrate with native SiO_2. b) Cleaned and deoxided Si substrate. c) Al and TiN deposited on Si. d) PMMA and PR deposited on Al/TiN film. e) PR UV-patterned and developed for the KID pattern. f) PMMA removed by ashing. g) KID features defined by Cl RIE. h) Device deep-cleaned by dedicated Cl purging process, ashing, and solvents. i) Nb deposited on Si and KID features. j) PMMA and thick PR deposited on Nb film. k) PR UV-patterned and developed for feedline pattern. l) PMMA removed by ashing. m) Feedline feature defined by F RIE, which also removes TiN. n) PR and PMMA removed by solvents, ashing, and PR strippers.

4.1.2 Substrate preparation

We start the fabrication from preparing the $\varnothing 3" \times 1$ mm substrate. The preparation consists of, in the order of application, visual inspection, solvent-immersed sonication, and vapor hydrofluoric acid (HF) surface deoxidation. The general rule of thumb for micro fabrication in a cleanroom is that, since not every step/process can be used as a stopping point, one should always prepare in advance all the following steps up to the next acceptable stopping point.

In particular for substrate preparation, this rule applies to steps that the wafer is exposed to open air that permits particle platting and surface oxidation. As explained previously, the former is particularly critical when humans work in close proximity. The former only applies to steps after the HF deoxidation. We try to avoid situations such as facing the (to-be-)instrumented surface upward in air without a pre-cleaned cover, and we always reserve and clean all the instruments, tools, and working spaces needed by the steps after the HF deoxidation up the metal deposition prior to applying HF.

Our substrate preparation process:

1. Load the wafer from the vendor container onto a 3-point wafer dipper and inspect for visible defects.
2. Sonicate the wafer in acetone at 3000 rpm, 1 min.
3. Sonicate the wafer in *another* beaker with isopropyl alcohol (IPA) at 3000 rpm, 1 min.
4. Sonicate the wafer in *another* beaker with deionized water (DI water) at 3000 rpm, 1 min.
5. Direct DI water rinse *by bench faucet*, 5–10 sec.
6. Blow-dry the wafer on a 95 °C hot plate and then cover it immediately to minimize air exposure time. The cover can be a clean cleanroom wipe or a pre-cleaned wafer holder lid.
7. Vapor HF surface SiO_2 removal, 1 min.
8. Mount the wafer onto the sputtering chuck.
9. Transfer the assembly into the load-lock chamber of the Al sputtering system.

Once the depressurization in the load-lock starts, the substrate preparation is complete. The fabrication process can be suspended indefinitely at this point. We provide a detailed list in Appx. C for other seldom emphasized but helpful practical tips.

4.1.3 Al sputtering

Our Al sputtering recipe for the KID film:

Parameter	Value	Unit
MDL LabView VI[1]	AlWire.vi	
Chamber pressure	$< 3 \times 10^{-9}$	mTorr
RF power	750	W
Ar flow	5	sccm
Pre-sput. time	1–2	min
Sput. pressure	1	mTorr
Sput. bias	420	V
Sput. current[2]	1.8	mA
Sput. rate[2]	50	nm/min
Sput. time	0.6	min
Sput. thickness	30	nm

[1] Virtual Instrument.
[2] Automatically reached under pre-set condition.

Table 4.1: The Al sputtering recipe for the KID film.

In order to minimize chamber contamination, the convention at the MDL H6 cleanroom is to open the main sputtering chamber to the attached load-lock only when the load-lock is pumped down to below 1×10^{-5} mTorr. We perform pre-sputtering to clean the target and to stabilize the plasma condition. It is done by starting the sputtering with a baffle larger than the plasma field between the wafer and the target that quickly moves away to start the main sputter. We typically pre-sputter for 1 min. but extend it to 2 min. if our process is the first Al sputtering of the day.

The parameters in the above table are typical values. Since a precise film thickness is critical for controlling the film's kinetic inductance, we always calculate the needed time to achieve the target thickness based on the latest Al film deposition parameters recorded for the system. We use the sputtering current, which is automatically reached by the recorded pre-set conditions, as the indication for the deposition rate. The current is proportional to the plasma flux and thus allows us to rescale the above sputtering rate and then determine the time needed by the target film thickness. We regularly achieve the targeted 30 nm thickness with better than ±0.5 nm variation across the 3" wafer.

Our TiN sputtering recipe for the Al protection layer:

Parameter	Value	Unit
MDL LabView VI	Titanium-Nitride.vi	
Chamber pressure	$< 3 \times 10^{-9}$	mTorr
RF power	1000	W
Ar flow	30	sccm
N_2 flow[1]	5	sccm
Ti flow	30	sccm
Ti pre-sput. time	0.5	min
TiN pre-sput. time	1–2	min
Sput. pressure	2	mTorr
Sput. bias	350	V
Sput. current[2]	2.86	mA
Sput. rate[2]	50	nm/min
Sput. time	0.1	min
Sput. thickness	5	nm

[1] Diatomic flow rate before entering the chamber.
[2] Automatically reached under pre-set condition.

Table 4.2: The TiN sputtering recipe for the Al-protection film.

TiN is much more robust against mechanical and chemical damages than Al and thus is chosen for the protection layer. We use a sputtering system equipped with both Al and Ti targets for the film deposition, so we can sputter Al and TiN in succession without extracting the wafer from vacuum. Since a few nm of TiN is sufficient to serve the purpose and can be reliably removed in the last RIE step, we do not time the TiN deposition as accurately as for Al. We estimate our typical TiN thickness to be 3-6 nm.

For future reference, it is worth mentioning that the above TiN sputtering recipe is developed and regularly calibrated by MDL colleagues for producing detector-grade high-quality TiN KIDs. One may obtain the desired Ti_xN_y stoichiometry by further fine-tuning the N_2 flow by at most ±0.3 sccm, otherwise the recipe is readily suitable for producing high kinetic inductance TiN KIDs.

4.1.4 KID pattern development

Our PR patterning recipe for the Al KID feature:

Parameter	Value	Unit
Dispenser	Dropper manual	
Disp. method	Dynamic	
MicroChem 495-PMMA C2		
Refractive index	1.49	
Disp. speed	75	rpm
Coating speed	1500	rpm
Coating spin time	60	sec
Thickness	160	nm
Baking temp.	115	°C
Baking time	$\gtrsim$3	min
Fujifilm GKR-6760 (PR6760) PR		
Refractive index		data unavail.
Disp. speed	75	rpm
Coating speed	1500	rpm
Coating spin time	60	sec
Thickness	820	nm
Baking temp.	115	°C
Baking time	1	min
UV-300 exposure		
Wavelength	300 (KrF)	nm
Power	25	mW/cm^2
Time	120	sec
Post-bake temp.	105	°C
Post-bake time	1	min

Table 4.3: The PR patterning parameters for the KID feature.

In the full sequence of applying the PR for Al patterning, we first spin-coat PMMA onto the Al/TiN film as a PR adhesion promoter, do a long soft-bake to cure the PMMA, then spin-coat PR onto the PMMA-covered wafer, and finally bake the PR. The above table lists the spinning speeds and times to achieve the targeted thicknesses. We also list the baking temperature and times found via extensive tests, to prevent under-curing, which can result in mixing of the PMMA and PR layers. In order to achieve a precise control of the following KID-defining RIE, we find a high quality construction of the PMMA/PR layer is arguably the most decisive factor. It should be free from rainbow-like interference fringes or uneven color distributions, which are the indications for 1) uneven liquid application or spin-coat distribution (circular/elliptic fringes), 2) local defects or particles (radial fringes), 3) mixing of under-cured PR/PMMA at the interface (patches), and 4) edge beads (dense fringes along the wafer edge). Fig. 4.2 provides example photos. We have

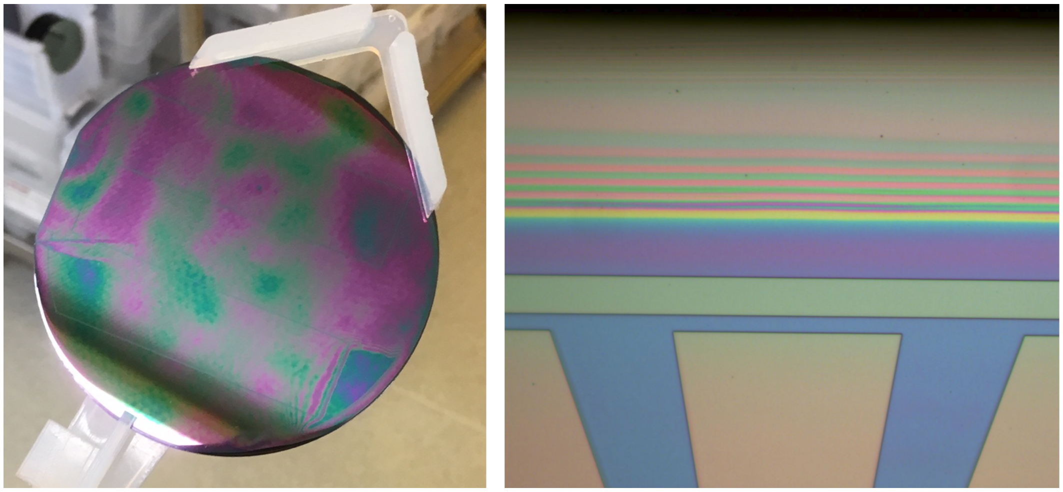


Figure 4.2: Left: The patch pattern suggests a nonuniform refractive index distribution, which is likely caused by the mixing of PR and PMMA or nonuniform layer thicknesses. Right: Edge beading causes the dense fringes at the wafer edge. The feature below the edge bead area is the 300-μm wide tapered wirebonding pad for the CPW central trace.

found that any of these PR/PMMA coating flaws are likely to result in poor RIE and thus cause irreversible damage to the device. So in this case, one should completely strip the PR/PMMA layer with solvent, repeat the wafer cleaning process, and then reconstruct the polymer layers. With the metal film already sputtered, one can not include the HF deoxidation as described for the wafer preparation cleaning; other steps are identical. If the fringes persist, we hope that the potential causes and symptoms listed above would help the reader effectively diagnose their processes and techniques.

Our current layout utilizes the meandering feedline pattern with input and output wire-bonding pads placed closely to the flat wafer edge (Fig. 3.19). Due to inconsistent profiling for the standard round wafer edge, it is our experience that some batches of Sil'tronix's 1-mm Si wafers are highly susceptible to edge bead issues. Such edge profile-related issue is manifested by ring fringes along the the edge and can lead to shorts between the CPW ground and central bonding pads for in current design (Fig. 4.2). While one can always apply post-fabrication dicing to remove the shorts, it is not a desirable solution as it can introduce damages through the additional step. It motivates us to replace the original i- and g-line PRs frequently utilized by (Moore, 2012; Cornell, 2018) by the current deep-UV (KrF) PR. i- and g-line PRs are typically more glutinous and therefore more prone to edge bead formation, while the deep-UV PR we select is thin and fluid. In addition to preventing edge bead formation, the fluidity allows us to apply the PR much more smoothly with normal droppers, as opposed to requiring filtered syringes for the viscous PRs. The change significantly improves the uniformity for our PR application and at the same time introduces much less particle contamination compared to syringes. In principle, the pattern definition by deep-UV (shorter wavelength) should also reduce edge diffraction hence provide a sharper pattern contrast and in turn higher-quality resonators; this advantage has not been explicitly checked. We also provide a list of practical techniques that we find seldom emphasized but helpful in in Appx. C for the readers' reference.

After PR/PMMA , the next step is to align the photo-mask for the Al pattern to the wafer. We find, similar to tool contact and PR application, the mask aligning process is also one of the major steps that can introduce particles from the mask to the wafer. However, limited by our requirement for large, undiced patterns, also for a high-pattern contrast and thus high-quality factor KIDs, hard-contact mask alignment[1] is currently the only practical option for our photolithography.[2] Depending on the fabricator's skill as well as the mask's alignment mark design, the alignment process requires exposing the instrumented wafer surface to open air for an extended period of time. In particular, the wafer is closely under the contact mask and the UV filter in the Karl Suss MA6 mask aligner's working space. It is thus important that these two instruments do not shed particles onto the wafer during operation. We always clean the masks using the acetone-IPA-DI water sonication procedure identical to our substrate preparation before every use to ensure no dust particle is transferred to the wafer during the contact. For the UV filter, we can not apply solvents, so we blow-clean the surface facing the mask carefully while keeping the surface facing downward to prevent particle plate-out. The UV filter and the mask then protects the wafer from particle plate-out once installed above the wafer. However, since the mask is transparent and has a slight gap from the filter, it is possible for particles to plate out on the top surface of the mask during aligning and subsequently be printed to the PR. Although it is likely that these particles can not cause a harmful high-contrast pattern due to out of focus exposure (offset by the mask thickness), we always gently blow-clean the mask again right before initiating UV exposure. We

[1]Contrasting with soft-contact alignment, hard-contact alignment requires a force to be applied to the mask to ensure the polymer layer to be patterned is compressed. The result is a better contrast but also the need to put the wafer the mask in firm contact.

[2]Recently the DMKID group is also developing the technique of adopting a stepper or an laser writer for the patterning, which avoids the need to place the wafer in contact with the mask. However, due to practical speed limitation, the technique is more suitable for fabricating smaller devices, e.g., cm-scale.

nevertheless suggest avoiding a strong blow that could agitate the particles around the working space. In addition to these highlighted major technique improvements, we again include more detailed instructions in Appx. C.

We typically pause at the completion of the UV PR patterning, i.e., successfully obtaining a visible pattern after the post-UV bake (Appx. C). We perform a detailed microscopic visual inspection of the entire feedline path. If no apparent >5-μm particle or defect in the inspected area, it is likely that the Al features will be reliably fabricated without leaving harmful features under the Nb feedline to be produced next. Since the cured PR/PMMA coat is thick and chemically stable for up to 3 days, if needed, we sometimes suspend our fabrication at this step and resume the next day. In this case, we store the wafers in commercial single-wafer holders, where the hard PR/PMMA coat provides the protection against physical damages and only accumulates dusts on the PR surface, which will be removed during RIE. Since the next appropriate stopping point is after depositing the Nb film, where the exposed Si surface and the Al features are again fully covered, we would choose to suspend the fabrication if the time does not permit completing the Al patterning and the Nb deposition.

Our recipes for removing the UV-exposed PR and then the PMMA underneath are:

Parameter	Value	Unit
PR development	Wet chemical solvent	
PMMA development	Soft plasma ashing	
PR development		
Developer	AZ 300 MIF	
Rinse time	60	sec
PMMA development		
Process #	7	
Step #	2	
O_2 flow	20	sccm
O_2 plasma gen. pres.	70	mTorr
O_2 RIE pres.	15	mTorr
O_2 RIE RF power	22	%[1]
PMMA etch rate	80	nm/min
PR etch rate	30	nm/min

[1] Set relative to the maximal power for MDL H6 asher, 13.56 MHz generator with a 600 W maximal power, driving a 10" circular plate with graphite cover.

Table 4.4: The recipes for developing the KID PR pattern and the PMMA-removal oxygen-plasma ashing.

We immerse the wafer into AZ-300 developer until the UV-exposed PR visibly dissolves. The reaction is typically done within 60 sec. but can occasionally take up to 70 sec. To ensure an uniform and complete AZ-300 application, we gently (without knocking the wafer inside) agitate the container to allow the PR to thoroughly mix with the flowing developer. A successful PR development leaves only the desired Al pattern in the PR/PMMA with an uniform color for each material. A counter example is given in Fig. 4.2, where the coloration does not follow the PR/PMMA thickness distribution based on the intended pattern, indicating an unsuccessful removal of the PR/PMMA. Possible causes to the result include: PR/PMMA mixing at their interface due to under-bake or under-cure, unexpected thickness variation due to wafer or SC film flatness issues, and/or mechanical or chemical damages to the polymer compounds. When obtaining such undesirable results, or the pattern does not develop within 70 sec., we strip the PR/PMMA layer and repeat the process. For successfully patterned wafers, we rinse off AZ-300 with DI water inside the developer container and then blow-dry the wafer on a room-temperature chemical bench. At this stage, dust particles attached to the wafer during drying can not leave a feature for the SC film, so we choose to dry the wafer at room temperature to avoid altering the PR/PMMA baking result. Prior to irreversibly patterning the SC film with RIE, we inspect the full wafer under the microscope for visible defects. If we find any defect that is large enough and at a location to invalidate the purpose of the device, and the detect is contained in the PR/PMMA layer, we strip the layer and repeat its construction. If the defect is in/under the SC film or the substrate, we abandon the wafer.

We need to remove the PMMA before etching the Al film with Cl plasma. The removal can be done separately in a dedicated oxygen ashing apparatus first or, since all MDL RIE systems are equipped with O_2, in the RIE chamber as an oxygen plasma pre-cleaning. However, according to our experience, we strongly prefer the former for several reasons. First, the MDL asher is a single-chamber, fast turnaround machine, which allows users to incrementally apply the ashing without waiting for the load-lock extraction time for the RIE systems. The PMMA ashing takes about 2 min. (Tab. 4.4), which we typically divide into several 20–60 sec. sessions interleaved with quick visual inspections. Second, like most typical RIE systems, MDL RIE systems use single-spot laser reflection to determine the progress of the material removal. Due to RIE systems' smaller plasma fields, the slow etching for the thick PMMA, and the small laser spot size relative to our full-wafer design, removing PMMA in RIE systems is difficult and has resulted in irreversible damages to our SC films. Finally, contrasting with the RIE systems that only follow pre-programmed recipes, the MDL asher allows on-the-fly plasma adjustments. As shown in the above recipe table, we prefer initiating the plasma at the nominal pressure (70 mTorr) for the process-7 step-2 recipe but quickly lowering the pressure to 15 mTorr depending on the plasma stabilization condition. This adjustment reduces the overall etching rate and therefore enhances PMMA removal control while preventing potential damage to the PR and the SC film. It also produces a more vertical PR pattern. We note that, although our PMMA-removal ashing recipe is tuned to exhibit a lowered etch rate for hardened organic compounds like the PRs, it still etches the PR at 30 nm/min., which is nonnegligible compared to the PMMA etch rate (80 nm/min., Tab. 4.4). We advise checking the remaining PR thickness when adopting the recipe to other processes.

We perform a final microscopic inspection with the now high-contrast PR pattern on the exposed Al/TiN film before permanently etching the film. In addition to confirming the overall pattern is produced defect-less, the high contrast allows us to inspect the radii at the corners, which are the indication of the photolithographic resolution. We typically achieve better than a fraction of μm radius at every corner, which we believe is critical for producing resonators with high quality factors.

4.1.5 Al KID definition

Etching of the Al KID layer consists of mainly executing the predefined program described below, which can be easily performed at high consistency with the fully automated RIE equipment. Nevertheless, one should be aware of and fully prepared for the strict deep cleaning procedure that must follow *immediately* after the RIE. If one delays the deep cleaning by more than 3 min., our experience is that the residual Cl etching agents always create random corrosion on the Al films as shown in Fig. 4.3. Resonators with such result perform very poorly and are unusable for a particle detector. Since the deep cleaning procedure requires certain chemicals, instruments, and wet bench work space, to avoid delaying the process, we always request exclusive use of the RIE instrument with all the necessary tools/spaces for the deep cleaning before starting the Al RIE step. We frequently use Kapton tape or wax to attach the 3" wafer to be etched to a 6" backing wafer that is fully coated with PR. It provides more balance to the wafer and prevents it from being blown into the RIE chamber by etching gases.

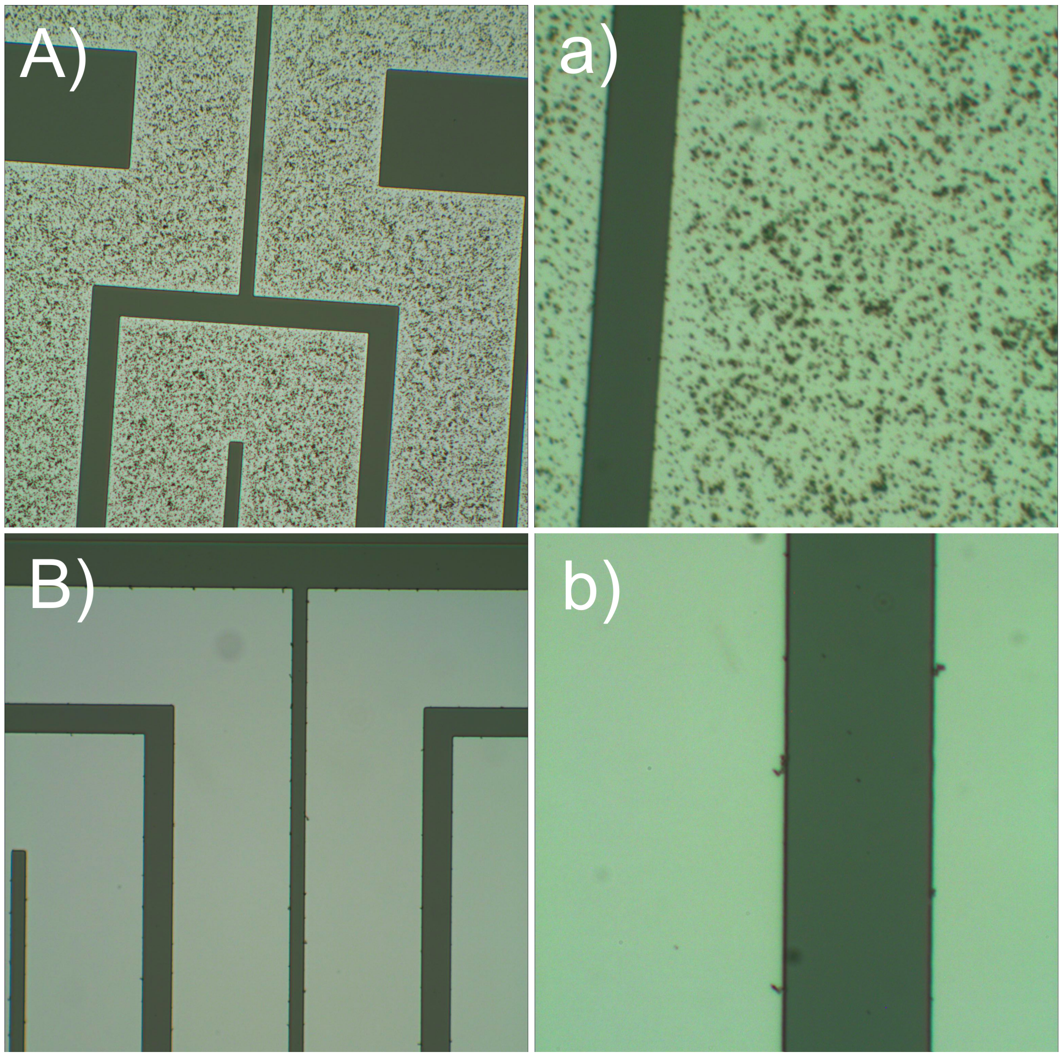


Figure 4.3: The photos on the right show the magnified regions for the corresponding photons on the left. A/a) Large-area surface cavities due to the inclusion of Feron-12 (CCl_2F_2) etching agent in a previous RIE recipe. B/b) Sidewall cavities due to residual Cl atoms attached to the film sidewalls. This photo result was produced by applying the nominal Cl-only recipe but with delayed/incomplete execution of the post-etch deep cleaning.

Our Al RIE recipe:

Parameter	Value	Unit
Program name	`bb_trilayer`	
Backside He cooling pres.	7	Torr
Oxygen organics cleaning/ashing		
O_2 flow	60	sccm
O_2 RIE pres.	8	mTorr
O_2 RIE RF power	100	W
O_2 RIE bias power	60	W
PMMA etch rate	2.3–2.6	nm/sec
Cl-based Al RIE		
BCl_3/Cl_2 flow	20/15	sccm
BCl_3/Cl_2 RIE pres.	10	mTorr
BCl_3/Cl_2 RIE RF power	350	W
BCl_3/Cl_2 RIE bias power	30	W
Al etch rate	1.1–1.2	nm/sec

Table 4.5: Our Cl-only RIE recipe. The O_2 ashing part in this table is included as intermediate steps in the `bb_trilayer` program, which is a flexible option for removing organic compounds or cleaning the wafer. We previously explained that one can use this option to remove PMMA but we do not prefer doing so. Based on previous discussion, if the PMMA layer is removed in advance, we skip this ashing step in `bb_trilayer` or apply it for 3–5 seconds if we feel the wafer needs further cleaning, e.g., due to reapplications of liquid chemicals for multiple PR/PMMA reconstructions. More explanation is found in the corresponding text.

We explain below our RIE sequence following the same step order, also labeled identically, as in the `bb_trilayer` program "steps" in MDL's PlasmaTherm Versaline RIE system; a more detailed discussion including practical techniques is offered in Appx. C. The Cl RIE program:

(a) Initialization: Load-lock evacuation and automatic wafer transfer.
We focus the end-point laser detector to where the Al film is to be removed, usually also with PR-protected features in the camera field for a visual comparison. We wait at this step until the post-RIE deep cleaning facilities are available.

(b) O_2 flow stabilization: Injecting O_2.

(c) O_2 plasma activation: Injecting RF power to initialize ashing plasma.
The chamber pressure, RF power, and RF bias are temporarily adjusted to 12 mTorr, 100 W, and 40 W, respectively, to activate the oxygen plasma. The program automatically continues to the next step in 10 sec.

(d) O_2 RIE/main ashing:

This step utilizes Tab. 4.5 parameters for an indefinite organic compound cleaning/ashing until manually continued to the next step. Since this ashing recipe is mainly for an overall cleaning for the devices before the main RIE, it is designed to have less differentiation (etch rates) between PR and PMMA. Therefore, we prefer only using it for wafer cleaning as explained previously.

(e) BCl_3/Cl_2 flow stabilization: Purging ashing plasma and injecting etching agent gases.

(f) BCl_3/Cl_2 RIE: Main Al etch.

Once started, the plasma RF power turns on and the Al etching begins immediately. The Cl plasma removes the thin TiN within half second and starts removing the Al film at the noted rate (Tab. 4.5). Based on our experience, the end-point signal should start decreasing at 25 sec. and plateau at 28 sec., with a total signal decrease of about 50%. The timing can vary by ± 1 sec., and we also allow the etch to continue for 1 more sec. to account for the etch rate variation across the wafer, as well as to confirm the signal plateauing is not due to noise. We strictly follow this well-tested timing scheme to prevent the over-etch that leads to surface and sidewall corrosion (Fig. 4.3).

For future reference, we have also tried other more commonly adopted Cl/F-mixed RIE recipes. The inclusion of fluorine, e.g., Freon-12 (CCl_2F_2), provides faster etch rates and produces more vertical profiles, but we find the few seconds-long process difficult to control with precision and oftentimes leads to the large-area cavities shown in Fig. 4.3. Aside from damaging the Al features with F, we also discovered by accident that, if the exposed Si surface is further etched by our current pure-Cl recipe for 10–20 sec., it can create the "black silicon" surface structure (Wikipedia contributors, 2021), which is an extreme surface dehomogenization that is likely detrimental to both the substrate phonon transport and the device's radiation and thermal properties. It is also worthwhile to note that, while BCl_3 etches Al more slowly than Cl_2, R. LeDuc found that the boron plasma can assist the penetration of the surface AlO_x and subsequently lead to much more uniform (horizontally) and vertical etches. We believe this uniformity and verticality is beneficial for our resonator quality.

(g) Chamber purging, end of process.

In this step, the RF power for maintaining the Cl plasma first turns off and then the gases in the chamber are pumped out. As soon as the chamber returns to the nominal vacuum level, the wafer is automatically transferred to the load-lock, and then the load-lock is vented with N_2 to atmospheric pressure.

We find that residual Cl atoms attached to the Al film can create corrosion cavities. To remove these Cl atoms, we transfer the device into the asher within 2 min. to perform a strong oxygen ashing for *at least* 2 min. We then rinse the entire device with flowing DI water for *at least* 2 min. to dilute and remove any remaining chlorine. We use MDL's general-purpose "high-pressure" deep oxygen cleaning recipe for this Cl cleaning:

Parameter	Value	Unit
Process #	6	
Step #	1	
O_2 flow	40	sccm
O_2 plasma gen. pres.	150	mTorr
O_2 RIE pres.	15	mTorr
O_2 RIE RF power	15	%[1]
Time	>2[2]	min
PMMA etch rate	—[2]	nm/min
PR etch rate	—[2]	nm/min

[1] Set relatively to the maximal power on MDL H6 asher, 13.56 MHz generator with a 600 W maximal power, driving a 10" circular plate with graphite cover.
[2] As long as needed.

Table 4.6: The recipe for the strong general-purpose oxygen ashing.

We oftentimes also apply Tab. 4.6 to remove PR, PMMA, and other kinds of organic compounds and polymers indifferently for deep surface cleaning. We want to minimize the time after the device acquires the active etchants until they are purged, so we always bring the RIE backing wafer with the device into the asher and then dismount the device afterward. In this case, we also hold the backing wafer by hands to save tooling time without the concern of damaging the device. After the strong O_2 ashing, we dismount the device from the backing wafer and rinse it with DI water *completely and continuously* for >2 min. To avoid damaging the newly fabricated Al pattern, instead of rinsing with a spray gun, we use a DI-water faucet adjusted to a gentle stream and introduce the water to a bare-Si region. We reduce the stream to be as small as possible that can still form a complete water film covering the entire wafer at a 45° angle. If the Cl purging process is performed thoroughly, at this stage the device should be free from the concern of corrosion and may be inspected under a microscope. Since we will immediately reintroduce the device to wet chemicals for PR/PMMA stripping, we do not dry the device thoroughly for this inspection.

4.1.6 Nb sputtering

We apply acetone, IPA, and then DI water sonication cleaning, 1 min. for each, identical to the substrate preparation, before sending the device into Nb sputtering. Readers may refer to the previous sections on the substrate preparation and the Al/TiN sputtering for detailed introductions. The differences between the cleaning procedures for the original substrate preparation and here include: First, we skip HF deoxidation, which would otherwise damage the Al KIDs. Second and also for the KIDs, we handle the device with extra care during the sonication and the mounting, dismounting, and the transfer of the device. We only allow tools and wafer holders to contact the areas that are far from the KIDs and also the CPW feature to be fabricated next. Third, we inspect the fabricated Al features under a microscope and compare the obtained corner radius to that previously measured for the PR. We also measure the film thickness and the sidewall profile, using only two designated KIDs, one at the center and the other at the edge, to minimize the KIDs that can be potentially damaged during the measurement. We proceed only if the above measurements are consistent with expectation and thus do not suggest an error. We also check along the KIDs for potential fatal defects or unetched Al patches that can block the feedline. If we notice such failures, depending on the situation and/or the purpose of the device, we sometimes perform repair photolithographies by covering the entire wafer with Kapton tape except the defected region, expose the local spot with microscope light, and then attempt to etch the defects in another RIE. The RIE can etch a few to tens nm into the substrate surrounding the Al patch, which can be accommodated by the 300 nm thickness of the Nb film to be deposited next. We prefer such repair even though it could damage a few KIDs in the region, compared to blocking the feedline and subsequently losing the entire device.

We use the following recipe to deposit 300 nm of Nb for the feedline:

Parameter	Value	Unit
MDL LabView VI	NbWire.vi	
Chamber pressure	$< 3 \times 10^{-9}$	mTorr
Ion milling		
Ar^+ mil. time	5	sec
Ar^+ mil. bias	155	V
Ar^+ mil. current	20	mA
Main Nb sputtering		
RF power	300	W
Ar flow	20	sccm
Pre-sput. time	30	sec
Sput. pressure	5	mTorr
Sput. bias	330	V
Sput. current	870	mA
Sput. rate	45	nm/min
Sput. time	6.67	min
Sput. thickness	300	nm

Table 4.7: The Nb sputtering recipe for the feedline film.

Compared to Al sputtering, we replace the HF deoxidation step prior to the Al deposition with a 5-sec. ion milling for the Nb deposition. The purpose is not to physically mill materials on the wafer (in fact it could be unwelcome for the KIDs) but to ensure that the water molecules attached to the device after the cleaning are fully expelled before depositing the Nb film, and also because we can not use HF on the Al KIDs. Since the thick Nb film is robust against mechanical damages and thus serves as a good protection for the Al KIDs underneath, we typically suspend the fabrication after depositing Nb and continue the next day.

4.1.7 Feedline pattern development

Parameter	Value	Unit
Dispenser	Dropper manual	
Disp. method	Dynamic	
MicroChem 495-PMMA C2		
(Identical to Al PR/PMMA patterning, c.f. previous)		
Fujifilm GKR-4602 (PR4602) PR		
Refractive index		data unavail.
Disp. speed	50	rpm
Distribution time	10	sec
Coating speed	3000	rpm
Coating spin time	60	sec
Thickness	3500	nm
Baking temp.	115	°C
Baking time	1	min
UV-300 exposure		
Wavelength	300 (KrF)	nm
Power	25	mW/cm^2
Time	160	sec
Post-bake temp.	105	°C
Post-bake time	1	min

Table 4.8: The recipe for developing the feedline PR pattern.

In general, most of the principles and the handling techniques introduced previously for the Al PR/PMMA patterning are applicable for the Nb process; readers may refer to the previous discussion for detailed information. The major change between the Al and the Nb PR constructions is the chosen PR. After testing many PR-RIE recipe combinations, we eventually settled on Fujifilm GKR-4602 with CHF_3-only RIE. The primary motivation for the recipe is that the ultra-thick GKR-4602 permits the use of a mild, slow RIE for the Nb feature definition. The slow process allows us to respond to the end-point laser signal under our timing-based etch control precisely, which prevents the damage to the Al KIDs underneath. However, the atypically thick PR also comes with special qualities that require dedicated handling techniques we developed. After we construct the PMMA layer identically to the Al fabrication, we directly pour the very viscous PR from the bottle to the center of the slow-spinning wafer. Similarly to Al fabrication, we adopt dynamic dispense for GKR-4602, but due to its high viscosity, we can only create a smooth dispensing stream by pouring steadily from the bottle rather than with a dropper or syringe. To achieve an uniform PR distribution, we lower the dynamic dispense spin[3] to 50 rpm and let the slow spinning continue for 10 more seconds after the accumulated PR at the wafer center reaches a diameter of 1"–1.5"; we then increase the spinning speed to the much higher coating speed. After the high-speed coating, the viscosity of the PR commonly leads to icicle-like spikes around the edge of the wafer. These spikes can generate stress and eventually crack the wafer from the edge during the long high-temperature UV exposure or RIE. To remove these spikes, we approach the wafer edge with an acetone-drenched cleanroom swab while the wafer spins at 5–10 rpm. It dulls the spikes extending out from the edge without applying too much force or touching the PR coat, which could damage/remove the PR on the edge and lead to a SC ring feature that shorts the CPW. Note that the removal is only doable before baking the PR. We suggest future DMKID fabricators practice this technique with dummy wafers. Because the GKR-4602 PR is infrequently used and thus its properties can suffer from aging, we further recommend practicing the full sequence of PR application, baking, UV exposure, post-baking, and development to test the recipes and adjust them when necessary.

After the completion of the PR/PMMA application, the following mask alignment, UV exposure, post-UV bake, wet PR development, and ashing for the PMMA development, in this written order, are all identical to the Al procedure already introduced. The only exception is that, due to the change of PR, the UV exposing time is extended to 160 sec. Also due to the much thicker PR, the quoted time and other parameters in the table are not as precise and should be tested in advance as explained previously. Since the sharpness of the Nb feedline is not as critical for the sensitivity of the detector, but we do want to avoid large defects ($> 5\ \mu$m) that may significantly alter the waveguide transmission, so we extend, as needed, the suggested times by up to 20% to ensure the full PR/PMMA hardening, exposure, and removal. Similarly, after the CPW feature is fully developed and the PMMA adhesion promoter ashed, we carefully confirm that the pattern is indeed produced without any visible ($\sim \mu$m) defect in the CPW gaps before proceeding to RIE. On the contratry, depending on the purpose of the device, we usually accept defects that can potentially short the feedline's ground trace to a KID or appears on a KID. This type of defects can at most invalidate the KID but is unlikely to affect the overall feedline transmission or other KIDs.

[3] i.e., the spinning speed during PR dispensing.

4.1.8 Nb feedline definition

Parameter	Value	Unit
Program name	`BB_PMMA_SiO2`	
Backside He cooling pres.	5	Torr
Oxygen organics cleaning/ashing		
O_2 flow	10	sccm
O_2 RIE pres.	10	mTorr
O_2 RIE RF power	150	W
O_2 RIE bias power	40	W
PMMA etch rate	2.3–2.6	nm/sec
Cl-based Al RIE		
CHF_3/O_2 flow	40/2	sccm
CHF_3/O_2 RIE pres.	10	mTorr
CHF_3/O_2 RIE RF power	400	W
CHF_3/O_2 RIE bias power	25 (35–40)[1]	W
Nb etch rate	0.63–0.71 (<0.35)[1]	nm/sec

[1] **If the plasma is ineffectively activated, adjust the bias power/etch rate according to the instruction detailed in the corresponding text.**

Table 4.9: The pure-CHF_3 RIE recipe for the current-generation DMKID feedline fabrication, *modified* from the original `BB_PMMA_SiO`$_2$ program. We advise users to double check the program details before execution. If the original `BB_PMMA_SiO`$_2$ program containing Cl is utilized, the KIDs are likely to be damaged as explained in the corresponding text.

RIE recipes for Nb or SiO_2 commonly include mixtures of fluorine/chlorine-combined compounds such as CF_xCl_{4-x}, CH_xCl_{4-x}, and CH_xF_{4-x}. The inclusion of both F and Cl helps with the etch rate, the sidewall verticality and therefore is especially popular when the film thickness is greater than 50 nm. Nevertheless, in our recipe shown in Tab. 4.9, we intentionally avoid any Cl-based compound to prevent damage to the Al KIDs once the Nb is fully etched; Al is otherwise relatively durable against F plasma. Without the assistance of chlorine, the reduced etch rate leads to a much longer etch time and correspondingly more tapered sidewalls, but due to the large dimension of our CPW design, it has not shown a noticeable impact to the feedline's functionality. In exchange, the reduced etch rate allows us to more accurately terminate the feedline etch for minimizing the exposure of the Al KIDs to the Cl etchant. We note that, although in principle florine exhibits too high a reduction potential to react with Al, in practice we find the F plasma still mildly but noticeably damages our Al films with the help of the RF energy. Therefore, in addition to weakening our etchant, we also lower the RF bias for plasma generation from the original `BB_PMMA_SiO2` setting to the quoted 25 W. The recipe and power adjustments together result in a 7–8 min. nominal etching time for our 300 nm Nb film. For future R&Ds, we find the adjustment occasionally initiates the plasma in another significantly inefficient state, resulting in a much weaker RIE that lasts for more than 15 min. While the chosen thick PR is still compatible with this type of abnormal process and protects the film, the unpredictability does create inconvenience that requires constantly pausing the RIE to extract the wafer from the chamber for visual inspections. We have tried and successfully fabricated test devices using CHF_3/CF_4-mixed RIE recipes with bias powers in the range of 35–40 W, which do seem to provide more reproducible and stable plasma conditions. We however have not rigorously pursued stabilization of this improved recipe since the current technique performs acceptably.

We again provide an explanation for our Nb RIE process following the program labeling for better referencing, and a more detailed version is found in Appx. C:

(a) Initialization: Load-lock evacuation and automatic wafer transfer.
We focus the end-point laser detector at a location where the the Al film will be revealed after the Nb is etched through, e.g., at a KID inductor, so we can obtain a precise timing to minimize the Al film's plasma exposure. Unlike in the KID-defining Cl RIE, where we wait for an unambiguous plateauing of the signal, here we terminate the RF power immediately when the steady rising signal changes noticeably (see (e) for detail). We typically also choose a camera field that includes Nb-covered bare substrate and PR-covered Nb so we can monitor the progress and ensure that the Si substrate and/or the Nb feedline are not etched unexpectedly.

(b) O_2 flow stabilization: Same as Cl RIE described in Sec. 4.1.5.

(c) O_2 RIE: Same as Cl RIE described in Sec. 4.1.5.

(d) CHF_3/O_2 flow stablization: Purging ashing plasma and injecting main RIE gases.
It has been shown that the oxygen in the recipe promotes temporary oxide formation on the sidewalls, which is easier for the main etchant(s) to react with than the original metal(s). We find including a small amount of oxygen improves the verticality of the RIE process.

(e) CHF_3/O_2 RIE: Main Nb etching.
As noted in (a), we focus the end-point laser at a location with Al under Nb. Since the Al film is more reflective than the Nb film, as the Nb is being thinned at a constant rate, the reflectance at the spot increases hence the steady rising signal. In principle, one should expect such signal to jump to a high value and then plateau when the Al film is exposed, but due to the slow etch rate of our recipe, the thin TiN cap on the Al film, the slow update of the etcher software, and also noise, the transition can take 1–2 sec. in a somewhat unpredictable trend. So, in order to minimize the plasma exposure to the KIDs, we choose to immediately terminate this step when the end-point laser signal shows a noticeable change to its steady rate of rise *regardless of the sign of the change*. The signal should continue to complete the expected trend in the following 1–2 sec. Due to a known plasma nonuniformity and the exact location of the laser spot, the Nb etch may be incomplete in the outermost few mm region of a 3" wafer. So, after terminating the RIE, we extract the wafer for a visual inspection and, if needed, continue the RIE at $\leqslant$3-sec. increments until the remaining Nb is visibly removed. Similar to our Cl RIE recipe, we find our F RIE recipe can produce black silicon if applied to bare substrate for an extended period of time.

(f) Chamber purging, end of process: Same as Cl RIE described in Sec. 4.1.5.

Ideally after the Nb feature definition, we should also apply the same deep cleaning treatment as for the Cl RIE, but due to florine's lower reduction potential relative to chlorine, we find it is not as time critical to purge our pure-F recipe with high-pressure oxygen ashing. At this stage the fabrication is essentially complete. To strip all the remaining organic compounds (PR/PMMA, organics formed during fabrication, etc.), we start with a 2-min. high-pressure ashing and interleave solvent cleaning, more ashing sessions, and sometimes strong PR-stripping agents like toluene, until we remove all the organic and polymer compounds from the surface of the wafer. If the design requires wafer dicing, we first inspect the final photolithography result after the cleaning, coat the instrumented surface(s) with an easily removable i-line PR for dust protection, dice, and then strip the PR with solvents. Finally, we transfer the completed device from JPL MDL H6 cleanroom to a non-cleanroom laboratory environment at Caltech for subsequent wire-bonding and characterization. We use a single-wafer polypropylene wafer carrier in a N_2-inflated sealed bag for transferring the device. We typically transfer the device and perform the wire-bonding and the first cryogenic characterization in a few days to a week after the fabrication. Fig. 4.4

shows the 80-KID Al-KID Nb-CPW detector fabricated with the technique detailed in this section, which we characterize in the following of this thesis.

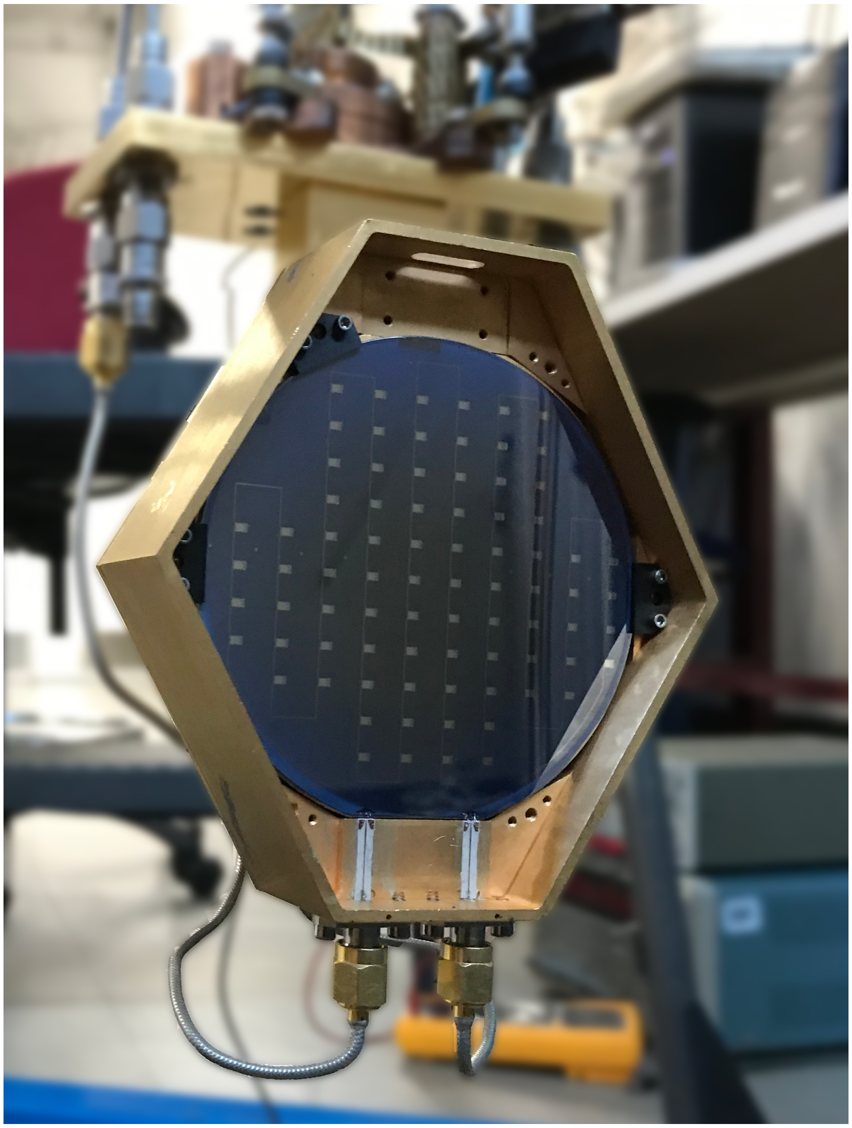


Figure 4.4: The 80-KID prototype detector with Al LE sCPS-KIDs, Nb CPW feedline and ground shields, on a $\varnothing 3" \times 1$ mm Si substrate. The two SMA cables connected from below provide the readout for the detector. The construction shown in the photo is the complete experimental configuration (lid removed for visibility), where the detector is housed by the gold-plated copper device holder (Moore, 2012) and mounted to the mixing chamber of an Oxford Kelvinox-25 DR.

4.2 Radio frequency performance

In the following of this chapter, we might occasionally refer the 80-KID prototype detector shown in Fig. 4.4 as the "YY device/detector," as it is internally dubbed. We apologize in advance if the reader finds such sloppy wordings.

4.2.1 Resonance data and modeling

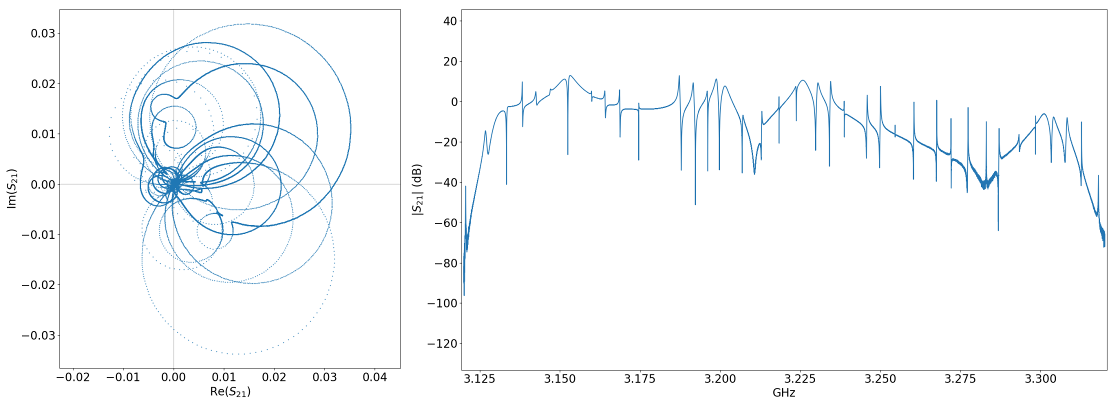


Figure 4.5: The complex in-phase/quadrature (IQ) (left) and the amplitude (right) S_{21} data for the 80-KID prototype detector measured at 60 mK without Eccosorb RF correction.

We acquired a broadband transmission (S_{21}) dataset for the 80-KID prototype detector using a HP 8720D vector network analyzer (VNA) to evaluate the detector's baseline radio-frequency performance; the data are shown in Fig. 4.5. We adjusted the VNA data range and density so that the dataset encompassed all the resonances and, for each resonance, a clearly profiled circular complex S_{21} trajectory as shown in Fig. 4.5. From these data, we are able to reliably extract the resonance parameters utilizing the fitting technique introduced in the next paragraph. We have explained in Ch. 3 that, due to the Eccosorb cooling difficulty, the data we present for the 80-KID Al-KID detector in this chapter were acquired without the Eccosorb long-range RF effect suppression. The resonances generally appear to be more distorted than the Nb-KIDs RF assessment device; we will specify the features that we regard as "distortions" later. Due to the RF distortion complexity, as well as the ongoing effort for the detector characterization by T. Aralis and K. Ramanathan, we adopt simulated results in certain discussions in the following for illustration purposes.

In order to obtain the VNA scan[4] at the highest signal-to-noise ratio (SNR), we

[4] To avoid confusion, we mean the unperturbed device transmission, as opposed to observing phonon pulses.

nominally read out the device at the highest possible power that does not observably alter the device's baseline RF response, which is a sign of generating a nonnegligible amount of extra QPs relative to other non-readout power sources. We always start the RF characterization procedure at a VNA readout power that is 8–15 dB lower than that causes one resonance to shift visibly. For the HP 8720D VNA we use, the smallest visible shift is set by the finest on-screen resolution of $\lesssim$10 Hz. For the prototype detector under discussion, we see such a frequency shift at -15– -12 dBm, so we take data at -25– -20 dBm, and we average the measurements for at least 10 times to suppress the measurement noise. For the RF characterization, our only principle criterion for an acceptable combination of the readout power, the density of data points, and the number of averaged S_{21} measurements is to permit a robust resonance fit for its parameter extraction without evidence of nonlinearity or QP generation.

After acquiring the VNA scan, we fit the resonance formula modified from (Gao, 2008)'s original formula to the complex S_{21} (IQ-plane) data to extract the characteristic parameters for each resonator. (Gao, 2008)'s resonance formula:

$$S_{21} = ae^{-i\omega t}\left[1 - \frac{\dfrac{Q_r}{Q_c}e^{i\phi_c}}{1 + 2iQ_r\dfrac{\Delta f}{f_r}}\right], \tag{4.1}$$

where $a = a(f)$ is a frequency-dependent real number describing the continuum (feedline) power attenuation, with a rotating phase delay term $e^{-i\omega t}$ for the finite time the readout signal travels from the transmitter to the receiver, the *real-number* Q_r and Q_c are the total and the coupling quality factors for the resonance, respectively, ϕ_c is an ad hoc phase factor whose meaning is discussed below, f_r is the resonant frequency, and we use Δf as shorthand for $(f - f_r)$. The quality factors follow the usual definition of inverse fractional energy loss per cycle, where Q_c is for the loss to the KID-feedline coupling mechanism, and Q_i is for the loss to all mechanisms internal to the resonator. We assume the total quality factor of the resonator follows

$$Q_r{}^{-1} = Q_c{}^{-1} + Q_i{}^{-1}, \tag{4.2}$$

i.e., dominated by Q_c and Q_i. ϕ_c is given the subscript of c based on (Gao, 2008)'s original argument that the ad hoc phase factor likely arises from a non-ideality in the coupling mechanism between the feedline and the KIDs. (Gao, 2008) adopted Eq. (4.2) to Eq. (4.1) while ignoring the potential dependence of Q_c and ϕ_c, which

Gao argued to both stem from the coupling mechanism but, as we suspect, utilized in practice only to phenomenologically stabilize data fitting. In Gao's original resonator network model, such a narrow on-resonance phase offset simply should not exist, but the data apparently required such a degree of freedom to be described. One may examine the effect of ϕ_c by setting $\phi_c = 0$ and find that Eq. (4.1) anticipates a strict attenuation to the readout signal when f approaches f_r ($\Delta f \to 0^{\pm}$). The result is effectively a "downward" Lorentzian shape in the $|S_{21}|$ spectrum, which is obviously incompatible with our data in Fig. 4.5. We present in Fig. 4.6 a more detailed Fig. 4.5 in 3.13–3.18 GHz, where the reader may find many of our KIDs exhibit partially downward (attenuating) and partially upward (amplifying) resonant shapes that indicate nonnegligible ϕ_c.

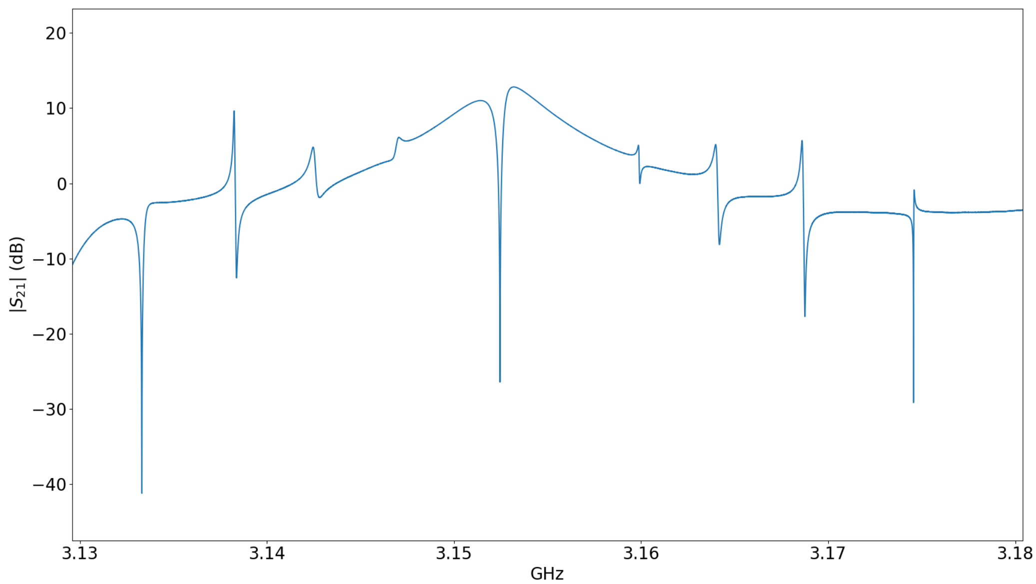


Figure 4.6: Detailed $|S_{21}|$ data in 3.13–3.18 GHz for the 80-KID prototype detector without Eccosorb RF correction.

As (Khalil et al., 2012) later suggested, if the phase term $e^{i\phi_c}$ is contributed by a nontrivial near-resonance input impedance mismatch in the feedline, in particular an equivalent capacitor as modeled in (Khalil et al., 2012), the appropriate energy conservation relation linking the intrinsic and the capacitive coupling losses should be

$$Q_r{}^{-1} = \left(\frac{|\hat{Q}_c|}{\cos\phi_c}\right)^{-1} + Q_i{}^{-1}, \tag{4.3}$$

where

$$\hat{Q}_c = |Q_c|e^{-i\phi_c}, \qquad \hat{Q}_c \in \mathbb{C}. \tag{4.4}$$

Comparing the above equations to Eq. (4.1) and Eq. (4.2), one realizes that (Khalil et al., 2012) effectively promotes the real-number Q_c in (Gao, 2008) to be a complex number that reduces to Eq. (4.2) at $\phi_c = 0$. (Khalil et al., 2012) argues that the applicability of such a modification reflects the fact that in reality the capacitive coupling may impose an impedance mismatch for the feedline at the resonant frequency, therefore the modification is only important near-resonance. Based on (Khalil et al., 2012)'s circuit model, one should consider, for the energy loss, not the total amplitude of the complex coupling but instead only its real part, hence the $\cos\phi_c$ term in Eq. (4.3). Consequentially, (Khalil et al., 2012) modifies Eq. (4.1) to

$$S_{21} = ae^{-i\omega t}\left[1 - \frac{\dfrac{Q_r}{|\hat{Q}_c|}e^{i\phi_c}}{1 + 2iQ_r\dfrac{\Delta f}{f_r}}\right], \tag{4.5}$$

which we adapt as our resonance model for the data fitting. It is worth noting that the (Khalil et al., 2012) model, so as the derived mathematical prescription, stems from a few assumptions, such as that the mismatched inline impedance is purely capacitive and should be intuitively much smaller than the intended coupling capacitance. Despite the fact that such an inline capacitance is to some extend a subjective choice for modeling and at the same time hard to predict/measure from first principle, currently the (Khalil et al., 2012) model is generally believed to be the closest description for practical high-quality SC resonators that are capacitively coupled to the readout lines. We are attempting to further generalize the (Khalil et al., 2012) model in the work to accommodate our data assuming a more general impedance mismatch than the extra capacitor of (Khalil et al., 2012).

4.2.2 Idealized S_{21} transformation

Fig. 4.7 presents a realization of Eq. (4.5) containing three resonances reconstructed based on the typical a, t, $\hat{Q}_c$, Q_r, and f_r extracted for the less distorted small-ϕ_c KIDs in the 80-KID prototype detector. We choose to adopt this simplified model in the following for illustrating the data processing procedure, the so-called *idealized S_{21} transformation*. As shown in the figure, the resonances usually span $O(10^{1-2})$ kHz, a result of having $Q_r \sim 10^5$ at the f_r of few GHz. Despite the data sparseness close to the resonant frequencies as marked the figure, with a realistic DAQ sampling rate of $O(10^{2-3})$ kHz, we are able to acquire sufficient data points for each resonance to allow a robust resonance fit and subsequently determine its RF parameters. Once the parameters are obtained, it is a common practice to remove the features that vary

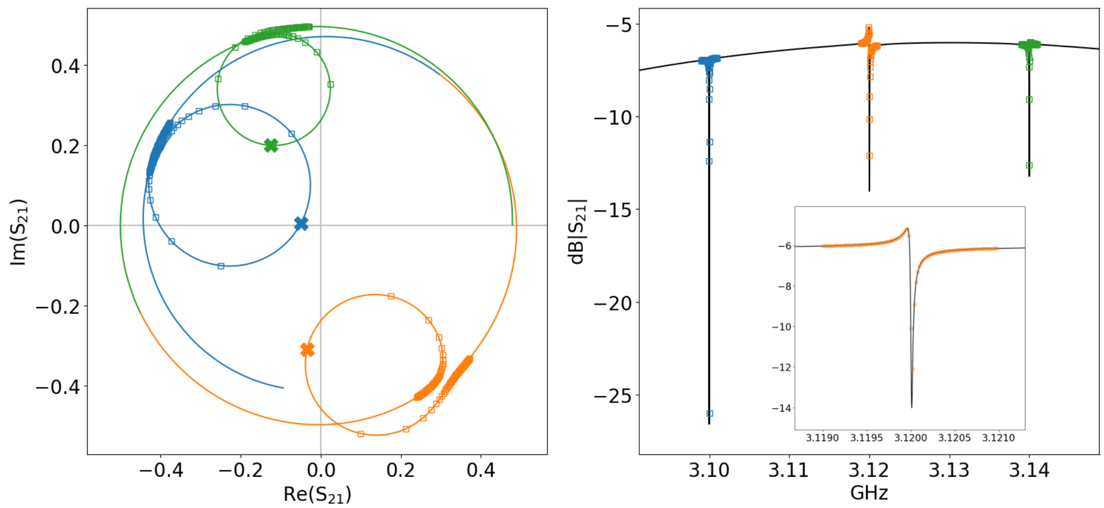


Figure 4.7: A simulated realization of Eq. (4.5) with 3 KID resonances, assuming typical parameter values extracted from the 80-KID prototype detector. Presented on the left and the right, respectively, are the IQ-plane and the amplitude S_{21} data. The $\times$ symbols in the left-hand side panel mark the on-resonance points for the resonances of the same colors, and the inset in the right-hand side panel is the zoom-in detail for the 3.12 GHz resonance. More detailed discussions may be found in the corresponding text.

KID-to-KID but are nevertheless irrelevant to the following resonator responsivity calibration, such as the feedline-related components. The treatment transfers the resonances to an idealized coordinate that provides an equal basis for the KIDs to be compared to each other. Fig. 4.8 illustrates step-by-step the transformation from the raw S_{21} to the "idealized" S_{21} by sequentially removing unnecessary features, which one may regard as "reverse-engineering" Eq. (4.5) term-by-term, eventually bringing all the resonance circles onto the positive real axis and crossing $(1, 0)$.

We begin the procedure by first removing the frequency-dependent continuum, i.e., the feedline transmission variation. We divide Eq. (4.5) and the complex S_{21} data by $a(f)$, so the baseline S_{21} for all the resonance circles falls on the unit circle, which represents an ideal feedline. In practice, since we process each resonance that occupies its corresponding narrow frequency range independently, we approximate the feedline attenuation by

$$\log a(f) = c_0 + c_1 f, \qquad c_0,\, c_1 \in \mathbb{R}, \tag{4.6}$$

where c_0 and c_1 are the fitted coefficients describing a straight-line continuum around the resonance in the conventional log-linear $20 \log_{10} |S_{21}(f)|$ plot. The criterion for

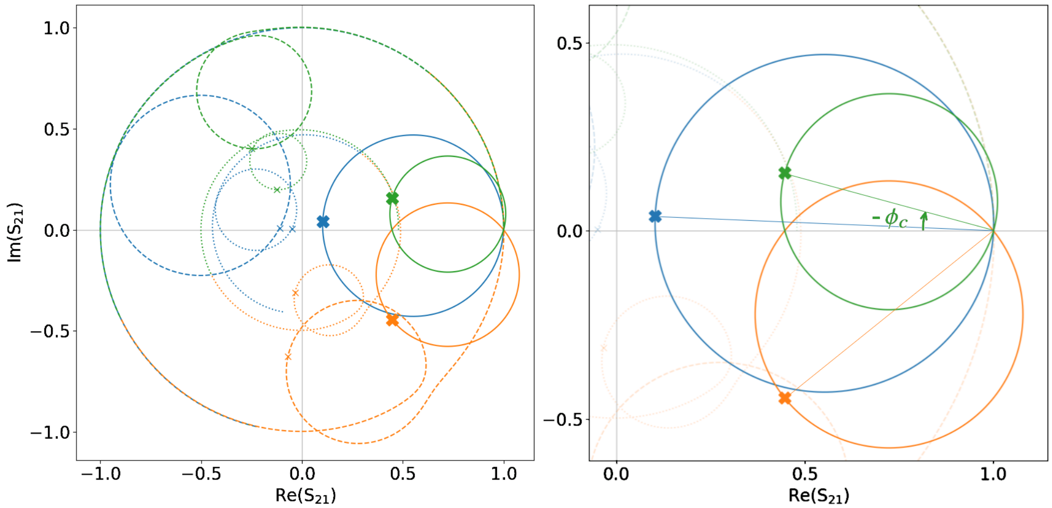


Figure 4.8: A step-by-step illustration for the idealized S_{21} transformation. The left-hand side panel shows the sequential modification, and the right-hand side panel is its zoom-in final result. On the left, the raw (Fig. 4.7), the feedline attenuation-removed, and the feedline phase delay-removed results are represented by dotted, dashed, and solid curves, respectively. The $\times$ symbols mark the resonance points for the curves/resonances of the same colors. On the right, the fully transformed results from the left-hand side panel are shown. The ϕ_c differences are denoted by the $(1,0)$–$\times$ line with respect to the x-axis, with other intermediate transformation results blurred into the background for better clarity. A detailed description is provided in the corresponding text.

the narrow frequency range over which we perform the data fitting is

$$\frac{|da|}{df} \ll \frac{|S_{21}(f_r)|}{f_r/Q_r}, \tag{4.7}$$

i.e., indicating the continuum varies much smaller than the size of the resonance. After correcting for the feedline attenuation, the radii of the resonance circles are subject only to the quality factors, which are uniquely associated to the resonators and, as we will show later, directly related to the resonators' responsivities. One may see that, by examining $f \to \infty$ and $f = f_r$ with the fact that $Q_r \sim O(10^5) \gg 1$, Eq. (4.5) suggests that the radius r for a resonance circle is (after removing $a(f)$)

$$r \to \frac{Q_r}{2|\hat{Q}_c|}, \qquad Q_r \gg 1. \tag{4.8}$$

Next, for better understanding the resonance properties and comparison, we rotate these continuum-removed resonance circles to the same location in the IQ plane, so that their off-resonance transmissions all sit at $(1,0)$. We eliminate the global phase rotation due to cable delay by further dividing the formula and data by $e^{-i\omega t}$. With a bit of rearrangement, the resonance model is now left with

$$S_{21} = 1 - \left(\frac{\frac{Q_r}{|Q_c|}}{1 + 2iQ_r \frac{\Delta f}{f_r}} \right) e^{i\phi_c}. \tag{4.9}$$

By setting $f \to \pm\infty$, i.e., far off-resonance, we find that all the circles described by Eq. (4.9) coincide at (1,0$^{\mp}$) as shown in Fig. 4.8, which is well-anticipated by the perfect off-resonance transmission now that the feedline attenuation is removed. For an easier explanation, we rewrite Eq. (4.9) using the frequency-dependent phase variable

$$e^{i\xi(f)} = \frac{1 - 2iQ_r \frac{\Delta f}{f_r}}{1 + 2iQ_r \frac{\Delta f}{f_r}} \tag{4.10}$$

and obtain (algebra in Appx. A)

$$S_{21} = 1 + \frac{Q_r}{2|\hat{Q}_c|} \left[1 + e^{i\xi(f)} \right] e^{i(\phi_c + \pi)}, \tag{4.11}$$

which shows that the idealized S_{21} is indeed a circle with a radius of

$$r = \frac{Q_r}{2|\hat{Q}_c|}. \tag{4.12}$$

Fig. 4.9 provides a step-by-step transformation picture to explain Eq. (4.11), where one starts with the

$$\frac{Q_r}{2|\hat{Q}_c|} \left[1 + e^{i\xi(f)} \right] \tag{4.13}$$

term in Eq. (4.11), which is a circle centered at $(0, r)$, and then the $e^{i(\phi_c + \pi)}$ term rotates the circle by $(\phi_c + \pi)$ around the origin, and finally the $+1$ term offsets the circle to crossing $(1, 0)$. The result, as shown by Fig. 4.8/4.9, is a circle with the radius r (Eq. (4.12)) that crosses (1,0) at far off-resonance, and the angle between the x-axis and the line connecting (1,0) and the transformed on-resonance point ($S_{21}(f_r)$) is equal to $-\phi_c$; Fig. 4.8 also illustrates the relation of the sign of ϕ_c and the direction for the rotation.

4.2.3 Resonance fitting technique

The above analysis for the idealized S_{21} transformation is not just a convention for standardizing resonance presentation, but it actually provides useful insights that

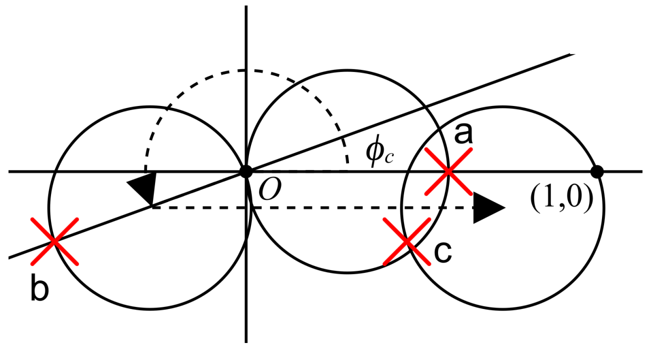


Figure 4.9: Following the arrow-dashed line, this figure illustrates the construction of Eq. (4.9) in the form of Eq. (4.11), starting from a) Eq. (4.13), and then b) rotating by $(\phi_c + \pi)$, and finally c) offsetting by $(+1, 0)$ to the idealized S_{21} location. The red $\times$ symbols mark the resonant frequency ($\xi = 0$), and the dashed line represents the trajectory of the center of the circle.

help with understanding the nature of the resonance data as well as the practical RF response analysis. Likewise, while it may appear somewhat technical, we hope that the following discussion for our resonance fitting techniques may also turn out to be pedagogical for readers who are not yet familiar with a RF (IQ plane)-style data analysis. Fig. 4.10 presents an example of our KID-resonance fitting result, demonstrating that we can reliably extract the RF parameters in Eq. (4.5) using the methodology introduced below. Readers who prefer to directly adopt our fitting scripts may choose to skip the following discussion and proceed to the next section.

So far we have been assuming that one may apply Eq. (4.5) as it is and straightforwardly obtain the parameters needed by the S_{21} idealization. On the contrary, one may also see that Eq. (4.5) requires simultaneously constraining six parameters in a highly nonlinear form, which is strictly divergent if not for the finite but still large Q_r. If the initial values for a, t, $|\hat{Q}_c|$, Q_r, ϕ_c, f_r are not chosen sufficiently close to their final values, common iterative fitting algorithms regularly do not converge. We therefore develop a procedure to determine these initial values with readily available information from the VNA scan. Among the parameters, the cable delay time t is the simplest to predict: By knowing the total length L_{cable} for the transmission line from the mixer at the transmitter side in the DAQ, through the entire system, and back to the mixer for the receiver, we estimate the cable delay using the group velocity for typical commercial coaxial cables,

$$t \approx \frac{L_{\text{cable}}}{v_{\text{cable}}} \sim \mathcal{O}(10)\ \text{ns}, \tag{4.14}$$

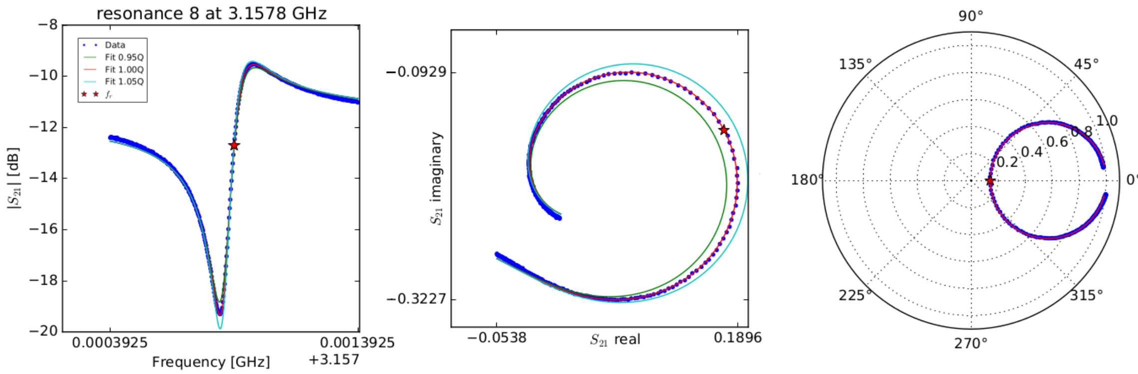


Figure 4.10: An example of the Eq. (4.5) S_{21} data fitting result. In the left-, middle, and right-hand side panels, the figures are the $|S_{21}(f)|$, raw IQ-plane $S_{21}(f)$, and the idealized $S_{21}(f)$ described by Eq. (4.9), respectively. The data, the best fit, the extracted resonant point, and the ±5% uncertainty boundaries are represented by blue dots, red curves, red stars, and different green curves (see legend). For this particular example, we obtain $f_r = 3157.89$ MHz, $Q_r = 2.34{\times}10^4$, $|\hat{Q}_c| = 2.72{\times}10^4$, $Q_i = 1.65 \times 10^5$, $\phi_c = -0.98°$, and $t = 53.35$ ns. Figure reproduced from a new version of our fitting script with uncertainty estimates by T. Aralis.

where

$$v_{\text{cable}} \approx 0.7c, \tag{4.15}$$

and c is the vacuum speed of light. Depending on the exact construction, we typically utilize a measurable total length of 5–10 m for the SMA cables, which dominates over the less well-measured transmission line lengths in the DAQ system and on the detector, so it is easy to provide an initial value that is accurate to a few percent.

For other parameters, it is most effective to start with the 1-D $|S_{21}(f)|$ spectrum for a quick yet sufficiently accurate estimate; the following discussion would be much easier to follow while cross referencing Fig. 4.11. We first estimate the transmission line attenuation function $a(f)$ by connecting the furthest off-resonance points with a straight line as shown by the blue dashed line in Fig. 4.11. We parameterize the straight line by

$$10 \log_{10} |S_{21}(f)| = c'_0 + c'_1 f, \tag{4.16}$$

where c'_i are determined analytically from data and have

$$c_i = \frac{c'_i \log 10}{20} \approx 0.12 c'_i, \quad i = 0,\ 1 \tag{4.17}$$

for Eq. (4.6). Note that Eq. (4.17) assumes that the S_{21} data is in units of amplitude but not power, so the denominator is 20, not 10.

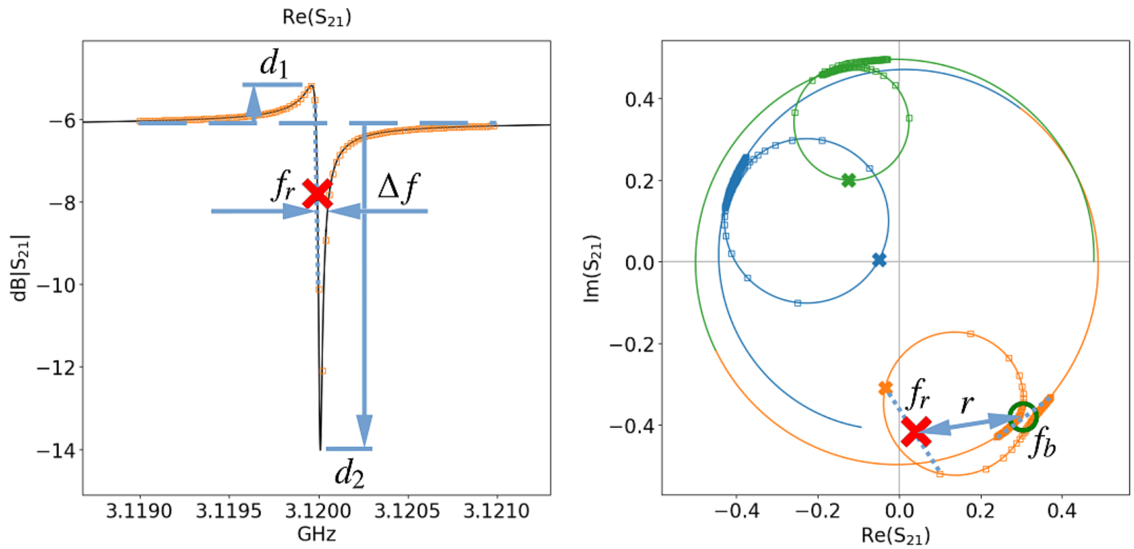


Figure 4.11: An illustration of the initial value estimation for the resonance fit. Left: Utilizing $|S_{21}(f)|$ (amplitude) information for estimating $a(f)$, ϕ_c, and Q_r; right: Utilizing complex S_{21} (IQ-plane) information for estimating f_r, r, and subsequently $|\hat{Q}_c|$. Due to the complexity of the procedure, readers are advised to refer to the corresponding text for an elaboration.

For estimating ϕ_c, we use the ratio of the distances d_1 and d_2 in Fig. 4.11 from the approximated $a(f)$ straight line to the highest and the lowest data points above and below the line, respectively. Notice that, unlike estimating $a(f)$, d_1 and d_2 are measured in linear units, not in dB. If we compare Fig. 4.11 to Fig. 4.8, we see that d_1 and d_2 are essentially the maximal deviations in the IQ plane from the continuum circle due to the resonance circling, which are manifested by the overshooting d_1 and the recessing d_2 in the $|S_{21}(f)|$ spectrum. It is also easy to understand that the overshooting and the recessing correspond to the "outside" and the "inside" of the IQ-plane continuum circle, respectively. The proportionality of the resonance circle falling inside and outside of the continuum, or equivalently the $x < 1$ and $x > 1$ extensions for the resonance circle in the idealized IQ plane (Fig. 4.8), respectively, may be estimated by

$$\tan \frac{|\phi_c|}{2} \approx \frac{d_1}{d_2}. \tag{4.18}$$

Eq. (4.18) is not exact in general but yields satisfying estimates close to $|\phi_c| = 0°$, 90°, 180° and becomes exact at these angles. One may check $|\phi_c| = 0°, 90°, 180°$ and realize that the angles correspond to when the resonance circle is fully inside, half inside and half outside, and fully outside of $0 < x < 1$ in the idealized plane, respectively. This check should also demonstrate that Eq. (4.18) is not exact mainly because of the curvature of the circle.

We need to determine the sign of ϕ_c after Eq. (4.18). To start, we first remind the reader, to trace along the IQ plane trajectory from low to high frequencies, the $S_{21}(f)$ trajectory always rotates clockwise due to the minus sign for the exponent of $e^{-i\omega t}$ in Eq. (4.5). We compare Fig. 4.11 to Fig. 4.9 and realize that, for a resonance circle rotating counterclockwise with respect to (1,0) in the idealized plane, i.e., a positive ϕ_c, the extension of the resonance circle outside the unit circle occurs for $f < f_r$. Correspondingly, the $|S_{21}|$ spectrum for a positive ϕ_c should exceed the continuum transmission for $f < f_r$. For example, in Fig. 4.11, d_1 is at a lower frequency relative to d_2, therefore in this case ϕ_c is positive, and vice versa for a negative ϕ_c.

We proceed from the "placement/rotation" parameters $a(f)$ and ϕ_c to the resonance's "shape" parameters–the quality factors Q_r and $|\hat{Q}_c|$. To estimate the quality factors, we take advantage of the geometric analysis presented earlier, in particular Eq. (4.12), which relates the radius of the resonance circle to its quality factors. To use Eq. (4.12), obviously we need to obtain an estimate for the radius, so we shift from analyzing the 1-D $|S_{21}|$ spectrum to the 2-D IQ-plane trajectory. We look for a practical algorithm that identifies the approximate location and boundary of the resonance circle and subsequently determines the radius. According to the characteristic resonance-circle variable in Eq. (4.10), if we sample the frequency space evenly for S_{21}, which is the most reasonable approach without prior knowledge of the device, the resonance circle will be profiled the densest per central angle ξ when $|\Delta f| \gg f_r$ and the sparsest when $|\Delta f| \to 0$; the phenomenon is clearly shown in Fig. 4.11. Inspired by this variable spacing, we estimate the resonant frequency by the mean of the frequencies for the two data points that are separated the furthest *in the IQ plane*. Such a data pair and its corresponding resonant frequency estimate are shown in Fig 4.11 by the blue dotted line and the red $\times$ symbol, respective.

While the above prescription may appear to be identically applicable for the $|S_{21}|$ spectrum, e.g., extracting the averaged f_r from the two furthest data points in $|S_{21}(f_r)|$, we note that the furthest data pair in the IQ plane is not necessarily the closest to the peak in $|S_{21}(f_r)|$ due to the variable ϕ_c. On the other hand, Eq. (4.11) proves that, with a denser sampling, our algorithm utilizing the IQ-plane information monotonically approaches the true $S_{21}(f_r)$ regardless of ϕ_c. Even in the worst case where $S_{21}(f_r)$ happens to coincide with a sampled data point under sparse sampling, which we purposely choose to be the case for Fig. 4.11, the estimated resonant point still does not deviate from the true value by a fraction of the size of the resonance circle. We also highlight the corresponding data pair in $|S_{21}|$ for the pair identified in the IQ plane, which the reader may see is not the closest pair to the peak(s) of the deformed Lorentzian shape nor the steepest or the flattest. We apply a similar method to estimate the "base" point, where the continuum and the resonance circles coincide. We estimate the base point by the average of the first and the last data points in the data series, marked by the green circle in Fig. 4.11. In principle, this estimate should be at the opposite location of the resonance point with respect to the center of the resonance circle. Due to the compression of the off-resonance data points, as long as the sampled frequency range reasonably encompasses the resonance circle, this estimate becomes insensitive to the exact frequencies of the first and the last data points and therefore provides a stable approximation for the base point. With both the resonance point and the base point determined, we straightforwardly infer from their separation the diameter ($2r$) of the circle, which is equal to $Q_r/|\hat{Q}_c|$ per Eg. (4.12). Fig. 4.11 demonstrates that our algorithm for estimating r is typically accurate to within a factor of 2, which is sufficiently accurate for initializing the fit.

So far we have explained our estimation for the initial values of $a(f)$, t, ϕ_c, f_r, and $Q_r/|\hat{Q}_c|$. Owing to the good agreement between the design simulation and the fabricated devices, it is usually sufficient to initialize Q_r by the designed value (10^5 for the prototype detector) with $|\hat{Q}_c| = Q_r/2r$. However, it is sometimes beneficial to manually measure the total quality factor Q_r to initialize both Q_r and $|\hat{Q}_c|$ with more constrained values, especially for characterizing previously untested designs. The conventional method for estimating the quality factor for a generic resonance is by

$$Q_r \approx \frac{f_r}{\Delta f_{-3\mathrm{dB}}}, \tag{4.19}$$

where $\Delta f_{-3\mathrm{dB}}$ is the width of the resonance, defined by the frequency span of the Lorentzian shape at 3 dB below the off-resonance value. In fact, one may show that Eq. (4.19) is exact when there is no resonance circle rotation ($\phi_c = 0$). In other words, the definition does not include the impedance mismatch that gives rise to the coupling phase ϕ_c in the (Khalil et al., 2012) model hence does not apply to resonances with non-trivial ϕ_c. Now that we understand ϕ_c manifests the local rotation of the resonance circle with respect to its base point, we follow the same concept for extracting ϕ_c with d_1 and d_2 (Eq. (4.18)) to modify Eq. (4.19) to

$$Q \approx \frac{f_r}{\Delta f_{\delta\mathrm{dB}}}, \tag{4.20}$$

where

$$\delta = \mathrm{dB}[1 - 10^{-3/10} \cos \phi_c]. \tag{4.21}$$

Eq. (4.21) modifies the 3-dB criterion by the rotation of the resonance circle toward the outside of the continuum circle, where the location and the frequency width for δ dB attenuation are marked by the horizontal arrows and the Δf in the $|S_{21}|$ panel of Fig. 4.11. Similarly to the estimation for ϕ_c, Eq. (4.21) takes the linear unit but not dB, because it is an effect in the linear IQ plane. Based on the definition of Eq. (4.21), Fig. 4.11, which exhibits $\phi_c \approx 39°$, is estimated with $\delta \approx 2.1$ dB but not 3 dB. The result agrees with our expectation for adjusting the 3 dB criterion and therefore helps with the fit. Note that Eq. (4.21) also yields positive δ when $\phi_c > \pi/2$, which corresponds to defining Δf *above* the off-resonance continuum. This result is also expected based on rotating the majority of the resonance circle to the outside of the continuum, which leads to a near-resonance transmission that is larger than $a(f)$ as if the readout signal is amplified.

Finally, we introduce the parameter-constraining sequence that we find the most effective based on trial and error experience. We find it is important to temporarily hold fixed some of the parameters, the ones with relatively reliable values, while allowing the fitter to vary the other parameters. After the floating parameter subset converges, we then swap the roles and fix these fitted values and vary the previously fixed parameters. The idea behind this approach is to avoid confusing the fitting algorithm with too many free parameters in the nonlinear model, which we find as critical as providing good initial values. We carry out the procedure multiple times, iteratively exchanging the fixed/floating parameter subsets, with more parameters relaxed into the floating group in succession while the fitter approaches the optimal result. Based on the same principle, we also find that it is more efficient to start by fitting the 1-D $|S_{21}|$ spectrum, effectively ignoring the phase degree of freedom, to establish a rough Lorentzian/resonance circle shape, including f_r, ϕ_c, Q_r, $|\hat{Q}_c|$. We then fully constrain the resonance circle shape (f_r, ϕ_c, Q_r, $|\hat{Q}_c|$) and fit for *only* the cable delay t in the IQ plane, practically rotating the fixed resonance circle shape to the closet location of the data. Hereafter we start relaxing parameters into the joint fit. Whether it is for the $|S_{21}|$ or the IQ-plane fit, we find that it is typically beneficial to group f_r and ϕ_c together, as they jointly determine the location and the above/below-continuum shape of the spectrum; Q_r and $|\hat{Q}_c|$ together, because they are responsible for the size of the resonance circle; ϕ_c alone, since it independently determines the rotation of the resonance circle as explained previously; and lastly we usually hold $a(f)$ fixed until every other parameter is sufficiently constrained, since the initial estimate (Eq. (4.17)) is generally quite accurate. It is also worth noting that, since the data points are much sparser near the resonance, it is often useful to let the fitting algorithm minimize in logarithmic (dB) scale first, so the model-data deviation near the resonance is emphasized relative to the numerous but less constraining off-resonance data points.

4.3 KID responsivity calibration

4.3.1 QP-generated signal

Recalling our introduction in Sec. 3.1, KID detects the QP density fluctuation Δn_{qp} in the SC film through observing its (kinetic) inductance change. If the energy source and its corresponding coupling mechanism causing Δn_{qp} are also known, one may then use the observed inductance change to deduce the energy deposited into the QP system and in turn the external energy emission. In practice, one observes the resonance modulation as the consequence of the inductance change, which (Gao, 2008) derived based on the Mattis-Bardeen theory to be given by

$$dS_{21} = \alpha|\gamma|\kappa\frac{Q_r^2}{Q_c}dn_{\mathrm{qp}}. \tag{4.22}$$

Eq. (4.22) relates the readout signal modulation dS_{21} to the QP density fluctuation dn_{qp}, where α is the phenomenological kinetic inductance fraction defined by Eq. (3.20), γ is a variable representing, effectively, the “sensitive” fraction of the film thickness due the film’s effective RF penetration depth, and κ is the main Mattis-Bardeen-based function encoding the dependence on physical properties including the film’s superconducting electronic density of states, its Cooper pair binding energy, the film temperature, and the readout signal frequency or equivalently the KID’s resonant frequency. Starting from this section, we will also use

$$Q_c := \frac{|\hat{Q}_c|}{\cos\phi_c} \tag{4.23}$$

to simplify the writing based on the discussion surrounding Eq. (4.3).

We begin introducing Eq. (4.22) by briefly recalling from Sec. 3.6 that α is the kinetic inductance fraction, defined by the fractional inductance of a KID that is contributed by the SC material’s kinetic inductance. Since kinetic inductance varies upon QP density fluctuations, this part of the total inductance is responsible for the detection, therefore α is directly associated to the responsivity of the KID. Depending on the exact inductor design, the kinetic inductance fraction usually resides in the range of 5–15% for sputtered Al films that are much thinner than their EM penetration depths. For γ, since it is a variable that depends on the actual relative values of the film thickness, its coherence length, and its effective EM penetration depth, one must perform the Mattis-Bardeen integrals numerically to determine γ. However, as in our case and also most reasonably designed modern KIDs, (Gao, 2008) showed that the film impedance Z_{eff} scales as

$$Z_{\text{eff}} \propto \lambda_{\text{eff}} \propto \frac{1}{d^2} \tag{4.24}$$

for thin films, whose film thicknesses d are much smaller than both the effective EM penetration depth λ_{eff} and the electron/QP mean free path ξ_{eff}. Contrasting thin films, thick films have

$$Z_{\text{eff}} \longrightarrow Z_{\text{eff}}(d = \lambda_{\text{eff}}). \tag{4.25}$$

Therefore, one reasonably aims to design a KID that exhibits a large film impedance, hence a large KID responsivity, utilizing the inverse quadratic enhancement in film thickness in the thin-film regime. Concerning γ, such a choice effectively allows the Mattis-Bardeen integrals to neglect the variation of the penetrating field, which leads to

$$\gamma \longrightarrow -1, \qquad d \ll \lambda_{\text{eff}} \ll \xi_{\text{eff}}, \tag{4.26}$$

i.e., a negligible factor in Eq. (4.22) for onward discussions.

κ encodes all the essence for practical KID applications. Assuming the Mattis-Bardeen theory for a general laminated structure that is thermalized to a temperature T much below T_c and is irradiated ("read out") by photons with the energy $\hbar\omega$ much below the film's QP bandgap energy Δ and its thermal excitation scale $k_{\text{B}}T$, i.e.,

$$\hbar\omega \ll k_{\text{B}}T \ll \Delta \lesssim k_{\text{B}}T_c, \tag{4.27}$$

where $\hbar$ and k_{B} are the Planck and Boltzmann constants, respectively, (Gao, 2008) rigorously derived

$$\kappa \coloneqq \kappa_1 + i\kappa_2, \qquad \kappa_1, \kappa_2 \in \mathbb{R}, \tag{4.28}$$

$$\begin{cases} \kappa_1 = \dfrac{1}{\pi N_0}\sqrt{\dfrac{2}{\pi k_{\text{B}}T\Delta_0}}\sinh(\xi)K_0(\xi) \\ \kappa_2 = \dfrac{1}{2N_0\Delta_0}\left(1 + \sqrt{\dfrac{2\Delta_0}{\pi k_{\text{B}}T}}e^{-\xi}I_0(\xi)\right) \end{cases}, \tag{4.29}$$

where N_0 is the electron density of states, Δ_0 is Δ at $T = 0$ K, I_0 and K_0 are the 0^{th}-order modified Bessel functions of the first and the second kinds, respectively, and the equations are represented by the characteristic energy scale

$$\xi = \frac{\hbar\omega}{2k_{\text{B}}T}. \tag{4.30}$$

For a well-controlled operation, we may reasonably expect that the manually injected readout signal dominates the photon irradiation, which monotonically targets the resonant frequency and therefore

$$\omega \approx 2\pi f_r. \tag{4.31}$$

Limited by the scope of this thesis as well as the completeness of Gao's work, we refer interested readers to (Gao, 2008) for further details of the above results.

We have prepared the models for the unperturbed baseline RF response and the dn_{qp}-perturbed response, namely Eq. (4.5) and Eq. (4.22), respectively; we now combine the two to be a full signal model for the detection and also the responsivity calibration. To begin, we first assume a general energy conservation relation

$$E = \delta n_{\text{qp}} V \Delta \approx \delta n_{\text{qp}} V \Delta_0, \tag{4.32}$$

where E is the instantaneous energy deposition, δn_{qp} is the corresponding QP density fluctuation, and V is the volume that accepts the energy and thus relates δn_{qp} to the total change of the number of QPs. Eq. (4.32) is a valid approximation when the time scale in discussion is much shorter then the QP recombination time (discussed later), E is small enough to not cause a significant effective QP system temperature change, and again $T \ll T_c$. We will use Eq. (4.32) in the following derivations whenever these conditions are satisfied, e.g., modeling the small-signal response. In order to connect Eq. (4.5) to Eq. (4.22), we need to convert Eq. (4.5) into a differential from for the transient response. Based on the analysis in Sec. 4.2, an apparent parametrization that fully characterizes a resonance is with the resonant frequency f_r and the quality factor Q_r, which describe the *location* and the *width* of the Lorentzian shape in $|S_{21}(f)|$, respectively. We have shown in Eq. (4.12) that the radius of the resonance circle is fully determined by the quality factors but independent of f_r. Eq. (4.10) and Eq. (4.11) together also show that f_r determines the circular angle variable ξ but does not participate in any other parts of Eq. (4.11). These observations combined suggest that an infinitesimal dS_{21} displacement caused by df_r is tangential to the resonance circle in the IQ plane; the argument is illustrated

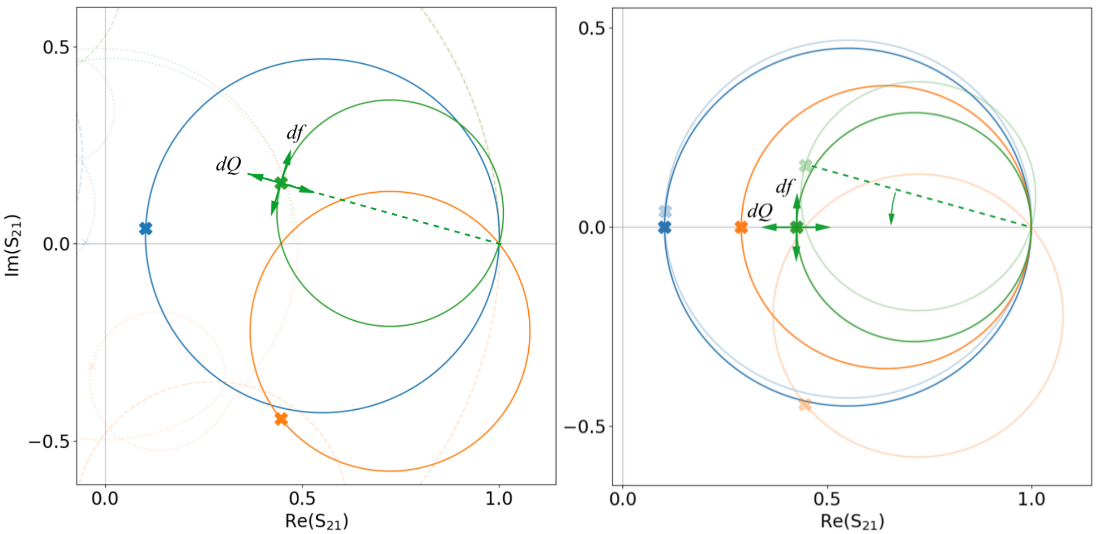


Figure 4.12: Left: The idealized S_{21} resonance circles identical to Fig. 4.8 (data points suppressed for clarity); right: Rotating the circles in the left-hand side panel by ϕ_c, so the dQ and the df directions align to the complex plane axes. Following the discussion in the corresponding text, when the radius of a resonance circle is perturbed due to a quality factor perturbation, the circle reduces its size toward $(1,0)$ along the dashed line connecting the resonance point and $(1,0)$. The fixed anchor point at $(1,0)$ is always required by Eq. (4.11), so the direction denoted by the dQ arrow always indicates a pure quality factor perturbation. On the other hand, if the angle parameter ξ is perturbed due to a resonance frequency perturbation, the S_{21} point should rotate along the resonance circle thus, for an infinitesimal perturbation, along the df direction as denoted. The right-hand side panel shows that, if ϕ_c is removed or exhibits a negligible value, the dQ and the df directions align to the real and imaginary axes of the complex IQ plane, respectively. The naming for the dQ and the df directions is due to the radius and ξ's exclusive dependences in $Q_{r(i)}$ and f_r, respectively. A more detailed explanation is found in the corresponding text.

in Fig. 4.12. We are therefore motivated to compute

$$dS_{21}(f_r, Q_i) = \frac{\partial S_{21}}{\partial(1/Q_i)} d\left(\frac{1}{Q_i}\right) + \frac{\partial S_{21}}{\partial(\Delta f/f_r)} d\left(\frac{\Delta f}{f_r}\right), \tag{4.33}$$

i.e., the infinitesimal displacement of $S_{21}(f)$ due to an infinitesimal deformation of the resonance circle as a function of the independent variables f_r and Q_r. Notice that, instead of directly perturbing f_r, we consider the normalized $\Delta f/f_r$ so to be consistent with Eq. (4.5), while the physical meaning is unchanged but with a conversion factor that can be obtained via chain rule. We also make the assumption/constraint that, within a reasonable limit of energy deposition that we model our detector, only the internal quality factor component in the total quality factor is changed, so we write the equation in terms of Q_i but not Q_r and again relate the two with chain rule.

Since it is the inverse of a quality factor that represents the dissipated energy, we consider $1/Q_i$ instead of Q_i. Similar to the choice of $\Delta f/f_r$, the difference between Q_r and $1/Q_i$ is, providing the assumptions, merely mathematical and convertible via chain rule and Eq. (4.3).

Before we substitute the analytical formula of S_{21} to compute Eq. (4.33), we note that Eq. (4.22) was derived considering only the resonator itself (Gao, 2008). It means that the coordinate system defining the real and the imaginary axes, i.e., κ_1 and κ_2 in Eq. (4.29), is the resonance circle's internal system, in which no feedline or coupling effects are included. So to make the IQ plane for the idealized S_{21} and Eq. (4.22)/(4.29) comparable, we need to further remove the ϕ_c rotation that misaligns the two coordinates as illustrated in Fig. 4.12. We modify Eq. (4.9) by

$$S_{21,\text{ideal}} \longrightarrow S_{21,\text{internal}} = [(S_{21,\text{ideal}} - 1)e^{-i\phi_c}] + 1 \tag{4.34}$$

$$= 1 - \frac{Q_r/Q_c}{1 + 2iQ_r(\Delta f/f_r)}, \tag{4.35}$$

i.e., rotate the circle by ϕ_c with respect to (1,0) as denoted in Fig. 4.12, so the *resonance point–(1,0)* connection (dashed line in Fig. 4.12) aligns to the $\mathbb{R}^+$ axis. With this treatment, the signal due to a resonant frequency perturbation or a quality factor perturbation is now purely imaginary or real, respectively. We then substitute Eq. (4.35) into Eq. (4.33) and arrive at (algebra in Appx. A)

$$\lim_{f\to f_r} dS_{21} = -\frac{Q_r^2}{Q_c}\left[d\left(\frac{1}{Q_i}\right) - 2id\left(\frac{\Delta f}{f_r}\right)\right]. \tag{4.36}$$

Since we have fully transformed the data and the corresponding RF response model into a comparable basis of the resonator's internal impedance perturbation, we simply compare Eq. (4.36) to Eq. (4.22), independently for the real (Q_i, "radial") and the imaginary (f_r, "tangential") parts of dS_{21}, also using Eq. (4.28), and obtain

$$\begin{cases} dn_{\text{qp},f_r} = \dfrac{-1}{\alpha\kappa_1} d\left(\dfrac{1}{Q_i}\right) \\[2ex] dn_{\text{qp},Q_i} = \dfrac{2}{\alpha\kappa_2} d\left(\dfrac{\Delta f}{f_r}\right). \end{cases} \tag{4.37}$$

Note that, for the above result, we also assume $|\gamma| = 1$ for the thin-film limit (Eq. (4.26)) and use the subscripts $n_{\text{qp},\underline{\boldsymbol{f_r}}}$ and $n_{\text{qp},\underline{\boldsymbol{Q_i}}}$ to denote whether the dn_{qp} signal is reconstructed from the f_r or the Q_i data.

4.3.2 Small-signal pulse readout

Despite the fact that Eq. (4.37) is in principle calculable and physically meaningful, the result is written as functions of perturbations in Q_i and f_r, which we have elaborated extensively that one needs to undertake a substantial data taking and fitting to acquire their values. Even more impractical, the RF parameter extraction technique developed in Sec. 4.2 only provides steady-state Q_i and f_r, which we have the leisure to sweep through the frequencies for profiling the full resonance and extracting the parameters. Contradicting the slow process, we anticipate the deformation of the resonance circle due to phonon pulses to occur instantaneously relative to the slow Q_i-&-f_r data acquisition, hence preventing a direct adoption of Eq. (4.37) to reconstruct dn_{qp} in real-time. We will rigorously justify the pulse instantaneity statement later when we introduce the temporal response of the detector. For now, one may simply consider the fact that SuperCDMS detectors usually obtain pulses at sub-msec. time scales, c.f. Ch. 2, while contrasting the short time scale is that it takes minutes for our pipeline to scan a resonance and subsequently extract the parameters.

We accept the fact that it is impractical to constantly scan through the entire resonance for tracking Q_i and f_r in real-time, while the circle is also shifting at a much faster speed, so we choose the second best option—the small-signal operation/approximation. In this case, instead of attempting to acquire data for the entire circle, we monitor just one "read-out" frequency f_{read} and its corresponding $S_{21}(f_{\mathrm{read}})$. Under the assumption that δQ_i^{-1} and $\delta(\Delta f/f_r)$ thus their corresponding δn_{qp} and the energy deposition in general are all small enough so that

$$\frac{\delta(\partial_x S_{21})}{\delta x} \gg \left.\frac{\partial^2 S_{21}}{\partial x^2}\right|_{f\approx f_{\mathrm{read}}}, \; x = \frac{1}{Q_i} \text{ or } \frac{\Delta f}{f_r}, \tag{4.38}$$

i.e., the x dependence of $\partial_x S_{21}$ within the small range of x variation is negligible, we may always take

$$\partial_x S_{21} \approx \partial_x S_{21}(f_{\mathrm{read}}). \tag{4.39}$$

The above prescription linearizes S_{21} within the small perturbation range associated to the small δn_{qp}. We further consider that, for an optimal responsivity, one typically chooses

$$f_{\mathrm{read}} \approx f_r, \tag{4.40}$$

therefore Eq. (4.36) becomes

$$\begin{cases} \mathbf{Re}(\delta S_{21}) \approx -\left.\dfrac{Q_r^2}{Q_c}\right|_{f_r} \delta\left(\dfrac{1}{Q_i}\right) \\ \mathbf{Im}(\delta S_{21}) \approx 2\left.\dfrac{Q_r^2}{Q_c}\right|_{f_r} \delta\left(\dfrac{\Delta f}{f_r}\right). \end{cases} \tag{4.41}$$

We then substitute Eq. (4.41) into Eq. (4.37) and finally obtain our **master equation for calibrating the RF signal to the pair-breaking pulse energy**:

$$\begin{cases} E_{f_r} = V\Delta_0 \cdot \dfrac{Q_c}{\alpha Q_r^2} \cdot \dfrac{\mathbf{Re}(\delta S_{21})}{\kappa_1} \\ E_{Q_i} = V\Delta_0 \cdot \dfrac{Q_c}{\alpha Q_r^2} \cdot \dfrac{\mathbf{Im}(\delta S_{21})}{\kappa_2}, \end{cases} \tag{4.42}$$

in which we also use Eq. (4.32) to convert the QP density to energy. We note again, for Eq. (4.42), **everything is evaluated as a constant at** $f = f_{\text{read}} \approx f_r$. These values must be acquired in advance to the pulse data taking by performing the steady-state RF calibration *once*.[5] It is worth noting that, the general derivation presented here also applies to $f_{\text{read}} \neq f_r$, e.g., unable to accurately adjust f_{read} to be close to f_r due to practical limitations. In this case, instead of having the $(-2)Q_r^2/Q_c$ prefactor from Eq. (4.41) for converting S_{21} and Q_i (f_r), one needs to calculate $dS_{21}(f_{\text{read}})$ using the full Eq. (A.5) in the appendix, tedious but nonetheless still yielding a calculable constant for Eq. (4.42).

We now inspect if we have obtained all the ingredients needed by Eq. (4.42). We have acquired Q_c and Q_r in the RF calibration, T is controlled and measured by the experimental apparatus, V is known by design (KID inductor area × film thickness), $\omega = 2\pi f_{\text{read}} \approx 2\pi f_r$, but we still lack the electron density of states N_0, the kinetic inductance fraction α, and the 0-K QP bandgap Δ_0; readers may reference Eq (4.29) for the inputs of κ. Among these yet discussed parameters, the density of states is well-formulated by the standard nonrelativistic quantum/statistical mechanics. Using the BCS theory, which relates T_c to the Cooper-pair mass for estimating the top boundary of the phase space, also providing the Fermi surface as the bottom of the phase space, N_0 is strictly calculable. In fact, N_0 are well-measured in the literature for most commonly adapted BCS superconductors and may be taken as a

[5] Or more realistically, at a frequency required by the RF re-calibration due to, for example, system drift.

constant intrinsic property for BCS SC materials. On the other hand, although in the BCS theory Δ_0 is also directly related to T_c, but owing to their linear dependence

$$\frac{2\Delta_0}{k_B T_c} \sim \mathcal{O}(1), \tag{4.43}$$

we expect Δ_0 to vary by 10–20% as a consequence of the practical T_c variation, or even more for some sensitive low-T_c materials. Note that the proportionality constant in the above relation, typically ≈ 3, is in principle uniquely calculable for every BCS SC, but it nevertheless varies between different SCs, so we write it in order-of-magnitude to just illustrate the idea.

For α, we recall the analysis in Sec. 3.6. By comparing the designed and the fabricated resonant frequencies, we obtained α_{Al} in 10–15% based on the Nb extrapolation. Despite the extrapolation is apparently inaccurate by itself, there are also precautions/uncertainties for relying on the design-fabricated f_r difference to estimate α as in Sec. 3.6. First, materials that are perfectly inductance-less, or equivalently possessing massless or scatter-less charge carriers, simply do not exist, so there is no direct method to experimentally verify the simulated resonant frequencies given by SONNET for such a material. Second, we found that f_r for the fabricated devices vary at an observable level, for both the design-to-fabricated variation of different KIDs in the same device as well as cooldown-to-cooldown variation for individual resonators; neither is unexpected. For the former, it is easy to understand considering Eq. (4.24), which indicates that the SC film impedance scales quadratically in thickness hence linearly in f_r by Eq. (3.11), and we know that it is not uncommon to measure film thickness variations at a few- to ten-percent level for sputtered films across ⌀3" areas. For the latter, the variation may be easily caused by flux trapping, which results in a different amount of free current that generates kinetic inductance, each time the resonator cools across T_c and regains superconductivity (Brink, 1995). Specifically for our system, in the absence of an artificial magnetic component in the cryostat, the dominant flux for our experiment is Earth's magnetic field attenuated by the cryogenic magnetic shielding. We therefore expect the flux to be predominately horizontal and strongly depend on the orientation of the device relative to the ground. Outside the Cryoperm magnetic shield, we measure the magnetic flux to be about 50 μT horizontally and an order-of-magnitude lower vertically, which is consistent with the expectation for Earth's magnetic field at the geographical location of Caltech.

4.3.3 Temperature sweep calibration

We realize, in order to achieve a robust S_{21}-to-energy calibration, we need to calibrate the input parameters for Eq. (4.42), especially α and Δ_0, individually for each KID, every time the detector is cooled to the SC state. One technique developed by (Gao, 2008) for such a calibration is utilizing thermally generated QPs to simulate the excess QP population upon (phonon) pulse detection. Unlike in an instant pulse detection, this temperature control technique allows a quasistatic process, so the calibration is done much more accurately at each temperature that corresponds to an equilibrated QP density. When we steadily raise the temperature of the detector to each designated temperature, we incrementally remove the KIDs' superconductivity. Assuming this suppression of superconductivity is dominated by the well-formulated thermal QP generation, we then extract the dependences for the QP density increase, by fitting a Mattis-Bardeen-based thermal model curve of the resonance to the temperature-controlled data. The α and Δ_0 parameters determining $n_{\mathrm{qp}}(T)$ are then adopted in Eq. (4.42) for the pulse detection. We have already explained that, limited by the single-f_{read} fast readout scheme, our model for energy calibration is only applicable for small signals. In other words, we expect in real applications the KIDs' superconductivity should only be diminished infinitesimally in the linear regime during the pulses. This statement then guarantees that the aforementioned temperature-sweep calibration is applicable in practice, since we know we may significantly interfere or even completely remove the KIDs' superconductivity by simply raising the temperature to a good fraction of T_c.

For the assumption that the thermal QPs dominate the QP population at equilibria, we have

$$n_{\mathrm{qp,th}} = 2N_0\sqrt{2\pi k_{\mathrm{B}} T \Delta_0}\, e^{\frac{-\Delta_0}{k_{\mathrm{B}} T}}, \tag{4.44}$$

the theoretical Boltzmannian QP density for a thermal equilibrium at the temperature T. Eq. (4.44) is nothing but the classical thermodynamic density for the QP gas, which exhibits the bandgap energy Δ, and the equation is multiplied by 2 for the two-QP contribution from each broken Cooper pair. However, despite Eq. (4.44)'s solid physics foundation and simplicity, we can not guarantee that the controlled thermal bath, i.e., the substrate, is the only energy source for the QP generation or the energy dissipation channel. Therefore, Eq. (4.44) provides a lower bound for the "ideal" n_{qp} situation when all other non-thermal pair-breaking sources are carefully mitigated rather than the most practical forecast. Nevertheless, if we vary the temperature adiabetically, or realistically slowly enough for the QP system to

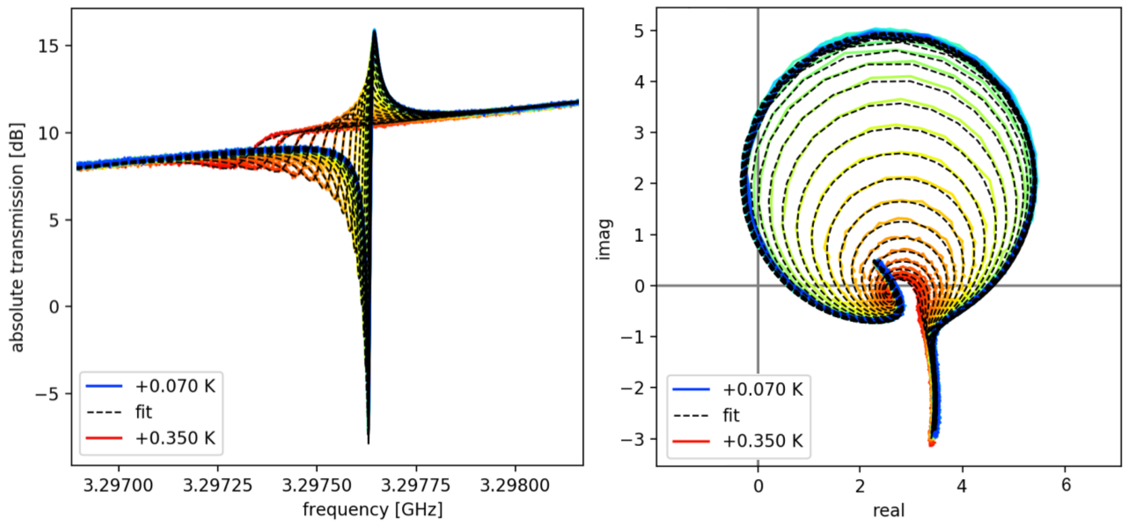


Figure 4.13: The temperature sweep calibration dataset for a KID resonance at 70–350 mK, shown respectively by the rainbow colors from blue to red at a 10 mK increment. The dashed curves represent the resonance fits (Eq. (4.5)), with the thick blue and red lines mark the start (70 mK) and the end (350 mK) of the series, respectively. For each resonance fit, a pair of Q_i^{-1} and f_r is extracted as one data point in Fig. 4.14. Original figure reprocessed by T. Aralis for a better color presentation.

fully equilibrate at each temperature, we can assume that the change is dominated by the difference in $n_{\mathrm{qp,th}}$. Under this assumption, we integrate Eq. (4.37) to be (algebra for $f_r(T)$ approximation in Appx. A)

$$\begin{cases} Q_i^{-1}(T) = -\alpha\kappa_1 n_{\mathrm{qp,th}} + Q_{i,0}^{-1} \\ f_r(T) \approx f_{r,0}\left(\dfrac{-\,\alpha\kappa_2 n_{\mathrm{qp,th}}}{2} + 1\right), \end{cases} \tag{4.45}$$

where we use $Q_{i,0}^{-1}$ and $f_{r,0}$, mathematically brought out as integration constants, to represent the nonnegligible, non-Mattis-Bardeen empirical resonator loss and the resonant frequency approaching 0 K, respectively. We also discard terms with values of zero at 0 K by definition, including $n_{\mathrm{qp,th}}$ and $\Delta f/f_r$; a step-by-step derivation is given in Appx. A. So, by taking multiple full-circle resonance frequency sweeps at different temperatures, typically ranging from the lowest achievable operating temperature to about 1/3 T_c, and then performing the RF calibration fit for each temperature to acquire $Q_i^{-1}(T)$ and $f_r(T)$ as functions of the controlled temperature, we then fit Eq. (4.45) to the $Q_i^{-1}(T)$ and $f_r(T)$ data to extract $Q_{i,0}$, $f_{r,0}$, and α, Δ_0, N_0 that are embedded in $\kappa_{1(2)}$. Fig. 4.13 shows an example for the temperature sweep dataset, whose corresponding Eq. (4.13) fitting result is provided in Fig. 4.14. Readers may also find our python-based fitting scripts on (Chang, n.d.).

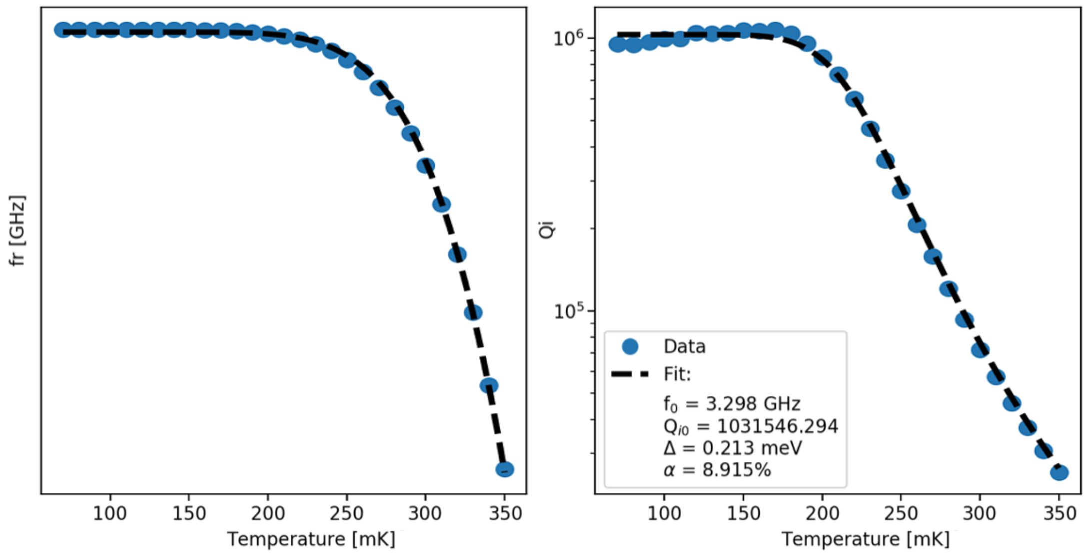


Figure 4.14: The Mattis-Bardeen fit (Eq. (4.45)) to the temperature sweep dataset presented in Fig. 4.13. The $Q_i^{-1}(T)$ and $f_r(T)$ results are shown in the right- and left-hand side panels, respectively. For this particular resonance, the extracted parameter values are listed in the inset in the $Q_i^{-1}(T)$ panel.

In terms of the practical execution of the Mattis-Bardeen temperature sweep calibration, due to N_0's small variation as explained previously, we find that it is sufficiently accurate to fix the value to

$$N_0 = 1.72 \times 10^4 \quad \left[\frac{1}{\mu\text{m}^3\,\mu\text{eV}}\right] \tag{4.46}$$

for aluminum thus subsequently remove it from the rather complex fitting formula to help with the fit. We verify the assumption of Eq. (4.46) retrospectively by relaxing N_0 after the other four floating parameters are first constrained without N_0's participation, and the results are generally consistent with Eq. (4.46) within fitting uncertainties. For the initial values of the gap energy and the kinetic inductance fraction, our experience is that it is typically sufficient to use

$$\begin{cases} \Delta_0 = 180 \pm 10 \quad [\mu\text{eV}] \\ \alpha = 0.10 \pm 0.03\ , \end{cases} \tag{4.47}$$

which we have also elaborated are anticipated to be accurate within ≈15% for typical LE KIDs instrumented from 20–40-nm sputtered aluminum films. Note that Eq. (4.47) is material- and resonator design-specific so is otherwise inapplicable for non-LE, non-Al, or non-thin film regime KIDs. Because of the complexity and the nonlinearity of Eq. (4.45), we find that it is generally helpful to initialize the fitting process with multiple combinations of α and Δ_0 that are randomly seeded in the suggested uncertainty ranges of Eq. (4.47). The results given by these random initial values then serve as cross-checks for excluding outliers and non-convergent fits. For $Q_{i,0}^{-1}$ and $f_{r,0}$, since thermal QPs are largely suppressed at $T \ll T_c$, the Q_i^{-1} and f_r extracted from the resonance circle at $T < T_c/10$ are usually sufficient to be the initial values. This argument is visualized in Fig. 4.14, where the reader may see that the fitted $Q_i^{-1}(T)$ and $f_r(T)$ curves are mostly independent of the temperature below 120 mK ($T_c \approx 1.2$ K).

4.3.4 Readout-power QP generation

This section provides a theoretical framework that we developed to study the readout power QP generation efficiency. The methodology was originally prepared for a dataset acquired by Dr. F. Defrance at ≈240 mK with different readout powers. The dataset did not yield a conclusive result with distinguishable readout power-generated QPs in the presence of thermal QPs. We encourage interested future readers to repeat the experiment at temperatures much below T_c to suppress the thermal QPs.

The existence of a non-zero empirical $Q_{i,0}$ (finite $Q_{i,0}^{-1}$) in Fig. 4.14 suggests there could be mechanism(s) dominating the QP generation at low temperatures other than the natural thermal generation. Among all potential energy sources, one inevitable source is the readout power, where the QPs can be generated by the GHz photons from the readout signal. Especially for DM experiments hosted by carefully built low-radiation facilities, such QP population could be nonnegligible when thermal QPs are highly suppressed.

One can easily demonstrate the hypothesis is viable by raising the readout power to continuously offset a KID resonance toward lower frequencies, which is consistent with an increased n_{qp}. To model the phenomenon, we assume the QP density at 0 K $n_{\text{qp},0}$ is dominated by the readout power-generated QP density $n_{\text{qp,read}}$, which in turn determines the observed $Q_{i,0}^{-1}$ and $f_{i,0}$. One can compare the GHz photon energy to Al's pair-breaking energy and conclude that their difference, roughly 1:40, must result in a highly suppressed QP generation process that involves many photons. Such a process is impossible to compute perturbatively from first principle, letting alone in practice the readout power coupling is also subject to practical details like the KID/feedline design or the surrounding RF structures.

On the other hand, we have demonstrated that we can unambiguously attribute the excess QPs to the readout power, so we are inspired to develop a data-driven technique to calibrate the readout power QP generation efficiency. We define an empirical readout power pair-breaking efficiency η_{read} by

$$\Gamma_{\text{read}} = \frac{\eta_{\text{read}} P_{\text{read}}}{\Delta}, \tag{4.48}$$

where Γ_{read} is the rate of QP generation due to readout power P_{read}. We start from the general generation-recombination (GR) equation

$$\Gamma_{\text{tot}} = R n_{\text{qp,tot}}^2 V, \tag{4.49}$$

where we use the subscript "tot" to denote the total amount, and R is the material's recombination constant. The recombination constant for aluminum is commonly found in the literature to be

$$R_{\text{Al}} \approx 10 \left[\frac{\mu\text{m}^3}{\text{s}}\right]. \tag{4.50}$$

We will also introduce our own method for calibrating R in the next section.

At a very low readout power that the QP generation is dominated by the thermal generation, the QP generation rate is given by

$$\Gamma_{\text{th}} = R n_{\text{qp,th}}^2 V. \tag{4.51}$$

When we raise the readout power to significantly impact the total QP density, Eq. (4.49) becomes

$$\Gamma_{\text{tot}} = \Gamma_{\text{read}} + \Gamma_{\text{th}} = \frac{\eta_{\text{read}} P_{\text{read}}}{\Delta} + R n_{\text{qp,th}}^2 V. \tag{4.52}$$

To be comparable to the data, we relate the readout power loss in the KID P_{KID} to the controllable DAQ readout power, the "generator" power P_g, by[6]

$$P_{\text{KID}} = \frac{2Q_r^2}{Q_i Q_c} P_g, \tag{4.53}$$

where we also assume $f \approx f_r$ to simplify the original formula. As pointed out previously, it is possible that not all the energy lost in the KID generates QPs, so we further modify Eq. (4.53) by

$$P_{\text{read}} = \frac{Q_i}{Q_{i,\text{qp}}} P_{\text{KID}}, \quad Q_{i,\text{qp}} \geq Q_i, \tag{4.54}$$

[6]Eq. (4.58) in (Gao, 2008).

where $Q_{i,\mathrm{qp}}$ represents specifically the QP-creating power, while Q_i encompasses all the loss mechanisms in the KID.

Combining Eq. (4.52)–Eq. (4.54), also choosing to operate at $T \ll T_c$ to enhance $n_{\mathrm{qp,read}}$'s dominance over $n_{\mathrm{qp,th}}$, we obtain

$$\eta_{\mathrm{read}} = \frac{\Delta_0 Q_c Q_{i,\mathrm{qp}}}{2P_g Q_r^2} RV(n_{\mathrm{qp,tot}}^2 - n_{\mathrm{qp,th}}^2). \tag{4.55}$$

Eq. (4.55) suggests that, with two datasets taken at different readout powers, we are able to write down Eq. (4.55) independently for the two readout powers and compare them to determine η_{read}. Ideally, if one of the two datasets is taken at a low readout power for the thermal $n_{\mathrm{qp,th}}$ to dominate $n_{\mathrm{qp,tot}}$ (Eq. (4.44)), and the other is at a high power for $n_{\mathrm{qp,read}}$ to dominate, the comparison is simplified to attributing the internal quality factor degradation to readout power-generated QPs. The internal quality factor change can then be converted to a QP density change, from where one can determine the readout power QP generation efficiency.

However, there are subtleties embedded in Eq. (4.55). Based on the relation of Q_i, $Q_{i,\mathrm{qp}}$, and $Q_{i,0}$, one might be tempted to calculate $Q_{i,\mathrm{qp}}$ by

$$Q_{i,\mathrm{qp}}^{-1} = Q_i^{-1} - Q_{i,0}^{-1} \tag{4.56}$$

using the Q_i^{-1} and $Q_{i,0}^{-1}$ fitted with Eq. (4.45). We compare Eq. (4.56) to Eq. (4.45) and find that the two equations combined suggest

$$Q_{i,\mathrm{qp}}^{-1} = -\alpha\kappa_1 n_{\mathrm{qp,\underline{\mathbf{th}}}}, \tag{4.57}$$

i.e., Eq. (4.45) assumes the QP population exhibits is predominantly thermal and is fully responsible for the QP-related readout power loss. In other words, Eq. (4.56) and Eq. (4.45) are simultaneously valid only when all the QPs are thermalize. By successfully fitting Eq. (4.45) to the temperature sweep data, we have effectively shown that the assumption is acceptable at low readout powers, but it remains to be shown for higher powers.

We anticipate that, with a high readout power constantly generating QPs, the QP energy distribution can be significantly altered from a thermal distribution. In this case, the internal quality factor can as well be degraded even at 0 K, i.e.,

$$Q_{i,0} = f[n_{\mathrm{qp}}(P_g)]. \tag{4.58}$$

Therefore, when we raise the readout power, we should consider

$$Q_{i,\mathrm{qp}}^{-1} = -\alpha\kappa_1 n_{\mathrm{qp,\underline{\mathbf{tot}}}} \neq -\alpha\kappa_1 n_{\mathrm{qp,\underline{\mathbf{th}}}} \tag{4.59}$$

due to the fact that $n_{\mathrm{qp},\underline{\mathbf{tot}}} \neq n_{\mathrm{qp},\underline{\mathbf{th}}}$ when P_g is high enough to achieve $n_{\mathrm{qp,read}} \gtrsim n_{\mathrm{qp,th}}$. In summary, the only valid parameter we can supply Eq. (4.55) based on Eq. (4.45) is $n_{\mathrm{qp,th}}$, which should be obtained from low-readout power temperature sweep data.

Since we are unable to objectively determine $Q_{i,\mathrm{qp}}$ for the high-power data, which represents the key effect of the readout power in Eq. (4.55), we make the assumption that the non-QP-related intrinsic resonator loss is sufficiently independent of the readout power. This assumption enables us to cancel the equal non-QP contribution in Q_i between the high- and low-power datasets. We should note that there already exists well-established conditions that can invalidate this assumption, e.g., the TLS loss, which is non-QP but strongly depends on the readout power. We proceed to lower the readout power until the fitted $Q_{i,0}$ decouples from P_g. We regard this readout power as the proper "low power" that provides not only $n_{\mathrm{qp,th}}$ but also the $Q_{i,0}$ value for, per the assumption, all readout powers. This $Q_{i,0}^{-1}$ allows us to write

$$Q_i^{-1} = Q_{i,\mathrm{qp}}^{-1} + Q_{i,0}^{-1} \tag{4.60}$$

$$= -\alpha\kappa_1 n_{\mathrm{qp,tot}} + Q_{i,0}^{-1} \tag{4.61}$$

$$= \left(-\alpha\kappa_1 n_{\mathrm{qp,th}}(T_{\mathrm{qp}}) + Q_{i,0}^{-1}(P_g)\right) + Q_{i,0}^{-1}, \tag{4.62}$$

in which we fix $Q_{i,0}^{-1}$ to the value obtained from the low-power data. The difference of the low-power $Q_{i,0}^{-1}$ and the total Q_i^{-1} measured at arbitrary readout powers then represents the $Q_{i,\mathrm{qp}}^{-1}$ due to all QP generation processes at the readout power. Eq. (4.61) suggests that, once we obtain the high-power $Q_{i,\mathrm{qp}}^{-1}$ combining the thermal- and the readout power-generated QPs, we can use α and κ_1 to compute the $n_{\mathrm{qp,tot}}$ at the high power, and then the input parameters of Eq. (4.55) are fully supplied.

4.4 Pulse detection

We have fully characterized the steady-state performance of the KIDs, so we proceed to model their transient behaviors. The following analysis will eventually lead to our pulse response model for the particle detection toward the end of this section. In the next section, we will then combine the pulse response with the noise performance to model the energy resolution for the detector.

4.4.1 Pulse shape

We begin from the general out-of-equilibrium GR equation

$$\frac{dn_{\mathrm{qp}}}{dt} = \frac{-\Gamma_{\mathrm{R}} + \Gamma_{\mathrm{G}}}{V} = -Rn_{\mathrm{qp}}^2 + \frac{\Gamma_{\mathrm{G}}}{V}, \tag{4.63}$$

where n_{qp} is the instantaneous QP density, and Γ_{G} and Γ_{R} are the total QP generation and recombination rates,[7] respectively. Assuming the "background" QP generation rate due to sources aside from the main pulses of interest varies much more slowly than the pulses, at the typical time scale of the pulses, we have the pulse-free equilibrium condition

$$\frac{dn_{\mathrm{qp}}}{dt} = 0 = -Rn_{\mathrm{qp,tot}}^2 + \frac{\Gamma_{\mathrm{G}}}{V}, \tag{4.64}$$

where we use the same notation $n_{\mathrm{qp,tot}}$ to emphasize that in this case the recombining QPs are essentially the steady-state QPs analyzed in the last section. Combining Eq. (4.63) and Eq. (4.64), we obtain the defining equation for the QP dynamics

$$\frac{dn_{\mathrm{qp}}}{dt} = -R(n_{\mathrm{qp}}^2 - n_{\mathrm{qp,tot}}^2), \tag{4.65}$$

which can be analytically solved to (algebra in Appx. A)

$$n_{\mathrm{qp}}(t) = \left(\frac{1+\chi}{1-\chi}\right) n_{\mathrm{qp,tot}}, \tag{4.66}$$

where

$$\chi(t) = \left(\frac{n_{\mathrm{qp,max}} - n_{\mathrm{qp,tot}}}{n_{\mathrm{qp,max}} + n_{\mathrm{qp,tot}}}\right) e^{-2Rn_{\mathrm{qp,tot}}t}. \tag{4.67}$$

In deriving the above result, we choose $t = 0$ to be the time for an instantaneous pulse energy injection and assume $n_{\mathrm{qp}}(t=0) = n_{\mathrm{qp,max}}$, the maximal instantaneous QP density the pulse creates. In the limits of very large and small pulses as defined below, Eq. (4.67) reduces to

$$\chi(t) \rightarrow \begin{cases} e^{-2Rn_{\mathrm{qp,tot}}t} & , \; n_{\mathrm{qp,max}} \gg n_{\mathrm{qp,tot}} \\ \left(\dfrac{\delta n_{\mathrm{qp}}}{2n_{\mathrm{qp,tot}}}\right) e^{-2Rn_{\mathrm{qp,tot}}t} & , \; \delta n_{\mathrm{qp}} = n_{\mathrm{qp,max}} - n_{\mathrm{qp,tot}} \rightarrow 0, \end{cases} \tag{4.68}$$

which yields the pulse shapes (algebra in Appx. A)

$$n_{\mathrm{qp}}(t) \rightarrow \begin{cases} (Rt)^{-1} & , \; n_{\mathrm{qp,max}} \gg n_{\mathrm{qp,tot}} \\ n_{\mathrm{qp,tot}} + \delta n_{\mathrm{qp}} e^{-2Rn_{\mathrm{qp,tot}}t} & , \; \delta n_{\mathrm{qp}} = n_{\mathrm{qp,max}} - n_{\mathrm{qp,tot}} \rightarrow 0. \end{cases} \tag{4.69}$$

For the large-pulse limit, we present the equation considering $t \ll 1$ so to preserve the large pulse characteristics at the early time; the full solution asymptotes to the small-pulse equation at later time as expected.

Eq. (4.69) shows that, when a large pulse that dominates the QP population is detected, one may simply determine the recombination constant R from the early

[7] Unit chosen to be consistent with Γ_{th} and Γ_{read}.

$\sim 1/t$ pulse decay epoch free from degeneracies with other factors. In practice, this R extraction is done by fitting a straight line of a slope of -1 to the logarithmic $n_{\rm qp}(t)$ data. This conclusion is physically expected since, at such a high QP density that is effectively insensitive to QP density change due to recombination, the rate of recombination should be solely determined by the intrinsic properties of the material, which is represented by the QP-QP scattering cross section R. At the same time, due to the need for two QPs to complete the interaction and therefore the quadratic dependence of $n_{\rm qp}$ in

$$\Gamma_{\rm R} = \frac{dn_{\rm qp}}{dt} = Rn_{\rm qp}^2, \tag{4.70}$$

we also anticipate $n_{\rm qp} \sim t^{-1}$. For the small-pulse case, Eq. (4.69) shows a well-anticipated exponential decay for the excessive QP part ($\delta n_{\rm qp}$) with a decay constant of

$$\tau_{\rm qp} = \frac{1}{2Rn_{\rm qp,tot}}. \tag{4.71}$$

Eq. (4.71) is just the time constant that arises in the small-perturbation version of Eq. (4.70) when $n_{\rm qp}(t)$ is dominated by the background population $n_{\rm qp,tot}$. This result motivates the common definition of the QP lifetime and therefore our notation of $\tau_{\rm qp}$.

Eq. (4.69) allows a well-defined two-step calibration for the recombination constant R and the background QP population $n_{\rm qp,tot}$: We first determine R using a large pulse, and then we substitute the obtained R value into the QP lifetime observed in a small single-exponential pulse to determine $n_{\rm qp,tot}$. Of course, one can always use the full solution in Eq. (4.66) to simultaneously determine R and $n_{\rm qp,tot}$ with arbitrary pulses. However, we should note that, while the calibration technique employs large pulses to constrain R in a straightforward fashion, it contradicts our small-signal-based calibration between S_{21} and $n_{\rm qp}$ derived in Sec. 4.3. Recall that the κ_1 and κ_2 factors obtained by the Mattis-Bardeen fit are strictly speaking only valid when $\delta n_{\rm qp} \ll n_{\rm qp,tot}$, therefore the recombination constant estimated by large pulses does bare some systematic uncertainty in $\kappa_{1(2)}$. Nevertheless, as (Gao, 2008) pointed out, in practice κ only vary by factors of $\lesssim 3$ as long as $T_{\rm qp} \ll T_c$, while in our case the instantaneity of the pulses should also prevent altering the effective $T_{\rm qp}$, so we believe that utilizing the κ factors introduces comparable or smaller uncertainties. Fig. 4.15 presents two examples of the preliminary pulses obtained by the 80-KID prototype detector, one through the nominal phonon-mediated detection scheme, and the other through a direct QP creation technique. To directly create QPs in a

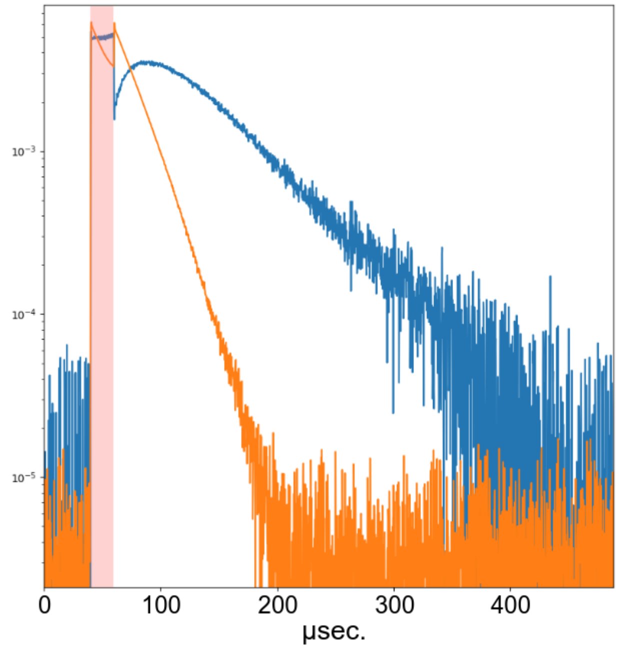


Figure 4.15: Example pulses obtained by in 80-KID prototype detector via the nominal phonon-mediated detection (blue) and the direct RF-pulsing energy injection (orange). The former and the latter exhibit decay time constants of ≈120 μs and ≈23 μs, respectively. The pink-shaded region represents the RF energy injection period, where the readout electronics are compromised and provide unphysical data as shown by the curves. The pulse sizes (y-axis) presented in this figure are in the raw DAQ unit. More information is found in the corresponding text and (Chang et al., 2019).

KID, we combine a strong monotonic RF pulse at the KID's resonant frequency into the readout signal, so the strong RF pulse is absorbed by the KID and creates QPs much like the readout-power QP generation mechanism introduced previously. In both pulse examples, we observe pulse shapes that are consistent with perfect single-pole exponential decays but with distinct decay time constants, ≈120 μs and ≈23 μs for the phonon-mediated and the RF-injection cases, respectively. O. Wen is currently pursuing a detailed investigation for the pulse shape, and more information about the RF energy injection technique is provided in (Chang et al., 2019) and the later part of this section.

Before rigorously including the phonon response into our model to complete the detection flow, we may already draw a few interesting conclusions for the current analysis with order-of-magnitude estimates. Considering the typical recombination constant in Eq. (4.50) and

$$n_{\mathrm{qp,tot}} \sim 10^{1-2}\ \mathrm{QP}/\mu\mathrm{m}^3 \tag{4.72}$$

achieved in the literature for aluminum KIDs, e.g., (de Visser et al., 2014), Eq. (4.71) suggests

$$\tau_{\rm qp} \sim 10^{2-3}\ \mu\text{sec.}, \tag{4.73}$$

a typical range of QP lifetimes observed for Al thin-film devices. We compare this typical QP lifetime to the expected pulse duration to examine our earlier instantaneous QP creation assumption. For the phonon-mediated detector design, the arrival of the phonon energy from the recoil is not instantaneous due to the finite-speed energy transportation with phonons. We therefore expect the incoming energy accumulates at a characteristic time scale that is manifested by the pulse rise time

$$\tau_{\rm rise} \sim \frac{t_{\rm sub}}{v_{\rm ph}}, \tag{4.74}$$

where $t_{\rm sub}$ is the typical size (thickness) of the detector, and $v_{\rm ph}$ is the speed of the phonons, i.e., the sound speed of the material. Whether the substrate is Si, Ge, or other proposed solid-state materials (Knapen et al., 2017; Kurinsky et al., 2019), we have

$$v_{\rm ph} \sim 10^{3-4}\ \text{m/s} \tag{4.75}$$

$$(v_{\rm ph,Si/GaAs/diamond} \approx 8/5/12\ \text{km/s}).$$

Taking

$$t_{\rm sub} \sim 0.1 - 1\ \text{cm} \tag{4.76}$$

for typical detector sizes, we find

$$\tau_{\rm rise} \sim 10^{-1-1}\ \mu\text{sec.}, \tag{4.77}$$

indicating

$$\tau_{\rm rise} \ll \tau_{\rm qp}. \tag{4.78}$$

So, unless a detector is made atypically large, our assumption for the instantaneous energy injection is valid for modeling the QP dynamics. More importantly, the time scale distinction supports that we should anticipate the phonon propagation to dominate the rise time, while the fall time is effectively $\tau_{\rm qp}$ in agreement with our earlier argument.

There are also derived design constraints given by the aforementioned characteristic pulse time scales. First, if the resonator does not exhibit a compatible ring-up/down time, the detection will be buffered hence yield an slower waveform than Eq. (4.74). We avoid such a condition by ensuring the resonator ringing time $\tau_{\rm ring}$ is shorter or comparable to $\tau_{\rm rise}$, so the resonator bandwidth encompasses the high-frequency rising of the incoming pulses, i.e.,

$$\tau_{\rm ring} \approx \frac{1}{\Delta f} \approx \frac{Q_r}{f_r} \lesssim \tau_{\rm rise}. \tag{4.79}$$

Note that we also do not want to unnecessarily widen the bandwidth to receive a higher noise. Providing the estimated $\tau_{\rm rise}$ and $f_r \sim$ GHz for typical KIDs, we find that, in order to ensure the rising edges of the phonon pulses are not distorted, we need

$$Q_r \lesssim 10^{4-5} \tag{4.80}$$

depending on the materials and the sizes of the detectors. Eq. (4.80) explains our earlier design choice in Sec. 3.5 (Eq. (3.15)). We then realize that, based on the fact that $Q_i > 10^6$ has been demonstrated achievable with careful fabrications, Eq. (4.80) indicates that the total quality factor of our design is dominated by the coupling quality factor. Although the coupling-dominated design is not the most efficient choice in terms of the readout power usage, it allows us to straightforwardly achieve the total quality factor adjustment by tuning the coupling capacitance, which in practice is easily done by adjusting the feedline-KID separation as in Sec. 3.5.

On the other hand, while we can always design $\tau_{\rm ring}$ to match $\tau_{\rm rise}$, there exist natural reasons for the signal fall time to deviate from $\tau_{\rm qp}$; one of which is when the phonons require a nonnegligible time to transfer their energy into QP generation relative to $\tau_{\rm qp}$. Such a buffered pair-breaking is manifested by slow-rising phonon pulses that lead to a continuous pair-breaking energy injection, while early QPs have started to recombine. Much similar to the prolonged signal rise time if the resonator ($\tau_{\rm ring}$) can not "catch up" with the energy injection ($\tau_{\rm rise}$), if the phonon absorption time $\tau_{\rm abs}$ has

$$\tau_{\rm abs} \lll \tau_{\rm qp} \tag{4.81}$$

and, phenomenologically (S. R. Golwala, 2010),

$$P_{\rm ph}(t) \sim e^{-\frac{t}{\tau_{\rm abs}}}, \tag{4.82}$$

where $P_{\rm ph}$ is the incoming phonon pair-breaking power, the fall time of the observed signal is modified by

$$\tau_{\rm qp} \longrightarrow \tau_{\rm qp} + \tau_{\rm abs}. \tag{4.83}$$

We will rigorously derive this conclusion in Sec. 4.6 following (S. R. Golwala, 2010)'s original derivation. In reality, the exact modification will depend on the practical details of the pulse shape and the temporal behavior of the phonons, such as the phonon absorption of surrounding materials including not only the KIDs but also all other materials in contact with the detector. T. Aralis and E. Lindeman are currently investigating the effect. We note that, while the reduced noise bandwidth due to the increased decay time constant is intuitively beneficial to the SNR, the buffered energy injection also reduces the pulse height and therefore causes a degradation in SNR. We will study the impact in the energy resolution due to these time constants in Sec. 4.6.

4.4.2 RF-pulsing calibration

We have developed a RF-pulsing calibration technique to test our pulse model. The technique utilizes controllable RF pulse injection to generate QPs and provides a systematic and flexible calibration scheme for both QP and phonon pulses. Fig. 4.16 illustrates the calibration scheme. By combining a short but strong monochromatic RF pulse at the designated KID's resonant frequency into the readout signal, we are able to create a (close-to-)instantaneous QP pulse event in the designated KID as modeled for Eq. (4.69). When the QPs generated by the RF pulse recombine, they subsequently release the energy in the form of phonons into the SC film and then the substrate. This energy transfer process then transforms each KID instrumented on the detector surface into a phonon pulse emitter. When we energize a "phonon pulser" KID, its own signal is determined by the KID's internal QP system dynamics, while for other "spectator" KIDs, their responses represent the designed phonon pulse detection, where we know the phonon source location is the well-defined pulser location. Contrasting traditional calibration techniques with external particle sources, the RF-pulsing technique allows us to physically locate, adjust in pulse energy, and trigger the events in a fully controlled fashion. Taking advantage of the high multiplexibility of our KID-based architecture, we plan to sequentially utilize each of KIDs in the detector as a phonon pulser and hopefully achieve an unprecedented fine calibration for the position dependence for phonon-mediated detectors.

We realize during the implementation of the RF pulsing calibration that there are a handful of practical limitations to the technique, which we are currently improving by fine-adjusting the electronic setup shown in Fig. 4.16. First, given the $Q_r \sim 10^5$ chosen for the prototype detector, we empirically find that the KIDs require pulses that are 10–20 μs in width and −30 to −20-dBm in power at the input of the device, modulo transmission attenuation uncertainty (feedline S_{21}), to inject sufficient RF

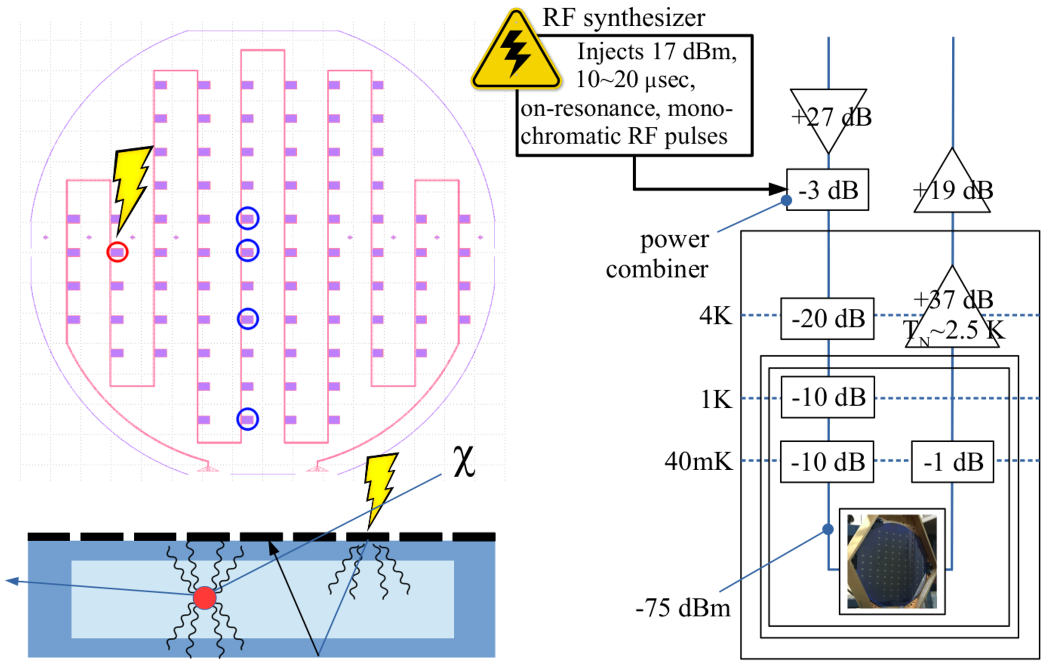


Figure 4.16: An illustration for the RF-pulsing calibration. On the left, the top layout shows the prototype detector with the RF-pulsed KID indicated by the yellow lightning symbol and the red circle. The blue circles mark the spectator KIDs for which we took data and present in Fig. 4.17. The bottom cartoon shows the corresponding side view of the detector, whose inner and outer regions are in light and dark blues, respectively. χ represents a DM particle that creates a recoil event at the red dot in the nominal inner region, whose phonon emission is represented by the wavy lines. The yellow lightning symbol in this cartoon also marks the RF-pulsed KID, which emits phonon into the substrate on the surface. The wiring schematic on the right illustrates the combination of the RF pulse into the readout system. The RF pulse is generated at the frequency of the red-circle KID by a RF synthesizer, details provided in the inset, and is injected into the readout system at room temperature with a power combiner at the input side (labeled). The schematic also shows our nominal readout amplification and attenuation setup. More details are provided in the corresponding text. Figure reproduced from (Chang et al., 2019).

energy for creating observable phonon pulses in the spectator KIDs. A quick calculation taking $Q_c \sim 10^5$, $Q_i \sim 10^6$, $V \approx 2 \times 10^5$ μm^3, $\Delta \approx 170$ eV, and also assuming an instant RF energy injection, yields

$$n_{\mathrm{qp}} \sim \frac{\eta_{\mathrm{read}} \left(\dfrac{Q_r^2}{Q_i Q_c} \right) P_{\mathrm{pulse}} t_{\mathrm{pulse}}}{\Delta V} \tag{4.84}$$

$$\approx (2-37) \times 10^6 \eta_{\mathrm{read}} \qquad [\mathrm{QP}/\mu\mathrm{m}^3] \tag{4.85}$$

in the pulser KID. Further assuming that the phonons emitted by this pulser KID are roughly equally delivered to all 80 KIDs and are subsequently observed at an single-KID phonon energy resolution of $\mathcal{O}(10)$ eV (Chang et al., 2019), which we will elaborate in the coming sections, the calculation suggests

$$\eta_{\rm read} \sim \mathcal{O}(10^{-5\sim-4}). \tag{4.86}$$

This order-of-magnitude estimate agrees with an $\eta_{\rm read} \lesssim 10^{-4}$ upper bound obtained in our earlier attempt of constraining $\eta_{\rm read}$ using Eq. (4.55) on the 240-mK power scan data.

The simple verification above suggests that the empirical minimal pulse strength for creating observable pulses is reasonable but nonetheless indicates other realistic issues. Since we install a total inline attenuation of 40 dB, split to different temperatures for thermal noise suppression (Fig. 4.16), plus an inevitable few-dB loss due to the RF combiner and the cable loss, the minimal power we need from the RF pulse synthesizer for generating observable signals is an atypically strong $\gtrsim$20 dBm. However, we use an Agilent-HP 83732B synthesizer in this experiment, which provides a maximal output power of 16–20 dBm with a high instability approaching the maximal output. In fact, it is due to this output limit that we have to extend the pulse duration into a non-instantaneous regime with respect to $\tau_{\rm ring}$ ($\tau_{\rm rise}$) so to accommodate the total injected energy. Also due to the same limitation, currently we are unable to unambiguously drive a phonon pulser KID into the strong recombination regime described by Eq. (4.69) to measure our own recombination constant, hence the single-pole decay shown earlier in Fig. 4.15. Secondly, due to the large pulse power, we realize that, even if we achieve such a strong RF energy injection by means such as employing amplifiers at the input, we are likely to damage the downstream electronics. In particular, the receiver circuit of the DAQ system consists of many electronic components that are specified with <0-dBm safe instantaneous power surges. As shown in Fig. 4.16, the readout system is located downstream of a cryogenic HEMT amplifier providing a gain of 37 dB and therefore expects a >10 dBm power surge. In order to protect the DAQ system, we limit ourselves from injecting greater than −1 dBm into the DAQ receiver, so we lower the pulse power by more than 10 dB. This weak pulsing then yields a small signal with too poor an SNR for self-triggering the pulse identification. Meanwhile, even though the phonon signals in the spectators show too weak a signal for triggering, the signal in the pulser due to the direct RF QP generation is much stronger and thus may serve as the global trigger for all the KIDs. Currently, we use this trigger

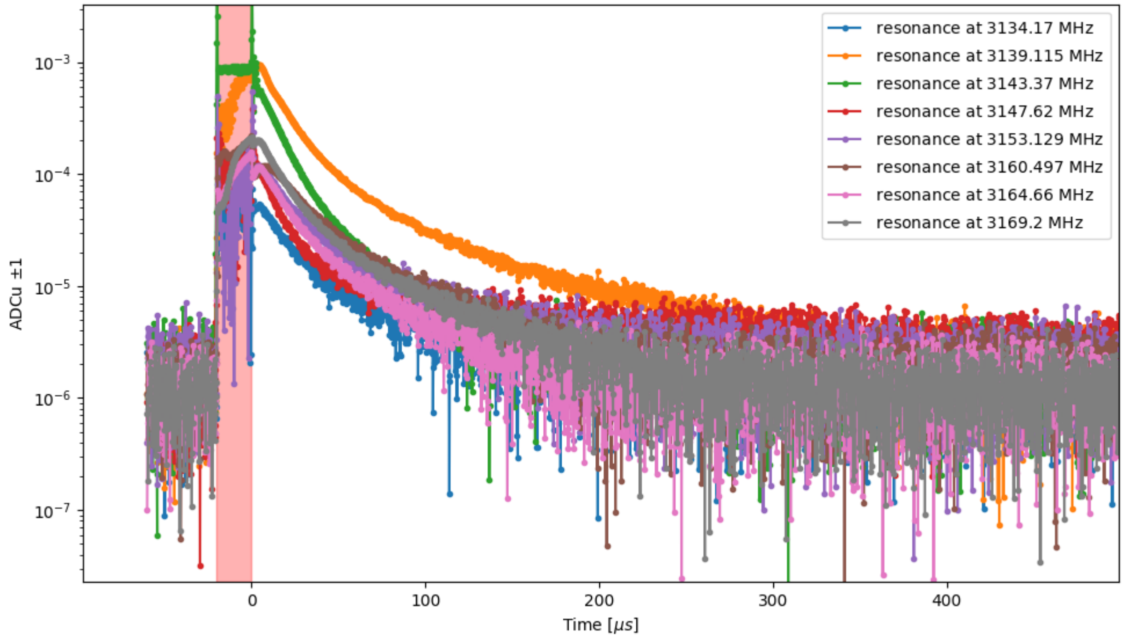


Figure 4.17: Example pulse templates obtained from the phonon-spectator KIDs in the 80-KID prototype detector. The pink-shaded region represents the RF energy injection duration, where the readout electronics is compromised and provides unphysical data as shown by the curves. The pulse sizes (y-axis) presented in this figure are in the raw DAQ unit. More information is found in the corresponding text and (Chang et al., 2019). Figure produced by T. Aralis.

scheme to simultaneously record the signals in all the KIDs, and, by averaging arbitrarily many[8] repetitive pulses for noise suppression, we are able to obtain a clear phonon pulse template for each spectator KID along with the already high-SNR QP pulse template for the pulser KID. Fig. 4.17 and Fig. 4.15 present the results for the phonon and QP pulses, respectively. Recalling that we discussed the necessity of a predefined pulse template for the OF calculation (Sec. 2.3), we will use these averaged phonon pulse templates later in the OF energy resolution calibration for our KID-based prototype detector.

Despite the preliminary status of the RF-pulsing calibration, we can still interpret the preliminary rise and fall times for both the QP and the phonon pulses. The rise and fall times observed for the pulser KID manifest the ringing time and the QP lifetime, respectively, and by comparing with their counterparts for the spectator KID, the difference for the rise times indicates the phonon propagation time from the pulser to the spectator, while the difference for the fall times may be contributed to effects such as the phonon absorption time. The data in Fig. 4.17 and Fig. 4.15

[8] We typically average 500–1000 pulses, which gives a satisfactory SNR in practically an instant.

show fall times for the spectators in the range of 50–200 μs, varying appreciably, while the pulser exhibits a fall time of 23 μs. Based on the observation, we find

$$\tau_{\mathrm{qp}} \approx 23\ \mu\mathrm{sec.}, \tag{4.87}$$

somewhat shorter than typical observations in the literature but not unreasonable if the film exhibits a relatively high $n_{\mathrm{qp,tot}}$ according to Eq. (4.71), and

$$\tau_{\mathrm{abs}} \approx 30 - 180\ \mu\mathrm{sec.} \tag{4.88}$$

Since the spectator KIDs are all located in close proximity as shown in Fig. 4.16, and in fact it should only take $\lesssim$10 μs for phonons to travel across the 3" substrate, the large variation in τ_{abs} can not be caused by the different distances between the pulser and the observing KIDs. The result therefore suggests other mechanisms to be the dominant causes for the high variation, e.g., low and/or variable phonon absorption efficiency for each KID. The absolute value of τ_{abs} that is in the range of $O(10^{1-2})$ μs is also nontrivial compared to the observed or the anticipated future τ_{qp} thus indicates the fall time modification discussed previously (Eq. (4.83)) is likely a realistic effect to be considered. T. Aralis is currently pursuing an improved model for phonon-SC film absorption and its consequential pulse shape and energy resolution modifications.

When we inspect the rising edges of the pulses, instead of extracting τ_{ring} and the phonon propagation delay, it becomes clear that the rising edges are contaminated by the pulsing. We find that the unphysical pulse shapes in the pulsing period clearly correlate with the strong RF pulse generation setting but are distorted in a yet-to-be understood manner. We believe that the distortion is most likely due to saturating the HEMT and/or the DAQ with an inappropriately large signal, despite the fact that that it is already lowered to avoid damage. The hypothesis is strongly supported by that we observe for the pulsed and non-pulsed readout configurations exhibiting distinguishable S_{21} equilibria and the $\vec{\kappa}$ trajectories in the IQ plane, i.e., out-of-equilibrium shifting direction upon δn_{qp} perturbations (introduced in the next section). Since the system requires a nontrivial amount of time to recover from the distortion, while currently the power limit also requires the RF pulse duration to last significantly into the rising periods of the signals, we are unable to extract useful information from the rise of the pulses. In addition, the general instability of the current proof-of-principle construction also suggests that it is preferable to repeat the experiment, also for more robust τ_{qp} and τ_{abs}, when the above issues are mitigated. We are currently implementing the so-called nulling signal, which is an identical copy of the troublesome RF pulse but exhibits an opposite phase, for canceling the RF surge before it enters the HEMT, e.g., at the output of the device. Sec. 4.7 presents the R&D status. We hope with the upgrade, the RF-pulsing calibration technique will be much more robust and versatile for future applications.

4.5 Noise performance

4.5.1 Noise analysis in IQ plane

When we read out the signal at $f_{\text{read}} \approx f_r$ in the absence of pulses, instead of acquiring a true constant of $S_{21}(f_r)$, we naturally observe that the obtained value fluctuates in time as

$$S_{21}(t) = S_{21}(f_r) + \delta S_{21}(t)|_{f \approx f_{\text{read}}}. \tag{4.89}$$

The variable δS_{21} component is the noise that limits our ability to perfectly determine $S_{21}(f_r)$ and defines the significance for later pulse detection. Eq. (4.89) implicitly assumes that, at precisely $f_{\text{read}} = f_r$, $S_{21}(f_r)$ is a constant in time and can be accurately determined in the absence of $\delta S_{21}(t)$. We will discuss the validity of the assumption and the techniques to enforce the condition later. The concept of noise confusion in the context of KID's RF-based detection scheme is most naturally visualized in the IQ plane as shown in Fig. 4.18. Instead of observing a singular point of $S_{21}(f_r)$ in the IQ plane, when sampled continuously in time, $S_{21}(t)$ forms a time-series cluster fluctuating with respect to the equilibrium, which is the expected true $S_{21}(f_r)$ according to Eq. (4.89). One can further represent this cluster by a contour that centers at the same equilibrium point and encloses a certain fraction of the data samples, such as ≈68.2% for a 1-σ contour. When the system state is temporarily driven away from the equilibrium, one may then determine with this noise contour the likelihood for such an event to be solely due to noise, or equivalently the significance for the event to be a signal of external interference that is not included in $\delta S_{21}(t)$.

Unlike other single-variable detection methods, an useful feature for an IQ plane-based signal is that the complex S_{21} detection is inherently 2-dimensional, containing not only the information of the absolute signal size, i.e., the energy, but also the directionality of the signal trajectory in the 2D plane (Fig. 4.18). To utilize this feature, a natural yet critical question is: What is the expected trajectory for the "real" QP generation? Eq. (4.42) gives the answer. By dividing the two equations therein, we obtain the relation

$$\frac{E_{f_r}}{E_{Q_i}} = \frac{\delta n_{\text{qp},f_r}}{\delta n_{\text{qp},Q_i}} = 1 = \frac{\mathbf{Re}(dS_{21})}{\mathbf{Im}(dS_{21})} \cdot \frac{\kappa_2}{\kappa_1}. \tag{4.90}$$

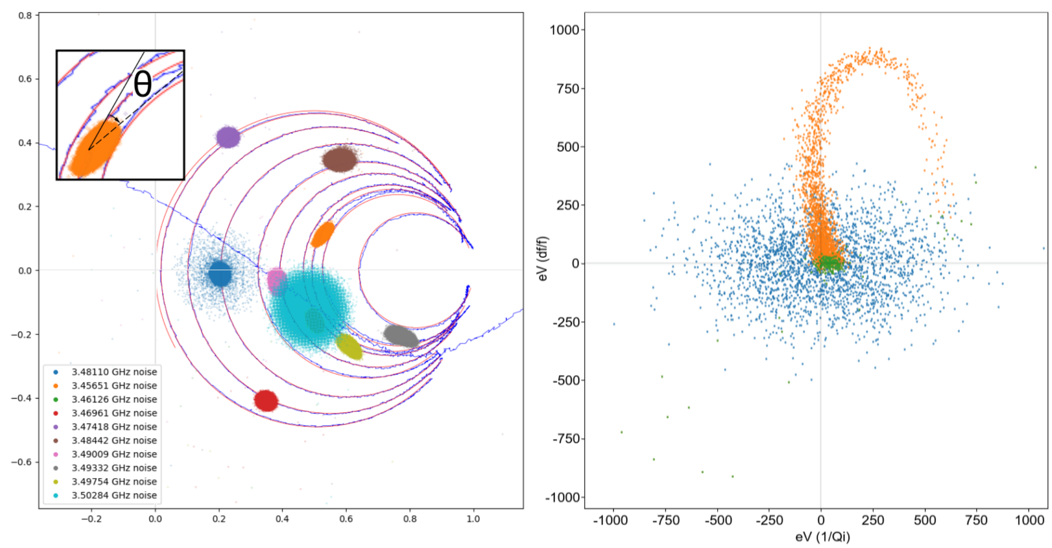


Figure 4.18: Left: Noise data streams acquired simultaneously from 10 different KIDs in the 80-KID prototype detector, all transformed to the idealized S_{21} IQ plane following Fig. 4.12. Red smooth circles represent the fitted idealized resonance circles. Due to practical misalignments of f_{read} and f_r, the data clusters are scattered on their corresponding resonance circles and are rarely in contact with the real axis given their noise ranges. In the inset, the solid line is the line tangent to the resonance circle at the selected noise cluster's equilibrium point, and the dashed line and θ show the direction defined by θ_{qp} (Eq. (4.91)). Right: The data streams acquired from the same KID previously utilized for direct RF pulsing (orange), pure noise (blue), and the averaged phonon pulse template in the RF-pulsing calibration (green). Both axes in this figure are calibrated into the QP generation energy according to Eq. (4.42). Due to the strong RF energy injection, the orange trajectory violates the small-signal linearization condition and thus exhibits a curved trajectory. Figure reproduced from (Chang et al., 2019).

We emphasize again that the subscripts f_r and Q_i for n_{qp} and E are just to denote that the energy (correspondingly the QP density) is independently calculated from the subscript data type along different axes. As true physical quantities, we expect the energy or its corresponding QP density to be independent of the data type utilized for extracting its value, hence the ratio of unity in the above equation. Eq. (4.90) then allows us to calculate the angle of the dS_{21} trajectory with respect to the imaginary axis, which is the idealized tangential df direction for the resonance circle (Fig. 4.12). Assuming the nominal $T \ll T_c$ and the small-signal operation conditions, Eq. (4.90) yields the angle

$$\theta_{\text{qp}} = \tan^{-1} \frac{\mathbf{Re}(dS_{12})}{\mathbf{Im}(dS_{12})} = \tan^{-1} \frac{\kappa_1}{\kappa_2}, \tag{4.91}$$

which we mark in Fig. 4.18 and Fig. 4.19. Specifically for SC aluminum, by substituting typical values

$$\begin{cases} T &= 20-100 \ \text{[mK]} \\ f_r &= 3-4 \ \text{[GHz]} \\ N_0 &= 1.72\times 10^{10} \ [1/\mu\text{m}^3\text{eV}] \\ \Delta_0 &= 170-190 \ [\mu\text{eV}] \end{cases} \tag{4.92}$$

into κ_1 and κ_2 (Eq. (4.29)), we find

$$\theta_{\text{qp,Al}} = 21-28^\circ. \tag{4.93}$$

We now understand that this angle is a characteristic property that is uniquely associated to aluminum, and similarly for every material. It is worth noting that, in the above calculation for $\theta_{\text{qp,Al}}$, the uncertainty range is predominately due to the assumed range of f_r for representing *all* the KIDs in our prototype detector. If we fix f_r to specific resonators with the typical fitted f_r uncertainty, the slow-varying κ functions at $T \ll T_c$ allow $\theta_{\text{qp,Al}}$ to be determined with an uncertainty of $\approx 1^\circ$, which is dominated by the assumed temperature and gap energy uncertainties in Eq. (4.92). These uncertainty sources are also quoted in Eq. (4.92) for representing typical values but are determined much more accurately for specific resonators, so in practical applications $\theta_{\text{qp,Al}}$ typically exhibits a sub-% uncertainty and therefore serves as a strong criterion for distinguishing QP activity from other non-QP signals.

Since the S_{21} trajectory due to real QP generation always follows θ_{qp} as illustrated in Fig. 4.19, we can safely attribute fluctuations that are orthogonal to the θ_{qp} direction to non-QP origins and manually remove them. Even for a noise that partially projects to the θ_{qp} direction, by comparing the internal coordinates of the resonance circles with respect to the global coordinate, i.e., the ϕ_c variation, we can also estimate the non-QP contribution to the noise that is shared among different resonators. In practice, based on the fact that the components involving in the signal generation, except the KID's QP system itself, are essentially the readout electronics, which all operate in the global continuum circle coordinate with no knowledge about the resonators, we can almost always attribute the noise orientated with the global coordinate to the readout electronics. These noise sources include the amplifiers and the DAQ RF equipment. We therefore name the directions given by the continuum circle coordinate at $S_{21}(f_r)$ the phase and the gain directions, whose difference will become clear when we discuss the physical mechanisms for the electronic noise generation. Thanks to the well-established working principles and calibration data

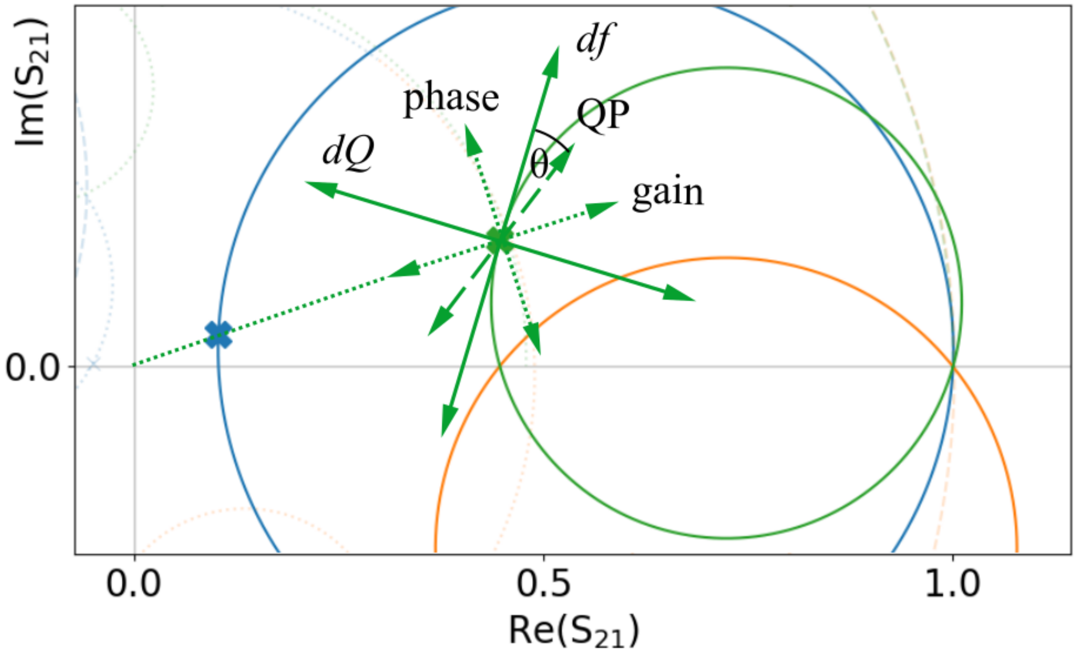


Figure 4.19: The directions of different noise sources, plotted at the ϕ_c-preserved idealized S_{21} readout point for the green resonance, consistent with Fig. 4.12 (left). The noise sources considered in this figure include QP-generated fluctuation (QP), KID-internal frequency fluctuation (df), KID-internal dissipation fluctuation (dQ), global-readout phase fluctuation (phase), and global-readout amplitude fluctuation (gain). df is tangential to the resonance circle at the readout point and is orthogonal to dQ; they belong to the KID-internal coordinate and are drawn by solid lines, c.f. Fig. 4.12. The gain direction always aligns to the global IQ-plane origin and in turn defines its orthogonal phase direction; they belong to the global-readout coordinate and are drawn by dotted lines. The QP direction is defined in the KID-internal coordinate by the θ_{qp} angle marked in the figure following Eq. (4.91). Notice that the global and KID-internal coordinates are degenerate when $\phi_c = 0$. Detailed discussion is found in the corresponding text.

for commercial readout electronics, we generally are able to predict their noise contributions and associate characteristic noise behaviors in the IQ plane to certain electronic components fairly accurately.

We categorize the electronic noise sources into three predictable classes: 1) Phase noise-dominated, 2) amplitude noise-dominated, and 3) IQ-insensitive power components. As the names suggest, these noise sources contribute noise energy predominantly via the components' instabilities in the phase, the amplitude, or the power that is equally carried in phase and amplitude. For the amplitude-type noise, we instead name it the gain noise so to emphasize that, in our readout system construction, the observed amplitude fluctuations are usually due to the instability in the amplifier gains that leads to an unstable amplitude detection. Since the electronics work independently of the KIDs thus affect S_{21} through the global signal transmission, their amplitude and phase directions for the state fluctuation also orientate with respect to the global IQ plane coordinate; Fig. 4.19 illustrates the orientation. Recalling the ϕ_c discussion for Fig. 4.8, Fig. 4.19 suggests that the resonance circle's internal coordinate that defines θ_{qp} is rotated relative to the gain and the phase directions by ϕ_c. Therefore, it is possible for a QP-generated trajectory to align with one of the electronic noise's principle axes when $-\phi_c \approx \theta_{\text{qp}}$ or $\pi - \theta_{\text{qp}}$. In contrast with the electronic noise, if the noise is produced by the instabilities in the KID but nevertheless not the QP system, the trajectories for such fluctuations should align with the resonator's internal coordinate, i.e., strictly following the orientation of the df-dQ coordinate at $S_{21}(f_{\text{read}})$ but not necessarily along θ_{qp}. One example for this type of KID-internal, non-QP fluctuation is the TLS fluctuation, which is generally believed to be an effective dielectric constant fluctuation in the KID's capacitor, likely due to the surface oxidation of the substrate. As the KID capacitance fluctuates, the consequence is a state fluctuation that ties to the resonance circle. We will discuss in more details in the following that, under sufficiently high readout powers, the TLS fluctuation may be suppressed in the *resonator's* amplitude direction, i.e., the dQ direction (Fig. 4.19), and leave the TLS noise observable only in df. In the following sections, we will also go through other expected physical noise sources one by one and eventually assign each of the noise directions in Fig. 4.19 a distinct physical meaning.

4.5.2 QP generation-recombination noise

We start from the fundamental QP GR noise to set an idealized noise performance, which serves as an inevitable lower bound as long as the QPs are utilized for the detection. In the absence of pulses, the only type of QP activities that causes noise is the natural thermodynamic fluctuation of the QP density at equilibrium. While the time-averaged QP density remains a constant due to balanced energy injection and dissipation, its instantaneous value still fluctuates in Poisson fashion due to spontaneous generation and recombination. Classically, such a fluctuation is only fully suppressed when the total energy of the system approaches zero, so according to our earlier discussion on the readout power QP generation, in principle we always have the QP GR noise even at 0 K, i.e., expecting

$$\langle n_{\text{qp,tot}} \rangle > 0, \tag{4.94}$$

where the angle bracket represents time-average, and we implicitly adopt $n_{\text{qp}} = n_{\text{qp,tot}}$ in the context of noise at equilibrium. Given proper thermodynamic conditions, such as temperature, equilibrated $\langle n_{\text{qp}} \rangle$, recombination constant, etc., one may understand that the problem is well-defined and fully solvable in standard statistical mechanics. Indeed, (Siegel, 2016) was able to present a rigorous derivation using only the Poisson statistics of the mean value λ, the equilibrated total number of QPs

$$\langle N_{\text{qp}} \rangle = \langle V n_{\text{qp,tot}} \rangle, \tag{4.95}$$

and the observation time interval Δt as variables. Using the notation/definition

$$S_x = \lim_{\Delta t \to \infty} \frac{|\tilde{x}(f)|^2}{\Delta t} \tag{4.96}$$

to present PSD-like quantities, where x is the time-domain quantity of interest and $\tilde{x}(f)$ is its Fourier-transformed counterpart, (Siegel, 2016) shows that, for the Poissonian probability of observing k GR events in t to $t + \Delta t$, i.e.,

$$\mathcal{P}(N_{\text{GR}}|_t^{t+\Delta t}) = \frac{(\lambda \Delta t)^k}{k!}, \qquad k \in \mathbb{Z}^*, \tag{4.97}$$

where N_{GR} is the total number of GR events observed since $t = 0$, one has

$$S_{dN_{\text{GR}}/dt}(f) = \lambda \tag{4.98}$$

for $f \neq 0$, i.e., neglecting the unobservable stationary QP background. This result is analogous to the Johnson-Nyquist noise for resistors, where the noise is due to the Poissonian scattering of electrons, so the white noise PSD is well consistent with our expectation.

Specifically for QP GR, by the simple relation

$$\lambda = \left\langle \frac{dN_{\text{GR}}}{dt} \right\rangle = \Gamma_{\text{GR}} = \frac{\Gamma_{\text{qp}}}{2}, \tag{4.99}$$

where the factor of 2 is to accommodate the different definitions for Γ_{qp}, the rate of creation/annihilation conventionally defined for individual QPs, as opposed to Γ_{GR}, which accounts for the GR "events" that require QP pairs, and also knowing

$$\Gamma_{\text{qp}} = V \left(R n_{\text{qp,tot}}^2 + \frac{n_{\text{qp,tot}}}{\tau_{\text{qp,max}}} \right), \tag{4.100}$$

which expresses the identical physics of Eq. (4.64) but, according to (Siegel, 2016), contains the extra phenomenological $\tau_{\text{qp,max}}$ for the empirical maximal QP lifetime, we obtain by chain rule

$$S_{\Gamma_{\rm qp}} = 2\left(\frac{\partial \Gamma_{\rm qp}}{\partial \left\langle dN_{\rm GR}/dt\right\rangle}\right)^2 S_{dN_{\rm GR}/dt} \tag{4.101}$$

$$= 2\times 2^2 \times \frac{V}{2}\left(Rn_{\rm qp,tot}^2 + \frac{n_{\rm qp,tot}}{\tau_{\rm qp,max}}\right)$$

$$= 4V\left(Rn_{\rm qp,tot}^2 + \frac{n_{\rm qp,tot}}{\tau_{\rm qp,max}}\right). \tag{4.102}$$

We manually add the leading factor of 2 to convert from the Fourier transform in $f \in (-\infty, \infty)$ to the conventional $f \in (0, \infty)$ for PSDs. More detailed discussions may be found in (Siegel, 2016).

We purposely carry the volume factor V throughout the derivation, which might appear unnecessary since the S_{21} signal only depends on the density of QPs. For the GR noise spectrum, instead of using $\Gamma_{\rm GR}$ that accounts for the total number of GR events in the entire SC film volume, the signal concerns the QP density and therefore the GR rate density

$$\gamma_{\rm GR} = \frac{\Gamma_{\rm GR}}{V}, \tag{4.103}$$

which then modifies Eq. (4.101) to

$$S_{\gamma_{\rm GR}} = \frac{S_{\Gamma_{\rm GR}}}{V^2} = \frac{4}{V}\left(Rn_{\rm qp,tot}^2 + \frac{n_{\rm qp,tot}}{\tau_{\rm qp,max}}\right). \tag{4.104}$$

Eq. (4.104) shows that, although it might appear unphysical at first sight, the GR noise as read out by a KID does depend on its absolute physical volume, therefore the unvanished V term in Eq. (4.104). We will interpret the physical meaning of such dependence later when we are ready. We can now apply the GR noise PSD to realistic situations, but before making numerical predictions, (Siegel, 2016) points out that, limited by the finite QP lifetime, the QP system can not oscillate arbitrarily fast in response to GR events that occur arbitrarily close in time. Such a natural lagging presents a “screening” effect for high-frequency GR occurrence and results in a single-pole “low-pass filtering” for $\delta n_{\rm qp}$ by

$$\delta\tilde{n}_{\rm qp}(f) = \frac{\tau_{\rm qp,eff}}{1 + i2\pi f\tau_{\rm qp,eff}}\sum_i \gamma_{{\rm qp},i}. \tag{4.105}$$

The detailed algebra for deriving Eq. (4.105) is found in (Siegel, 2016). Eq. (4.105) sums over all sources i for QP generation rate densities $\gamma_{{\rm qp},i}$, and $\tau_{\rm qp,eff}$ may be determined by

$$\tau_{\text{qp,eff}} d\Gamma_{\text{qp}} = V dn_{\text{qp}} \tag{4.106}$$

$$\Rightarrow \frac{1}{\tau_{\text{qp,eff}}} = 2Rn_{\text{qp,tot}} + \frac{1}{\tau_{\text{qp,max}}} \tag{4.107}$$

following Eq. (4.100), which finally yields the general QP GR noise PSD for our interest,

$$S_{\delta n_{\text{qp,tot}}} = \frac{\tau_{\text{qp,eff}}^2}{1 + (2\pi f \tau_{\text{qp,eff}})^2} \sum_i S_{\gamma_i}. \tag{4.108}$$

We have stated that we expect the GR noise to be the only type of noise in our system that is contributed by QPs, so we simply combine Eq. (4.108) and Eq. (4.104) to be

$$S_{\delta n_{\text{qp,tot}}} = \frac{\tau_{\text{qp,eff}}^2}{1 + (2\pi f \tau_{\text{qp,eff}})^2} \times \frac{4}{V} \left(R n_{\text{qp,tot}}^2 + \frac{n_{\text{qp,tot}}}{\tau_{\text{qp,max}}} \right), \tag{4.109}$$

which gives a white spectrum of power density $\left|S_{\delta n_{\text{qp,tot}}}\right|$ up to the QP cutoff frequency $f_{\text{qp,cut}}$. Note that we use the notation of $\left|S_{\delta n_{\text{qp,tot}}}\right|$ not as the literal absolute value of Eq. (4.109) but for representing the constant white PSD height to be derived below. Based on the fact that we currently observe $\tau_{\text{qp}} \lesssim O(10^2)$ μs, while $\tau_{\text{qp}} \sim$ ms has been demonstrated in multiple occasions in the literature, we conclude, for our current setup,

$$\tau_{\text{qp,eff}} \ll \tau_{\text{qp,max}} \;\Leftrightarrow\; \tau_{\text{qp,eff}} \approx \frac{1}{2Rn_{\text{qp,tot}}}, \tag{4.110}$$

i.e., the generation-recombination is currently dominated by an unbuffered random process given by Eq. (4.71) and therefore follows the physics assumptions for Eq. (4.64). We use this simplified $\tau_{\text{qp,eff}}$ to reorganize Eq. (4.109) into a cleaner form

$$S_{\delta n_{\text{qp,tot}}} = \frac{\left|S_{\delta n_{\text{qp,tot}}}\right|}{1 + (2\pi f / f_{\text{qp,cut}})^2}, \tag{4.111}$$

where

$$\begin{cases} \left|S_{\delta n_{\text{qp,tot}}}\right| & = \tau_{\text{qp,eff}}^2 S_{\gamma \text{GR}} \approx \dfrac{1}{RV} \\ f_{\text{qp,cut}} & = \dfrac{1}{\tau_{\text{qp,eff}}} \approx 2Rn_{\text{qp,tot}}; \end{cases} \tag{4.112}$$

notice that we keep $f_{\text{qp,cut}}$ in Hz with the factor of 2π. The above result gives a quick estimate for the total GR noise power

$$\delta n_{\text{qp,tot}} \approx \sqrt{f_{\text{qp,cut}} \cdot \left|S_{\delta n_{\text{qp,tot}}}\right|} \approx O(1) \times \sqrt{\frac{n_{\text{qp,tot}}}{V}}. \tag{4.113}$$

It is satisfying to see that, starting from the simple Poissonian assumption, we have arrived at the expected relation

$$\delta n_{\text{qp,tot}} \propto \sqrt{n_{\text{qp,tot}}} \tag{4.114}$$

for a general random process. Interestingly, Eq. (4.112) shows that the PSD is a constant that only depends on the recombination constant, an intrinsic property of the material that presumably varies negligibly, but it is the noise bandwidth, or the cut-off frequency, that introduces the QP density dependence into the noise power. In fact, one may rigorously include the roll-off of the filtered spectrum and show that

$$\delta n^2_{\text{qp,tot}} = \int_0^\infty S_{\delta \text{n}_{\text{qp,tot}}} df = \frac{n_{\text{qp,tot}}}{V}, \tag{4.115}$$

i.e., the loose $O(1)$ factor in Eq. (4.113) is exactly 1. However, (Siegel, 2016) speculates that there does exist possible realistic conditions that may introduce the $O(1)$ factor in Eq. (4.113), such as the QP nonuniformity or the current density nonuniformity discussed in Sec. 3.5. So, we think Eq. (4.113) is appropriately presented considering practical uncertainties. We can now interpret the confusing volumetric dependence for $S_{\gamma_{\text{GR}}}$ in Eq. (4.104). In reality, when we observe the resonator's surface inductance change due to QP *density* fluctuation, the QPs are created simultaneously in the entire film volume. We can clarify the concept by multiplying V to Eq. (4.113) on both sides to recover

$$\delta N_{\text{qp,tot}} \approx O(1) \times \sqrt{N_{\text{qp,tot}}}\,, \tag{4.116}$$

where the V dependence vanishes, showing the GR noise does follow the exact Poissonian fluctuation. However, it is for the total number of QPs in the KID, not the density. One may understand it as sampling the QP density via the S_{21} signal, but instead of probing an unit volume, the sampling is done simultaneously for the entire KID volume (V unit volume), so the variation is suppressed by $\sqrt{V}$ and therefore the factor of $1/\sqrt{V}$ in Eq. (4.113).

We proceed to the numerical predictions now that we have fully understood the model. Using the R_{Al} in Eq. (4.50), $\tau_{\text{qp,eff}} \approx 100\ \mu$s, and the typical

$$V \approx 3.5 \times 10^4 \qquad [\mu\text{m}^3] \tag{4.117}$$

for the KID inductors in the 80-KID prototype detector, Eq. (4.112) and Eq. (4.113) yield

$$\begin{aligned} \left|S_{\delta \text{n}_{\text{qp,tot}}}\right| &\approx 3 \times 10^{-6} \quad \left[(\text{QP}/\mu\text{m}^3)^2/\text{Hz}\right] \\ \delta n_{\text{qp}} &\approx 7 \times 10^{-2} \quad [\text{QP}/\mu\text{m}^3] \\ n_{\text{qp,tot}} &\approx 2 \times 10^{2} \quad \left[\text{QP}/\mu\text{m}^3\right], \end{aligned} \tag{4.118}$$

a result that is consistent with our expectation from the RF pulsing calibration. We convert the QP GR noise to its equivalent noise power in S_{21} for easier comparisons with other noise sources, for which we compute

$$\left|S_{\delta \mathrm{S}_{21}}\right| = \left(\frac{\partial \delta S_{21}}{\partial \delta n_{\mathrm{qp}}}\right)^2 \left|S_{\delta \mathrm{n}_{\mathrm{qp,tot}}}\right|. \tag{4.119}$$

The squared conversion factor is obtained with Eq. (4.5) and Eq. (4.22) by chain rule, which gives (cont'd, see Appx. A for detailed algebra)

$$= \left|\underbrace{\alpha|\gamma|\kappa\frac{Q_r^2}{Q_c}}_{S_{21}/n_{\mathrm{qp}}} \cdot \underbrace{\left(\frac{Q_r}{Q_c}\chi_{\mathrm{qp}}\right)}_{\approx 1}\right|^2 \left|S_{\delta \mathrm{n}_{\mathrm{qp,tot}}}\right|. \tag{4.120}$$

For future reference, we include

$$\chi_{\mathrm{qp}} = \frac{Q_i}{Q_{i,\mathrm{qp}}} \leq 1 \tag{4.121}$$

to represent the QP generation efficiency for the internal loss following (Zmuidzinas, 2009). We nevertheless take the whole parenthesized term containing χ_{qp} to be unity as noted below, based on our Q_c-dominated Q_r and assuming $\chi_{\mathrm{qp}} \sim 1$. Using $|\gamma| = 1$, $\alpha \lesssim 0.1$, $Q_r \approx Q_c \approx 10^5$ following Eq. (4.26), Eq. (4.47), Eq. (4.80) and their corresponding discussions, respectively, and $\kappa \lesssim 5 \times 10^{-7}$ given by Eq. (4.28) using typical Al parameters in Eq. (4.92), we obtain

$$\left|S_{\delta \mathrm{S}_{21}}\right| \lesssim 8 \times 10^{-11} \quad [(V_{\mathrm{out}}/V_{\mathrm{in}})^2/\mathrm{Hz}]. \tag{4.122}$$

Note that, due to the unit of n_{qp}^2 for $S_{\delta n_{\mathrm{qp}}}$, $S_{\delta S_{21}}$ is automatically in the desired unit of S_{21}^2.

We compare the natural GR noise bandwidth $f_{\mathrm{qp,cut}}$ given by the QP lifetime to the designed resonator noise bandwidth,

$$\Delta f_r = \frac{\pi f_r}{Q_r} \gtrsim 1 \times 10^5 \gg f_{\mathrm{qp,cut}}, \tag{4.123}$$

and find that the GR noise spectrum is fully included. Assuming our typical input power for the prototype detector

$$P_g \lesssim -60 \text{ dBm} \quad (10^{-9} \text{ W}), \tag{4.124}$$

we estimate the power fluctuation at the output of the resonator due to spontaneous QP GR, in terms of its equivalent noise temperature $T_{n,\mathrm{GR}}$, by

$$k_{\mathrm{B}} T_{n,\mathrm{GR}} \Delta f_r \approx \left[\frac{Q_r^2}{Q_i^2} P_g\right] \cdot \left|S_{\delta S_{21}}\right| \frac{f_{\mathrm{qp,cut}}}{2\pi}, \tag{4.125}$$

which rearranges to the equivalent noise temperature for the QP GR noise

$$\begin{aligned} T_{n,\mathrm{GR}} \lesssim\ & 0.9\ \mathrm{K} \\ & \times \left[\left(\frac{\alpha}{0.1}\right)^2 \left(\frac{|\kappa|}{5\times 10^{-7}\ \mu\mathrm{m}^3}\right)^2 \left(\frac{Q_r \approx Q_c}{10^5}\right)^2 \left(\frac{Q_i/Q_r}{10}\right)^{-2} \left(\frac{P_{\mathrm{read}}}{10^{-9}\ \mathrm{W}}\right)^1\right. \\ & \left.\left(\frac{f_{\mathrm{qp,cut}} \triangleq \tau_{\mathrm{qp,eff}}^{-1} \propto n_{\mathrm{qp,tot}}}{10^4\ \mathrm{Hz}}\right)^1 \left(\frac{\Delta f_r}{10^5\ \mathrm{rad}\cdot\mathrm{Hz}}\right)^{-1} \left(\times\frac{2\pi\Delta f_r}{f_{\mathrm{qp,cut}}}\right)^{\text{see later}}\right]. \end{aligned} \tag{4.126}$$

There are important remarks to be made for the above forecast. First, instead of using the off-resonance readout power P_{read} as the transmitted power, we consider the on-resonance power transmission that is attenuated by $\left(Q_r^2/Q_i^2\right)$ as shown by the bracket in Eq. (4.125); this is the appropriate power that we utilize for observation at the output of the resonator. The derivation for the $\left(Q_r^2/Q_i^2\right)$ factor is given in Appx. A, where we also assume the transmission-line loss α and the coupling phase ϕ_c are negligible. Second, to make the prediction more accurate, instead of using $f_{\mathrm{qp,cut}} = \tau_{\mathrm{qp}}^{-1}$ for the GR noise bandwidth, we keep the factor of 2π so the GR noise bandwidth is consistent with the conventional 3 dB-attenuated bandwidth definition. Third, although Eq. (4.125) seems to suggest that the GR noise behaves much like a thermalized resistor, exhibiting a white noise PSD of $k_{\mathrm{B}} T_n$ that extends to very high frequencies[9], one should keep in mind that the QP GR white spectrum only extends up to $f_{\mathrm{qp,cut}}$. We have also argued that the resonator noise bandwidth is much wider than $f_{\mathrm{qp,cut}}$, so the GR noise PSD practically only occupies the low-frequency region of the entire band of observation. In this case, one should not misinterpret that Eq. (4.125) suggests a true Johnson-Nyquest-like thermal spectrum of a PSD of $k_{\mathrm{B}} T_{n,\mathrm{GR}}$, but rather an *averaged* PSD by distributing the total GR noise power that concentrates in $f < f_{\mathrm{qp,cut}}$ to the wider band $f < \Delta f_r$. Therefore, we should present the GR noise temperature by

$$T_{n,\mathrm{GR}}(f) \approx \begin{cases} T_{n,\mathrm{GR}} \times \dfrac{2\pi\Delta f_r}{f_{\mathrm{qp,cut}}} & ,\ f < \dfrac{f_{\mathrm{qp,cut}}}{2\pi} \\ 0 & ,\ \text{otherwise,} \end{cases} \tag{4.127}$$

[9]Realistically up to the electron scattering/relaxation frequency in a resistor.

where the $T_{n,\mathrm{GR}}$ is the constant noise temperature obtained in Eq. (4.125). We will discuss later in our OF-based energy resolution analysis that such a concentrated noise is undesirable compared to a true uniform thermal spectrum, because in this case the OF is less effective in rejecting noise at higher frequencies. Fourth, besides the factor-of-unity correction for unknown nonuniformity effects, we collect all the assumed relevant values into Eq. (4.126), so the reader may rescale the forecast based on their own conditions. One may see that the 0.9-K forecast could easily vary by an order of magnitude due to reasonable adjustments to the assumed inputs. We emphasize again that, while we provide a prediction of power that uses the absolute value of κ, being a QP-generated noise, its corresponding δS_{21} trajectory obeys the characteristic θ_{qp}, so for the df (dQ) signal, the noise size is reduced by $\cos^2\theta_{\mathrm{qp}}$ ($\sin^2\theta_{\mathrm{qp}}$).

Fifth and finally, while the calculated equivalent GR noise presented as δS_{21} varies depending on many parameters, we are always most interested in the actual energy deposited in the KID for QP creation. From this perspective, Eq. (4.112) and Eq. (4.113) already show that the equivalent pair-breaking energy due to GR noise is simply

$$\delta E_{\mathrm{GR}} = V\delta n_{\mathrm{qp,tot}} \cdot \Delta \approx O(1) \times \Delta\sqrt{V n_{\mathrm{qp,tot}}}. \tag{4.128}$$

Substituting the same Δ_{Al} and V used previously, we find

$$\delta E_{\mathrm{GR,Al}} \approx 0.4\ \mathrm{eV}. \tag{4.129}$$

We again assume no spectral filtering for the GR noise because, one, it requires a much reduced Δf_r to evade the natural GR spectrum and, two, $f_{\mathrm{qp,cut}}$ per Eq. (4.112) is essentially the OF noise bandwidth if dominated by the QP lifetime, therefore making the OF ineffective in rejecting the GR noise (see later). On the other hand, Eq. (4.128) also shows that, as long as the equilibrated QP density is held constant, which presumably is the case when the environmental energy injection is stable and dominates the background QP generation, the GR noise with respect to the signal energy is suppressed by $\sqrt{V}$ for reduced KID form factors. It is expected since the pair-breaking signal scales linearly with the volume, while the GR noise increases with the square-root of the volume (Eq. (4.113)). Finally, if the environmental condition balancing $n_{\mathrm{qp,tot}}$ is improved, $\delta E_{\mathrm{GR}} \propto \sqrt{n_{\mathrm{qp,tot}}}$ always helps as anticipated.

4.5.3 HEMT noise: High-/low-temperature calibration

Having established the fundamental GR noise limit, we now study the artificial noise sources for comparison. We begin with the basic readout electronic noise, including the Ettus Research USRP DAQ system and the HEMT low-noise amplifier located at the immediate output of the detector at 4 K (Fig. 4.16). Since the HEMT amplifier contributes noise predominately through its amplification gain instability by orders-of-magnitude due to the high gain over its own phase-less thermal noise, we treat the HEMT as a pure amplitude noise source, shown by the "gain" direction in Fig. 4.19. In the $\mathcal{O}(1)$-GHz range, commercial HEMTs generally exhibit noise PSDs that are predominately white and structureless, so the manufacturers usually specify the noise levels by the noise temperatures, in the typical range of

$$T_{n,\mathrm{HEMT}} = 1.5 - 7 \text{ K} \tag{4.130}$$

with typical gains of

$$G_{\mathrm{HEMT}} = 25 - 40 \text{ dB}. \tag{4.131}$$

For the HEMT models that we acquired from Cosmic Microwave Technology (Inc., n.d.)[10], CMT quotes the noise temperatures of ≈2 K and ≈4 K with 37 dB and 27 dB gains, respectively. Comparing to the GR noise derived earlier (Eq. (4.126)) and also considering the ongoing facility upgrade (Sec. 4.7) that aims to reduce the background QP population thus also the GR noise, we anticipate that the HEMT noise to dominate the GR noise under nominal conditions. Therefore, we perform the work introduced in this section to obtain a more accurate, amplifier-specific, frequency-dependent gain and noise temperature calibration for each of our CMT HEMTs. To follow the discussion below more easily, the reader may reference our calibration result in Fig. 4.20 and the electronics diagram in Fig. 4.16.[11]

We measure the noise PSDs by a HP-Agilent 8563E spectral analyzer set to a bandwidth of 1 MHz. For each PSD spectrum, we average 200 measurements using the spectral analyzer's built-in function for noise suppression, which we find sufficient for allowing the PSD fluctuation to be limited by the irreducible fluctuation the spectral analyzer receives. It is unclear whether the irreducible fluctuation is due to the external fluctuation of the lab EM environment, the internal fluctuation of the data acquisition equipment, the circuit being measured, or the spectral analyzer's averaging algorithm based on the monitor-shown data curves; we suspect the last being the cause. We first terminate the input of the spectral analyzer with a 50-Ω terminator to measure its own instrumental noise. Knowing the terminator emits

[10] Obsolete models that cannot be found online.

[11] The room-temperature amplifier(s) utilized in Fig. 4.16 at the output of the cryogenic system is modified, see following discussion.

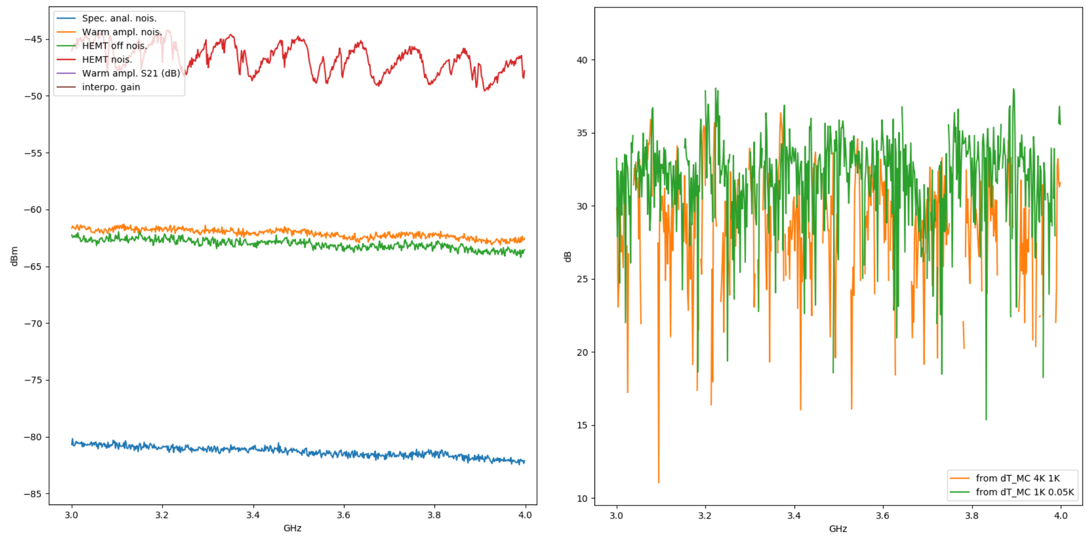


Figure 4.20: Left: From bottom to top, the noise PSD data for HP-Agilent 8563E spectral analyzer (blue), two ZVA-183-S+ room-temperature amplifiers connected to the cryostat (green), two ZVA-183-S+ terminated at room temperature (orange), CMT HEMT (red). Right: HEMT gain profiles determined independently by the 4 K/1 K (orange) and 1 K/50 mK (green) MC temperature differences. For this particular result, the ZVA amplifier gain is separately constrained by the PSD difference when the ZVA amplifiers are terminated at room temperature or by the cryostat at 4K. Detailed explanation is provided in the corresponding text.

thermal noise at a temperature of 297 K and using the Johnson-Nyquist noise PSD

$$P(f) = k_B T_n \Delta f_{BW}, \tag{4.132}$$

where $P(f)$ is the noise power in $(f, f + df)$, T_n is the (equivalent) temperature of the noise source, f_{BW} is noise bandwidth, we obtain

$$P_i(f) = k_B(297 + T_{n,8563E})\Delta f, \tag{4.133}$$

where $T_{n,8563E}$ is the frequency-dependent noise temperature of the spectral analyzer, and $\Delta f = 1$ MHz is the designated integration bandwidth. For the spectral analyzer's near-constant $P_i(f) \approx -82$ dBm/Hz, we find

$$T_{n,8563E} \approx 4.6 \times 10^5 \text{ K}. \tag{4.134}$$

If we directly apply the HEMT output to the spectral analyzer, we expect

$$P_{amp}(f) = G \times P(f), \tag{4.135}$$

where P_{amp} is the output power of an amplifier, G is its gain, and in this case the T_n in $P(f)$ is the equivalent noise temperature of the amplifier. Given the gain and noise temperature data provided by CMT, we anticipate adding an excess noise of $G_{\text{HEMT}} T_{n,\text{HEMT}} \sim \mathcal{O}(10^{\leq 4})$ K-equivalent to the 5×10^5-K spectral analyzer noise, which is unresolvable in the presence of measurement uncertainty.

So instead, we apply two Mini-Circuits ZVA-183-S+ room-temperature amplifiers in series to ensure that the HEMT noise dominates the spectral analyzer by 2–3 orders-of-magnitude. These ZVA room-temperature amplifiers are quoted with a ≈52 dB combined gain. Before utilizing the ZVA amplifiers to amplify the HEMT output, we again measure the added noise due to the ZVA amplifiers. We terminate the input of the ZVA amplifier pair with a 50-Ω terminator, and again assuming a 297 K room temperature, equal $T_{n,\text{ZVA}}$ and G_{ZVA} for *each* ZVA amplifier, we measure

$$P_{\text{ii}} = k_{\text{B}} \{ G_{\text{ZVA2}} [G_{\text{ZVA1}} (297 + T_{n,\text{ZVA1}}) + T_{n,\text{ZVA2}}] + T_{n,\text{8563E}} \} \Delta f. \tag{4.136}$$

We write the equation from the inner bracket to the outer bracket corresponding to the signal flow from upstream to downstream, cascading together the noise sources (temperatures) that are amplified by the corresponding instrumental gains. Note that, while we distinguish ZVA$\underline{1}$ and ZVA$\underline{2}$ in the equation to help with understanding the signal flow and the notation correspondance, in reality their values are similar and not measured separatedly. Per the equal $T_{n,\text{ZVA}}$ and G_{ZVA} assumption, we rearrange the above equation to (cont'd)

$$= k_{\text{B}} \{ [G_{\text{ZVA}}^2 (297 + T_{n,\text{ZVA}}) + \cancel{G_{\text{ZVA}} T_{n,\text{ZVA}}}] + T_{n,\text{8563E}} \} \Delta f. \tag{4.137}$$

Eq. (4.137) shows that the second ZVA amplifier is subdominant by a factor of $G_{\text{ZVA}} \approx 10^{2.6}$ relative to the preceding one and thus is negligible in practice. As shown in Fig. 4.20, we measure $P_{\text{ii}} \approx -62$ dBm/Hz, which, assuming the manufacturer-specified 26 dB gain for each ZVA amplifier, yields

$$T_{n,\text{ZVA}} \approx 286 \text{ K}. \tag{4.138}$$

The obtained noise temperature is consistent with the data-sheet value based on the specified 26 dB amplifier gain. In the following, we will show that our calibration methodology can in principle simultaneously determine G_{ZVA} and $T_{n,\text{ZVA}}$ without introducing predefined values or assumptions/approximations.

Next, we add the HEMT in front of the room-temperature readout, now that all the supporting instruments are modeled and calibrated. Although we heat-sink the HEMT amplifier to 4 K, its input connects to the DMKID device that is at the mixing chamber (MC) temperature T_{MC} of the DR, which varies between 50 mK, 1 K, and

4 K at different stages in the refrigerator operation. Since the upstream noise before the DMKID device is highly attenuated before the device, we assume that, when ignoring near-resonance data, the MC thermal noise dominates the noise input of the HEMT. We therefore write

$$P_{\text{iii}} = k_{\text{b}}[\![G_{\text{ZVA}}\{G_{\text{ZVA}}[G_{\text{HEMT}}(T_{n,\text{MC}}+ \\ T_{n,\text{HEMT}}) + T_{n,\text{ZVA}}] + T_{n,\text{ZVA}}\} + T_{n,8563\text{E}}]\!]\Delta f \quad (4.139)$$

based on the same concept of Eq. (4.136). Eq. (4.139) contains four unknowns: G_{HEMT}, G_{ZVA}, T_{HEMT}, and T_{ZVA}, while $T_{n,8563E}$ has been separately determined in Eq. (4.133). We also have four linearly independent equations, one from the ZVA measurement Eq. (4.136) and three from Eq. (4.139) when $T_{\text{MC}} = 50$ mK, 1 K, and 4 K, so mathematically all the unknowns are solvable. In fact, one may understand the conclusion by either comparing Eq. (4.136) and Eq. (4.139), or simply understanding that, physically, the experimental configurations for the ZVA amplifier measurement and the HEMT noise measurement only differ by the injected signal for the ZVA pair. The former utilizes a room-temperature terminator, while the latter accepts the cryogenic HEMT system, i.e., suggesting that Eq. (4.136) and Eq. (4.139) would be equal if the HEMT-amplified cryogenic noise contribution is equal to the 50-Ω terminator at room temperature, or

$$P_{\text{ii}} = P_{\text{iii}} \Leftrightarrow G_{\text{HEMT}}(T_{n,\text{MC}} + T_{n,\text{HEMT}}) = 297. \quad (4.140)$$

In practice, since we already know that $G_{\text{HEMT}} \sim 10^3$ and $(T_{n,\text{MC}} + T_{n,\text{HEMT}}) \approx 5$–10 K, while $T_{n,\text{ZVA}} \approx 140 - 230$ K, which is only a few % of the expected HEMT-amplified equivalent noise temperature $[G_{\text{HEMT}}(T_{n,\text{MC}} + T_{n,\text{HEMT}})]$, so the calibration precision (relative accuracy) for $T_{n,\text{HEMT}}$ and G_{HEMT} effectively only depends on the robustness of the $T_{n,\text{MC}}$ data, but their accuracies (absolute values) do scale linearly with the calibrated $T_{n,\text{ZVA}}$ and G_{ZVA}.

Algebraically, the above analysis suggests that we may differentiate the P_{iii} equations for different MC temperatures and use these known temperature differences to constrain the non-HEMT-related terms, e.g.,

$$P_{\text{iii}}|_{4\text{K}} - P_{\text{iii}}|_{1\text{K}} = k_{\text{B}}G^2_{\text{ZVA}}G_{\text{HEMT}}(4-1)\Delta f \quad (4.141)$$

$$\Leftrightarrow$$

$$k_{\text{B}}G^2_{\text{ZVA}}\Delta f = \frac{P_{\text{iii}}|_{4\text{K}} - P_{\text{iii}}|_{1\text{K}}}{(4-1)G_{\text{HEMT}}}, \quad (4.142)$$

and then use the 297 K terminator as a "standard candle" of thermal noise power to subsequently calibrate the absolute scale of the HEMT-related terms (P_{ii}), e.g.,

$$P_{\mathrm{iii}}|_{4\mathrm{K}} - P_{\mathrm{ii}} = k_{\mathrm{B}} G_{\mathrm{ZVA}}^2 [G_{\mathrm{HEMT}}(4 + T_{n,\mathrm{HEMT}}) - 297]\Delta f \tag{4.143}$$

$$\Leftrightarrow$$

$$\begin{cases} \dfrac{P_{\mathrm{iii}}|_{4\mathrm{K}} - P_{\mathrm{ii}}}{P_{\mathrm{iii}}|_{4\mathrm{K}} - P_{\mathrm{iii}}|_{1\mathrm{K}}} = \dfrac{G_{\mathrm{HEMT}}(4 + T_{n,\mathrm{HEMT}}) - 297}{(4-1)G_{\mathrm{HEMT}}} \\ \dfrac{P_{\mathrm{iii}}|_{1\mathrm{K}} - P_{\mathrm{ii}}}{P_{\mathrm{iii}}|_{1\mathrm{K}} - P_{\mathrm{iii}}|_{50\mathrm{mK}}} = \dfrac{G_{\mathrm{HEMT}}(1 + T_{n,\mathrm{HEMT}}) - 297}{(1-0.05)G_{\mathrm{HEMT}}} \end{cases}. \tag{4.144}$$

This technique is internally dubbed the "high-/low-temperature calibration" for its comparison-based logic. Note that the three different $T_{n,\mathrm{MC}}$ only provide two linearly independent temperature differences as shown in Eq. (4.144), therefore two independent constraints that minimally satisfy our need for solving the equations. We realize from the above simultaneous equations that, with more datasets taken at different $T_{n,\mathrm{MC}}$, we may over-constrain $T_{n,\mathrm{HEMT}}$ and G_{HEMT}. In practice, the over-constraint is equivalent to suppressing the measurement uncertainty across datasets, or effectively "fitting" for the best representations of $T_{n,\mathrm{HEMT}}(f)$ and $G_{\mathrm{HEMT}}(f)$. However, due to our choice for the HEMT noise to dominate the ZVA amplifiers for a more robust HEMT calibration, this technique is not as accurate for calibrating the ZVA amplifiers. One may examine the effect by substituting the obtained $T_{n,\mathrm{HEMT}}$ and G_{HEMT} into Eq. (4.143) to compute G_{ZVA}. Due to the large difference of their noise powers, the left-hand side of the equation is dominated by P_{iii}, while the right-hand side is also dominated by the amplified HEMT noise temperature term over room temperature, reflecting that the configuration is designed for calibrating the HEMT but is inaccurate for the ZVA amplifiers. Unless the obtained HEMT parameters are highly accurate, the ZVA-related terms are only of fractional sizes of the HEMT terms and would be severely smeared by the uncertainties in $T_{n,\mathrm{HEMT}}$ and G_{HEMT}.

Understanding that the result is physically due to calibrating the P_{ii} room-temperature system using a much larger effective temperature $G_{\mathrm{HEMT}} \times T_{n,\mathrm{HEMT}}$, we realize that it is more straightforward to simply calibrate the ZVA amplifiers by varying the temperature of the calibration source in a comparable range of $T_{n,\mathrm{ZVA}}$, hence not only avoiding a high relative uncertainty but also fully decoupling the P_{ii} and P_{iii} equations. So, in addition to thermalizing the terminator at room temperature, one may as well cool it to the liquid nitrogen temperature (77 K) or other available temperatures and then carry out the calibration identically to the HEMT high-/low-temperature

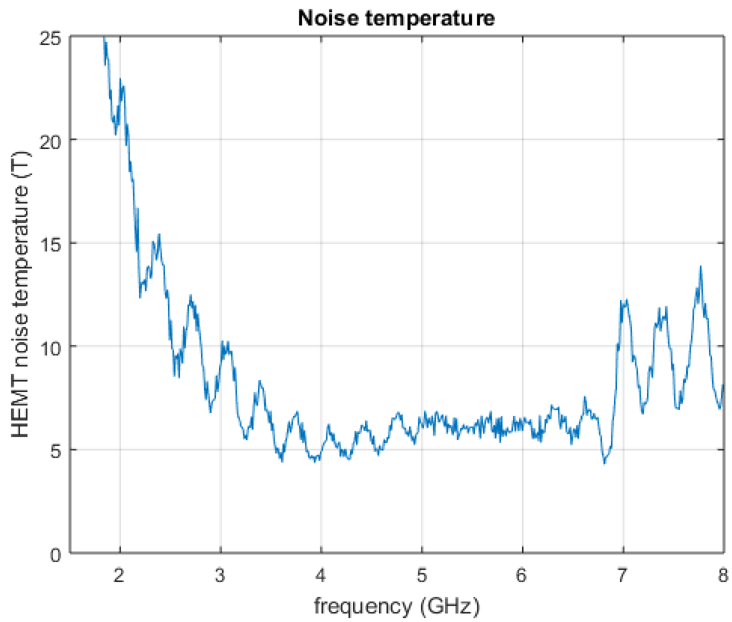


Figure 4.21: The improved HEMT noise temperature measurement by O. Wen.

calibration. We therefore choose to present the result given by this independent high-/low-temperature calibration for the ZVA amplifiers in Fig. 4.20. However, since the calibration technique we develop in this section was originally inspired by a dataset acquired for a different purpose at the 4 K standby, 1 K ^{3}He/^{4}He-mix condensing, and the 50 mK base temperature phases of an usual DR operation, we had to use the 4 K standby-phase noise emission from the entire cryogenic system as the low-temperature calibration source for the ZVA amplifiers. Due to the fact that the $T_{n,\mathrm{MC}}$ data, especially for the mix-condensing phase, were only recorded at a $\approx$70% accuracy, the dataset did not yield useful numerical results surpassing the CMT calibration. We nevertheless demonstrate that the methodology is valid with a noisy but consistent result in Fig. 4.20. O. Wen later repeated a dedicated calibration, where a carefully designed thermal-mechanical system was used with much more data points taken to over-constrain the system with linear fits. The much improved result by O. Wen is shown in Fig. 4.21.

4.5.4 Combined electronic noise

We now combine the HEMT noise with other components in the full readout chain. First of all, we should point out that defining the noise temperature T_n for an amplifier based on the Johnson-Nyguist noise (Eq. (4.132)) does not imply that the amplifier exhibits a physical temperature of T_n. On the contrary, one generally needs to model the equivalent circuit as a perfectly thermalized resistor that emits the classical

Johnson-Nyguist noise at T_n, in parallel with the main signal amplification chain that is noiseless. In reality, since amplifier users normally expect good products to yield much larger signals over the environmental backgrounds, in other words large gains, such an equivalent T_n is usually more directly related to the intrinsic noise sources of the physics principle that achieves the amplification but not the device's physical temperature, otherwise it is probably by-definition a bad amplifier. Taking the HEMT as an example, it generates noise predominately due to the gain instability that is derived from the unstable power supply controlling the gain, so we were able to conclude that the HEMT noise is always in-phase with the amplified signal thus aligns to the gain direction in Fig. 4.19. The argument also explains that amplification mechanism dependence naturally leads to frequency-dependent noise temperatures and gains.

Understanding that the noise temperature for an amplifier is not physical but rather artificially defined based on the noise power $P_{n,\text{amp}}$ and gain G_{amp} helps in contrasting with the attenuator model, which is oftentimes confusing especially for people who falsely expect $T_{n,\text{amp}}$ to physically exist in the amplifier. For an amplifier to output an energy larger than the input signal, the process only conserves energy with the addition of an active energy supply. The supplied energy is turned into the output energy by the amplification mechanism in a correlated fashion with the input signal. So, since energy-wise the input signal contributes negligibly to the output at large gains, instead of picturing physically multiplying the input energy to be the output, it is more appropriate to understand an amplifier as outputting the strong energy supply using the input signal as a control throttle. One may then appreciate that the output noise of an amplifier is caused by the combined instability of the throttling and the energy supply but just equated to an equivalent noise temperature. The picture is very different for an attenuator. Being a *passive* component, we use the attenuator's natural loss to spontaneously "leak" the real input energy from the signal to achieve attenuation. As a physical object in thermal equilibrium with the environment, a passive attenuator emits EM radiation as noise by the exact thermal process underlying the Johnson-Nyquist noise. Contrasting an amplifier, the noise power of an attenuator is therefore strictly correlated with its physical temperature.

We again use an equivalent thermal radiator at the input of the attenuator, but as opposed to the amplifier model, we do know at the input of the attenuator, there is a real space that thermalizes at a physical temperature T_0 and radiates at $k_B T_0$ [W/Hz]. This emission travels through the attenuator and is attenuated by the attenuation factor A and presents the remaining power $k_B T_0/A$ [W/Hz] to the space at the exit of the attenuator. Empirically, we know that the space at the exit accepts the attenuated radiation while equilibrates with the environment to be at T_0, a physical temperature, through other heating and dissipation channels. Energy conservation suggests that the attenuator absorbs the power

$$dP = k_B T_0 \left(1 - \frac{1}{A}\right) df. \tag{4.145}$$

The fact that the attenuator is at a thermal equilibrium requires this power that does not participate in thermalizing the attenuator to be presented as the noise power to the downstream. We consider a similar circuit model as the amplifier, consisting of a thermal radiator of an equivalent noise temperature $T_{n,\mathrm{att}}$ *at the input* and a noiseless attenuator of an attenuation factor A, so we write

$$\frac{k_B T_{n,\mathrm{att}}}{A} df = dP = k_B T_0 \left(1 - \frac{1}{A}\right), \tag{4.146}$$

which gives

$$dP_{n,\mathrm{att}}(f) = G_{\mathrm{att}}(k_B T_{n,\mathrm{att}}) df \tag{4.147}$$

with

$$\begin{cases} T_{n,\mathrm{att}} = T_0(A-1) \\ G_{\mathrm{att}} = A^{-1} \leq 1, \end{cases} \tag{4.148}$$

where $P_{n,\mathrm{att}}$ is the noise power emitted at the output of the attenuator, G_{att} is the "gain" of the attenuator that we define by A^{-1} for later notation consistency, and $T_{n,\mathrm{att}}$ is the equivalent noise temperature *at the input* of the attenuate, which is determined by the attenuator's physical temperature T_0 and the attenuation A at thermal equilibrium. Similar to the amplifier, the attenuation A (G_{att}) also depends on frequency according to the attenuation mechanism. Eq. (4.148) exhibits

$$dP_{n,\mathrm{att}} \rightarrow \begin{cases} k_B T_0 df & , A \gg 1 \\ 0 & , A \ll 1 \end{cases}, \tag{4.149}$$

which is exactly what we expect: When the attenuation is very large and absorbs all the input energy, the input energy is "terminated" and converted into the thermal noise $k_B T_0$ of a thermalized resistor/terminator. For very small attenuation, the signal is transmitted losslessly therefore does not dissipate energy to excite the construction of the "transmission line" to radiate noise. We now realize that the attenuator model applies not only to practical attenuators but also general passive devices that may or may not be lossy.

We obtain the general formula for all the devices in the readout chain,

$$dP_n(f) = G(f) \cdot k_B T_n(f) \cdot df, \tag{4.150}$$

where we drop the subscript distinguishing amplifiers and attenuators, knowing that for active components we may either calibrate the noise temperatures and gains as done for the HEMT or refer to manufacturer data. For passive components, Eq. (4.148) always holds at thermal equilibrium with the attenuation (insertion loss) given by the data sheets. More importantly, the unified equation treats every component as an input-to-output transducer with a well-defined transfer function $G(f)$ and an extra thermal source $k_B T_n(f)$ at the input, so we have

$$|S_{\mathrm{out}}(f)|^2 = G(f) \left[|S_{\mathrm{in}}(f)|^2 + k_B T_n(f) \right], \tag{4.151}$$

where S is the input/output signal spectral density as denoted by the subscript. When considering only the readout apparatus for modeling the baseline readout noise before installing the detector, for any specific component in the chain, its input signal S_{in} is basically the modulated combined noise from all upstream components. We label the components from the upstream to the downstream by $j = 1, 2, 3, ...,$ so for each upstream component $j = x$, its noise contribution to the input of the component $j = y$ in discussion is

$$dP_x = \left(\prod_{j=x}^{y-1} G_j \right) k_B T_{n,x} df. \tag{4.152}$$

The total input noise PSD for the y-th component, not including its own noise generation, is

$$\left|S_{\mathrm{in},y}\right|^2 = \sum_{x=1}^{y-1} \frac{dP_x}{df} \tag{4.153}$$

$$= k_B \times \left[\!\left[G_{y-1} \left\{ ... G_3 \left[G_2 \left(G_1 T_{n,1} + T_{n,2} \right) + T_{n,3} \right] ... + T_{n,y-2} \right\} + T_{n,y-1} \right]\!\right], \tag{4.154}$$

or sometimes more convenient to use the equivalent noise temperature at the input by

$$T_{\mathrm{in},y} = \frac{\left|S_{\mathrm{in},y}\right|^2}{k_B} \tag{4.155}$$

$$= G_{y-1} \left\{ ... G_3 \left[G_2 \left(G_1 T_{n,1} + T_{n,2} \right) + T_{n,3} \right] ... + T_{n,y-2} \right\} + T_{n,y-1}. \tag{4.156}$$

Eq. (4.156) has exactly the same structure of Eq. (4.139) and may be understood identically: From the inner most bracket to the outer layers, each layer represents that the input signal combines with the component's own noise and is subsequently transduced together by the component's gain, and then the combined output travels to the next component/bracket layer to continue as its input. We concluded previously that we are generally able to flexibly adjust the amplification and attenuation configuration after the HEMT so to ensure that the noise the DAQ receives is HEMT-dominated. We now realize that we should revise the statement to "by the combined output of" the HEMT, which includes the possibility that the HEMT output nonnegligibly contains the upstream noise entering the HEMT with its own noise.

Based on Eq. (4.156) and the circuit diagrams provided by Ettus Research, A. Villalpando is able to breakdown the transmitter (Tx) readout signal-generation electronics of our USRP (Research, n.d.) DAQ system to individual RF components, collect the components' $G(f)$ and $T_n(f)$ specified by the manufacturers, and together with the attenuators installed at different temperatures in the DR (Fig. 4.16), predict the T_{in} at the inputs of the DR and the HEMT. Without a DMKID device, the expected equivalent noise temperatures are

$$\begin{aligned} T_{\text{in,DR}} &\approx 5.4 \times 10^5 \text{ K} \\ T_{\text{in,HEMT}} &\approx 43 \text{ K}, \end{aligned} \tag{4.157}$$

in which the programmable Tx amplifier set in CBX-120, consisting of a programmable 0–31.5-dB attenuator and two constant-gain amplifiers providing an equal counter gain, contributes the dominant 99.1% and 98.2% of the noise powers for the DR and the HEMT cases, respectively. Considering our HEMT exhibits a 5–10 K noise temperature in 3–4 GHz (Fig. 4.21), Eq. (4.157) seems to suggest that the noise observed after the HEMT is dominated by the upstream electronics but not the HEMT itself. However, when the readout electronics are run with a DMKID detector, where only on-resonance transmission is utilized for detection, the KIDs also contribute the $(Q_r/Q_i)^2 \sim 10^{-2}$ resonator attenuation factor derived in Eq. (4.125). Therefore, instead of having a 43 K upstream noise, we predict[12]

$$T_{\text{in,HEMT}} \rightarrow \left(\frac{Q_r}{Q_i}\right)^2 \times T_{\text{in,HEMT}} \sim 0.4\text{K}, \tag{4.158}$$

[12] Note for future readers: This contradicts A. Villapando's preliminary conclusion.

which is subdominant to the HEMT noise. Note that, since this upstream noise injection is dominated by the passive devices[13], which contribute the noise via a random thermal process independently to the signal, the noise energy distributes equally in both the phase and gain directions in Fig. 4.19. It however does not alter the expectation of observing a gain-dominated noise after the HEMT, since the output noise is dominated by the HEMT but not the upstream thermal noise.

In addition to the phase-less thermal noise, we notice that A. Villapando's work only considers the front-end electronics of the Tx circuit and the input-side cryogenic components. However, due to the practical need for the up-/down-mixing of the RF signal with the GHz local oscillator (LO) to/back from the resonant frequencies, we anticipate a nonnegligible additional phase noise due to the phase misalignment of the up- and down-conversions. According to the data sheet for the MAX2870 synthesizer the CBX-120 RF daughterboard is equipped with (Research, n.d.), the synthesizer exhibits a phase noise of -140 to -120 dBc/Hz. Given than the RF generator has a full range of ± 1 V, we anticipate the maximal mixer-generated phase noise of

$$dP_{n,\mathrm{LO}} = \frac{(1\ \mathrm{V})^2}{2 \cdot 50\ \Omega} \times 10^{-14 \sim -12} df \tag{4.159}$$

$$= 10^{-16 \sim -14} df\ [\mathrm{W}] \tag{4.160}$$

at the output of the mixer. We convert the estimate into an equivalent noise temperature of

$$T_{n,\mathrm{LO}} = \frac{dP_{n,\mathrm{LO}}/df}{k_\mathrm{B}} \approx 7.2 \times 10^{6 \sim 8}\ \mathrm{K}. \tag{4.161}$$

Notice that this synthesizer-generated noise is due to pure phase instability and therefore fluctuates in the electronic phase noise direction. We include this phase noise into A. Villapando's preliminary work and, after attenuated by the Tx front-end and cryogenic electronics, find a phase noise-dominated HEMT-input noise temperature

$$T_{\mathrm{in,HEMT}} \sim 0.7 - 22\ \mathrm{K}. \tag{4.162}$$

Unlike our previous $T_{\mathrm{in,HEMT}} \approx 0.4$ K prediction that concerns only the phase-less thermal noise sources, the revised prediction in Eq. (4.162) is due to the LO phase instability and therefore is in the electronic phase noise direction.

In summary for the combined electronic noise study, we predict that the noise at the output of the DR, in the presence of the detector, is dominated by the HEMT

[13] Modulo the complexity that its variability is also contributed by other two amplifiers in the Tx circuit, which contribute noise in the gain direction.

gain instability over the USRP DAQ variable attenuator in the electronic amplitude (gain) noise direction, while in the electronic phase noise direction, the noise is dominated by the synthesizer-generated phase noise over the USRP attenuator. The HEMT noise PSD is predominately white, while the mixer phase noise follows the $1/f$ PSD due to the phase misalignment mechanism thus increases toward lower frequencies. Finally, for the background QP density given by the current experimental construction, we expect the amplitude of the QP GR noise to be smaller than the HEMT-dominated gain-direction noise by an order-of-unity, therefore potentially observable, but the effective mixing of the two depends on the relative angle of θ_{qp} and ϕ_c.

4.5.5 Correlated noise removal

The fact that the electronic noise and the KID-generated signal (or GR noise) apply distinctly to the global or the local resonance circle coordinates, respectively, inspires us to attempt to estimate and subsequently remove the electronic noise embedded in the QP signal. We anticipate it to be viable based on comparing the on- and off-resonance signals. Leveraging the distinction that the high-quality resonators only participate in the transmission within their narrow frequency bands

$$\Delta f \lesssim f_r/Q_r \sim O(10^{4-5})\ \mathrm{Hz}, \tag{4.163}$$

while wide-band readout electronics typically exhibit frequency dependence at the scales comparable to their suitable frequency ranges, i.e.,

$$\frac{dG(f)}{df} \sim \frac{O(1)}{f_r} \tag{4.164}$$

$$\Rightarrow\ \Delta f \lesssim O(f_r) \sim O(10^{8-9})\ \mathrm{Hz}, \tag{4.165}$$

we realize that if we read out at a "noise-monitoring" frequency $f_{\mathrm{monit}} = f_r \pm \Delta f$ that is in

$$\frac{f_r}{Q_r} \ll\ \Delta f \ll f_r \tag{4.166}$$

$$\Rightarrow O(10^6) \lesssim\ \Delta f \lesssim O(10^7)\ [\mathrm{Hz}], \tag{4.167}$$

i.e., significantly off-resonance while the electronics still behave approximately the same as on-resonance, we can utilize the off-resonance signal as an electronics noise proxy to remove the identical contamination in the on-resonance signal. According to Eq. (4.5) and this assumption for the on- and off-resonance signal compositions, one can straightforwardly compute the ratio of the on-/off-resonance transmissions by

$$\frac{S_{21,\mathrm{on}}}{S_{21,\mathrm{off}}} = \frac{S_{21}(f_{\mathrm{read}})}{S_{21}(f_{\mathrm{monit}})} \tag{4.168}$$

$$= \frac{a(f_{\mathrm{read}})}{a(f_{\mathrm{monit}})} e^{-2\pi i (f_{\mathrm{read}} - f_{\mathrm{monti}})t} \left[1 - \frac{\frac{Q_r}{Q_c} e^{i\phi_c}}{1 + 2iQ_r \frac{f_{\mathrm{read}} - f_r}{f_r}} \right], \tag{4.169}$$

where we use the subscripts on/off to denote on-/off-resonance. The electronic noise "cleaning" for the main resonator signal is then simply subtracting the noise proxy rescaled to the near-resonance signal size,

$$S_{21,\mathrm{on}}(t) \underset{\mathrm{cleaning}}{\longrightarrow} S_{21,\mathrm{on}}(t) - \left(S_{21,\mathrm{off}}(t) \times \frac{S_{21,\mathrm{on}}}{S_{21,\mathrm{off}}} \right), \tag{4.170}$$

where we use $S_{21}(t)$ to specify noisy instantaneous signal timestreams and S_{21} without (t) for the long-time-averaged unperturbed transmissions.

The simple subtraction cleaning method above has a flaw. Based on Eq. (4.166) and the prerequisite DAQ precision for frequency-profiling the resonances, the terms in Eq. (4.169) that differentiate f_{monit} and f_{read} are resolvable and thus introduce manageable uncertainties, which also allows us to omit the square bracket in Eq. (4.5) for $S_{21,\mathrm{off}}$. However, we purposely write down the full equation for the resonator signal $S_{21,\mathrm{on}}(f_{\mathrm{read}} \approx f_r)$ to show that, in the denominator, the uncertainty is inflated by the large Q_r prefactor relative to the 1 at front, resulting in a highly uncertain value unless

$$\delta(\Delta f / f_r) \ll Q_r^{-1} \;\Leftrightarrow\; \delta(\Delta f) \ll \mathcal{O}(10)\ \mathrm{KHz}. \tag{4.171}$$

Unfortunately, in order to achieve such a $\lesssim$1-KHz frequency accuracy and stability in practice, not only do we need to calibrate the USRP electronics to a challenging high stability for its grade of products, but we also need to determine an accurate f_r in the resonance fit to be referenced, which has been demonstrated unpractical. This analysis inspires us to search for an empirical method to determine the appropriate $S_{21,\mathrm{on}}/S_{21,\mathrm{off}}$ conversion factor, and then the cleaning procedure as described by Eq. (4.170) still applies.

While O. Wen is currently leading the effort for developing such an empirical approach, we introduce the basic concept for the cleaning in the following. We first define

$$S_{21,\mathrm{on}}(t) = S_{21,\mathrm{s}}(t) + S_{21,\mathrm{nKID}}(t) + S_{21,\mathrm{nEle,on}}(t) \tag{4.172}$$

$$= \sum\!\!\!\!\!\!\int_{-\infty}^{\infty} \left[\tilde{S}_{21,\mathrm{s}}(f) + \tilde{S}_{21,\mathrm{nKID}}(f) + \tilde{S}_{21,\mathrm{nEle,on}}(f)\right] e^{2\pi i f t}\delta f \tag{4.173}$$

$$S_{21,\mathrm{off}}(t) = S_{21,\mathrm{nEle,off}}(t) \tag{4.174}$$

$$= \sum\!\!\!\!\!\!\int_{-\infty}^{\infty} \left[\tilde{S}_{21,\mathrm{nEle,off}}(f)\right] e^{2\pi i f t}\delta f, \tag{4.175}$$

where the S_{21} timestreams for the transient signal, the shared electronic noise, and the KID-generated noise are specified by, respectively, the subscripts s, nEle, and nKID according to our on-/off-resonance signal composition assumption. We also define their corresponding Fourier integrals given by the mode amplitudes $\tilde{S}_{21}$. To quantify the common part in the on-/off-resonance signals, we substitute the signal definitions into the normalized standard cross-correlation coefficient

$$\frac{\left\langle S_{21,\mathrm{off}}(t) \middle| S_{21,\mathrm{on}}(t)\right\rangle}{\sqrt{\left\langle S_{21,\mathrm{off}}(t)\right\rangle \left\langle S_{21,\mathrm{on}}(t)\right\rangle}} = \frac{\sum\!\!\!\!\!\!\int_{-\infty}^{\infty} S^*_{21,\mathrm{off}}(t) S_{21,\mathrm{on}}(t)\delta t}{\sqrt{\sum\!\!\!\!\!\!\int_{-\infty}^{\infty} \left|S_{21,\mathrm{off}}(t)\right|^2 \delta t \sum\!\!\!\!\!\!\int_{-\infty}^{\infty} \left|S_{21,\mathrm{on}}(t)\right|^2 \delta t}}, \tag{4.176}$$

which effectively calculates the angle (cosine) between the two vectors $S_{21,\mathrm{on}}(t)$ and $S_{21,\mathrm{off}}(t)$ in the Hilbert space of t. Providing the condition that the anticipated pulses only occupy a small fraction of the timestream relative to the near-infinite data length, all the signal terms $\tilde{S}_{21,\mathrm{s}}$ in Eq. (4.176) are suppressed by $\sqrt{t}$ relative to the noise terms and hence become negligible at large t, and the correlation coefficient reduces to (cont'd)

$$= \frac{\sum\!\!\!\!\!\!\int_{-\infty}^{\infty} \left[\tilde{S}^*_{21,\mathrm{nEle,off}} \left(\tilde{S}_{21,\mathrm{nKID}} + \tilde{S}_{21,\mathrm{nEle,on}}\right)\right] \delta f}{\sqrt{\sum\!\!\!\!\!\!\int_{-\infty}^{\infty} \left|\tilde{S}_{21,\mathrm{nEle,off}}\right|^2 df \sum\!\!\!\!\!\!\int_{-\infty}^{\infty} \left[\left|\tilde{S}_{21,\mathrm{nKID}}\right|^2 + \left|\tilde{S}_{21,\mathrm{nEle,on}}\right|^2 + 2\mathbf{Re}\left(\tilde{S}^*_{21,\mathrm{nKID}} \tilde{S}_{21,\mathrm{nEle,on}}\right)\right] \delta f}}. \tag{4.177}$$

In deriving Eq. (4.177), we invoke that, at large t, oscillating terms due to multiplying different frequency modes in $S_{21,\mathrm{on}}$ and $S_{21,\mathrm{off}}$ average to zero, leaving only the products of equal-frequency terms whose $e^{2\pi i (f-f')t}$ term vanishes. The isolated unbounded $\int dt$ integrals in the numerator and the denominator then cancel each other, manifesting that the system is held at equilibrium at every instant.

Eq. (4.177) shows that the correlation coefficient depends not only on the on-/off-resonance electronic noise terms but also the KID-generated noise. In fact, the result is symmetric under

$$\tilde{S}_{21,\mathrm{nEle,on}} \leftrightarrow \tilde{S}_{21,\mathrm{nKID}} \tag{4.178}$$

and treats the off-resonance contribution $\tilde{S}_{21,\mathrm{nEle,off}}$ as an isolated multiplier. It indicates that we can in principle extract not just a shared contamination from $S_{21,\mathrm{on}}(t)$ but also any timestream as defined by $S_{21,\mathrm{off}}(t)$, which is not surprising since in practice we have no objective method to identify the shared noise in $S_{21,\mathrm{on}}(t)$ except assuming a subjective $S_{21,\mathrm{off}}(t)$. The on-/off-resonance signal composition assumption nevertheless provides a physical justification for utilizing $S_{21,\mathrm{off}}(t)$. Meanwhile, the symmetry under Eq. (4.178) suggests that the correlation coefficient given by Eq. (4.177) is only an unbiased indicator for the intended electronic noise removal when

$$\left|\tilde{S}_{21,\mathrm{nEle,on}}\right| \gg \left|\tilde{S}_{21,\mathrm{nKID}}\right|, \tag{4.179}$$

which, if assumed, yields the correlation coefficient (cont'd)

$$= \frac{\displaystyle\sum\!\!\!\!\!\!\int_{-\infty}^{\infty} \tilde{S}^{*}_{21,\mathrm{nEle,off}} \tilde{S}_{21,\mathrm{nEle,on}} \delta f}{\sqrt{\displaystyle\sum\!\!\!\!\!\!\int_{-\infty}^{\infty} \left|\tilde{S}_{21,\mathrm{nEle,off}}\right|^2 df \sum\!\!\!\!\!\!\int_{-\infty}^{\infty} \left|\tilde{S}_{21,\mathrm{nEle,on}}\right|^2 df}}. \tag{4.180}$$

Eq. (4.180) is nothing but the relative phase between $S_{21,\mathrm{nEle,on}}$ and $S_{21,\mathrm{nEle,off}}$, so we insert their relative amplitude and recover Eq. (4.170), i.e.,

$$S_{21,\mathrm{on}}(t) \xrightarrow[\text{cleaning}]{} \\ S_{21,\mathrm{on}}(t) - \left(S_{21,\mathrm{off}}(t) \times \underbrace{\frac{\left\langle S_{21,\mathrm{off}}(t) \middle| S_{21,\mathrm{on}}(t) \right\rangle}{\sqrt{\left\langle S_{21,\mathrm{off}}(t) \right\rangle \left\langle S_{21,\mathrm{on}}(t) \right\rangle}}}_{\text{phase}} \underbrace{\frac{\left|S_{21,\mathrm{on}}(t)\right|}{\left|S_{21,\mathrm{off}}(t)\right|}}_{\text{amplitude}} \right). \tag{4.181}$$

We note again that this result holds only under the condition that the combined noise proxy as defined by $S_{21,\mathrm{off}}$ dominates the noise composition, in particular dominating the KID-generated noise.

Now that we have shown the empirical timestream cross-correlation and the model-based $S_{21,\mathrm{on}}/S_{21,\mathrm{off}}$ conversion are mathematically equivalent, we compare the two approaches to argue for why the former is preferred. Indeed, if all the prerequisites are satisfied, both methods, namely Eq. (4.170) and Eq. (4.181), should perform exactly the same. However, for the model-based method, the $S_{21,\text{on}}/S_{21,\text{off}}$ factor is only a good representation for the conversion if it is not biased. In practice, one may suppress the random noise with a long averaging, but as discussed surrounding Eq. (4.171), we have demonstrated that the remaining systematic uncertainty is sufficient to invalidate the approach. On the contrary, cross-correlating the timestreams empirically accommodates the systematic uncertainty, while by-definition the Fourier transforms are to be performed on long enough data to suppress random-phase terms, which equivalently achieves the noise suppressing of averaging long timestreams. However, there is an obvious weakness to the cross-correlation technique. As the derivation requires, one always needs to make a subjective assumption for the noise proxy. In situations where the electronic noise does not necessarily dominate the total noise, nor can one reliably isolate the electronic noise for calculating the correlation coefficient, Eq. (4.181) still takes the assumption that $S_{21,\text{on}}(t)$ is dominated by the noise proxy for the cleaning regardless, therefore understandably introducing artificial biases.

Fig. 4.22 presents a preliminary result for the cross-correlation electronic noise removal using the monitor signal. We find that the noise removal technique provides more than an order-of-magnitude overall reduction in the noise PSDs for the prototype detector of this thesis. In addition to the overall improvement, the technique effectively suppresses the visible $1/f$ low-frequency noise in both the dQ and df signals, as well as the structural high-frequency noise in the df signal. After the noise cleaning, we obtain much improved white noise spectra except for the 0.1-MHz structure in dQ, which we suspect is due to unidentified environmental or instrumental RF pickups. To avoid unnecessary bias, we only consider PSDs below 0.2 MHz for the cleaning data processing and therefore the artificial cutoffs in Fig. 4.22. Recalling that we predict the noise to be dominated by the white HEMT amplifier noise in the gain direction, potentially raised by the GR noise below $f_{\text{qp,cut}}$, while a significantly higher $1/f$ mixer phase noise is anticipated in the phase direction. From the fact that the post-cleaning noise PSDs are generally white and without a noticeable discontinuity at $f_{\text{qp,cut}} \sim 10^{4-5}$ Hz, we conclude that our pre-cleaning noise is dominated by the $1/f$ electronic phase noise mixed into the dQ and df signals, and the post-cleaning noise is dominated by the HEMT. Also from the fact that we successfully mitigate the low-frequency $1/f$ noise with the monitor signal that is external to the KID, especially in the df direction, we exclude the possibility that the TLS noise is significant in the observed total noise (see next section).

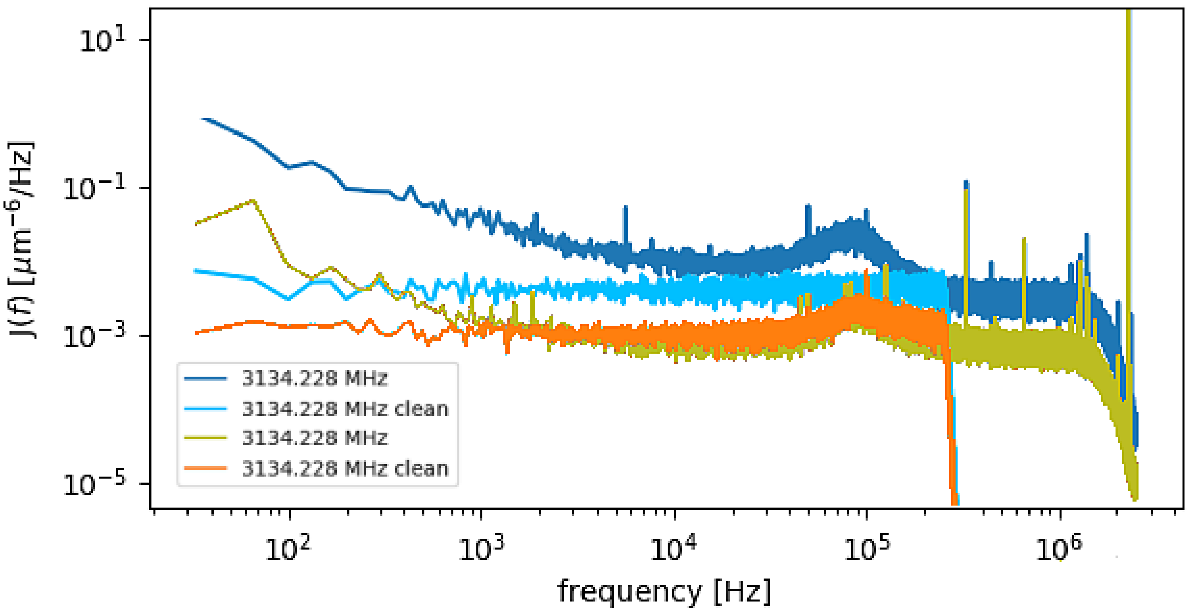


Figure 4.22: An example preliminary result for the cross-correlation electronic noise removal with monitor signal. The data shown in the figure are the noise for the f_r = 3.134-GHz KID in the 80-KID prototype detector, acquired at a readout power of -25 dBm at the cryostat input, the highest readout power without an apparent readout-power QP dominance. We choose the monitor signal frequency to be 4 FWHM (full width at half max) lower than f_r. The curves in the figure represent the noise PSDs for the df signal before (dark blue) and after (light blue) the noise removal, and the dQ signal before (green) and after (orange) the noise removal. The PSD unit is calibrated to the QP density based on the procedure detailed in the RF and the temperature-sweep calibrations. We low-pass filter the data at 0.2 MHz in advance, a bandwidth we deem sufficient for the phonon pulse detection based on our pulse calibration and energy resolution model (Sec. 4.6). The result shows, in both the dQ and df directions, we successfully suppress the $1/f$ noise populating <10 kHz to be the white noise PSDs, while the 0.1-MHz structure is also mitigated in the df direction but remains in dQ. For this particular result, we thank A. Villapando and T. Aralis, respectively, for helping with implementing the algorithm and data taking.

For future reference, it is useful to note that recently a new DMKID detector utilizing Nb-proximitized KID capacitors has obtained a dominating $1/f$ noise only in the df direction post-noise removal, which is suppressed at higher readout powers. It strongly suggests a TLS noise contribution by the proximitized capacitors. According to previous elaboration, due to the dominant KID-internal noise component, we anticipate and indeed find that our simple cross-correlation cleaning technique without electronic noise isolation is ineffective for this new device. O. Wen is currently leading the effort for developing a more advanced technique for obtaining a proper ($S_{21,\text{on}}/S_{21,\text{off}}$) conversion factor under high-TLS noise.

4.5.6 Two-level system noise

We briefly introduce TLS noise at the end of this section, since it is a well-established phenomenon that is frequently observed in SC resonators. Although we have demonstrated that it is a negligible effect in our Al-KID prototype detector, we do acquire a nontrivial TLS contribution in our recent Al/Nb-hybrid KIDs. Limited by the specialization of the topic, especially when the TLS noise is not observed in the main detector for this thesis, we provide a practical summary for its physics principle with useful diagnostic observables and refer the reader to in-depth references cited below. As the name TLS suggests, the phenomenon is caused by quantum systems "tunneling" between two distinct (meta-)stable states or energy levels. Such a phenomenon is therefore increasingly significant when the system is cooled to temperatures comparable or below its typical state separation (divided by k_B), otherwise the density function of states should spread thermally, i.e., classically, across all possible states and subsequently suppresses the quantized two-level phenomenon. The tunneling model also naturally expects the tunneling rate to depend on the lower (ground) state's occupation, which in turn suggests the temperature and/or the relevant external energy injections to be the dependence.

Specifically for KIDs, so far one well-recognized method to realize such a TLS is through the inhomogeneity of the supposedly pure-substance dielectric materials, while other possibilities are being actively investigated. It has been shown recently that the native oxides on the surfaces or interfaces of the substrates and the SC films provide the inhomogeneity for the TLS. With the resonating RF photons constantly experiencing two different dielectric materials, the dielectric constant and inevitably the derived resonance property then exhibit a tunneling junction-like noise behavior. Based on the state-tunneling picture, (Phillips, 1987) provided a rigorous quantum mechanical derivation that yields

$$\frac{\Delta\epsilon}{\epsilon} = \frac{-2\delta}{\pi}\left\{\mathbf{Re}\left[\Psi\left(\frac{1}{2}+\frac{\xi}{\pi i}\right)\right] - \log(2\xi)\right\}, \tag{4.182}$$

where ϵ is the material's macroscopic ground state dielectric constant, δ is its loss tangent due to the TLS dissipation, Ψ is the complex di-gamma function, and ξ is the usual thermodynamic energy variable defined in Eq. (4.29), which encodes the RF excitation with respect to the temperature. Considering that the KID capacitor's RF photon field is predominately between the SC electrodes thus experiences the dielectric TLS more strongly than the inductor, which confines the current in the SC filme, (Gao et al., 2008) argued that the TLS may lead to a fluctuating capacitance for the capacitor and in turn a fluctuating KID resonance (Gao et al., 2008). was able to propagate the tunneling fluctuation in Eq. (4.182 to the capacitance and showed that the KID observables are modified by

$$\frac{1}{Q_i} = F\delta \times \frac{\tanh \xi}{\sqrt{1 + E^2/E_c^2}} \tag{4.183}$$

$$\frac{f_r - f_r|_{T=0}}{f_r|_{T=0}} = F\delta \times \frac{1}{\pi}\left\{\mathbf{Re}\left[\Psi\left(\frac{1}{2} + \frac{\xi}{\pi i}\right)\right] - \log(2\xi)\right\}, \tag{4.184}$$

where E is the electric field strength between the electrodes of the capacitor (IDC fingers), E_c is a phenomenological critical field that, if exceeded, maintains the state at the top level without fluctuation, F is a phenomenological variable Gao postulated for the fractional TLS/oxide contamination to the pristine substrate, and we rearrange the product of $F\delta$, the maximal possible TLS loss, to the front of the equations to distinguish the temperature- and the RF field-dependences.

In the above result, the total RF energy E^2 scales against the loss Q_i^{-1} as one expects for suppressing the two-level tunneling at larger excitations. However, the equations also suggest that the frequency shift due to TLS dissipation is fixed regardless of the adjustable readout power, as long as the photon energy for the resonator, f_r or ξ, is determined. Experimentally, a consistent noise suppression has also been demonstrated on typical GHz-range resonators possessing typical IDC capacitors under reasonable readout powers, where noticeable excess noises are frequently obtained only in the df noise but not in dQ. However, in order to contribute the additional df noise to TLS, one needs a plausible criterion to justify the association. Although currently the micro-physics for TLS noise is very much under development thus is difficult to make first-principle predictions, according to the hypothesized tunneling picture, it is generally anticipated that the TLS noise exhibits an $1/f$-like noise PSD that is prominent at low frequencies, i.e.,

$$S_{\delta f_r/f_r} \propto f^{-\alpha} \qquad , \alpha \approx \frac{1}{2}, \tag{4.185}$$

which is indeed consistent with the observations. According to (Gao et al., 2008), if one makes an educated assumption that a similar saturation to the TLS loss at high readout powers also applies to the fluctuation of the dielectric constant $\delta|\Delta\epsilon(t)|$, in the regime where the TLS noise is fully suppressed in dQ by a high readout power

(E^2), one has

$$S_{\delta\Delta\epsilon} \propto \frac{1}{\sqrt{1+E^2/E_c^2}} \sim \frac{1}{E}, \tag{4.186}$$

which yields

$$S_{\delta f_r/f_r} \propto \frac{1}{E} \propto \frac{1}{P_g^{1/2}}. \tag{4.187}$$

Despite the somewhat large uncertainty, it has also been observed in multiple occasions that (Siegel, 2016)

$$S_{\delta f_r/f_r} \propto \frac{1}{T^{1-2}}. \tag{4.188}$$

So in summary, one may diagnose whether the excess df noise is due to TLS by

$$S_{\delta f_r/f_r,\mathrm{TLS}} \propto \frac{1}{f^\alpha P_{\mathrm{read}}^\beta T^\gamma} \qquad \text{, with} \begin{cases} \alpha \approx 0.5 \\ \beta \approx 0.5 \\ \gamma \approx 1-2 \end{cases}. \tag{4.189}$$

It is worth mentioning that, in multiple recent studies, the obtained pure-df noise PSDs that agree with Eq. (4.189)'s empirical criteria unambiguously correlate with, or even controllable by, physical manipulations to the substrates' dielectric inhomogeneities, such as via atomic or X-ray microscopy characterizations, controlled oxide growth, or post-fabrication treatments (Virginia et al., 2020). These studies provide strong supports for the diagnosis technique as well as the hypothesized physics origin for the TLS noise.

4.6 Energy resolution

4.6.1 Pulse shape model

The following content largely references Golwala and Zmuidinas' published or unpublished theoretical works (S. R. Golwala, 2010; Zmuidzinas, 2009; Zmuidzinas, 2012).

We first examine the OF resolution for a perfect one-sided exponential-decay pulse. The reader can refer to Appx. B for further discussion of OF and our convention of notation. We consider a peak-normalization signal template

$$\hat{S}(t) = \begin{cases} e^{-t/\tau} & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{4.190}$$

which, according to the convention of Eq. (B.10), has

$$\tilde{\hat{S}}(f) = \frac{\tau}{2\pi i f\tau + 1} \sim \tau \times \begin{cases} 1 & , 2\pi f \ll \tau^{-1} \\ f^{-1} & , 2\pi f \gg \tau^{-1}. \end{cases} \tag{4.191}$$

Notice that we now reuse τ to denote the decay time constant for the signal, in the hope that it will be linked to the physical time constants discussed in Sec. 4.4. This τ should not be confused with the OF time offset, which we have argued is generally negligible and therefore omitted in the following discussions. Based on the in-depth elaboration in Sec. 4.5, we concluded that we anticipate that, after cross-correlating the noise monitor signal, our noise is dominated by the HEMT in the electronic gain direction (Fig. 4.22). Regardless of our final choice for the readout scheme, e.g., 1D/2D readout, noise removal techniques, etc., we have explained that theoretically the aforementioned dominant noise sources all exhibit white noise spectra therefore suggest that we can always express the noise by an equivalent noise temperature

$$\left|\tilde{N}(f)\right|^2 = \frac{k_\mathrm{B}T_n}{2}; \tag{4.192}$$

the factor of 1/2 for is for defining $\tilde{N}$ in $f \in (-\infty, \infty)$. We substitute the signal template and the noise assumptions Eq. (4.191) and Eq. (4.192) into the OF resolution expression in Eq. (B.20) and obtain

$$\sigma_{A'} = \sqrt{\frac{k_\mathrm{B}T_n}{\tau}}. \tag{4.193}$$

Eq. (4.193) is the simplest, yet oftentimes realistic enough for a first approximation, standard OF baseline resolution for a single-time constant exponential-decay pulse on a white noise of an equivalent temperature T_n. Thanks to Eq. (4.191)'s asymptotic behavior at much above and below $f \sim 1/\tau$, one may understand the neat form of Eq. (4.193) as a constant noise PSD of $k_\mathrm{B}T_n$ being accumulated from DC up to $1/\tau$ and then square-rooted to indicate the size of its corresponding uncertainty. The result is an easily appreciable concept, namely a noise "strength" of $k_\mathrm{B}T_n$ with a finite "bandwidth" of $1/\tau$, and provides a quick estimate for the respected resolution based on assumed dominant noise sources and signal time constants. We therefore take it as the starting point for the resolution modeling.

We mentioned in the pulse modeling analysis that, if the phonons can not transmit the recoil energy immediately into the KID QP system, the signal pulses are expected to be boffered and delayed. Recall that in Eq. (4.69) we assume a sudden appearance of out-of-equilibrium QPs at $t = 0$, which suggests

$$\tau = \tau_\mathrm{qp} = \frac{1}{2Rn_\mathrm{qp}} \tag{4.194}$$

for Eq. (4.191). If the phonon energy insertion for pair-breaking requires a non-negligible amount of time, we should consider a more complex pulse shape with a finite-slope rising edge and likely a altered decay time constant(s). At the moment, T. Aralis and E. Lindeman are investigating the realistic phonon pulse shape with a first-principle approach and in a Monte Carlo simulation, respectively. To illustrate the implementation of the phonon pulse shape for future reference, we follow (S. R. Golwala, 2010)'s toy model, which postulates, instead of a δ-function phonon energy deposition, a QP generation with an exponential-decay time profile[14]

$$\delta\Gamma_{\mathrm{G}} = \frac{\eta_{\mathrm{ph}} E_{\mathrm{ph}}}{\Delta} \frac{e^{-t/\tau_{\mathrm{abs}}}}{\tau_{\mathrm{abs}}}, \tag{4.195}$$

where η_{ph} is a phenomenological pair-breaking efficiency for the total phonon energy E_{ph}, and we use the small variation of the total QP generation Γ_{G} to represent the observable out-of-equilibrium signal generation. We rewrite Eq. (4.63) using the above assumption also for a small perturbation, which yields

$$\frac{d\delta n_{\mathrm{qp}}}{dt} = -\frac{\delta n_{\mathrm{qp}}}{\tau_{\mathrm{qp}}} + \frac{\eta_{\mathrm{ph}} E_{\mathrm{ph}}}{\Delta V} \frac{e^{-t/\tau_{\mathrm{abs}}}}{\tau_{\mathrm{abs}}}. \tag{4.196}$$

The above small-variation GR equation is solved to (algebra in Appx. A)

$$\delta\tilde{n}_{\mathrm{qp}}(f) = \frac{\eta_{\mathrm{ph}} E_{\mathrm{ph}}}{\Delta V} \cdot \underbrace{\frac{\tau_{\mathrm{qp}}}{\tau_{\mathrm{abs}} - \tau_{\mathrm{qp}}} \left[\frac{\tau_{\mathrm{abs}}}{2\pi i f \tau_{\mathrm{abs}} + 1} - \overbrace{\frac{\tau_{\mathrm{qp}}}{2\pi i f \tau_{\mathrm{qp}} + 1}}^{\mathbf{X}(\tau_{\mathrm{abs}}, \tau_{\mathrm{qp}})} \right]}_{\mathbf{Y}(\tau_{\mathrm{abs}}, \tau_{\mathrm{qp}})} \tag{4.197}$$

$$\Rightarrow \quad \delta n_{\mathrm{qp}}(t) = \frac{\eta_{\mathrm{ph}} E_{\mathrm{ph}}}{\Delta V} \cdot \frac{\tau_{\mathrm{qp}}}{\tau_{\mathrm{abs}} - \tau_{\mathrm{qp}}} \cdot \left[e^{-t/\tau_{\mathrm{abs}}} - e^{-t/\tau_{\mathrm{qp}}} \right]. \tag{4.198}$$

We define the $\mathbf{X}(\tau_{\mathrm{abs}}, \tau_{\mathrm{qp}})$ and the $\mathbf{Y}(\tau_{\mathrm{abs}}, \tau_{\mathrm{qp}})$ sections of the equation for later convenience.

We compare the frequency-domain solution in Eq. (4.197) to Eq. (4.191) and find that, at frequencies much below τ_{qp}^{-1} and τ_{abs}^{-1}, the finite pair-breaking time rescales the pulse shape by

$$\tilde{\tilde{S}} \longrightarrow \tilde{\tilde{S}} \times \frac{|\mathbf{Y}|}{|\mathbf{X}|}, \tag{4.199}$$

which then modifies the OF basedline resolution by

$$\sigma_{A'} = \sqrt{\frac{k_{\mathrm{B}} T_n}{\tau_{\mathrm{qp}}}} \longrightarrow \sigma_{A'} = \sqrt{\frac{k_{\mathrm{B}} T_n}{\tau_{\mathrm{qp}}}} \times \left[\frac{\displaystyle\sum\!\!\!\!\!\!\int |\mathbf{Y}|^2 \,\delta f}{\displaystyle\sum\!\!\!\!\!\!\int |\mathbf{X}|^2 \,\delta f} \right]^{-1/2}. \tag{4.200}$$

[14]Notice that we change (S. R. Golwala, 2010)'s normalization convention for $\delta\Gamma_{\mathrm{G}}$ so to have $\int_0^\infty \delta\Gamma_{\mathrm{G}} = \frac{\eta_{\mathrm{ph}} E_{\mathrm{ph}}}{\Delta}$.

It reality, a robust determination of Eq. (4.200) almost certainly requires numerical simulations due to unique pulse shapes for different detector designs. Nevertheless, as $\mathcal{O}(10^2)$-μsec. τ_{abs} have been observed in SuperCDMS detectors, whose geometries and material compositions are very similar to our detectors, and at the same time we have also obtained our current $\tau_{\mathrm{qp}} \approx 23$ μsec. with preliminary τ_{abs} as high as 180 μsec. (Eq. (4.87)), we assume

$$\tau_{\mathrm{qp}} \ll \tau_{\mathrm{abs}}, \tag{4.201}$$

which simplifies the above result to (algebra in Appx. A)

$$\left[\frac{\displaystyle\oint\!\!\!\!\!\!\sum |\mathbf{Y}|^2\,\delta f}{\displaystyle\oint\!\!\!\!\!\!\sum |\mathbf{X}|^2\,\delta f}\right]^{-1/2} \approx \sqrt{\frac{\tau_{\mathrm{abs}}}{\tau_{\mathrm{qp}}}}. \tag{4.202}$$

With the pulse shape effect included into the model, we substitute the pulse and the noise data presented previously and arrived at our first resolution estimate

$$\sigma_{A'} = \sqrt{\frac{k_{\mathrm{B}}T_n}{\tau_{\mathrm{qp}}}} \cdot \sqrt{\frac{\tau_{\mathrm{abs}}}{\tau_{\mathrm{qp}}}} \tag{4.203}$$

$$\approx 2.3 \times 10^{-9}\,\sqrt{\mathrm{W}} \quad \times \left[\left(\frac{T_n}{4\ \mathrm{K}}\right)^{1/2} \left(\frac{\tau_{\mathrm{qp}}}{100\ \mu\mathrm{s}}\right)^{-1} \left(\frac{\tau_{\mathrm{abs}}}{1\ \mathrm{ms}}\right)^{1/2}\right] \tag{4.204}$$

in the unit of a fluctuating noise amplitude.

4.6.2 Energy resolution model

We continue to translate the noise power-calibrated Eq. (4.203) using the simple relation

$$\sigma_y = \sigma_x \cdot \frac{\partial \delta y}{\partial \delta x}, \tag{4.205}$$

which transforms the resolution formula into other units corresponding to different physical quantities of interest, such as the number of QPs or the incoming phonon energy. We should emphasize that, because the dependence of y may implicitly depend on x, instead of using $\partial y/\partial x$ for the chain-rule transformation, one should pay attention to the difference between $\partial\delta y/\partial\delta x$ and $\partial y/\partial x$ and ensure that the former is calculated for generating observable signals. One example for such difference is $\partial\delta S_{21}/\partial\delta n_{\mathrm{qp}}$ and $\partial S_{21}/\partial n_{\mathrm{qp}}$ in Eq. (4.119) and Eq. (4.22), respectively. In the following, we will first transform the $\sigma_{A'}$ previously obtained in Eq. (4.203) for a pure noise power to the variation in S_{21}, by comparing Eq. (4.203) to the applied readout power. We then convert the $\sigma_{S_{21}}$ into its corresponding $\sigma_{n_{\text{qp}}}$, the equivalent resolution for the QP density fluctuation in the KID, which physically indicates the resolvable number of QPs using the readout power on the assumed noise condition. Finally, we will relate $\sigma_{n_{\text{qp}}}$ to the phonon pair-breaking energy deposition to complete our energy resolution model, where we will invoke a phenomenological model for the phonon energy deposition from the substrate into the QP system.

4.6.2.1 RF signal resolution

We link the applied readout power to the signal/noise amplitude by

$$P_g = \frac{A^2}{4}, \tag{4.206}$$

where A is the amplitude in the unit of $\sqrt{\text{W}}$ (SI). The relation converts the amplitude resolution

$$\sigma_{S_{21}} = \sigma_{A'} \cdot \frac{\partial \delta S_{21}}{\partial \delta A'} = \sigma_{A'} \cdot \frac{1}{A} = \sigma_{A'} \cdot \sqrt{\frac{1}{4P_g}} \tag{4.207}$$

$$\approx 3.7 \times 10^{-5} \sqrt{\text{W}/\text{W}} \quad \times \left[\left(\frac{P_g}{10^{-9}\ \text{W}\ (-60\ \text{dBm})} \right)^{-1/2} \right]. \tag{4.208}$$

4.6.2.2 QP density resolution

We already derived $\partial \delta n_{\text{qp}} / \partial \delta S_{21}$ in Eq. (4.119) and its corresponding Appendix content for the GR noise model, so we simply substitute the formula and obtain

$$\sigma_{\text{qp}} = \sigma_{S_{21}} \cdot \frac{\partial \delta n_{\text{qp}}}{\partial \delta S_{21}} \tag{4.209}$$

$$= \sigma_{S_{21}} \cdot n_{\text{qp}} \frac{Q_c Q_i}{Q_r^2} \frac{1}{\chi_{\text{qp}}} \tag{4.210}$$

$$\approx 7.4 \times 10^{-2}\ \text{QP}/\mu\text{m}^3 \times \left[\left(\frac{n_{\text{qp}}}{200\ \text{QP}/\mu\text{m}^2} \right)^1 \left(\frac{Q_r \approx Q_c}{1 \times 10^5} \right)^{-1} \left(\frac{Q_i}{1 \times 10^6} \right)^1 \left(\frac{\chi_{\text{qp}}}{1.0} \right)^{-1} \right]. \tag{4.211}$$

Alternatively, we may also use the Mattis-Bardeen equation (Eq. (4.22)) that assumes the QP system is at a thermal equilibrium with the bath, which yields (cont'd)

$$= \sigma_{S_{21}} \cdot \frac{Q_c^2}{\alpha\kappa|\gamma|Q_r^3\chi_{\mathrm{qp}}} \tag{4.212}$$

$$\approx 7.4 \times 10^{-3}\ \mathrm{QP}/\mu\mathrm{m}^3$$

$$\times \left[\left(\frac{\alpha}{0.1}\right)^{-1} \left(\frac{\kappa}{5\times 10^{-7}}\right)^{-1} \left(\frac{Q_r \approx Q_c}{1\times 10^5}\right)^{-1} \left(\frac{\chi_{\mathrm{qp}}}{1.0}\right)^{-1} \right]. \tag{4.213}$$

The resolution given by the empirical QP density is an order-of-magnitude higher than assuming the Mattis-Bardeen equation, which assumes for κ that the QPs are at the thermal bath temperature. Recall that we found in both the pulse detection (Sec. 4.4) and the GR noise (Sec. 4.5) analyses that our current n_{qp} is an order-of-magnitude higher than which anticipated by the Mattis-Bardeen equation as well as the low values obtained in the literature. Providing the resolution scales linearly in n_{qp}, the above result is consistent with previous conclusion that the QP system is currently subjected to an unidentified energy source that holds the QP density higher than which allowed by the bath temperature. Although it is not immediately a close-to-optimal result, it encouragingly indicates that we may potentially achieve an order-of-magnitude resolution improvement by mitigating the unidentified energy source. We will discuss the potential mechanisms that elevates our current n_{qp} and introduce ongoing mitigation works later.

4.6.2.3 Phonon energy resolution

We assume that the phonons deliver a total energy of δE_{ph} at a pair-breaking efficiency η_{ph} that is uniformly distributed to all N_{KID} KIDs, i.e.,

$$V\delta n_{\mathrm{qp}}\Delta = \frac{\eta_{\mathrm{ph}}\delta E_{\mathrm{ph}}}{N_{\mathrm{KID}}}, \tag{4.214}$$

so

$$\sigma_{E_{\mathrm{ph}}} = \sigma_{n_{\mathrm{qp}}} \cdot \frac{\partial \delta E_{\mathrm{ph}}}{\partial \delta n_{\mathrm{qp}}} = \sigma_{n_{\mathrm{qp}}} \cdot \frac{V\Delta N_{\mathrm{KID}}}{\eta_{\mathrm{ph}}} \tag{4.215}$$

$$\approx 15.6 - 156\ \mathrm{eV}$$

$$\times \left[\left(\frac{V}{3.5\times 10^4\ \mu\mathrm{m}^3}\right)^{1} \left(\frac{\Delta}{180\ \mu\mathrm{eV}}\right)^{1} \left(\frac{N_{\mathrm{KID}}}{80}\right)^{1} \left(\frac{\eta_{\mathrm{ph}}}{0.24}\right)^{-1} \right], \tag{4.216}$$

where the lower and the higher bounds are calculated for the idealized Mattis-Bardeen and the empirical n_{qp} conditions, respectively, and we consider an example value of $\eta_{\rm ph} \approx 0.24$ based on SuperCDMS detectors' ≈30% Al fin phonon absorption efficiency. We reduce the 30% efficiency to 24% because ≈22% of the Al coverage for our prototype detector is the capacitors, which we assume consume phonons identically but do not generate signals. However, we should point out a difference between SuperCDMS Al fin and our KID film, which could potentially invalidate the efficiency assumption in an unfavorable way for our KID-based detector. Latest-generation SuperCDMS detectors typically employ 300–600 nm thick Al films, a good fraction of the phonon mean free path in Al, in order to ensure higher phonon collection efficiencies, i.e., phonons interact with Cooper pairs before leaving the film volume. Contrasting the thick films, in order to leverage the $(\lambda_{\rm eff}/d)^2$ kinetic inductance enhancement in the thin-film regime, we always choose 20-40 nm thick films for the KIDs. We are currently investigating if our much thinner films would impose a significant suppression to $\eta_{\rm ph}$. Also notice that, in contrast of (Moore et al., 2012; S. R. Golwala, 2010), we do not consider for Eq. (4.215) an additional factor of $1/\sqrt{N_{\rm KID}}$ due to repeating the energy measurement by $N_{\rm KID}$ times with $N_{\rm KID}$ KIDs for noise suppression. It is in principle a valid assumption, but due to the ongoing calibration work for all the KIDs in the prototype detector, the noise suppression by combining the measurements has not yet been performed in practice.

4.6.2.4 Pair-breaking energy resolution

In addition to the above results, we provide one more type of the resolution that is useful for later references, the so-called pair-breaking energy resolution $\sigma_{E_{\rm qp}}$. It is defined as the resolution for the energy that has really generated QPs, which then allows cross-comparing different KIDs' *inductor-only* intrinsic sensitivities regardless of the variation due to the energy insertion mechanisms or the KID/detector applications. We calculate $\sigma_{E_{\rm qp}}$ from Eq. (4.215) by setting $N_{\rm KID}$ and $\eta_{\rm ph}$ to be 1 or multiplying Eq. (4.209) with $V\Delta$, i.e.,

$$\sigma_{E_{\rm qp}} = \sigma_E \Big/ \frac{\eta_{\rm ph}}{N_{\rm KID}} = \sigma_{n_{\rm qp}} \cdot V\Delta \approx 0.046 - 0.46\ \text{eV}. \tag{4.217}$$

Based on the methodology detailed so far in this thesis, T. Aralis and K. Ramanathan are currently pursuing a complete energy resolution analysis for the 80-KID prototype detector of this thesis. They recently present the latest results in (Ramanathan et al., 2021), summarized in Fig. 4.23, in which they demonstrate that the obtained $\sigma_{E_{\rm qp}}$ are indeed improvable by an order-of-magnitude with the correlated noise removal as anticipated by Fig. 4.22. The energy resolutions after the noise removal

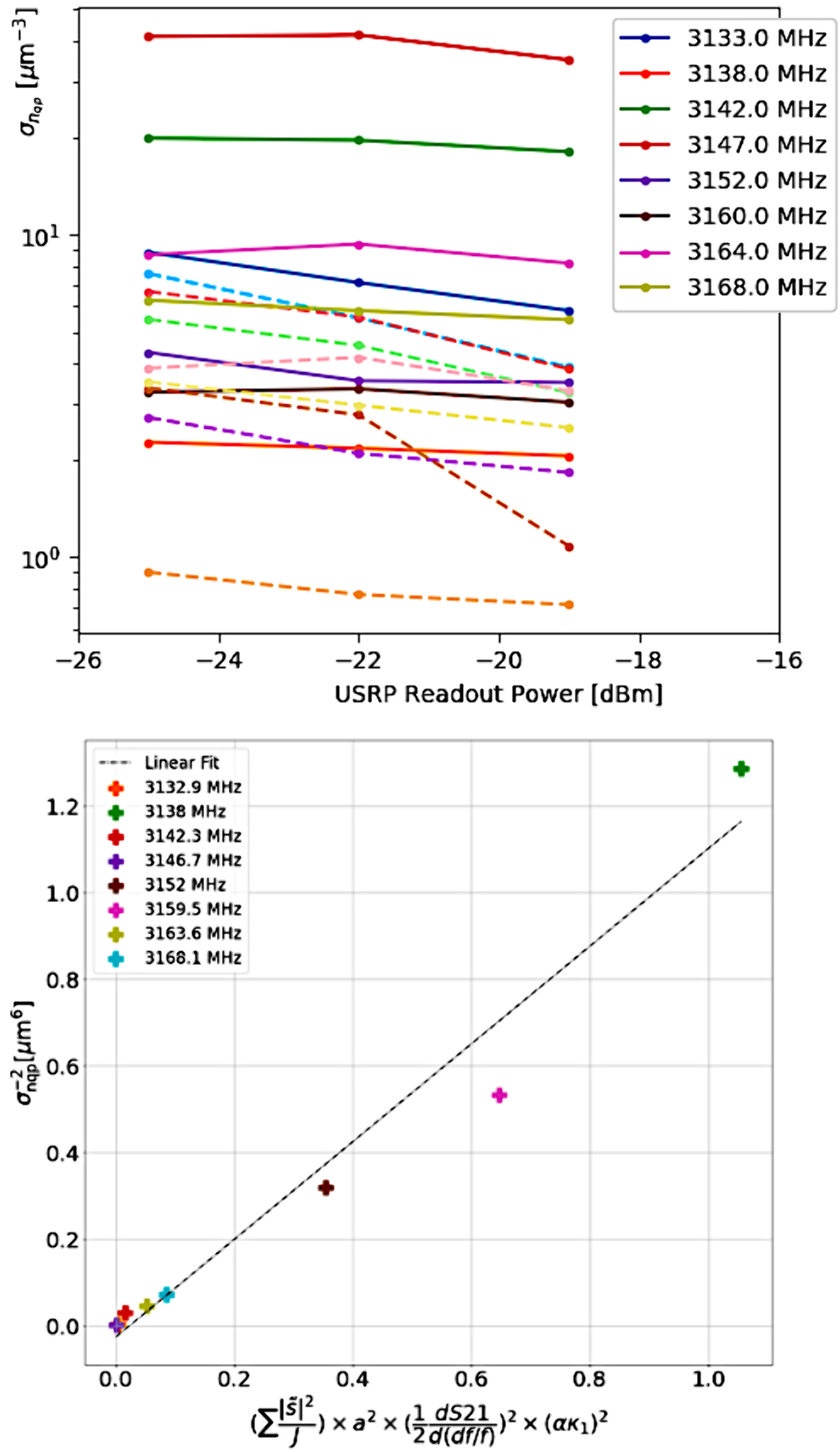


Figure 4.23: The preliminary energy resolutions obtained from the selected KIDs in the 80-KID prototype detector, whose corresponding resonant frequencies and marker colors are listed in the legend. The top and bottom figures present the resolutions against the readout power and the resolution model parameters, respectively. The solid and the dashed curves in the top figure represent the OF resolutions calculated with and without the correlated noise removal, respectively. The black dashed line in the bottom figure is a diagonal line that represents the predicted correspondence of the noted x- and y-axis quantities based on our resolution model. More details are found in the corresponding text and (Ramanathan et al., 2021). Figures produced by T. Aralis.

locate in the predicted range of Eq. (4.213) to Eq. (4.211) as shown in Fig. 4.23. While currently the observed resolutions for different KIDs scatter by roughly an order-of-magnitude as in the figure, they demonstrate by plotting the obtained resolutions against the dependence given by Eq. (4.219) that the resolutions indeed follow the model developed in this thesis.

4.6.3 Optimization

In parallel with the above empirical resolution modeling based on the obtained data, we reorganize the model to be a generic resolution formula that is suitable for general KID-based phonon-mediated detectors, as well as for aiding future detector designs. For later convenience, we first collect all the resolution equations introduced previously and rewrite

$$\sigma_E = \frac{\partial \delta E}{\partial \delta n_{\mathrm{qp}}} \cdot \frac{\partial \delta n_{\mathrm{qp}}}{\partial \delta S_{21}} \cdot \frac{\partial \delta S_{21}}{\partial \delta A'} \cdot \sigma_{A'} \tag{4.218}$$

$$= \frac{V \Delta N_{\mathrm{KID}}}{\eta_{\mathrm{ph}}} \cdot n_{\mathrm{qp}} \frac{Q_c Q_i}{Q_r^2} \frac{1}{\chi_{\mathrm{qp}}} \cdot \sqrt{\frac{1}{4P_g}} \cdot \sqrt{\frac{\tau_{\mathrm{abs}}}{\tau_{\mathrm{qp}}}} \sqrt{\frac{k_{\mathrm{B}} T_n}{\tau_{\mathrm{qp}}}}. \tag{4.219}$$

For situations where the devices have been fabricated, we physically measure the parameters in Eq. (4.219) and use it for a theoretical resolution prediction. However, if we are in the design phase for a new detector, where we are allowed to choose the parameters that are otherwise determined after the fabrication, Eq. (4.219) becomes ambiguous thus requires us to retreat to the physical assumptions leading to the optimization for these parameters. In particular, we consider three parameters in Eq. (4.219) that are related to more fundamental physical quantities: 1) The coupling/total quality factor $Q_{c/r}$, 2) the phonon energy absorption time τ_{abs}, and 3) the dominant QP population n_{qp}. As we will elaborate below, the independent optimizations for these parameters involve many other parameters in Eq. (4.219) due to the assumed QP-generation mechanisms, therefore together they indicate an optimal operation scheme in the end. Due to the close relation of 1) and 2), we will also combine their discussions in the following.

4.6.3.1 Quality factor and sensor coverage

According to the analysis surrounding Eq. (4.80), we rewrite

$$\frac{t_{\mathrm{sub}}}{v_{\mathrm{ph}}} \approx t_{\mathrm{rise}} \approx t_{\mathrm{ring}} \approx \frac{Q_r \approx Q_c}{\pi f_r} \quad (\because Q_i \gg Q_c), \tag{4.220}$$

which determines the optimization for the Q-factors based on matching the resonator response to the geometry and the material of the detector (S. R. Golwala, 2010). Based on the fact that it should only take $\sim\mu$sec. for phonons to arrive at the instrumented surface, which apparently contradicts the observed $\lesssim$msec. $\tau_{\rm abs}$ (Sec. 4.4), we believe that, instead of being immediately absorbed by the KIDs, the phonons remain in the substrate and travel multiple times of $t_{\rm sub}$ until they are finally absorbed by the KIDs or other phonon-sinking materials after many encounters with the KIDs. Adopting (S. R. Golwala, 2010)'s parametrization,[15], we assume the phenomenological probability for a phonon to create QPs in the KID for each encounter $\mathcal{P}_{\rm abs}$, which anticipates a phonon absorption time scale

$$\tau_{\rm abs} = \frac{2t_{\rm sub}}{v_{\rm ph}} \times \frac{1}{\mathcal{P}_{\rm abs}\eta_{\rm fill}} = \frac{2\tau_{\rm rise}}{\mathcal{P}_{\rm abs}\eta_{\rm fill}}. \tag{4.221}$$

Note that the phenomenological definition of $\mathcal{P}_{\rm abs}$ encompasses not only the probability for the phonon to penetrate the substrate-KID interface, but also the absorption probability by the KID's QP system before leaving the film volume, and potentially other yet identified effects. The first term in Eq. (4.221) represents the approximate time for the phonon to travel a full round-trip of the substrate to the instrumented surface, in other words we assume a single-side instrumented detector. The second term represents the combined phonon absorption probability when it arrives at the surface, where $\eta_{\rm fill}$ is the fractional surface film coverage. We also relate $\tau_{\rm abs}$ to $\tau_{\rm rise}$ in the second equation using Eq. (4.220).

We examine the model for the 80-KID prototype detector, which has an Al-KID coverage of

$$\eta_{\rm fill} \approx 2.9\% \tag{4.222}$$

for one of the $\varnothing$3" substrate surfaces; the other surface is uninstrumented. Given $t_{\rm sub}$ = 1 mm, $v_{\rm ph}$ = 8433 m/s for Si, and also the observed $\lesssim$ms $\tau_{\rm abs}$, Eq. (4.221) suggests

$$\mathcal{P}_{\rm abs} \sim \mathcal{O}(1)\%, \tag{4.223}$$

which is consistent with a preliminary Monte Carlo simulation result based on (S. B. Kaplan et al., 1976) by E. Lindeman. We reserve the above determination procedure for $\mathcal{P}_{\rm abs}$ to optimize $\eta_{\rm fill}$. In principle, by knowing the detector-dependent pulse rise time (Eq. (4.220)) and the phonon absorption probability, Eq. (4.221) yields the optimal sensor coverage $\eta_{\rm fill}$ for the targeted $\tau_{\rm abs}$. We note that, while it may seem

[15]Notice minor modifications and the choice of convention if to directly compare with (S. R. Golwala, 2010) such as the factor of 2 for $\tau_{\rm abs}$.

straightforward that a smaller τ_{abs} gives a smaller resolution per Eq. (4.219) thus is always preferred, which in turn prefers a detector that is covered as much as possible per Eq. (4.221), increasing the film coverage in practice likely also alters the volume V in Eq. (4.219) and potentially the needed readout power P_g and the resulting background QP density n_{qp}. Therefore, to optimize the energy resolution, all these factors need to be taken into account based on their relative impacts given by Eq. (4.219).

4.6.3.2 Readout power optimization

Zmuidzinas and colleagues (2009) realized during the early KID development that, if a KID is under no steady-state external energy insertion except its own readout power, one may in principle arbitrarily raise the readout power to improve the resolution by $\sqrt{1/P_g}$ as in Eq. (4.207). It is of course a conceptualized guideline but not realistic. The physical implication of the statement is that, if the external energy loading to the KID is sufficiently low with respect to the indispensable readout energy, the KID should follow the postulated resolution enhancement up to a highest beneficial P_g, where the unconsidered effects introduced by the readout power itself begins to hinder the improvement. One expected example for such effects is when the readout power is raised to an absolute dominance, in which case the resolution asymptotes to be, not arbitrarily reduced, but limited by the fluctuation carried by the readout signal itself. There is, however, another more realistic limit that we have discussed in-detail in Sec. 4.3–the readout-power QP generation. We now study its impact on the resolution.

Supposed that we are allowed to raise the readout power to, not necessarily dominating the combined noise, but sufficiently high so that the readout-power generated n_{qp} consists the majority of the equilibrated total $n_{\text{qp,tot}}$. We have introduced the model in Eq. (4.49) and Eq. (4.50):

$$\Gamma_{\text{read}} = \frac{\eta_{\text{read}}}{\Delta} \cdot \chi_{\text{qp}} \cdot \frac{2Q_r^2}{Q_i Q_c} \cdot P_g \approx R n_{\text{qp,tot}}^2 V. \tag{4.224}$$

The terms from the right to the left in the above equation convert P_g to the resonator-coupled energy, the QP system-coupled energy, and the pair-breaking efficiency-modulated number of QPs, respectively. Since in this model we purposely create with the readout power a much elevated QP density, which we expect to result in a free recombination-dominated condition but not by the irreducible $\tau_{\text{qp,max}}^{-1}$ of Eq.

(4.107), we relate $\tau_{\text{qp,eff}}$ to the above result by Eq. (4.110) and obtain

$$\begin{aligned} n_{\text{qp}} &\approx \sqrt{\frac{\eta_{\text{read}}}{VR\Delta}\chi_{\text{qp}}\frac{2Q_r}{Q_i}P_g} && (\Leftarrow \Gamma_{\text{G}} \approx \Gamma_{\text{read}}) \\ &\approx \frac{1}{2R\tau_{\text{qp}}} && (\Leftarrow 2Rn_{\text{qp}} \gg \tau_{\text{qp,max}}^{-1}), \end{aligned} \tag{4.225}$$

where we also take $Q_c \approx Q_r$ for our particular design to simplify the result. The first and the second approximations in Eq. (4.225) represent the Γ_{read}-dominated and the free recombination regimes, respectively, therefore we have linked P_g to n_{qp} and τ_{qp} for Eq. (4.219). Before fully transforming the resolution model with the above assumptions and equations, we first discuss which regime we are in for Eq. (4.225). Assuming all the same typical values used in the GR noise modeling (Eq. (4.119) & Eq. (4.125)) and recalling that we find that η_{read} is in the range of $\lesssim \mathcal{O}(10^{-2})$, we assume $\eta_{\text{read}} = 0.01$ and $\chi_{\text{qp}} = 1$ since no noticeable non-QP loss mechanism has been observed for our prototype detector, and we find Eq. (4.225) predicts

$$n_{\text{qp(,read)}} \approx 4.4 \times 10^2. \tag{4.226}$$

Considering the uncertainties in the assumed inputs, this result is surprisingly consistent with the observation within a factor of ≈2. It provides a strong support that we are operating under the Γ_{read}-dominated condition, which also demonstrates that raising the readout power to dominate the QP GR rate is generally achievable in practice.

Without making further assumption for the QP lifetime, we substitute the first part of Eq. (4.225) that relates P_g and n_{qp}, together with the τ_{rise} and τ_{abs} models in Eq. (4.220) and Eq. (4.221), into Eq. (4.219) and obtain (still use $Q_r \approx Q_c$)

$$\sigma_E = \frac{V\Delta N_{\text{KID}}}{\eta_{\text{ph}}} \cdot \frac{Q_c Q_i}{Q_r^2}\frac{1}{\chi_{\text{qp}}}\sqrt{\frac{\eta_{\text{read}}}{VR\Delta}\chi_{\text{qp}}\frac{Q_r}{2Q_i}} \cdot \sqrt{\frac{2Q_c}{\pi f_r \mathcal{P}_{\text{abs}}\eta_{\text{fill}}\tau_{\text{qp}}}}\sqrt{\frac{k_{\text{B}}T_n}{\tau_{\text{qp}}}} \tag{4.227}$$

$$= \frac{\sqrt{V\Delta}N_{\text{KID}}}{\eta_{\text{ph}}} \cdot \underbrace{\sqrt{\frac{\eta_{\text{read}}Q_i}{R\chi_{\text{qp}}}}}_{\text{readout power}} \cdot \underbrace{\sqrt{\frac{k_{\text{B}}T_n}{\pi f_r \mathcal{P}_{\text{abs}}\eta_{\text{fill}}\tau_{\text{qp}}^2}}}_{\text{optimal filter}}. \tag{4.228}$$

We preserve in the first line the term-by-term correspondence with Eq. (4.219) so to help with interpreting the physical meaning of each term in the future, and we combine all the repeating parameters for the second line. We find that the explicit dependence of P_g is canceled, leaving only the parameters that participate the determination of the readout power (Q_i, η_{read}, ...) but are fixed for specific KIDs in the term noted by *readout power*. We realize that the cancellation is between the P_g in the denominator of $\sigma_{S_{21}}$ and the $n_{\text{qp}}(P_g)$ in $\partial\delta n_{\text{qp}}/\partial\delta S_{21}$. It indicates that, in the Γ_{read}-dominated regime, although one may always raise the readout power to achieve a larger signal size, the KID responsivity is suppressed by the same amount due to extra QPs created by the readout power, which is calculated by exactly the remaining parameters in the *readout power* term. Somewhat counter-intuitively, in this case the parameters that are nominally regarded as KID responsivity indices now positively correlate with the resolution, which reflects the undesirable reduction to the available signal at the output of the KID due to the energy consumption of the KID's QP system.

It might appear to the reader that the above analysis leads to the conclusion that one may carelessly apply a very large readout power, e.g., as high as possible without altering the refrigerator operation, so to ensure that σ_E saturates at the maximum as P_g is continuously removed from the equation in the Γ_{read}-dominated regime. It is not true if we further consider the τ_{qp}-P_g dependence that we have been ignoring so far. Eq. (4.228) clearly shows that, if one raises the readout power to create so many QPs that drive the system into the strong recombination regime, the suppressed τ_{qp} will start to degrade the resolution as one would naturally expect. In fact, it is not difficult to see from Eq. (4.107) that, given that a finite $\tau_{\text{qp,max}}$ always applies even at near 0 K due to practical material qualities, the criterion for decoupling the readout powder from affecting the QP lifetime is to have

$$2Rn_{\text{qp}}(p_g) \ll \frac{1}{\tau_{\text{qp,max}}}. \tag{4.229}$$

We then consider the longest QP lifetimes observed in the literature, roughly 2 ms, and find

$$n_{\text{qp,min}} = \frac{1}{2R\tau_{\text{qp,max}}} \gtrsim 25\ \text{QP}/\mu\text{m}^3, \tag{4.230}$$

where $n_{\text{qp,min}}$ is the equivalent minimal QP density corresponding to $\tau_{\text{qp,max}}$. This minimal QP lifetime may be set by the lowest possible quiescent QP population due to environmental energy insertion, whether thermal or non-thermal, or simply due the material's intrinsic properties if the environmental load is truly negligible.

Apparently, we currently have $n_{\text{qp}} \gg n_{\text{qp,min}}$, so we continue to include the second part of Eq. (4.225) into Eq. (4.228) for the strong recombination regime and obtain

$$\sigma_E = \frac{\sqrt{V\Delta}N_{\rm KID}}{\eta_{\rm ph}} \cdot \sqrt{\frac{\eta_{\rm read}Q_i}{R\chi_{\rm qp}}} \cdot \sqrt{\frac{k_{\rm B}T_n}{\pi f_r \mathcal{P}_{\rm abs}\eta_{\rm fill}}} \sqrt{\frac{4R\eta_{\rm read}}{V\Delta}\chi_{\rm qp}\frac{2Q_r}{Q_i}P_g} \tag{4.231}$$

$$= \frac{N_{\rm KID}}{\eta_{\rm ph}} \cdot \sqrt{\frac{8k_{\rm B}T_n\eta_{\rm read}^2 Q_r P_g}{\pi f_r \mathcal{P}_{\rm abs}\eta_{\rm fill}}}; \tag{4.232}$$

we again preserve the order of terms in the first line for future reference. Since the strong recombination regime is apparently not the ideal condition to operate the detector, there is not much to comment on for the resolution optimization. The main conclusion from this analysis is that the result indeed exhibits $\sigma_E \propto \sqrt{P_g}$ as expected and, interestingly, reduces to a set of parameters that are extrinsic to the KID, as opposed to Q_i, Δ, $\chi_{\rm qp}$, etc. Since the resolution degradation due to a high readout power is a consequence of the QP lifetime suppression, which is functionally introduced into the resolution formula via the noise bandwidth terms,[16] in principle we are able to observe the effect in read-time as suppressing the exponential decay time constant of the pulses. We may then quickly set a rough upper bound for the appropriate P_g. To further refine the optimal readout power, we may scan the OF resolution as a function of the readout power slightly below and close to the limit, e.g., a few to 10 dB below, and empirically identify the optimized P_g.

It is still worth noting that, if we rigorously optimize $\sigma_E(P_g)$ for Eq. (4.219) by

$$0 = \frac{d}{dP_g}\sigma_E \propto \frac{d}{dP_g}\left(\frac{n_{\rm qp}}{\sqrt{P_g}\tau_{\rm qp}}\right) \tag{4.233}$$

$$= \frac{d}{dP_g}\left[\frac{n_{\rm qp}}{\sqrt{P_g}} \cdot 2R\left(n_{\rm qp} + n_{\rm qp,min}\right)\right], \tag{4.234}$$

the algebra yields the expected (algebra in Appx. A)

$$n_{\rm qp} \ll n_{\rm qp,min}, \tag{4.235}$$

or explicitly

$$\frac{\eta_{\rm read}}{RV\Delta}\chi_{\rm qp}\frac{2Q_r}{Q_i}P_g \lll \left(\frac{1}{2R\tau_{\rm qp,max}}\right)^2. \tag{4.236}$$

We again substitute the typical parameter values for our prototype detector and indeed find that the optimal P_g for $\tau_{\rm qp} \approx 100\ \mu$s is

$$P_g \approx (-59\ \text{dBm} = 1.3\ \text{nW}) \times \left[\left(\frac{\eta_{\rm read}}{1\times 10^{-2}}\right)^{-1}\left(\frac{\tau_{\rm qp}}{100\ \mu\text{sec.}}\right)^{-2}\right]; \tag{4.237}$$

[16] Both the nominal noise bandwidth and the delayed energy insertion correction $\tau_{\rm qp}^{1/2}$.

here we only highlight the relevant dependences and refer the reader to the previous resolution results for other identical dependences. We have never rigorously optimized our readout power beyond empirically probing the highest power that does not create a visible resonance offset on the VNA monitor, but the theoretical optimal readout power in Eq. (4.237) turns out to be highly consistent with the $P_g \approx -60$ dBm that we identify. Recently, T. Aralis also tentatively carries out a readout power optimization test and find that the obtained OF resolution is indeed independent of P_g in -65~-59 dBm.

4.6.4 Future prospect

While we are actively investigating our current resolution results, we would like to conclude the section with encouraging future resolution upgrades. Now that we have discussed in great details for the optimization, we begin this section by assuming the optimal operation condition prescribed previously. In summary, one should apply a readout power as strong as possible, so with a carefully mitigated environmental energy load, hopefully the KID is easily dominated by the readout-power QP generation at a low readout power. The QP population created by the external energy injection, hence the required readout power to achieve the Γ_{read} dominance, should ideally be lower than the effective QP population given by the maximal QP lifetime. Combining the two conditions, one practically applies a readout power that generates a readout-contributed QP density that is equal or slightly lower than the effective $n_{\text{qp,min}}$ set by $\tau_{\text{qp,max}}$. In this ideal condition, $\tau_{\text{qp,max}}$ is preferably long for a narrower noise bandwidth, which in turn poses a stronger upper bound for P_g. Once we achieve the prescribed optimal operation, providing that the coupling (total) quality factor is also adjusted based on Eq. (4.220) and Eq. (4.221), we take $\tau_{\text{qp}} \approx \tau_{\text{qp,max}}$ for Eq. (4.228) and predict

$$\sigma_{E_{\text{ph}}} = \left(\frac{\sqrt{V}\Delta N_{\text{KID}}}{\eta_{\text{ph}}} \cdot \sqrt{\frac{\eta_{\text{read}} Q_i}{R \chi_{\text{qp}}}} \cdot \sqrt{\frac{k_B T_n}{\pi f_r \mathcal{P}_{\text{abs}} \eta_{\text{fill}} \tau_{\text{qp,max}}^2}} \right) \frac{1}{\sqrt{N_{\text{KID}}}} \tag{4.238}$$

$$\approx 19.4 \text{ eV} \tag{4.239}$$

$$\times \left[\left(\frac{V}{3.5 \times 10^4\ \mu\text{m}^3} \right)^{1/2} \left(\frac{\Delta}{180\ \mu\text{eV}} \right)^{1/2} \left(\frac{N_{\text{KID}}}{100} \right)^{1/2} \left(\frac{\eta_{\text{ph}}}{0.3} \right)^{-1} \right.$$
$$\left(\frac{\eta_{\text{read}}}{1 \times 10^{-2}} \right)^{1/2} \left(\frac{Q_i}{1 \times 10^6} \right)^{1/2} \left(\frac{R}{10\ \text{Hz} \cdot \mu\text{m}^3} \right)^{-1/2} \left(\frac{\chi_{\text{qp}}}{1} \right)^{-1/2}$$
$$\left. \left(\frac{T_n}{4\ \text{K}} \right)^{1/2} \left(\frac{f_r}{3.5\ \text{GHz}} \right)^{-1/2} \left(\frac{\mathcal{P}_{\text{abs}}}{5 \times 10^{-3}} \right)^{-1/2} \left(\frac{\eta_{\text{fill}}}{2.9 \times 10^{-2}} \right)^{-1/2} \left(\frac{\tau_{\text{qp}}}{2.0\ \text{msec.}} \right)^{-1} \right].$$

Notice that now we add the $1/\sqrt{N_{\mathrm{KID}}}$ factor to the parenthesized original Eq. (4.228), indicating that we assume achieving a robust cross-calibration among KIDs, which justifies averaging the individual measurements for noise cancellation. We also readopt $\eta_{\mathrm{ph}} \approx 0.3$ but not the value adjusted by utilizing Al capacitors, in the hope that the capacitors will be fabricated separately with higher-$\Delta(T_c)$ materials that are insensitive to phonons. O. Wen is currently refining this approach with improved fabrication techniques.

Other actively ongoing or planned DMKID R&D efforts that will improve the resolution but have not yet been discussed so far include:

1. Adapting quantum noise-limited amplifiers, which by-definition exhibit a "single-photon" noise, or more realistically

$$k_{\mathrm{B}} T_n = n_\gamma \cdot \frac{h f_r}{2}, \qquad n_\gamma \in \mathbb{N}, \tag{4.240}$$

where n_γ is the integer number of noise photons with $n_\gamma \leq 3$ but seldom perfectly 1 in realistic applications.

2. Adapting low-T_c materials for the KIDs, which, assuming the BCS model, is expected to reduce the pair-breaking bandgap by roughly

$$\Delta \propto T_c. \tag{4.241}$$

For low-T_c materials such as AlMn or Hf (Mazin et al., 2020), which have been demonstrated with T_c of <200 mK, the reduction in Δ relative to Al is roughly an order-of-magnitude.

Therefore, assuming $n_\gamma = 2$ and $T_c = 150$ mK, we anticipate further reducing the resolution from Eq. (4.239) to

$$\sigma_{E_{\mathrm{ph}}} \approx 1.3 \text{ eV}, \tag{4.242}$$

with which we conclude the resolution modeling work of this thesis. In the next section, we will introduce the ongoing efforts for realizing the above upgrades.

4.7 Facility upgrade for resolution reduction

Eq. (4.238) shows that, in order to further reduce the energy resolution, the potential approaches generally fall in the categories:

1. The design of the KID and the integrated detector layout. From the perspective of an individual KID, this approach includes the volume V, the phonon absorption

efficiency $\eta_{\rm ph}$, the resonant frequency f_r, and potentially also the RF characteristics such as the internal quality factor Q_i and the readout-power pair-breaking efficiency $\eta_{\rm read}$, due to different resonator designs. In terms of modifying the integrated detector layout, one can optimize the total number of KIDs $N_{\rm KID}$, the sensitive phonon-absorbing material coverage $\eta_{\rm fill}$, and again the RF-related parameters if considering not only the KIDs but also the feedline and the mechanical structures forming the RF environment.

2. The SC film and the micro fabrication, essentially the solid-state physics of the detector. If we fine-tune the fabrication techniques or even switch to other low-T_c materials, we adjust the QP bandgap Δ, the internal QP-generation efficiency $\chi_{\rm qp}$ (and of course Q_i), potentially the QP lifetime $\tau_{\rm qp}$ if it is limited by the material property, and with further understanding and control of the SC film-substrate interface, the pair-breaking probability $\mathcal{P}_{\rm abs}$ and hopefully also $\eta_{\rm ph}$, e.g., by adjusting the film thickness or the SC film-substrate interface condition. These proposals are realistically effective likely only after we carefully redesign the detector holder mechanics, so the phonon loss to the non-KID materials is largely mitigated.

3. Lastly and the most apparently, if we are able to *maintain the amplifier noise-dominated condition*, lowering the amplifier noise temperature T_n will directly benefit the resolution by $\sigma_{E_{\rm ph}} \propto \sqrt{T_n}$.

4.7.1 Kinetic inductance parametric amplifier

We introduce our ongoing works for implementing the quantum noise-limited kinetic inductance parametric amplifier (KIPA), developed by (Klimovich et al., 2019) and colleagues at Caltech and NASA JPL (Klimovich et al., 2019). demonstrated (2019) that the KIPA exhibits an absolute quantum noise-limited noise performance of $n_\gamma = 1$ in the range of a few GHz. Given our typical 3–4-GHz resonant frequencies, the quantum-limited noise performance is equivalent to the noise temperature of

$$T_{n,\rm KIPA} \approx 72 - 96\ \rm mK, \tag{4.243}$$

an astonishing near two orders-of-magnitude reduction in noise temperature compared to our 4 K HEMT low-noise amplifiers (LNAs). We are greatly benefited by the established installation and operation schemes by (Klimovich et al., 2019), for which our current plan is to duplicate as much of (Klimovich et al., 2019)'s experimental construction and experience with minimal necessary adjustments to satisfy our specific requirements (Zobrist et al., 2019). also adapt the same approach for a KID-based astronomical instrument and share with us useful experiences and guidance for practical KIPA implementation.

One distinction between (Klimovich et al., 2019) and our system is that (Klimovich et al., 2019)'s system is dedicated to characterizing the KIPA by itself, while we apparently install the detector in front of the KIPA to utilize its amplification. In (Klimovich et al., 2019)'s design, except those components that we also duplicate for driving the KIPA (next paragraph), the upstream circuit of the KIPA nominally consists of only coaxial cables and attenuators. According to our previous readout noise analysis (Sec. 4.5), as long as a sufficiently strong readout signal is available, one is always able to design the attenuation so that the input noise for the KIPA is dominated by the preceding lowest-temperature attenuator; this is exactly the approach in (Klimovich et al., 2019). Since the mounting temperature of the KIPA is only a fraction of Eq. (4.243), the added noise is subdominant to the KIPA and therefore permits an unambiguous demonstration for KIPA's quantum noise-limited performance. On the contrary, we elaborated in Eq. (4.126) and the surrounding discussion that, if we install our detector preceding the KIPA at the same temperature, the dominant input noise becomes the QP GR noise from the KID(s). Based on Eq. (4.126), our current QP GR noise may be as high as $\lesssim$1 K-equivalent thus, although currently subdominant to the HEMT, unfortunately dominates the KIPA by an order-of-magnitude. In other words, we should simultaneously suppress the QP GR noise by at least an order-of-magnitude, so the LNA noise-dominated condition is maintained for us to take the full advantage of the KIPA.

We provide here a brief theoretical justification for reducing the GR noise that supports the KIPA installation plan, and we will formally introduce our practical effort for the GR noise suppression work in the next section. We inspect the adjustable parameters in Eq. (4.126) for the GR noise suppression. Without redesigning or refabricating the detector, the only two adjustable parameters are the equilibrium QP density $n_{\mathrm{qp,tot}}$ and the readout power P_{read}, and we have explained that, to achieve the optimal resolution, we aim for applying a readout power that mildly dominates the quiescent QP population and therefore couples these two parameters. We then realize that, if we can reduce $n_{\mathrm{qp,tot}}$, not only do we suppress the GR noise linearly by reducing its intrinsic bandwidth $f_{\mathrm{qp,cut}}$ according to Eq. (4.126), but since the optimal readout power is reduced correspondingly, we further lower the optimal QP density and in turn also the GR noise by the linear optimal P_{read} dependence. To show the allowed optimal readout power reduction for the suppressed $n_{\mathrm{qp,tot}}$, we combine Eq. (4.237) and Eq. (4.225) and find that the optimal readout power should be lowered by

$$P_{\text{read}} \propto n_{\text{qp,tot}}^2. \tag{4.244}$$

Substituting the above result into Eq. (4.126) so to be considered with the intrinsic GR noise bandwidth, we find that the equivalent GR noise temperature is suppressed by

$$T_{n,\text{GR}} \propto n_{\text{qp,tot}}^3 \tag{4.245}$$

under the optimal readout condition. Based on the fact that we currently observe a QP density that is 10–20 times higher than the lowest achievable values in the literature, Eq. (4.245) suggests that the $O(10)$ GR noise suppression required by the KIPA is realistic with ample surplus. Since a quantum-limited amplifier provides theoretically the lowest possible noise, once we achieve the KIPA-compatible low n_{qp}, we can safely remove the GR noise from future noise concerns. Even for more advanced techniques evading the quantum limit like the squeezed readout, which typically offer order-of-unity further noise suppressions in recent demonstrations, the theoretical GR noise suppression given by Eq. (4.245) is still orders-of-magnitude away from being relevant in practice. It is worth noting that, aside from allowing better LNA options, lowering the $n_{\text{qp,tot}}$ also linearly raises the QP lifetime per Eq. (4.71) and therefore directly impacts the resolution through the OF noise bandwidth by (Eq. (4.238))

$$\sigma_{\text{E}} \propto \tau_{\text{qp,tot}}^{-1} \propto n_{\text{qp}}. \tag{4.246}$$

In order to introduce the project in an understandable order, without going into much detail until the next section, we state for now that the techniques for suppressing the quiescent $n_{\text{qp,tot}}$ generally involve decoupling the device from undesirable QP-generating external energy sources, whether the energy is transmitted through radiation, thermal/electronic conduction, mechanical forces, or any other channels. In practice, engineering for decoupling these energy transmissions always implies carefully modifying the existing experimental infrastructure, so the energy is absorbed and then dissipated before it reaches the device. In parallel with implementing the decoupling mechanisms, to support the KIPA at its quantum-limited performance, (Klimovich et al., 2019) also suggests a substantial upgrade to our current readout electronic infrastructure. Therefore, the facility upgrades required by both purposes motivate an integrated renovation for our system. Fig. 4.24 (input circuit) and Fig. 4.25 (output circuit) present the electronic schematics for the upgrade, with Tab. 4.10 lists the electronic components selected specifically for our system. The major

Component	Range (GHz)	Insertion/return loss (dB)	Isolation (dB)	Comment
Anritsu K250	0.1–0.5	<1.2^{T}/ >15	>20	
	0.5–20	<1.2^{T}/ >15	40	
	20–40	<1.2^{T}/ >10	40	
Marki DPX-1114	0–11	1.2^{T}/ 12–20^{T}	25–40^{T}	LPF port
	14–28	1.2^{T}/ 8–12^{T}	40–50^{T}	HPF port
	0–28	0.35–0.51[1] / 24–42[1]	-	Common port
Marki PD-0218	2–18	1^{T}(+3)/25^{T}	15–22^{T}	
Radiall R595.333.415	0–6	0.2–0.3^{M}/ 23.1[2]	85	Insertion loss dominated by dielectric loss
Radiall R595.333.115	0–6	0.2–0.3^{M}/ 23.1[2]	85	
QuinStar QCJ-G0400801AM	4–8	0.2^{M}/ 9.5^{m}[2]	19^{m}	Shielded against 0.15T
Marki FLP-1250	0–11.5	0.6^{T}/ 15–20^{T}	-	
	15.5	25–30^{T}/ 0.4[3]	-	
	19.5–32	45–50^{T}/ 0.9–1.9[4]	-	
CMT CIT412	4	+30–32^{T}/ 20–25^{T}	-	
	8	+30–32^{T}/ 15–18^{T}	-	
	12	+30–32^{T}/ 19–21^{T}	-	

^{T,M,m} Typical, maximal, minimal.

[1] S(LPF-com) and S(com-com) in 3–5 GHz, see test data.

[2] S11 calculated from VSWR.

[3] See S11 in test data.

Table 4.10: Summary table for DMKID KIPA readout electronics. More details may be found in the corresponding text and Fig. 4.24/4.25.

upgrades to the system aside from the apparent KIPAs are the two room-to-base-temperature feedthrough sub-circuits in the input schematic, labeled as the "KIPA DC bias" and the "KIPA pump," and the two reversed feedthrough sub-circuits in the output schematic colored correspondingly. Unlike in the input schematic, the output sets terminate at different cryogenic stages but do not return to the room temperature.

For an in-depth introduction to the the functionalities and the underlying physics of the DC bias and the pump sub-circuits in the KIPA operation, we recommend (Eom et al., 2012; Zmuidzinas, 2012) for comprehensive discussions and an early realization presented in (Eom et al., 2012). For more recent developments and the detailed description for the particular KIPAs that we acquire for the current DMKID application, we recommend N. Klimovich, S. Shu, and P. K. Day's works on the topic since 2018, especially N. Klimobich's PhD dissertation.[17] In summary, KIPA operates as a RF waveguide that exhibits an extreme dispersion relation due to the

[17] Expected to complete by 2022–2023.

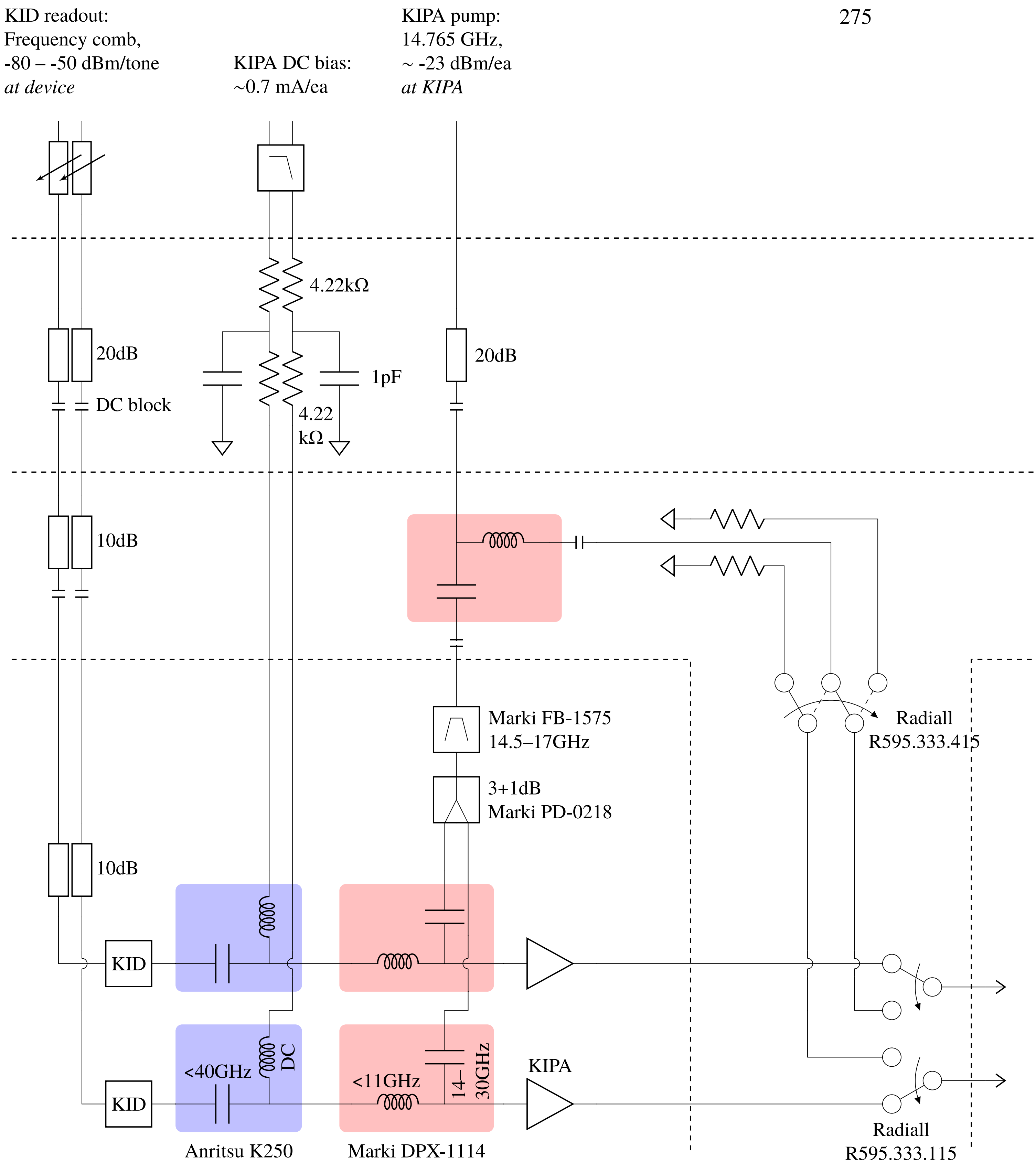


Figure 4.24: The input circuit diagram for DMKID detectors with the KIPA readout. The dashed lines mark different temperature stages, from top to bottom, room temperature environment, 4 K LHe stage, 0.9 K still stage, and the base temperature stage. The arrows continue to Fig. 4.25.

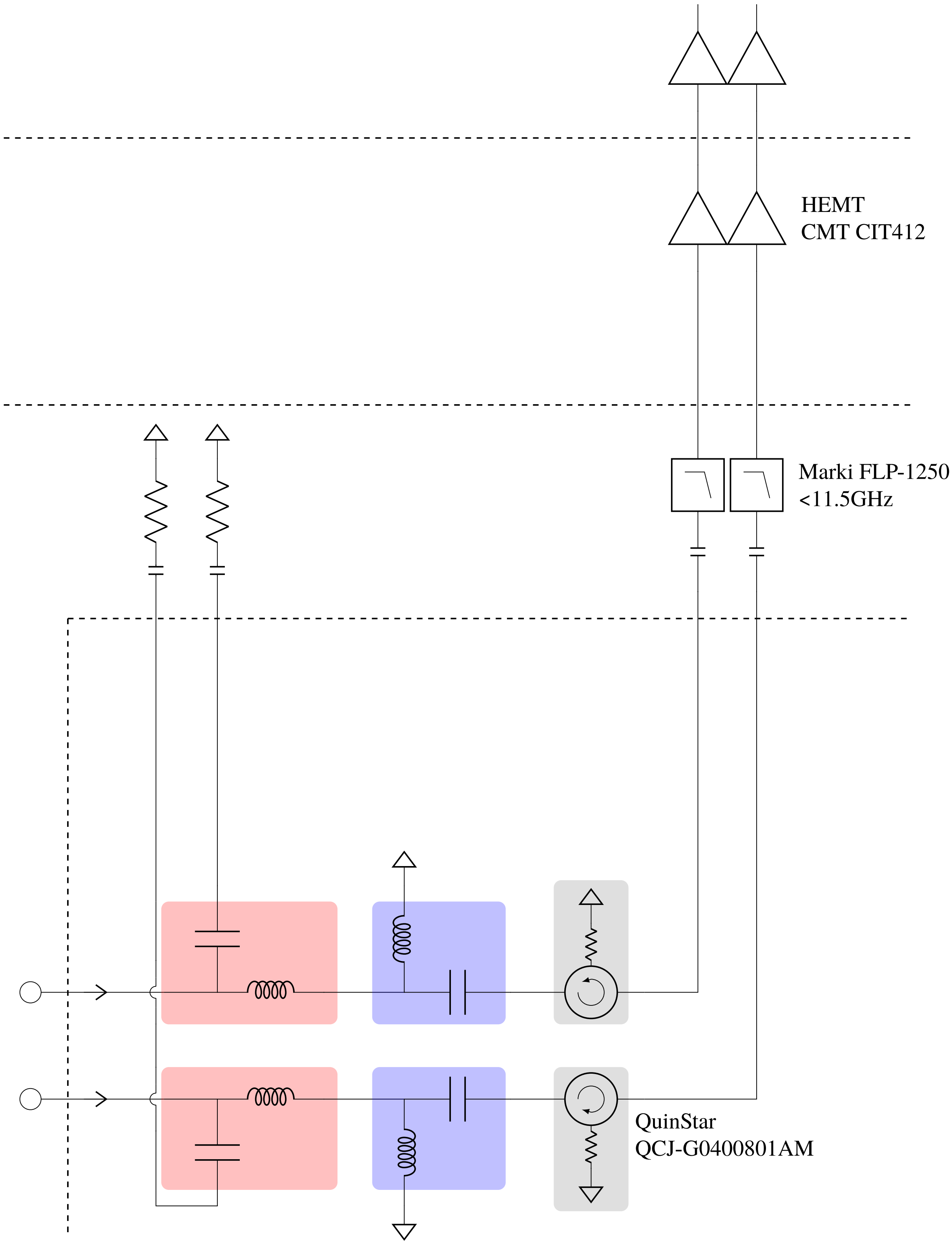


Figure 4.25: The output circuit diagram for DMKID detectors with the KIPA readout. The dashed lines mark different temperature stages, from top to bottom, room temperature environment, 4 K LHe stage, 0.9 K still stage, and the base temperature stage. The arrows continue from Fig. 4.24.

nonlinearity of the kinetic inductance and therefore allows the realization of a parametric amplification similar to the Kerr effect for optical fibers (Eom et al., 2012). To match the much raised kinetic inductance, our KIPA as designed by (Klimovich et al., 2019) utilizes an in-waveguide continuous IDC design to construct a 50 Ω CPW. The waveguide then transmits at a group velocity that is orders-of-magnitude reduced from the vacuum speed of light with a steep dispersion relation. The net consequence after different modes travel multiple wavelengths to the output of the waveguide is then, identical to the Kerr effect, a redistribution of the energies originally carried by different frequencies. Therefore, by manually supplying a monotonic "pump" signal carrying a large energy into the original signal, the waveguide effectively serves as an amplifier that energizes the original signal with the injected pump energy. In fact, this energy distribution S-matrix is just the gain profile for the amplifier, which is "parameterized" by the controllable nonlinearity or equivalently the dispersion relation. One can draw an analogy from the commercial semiconductor amplifiers: The KIPA pump RF energy insertion is the usual source/drain combination that supplies the energy for the amplification. Since the SC waveguide's kinetic inductance is nonlinearily enhanced by the increased DC current flowing in the SC film, the KIPA DC bias current realizes the parametrization and thus is regarded as the usual gate/base combination that modulates the channeling of the signal through the amplifier. Limited by the scope of this thesis, we unfortunately can only provide this unsatisfying analogy for such an intriguing topic on its own, so we highly recommend the interested readers to (Eom et al., 2012; Zmuidzinas, 2012) for further readings.

Specifically for our KIPAs, in order to achieve a high gains with large DC biases, the CPW waveguides are fabricated of NbTiN, a SC material known to carry high critical currents, $T_c > 10$ K, and exhibit high kinetic inductances as well as high DC-current nonlinearities. As shown in Fig. 4.24, our KIPAs nominally operate at a 0.7 mA DC bias and a 5 μW (-23 dBm) pump power carried at ≈15-GHz, which is far from our KID resonant frequencies. The pump and the DC bias are combined into the signal through, respectively, a diplexer and a bias-Tee, shown in the figure by the blue and the pink rectangles. Depending on the fabricated qualities of specific KIPAs, the resulting gains vary in a typical range of 15–25 dB in 1–5 GHz, with the advantage that one may flexibly adjust the amplification bands by rearranging the pump frequencies with respect to the detectors' resonator bands. In the DC bias feedthoughs, we utilize various low-pass filters at room temperature and at 4 K to improve the DC quality and in turn the gain stability similar to the

HEMT gate control. In the pump feedthrough, we attenuate the thermal noise at 4 K based on the same principle for the readout signal (Sec. 4.5), and we band-pass-filter the 15 GHz pump signal at the base temperature and branch it to each KIPA using a RF splitter. Instead of splitting the pump signal at the base temperature, we can in principle install multiple parallel pumps independently for the KIPAs, but we preliminarily choose the splitting scheme so to minimize the physical cutouts required by the number of coaxial cables penetrating the 1-K Cryoperm magnetic shield. On the contrary, we do not anticipate the thin DC wires to present such a constraint, therefore we choose the simpler parallel DC bias scheme as in Fig. 4.24. Notice that, compared to our original readout circuit design in Fig. 4.16, we remove the 1 dB attenuator between the detector and the HEMT LNA. We originally employ this attenuator to damp the potential standing wave formation between the detector and the downstream components, most likely between the inherently impedance-mismatched KIDs and the HEMT input. For the KIPA readout, this attenuator would catastrophically dissipate the strong DC bias current at the base-temperature stage of the cryostat.

In the lower-right of Fig. 4.24, there is a sub-circuit consisting of three RF switches, two 1 K terminators, and a central line that branches from the KIPA pump feedthrough at 1 K by a diplexer. The signal lines of this sub-circuit join the main signals at the KIPA outputs. Recalling Eq. (4.144) and the corresponding 4 K/1 K noise temperature calibration technique, one may understand that this sub-circuit provides 1 K thermal calibration references for the KIPAs using the two 1 K terminators. We choose the identical diplexer for combining the pump tones into the signals for the central line of this sub-circuit (notice the high-/low-frequency port orientation), so in principle the intended pump tone passes this branching diplexer and reaches the signal lines at the bottom of Fig. 4.24 for driving the KIPAs. The low-frequency port for this branching diplexer, however, directs the low-frequency signal in the pump feedthrough that does not contain the pump tone to the RF switch sets as another noise calibration reference. Due to our principle for installing attenuators, the pure-noise input of this diplexer is dominated by the 20 dB 4 K attenuator above it and therefore presents to the calibration sub-circuit a 4 K thermal spectrum up to the low-frequency cutoff of the diplexer. Again, since we select the identical diplexer for combining the KID signal and the pump, we know that the low-frequency band of this diplexer covers the KID frequencies, therefore the branched partial 4 K thermal spectrum also covers the KID frequencies and serves as an appropriate 4 K calibration reference for the KIPA at the KID resonant frequencies. The advantage of adapting this rather non-straightforward noise branching technique is that, since we mount the additional diplexer at 1 K, we avoid installing yet another 4 K-to-1 K cable that requires a magnetic shield cutout, which is otherwise needed by a simple 4 K thermal noise source. However, we should note that, in order to achieve an accurate calibration with this technique, we must carefully account for the losses of the diplexer and the switches (Tab. 4.10), as well as an accurate noise temperature assumption for the supposedly 4 K 20 dB attenuator. If the upstream noise up to the attenuator is high, e.g., 300 K without loss or a thermalization to a lower temperature, the total output noise of the attenuator is about 7 K-equivalent but not 4 K, which apparently introduces a nonnegligible uncertainty to the calibration. We can use the 1 K terminators to calibrate the branch line noise. Alternatively, we can in principle replace the 20 dB attenuator with a stronger one so to enforce the noise emission at 4 K. However, assuming that the KIPA does exhibit a near quantum-limited $\lesssim$100-mK $T_{n,\text{KIPA}}$ with a 25 dB gain, the expected LNA output noise temperature that we intend to calibrate is much higher than a few Kelvin, i.e., $\lesssim$35 K, so we determine that, for a preliminary plan, a calibration reference with a higher noise temperature is beneficial as long as it can be calibrated independently. The calibration technique detailed above is scheduled to be verified and fine-tuned in a 4 K R&D cryostat in Summer–Fall, 2021, in parallel with the DR mechanical renovation for hosting the KIPA and associated hardware.

Once we properly construct the readout circuit as in Fig. 4.24, we in principle can send the KIPA-amplified signal to the original HEMT-equipped downstream readout system. However, (Klimovich et al., 2019) point out based on their quantum-limit noise calibration experience that, in order to achieve an observable KIPA-limited noise performance, it is crucial that one carefully inspects the entire readout circuit and mitigates every possible source of interference. In particular, (Klimovich et al., 2019) found that the resonances formed in front of or across the KIPA drain the pump energy and suppress the KIPA amplification. Following (Klimovich et al., 2019)'s approach, we employ the identical bias-Tee and diplexer pairs in the output circuit diagram (Fig. 4.25) mirroring the DC bias- and pump-combining sets in the input circuit diagram, so we manually remove the strong DC and RF KIPA-driving energy that we manually inject into the main signal lines. With any small downstream impedance mismatches, such as at the HEMT inputs,[18] the strong energy would create a large backward flow that completely destroy the operation

[18] HEMTs block backward transmissions, i.e., S_{12}, so impedance mismatches after the HEMTs are less likely to cause this issue.

of the KIPAs. Besides this main consideration, most types of the commercial electronics downstream to the KIPAs simply can not sustain such a high energy without being compressed or permanently damaged. However, even after manually removing the DC and the RF KIPA drivers, (Klimovich et al., 2019) still found that the noise performance of KIPA is far from the theoretical expectation, namely the $\hbar\omega/2$ quantum limit. One major achievement by (Klimovich et al., 2019) since (Eom et al., 2012)'s early demonstration is that (Klimovich et al., 2019) was able to identify the resonances generated by small backward transmissions, formed by impedance-mismatched nodes across the KIPA, as the main cause to the degraded noise permanence. These useless resonances drain the pump energy into amplifying themselves but not the indented signal. Furthermore, after finding that the HEMT input to be the most mismatched node that caused the dominant reflection, (Klimovich et al., 2019) eventually demonstrated an unambiguous quantum-limited noise performance by utilizing an isolator at the input of the HEMT. The isolates are shown in Fig. 4.25 by the gray rectangles. The disadvantage for (Klimovich et al., 2019)'s approach is that, since commercial isolators generally use ferrite-based magnetic compounds to terminate the circulated wave, all the isolators inside the 1 K Cryoperm magnetic shield must be further shielded to prevent interfering with SC devices. While these shields already require a substantial space with extra cooling power/time, the requirement is made even more constraining by the fact that the ferrite terminators in the isolators are also thermal noise emitters, so ideally these bulky isolators should be mounted at the base temperature to best suppress their backward-emitting thermal noise. In addition to the isolators, we also install DC breaks and low-pass filters to dissipate the residual and thus resonance-forming DC bias and pump before the HEMTs. Contrasting the isolators, these DC breaks and low-pass filters do not need to be at the base temperature, so we install them at the much wider 1 K space outside the magnetic shield.

Finally, we discuss alternative future plans before concluding the readout circuit upgrade discussion. First, commercial RF switches typically utilize mechanical mechanisms that literally rotate the SMA connectors to achieve the switching. Whether the rotation is done by mechanical or electromagnetic forces, the frictional (plus Joule heating for the latter) energy dissipation generally much exceeds the cooling powers of commercial DRs even for the most powerful ones. Specifically for the Radiall magnetic-mechanical switches that we select (Tab. 4.10), the heat dissipation is about 3.4–3.8 W, or a 50–60 mJ integrated dissipation for each switch. The authors of (Zobrist et al., 2019) shared with us that, if installed at the mixing chamber temperature, these switches require at least 4 hours for the refrigerator to recover from the switching thermal shocks. For our DR with its 25 μW cooling power, we believe that the thermal shocks will simply stop the dilution circulation. Therefore, we decided to heat-sink the switches and also the branching diplexer for the 4 K calibration to the 1 K stage (Fig. 4.24). Due to the same cooling power constraint, we dissipate the 5 μW/ea pump signals to 1 K but not the MC as shown in Fig. 4.25. However, this constraint does not apply to the DC biases, so we simply short them to the chassis, i.e., the refrigerator's base temperature mechanical structure. Comparing to straightforwardly installing the above instruments at the base temperature, the additional base-to-1 K wires are expected to introduce a noticeable but nonetheless manageable heat load to the MC. Second, we decide to install the DC blocks and the low-pass filters that cancel the HEMT reflection at 1 K for but not the MC. In principle, this choice also helps with saving the MC cooling power from compensating the filtered energy, but the main motivation is to save the MC space that is now foreseeably crowded by many additional electronics. We have explained that we require the isolators to be at the base temperature so to reduce their backward thermal noise emission. If the requirement is met, we expect that any backward-emitting noise or unwanted signals to be isolated by the isolators, which then convert the absorbed energy into MC heating and the isolators' base-temperature noise. The authors of (Klimovich et al., 2019) shared based on their extensive fine-tuning experiences that the best KIPA noise performance occurs when single-junction isolators are utilized (see Tab. 4.10, 19 dB insertion loss), contradicting their original intuition that the double-/multi-junction models of the same product that we choose offer stronger isolations and therefore expectantly a better KIPA performance. Their hypothesis is that, even though an increasing isolation provides a better HEMT reflection blockage, the higher insertion loss is more likely to promote the formation of other standing-wave loops and result in the same issue the HEMT causes. Third, we anticipate that the KIDs, which are inherently impedance-mismatched nodes, to cause reflections and potentially promote the same standing-wave formation. In particular, we worry that the resonances may form between the KIDs and the SMA connections of the bias-Tees, the diplexers, and the KIPAs in the input-circuit diagram. This was not an issue and has not been tested in (Klimovich et al., 2019) due to the absence of an actual detector. Not only do we know that these resonances will suppress the KIPA amplification, they are also likely to be highly detrimental to the sensitivities of the KIDs. To prevent this effect, we reserve the MC space and design the mechanical fixtures for installing additional isolators between each KID device and its KIPA. Finally, de Visser et. al. suggested us to install Eccorsorb-based coaxial thermal noise filters at the base-temperature entrances of the signal paths and their corresponding base-temperature exits, i.e., before the KIDs and after the isolators, respectively. The purpose is to prevent forward and backward thermal noise from higher temperatures to create excessive QPs in the KIDs. Since these filters are not directly related to the KIPA installation and are relatively unlikely to cause issues with other components, we suppress them in the circuit diagrams.

4.7.2 Ultra-low radiation facility

We seek for a cryogenic material, mechanical, and electronic facility renovation that reduces the quiescent QP density in the KIDs and simultaneously physically accommodates all the aforementioned electronic components for the KIPA installation (Baselmans et al., 2012; de Visser et al., 2014). demonstrated that one dominant channel for depositing unintended pair-breaking energy into sensitive SC devices is through the thermal radiation in the cryostats. Although typically cryogenic systems are designed in the "Russian doll-like" multi-layer encapsulation structure, which seemingly provides multiple hermetic radiation shields for the devices in the coldest inner-most chamber, it was noticed during the early KID development that the KIDs' quiescent QP densities unambiguously associated to well-identified thermal radiation sources that located many layers outside of the detector chambers (de Visser et al., 2014). Based on basic thermal physics, everything with a finite temperature radiates as a blackbody, therefore the devices are expected to accept certain levels of thermal radiation of different temperatures, since they must be realistically surrounded by physical experimental apparatus. However, if the detector chamber is truly hermetic and maintains its temperature, no matter how strong the outside radiation source is, the blackbody photon bath that the detector is in should remain unchanged, which contradicts the observation. It proves that the Russian doll structure is in fact not as hermetic as needed to prevent the thermal photons emitted by the outside sources or the warmer outer chambers to enter the detector chamber before being absorbed and thus thermalized. It also suggests that the photons causing the KID performance variation do not belong to the base-temperature thermal radiation bath that the KIDs are in. It is then easy to understand that, for photons to penetrate so many visually hermetic containers, they undertake numerous scatterings and reflections, effectively bleeding through tiny cracks and mechanical contact slits in the shielding structures like a leaking gas. Due to the importance of the phenomenon, we have a full chapter later in this thesis dedicated to the physics principle, the simulation, and the experimental calibration for the thermal photons in cryogenic systems. For introducing the facility upgrade, the above pictorial explanation is sufficient to guide the design of our new cryogenic infrastructure.

Based on the above picture that the blackbody photons are heavily confined in their corresponding cryostat chambers are are only allowed to occasionally leave the enclosures after countless random reflections on the cavity surfaces, (Baselmans et al., 2012) demonstrated that one can effectively suppress these non-thermalized, heavily scattered "leaking" photons with a "box in box" device housing strategy. Shown in Fig. 3.5, instead of holding their device in a container that directly locates in the open thermal-photon environment, i.e., simply mounting the detector holder on the base-temperature plate, the authors used another wider container to fully enclose the original device holder and then placed the assembly, which they dubbed the box in a box, in the open cryostat space. Suppose there exists a small probability $\mathcal{P}$ for a photon to undertake numerous reflections on the cavity walls until entering the next chamber through a small connecting channel, by adding more of the box-in-box enclosures, the resulting probability for the photon to survive the passage into the device space is exponentially suppressed by $\sim\mathcal{P}^n$, where n is the number of chambers to penetrate. (Baselmans et al., 2012) therefore proposed that, in order to leverage the reflection-controlled leaking probability, in addition to adding more enclosures and thus raising n, one should also 1) carefully redesign the mechanical slits in the containers, so the number of reflections needed for the photon to leave the container increases, and to ensure the most effective confinement, 2) the mechanical slits and the cutouts for cable feedthroughs should all misalign to each other to prevent a straight line-of-sight photon penetration. These techniques aim to increase p, which is generally more practical than increasing n considering the limited cryostat space. Due to the concept, we name the space between the inner and the outer containers the scattering chamber, and the inner container space the detector chamber. In order to fully leverage such a reflection-based radiation shielding scheme, (Baselmans et al., 2012) also proposed coating the scattering chamber surfaces, namely the inner and the outer walls of the outer and the inner containers, respectively, with a highly absorbing material in the anticipated thermal radiation frequency range. With the radiation-absorbing coat, the base probability $\mathcal{P}$ is further suppressed by $\sim (R_{\text{black}}/R_{\text{Cu}})^N$, where R_{black} and R_{Cu} are the reflectances of the coating material and the unpainted gold-plated copper surface, respectively. Know that gold or copper exhibit $R > 99$ % in the long-IR to sub-/mm-wave range, while dedicated IR-absorbing materials can easily achieve $R < 1$ %, the suppression in p is orders-of-magnitude, which is further raised by the power of n for the total shielding.

The authors of (Baselmans et al., 2012) shared with us through private conversations that the specific IR/mm-wave-absorbing material that they employed was a mixture of 1) Epotex-920 epoxy resin as the mixture matrix mixed with 2) 3% of Lampblack by weight. They combined this fluid mixture with crushed irregular SiC grains at the sizes of about 1 mm and then applied the mixture to the surfaces. Lampblack is also called the carbon black, which is a form of fine pure carbon powder sieved by at least #350 mesh thus is <30 μm in particle size and commonly below 1 μm in commercial products. The authors noted that the recipe was derived from earlier experiences of (Diez et al., 2000; Hargrave et al., 2000; Lamb, 1996) for similar sub-mm wave absorption purposes and may still be improved depending on applications. In addition to the recipe, (Baselmans et al., 2012) also emphasized that it is crucial for the coating to be a few times thicker then the indented wavelengths, so the photons experience the absorption that is dominated by the bulk lossiness of the compound but not the interface properties of materials. Given that the pair-breaking energy for SC Al ($2\Delta \approx 360$ μeV) corresponds to a single-photon wavelength of $\approx$3.4 mm in vacuum, (Baselmans et al., 2012) constructed their coating to be about 3 mm in thickness. Considering the epoxy refractive index and that the epoxy-metal interface under the coating also reflects the photons, the effective photon path in the compound is more than 3 times of the wavelength (Baselmans et al., 2012). then added the SiC grains at the sizes that are not small relative to the wavelengths of interest so to make the compound an inhomogeneous medium for the photons, which promotes the random photon scattering and subsequently the effective photon path in the medium. The visibly irregular surface of the coating also helps with suppressing the refection at the compound-vacuum interface (Baselmans et al., 2012). pointed out that, in addition to enhancing the absorption at <500 GHz ($\approx$4.8 K blackbody radiation peak) relative to the commonly applied Stycast, the fine Lampblack powder also greatly helps with the mixture's adhesion to smooth copper or gold surfaces. In practice, however, the viscous liquid-solid mixture was still found challenging for constructing the full 3 mm layer in one application. Instead, the authors successfully constructed the coating by dividing the full process into a few steps, where they incrementally stacked and fully cured each 0.5–1-mm layer until it reached the targeted thickness. We were also suggested by the authors to irregularize the surface of each intermediate layer with a pencil, which also contains carbon and seemed to help with the layer adhesion and potentially also the internal refraction of the coating.

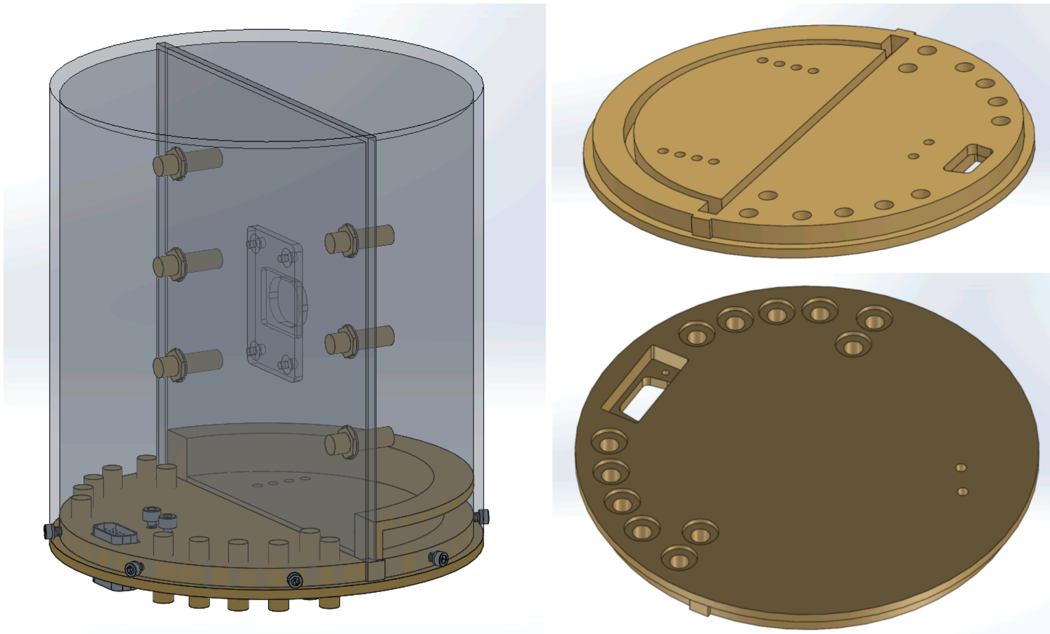


Figure 4.26: DMKID ultra-low radiation detector housing designed for Caltech Oxford Kelvinox 25 DR. Left: The full construction including the can with the guarding chamber, the middle plate, and the bottom feedthrough lid. Top right: The inner side of the bottom feedthrough lid, showing the slots for the guarding ring and the middle plate. Bottom right: The outer side of the bottom feedthrough lid, showing the indents for the feedthrough connector flanges. A detailed description is provided in the corresponding text.

Fig. 4.26 presents the design of our box-in-a-box ultra-low radiation detector container. Unlike in (Baselmans et al., 2012), where the box-in-a-box structure is realized with the two apparent rectangular boxes enclosing the detector (Fig. 3.5), our design is based on our reinterpretation for the underlying concept and not as apparent as a box in another box. Due to our DR's limited base-temperature space for fitting not only the detectors but also the KIPA-readout electronics discussed previously, we realize that it is impossible for our DR to offer the space for two independent enclosures and still offer a sufficient space for the electronics. So instead, we retreat to the design concept that requires all the photons bleeding into the layered containers to be first confined and heavily scattered in a dedicated scatter chamber, before these confined photons find the mechanical channels leading into the main detector chamber. As shown in Fig. 4.26, if ignoring the electronic feedthroughs for a moment, the only possible channel for the outside photons to enter the container is through the contact slit between the bottom lid and the cylindrical sidewall of the can. The top lid is hermetically brazed to the sidewall. With the vertical middle plate separating the in-can space, the half-cylindrical space on the left-hand side in the figure becomes the scattering chamber for the photons entering from the left half of the bottom lid slit. However, if we then place the detectors in the right-hand side chamber, the photons entering the can from the other half of the slit can directly enter the main detector chamber without pre-scattering. Therefore, we create the half circular, half inch-wide "guarding ring" chamber along the right-hand side half of the bottom lid slit, so this half inch-wide chamber serves as the scatter chamber for the photons entering from the right. As in (Baselmans et al., 2012), we coat the inner surfaces of the half-cylindrical chamber on the left-hand side and the guarding ring chamber with the IR-absorbing material. It is due to the thickness requirement discussed previously for an effective IR-absorbing coat and also the practicality for applying it that we determine the width of the guarding ring space. In addition to these two scattering chambers, we also add two pins to the bottom lid that align to the middle plate, which enter the middle plate when assembled. They block the straight pathways for the photons entering the construction from neither the left- or the right-hand side but in the plane of the middle plate and therefore avoiding the scattering chambers.

Fig. 4.27 presents the fabricated work. In the design, all the mechanical contacts that define the isolated spaces use slotted and/or lipped matings as shown in Fig. 4.26 and Fig. 4.27. These locations include the bottom lid's contact with the cylindrical sidewall, the guarding ring, the middle plate, and all the bolting and feedthrough connector indents on the bottom lid and the middle plate surfaces. We consider the 50 μin (1.3 μm) gold-plating thickness for the dimensions so to obtain <10 μm mating tolerances for the fabricated contact slits, which ensures that the sub-mm thermal photons can only penetrate these isolated space barriers in extremely suppressed evanescent modes.[19] The slots and lips also require the photon propagation to undertake at least two 90° turns thus further suppress the radiation hermiticity. The complex main cylindrical can, with the guarding ring and the slot on the inner surface for the middle board, is fabricated by first brazing the half-circular guarding ring piece, which is machined from a separated raw material, to the inside of a standard cylinder, and then the long vertical slots for the middle plate are engraved by wire-EDM (electrical discharging machining), and finally a disc piece with a straight slot in the middle is brazed to the top of the cylinder. The brazing joints are all done to a vacuum-hermetic grade thus also ensure radiation hermiticity. If we find the entire construction is made overly hermetic for a smooth

[19] See next chapter for a quantitative analysis.

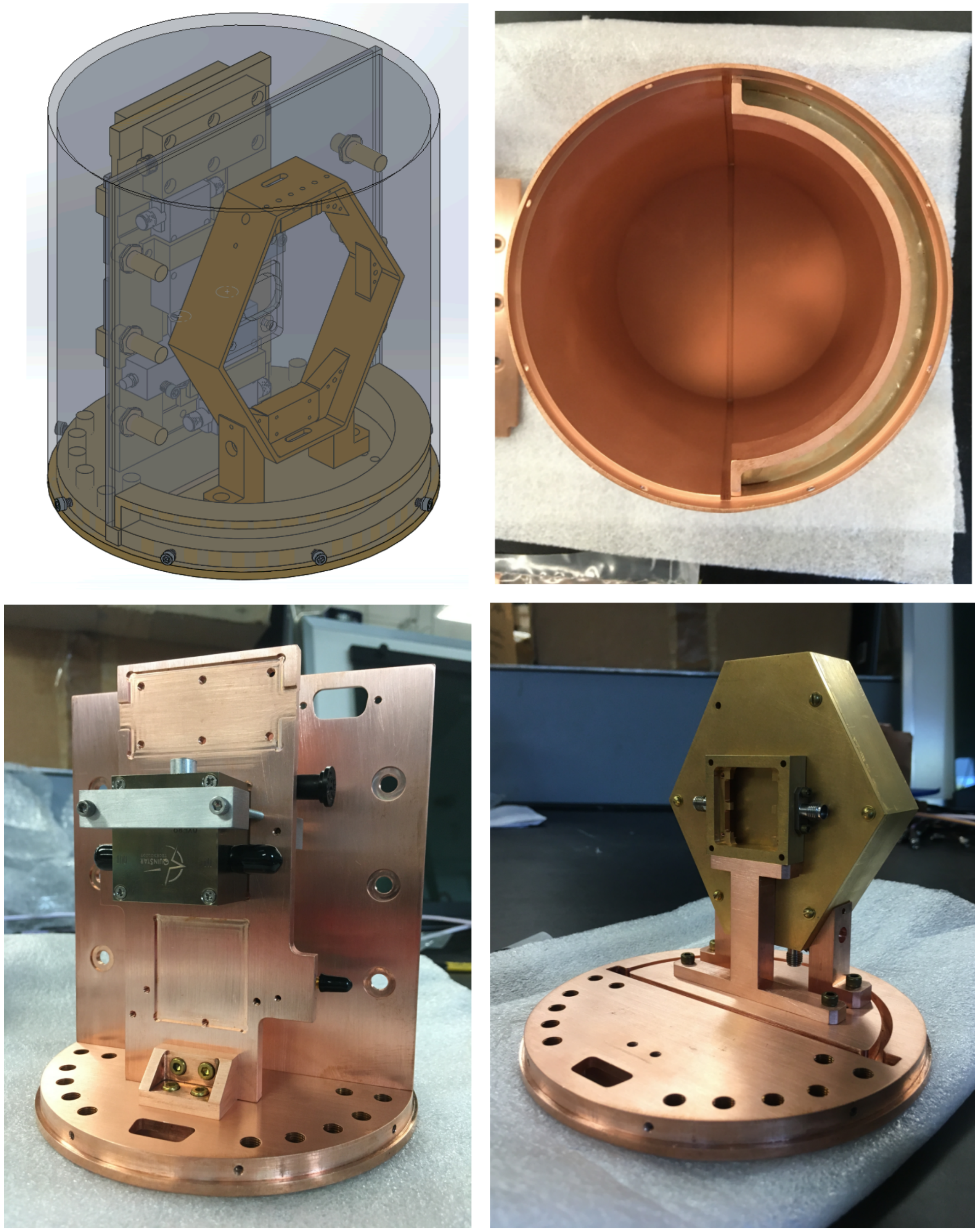


Figure 4.27: Fabricated (before gold-plating) DMKID ultra-low radiation detector housing for Caltech Oxford Kelvinox 25 DR. Top left: The full construction including the complete detector housing, the KIPA electronics on the left, and the 3" detector holder on the right. Top right: The brazed guarding ring, the slotted sidewall, and the slotted top lid. Bottom left: The bottom lid with the middle plate inserted to the slot. The photo shows the electronics chamber side with the electronics fixture board and a magnetically shielded QuinStar isolator. Bottom right: The 3" (back, large) and the low-threshold (front, small) detector holders assembled to the bottom plate. More description may be found in the corresponding text.

evacuation, we plan to retrofit 0.5 mm wide/deep slots on the surface of the bottom lid's lip at both the detector and the non-detector sides, so these surface slots form pumping channels when the bottom lid assembles with the can. We plan to apply the same technique for the middle plate slot in the bottom lid, so the detector chamber may be evacuated through the large scattering chamber. Nevertheless, since adding channels is always disfavored by the radiation tightness, we will only do so if it is later proven necessary, e.g., showing a virtual leak.

Our design concept utilizing the middle plate and the guarding ring chamber allows us to efficiently use the space to install the KIPA-driving electronics in the large half cylindrical scattering chamber, including the two pairs of the bias-Tees and the diplexers with the KIPAs in Fig. 4.24. They connect to the two detectors in the detector chamber via the feedthrough connectors on the middle plate and the input/output readout wiring via similar feedthroughs on the bottom lid. As shown in Fig. 4.27, these electronics support operating a 3" detector and a small low-threshold detector simultaneously. Among all the electronics that we plan to mount at the base temperature, the KIPAs are SC devices and therefore are prioritized to be contained in the low-radiation housing for a suppressed radiation load. Due to a sufficient space, it is then more straightforward to also contain the DC bias- and the pump-combining RF electronics in the large scattering chamber for a simpler wiring. In principle, these components participate the KIPA operation more directly and therefore may also be benefited, in terms of reducing their instabilities and thus the KIPA noise, by being inside the radiation-hermetic can. After placing the KIPAs and the above KIPA drivers inside the large scattering chamber, there is still space for two isolators equipped with heavy flux shields, so we prepare their fixtures as in Fig. 4.27, in case we find that we need to install isolators between the KIPAs and the KIDs for suppressing the resonance formation. These fixtures may otherwise be used for the isolators in the output circuit diagram (Fig. 4.25), but since the calibration system's RF switches require the wiring between the KIPAs and these isolators to run to 1 K and then back to MC, we do not find it apparently beneficial to install these isolators inside the low-radiation housing relative to the nominal locations outside the construction. For feeding the coaxial cable signals through the enclosures, we use threaded holes for the SMA feedthrough connectors to achieve a better radiation hermiticity compared to the typical through-hole connector mounting. We limit the connector locations to the electronics chamber side to prevent photons from entering the detector chamber via the connector mounting. For the DC wiring, we use a DB-15 male-to-male gender converter for the feedthrough, because without threading, we decide to embed the relatively thick gender converter into a matched indent in the bottom lid to provide a photon labyrinth.

At the moment, the remaining works for the construction of the ultra-low radiation housing include: 1) Determining the necessity of the pump-out channels and the corresponding machining, 2) modifying the indents around the SMA feedthrough connectors to ensure sufficient connector extrusions for making cable connections, 3) modifying the DB-15 gender converter's bolting scheme to improve its mounting mechanical stability. Since the DC bias' stability is the main cause to the amplifier noise, the improved mounting stability ensures that the KIPA noise is not dominated by mechanically unstable connections. Finally after the above machining works, 4) gold-plating the mechanical parts and 5) applying the IR-absorbing compound to the scattering chambers.

Chapter 5

CRYOGENIC THERMAL RADIATION MITIGATION

5.1 Motivation

In 2012, Domange et. al. (2012) identified ionizable D^{-} and/or A^{+} impurity sites in cryogenic Ge that may be as shallow as 0.75 ± 0.02 meV. Based on thermally stimulated leakage current data taken at different electric field biases, the authors found the current was emitted predominantly by trapping sites with single charges trapped by neutral atoms' screened $1/r^4$ Coulomb potential. Using the pulse shape with Ramo's theorem, the authors were able to deduce a current-emitting site density from the observation, which then indicated a trapping cross section. However, the study was unable to distinguish the trapping species, whether they were D^{-}, A^{+}, or other unidentified impurities. The deduced density was also orders-of-magnitude lower than the results by other works probing the trapping sites with different methods. One possible resolution the author proposed is the existence of trapping sites that were too deep to be probed by the thermal stimulation technique. These deeper sites could still be measured by other methods and thus offer an explanation to the difference.

We are motivated by the identification of the 0.75 meV trapping sites in Ge, as the depth corresponds to the peak energy of blackbody (BB) emission at 3.1 K. Even in a seemingly dark cryogenic environment, BB radiation at this temperature is present and therefore can cause a leakage current background via photoionization. We will show later this mechanism is also relevant for Si. The actual importance of such leakage current generation depends on various factors, such as the amount and the spectrum of the emitted BB radiation, the way the radiation propagates to the detectors, the actual impurities in the detectors, and finally the impurities' interactions with the radiation. All these factors contain certain details that are difficult to be modeled analytically. The leakage current mechanism also suggests the deeper sites Domange et. al. hypothesized should be less relevant to our study depending on the exact depths of the sites.

Optimistically, one can argue that, since in most cases the devices are fully enclosed by *mechanically* hermetic metal structures heat-sunk to 10^1–10^2 mK, photons that are energetic enough to liberate the trapped charges should be nonexistent in the device chambers. Nevertheless, as we have already learned in Sec. 4.7, without using special mechanical design and/or fabrication techniques, mechanical hermeticity typically implies 10^0–10^2-μm contact slits depending on the manufacturing quality, while geometrical/assembling misalignment can easily lead to 10^2–10^3-μm gaps. These gap sizes are not small enough compared to the trans-mm[1] BB photon wavelengths to make the enclosures truly hermetic in the cryogenic BB radiation environment. At the same time, copper is known to be an almost perfect reflector in the trans-mm range. If one makes a gross assumption that all the BB radiation emitted by 4-K surfaces ends up arriving at the crystals and photo-ionizing the impurity sites, a quick estimate yields:

$$
\begin{aligned}
E &= \epsilon\sigma A T^4 \qquad (5.1)\\
\sigma &= 5.67\times 10^{-8} \quad \mathrm{W/m^2\cdot K^4}\\
&= 3.54\times 10^{7} \quad \mathrm{eV/s\cdot cm^2\cdot K^4},
\end{aligned}
$$

where σ is the Stefan–Boltzmann constant, and E, ϵ, A, T are the emission power, emissivity, surface area, and temperature of the BB, respectively. Taking $T = 4$ K, $A \sim 10^3$ cm^2, and $\epsilon = 1.0$ for simplicity, we obtain $E = 1.45$ μW, corresponding to 1.93 mA if further divided by the 0.75-meV ionization bandgap. Such a mA-level leakage current contradicts typical observations by more than 15 orders-of-magnitude (Agnese et al., 2017; Amaral et al., 2020a). It shows that, if the hypothesized BB radiation leakage current generation is the dominant mechanism, the aforementioned factors from photon generation to photoionization all need to be modeled in-detail. Together, these factors should lead to a large but still accurately estimated reduction to the emitted radiation for the hypothesis to be comparable with observation.

However, while it is possible to calculate the BB surface emission and the photoinoization cross section analytically, it is impossible in practice to calculate the fraction of the radiation reaching the detectors. It is difficult due to the complex geometry of the experiment and, more importantly, the mechanical slits that are supposed to act like hermetic seals but do not in reality. Therefore, in order to make predictions that are comparable to measurements, as well as to provide a guiding tool for future background mitigation, we construct a fully first-principle Monte Carlo simulation framework. The simulation uses only modelable/measurable microphysics with accurate CAD-generated structures, so it can be flexibly adapted to different physical installations.

[1] Several hundred to a few thousand μm.

5.2 Physics elements

Every step in the journey of a BB photon, from emission to photoionization, is well understood in terms of physics principle but nevertheless introduces free parameters that depend on actual conditions. The process begins by having a BB emit a photon, in which the photon energy and direction are well established by quantum physics as functions of surface temperature, spatial orientation, and roughness/emissivity. We define a space the photon travels in as "open" if its boundaries are far enough relative to the photon's wavelength and "free" if it is free of scattering or reflection aside from at the boundaries. Once the BB photon is emitted into a free open space, it follows a straight-line path as a point-like particle and experiences classical boundary effects including reflection, absorption, and transmission.

The *open* space condition is violated when the size of the space approaches the photon wavelength. For 10s K BB photons, this size is 10^2–10^3 μm. In such spaces, the photon propagates as an EM wave guided by the boundaries rather than a point-like particle. An example would be where the mechanical structures are designed to be hermetic but do not use machining or polishing methods that permit (sub-)μm mating tolerances. There are two conditions that require modifying the *free* space condition. First, when a photon travels in a dielectric material with a nonnegligible opacity, we use a photon frequency-dependent mean free path to determine the absorption of the photon. We model all dielectric materials by this method, since the materials we are interested in are all homogeneous at the scale of the BB photon wavelengths. Such homogeneity allows photons to travel without scattering and thus makes the dielectric material-filled spaces free spaces in our model. The second modification to free-space propagation occurs when the photon arrives at the boundaries. We calculate the classical probabilities and trajectories of reflection, refraction, and absorption based on the incoming photon's direction, frequency, and the refractive indices of the media.

Finally, when the photon arrives in the detector crystals, it can interact with an impurity site, photoionize the bound charge, and lead to leakage current. The general photoionization framework for the interaction is well understood in QM scattering theory, which can be calculated given a *realistic model* for the depth and shape of the trapping potential. In the following sections, we will discuss in sequence the physics underlying each of of the steps listed above, which will then be combined into the simulation presented in Sec. 5.3.

5.2.1 Blackbody emission

Planck's law gives the BB spectral radiance

$$B(\nu, T) = \frac{2h\nu^3}{c^2}\frac{1}{e^{h\nu/k_BT} - 1}, \tag{5.2}$$

where B is the radiance in the (SI) unit of $(\mathrm{W/Hz})/(\mathrm{sr}\cdot\mathrm{m}^2)$, i.e., radiation power per unit photon frequency, radiating from unit area on the BB into unit solid angle, h and k_B are the Planck and Boltzmann constants, c is the speed of light, and ν is the frequency of the photons. To generate photons as quantized particles for later simulation, instead of using Eq. (5.2), it is more convenient to have a number density description in energy, which may be obtained by dividing frequency bins with photon energy $E_\gamma = h\nu$, giving

$$N(E_\gamma, T) = \frac{2{E_\gamma}^2}{h^2c^2}\frac{1}{e^{E_\gamma/k_BT} - 1}, \tag{5.3}$$

where N is now the photon flux per unit energy per $(\mathrm{sr}\cdot\mathrm{m}^2)$. We assume the emitting BB has a well-defined T but has an energy-dependent emissivity $\epsilon(E_\gamma)$. For such a BB, photons are emitted evenly toward every direction with no polarization preference. Therefore, the total photon emission rate from this BB becomes

$$N_{\mathrm{BB}}(T, A, \epsilon(E_\gamma)) = 2\pi A\int_0^\infty \epsilon(E_\gamma)\frac{2{E_\gamma}^2}{c^2}\frac{1}{e^{E_\gamma/k_BT} - 1}dE_\gamma, \tag{5.4}$$

where 2π arises from the hemisphere the BB radiates into. We will explain our practical implementation of Eq. (5.3) and Eq. (5.4) in Sec. 5.3, where the simulation incorporates multiple BBs of different geometries, temperatures, and emissivities.

5.2.2 Open space photon propagation

We define a space as an open space if its typical boundary separation is much larger than the wavelength in discussion. In an open space, photons travel unaffected by boundary effects, hence strictly following their momentum directions without diffraction.

We are most interested in the photoionization of sub-to-1-meV impurity states (Domange et al., 2012), which correspond to wavelengths less than 1.9 mm (Tab. 5.1). Based on this limit, and also that we regard a space that is 10 times wider than the wavelength in discussion as reliably open, we treat any space that is wider than 2 cm as fully open in the simulation. For higher-energy photons with shorter wavelengths, we further choose more stringent criteria shown in Tab. 5.1 to avoid utilizing computational resources on an unnecessary simulation for wave-like propagation.

E_γ [meV]	T_{peak} [K]	λ_{peak} [μm]	L [μm]
<0.6	<1.5	>1932	—
0.6–2.1	1.5–5	1932–580	20000
2.1–21	5–50	580–58	700
21–171	50–400	58–7	70
>171	>400	< 7	0

Table 5.1: Open space criterion as a function of photon energy. The temperatures shown in the table are the BB temperatures giving peak emission at the corresponding energies. We group the energies into 5 groups based on the typical temperatures present in cryostats. Currently, we do not generate photons with energies lower than 0.6 meV due to their inability to create photoionization according to (Domange et al., 2012). We choose the open space critera L to be roughly an order-of-magnitue larger than the largest wavelengthes for their corresponding energy ranges.

It is useful to relate photon energy to its corresponding BB temperature, so we can determine which mechanical structures' emission is more critical for the photoionization. Wien's Displacement Law relates the peak BB photon energy $E_{\gamma,\text{peak}}$ and thus the corresponding wavelength λ_{peak} to the BB temperature T by

$$\begin{aligned} E_{\gamma,\text{peak}} &\approx 0.43T \quad \left[\text{meV}\cdot\text{K}^0\right], \text{ or} \\ \lambda_{\text{peak}} &= 2898/T \quad \left[\mu\text{m}/\text{K}^0\right]. \end{aligned} \tag{5.5}$$

Considering the 1.9 mm wavelength limit, Eq. (5.5) suggests that structures at about 1.5 K or higher are responsible for the photoionization.

We model open-space photon propagation with a frequency-dependent mean free path when the space is filled with a dielectric material. This method is justified by the fact that the materials we plan to include all have refractive index variations at the length scales much longer than cryogenic BB photon wavelengths. These materials include polymer solids like polyethylene or Cirlex, Si and Ge crystals, and the high-vacuum cryostat space[2]. The simulation calculates with the mean free path a survival probability for the photon after each "step," a fixed distance of traveling much smaller than any physical length scales in the simulation. At the end of each step, the simulation then determines whether the photon should be terminated based on the probability.

The same argument of slow-varying refractive index also allows us to model the transmission and the reflection directions at material boundaries by the classical law

[2]For simplicity, we currently treat it as a perfect vacuum, with a frequency-independent refractive index of unity.

of reflection, i.e., the Snell's law

$$\theta_i = \theta_r, \tag{5.6}$$

$$n_1 \sin\theta_i = n_2 \sin\theta_t, \tag{5.7}$$

where θ_i, θ_r, θ_t are the angles of incident, reflection, and transmission, respectively, and n_1, n_2 are the indices of refraction on the incident and the transmission sides, respectively. The uncertainty of the outgoing photon direction due to surface roughness is added phenomenologically to the nominal direction given by the Snell's law through

$$\hat{k} \to \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sin\theta & 0 & \cos\theta \\ 0 & 1 & 0 \\ -\cos\theta & 0 & \sin\theta \end{pmatrix} \hat{k}, \tag{5.8}$$

where $\hat{k}$ is the original direction before the roughness adjustment, ϕ is a random angle evenly distributed in 0 to 2π, and θ is a Gaussian-distributed angle of smearing following the probability density function

$$\mathcal{P}(\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\theta^2}{2\sigma^2}}. \tag{5.9}$$

We use $\sigma = 3°$ for polished metallic surfaces and $\sigma = 10°$ for non-metallic or metallic but unpolished post-machining surfaces (Bergstrom, 2008). Empirically, we find varying σ in $1°$ to $20°$, corresponding to a finely polished mirror to visibly dull surfaces in practice, has no impact to the simulated detected spectra.

5.2.3 Bounded space photon propagation

We define a bounded space as non-open following the same criteria in Tab. 5.1. Although the radiation/photon field is always governed by Maxwell's Equations, in practice one can only analytically calculate the field distribution when the field boundary is sufficiently negligible, i.e., "open" in our setup, or exhibits a high degree of symmetry for simplification. For a realistic free-formed boundary with well-established bulk governing equations, one typically applies the finite element analysis (FEA) technique that divides the bounded space into small pieces, exchanging one general but uncalculable equation with a system of calculable, piecewise-linear solutions that share boundary conditions. We will discuss in Sec. 5.3 our implementation for FEA.

Nevertheless, after inspecting the experimental constructions we are interested in, we realize there is one important type of bounded spaces that can be solved analytically, which is the mechanical contact surface that can be approximated by a semi-open parallel-plate metallic channel. These imperfect contact seals enable BB photons to penetrate the supposedly isolated spaces to generate the leakage current. Even though these seals' actual profiles may vary, we can obtain physics insight into mitigation strategies by first understanding the parallel-plate construction as an idealized model. Among other useful details, we will show in the end that the transverse-electric field (TE) mode transmission exhibits a cutout frequency while the transverse-magnetic field (TM) mode does not. This is critical because it suggests one would always obtain certain level of radiation penetration through a contact surface seal regardless of the seal's tightness.

5.2.3.1 TE mode in parallel-plane waveguide

Fig. 5.1 presents the side view of a channel formed by two parallel planes (hashed region) at a separation d. The boundary planes extend much longer in $\hat{y}$ than the trans-mm wavelength of interest and hence can be regarded as infinitely extended. We assume d is of the same order-of-magnitude of the wavelength or smaller, so a significant wave-like transmission that we intend to model occurs in the channel. Typically, mechanical structures in cryogenic experiments forming these tightly sealed parallel-plane channels are at the sizes of cm or larger and are generally made of high-RRR (residual resistance ratio) copper. At the trans-mm wavelengths, the conductivity of high-RRR Cu yields sub-μm skin depths, effectively making the mechanical structures impenetrable by the radiation. We therefore approximate the parallel planes' $\hat{z}$ extensions to be infinite. Without losing the general electromagnetic characteristics of the model, the above conditions allow us to simply the model by assuming the boundary planes are made of perfect conductors and the channel is under vacuum. The perfect conductor assumption prevents the transmitted radiation to exhibit non-zero components parallel to the boundaries, and the two assumptions together remove the mild decaying of the transmitted field along $\hat{x}$. We will present in Sec. 5.3 an FEA study including these omitted effects.

We fix the axes and the origin as shown in Fig. 5.1: The channel is bounded in $\hat{z}$ at a separation of d, extends in $\hat{y}$, and propagates the incoming radiation in $\hat{x}$. This configuration becomes a standard waveguide problem that is solved in EM textbooks, e.g., (Jackson, 1998). One can separate the field solutions into the **transverse-electric field (TE) modes** and the **transverse-magnetic field (TM) modes**. The former exhibits electric fields in the transverse direction of the transmission ($\hat{x}$),

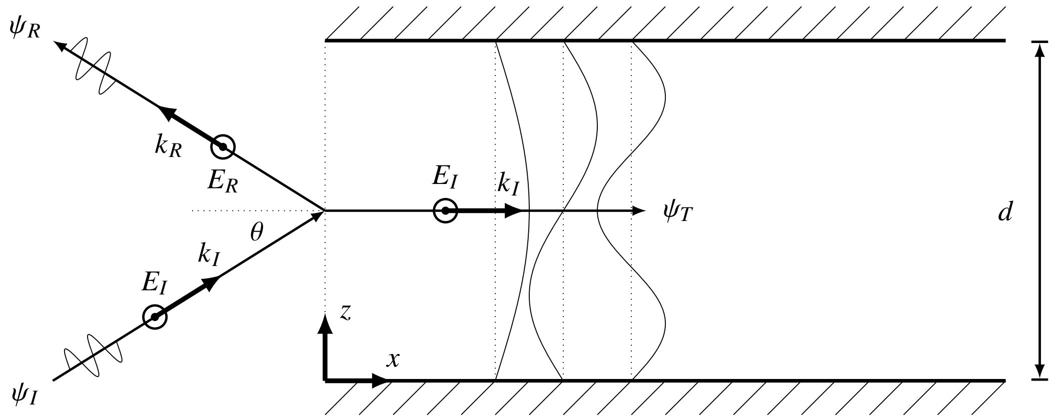


Figure 5.1: The parallel-plate waveguide model for the TE-mode propagation. The adopted coordinate system is shown by the arrows in the bottom-left corner of the waveguide, and the hashed regions represent conductive materials (semi-)infinitely extending in $\hat{y}$ (xz plane). Detailed description is found in the corresponding text.

while the latter exhibits the magnetic fields in the transmission direction. The TE-mode solution allowed by the boundary contrition is

$$\vec{E} = E_0 \sin(k_z z) e^{i k_x x} \hat{y}, \tag{5.10}$$

where $\vec{E}$ is the resulting electric field, E_0 is the field amplitude determined by the incoming wave, and k_x and k_z are the propagation constants in $\hat{x}$ and $\hat{z}$, respectively. Furthermore, the boundary condition requires

$$k_z = \frac{n\pi}{d} \quad , \ n = 1, 2, 3, ..., \tag{5.11}$$

for the transmission to be sustainable.

At a given angular frequency ω, since $|\vec{k}| = \omega/c$ is a constant, Eq. (5.11) leads to

$$k_x = \sqrt{|\vec{k}|^2 - k_z^2} = \sqrt{\left(\frac{\omega}{c}\right)^2 - \left(\frac{n\pi}{d}\right)^2} \quad , \ n = 1, 2, 3, ... \tag{5.12}$$

Therefore, to have a sustainable transmission corresponding to a real-number k_x, ω has the lower bounds of

$$\omega_c = c\left(\frac{n\pi}{d}\right) \quad , \ n = 1, 2, 3, ..., \tag{5.13}$$

for different modes (n). This is the cutoff frequency for a guided TE-mode transmission. Below the cutoff frequency, the transmission decays exponentially at the decay constant of $\mathrm{Re}(k_x)$. For $d \sim \mu$m, Eq. (5.13) yields

$$\frac{\omega_c}{2\pi} \sim n \times 150 \quad \text{GHz}, \tag{5.14}$$

which corresponds to the peak emission of BBs at $\approx 1.5n$ K. The result shows, in reality, the mechanical contact seals can permit BB radiation in the relevant cryogenic temperature range.

5.2.3.2 TM mode in parallel-plane waveguide

The field solution for the TM mode is

$$\vec{B} = B_0 \cos(k_z z) e^{ik_x x} \hat{y}, \tag{5.15}$$

where $\vec{B}$ and B_0 are the magnetic field and its amplitude, respectively (Jackson, 1998). The result is different from the TE-mode solution in that the magnetic field peaks at the boundaries as opposed to the vanishing electric field in Eq. (5.10). It is because the out-of-plane current must vanish at the boundaries, for which one can calculate using Ampère's law and realize the curl of the magnetic field requires Eq. (5.15) to have only the $\cos(k_z z)\hat{y}$ term. One can then determine the corresponding electric field in the waveguide with $\nabla \times \vec{B} = \partial_t \vec{E}/c^2$, which yields

$$\vec{E} = \frac{B_0 c^2}{\omega} \left[k_x \cos(k_z z)\vec{z} - ik_z \sin(k_z z)\vec{x}\right] e^{ik_x x}. \tag{5.16}$$

Similar to the TE mode, the boundary condition allows only discrete standing-wave modes in $\hat{z}$ with

$$k_z = \frac{m\pi}{d} \quad , \; n = 0, 1, 2, ... \tag{5.17}$$

This k_z then yields the cutoff frequency

$$\omega_c = c\left(\frac{m\pi}{d}\right) \quad , \; m = 0, 1, 2, ..., \tag{5.18}$$

following the same principle of Eq. (5.12).

Notice a critical distinction between the TM mode and the TE mode is that the TM-mode index m includes 0. The reason is that, when $k_z = 0$ at $m = 0$, the cosine term in the TM field solution does not vanish to yield a trivial solution. On the contrary, the TE-mode solution in Eq. (5.10) does become trivial at $k_z = 0$, so $n = 0$ is not allowed. For this special case of $m = 0$, the magnetic field is in $\hat{y}$ and the electric field is in $\hat{z}$; both are transverse to the transmission direction ($\hat{x}$) while neither oscillates in the transverse plane. This field configuration is a **TEM** (transverse E-and-M) transmission that resembles a plane wave propagating along the waveguide, which is normally prohibited in hollow waveguides but allowed in our case by the TM-TEM degeneracy. Eq. (5.18) shows, for $m = 0$, there is no cutoff frequency to prevent radiation from penetrating the mechanical seals in a TEM mode.[3]

We compare the TM field solution (Eq. (5.16)) to that for a waveguide fully enclosed in the transverse plane and then realize the $m = 0$ mode is only allowed because of the infinite $\hat{y}$ extension of the parallel planes (Jackson, 1998). Although such an idealization is not the case in practice, we argue the model is still applicable, so the lack of a cutoff frequency is a real concern. Since we are mainly interested in the 0.75 meV impurity energy level corresponding to trans-mm or shorter wavelengths, while the cutoff frequency is set by the size of the waveguide that is generally $\mathcal{O}$(cm), we can conclude the mechanical seals are effectively infinite and do not exhibit cutoff frequencies for the BB radiation range of interest.

5.2.4 Impurity site photoionization

Although (Domange et al., 2012)'s finding for the shallow impurity states motivates this study, the information obtained in the reference is not sufficient for our purpose. Specifically, we want to be able to generate leakage current in the simulation event by event. For this purpose, we need a microscopic photoionization model, where the BB radiation and the impurity state interact as individual particles. The needed information lacking in (Domange et al., 2012) is therefore a calculable photoionization cross section that depends on the properties of the actual participating photon and impurity particle.

Nevertheless, there is still useful information in (Domange et al., 2012) that we utilize as the starting point for developing our own model. In particular, by adjusting the bias field of the sample, Domange et. al. observed a charge emission energy barrier shift that was consistent with (Lax, 1960)'s model prediction. Lax's model describes a neutral dopant (D^0 or A^0) capturing an external charge (D^- or A^+) due to polarizing the dopant when the charge is located in vicinity; we will refer to this external charge as *overcharge* in the following. Lax concluded, under such polarization binding, the trapped charge is bound by a potential well with a r^{-4} shape, r being the distance of the dopant and the charge. Domange et. al. then derived an expected emission energy change from the r^{-4} shape and showed it was consistent with their observation. However, while Domange et. al. interpreted the result as a proof of Lax's model, the 0.75 meV binding energy measured independently was in significant contradiction with the energy scale Lax originally postulated.

[3] Note that it is also due to the randomness of the BB radiation, which inevitably has components in the TEM-mode orientation to excite the transmission.

Below we discuss the development of our photoionization cross section model based on (Domange et al., 2012) and (Lax, 1960). In the end, the result depends on fundamental properties of the photon, the impurity particle, and the substrate material. Therefore, in addition to Ge that was studied in (Domange et al., 2012), we can adopt the result for Si or other detector materials of interest in the future.

5.2.4.1 Generic photoionization

Before we include the overcharge, the neutral dopant in Lax's model is identical to a standard QM hydrogen-like atom, which is described by the generic non-relativistic Hamiltonian

$$\hat{\mathrm{H}} = \frac{(\hat{p} + e\hat{A}/c)^2}{2m_e} - e\hat{\phi}(r), \tag{5.19}$$

where $\vec{p}$, $\vec{r}$, m_e, and e are the electron's momentum, position, mass, and charge, respectively, $r = |\vec{r}|$ by setting the nucleus to be the origin, A and ϕ are the vector and the scalar potentials, and we use hatted symbols to distinguish QM operators. We assume ϕ is spherically symmetric. One standard technique for solving Eq. (5.19) is treating the external photon as a plane wave for $\hat{A}$. Since the photon is originated from much further than the size of the system, one can then expend the Hamiltonian into a perturbation series and keep the leading term under the condition

$$\left|\frac{eA_0}{cp}\right| \ll 1, \tag{5.20}$$

where A_0 and p are the magnitudes of $\langle\hat{A}\rangle$ and $\langle\hat{p}\rangle$, respectively. Since we are modeling photo-absorption, we can omit the outgoing *emission* wave and rewrite Eq. (5.19) as

$$\hat{\mathrm{H}} = \left[\frac{\hat{p}^2}{2m_e} - e\hat{\phi}(r)\right]_{\mathrm{o}} + \left[\frac{eA_0 e^{i(\vec{k}\cdot\vec{r}-\omega t)}\hat{p}}{2m_e c}\right]_{\mathrm{i}}, \tag{5.21}$$

where the subscripts o and i denote the original and the perturbative interaction Hamiltonians, respectively. Detailed discussions for the above formalism can be found in (Shankar, n.d.; Sakurai, n.d.).

We know the BB photon is trans-mm (λ) while the impurity system is of atomic scale (r), therefore

$$|\vec{k}\cdot\vec{r}| \sim \left|\frac{2\pi r}{\lambda}\right| \sim \frac{\mathrm{nm}}{\mathrm{mm}} \ll 1 \Rightarrow e^{i\vec{k}\cdot\vec{r}} \approx 1. \tag{5.22}$$

The condition allows us to adopt the dipole approximation to further reduce the interaction to

$$\hat{\mathrm{H}}_{\mathrm{i}} \approx \frac{eA_0\hat{p}}{2m_e c}e^{-i\omega t} = -\vec{\mu}_e \cdot \vec{E}, \tag{5.23}$$

where μ_e is the electronic system's dipole moment, and $\vec{E}$ is the electric field in the region due to $\hat{A}$. In this regime, the interaction Hamiltonian is just the dipole energy $\mu_e E$ (Shankar, n.d.; Sakurai, n.d.).

With the dipole transition Hamiltonian above, we can calculate the photo-absorption perturbatively by the Dyson series formalism. We define the transition amplitude from the initial state $|i\rangle$ to the final state $|f\rangle$ as

$$d_{fi} = \frac{-i}{\hbar}\left\langle f \left|\hat{\mathrm{H}}_{\mathrm{i}}\right| i\right\rangle, \tag{5.24}$$

which gives the Fermi's Golden rule transition rate (Shankar, n.d.)

$$\mathrm{R}_{\mathrm{fi}} = 2\pi\hbar|d_{fi}|^2\delta(\Delta E - \omega\hbar). \tag{5.25}$$

We approximate the initial and the final states by

$$|i\rangle \approx \frac{e^{-r/r_0}}{(\pi r_0^3)^{1/2}} \tag{5.26}$$

and

$$|f\rangle \approx \frac{e^{-i\vec{k}_f\cdot\vec{r}}}{(2\pi\hbar)^{3/2}}. \tag{5.27}$$

Eq. (5.26) assumes the overcharge electron is initially on a S-shell of a radius of r_0, which we justify by the low temperature of the detector substrate relative to its bandgap. Eq. (5.27) assumes the electron scattered into the conduction band can be approximated by a free particle, therefore the plane wave with a momentum of $\hbar\vec{k}_f$. Substituting these initial and final states into Eq. (5.24) and Eq. (5.25), we obtain

$$\left|\left\langle f \left|\hat{\mathrm{H}}_{\mathrm{i}}\right| i\right\rangle\right| = \ldots \text{ (algebra in Appx. A)} \tag{5.28}$$

$$= \frac{eA_0\cdot\hbar\vec{k}_f}{2m_e c}\cdot\frac{1}{(\pi r_0^3)^{1/2}}\cdot\frac{2\pi}{(2\pi\hbar)^{3/2}}\cdot\frac{4r_0^3}{(k_f^2 r_0^2+1)^2} \tag{5.29}$$

and then

$$\mathrm{R}_{\mathrm{fi}} = \frac{2\pi}{\hbar}\left(\frac{e\hbar}{2m_e c}\right)^2\frac{1}{\pi r_0^3(2\pi\hbar)^3}\cdot\frac{|\vec{A}_0\cdot\vec{k}_f|^2 64\pi^2 r_0^6}{(k_f^2 r_0^2+1)^4}\delta(\Delta E - \hbar\omega). \tag{5.30}$$

Eq. (5.30) is the differential rate for a photon with energy $\hbar\omega$ to scatter an electron with an initial energy E_i into a final state of energy E_f, $\Delta E = E_f - E_i$, toward the momentum direction of $\vec{k}_f$.

Since our detector can not resolve $\vec{k}_f$, we integrate the scattering rate for all final momentum directions by

$$\begin{aligned} \mathrm{R}_{k_f} &= \int \mathrm{R}_{\mathrm{fi}}\,(\hbar k_f)^2 d\Omega \qquad (5.31) \\ &= ...(\text{constants})... \int (A_0 k_f \cos\theta)^2\,(\hbar k_f)^2 d\cos\theta d\phi \\ &= \frac{2\pi}{\hbar}\left(\frac{e\hbar}{2m_e c}\right)^2 \frac{1}{\pi r_0^3 (2\pi\hbar)^3} \cdot \frac{A_0^2 k_f^4 64\pi^2 r_0^6 \hbar^2}{(k_f^2 r_0^2 + 1)^4} \cdot \frac{4\pi}{3}\delta(\frac{\hbar^2 k_f^2}{2m_e} - E_i - \hbar\omega). \qquad (5.32) \end{aligned}$$

We rewrite the formula with a general binding energy

$$U = -E_i, \qquad (5.33)$$

which will be replaced by a calculable model next, and then integrate over final state energy to be the total absorption rate

$$\begin{aligned} \mathrm{R}(\omega) &= \int R_{k_f}\, dk_f \qquad (5.34) \\ &= \frac{2\pi}{\hbar}\left(\frac{e\hbar}{2m_e c}\right)^2 \frac{1}{\pi r_0^3 (2\pi\hbar)^3} \cdot \frac{A_0^2 64\pi^2 r_0^6 \sqrt{2m_e(\hbar\omega - U)/\hbar^2}^3}{(\sqrt{2m_e(\hbar\omega - U)/\hbar^2}^2 r_0^2 + 1)^4} \cdot m_e\hbar \cdot \frac{4\pi}{3}. \end{aligned}$$

$$(5.35)$$

For calculating the above integration, one can use

$$\delta(\frac{\hbar^2 k_f^2}{2m_e} - E_i - \hbar\omega) = \frac{m_e}{\hbar k_f}\delta(k_f - \sqrt{2m_e(\hbar\omega - U)/\hbar^2}) \qquad (5.36)$$

to convert between the energy and the momentum conservation relations given by Fermi's Golden rule.

Finally, we derive the sought photon energy-dependent photoionization cross section $\sigma(\omega)$ by equating the photon and the EM wave energy absorption rates,

$$\hbar\omega \cdot \mathrm{R}(\omega) = \langle S \rangle\, \sigma(\omega), \qquad (5.37)$$

where

$$\langle S \rangle = \frac{\omega^2 A_0^2}{2 Z_0 c^2} \qquad (5.38)$$

is the time-averaged classical Poynting vector for the plane wave $\hat{A}$, and Z_0 is the vacuum impedance. Eq. (5.37) yields

$$\begin{aligned}\sigma(E_\gamma) &= \ldots \text{ (algebra in Appx. A)} \\ &= \frac{32 Z_0 e^2 r_0^3 \hbar^6}{3 m_e E_\gamma} \frac{[2m_e(\hbar\omega - U)]^{3/2}}{(2m_e(\hbar\omega - U) r_0^2 + \hbar^2)^4}. \end{aligned} \tag{5.39}$$

Notice that, while the generic form of Eq. (5.39) may suggest the binding energy and the initial-state radius are independent, they are not in Lax's model due to the r^{-4} potential. Derivation in this subsection is based on (Shankar, n.d.; Sakurai, n.d.) with modifications made for our specific case. The reader may find more detailed discussions therein for the standard QM scattering theory and the algebra omitted in our derivation.

5.2.4.2 Binding potential

We assume a general spherical binding potential in deriving Eq. (5.39). Next, we build a model specifically for the impurity potential (Domange et al., 2012; Lax, 1960) identified. In particular, we seek for calculable descriptions for the initial-state orbital radius r_0 and the binding energy U to be used in Eq. (5.39).

While the change of the charge emission barrier Domange et. al. observed is consistent with that expected by a r^{-4} potential, Domange et. al. could not conclude whether the observation was indeed due to the binding system Lax proposed. One reason is that Domange et. al. measured the charge emission energy by thermal stimulation, which is only sensitive to a subgroup of the impurity class Lax's model describes. Furthermore, Lax's original derivation used certain atomic and material parameters that lacked reliable experimental data. Lax was only able to speculate a *natural* energy scale anticipated by the general picture of the model, which turned out to be orders-of-magnitude higher than Domange et. al.'s 0.75 meV binding energy observation.

According to (Lax, 1960), an initially neutral impurity site can be slightly polarized to the level of $q_{\text{ext}}/\epsilon_r$ upon the introduction of an external charge q_{ext}, where ϵ_r is the material's dielectric constant. The polarized impurity site then exhibits an attractive dipole potential that binds the external charge to form an over-charged bound system. The polarizing field by the external charge is given by

$$E_{\text{pol}} = \frac{1}{4\pi\epsilon} \frac{q_{\text{ext}}}{r^2}, \tag{5.40}$$

where r is the distance between the atom and the external charge. Assuming E_{pol} is small compared to the electron cloud's binding field, the neutral atom acquires a polarized dipole

$$\mu_{e,\text{pol}} = \alpha E_{\text{pol}} = \frac{\alpha}{4\pi\epsilon}\frac{q_{\text{ext}}}{r^2}, \tag{5.41}$$

where α is the phenomenological classical atomic polarizability (Jackson, 1998). Note that this dipole always aligns its axis to the external charge. The induced dipole in turn creates an *attractive* dipole potential

$$\phi(r) = -\frac{\mu_{e,\text{pol}}}{4\pi\epsilon r^2} = -\frac{\alpha q_{\text{ext}}}{(4\pi\epsilon)^2 r^4} \tag{5.42}$$

for the external charge. Based on (Lax, 1960)'s binding mechanism picture, we have independently derived in the above a binding field that exhibits a spherical r^{-4} shape Lax (1960) and Domange et. al. (2012) proposed/observed.

We can obtain well-defined U and r_0 with the above result. According to the classical virial theorem, a spherical r^{-n} potential has

$$E_k = \frac{-n}{2}V, \tag{5.43}$$

where E_k and V are the orbiting object's kinetic and potential energies, respectively. We adopt Bohr's angular momentum quantization and assume the bound charge is always in its ground state due to the low temperature, i.e.,

$$E_k = \frac{(\hbar/r)^2}{2m_e}. \tag{5.44}$$

The virial theorem then yields the characteristic radius, r_0, and the corresponding potential energy, V_0:

$$\frac{(\hbar/r)^2}{2m_e} = \frac{2\alpha q_{\text{ext}}^2}{(4\pi\epsilon)^2 r^4}\Bigg|_{r=r_0} \Rightarrow \begin{cases} r_0 = \dfrac{2q_{\text{ext}}}{4\pi\epsilon\hbar}\sqrt{\alpha m_e} \\ V_0 = q_{\text{ext}}\phi(r_0) = -\dfrac{\hbar^4(4\pi\epsilon)^2}{16 q_{\text{ext}}^2 \alpha m_e^2} \end{cases}. \tag{5.45}$$

Eq. (5.45) yields the bound state energy

$$E_0 = E_k(r_0) + V_0 = \frac{\hbar^4(4\pi\epsilon)^2}{16 q_{\text{ext}}^2 \alpha m_e^2}. \tag{5.46}$$

A worrisome feature of our model is that the total energy for the over-charged system at equilibrium is positive, which indicates the binding system is actually unstable. Therefore, instead of requiring an incoming BB photon to contribute an energy comparable to the size of the potential well,[4] a much smaller energy can destabilize the system and subsequently lead to photoionization. In fact, it has been rigorously proven in classical mechanics that, for any generic r^{-n} type attractive potentials, $n < 2$ is required for a system to be intrinsically stable; one can show this by requiring $E_k + V < 0$ in Eq. (5.43). Therefore, we recognize the discovered instability as an inherent property for the r^{-4} potential but not an error.

Next, we discuss the appropriate polarizability to be used in the above results. The neutral atom is polarized when its electron cloud is displaced from the original location. Following the small external field assumption, to leading order, a bound charge in the cloud is under a force of $m_e \omega_0^2 d_{\text{pol}}$, where m_e is the charge's mass, ω_0 is its natural damping frequency near the minimum of the binding potential, and d_{pol} is the acquired displacement (Jackson, 1998). Force balance requires

$$m_e \omega_0^2 d_{\text{pol}} = e E_{\text{pol}}, \tag{5.47}$$

with which we can rewrite Eq. (5.41) as

$$\mu_{e,\text{pol}} = e d_{\text{pol}} = \frac{e^2 E_{\text{pol}}}{m_e \omega_0^2}. \tag{5.48}$$

We compare Eq. (5.41) and Eq. (5.48) and find the polarizability *for a single bound charge* is

$$\alpha_e = \frac{e^2}{m_e \omega_0^2}. \tag{5.49}$$

Traditionally, to connect α_e to $\alpha_{(\text{atom})}$, one continues to sum up all different charge species in the atom with their corresponding damping frequencies and number densities (Jackson, 1998). However, at temperatures much below the binding energy divided by k_{B}, while the polarizing field is weak, Lax argued that there should be effectively only one electron, an outermost-shell electron, to be ionized. This assumption gives

$$\hbar \omega_0 \approx I, \tag{5.50}$$

where I is the ionization energy, and ω_0 is for the outermost electron. Lax therefore suggested the approximation of

$$\alpha \approx \alpha_e \approx \frac{e^2 \hbar^2}{m_e I^2}. \tag{5.51}$$

[4] There is no longer a "well" as postulated by (Lax, 1960).

Lax pointed out that, in order to reflect realistic dynamics, one should use the effective mass of the charge carrier in Eq. (5.51), not the free electron mass, and the approximations may only be accurate to order-of-magnitude.

We compare Lax's model for the polarizability to the classical atomic/molecular polarizability, which is defined by

$$\mathrm{P} = N\mu_{e,\mathrm{pol}} = \epsilon_0 \chi_e E_{\mathrm{mac}}, \tag{5.52}$$

where P is the volumetric polarization, N is the number density of atoms, ϵ_0 is the vacuum permittivity, χ_e is the material's macroscopic electric susceptibility, which is more commonly presented in the literature by

$$\epsilon_r = \chi_e + 1, \tag{5.53}$$

and E_{mac} is the macroscopic electric field. The key difference between the classical polarizability and Lax's approach is that the former is defined by the global field, which includes not only the polarizing field in discussion but also the field from the surrounding atoms that are also polarized. A widely adopted version of such classical polarizability is the Clausius-Mossotti model (Aspnes, 1982), which has

$$\alpha = \frac{\epsilon_0}{N} \cdot \frac{(\epsilon_r - 1)}{(\epsilon_r + 2)/3}. \tag{5.54}$$

Since neither Lax's approach nor the Clausius-Mossotti model considers a QM wavefunction, while ionization is inherently a QM phenomenon, we can not determine which is more realistic or accurate from the derivations. Therefore, we will compare the two approaches when we compute measurable quantities with different α.

We use the parameters listed in Tab. 5.2 to calculate the numerical predictions of our model and summarize the results in Tab. 5.3. We compare the predictions given by both the polarizabilities based on the classical permittivities ϵ_r and the single-atom ionization energies I, i.e., Eq. (5.54) and Eq. (5.51), respectively. We find the classical macroscopic approach yields much smaller polarizabilities than (Lax, 1960)'s microscopic model, leading to unrealistic radii that are smaller than the impurity site's native electron cloud. The corresponding binding potentials are also much higher than the atoms' nominal ionization energies.

These apparent contradictions may be explained by the orientation polarizability ignored in the Clausius-Mossotti theory. According to the Langevin-Debye model

	Si	Ge
Dielectric constant	11.68	16.2
Atomic density (10^{22}/cm^2)	4.99	4.54
Ionization energy (eV)	1.1	0.7
D^- density (10^{10}/cm^3)	415	8.3
A^+ density (10^{10}/cm^3)	70	1.4
DOS electron eff. mass/m_0	1.08	0.56
DOS hole eff. mass/m_0	0.57	0.29

Table 5.2: The Si and Ge material parameters for the corresponding calculations in Sec. 5.2, acquired from (Van Zeghbroeck, 1997; Phipps, 2016). m_0 is the *free* electron mass not to be confused with m_e in the formulae. The adoption of the listed values are explained in the corresponding text.

(Debye, 1929), if the induced dipoles are allowed to thermodynamically redistribute with respect to the external field, the effective dipole density in Eq. (5.54) should scale with the temperature by $N \propto T$. This relation manifests the expectation that fewer dipoles should contribute to the observed global field at lower temperatures. It also shows that the Clausius-Mossotti polarizability can become unphysically large at low temperatures due to its inverse proportionality in N, a condition commonly called the *Mossotti catastrophe*. In reality, material imperfections and QM effects limit the dipole alignment at low temperatures and prevent the Mossotti catastrophe, leading to different models that predict much increased but finite polarizabilities similar to (Lax, 1960), e.g., (Onsager, 1936; Kirkwood, 1939). We therefore determine that Lax's model is more appropriate for our application and proceed with Eq. (5.51) for α.

Fig. 5.2 presents the total energy for the bound external charge in Si and Ge according Eq. (5.46). The potentials fall off toward $r = 0$, suggesting the bound charge could fall in toward the impurity. However, such behavior would lead to a strongly polarized system that contradicts our assumption. In reality, this contradiction is prevented by the fact that the bound charge would experience a strong Coulomb repulsion approaching the inner shell electrons ($r \approx 1.1$–1.2 Å) indicated by the hashed region in Fig. 5.2. Shown in the bottom panel of Fig. 5.2, we phenomenologically include a repulsive r^{-8} potential based on the Lennard-Jones model to represent the atomic cloud (J. Jones, 1924a; J. Jones, 1924b; Lennard-Jones, 1931). With such a modification, the potential becomes positive close to the atom and forms a shallow dip near the atom's native electron cloud. We identify this shallow dip as the meta-stable state that binds the external charge.

(Real I / Classical ϵ_r)	Si	Ge
Electron		
α/ϵ_0 (Å^3)	1047.57 / 46.91	4988.96 / 57.20
r_0 (Å)	2.24 / 0.47	2.54 / 0.27
V_0 (eV)	-0.35 / -7.80	-0.52 / -45.74
E_0 (eV)	0.35 / 7.80	0.52 / 45.74
Hole		
α/ϵ_0 (Å^3)	1984.88 / 46.91	9633.85 / 57.20
r_0 (Å)	2.24 / 0.34	2.54 / 0.20
V_0 (eV)	-0.66 / -27.99	-1.01 / -170.59
E_0 (eV)	0.66 / 27.99	1.01 / 170.59

Table 5.3: The trapped charge's orbital radii, potential energies, and total energies for Si and Ge, calculated from the material parameters listed in Tab. 5.2. The values for "Emperical I" and "Classical ϵ_r" correspond to the results calculated with the polarizabilities obtained from the ionization energy data and the classical permittivities, respectively.

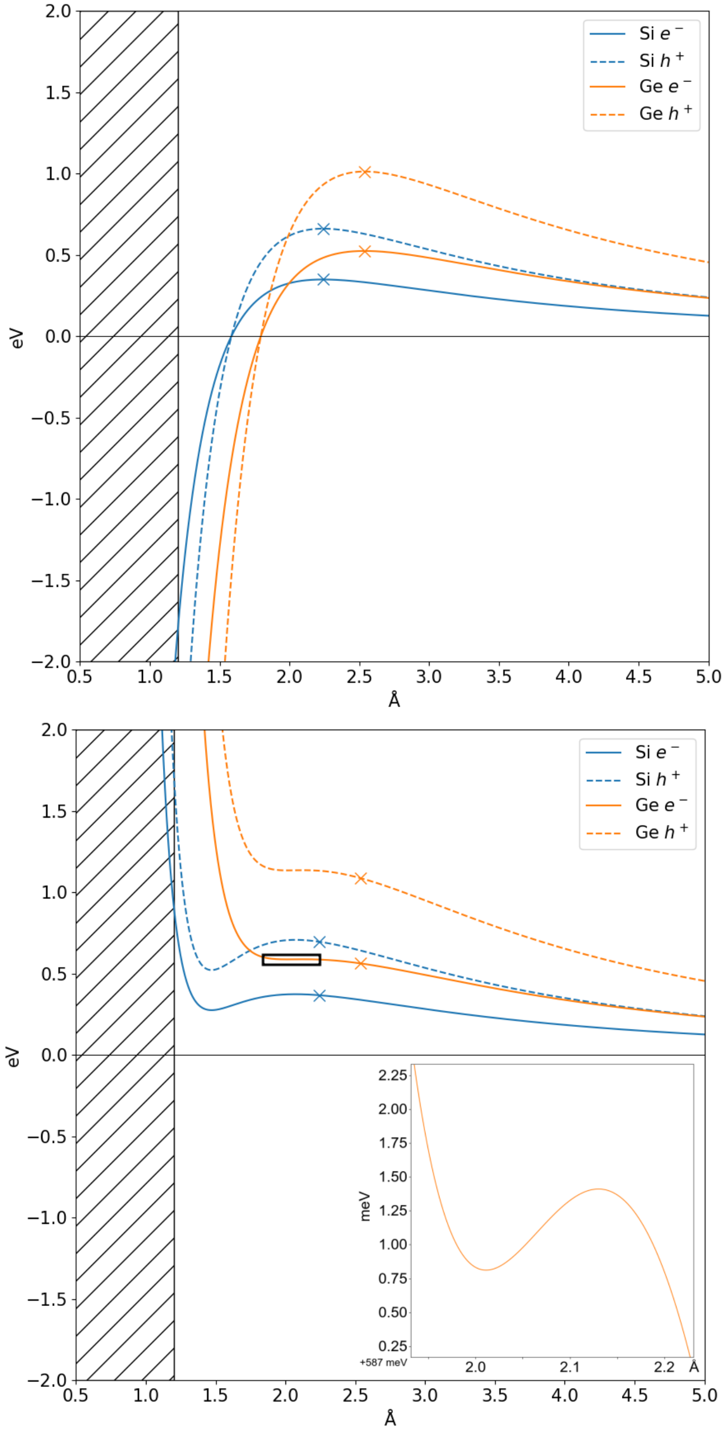


Figure 5.2: The total energy of the trapped external charge versus radius for an electron (solid) or hole (dashed) in Si (blue) and Ge (orange). The top and bottom panels show, respectively, the results for our original model and further including the phenominological Lennard-Jones atomic electron cloud. The shaded regions are roughly the forbidden space occupied by the atom's native electron cloud. The inset in the bottom panel corresponds to the box-highlighted Ge-electron meta-stable state, which we use to constrain the Lennard-Jones potential coefficient. To keep the figure legible, we only show the results obtained using the polarizabilies calculated from the empirical ionization energies. The interested readers should be able to fully reconstruct the omitted curves based on the corresponding discussions. Detailed explanation may be found in the in corresponding text.

We include a repulsive "core" component that exhibits twice the power in r than the dipole attraction into Eq. (5.45) based on the Lennard-Jones model,

$$E = E_k|_{r_0} - k\alpha \left[\frac{e}{r^4} - \beta \frac{Ze}{r^8} \right], \tag{5.55}$$

where $k = e/(4\pi\epsilon)^2$ and the polarizability α were introduced previously, and β is utilized phenomenologically to represent the relative strength of the screened-Coulomb core to the dipole attraction. To determine β, we require the meta-stable state for Ge to yield the binding energy of 750 μeV (Domange et al., 2012) observed. The inset in Fig. 5.2 shows the fitting for β. Although (Domange et al., 2012) did not distinguish electron and hole, (Phipps, 2016) points out that the charge carriers in Ge are predominantly electrons, so we choose to adjust the electron curve in Fig. 5.2 and obtain

$$\beta \approx 0.095\ [\text{Å}^{-4}], \tag{5.56}$$

which is much smaller than unity as expected for a Coulomb-screened neutral atom. Without further data constraints, we assume this β applies to hole in Ge and both charge carriers in Si to calculate their results in Fig. 5.2.

The inclusion of repulsive cores yields meta-stable state depths as the binding energies for Eg. (5.39), so with the r_0 determined previously, we can calculate the photoionization cross sections shown in Fig. 5.3. We further convert the cross sections into absorption mean free paths by

$$L = \frac{1}{N\sigma}. \tag{5.57}$$

To be comparable with applications insensitive to charge carrier species, the cross sections need to be weighted by their concentrations. For Ge, (Phipps, 2016) measured the concentrations that we quoted in Tab. 5.2. For Si, (Phipps, 2016) only states that the concentrations are "believed to be 10–100 times higher" than Ge. So, we assume both the electron and the hole exhibit concentrations that are 50 times higher than their Ge counterparts. Finally, we substitute these U and r_0 into Eq. (5.39) to obtain the photoionization cross sections shown in Fig. 5.3.

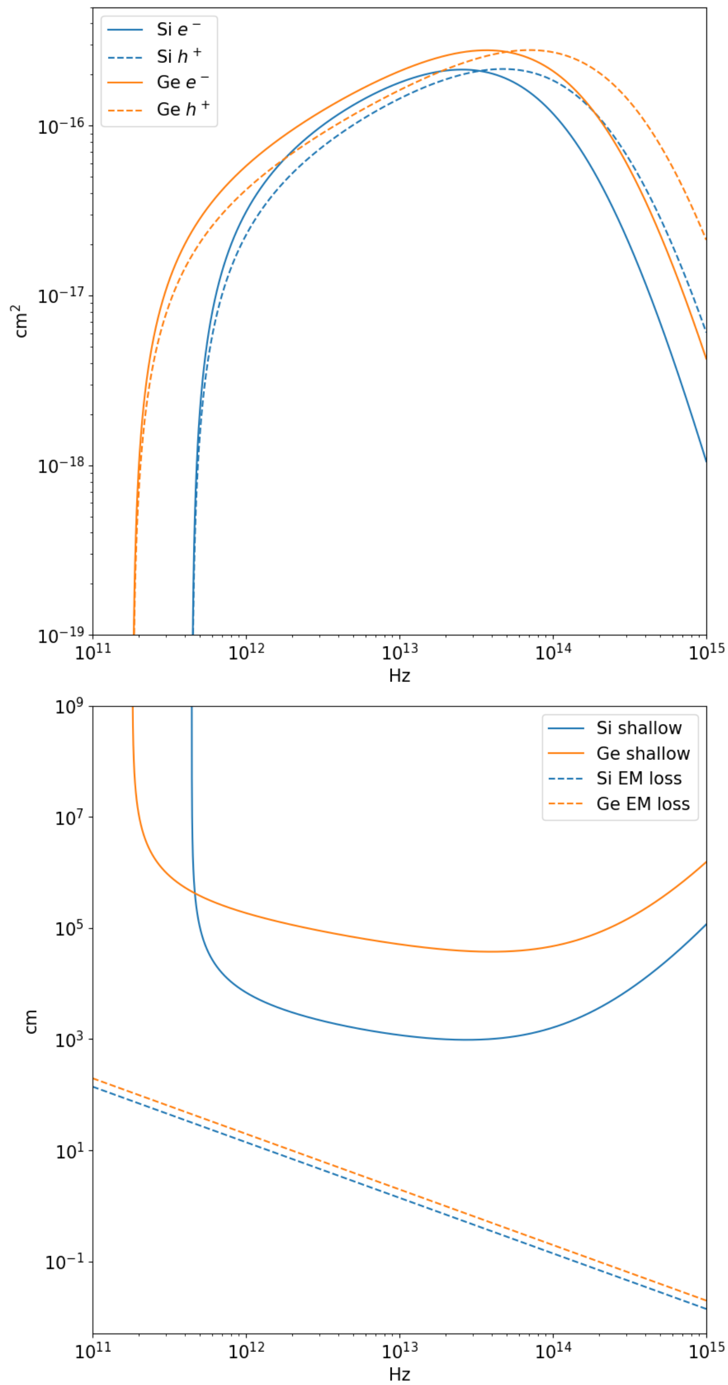


Figure 5.3: The BB-photon shallow trapping site photoionization cross section (top) and mean free path (bottom) in Si and Ge crystals, plotted against the photon frequency using the same style of Fig. 5.2. In the bottom panel, instead of showing photon partial mean free paths separately for electrons and holes, we combine both based on the assumed relative populations for a more meaningful presentation, represented by the solid lines. Also shown in the bottom panel by the dashed lines for reference is the total photon mean free path derived from the material's loss tangent data (Krupka et al., 2006; Datta et al., 2013; Kopas et al., 2020), including all photon absorption mechanisms, not just the shallow impurities. Detailed explanation may be found in the corresponding text.

5.3 Photon simulation

In the last section, we discussed in detail the physics elements to be included in the photon simulation. Considering the nature of these elements, we need two distinct systems: 1) A "particle-like" simulation, which performs a time-slice finite-stepping routine for every particle to travel from its generation to absorption, and 2) a "wave-like" simulation, which performs a full EM field calculation when the photon wavelength approaches the size of the mechanical-optical cavity it is in. In the following section, we first introduce our constructions for both systems separately, and then we discuss the status of integrating the two systems to deliver predictions that can be compared to experimental data.

5.3.1 Particle-like simulation

5.3.1.1 Model construction

We utilize the Geant4 simulation toolkit widely adopted by the particle physics community as the foundation of our particle-like simulation program (Agostinelli et al., 2003; Geant4, n.d.). In Geant4, particles are point-like and propagate in finite ray-tracing steps, so it is applicable for the open space defined in Tab. 5.1. Since in practice the BB photons can not produce any types of free secondaries with their low energy, we use the built-in `G4OpticalPhoton` class for these BB photons and disable all other particles and interactions. Unlike the physical "photon," the optical photon in Geant4 is an artificial particle with no associated particle interactions and propagates based on user-defined boundaries and medium conditions. Between boundaries, an optical photon can only travel in a straight line or be absorbed. In parallel with `G4OpticalPhoton`, we adapt external packages or construct custom-made ones to satisfy the physics requirements of our application.

In traditional Geant4-based applications, particles of interest are typically energetic enough to travel through structural materials like aluminum, copper, polymers and therefore allow Geant4 to ignore intricate mechanical structures. However, for the low-energy BB photons, all metals are essentially opaque and can not be ignored, mm-thick dielectric materials can cause significant attenuation, and mechanical cutouts and slits as narrow as 10s of μm need to be accurately incorporated for photon propagation. Due to such requirements, it is impossible in practice to construct our model using the parametrized simple geometries offered by Geant4. We therefore take the approach of first constructing the full mechanical model in a professional CAD (computer-aided design) program and then duplicating it into our simulation with the help of external tools.

For the results presented in the following, we use (SOLIDWORKS, n.d.) 2017 for CAD model construction and export the models in ASCII `.STL` format at the highest translational and angular resolutions in SOLIDWORKS, ≈7 μm and ≈0.5°, respectively. We import the `.STL` files into Geant4-based simulations using the CADMesh tool (2$^{\text{nd}}$ ver.) as `TessellatedMesh` type objects (Poole et al., 2012a; Poole et al., 2012b). The standard syntax and technical details for CADMesh can be found in (**CADMesh_github**). To ensure an accurate placement for the mechanical parts in the imported models, we assemble all the parts in each project to be a SOLIDWORKS *assembly* for the `.STL` file export, where we disable the default "translate to positive space" option and save the assembly by individual pieces rather than a single object. Such settings give individual `.STL` file for each mechanical part at its correct location in the global coordinate, allowing us to place each part via CADMesh independently based on the purpose of the simulation, e.g., removing certain components for test or design studies.

We further group mechanical parts that are frequently "installed" together into different modules, which we call the sub-detectors in the script. These sub-detectors include the main cryostat assembly, the BB radiation shield, the particle detector module, etc. This method makes the script more readable, and the user can also more easily adjust the simulation construction without introducing errors. We design each sub-detector to be a standalone C++ class that provides a `G4AssemblyVolume`, which in fact is the object the script uses to build `CADMesh::TessellatedMesh`, not the individual parts. All these classes are initialized in the standard Geant4 `DetectorConstruction` script, so the user can call each `G4AssemblyVolume`'s `MakeImprint` function to install the corresponding sub-detector based on the purpose of the simulation. Since we already place all the mechanical parts at the correct absolute locations in the global space, we always use a trivial (0,0,0) translation vector for `MakeImprint`.

For readers attempting to reproduce our method, we reported to the author of CADMesh that, by the time this chapter was written (Jan. 2021), there was a bug for the program to properly utilize SOLIDWORKS-generated ASCII `.STL` files. The corresponding error message during initializing the simulation is

```
STL files start with "solid".
Make sure you are using an ASCII. STL file.
```

We were able to circumvent the issue by manually editing the `.STL` files in a text editor. The "`solid`" keyword in the first line of the files can not have any leading characters like a space, a tab, etc., and has to be separated from the rest of the line by a single space. The next word after `solid` is taken as the name of the part, which also can not contain any special characters.

5.3.1.2 Material optical property

We collect all the material data that a simulation uses into a "`[project name]Material`" class similar to the sub-detector modules. This material database class is initialized together with the sub-detector classes, so the materials, i.e., the material class' class members, cab be invoked in the sub-detector construction. Specifically, these materials are called by the `G4LogicalVolume` objects in the sub-detector `G4AssemblyVolume` objects. The primary motivation for utilizing a shared material database is so that we can avoid repeatedly instantiating the same material for multiple sub-detectors, which involves manually copying the lengthy data and can lead to inconsistencies especially during data updates. The method also makes the script more readable, maintainable and in principle requires less run-time memory for the simulation.

Fig. 5.4 summarizes the optical properties of the materials currently in our material database: Vacuum, oxygen-free high-thermal conductivity (OFHC) copper, Cirlex, printed circuit board (PCB), crystalline silicon and germanium, and perfect reflector and absorber. This list includes the majority of the materials in practical cryogenic experiments, except the last is for testing idealized scenarios. Cirlex is included because SuperCDMS and DMKID utilize Cirlex for detector clamps. We currently take Cirlex as a general approximation for high-density polymers and use its optical properties for other polymers such as Kapton. For crystalline silicon and germanium serving as detectors, we apply further treatments in addition to selecting the corresponding materials to better model the photoionization. We introduce the treatments in a dedicated section later.

We have also compiled from the literature the photon mean free path data for 20 other materials frequently employed as radiation shields in microwave- to mm wave-range applications. These materials have not yet been included into the simulation database but can be implemented similar to existing materials. Fig. 5.5 summarizes the results for the reader's quick reference, with detailed frequency dependent data found in the corresponding citations (see figure caption).

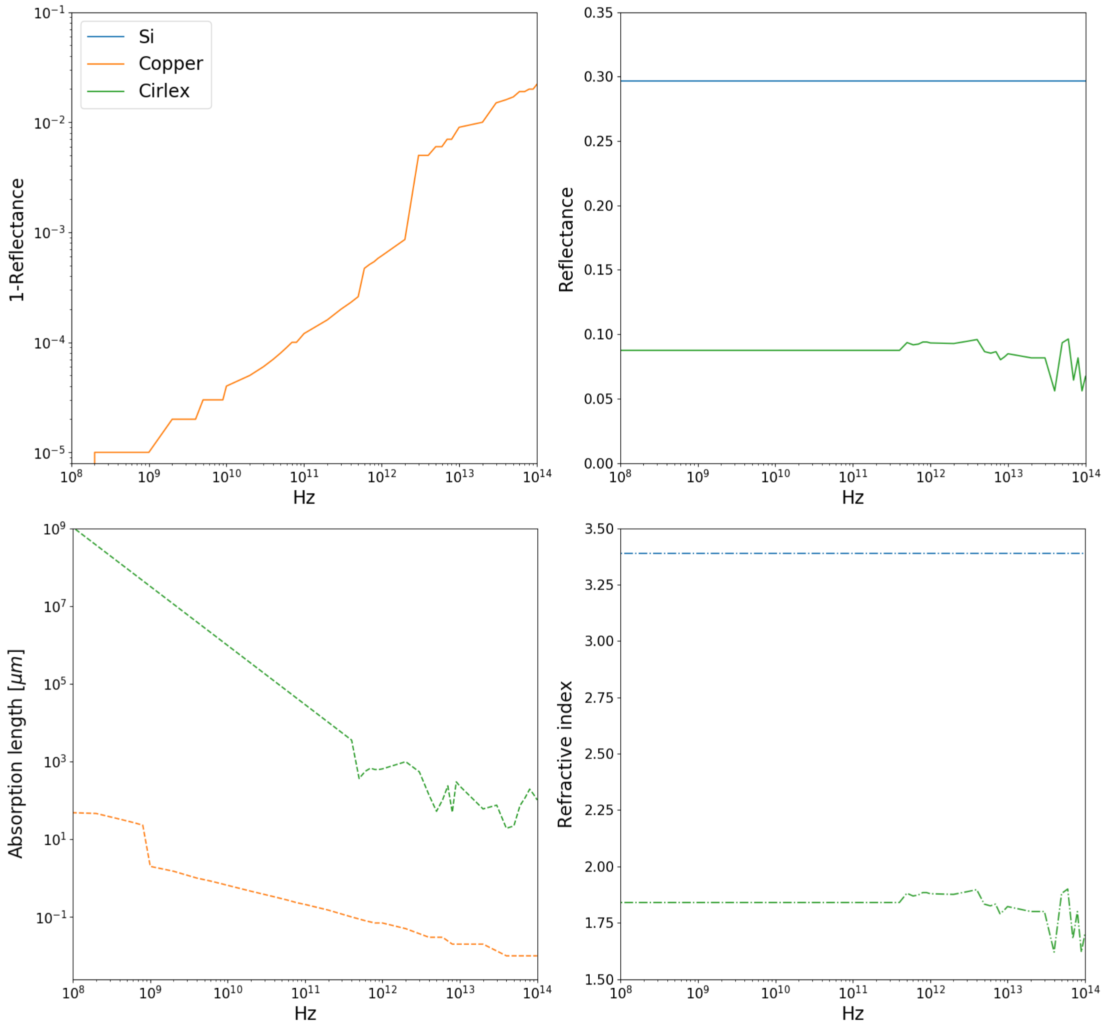


Figure 5.4: The frequency-dependent material properties currently used in our BB photon simulations. Blue, orange, and green curves represent Si, copper, and Cirlex (polymer proxy), respectively. Due to copper's high reflectance, we present the reflectance's difference from unity in a separate panel (top-left) for clarity. GEANT4 automatically calculates dielectric materials' reflectances using Fresnel's equations, so the results depend not only on the frequency but also the incident angle and the polarization. As an example, we show in the top-right panel the normal incidence results that are independent of polarization. Data shown in this figure for Si are not used for detectors.

There is a critical disadvantage of the current GEANT4 optical photon framework that prevents us from straightforwardly applying the above materials. Nominally, the program calculates the photon's behavior at the boundary (reflection, transmission, absorption) using the classical Fresnel's equations, or the user can overwrite the calculated optical properties by attaching a customized surface to every object. We

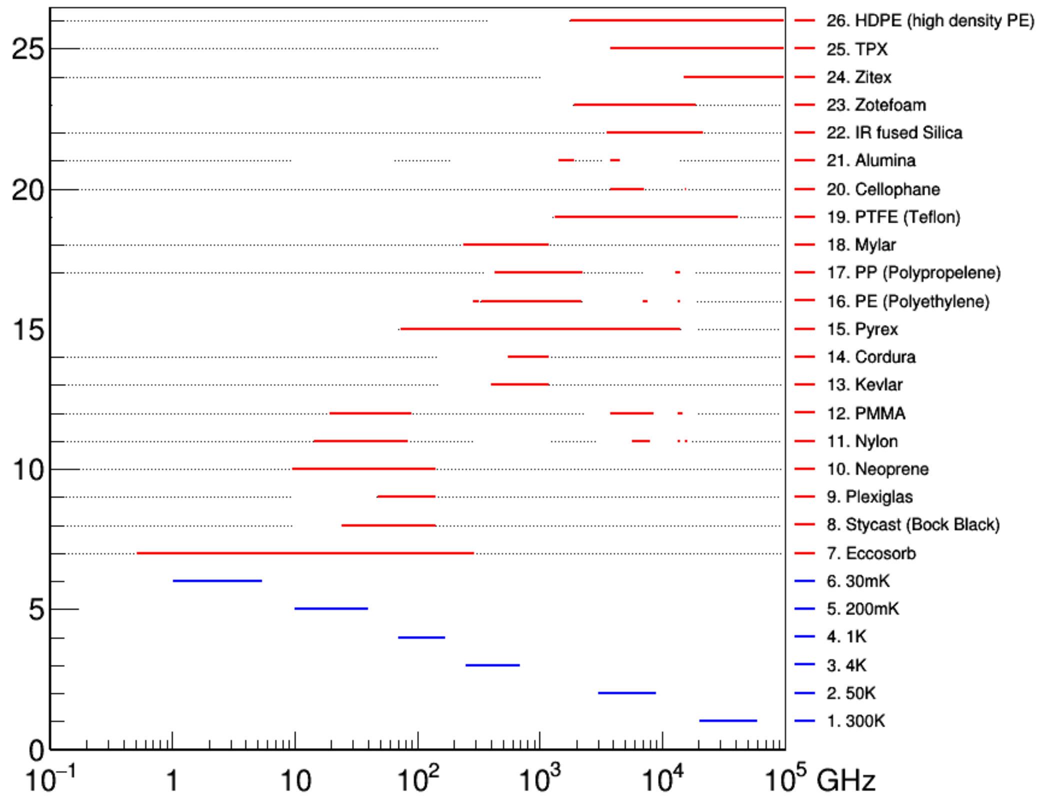


Figure 5.5: A summary of frequency-dependent photon mean free path data for trans-mm wave absorbers. Materials are labeled by their y values listed on the right. The blue lines (1–6) represent the major BB spectral ranges exceeding FWHM for the corresponding BB temperatures (Eq. (5.2)). The red lines (7–26) represent the frequency ranges that the materials exhibit photon mean free paths less then 1 cm, otherwise the ranges are left blank. The dotted lines represent undetermined ranges that lack data. Due to the high number of references for creating this figure and the unpublished status of some of them, we collect all the references as a data folder on (Chang, n.d.), where the reader can find detailed measurements.

realize that, while the Fresnel's equations are in principle correct, when calculating metal reflectance, the result is sensitive to the imaginary part of the complex refractive index that represents the metal's radiation dissipation. Especially at cryogenic temperatures, practical metal quality variation can result in orders-of-magnitude different imaginary parts of refractive indices and thus a highly uncertain reflectance from the Fresnel's equations. Moreover, such dependence becomes more serious when the frequency approaches the metal's plasma frequency, which for most metals in our simulation (Cu, Al, etc.) is in $\mathcal{O}(1)$ THz (Ordal et al., 1985) and coincide with the major frequency range for $\mathcal{O}(10)$-K BB radiation (Fig. 5.5). So, we conclude Fresnel's equations are unreliable for our application. In addition, Materials' practical surface conditions such as roughness or oxidation can also impact objects' actual surface optical properties but are not considered by the Fresnel's equations. Therefore, we decide to adopt customized optical surfaces for all the objects in our simulation.

However, assigning individual surfaces to objects using empirical data poses another challenge in practice. As explained previously, we use detailed CAD models to preserve the necessary complexity for our simulations. Since in Geant4's current design one needs to create an independent optical surface for every object, our complex models with large numbers of mechanical parts effectively lead to C++ scripts that are impossible to write or maintain. We then realize, despite the large numbers of parts, the parts are predominantly made of a handful of common materials listed above. Furthermore, since we typically fabricate parts of the same material using the same technique, these parts are expected to exhibit common surface optical properties.

We therefore manually modify Geant4's source code, so the surface properties that were exclusively callable by `G4LogicalSurface` can be passed to `G4Material`. When an optical photon travels through an interface with such a material on the stepping endpoint side, `G4LogicalSurface` processes takes place to determine the boundary interaction. With such a modification, we only need to maintain the surface properties with the corresponding materials in the shared material database without even constructing a `G4LogicalSurface`. Fig. 5.6 shows the result. We are able to construct the entire unsimplified SuperCDMS SNOLAB cryogenic system with fully defined optical properties with a short script, which is otherwise impossible for the standard SuperCDMS simulation (Kelsey et al., 2021). To build our modification, readers can download the modified `G4OpBoundaryProcess.cc` from (Chang, n.d.) to replace the corresponding script in Geant4 source code and then recompile Geant4.

5.3.1.3 BB photon generation

We have created a C++ toolkit for users to generate BB photons as the primary particle in Geant4 simulations. The toolkit consists of two parts: One, determining the primary particle's energy and, two, specifying its initial location and direction. The toolkit samples these initial conditions from user-defined probability distribution functions (PDFs). For idealized BB radiation, the appropriate energy PDF is Eq. (5.3). One can also use other PDFs to include emissivity's frequency

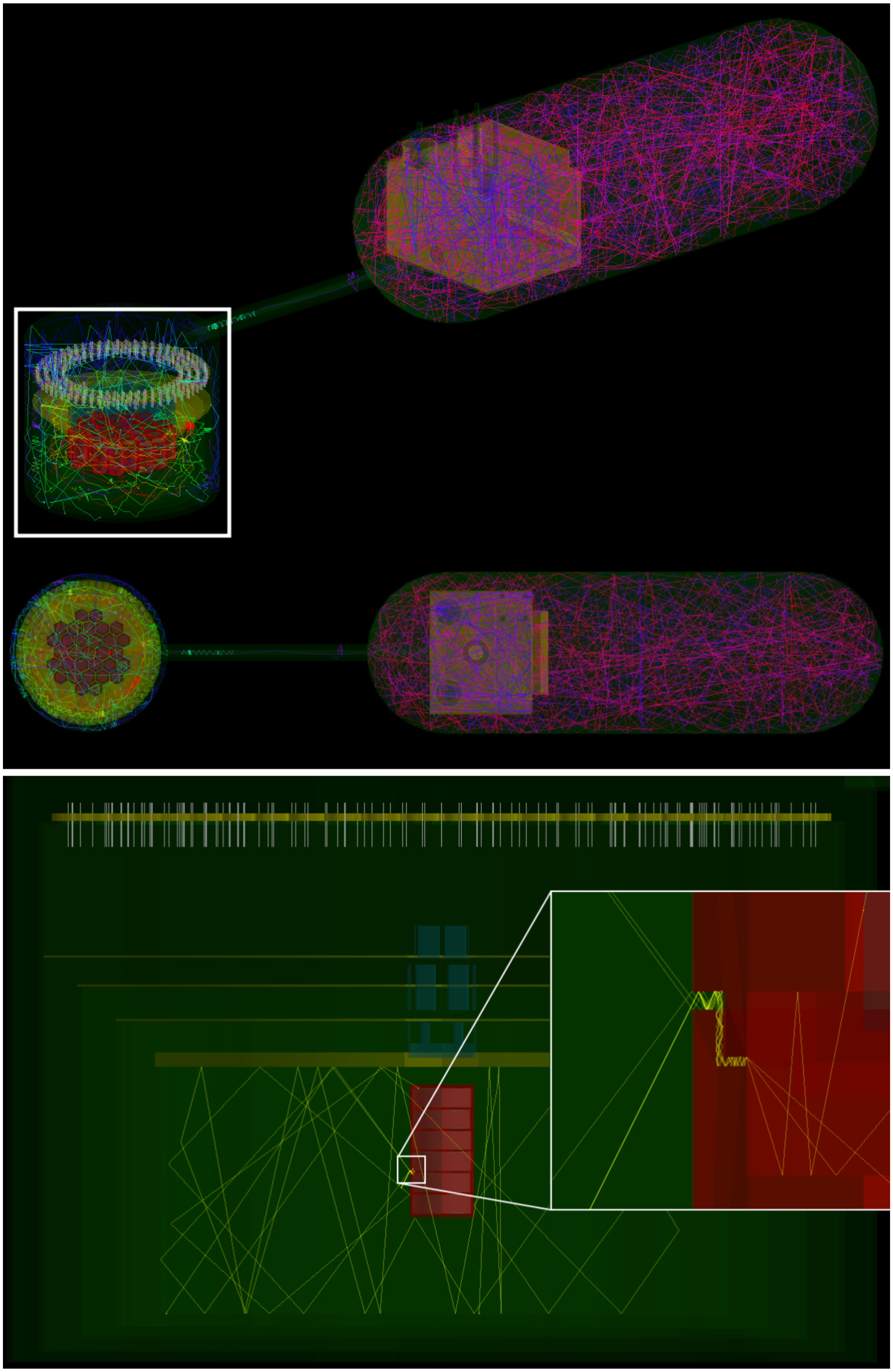


Figure 5.6: The GEANT4 simulation screenshots for the full SuperCDMS SNOLAB cryogenic system (top) and a zoom-in example (bottom) with irrelevant mechanical structures suppressed for visual clarity. The colored tracks represent the photons with corresponding temperatures E_γ/k_B, roughly 200–250 K, 3–10 K, <0.5 K for blue-purple, green, and yellow-red tracks, respectively. The complete view shows the electronics tank (horizontal on the right) is dominated by warmer photons generated by the electronics, while the main experimental apparatus chamber (vertical cylinder on the left) is dominated by <10-K photons as expected. Shown in the bottom panel, corresponding to the indicated region in the top panel, is an example event where the photon penetrates the 15-μm detector housing slit, formed between the lid and the sidewall, into the detector chamber.

dependence or for non-BB emitters. Location and direction PDFs depend on the geometry, the surface condition, and the spatial orientation of the radiator and thus do not have simple analytical forms.

For non-thermalized emitters, energy and location/direction PDFs can be correlated. However, we are mainly interested in applications consisted of sealed chambers, where we expect the photons to experience high numbers of reflections before penetrating the mechanical seals separating the chambers. In this case, we expect the practical reflection to randomize the initial location and direction information when the photons arrive at the detector. Therefore, we expect the BB radiation the detector experiences to be insensitive to the correlation between these initial conditions, which allows us to sample the PDFs independently.

For model construction, we assume every object in the cryostat is a perfect BB at its nominal temperature. Being an idealized BB, every location on the surface of the object exhibits the same temperature and emits BB photons at an equal probability. For a given object, this property again suggests we can sample from the energy PDF and the location/direction PDFs independently without worrying their correlation, such as due to temperature variation on the surface. We therefore sample from the BB spectrum given by the object's nominal temperature first, and then we randomly select every location evenly on the surface of the object to be the photon origin. To determine the emission direction, we use the fact that there is no preferred direction or polarization for BB radiation, which allows us to select a random direction evenly toward the hemisphere above the photon origin. Our algorithm determines the direction by first selecting random azimuthal and polar angles for the upper hemisphere and then rotating this direction by the difference of the zenith and the normal direction of the photon origin.

Below we detail the code structure for the BB photon toolkit for readers' adoption or future improvement, and we have also prepared sample scripts on (Chang, n.d.). We currently find it challenging to acquire the normal vector information for `TessellatedMesh` objects in CADMesh, so instead we prepare several parameterized geometries commonly seen in the simulations to approximate the BB surfaces. These geometries include polyhedrons, sphere, cylinder, round or rectangular surfaces, whose normal vectors can be calculated analytically for every chosen photon origin on the surfaces. When constructing an actual simulation, we enclose major radiating objects, such as the cryostat chambers or dedicated BB radiators, by their most similar proxy geometries. For objects with intricate structures that can not be represented exactly, we calculate the objects' precise surface areas in SolidWorks and enclose them by the same proxy surface areas. Based on the argument that the direction and location information is mostly lost after numerous reflections, we argue the approximation is sufficiently accurate in generating the BB radiation environment. Even when an object is too complex to be described by one of the proxy geometries, we find it always possible to construct its full surface by a few of the simplified geometries to a satisfactory precision.

For assigning photon energy in the standard GEANT4 `PrimaryGenerationAction` class, the toolkit has two header-source pairs:

`BBEvt.hh/cc`
`GetBBSpecCDF.hh/cc`.

The `BBEvt` "BB event" class is the information carrier for individual BB photons, containing 3-vectors for position, momentum, and a numerical value for the energy. The `GetBBSpecCDF` "get BB spectral CDF" class provides the functions for calculating and sampling energies from Eq. (5.3) or user-defined energy spectra. For defining the BB geometries, their spatial orientations, and employing them into the simulations, the toolkit has:

`GeometricSurface.hh/cc`
`ThermalSurface.hh/cc`

The `GeometricSurface` class is the information carrier for individual geometric surfaces, containing a list of vectors, matrices, and boolean identifiers to describe the surface's shape, location, rotation, and direction of emission (inward, outward, or both).

The `ThermalSurface` class constructs the actual BB surfaces with the previous classes. It contains the members

`Add[Shape]Surface()`
`GetEvt()`

and needs to be initialized with a photon energy PDF and an emissivity.[5] A `ThermalSurface` can be a collection of many geometric surfaces as needed to represent the shape of the BB. The user can invoke the `Add[Shape]Surface()` function to register different geometries to the `ThermalSurface`. Afterwards, the user can invoke the `GetEvt()` function repeatedly to generate BB events based on

[5] Set to unity if not specified.

the defined properties. `GetEvt()` returns a `BBEvt` that one can pass the member parameters (energy, position, etc.) to the standard `G4ParticleGun` to generate an event.

Lastly, to incorporate multiple thermal surfaces of different temperatures in `PrimaryGeneratorAction`, `ThermalSurface` offers the members

$$\texttt{temp},\quad \texttt{area},\quad \texttt{effArea}$$

for its temperature, area, and the emissivity-weighted effective area, respectively. The user can repeatedly choose which `ThermalSurface` is to emit a BB photon using these parameters. Since the average photon energy for a BB is proportional to its temperature, i.e.,

$$\langle E_\gamma \rangle = \frac{360 \cdot \zeta(5)}{\pi^4} \cdot k_\mathrm{B}T \approx 3.83 k_\mathrm{B}T, \tag{5.58}$$

the Stefan-Boltzmann law suggests the emitting number rate is proportional to the BB's effective area and T^3, i.e.,

$$\begin{aligned} N_\mathrm{BB}(T) &= \frac{\epsilon\sigma A T^4}{3.83 k_\mathrm{B}T} && (5.59) \\ &\approx (1.07 \times 10^{11}\ [\mathrm{s^{-1}cm^{-2}K^{-3}}]) \cdot \epsilon A T^3. && (5.60) \end{aligned}$$

Thus, if all surfaces in a simulation are idealized BB surfaces, the user should choose the emitting surfaces by the radio of

$$\texttt{effArea}^*\texttt{pow}(\texttt{temp}, 3). \tag{5.61}$$

Fig. 5.7 presents a result of incorporating multiple `ThermalSurface`s of different temperatures.

5.3.1.4 Si/Ge crystal photoionization

We discuss in the following our practical code implementation based on the trans-mm photon D^-/A^+ photoionization model derived previously. Since GEANT4 performs a finite stepping style simulation, we are particularly interested in deriving a photon absorption mean free path model, which GEANT4 can utilize to determine the survival of particles after each step. We then record the absorption events for further analysis. In the end, we realize that the empirical data found in the literature suggest Si and Ge crystals absorb trans-mm photons much more strongly than the D^-/A^+ contribution alone. Therefore, it is more appropriate in physics, and in fact simpler

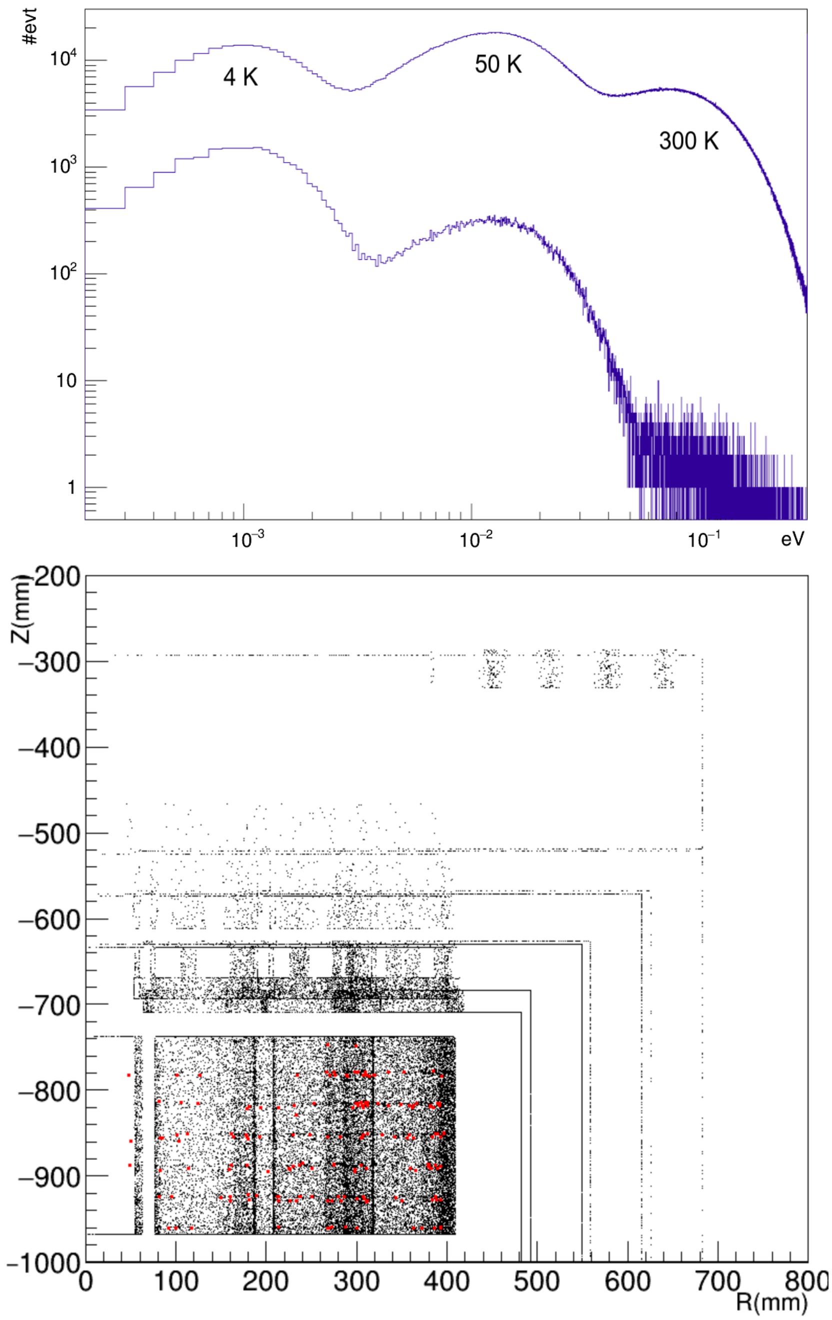


Figure 5.7: An example BB photon simulation based on an early SuperCDMS SNOLAB cryostat mechanical and thermal design. The top panel shows the photon energy spectra at the generation (upper curve) and entering the base-temperature can (lower curve). We label the spectral peaks by their contributing temperatures, 4 K, 50 K, 300 K, respectively assumed for the electronic feedthrough mechanics, the cryostat's outer can, and the electronics tank (Agnese et al., 2017). The result clearly shows that the photons emitted by structures further from the base-temperature can is attenuated more. The bottom panel shows the locations (black dots) the photons are absorbed, as a side view of the highlighted region in Fig. 5.6. We plot the result by the radius (R) from the center of the cryostat based on the symmetric design. Since most photons are absorbed in the shallow regions of materials, the black dots collectively visualize the entire mechanical structure. The large red dots highlight the absorption in the detector crystals, which are the type of the detector housing penetration events described in Fig. 5.6.

in practical implementation, to decouple the photon absorption in simulation from the photoinozation (leakage current) data generation that is done with the simulation data.

Although we hypothesize the D^-/A^+ photoionization (Fig. 5.3) to be the primary leakage current generation mechanism, these impurities likely are not the dominant source for the total photon absorption in Si and Ge. As explained previously, (Domange et al., 2012)'s result suggests there exists other deeper sites at much higher densities. These deeper sites can in principle also absorb BB photons even though (Domange et al., 2012)'s experiment did not directly demonstrate photo-absorption. We are motivated by the finding to search for empirical data in the trans-mm range for the simulation to determine photons' passage in detector Si/Ge crystals more realistically.

We find trans-mm wave absorption in semiconductors has been actively studied in the field of microwave engineering for many years, from where we can obtain plenty of empirical measurements. Instead of presenting the photon absorption as mean free path, it is more common in the field to use loss tangent, which is defined by the power attenuation

$$\frac{P(x)}{P(0)} = e^{-\delta k x}, \tag{5.62}$$

where P is the radiation power at the distance x into the material, and δ is the loss "angle" parameter. On the other hand, a mean free path specifies the particle survival rate by

$$\frac{\Phi(x)}{\Phi(0)} = e^{-x/L}, \tag{5.63}$$

where Φ is the particle flux at x, and L is the mean free path. One can relate the two equations by

$$P = \Phi E_\gamma \tag{5.64}$$

and find

$$L = \frac{1}{\delta k} = \frac{c}{\delta n \omega}, \tag{5.65}$$

where we use n for the refractive index of the medium and keep c as the speed of light in vacuum. For lossy materials, one can write the permittivity ϵ as a complex number, whose real part manifests the propagation, and the imaginary part gives rise to the exponential decay for the loss. It is thus natural to present ϵ by the tangent value of its principle angle, i.e.,

$$\tan\delta \approx \frac{\mathrm{Im}(\epsilon)}{\mathrm{Re}(\epsilon)} = \tan\angle\epsilon, \tag{5.66}$$

which gives the ratio of the lossy and the propagation parts as the fractional loss per wavelength of propagation (Jackson, 1998).

Notice that we use approximation for Eq. (5.66) because, in general,

$$\tan\delta = \frac{\omega\,\mathrm{Im}(\epsilon) + \sigma}{\omega\,\mathrm{Re}(\epsilon)}, \tag{5.67}$$

where σ is the conductivity of the material (Jackson, 1998). For high-purity Si and Ge at cryogenic temperatures, typically

$$\sigma \approx 10 - 20 \quad [\mathrm{mS/m}], \tag{5.68}$$

and in 0.1–10 THz,

$$\omega\,\mathrm{Im}(\epsilon) \approx 10^{-2} - 10^{3} \quad [\mathrm{mS/m}]. \tag{5.69}$$

We compare the above values and realize

$$\omega\,\mathrm{Im}(\epsilon) \gg \sigma \tag{5.70}$$

does not necessarily hold to support Eq. (5.66)'s approximation. This potential inapplicability for the usual definition of loss tangent is partially due to our moderate frequencies of interest, i.e., not as high as IR or visible light for the crystals to be "opaque," and partially due to the fact that semiconductors exhibit nonnegligible conductivities. Moreover, we should note that ϵ and σ are derived from the electronic system physics, which is inherently dependent of temperature, incident photon frequency, crystal quality, etc., and can vary by orders-of-magnitude sample-to-sample. We focus on developing the simulation infrastructure in this thesis and remind future readers that, based on the above remark, if possible, they should always utilize empirical loss tangent data for the actual Si/Ge sample being studied.

Due to the current unavailability of such data for us, we use loss tangents in the literature to continue the simulation development. For Si, we adopt the measurements by (Krupka et al., 2006; Datta et al., 2013), which show the loss tangents stop being significantly (logarithmically) sensitive to the temperature after charge carriers "freeze out" below 25 K. The observed asymptotic values in the references, however, differ by a factor of a few at the 10^{-4} level. We choose the measurement obtained for 6–17 GHz by (Krupka et al., 2006), which shows a decreasing trend toward higher frequencies into our main frequency range of interest. Similar measurements by (Datta et al., 2013) at frequencies closer to our range of interest, 240–1800 GHz, suggest the loss tangents deviate from the decreasing trend (Krupka et al., 2006) obtained and asymptote to as low as 7×10^{-5} in certain samples. Based on these results, we choose

$$\tan \delta_{\mathrm{Si}} \coloneqq 1 \times 10^{-4} \tag{5.71}$$

as a reasonable representation.

For Ge, we follow the results obtained by (Kopas et al., 2020) and choose

$$\tan \delta_{\mathrm{Ge}} \coloneqq 6 \times 10^{-5}. \tag{5.72}$$

(Kopas et al., 2020) obtained the Ge data at 40 mK using superconducting Nb microwave resonators instrumented on the crystals, therefore at frequencies much lower than the BB photons. By probing different depths into the crystal sample with different resonator designs, (Kopas et al., 2020) concluded that the lossiness was dominated by the surface contamination within about 20 nm. Considering that we need a loss tangent that represents the bulk of the crystal, we choose the smallest value obtained in the reference.

We substitute the above loss tangents into Eq. (5.65) with $n = \sqrt{\epsilon_r}$ from Tab. 5.2 to calculate the photon mean free paths for Si and Ge, shown in Fig. 5.3. These empirical mean free paths are lower than that derived from our impurity site model by 3–4 orders-of-magnitude, which is consistent with (Domange et al., 2012)'s conclusion that, to explain data, the total photon absorption is likely dominated by sources other than the shallow impurities by 3–4 orders-of-magnitude. It is encouraging that we are able to arrive at such consistency with data from a purely first-principle model for the shallow impurities. In terms of simulation implementation, the dominance of the empirical mean free paths allow us to directly adopt them for Geant4 to determine photon track termination. We record these terminated photons and then randomly determine whether a photon leads to a leakage current event using the ratio of the empirical and the shallow impurity-only mean free paths at the photon frequency.

We have two remarks for the methodology demonstrated above. First, the empirical mean free paths span from cm to m in our frequency range of interest, suggesting our cm-size detectors are nearly transparent to the BB photons. Especially for the few-Kelvin (sub-THz) photons that are most relevant to the shallow site photoionization, i.e., >750 μeV and emitted most closely, the probabilities for a photon to be absorbed in a few-cm crystal or on a copper surface (Fig. 5.4) per incident are comparable. Since neither dominates the photon absorption, we expect the simulation results to be sensitive to the accuracy of both inputs. Second, the results in Fig. 5.3 should be regarded only accurate to order-of-magnitude at its present state. Recall that our model requires a repulsive atomic cloud component to cancel the dipole attraction to form the shallow binding state, where both components are much larger than the resulting binding state. We demonstrated the idea with a phenomenological Lennard-Jones model tuned to yield the observed binding energy, but we expect the result remains sensitive to this large-component cancellation. Therefore, more studies are needed to examine whether out treatment is sufficiently accurate, or a more sophisticated full-Hamiltonian calculation is necessary for the atomic cloud repulsion. At the same time, the loss tangents can also vary appreciably from the above example values sample-to-sample.

5.3.1.5 DAQ code

We briefly introduce our current DAQ code structure for interested readers. Due to frequent mentioning of terminologies, it may be easier to understand the following side-by-side with the actual script. We utilize the ROOT data analysis framework (release 6.16/00, Brun and Rademakers, 1997) for the simulation data acquisition. It is C++-based like GEANT4 and makes the code integration more straightforward. We create a control class

`ROOT_data_record`

for the simulation program to interface with ROOT functions in an organized fashion. `ROOT_data_record` offers proxy functions for instantiating and manipulating ROOT objects such as files (`TFile`), histograms (`TH1F`), n-tuples (`TTree`), etc.; sample header- and source-code files can be found on (Chang, n.d.). We prefer compiling ROOT into our simulation projects with `Makefile`, and a sample `CMakeLists.txt` is also offered on (Chang, n.d.).

Although we choose ROOT for a direct data acquisition from the C++-based simulations, we convert all the ROOT `TTree` data into Python formats, so we can better exchange data and take advantage of available analysis tools with the wider Python community. We find it is the cleanest to convert ROOT `TTree` n-tuples into Python `pandas` n-tuples by

`root_pandas.read_root('[file name].root')`.

We declare a C++ `struct` object, `ABSPoint`, in the header

`run_ROOT_data.hh`

to store the photon absorption point data. It contains a series of numbers for event energy, momentum, position, several integers for the times of reflection, tags for the physical volume terminating the photon, and other general information. We use `ROOT_data_record` to create a `TFile` for each simulation run in the standard `BeginOfRunAction()`, which contains a `TTree` of the same n-tuple structure of `ABSPoint`. A global variable using the same `ABSPoint` structure is employed in parallel to track the photon information. We create this variable at the beginning and update the photon information in the main part of the standard `SteppingAction` source code. So, in `UserSteppingAction()`, if the particle is an optical photon and is terminated by a `G4OpticalPhoton` process, we save the end-of-step information to the global `ABSPoint` variable and fill it to the `TTree` at the end of the script. At the end of the run, we write the `TTree` to the `TFile`, so the file contains a `TTree` that contains all the photon absorption events.

5.3.2 Wave-like simulation

The rule one specifies for a wave (field) simulation is the converse of that for a particle simulation. Instead of specifying local properties as detailed above for GEANT4 simulations, one specifies the boundary conditions to solve for the field configuration based on the governing physics. For BB photons much below the electron pair-production energy threshold, this physics is the classical electromagnetism that is summarized by Maxwell's Equations.

Modern commercial programs can solve Maxwell's Equations at high precision, e.g., Simulia CST (computer simulation technology), Ansys HFSS (high-frequency structure simulator), Sonnet, COMSOL Multiphysics, etc. (Ansys, n.d.; Simulia CST Studio Suite, n.d.; SONNET, n.d.; COMSOL Multiphysics, n.d.). These programs adopt the finite-element analysis (FEA) technique to solve the boundary value problems. In an FEA, arbitrary user-defined volumes are first "meshed" into numerous cell elements at length scales much smaller than the expected local field variation scales. Within each mesh cell, Maxwell's equations are linearized to connect the boundary values from one side to the other. Collectively, connections across the cells yield the rule for connecting the specified conditions at the macroscopic boundaries, which in turn determines an unique field solution.

We choose HFSS[6] for our BB radiation simulation. Comparing the programs, COMSOL Multiphysics is specialized in integrating different physics into one simulation,

[6] Recently moved under ANSYS Electromagnetic Suite 19.2.0.

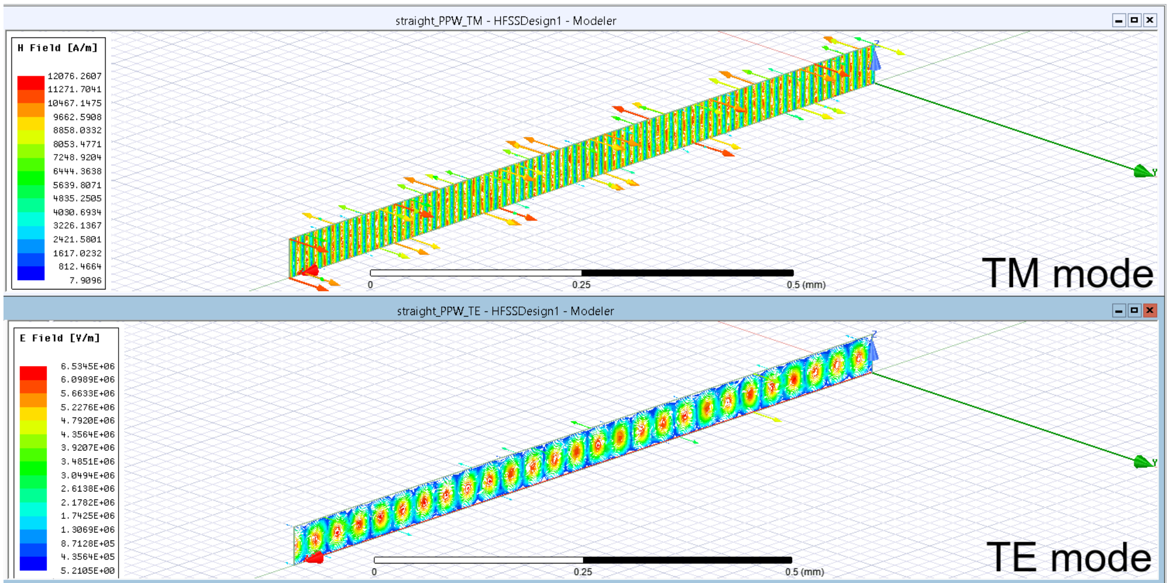


Figure 5.8: The top and the bottom panels present the HFSS-simulated TM and TE field configurations using PMC and PEC xz planes, respectively. In the TM (TE)-mode figure, colored surfaces and arrows represent equal-magnetic (electric) field strength surfaces and directions. These results are for 5 GHz, which is well below the 3 THz cutoff frequency, so both results are dominated by the lowest modes. The absolute field strengths in the results vary with the designated excitation and therefore is only meaningful relatively within the simulations.

such as fluid dynamics, mechanics, EM, etc. However, it is generally known to be less accurate and require more computation resources for high-frequency[7] EM simulations. Sonnet, on the other hand, is specialized in high-frequency EM and provides an outstanding precision. However, its modeling workflow and the style of result presentation are tailored for "laminated 3D[8]" devices, such as PCBs and microchips, which we find difficult to work with for our complex 3D models. Both CST and HFSS provide flexible modeling user interfaces to meet our needs. Due to different underlying algorithms for calculating Maxwell's equations, CST is typically more accurate in time-domain simulations, while HFSS is superior in frequency-domain simulations. Since we are primarily interested in steady-state power transmission, we choose HFSS over CST for this work.

It is straightforward to model the parallel-plane waveguide discussed in Sec. 5.2 in HFSS. Fig. 5.8 shows a simple example, where two conductive xy planes extending to infinity in $\hat{y}$ are separated in $\hat{z}$ by 50 μm; the reader can compare this model to

[7]Higher than RF.
[8]Or "2.5D" in industrial jargon.

Fig. 5.1 in the following discussion. To model the mechanical seals, we assign the *finite conductivity boundary* condition with a conductivity of

$$\sigma_{\text{Cu,HFSS}} \coloneqq 5 \times 10^9 \ \ [\text{S/m}] \tag{5.73}$$

to the surfaces. We compare early results assuming this high but finite conductivity or utilizing *perfect electric conductor (PEC)* boundaries and can conclude the difference is negligible in practice. We attach two *wave ports* with the same integration vectors to the end surfaces to represent the incoming and the outgoing waves. We do not set impedance normalization for the ports. For the infinitely extended direction ($\hat{y}$ in our model) that should not form nodes, we choose the waveguide width to be much smaller than the wavelengths of interest following HFSS' recommendation. This technique suppresses artificial resonances forming in the open direction and also saves computation time. We use 1 μm in $\hat{y}$ for all the following simulations, which prevents forming unphysical nodes up to 150 THz, i.e.,

$$\lambda/2 = 1 \ [\mu\text{m}] \text{ in vacuum} \quad \Leftrightarrow \quad f = 150 \ [\text{THz}]. \tag{5.74}$$

We assign the *perfect magnetic conductor (PMC)* or the *perfect electric conductor (PEC)* boundary conditions to the two xz boundaries that represent the waveguide's open direction. PMC enables HFSS users to construct surfaces that only exhibit perpendicular magnetic fields. Even thought PMCs do not exist in reality, it allows users to separate certain modes satisfying the condition from the mixed results and thus is frequently adopted. Contrasting with the PMC is the PEC that only exhibits perpendicular electric fields, which does exist in reality and serves as a good approximation for high-conductivity materials. One can compare these boundary conditions to the derivations in Sec. 5.2.3 and realize the PMC and the PEC, respectively, require the results to be purely TM or TE modes.

Fig. 5.9 shows the result for the above model, which yield a transmission of unity down to DC with PMC boundaries, while PEC boundaries give a cutoff at 3 THz corresponding to the height of the waveguide. This result is consistent with the derivation in Sec. 5.2.3. We can already conclude from this simple model that, for BB radiation at below ≈30 K (below a few THz), typical mechanical contact slits that are narrower than 100 μm only permits the radiation in the lowest TE and TM modes, and most of it is in the TEM mode below the TE mode cutoff frequency. We explained this expectation in the Sec. 5.2.3 derivation and have reproduced it here.

With such understandings in simulation technique as well as the actual situation in hand, we develop in the following our radiation mitigation strategies. We first test

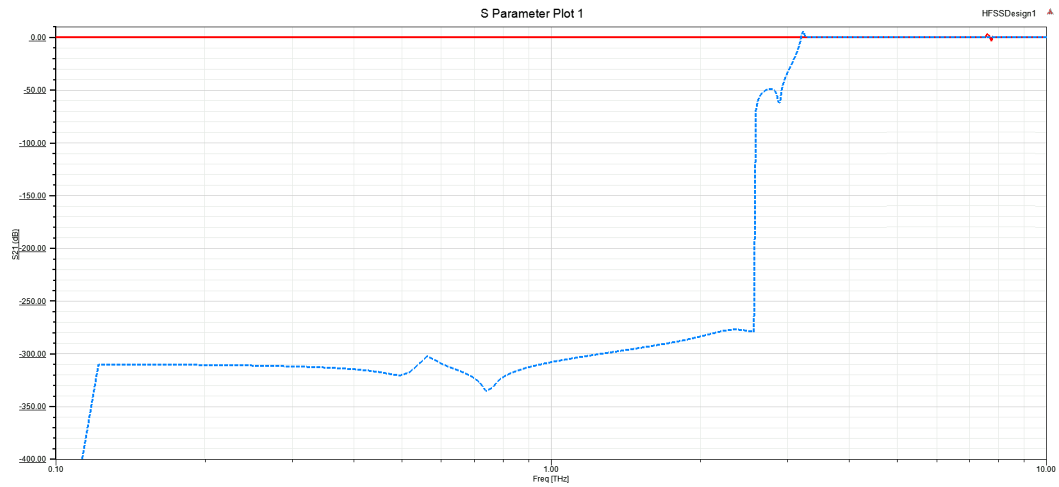


Figure 5.9: The upper red solid and the lower blue dashed curves show the HFSS-simulated TM- and TE-mode transmissions (S_{21}) in the 1 mm wide waveguide using PMC and PEC xz planes, respectively. Detailed descriptions for the waveguide construction and analysis are found in the corresponding text.

implementing a simple discontinuity in the waveguide (Fig.5.10), which is relatively straightforward to retrofit. For example, if found needed, one can add such step-like widening into detector housing seals by electrical discharge machining (EDM) or end milling. Fig. 5.10 shows the simulations for sudden height perturbations in the middle of a 1 mm (length) $\times$ 50 μm (height) parallel-plane waveguide, one expanding to 100 μm and the other to 73 μm. In both cases, extra k_z components are developed at the perturbations to satisfy the new boundary conditions, which can be seen from the vertical ($\hat{z}$) movements of the field contours down the transmission.

Shown in Fig. 5.11 is a comparison of the transmissions (S_{21}) for the original straight and the perturbed waveguides. The perturbation results in a series of narrow resonance-like structures above 2 THz, while the TEM transmission is mostly unaffected, except a few structures in 1–2 THz caused by the particular geometry of this example. Fig. 5.11 indicates that the technique is only effective above the TE cut-off frequency, which can be interpreted as selectively allowing the matching guided modes of the upstream and the downstream sections. It also explains that, since TEM transmission does not form resonances between the perturbed boundaries, it can remain largely unaffected. Therefore, we conclude the height perturbation technique alone is not useful because we are mainly interested in shielding TEM transmission.

Next, we test the 90° turn shown in Fig. 5.12. While an isolated turn as in Fig.

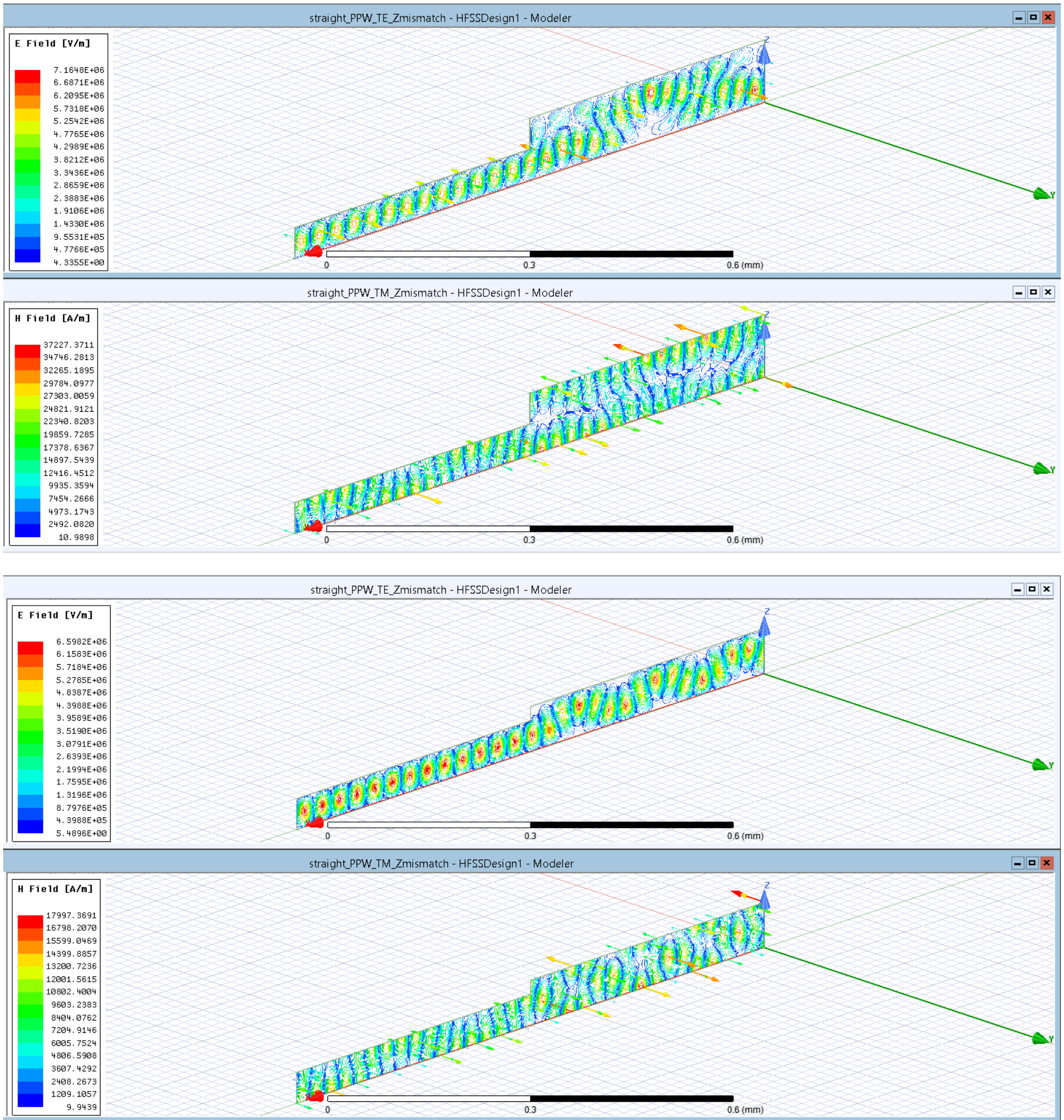


Figure 5.10: The 5 GHz field visualizations for 100 μm (top two) and 73 μm (bottom two) perturbations. For each result, the upper and the lower figures show the TE and the TM modes, respective. The color/arrow scheme is identical to Fig. 5.8.

5.12 is not practical to implement, one can combine multiple such turns to form a stair shape or a meander, which then become practical to be incorporated into the mechanical seals. So, we first study this single turn to understand its radiation shielding mechanism, and then we will present the result for a practical design next. The model we test is identical to the previous 1 mm $\times$ 50 μm waveguide but has a 90° angle in the middle. We find the transmission is further suppressed compared

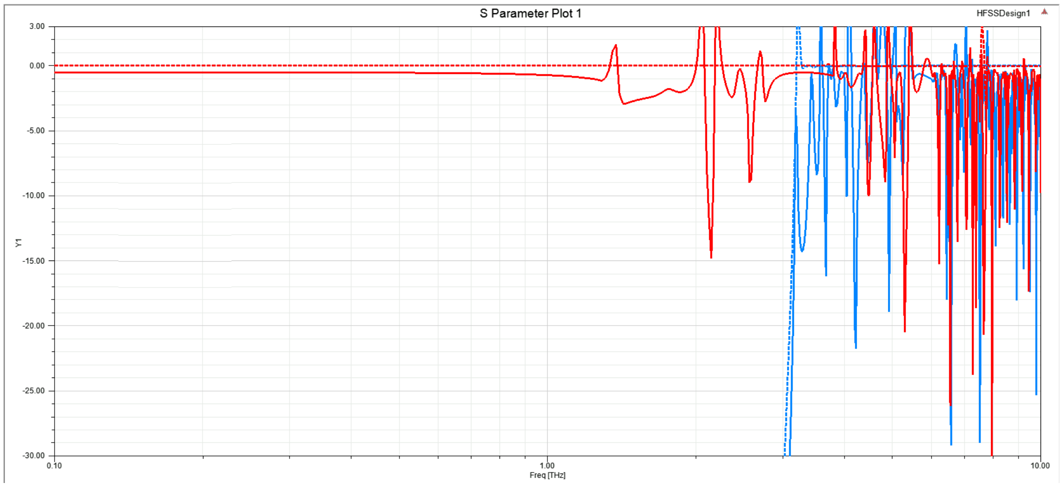


Figure 5.11: The TM (red) and TE (blue) transmissions (S_{21}) in the width-perturbed parallel-plane waveguides. The dashed and the solid lines represent the waveguides without the perturbation and widened to 100 μm, respectively. We choose to omit the 73 μm perturbation result for visual clarity, since it is similar to the 100 μm case.

to that in the height-perturbed straight waveguide. The field visualization shows that the originally quantized k_z becomes the wavevector along the new propagation direction, while the originally unbounded[9] k_x parallel to the waveguide now needs to satisfy the boundary conditions. Evidently, such a change of boundary conditions is more drastic than perturbing a dimension already requiring wavevector quantization and therefore results in a stronger transmission suppression. However, due to the same reason that the TEM transmission is insensitive to the boundaries, transmission below the TE cutoff frequency is still unaffected.

We summarize our wave-like propagation study to be a BB photon shielding strategy for mechanical seals. First, the TE cutoff frequency is a powerful shielding mechanism to be included. Because the cutoff frequency is controlled by the size of the waveguide, we propose utilizing accurate machining with a careful tolerance design to achieve $\lesssim$10-μm seal gaps. At reasonable prices, this precision can be achieved on our typical mechanical parts by industrial-grade milling. The narrow gap guarantees the cutoff frequency shields the radiation from BBs up to $\approx$150 K. A narrow section of this size should be as long as possible to suppress the evanescent transmission. Second, since the amount of radiation entering the seals is roughly proportional to the seals' entrance sizes, one should also apply the same precision fitting to minimize the seal entrances. Third, we propose including a waveguide

[9]Note that the values of k_x are related to k_z to obey the speed of light, so they appear to be also quantized. It is however not due to a constraint of the boundaries.

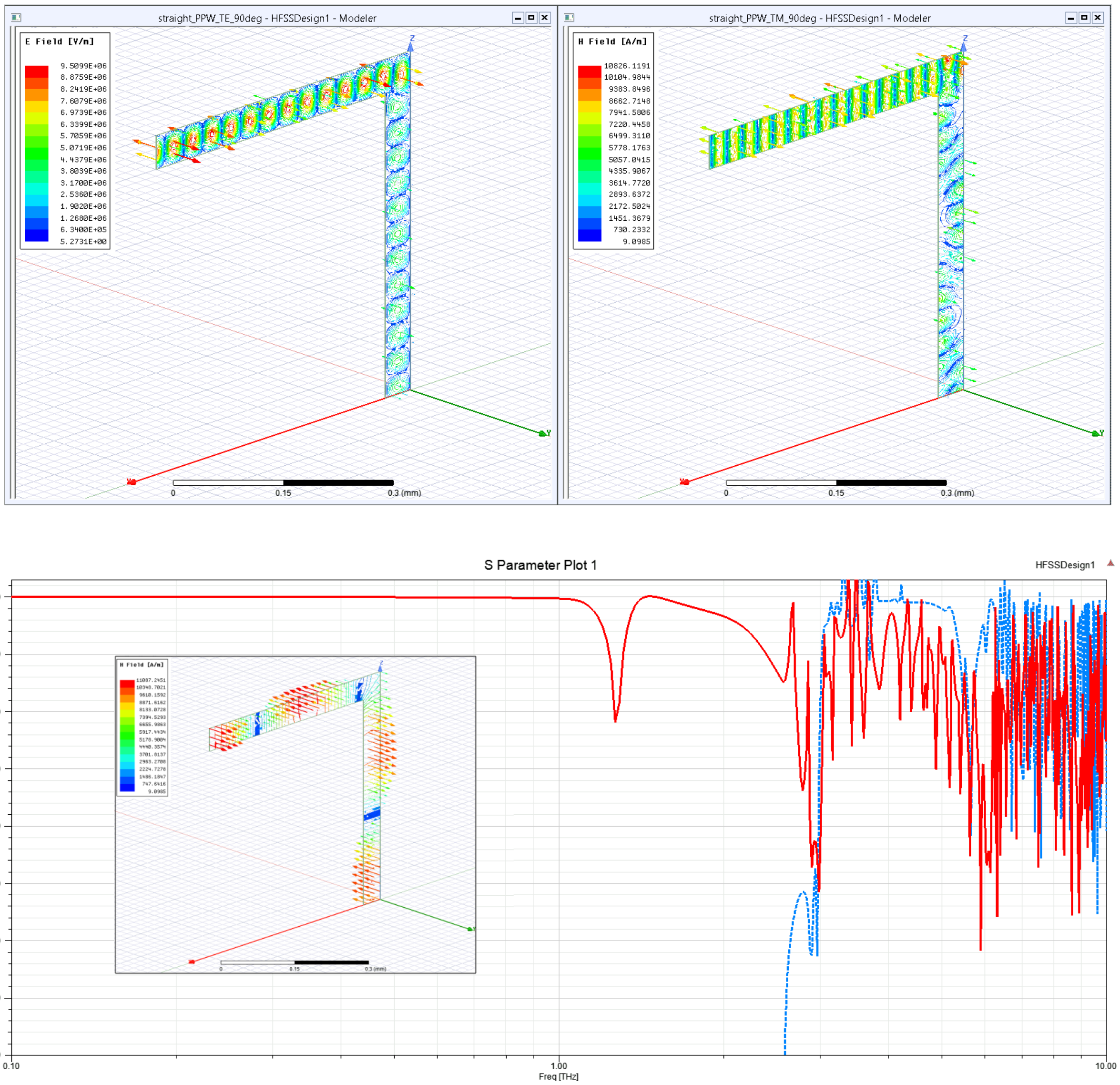


Figure 5.12: The simulation results for the TE- (top left) and the TM-mode (top right) transmissions in a parallel-plane waveguide with a 90° turn, using the same coloring scheme of Fig. 5.8. The bottom panel shows the corresponding S_{21} for the TE (blue dotted) and the TM (red solid) modes, with the inset showing the unaffected TEM field distribution.

section that is 1–2 times wider than the major wavelength to be shielded. Although we recommend implementing the cutoff frequency, we have also explained this technique would not be effective for the TEM transmission. So, this wide section helps with shielding the radiation entering the seal in TEM mode by scattering it into $m \geq 1$ TM modes, which is sensitive to the size of the waveguide and can be shielded by narrower downstream sections.

Fig. 5.13 gives an example (top-left figure) of the proposed photon shielding strategy. The structure is mechanically supported by the contacts at the entrance and the exit of the seal. We assume the contacts to be 20 μm high and 500 μm long, which sets a TE cutoff frequency corresponding to a BB temperature of $\approx$ 75 K. Immediately after the tight entrance is a 500 μm high section that scatters BB radiation higher than 350 mK into non-TEM modes. Then, two intermediate sections at 100 μm (vertical) and 200 μm (horizontal) are used to qualitatively represent typical machining results for a sliding fit. Fig. 5.13 shows a non-TEM transmission is indeed induced in the 500 μm section, while in the reference model the propagation is TEM throughout the waveguide. In the S_{21} result, the proposed design enhances the shielding below THz due to the wide section concept.

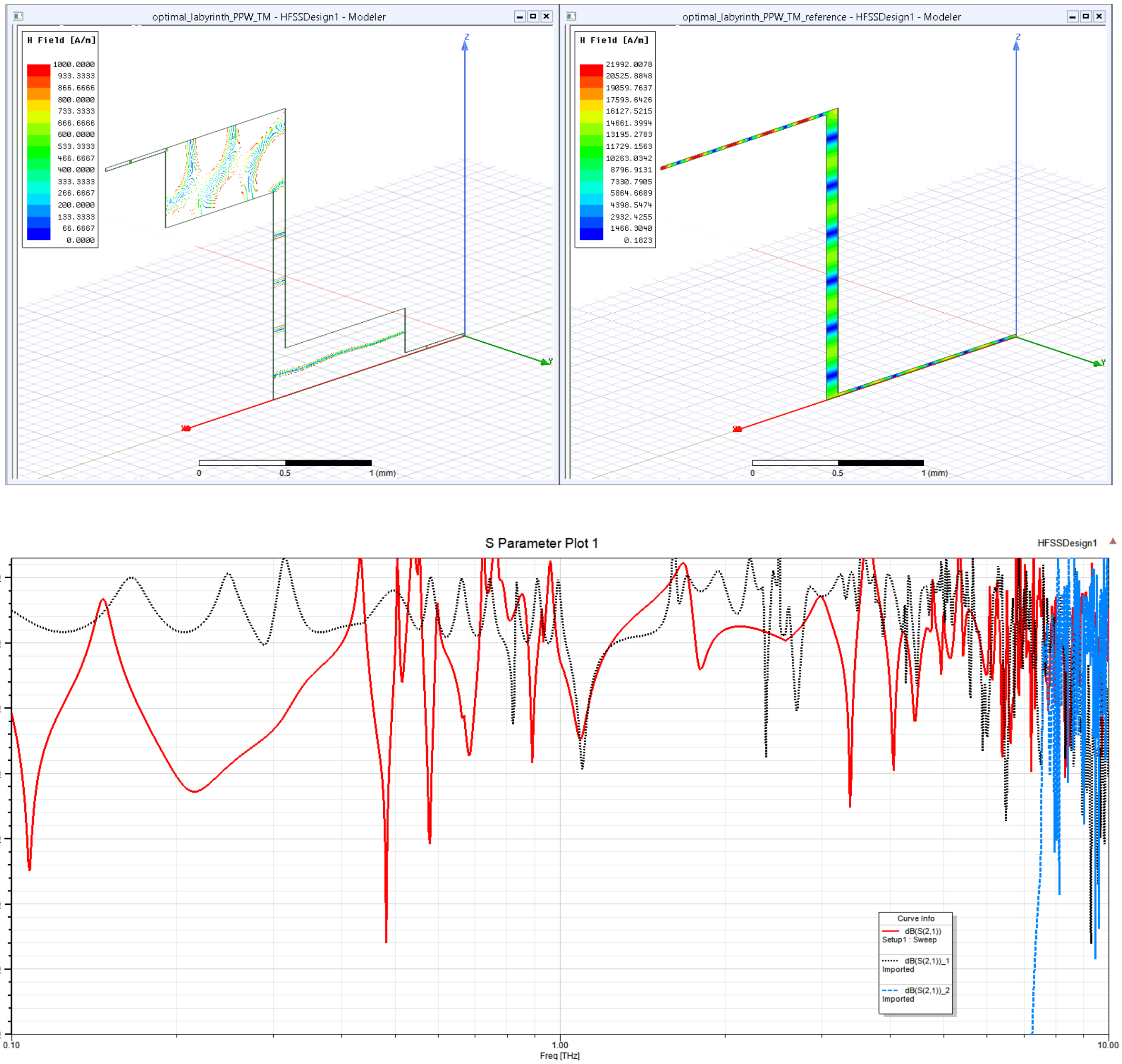


Figure 5.13: An example design (top-left) and its resulting S_{21} (bottom) based on our mechanical seal photon shielding strategy. From the entrance (left) to the exit (right) of the seal, the widths are 20 μm, 500 μm, 100 μm, 200 μm, and 20 μm. The 20 μm sections are 500 μm long, representing the designed mechanical contact supports. Other sections are 1000 μm long. The top-right figure is a reference design with the same overall geometry but is 20 μm wide throughout the waveguide, except the vertical section is assumed to be 100 μm for a typical sliding fit. Both field visualizations are for 500 GHz, corresponding to the peak emission from a $\approx$5 K BB. Both figures present the TM field, since the TE field is highly suppressed by the cutoff frequency and thus is not interesting. The S_{21} figure shows the proposed design's TM and TE transmissions by the red solid and blue dashed lines, respectively, while the black dotted line is for the reference design's TM mode. The reference design's TE transmission is similar to the example design's so is removed for clarity.

5.4 Calibration plan

We want to calibrate the free parameters in our model currently chosen based on approximations or reasonable guesses, including the structural materials' optical properties, the loss tangents for our Si and Ge crystals, and the model inputs for the shallow impurity photoionization. These parameters can be categorized into extrinsic and intrinsic properties of the detector. The former considers the photon propagation until illuminating the crystal, and the latter determines the leakage current generation from the illumination.

In this section, we design a BB photon calibration system to independently measure the former, i.e., the radiation exposure of the detectors, whose data can be used to calibrate the simulation. Once we calibrate the simulation, it can then provide predictions to be compared to the leakage current data to subsequently calibrate our leakage current generation model. We plan to perform the experiment in the NEXUS DR (Z. Hong et al., 2019) in summer–fall, 2021, and prospectively other SuperCDMS test facilities after the NEXUS experiment. After validating the system in NEXUS DR, we look forward to deploy it alongside future SuperCDMS detectors to achieve a synchronized leakage current characterization.

5.4.1 BB radiation source

5.4.1.1 Thermal-mechanical model

We build a temperature-controlled BB radiation source for calibration. It is designed to emit in our primary temperature range of interest when installed on 4 K mechanics, with a flexibility up to 20 K, and can dominate the background BB radiation from the cryostat mechanics. NEXUS DR has a 4 K payload space of $\varnothing$12" $\times$ 18" (Fig. 5.16). Assuming the material of the payload space exhibits an emissivity of unity and targeting a BB temperature of $\approx$20 K, Stefan-Boltzmann law suggests the BB can dominate the background radiation energy with a surface area much larger than 7 cm^2. This is a conservative estimate since the cryostat materials, mostly Cu, are expected to exhibit typical emissivities of $<$ 0.1. If needed, we can also install a radiation shield around the photon detector that is heat-sunk to 1 K to further reduce the 4 K background.

We build the physical BB radiation source by heating a specialized BB material with commercial heaters. We use the TK RAM material by Terahertz to approximate the BB (Terahertz, n.d.), which is a common choice of "tessellating radar absorbing materials" in the sub/mm-wave range and is available in convenient tile shapes. The mechanics the TK RAM tiles attache to is made of OFHC Cu, whose high thermal conductivity enables a responsive temperature control and an uniform BB temperature. We use the epoxy adhesive (Stycast, n.d.) commonly utilized in cryogenic applications to attach the TK RAM tiles to the Cu mechanics. It offers a good thermal link for the BB and the mechanics at cryogenic temperatures, and it is mechanically robust over thermal cycles due to its thermal expansion coefficient similar to Cu. We use carbon fiber-reinforced polymer (CFRP) to mechanically attach the above assembly to the DR's 4 K mechanics. Due to the low thermal conductivity of CFRP, the BB can be heated to target temperatures substantially higher than 4 K without conducting too much heat to the DR. CFRP also exhibits outstanding mechanical properties to allow supporting the assembly by a small quantity, which helps with reducing the heat load onto the DR. We attach commercial thermal sensors and resistive heaters to the Cu body of the BB assembly to monitor and control the BB temperature.

In order to operate the BB radiation source in practice, the CFRP "thermal break" needs to provide sufficient thermal resistance to decouple the instrument's heat dissipation. At the same time, the thermal break also needs to provide sufficient thermal conduction so the instrument can spontaneously cool to 4 K in a reasonable time via natural cooling. We aim to design a thermal break that conducts much less heat than the cryostat's cooling power for its 4 K mechanics. The requirement allows assuming the 4 K structure is held at a constant temperature (T_0) regardless of the BB temperature (T_{BB}).

We picture a slender column-shaped thermal break of a height h that raises the instrument from the cryostat 4 K floor. The geometry can be approximated by a 1D boundary value problem, because the transverse size of the slender column is small compared to its height. Along the height direction $\hat{x}$, the thermally conducted power is given by

$$\dot{Q}(x) = \frac{dQ}{dt}(x) = \kappa[T(x)]A(x)\frac{dT}{dx}, \tag{5.75}$$

where κ is the thermal conductivity as a function of T, and A is the cross section area of the column. Notice the above equation is for a general case, where the thermal conductivity can be a function of temperature that varies along x, and the cross section area can also vary along x. At equilibrium, we have

$$\dot{Q}(x) = \dot{Q}, \tag{5.76}$$

i.e., a constant throughout the thermal break, so we solve for

$$\dot{Q} = \frac{\int_{T_0}^{T_{\rm BB}} \kappa(T) dT}{\int_0^h \frac{dx}{A(x)}} \tag{5.77}$$

to determine the heat conducted to the cryostat.

It is useful to define the thermal resistance $R_{\rm th}$ based on the thermal/electric conduction analogy

$$\dot{Q} \longleftrightarrow I \tag{5.78}$$

$$\Delta T \longleftrightarrow V \tag{5.79}$$

$$R_{\rm th} \longleftrightarrow R, \tag{5.80}$$

which gives

$$\dot{Q} = \frac{\Delta T}{R_{\rm th}} \quad \rightarrow \quad R_{\rm th} = \frac{T_{\rm BB} - T_0}{(\int_{T_0}^{T_{\rm BB}} \kappa(T) dT)/(\int_0^h \frac{dx}{A(x)})}. \tag{5.81}$$

We take a rough fit to the data in (Hartwig et al., 1984; Kellaris et al., 2014; Runyan et al., 2008) and obtain

$$\kappa \approx e^{(0.12T - 3.45)} \quad [\mathrm{W\,m^{-1}\,K^{-1}}] \tag{5.82}$$

to represent the thermal conductivity for CFRP, which is accurate to a factor of a few due to the data variation. Letting $T_0 = 4$ K, $T_{\rm BB} = 20$ K, and using Eq. (5.82), Eq. (5.81) yield

$$R_{\rm th} = 2.49\ [\mathrm{K/W \cdot m}] \times \left(\frac{h}{A}\right). \tag{5.83}$$

This result is useful since it is directly related to the mechanical design parameters h and A.

We write the energy conservation condition at equilibrium with $R_{\rm th}$:

$$P = I^2 R = P_B + \dot{Q} = \epsilon \sigma T_{\rm BB}^4 S + \frac{T_{\rm BB} - T_0}{R_{\rm th}}, \tag{5.84}$$

where P is the BB heating power supplied through I and R, the DC current and the resistance of the heater, respectively, and we assume the Joule heating is dissipated by the conductive and the radiative energy losses $\dot{Q}$ and P_B, respectively, and S is the BB's surface area. We assume

$$S \approx 100 \quad [\mathrm{cm}^2] \tag{5.85}$$

based on previous argument that we need a surface area much larger than 7 cm^2. Taking $\epsilon = 1.0$ for the BB, we obtain

$$P_B \approx 90 \quad [\mu\mathrm{W}], \tag{5.86}$$

a radiation dissipation that is negligible compared to the conductive dissipation by Eq. (5.83) at realistic $(A/h) \sim$ cm and $\Delta T \sim$ K. Eq. (5.84) indicates the driving current is in a reasonable range of $\mathcal{O}(1)$ mA, e.g., 2.5 mA for a 10 kΩ heater.

Next, we calculate the thermal time constant of the CFRP-isolated BB to ensure the instrument can cool back to 4 K in a reasonable amount of time. At every instant,

$$\dot{Q}dt = c_V \rho V dT, \tag{5.87}$$

where c_V is the BB's heat capacity[10], ρ is the density, and V is the volume. Notice that, before reaching equilibrium, $\dot{Q}$ can be a function of x along the CFRP column, so for simplicity, in the following we only write these quantities to represent the BB part, i.e., at $x = h$. Combining Eq. (5.87) and Eq. (5.77), we obtain the general formula

$$\frac{A}{\rho V h} dt = \frac{c_V[T(t)]}{\int_{T_0}^{T(t)} \kappa(T')dT'} dT \tag{5.88}$$

for the cooling time constant.

This differential-integral equation is difficult to solve analytically because both c_V and κ are the functions of temperature that implicitly depend on the instantaneous time t. Alternatively, we can estimate the cooling time constant assuming every location in the CFRP column is at T_{BB}, knowing the thermal conductivity varies within order-of-magnitude in T_0 to T_{BB}. $T(t = 0) = T_{\mathrm{BB}}$ reduces Eq. (5.88) to

$$\frac{dT}{T_{\mathrm{BB}} - T_0} = \frac{dt}{\rho V (c_V R_{\mathrm{th}})|_{T_{\mathrm{BB}}}}, \tag{5.89}$$

where c_V and R_{th} are constants due to the temperature assumption. We solve for

$$\Delta T(t \ll \tau) \sim e^{-t/\tau} \tag{5.90}$$

and find

$$\tau = c_V \rho V R_{\mathrm{th}}. \tag{5.91}$$

To put the above result into perspective, we assume the copper structure of the BB assembly is made of the thinnest mechanically sound plates, 0.5 cm in thickness, to

[10] Using subscript V for "constant volume" and to distinguish it from the speed of light.

minimize the thermal time constant. These plates are designed to have a surface area of 100 cm^2 for attaching the BB tiles of the same area, thus $V \approx 50\ \mathrm{cm}^3$. According to (Simon et al., 1992), OHFC copper has $c_V \approx 7.51\ [\mathrm{J} \cdot \mathrm{kg}^{-1} \cdot \mathrm{K}^{-1}]$ at 20 K and $\rho \approx 8.96\ [\mathrm{g} \cdot \mathrm{cm}^{-3}]$. Assuming $(A/h) \sim \mathrm{cm}$, we find

$$\tau \approx 840 \quad [\text{sec.}], \tag{5.92}$$

which is a manageable time. In practice, we would always wait for a few times of the time constant to ensure the instrument has fully cooled to the base temperature, so the above calculation is sufficiently accurate for us to conclude that the BB can reliably cool to 4 K spontaneously within 2 hours.

5.4.1.2 Detailed design, construction, and testing

We design and fabricate our BB radiation source shown in Fig. 5.14 based on the above modeling work; detailed mechanical drawings and `.STL` files are found on (Chang, n.d.). In addition to the properties calculated above, we look for a practical design that is compatible with the existing mechanical and electronic infrastructures in NEXUS DR. Hopefully the design is also relatively easy to be replicated and installed in other SuperCDMS test facilities.

We expect the entire assembly to be installed to have a surface area of 150–200 cm^2 based on the planned 100 cm^2 BB surface, which should yield a realistic instrument with 5–10-cm side widths. Aiming for an ideal scenario that the BB source can be run with a full SuperCDMS tower installed in NEXUS DR, we conclude from the constrained space usage that the only available space for the BB is between the DR's 4 K electronic feedthrough cap and the HEMT cards; Fig. 5.14 shows this space. We then conclude, for this space, the only location to mount the BB radiation source is the lid of the 4 K feedthrough cap, because the cap sidewalls are mostly occupied by the feedthrough PCBs, and the tower mechanics was designed for sub-mW unanticipated heat load, which is much lower than the BB source's 10^2-mW disspation. After several design iterations, we also realized an instrument with comparable width and height, rather than extended in one direction like a disc, would utilize the limited space more efficiently while offering the intended BB surface area.

Fig. 5.14 shows the final design. The main BB is a hexagonal prism chosen to approximate a cylinder but more easily fabricated and assembled in practice. The prism design enables fabricating the instrument with stock plates to be a lightweight hollow structure, and the flat surfaces also make attaching the BB tiles easier. The diameter and the height of the hexagonal prism are ≈3", yielding a surface area of ≈180 cm^2. It is supported by eight ⌀0.48 mm CFRP cylinders with a total cross section area of 0.3 cm^2. The CFRP cylinders hold the BB assembly at 0.6 cm above the instrument's circular Al mounting pedestal, so $(A/h) = 0.5$ cm. We choose Al for the pedestal for better mechanical robustness and machinability relative to Cu, knowing the added thermal resistance is negligible compared to CFRP's and thus does not affect our thermal model. The final design allows a clearance of ≈13 mm, much larger than the wavelengths of interest, between the BB surfaces and their parallel HEMT cards, which can be utilized for future additions of sub/mm-wave filtering materials to restrict the spectrum for calibration.

We prepared the BB surface by first assembling small TK RAM tiles into a wide plate and then laser cutting the plate into the hexagonal prism's top and sidewall shapes. Some of the pyramids along the cuts, i.e., on the edges of the cut shapes, were melted as shown in Fig. 5.15 due to high temperature. We attached these shaped BB tile plates and the CFRP cylinders to the metallic parts with Stycast to complete the assembly. The fabricated final product weights 550±5 g for the BB assembly, ≈50 g non-copper, i.e., BB tiles and Stycast. The obtained BB surface area, 160±1 cm^2, is slightly smaller than anticipated due to leaving edge space for screw holes. The length of CFRP thermal breaks, 0.64 cm, is also slightly higher than designed due to excess Stycast. Based on our thermal model, we expect the fabricated BB radiation source to exhibit

$$
\begin{array}{rcll}
P_B & \approx & 144 & [\mu\mathrm{W}] \\
h/A & \approx & 0.43 & [\mathrm{cm}^{-1}] \\
R_{\mathrm{th}} & \approx & 5.76 \times 10^2 & [\mathrm{K/W}] \\
\dot{Q} & \approx & 28 & [\mathrm{mW}] \\
\tau & \approx & 2.4 \times 10^3 & [\mathrm{sec.}].
\end{array}
$$

We tested the design in the "BEMCO" 4 K cryostat in Cahill, Caltech. Fig. 5.15 shows the test wiring and the result. We attached two Lake Shore silicon diode thermometers to the instrument's Al pedestal and the side of the top BB as shown in the photo. We used threaded hole and screw to firmly attach the thermometer to the Al pedestal. However, we were unable to produce a threaded hole in the soft TK RAM material for screwing. So, we bolted the thermometer to a 1×2 cm^2 copper sheet and then fixed the copper sheet to the flat side surface of the BB with

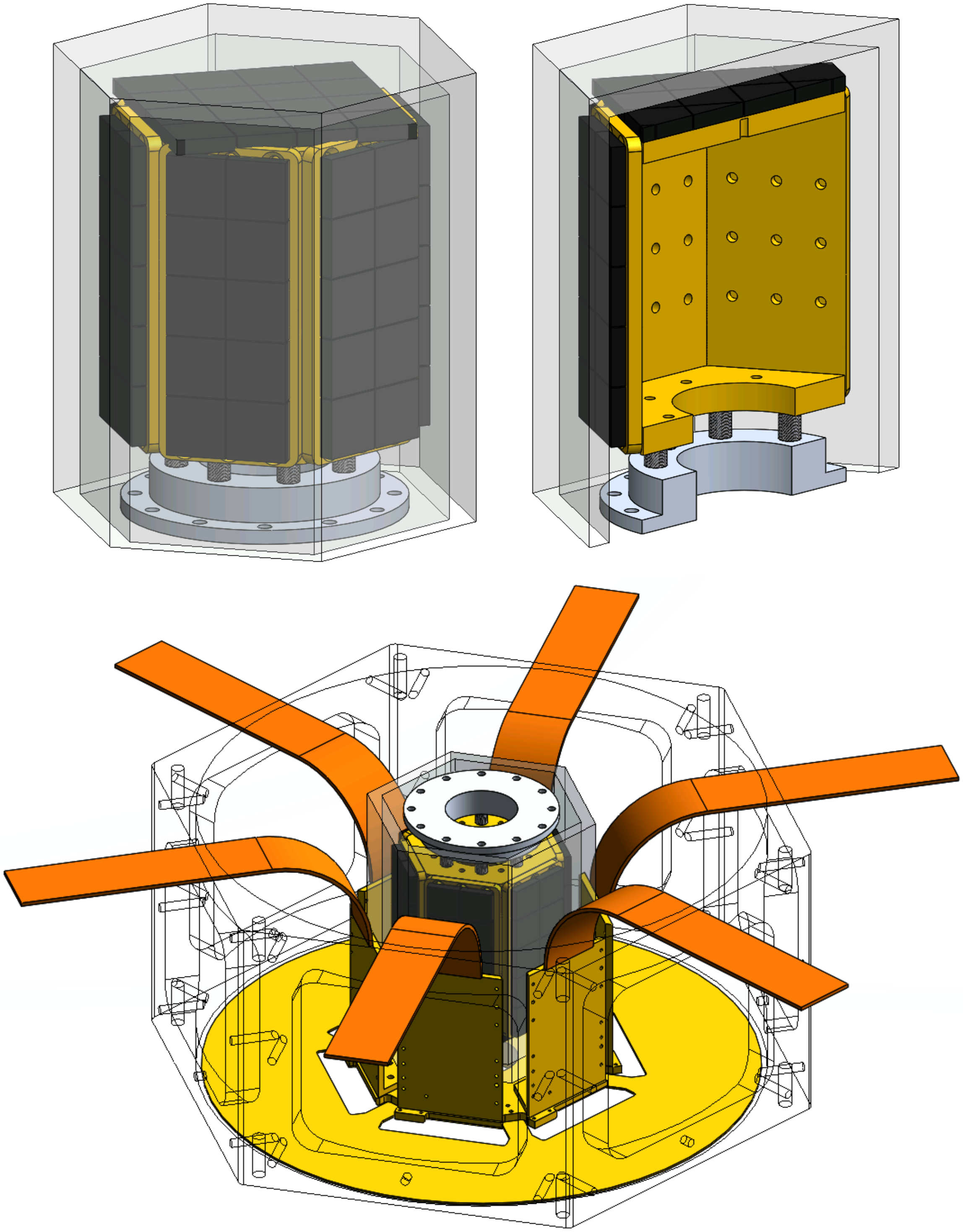


Figure 5.14: Top: The BB radiation source. Bottom: The full assembly with the NEXUS 4 K electronic feedthrough cap (wire frame) and the SuperCDMS SNOLAB tower's 4 K structure. Copper, aluminum, and Kapton are colored in yellow, gray, and orange, respectively. Black square tiles and rods are the TK RAM tiles and the CFRP columns. The transparent hexagonal prism covering the instrument represents the space reserved for sub/mm-wave filtering materials. The blind hole arrays on the inside of the instrument's copper body are for attaching thermometers and heaters, whose wiring runs through the circular cutout at the bottom of the hexagonal prism and the gaps between the CFRP columns. The bottom assembly is oriented upside-down for visual clarity, so in reality the instrument stands upright on the feedthrough cap rather than suspended upside-down as in the figure.

VGE-7031 varnish and Kapton tape. To control the BB temperature, we attached two 100 Ω resistive heaters, connected in parallel, to the inner sides of two opposite copper plates in the hexagonal prism.

Fig. 5.15 shows the BB cooled much more slowly than the Al pedestal as expected. However, it also shows the BB approached an equilibrium temperature of about 8 K but not 4 K. To investigate the issue, we applied 1 V to the heaters for an hour and then let the structure cool spontaneously to the base temperature. We obtained rough thermal time constants of 1.7 hour and 2.0 hour for the heating and the cooling periods, respective, by fitting simple exponential functions to the data as shown in the figure. These time constants are higher than designed by a factor of $\approx$3. In the presence of a typical lead resistance of 10s Ω, we estimate that a heating power of $\approx$10 mW were applied to the BB. Such a power is within order of magnitude but lower than our model prediction for raising the BB temperature to the observed 20–22 K.

When we opened the cryostat, the copper sheet connecting the thermometer had almost detached from the BB tile and was barely touching with the help of the Kapton tape. The result offers a plausible explanation to the BB's high equilibrium temperature, the tripled thermal time constant, and the lower-than-expected heating power to achieve the target temperature. Despite the unfortunate outcome, we still have demonstrated the instrument performs acceptably well for the intended calibration experiment. We have planned for another test with improved thermometer mounting, but the test acquired a low priority due to that the utility of the instrument has been verified.

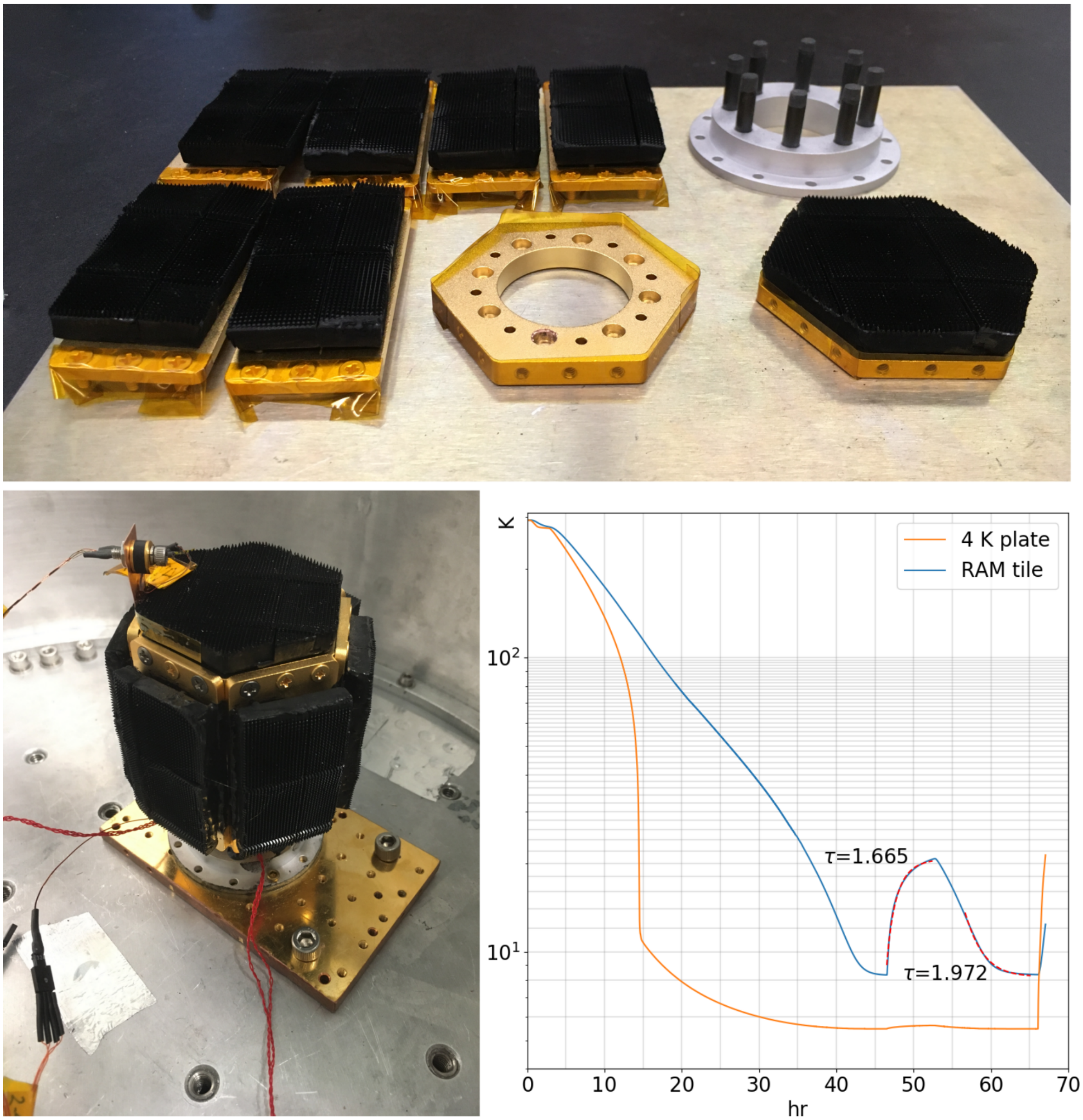


Figure 5.15: The fabricated BB radiation source before (top) and after (bottom) assembling. The top photo shows the parts during Stycast curing, where we used Kapton tapes to protect the screw holes. These tapes were removed before assembling. The bottom photo shows the assembled instrument with the test wiring ready to be run in the BEMCO cryostat. The red wires connected the heaters; the 4-lead jack connected the thermometer on the Al pedestal; the 2-lead wire connected the BB thermometer on the top, which was attached with a small copper sheet, GE varnish, and Kapton tape. In the bottom right graph, the orange and the blue temperature profiles were taken by the the Al pedestal and the BB thermometers, respectively. The red dashed curves represent rough exponential fits to the heating and the cooling data, with the extracted thermal time constants in the unit of hour noted on the side.

5.4.2 NEXUS DR overview

The NEXUS cryostat is a CryoConcept HEXA 200 dry DR, utilizing a pulse tube cooler for 50-K and 4-K $^{3/4}$He-cryogen pre-cooling, with >250 μW and 10 μW designed cooling powers at 100 mK and 20 mK MC temperatures, respectively, and its 4-K plate is expected to provide a >300 mW cooling power. Fig. 5.16 shows the cross-section side view of the DR: A 15"-thick (vertical) lead shield is installed above the detectors between the payload chamber and the 70-mK cold plate, heat-sunk to 1 K via copper columns through cutouts on the cold plate, to provide a near-vertical energetic particle shielding for the payloads. The payload chamber is roughly a $\varnothing$12" $\times$ 18" cylindrical space formed by the 1-K can, measured from the bottom of the 1-K can to the MC-temperature plate below the lead shield. The mechanical structure below the lead shield is primarily designed to be able to host a full SuperCDMS SNOLAB detector tower (Agnese et al., 2017). When hosting the SNOLAB tower, the tower is flipped upside-down relative to the orientation for the SNOLAB "SNOBOX" cryostat, so the detector stack is closer to the lead shield for a wider angle of shielding, and the SQUID cards are below the detectors, and then the 4 K HEMT cards are at the bottom (Z. Hong et al., 2019). In this configuration, all the bottom lids of the cans at and below 4 K are removed, so different temperature stages of the tower can be mounted to the corresponding refrigerator plates. To provide the HEMT cards with electronic connection, a hexagonal cap with 3" $\times$ 3.5" cutouts on all sidewalls is assembled to the bottom of the 4 K can, surrounding the HEMT cards with PCB feedthrough boards covering the cutouts.

For the BB radiation environment modeling, we are particularly interested in the completeness and accuracy in modeling the mechanical openings and the realistic radiation budget, so we require all the cutouts and screw holes on the standard 4-K plate above the lead shield be carefully sealed by copper foils/tapes and screws during BB calibration experiments, so to ensure good blocking of photons from higher temperatures. According to Fig. 5.5, polymer materials such as the PCBs are generally semi-transparent in the trans-mm wave range, so we may also expect the photons from higher temperatures to travel through the space between the 4-K and 50-K cans, reaching the bottom of the cans, and subsequently enter the detector chamber through the feedthrough PCBs. In practice, however, the NEXUS electronic feedthrough boards are all coated with full copper ground planes at an 1-oz thickness ($\approx$35 μm), which is much thicker than the nm-scale skin depths for trans-mm waves (Fig. 5.19). We therefore assume that the 4-K space, from the foil-sealed 4-K plate on top to the electronic feedthrough cap at the bottom, is a

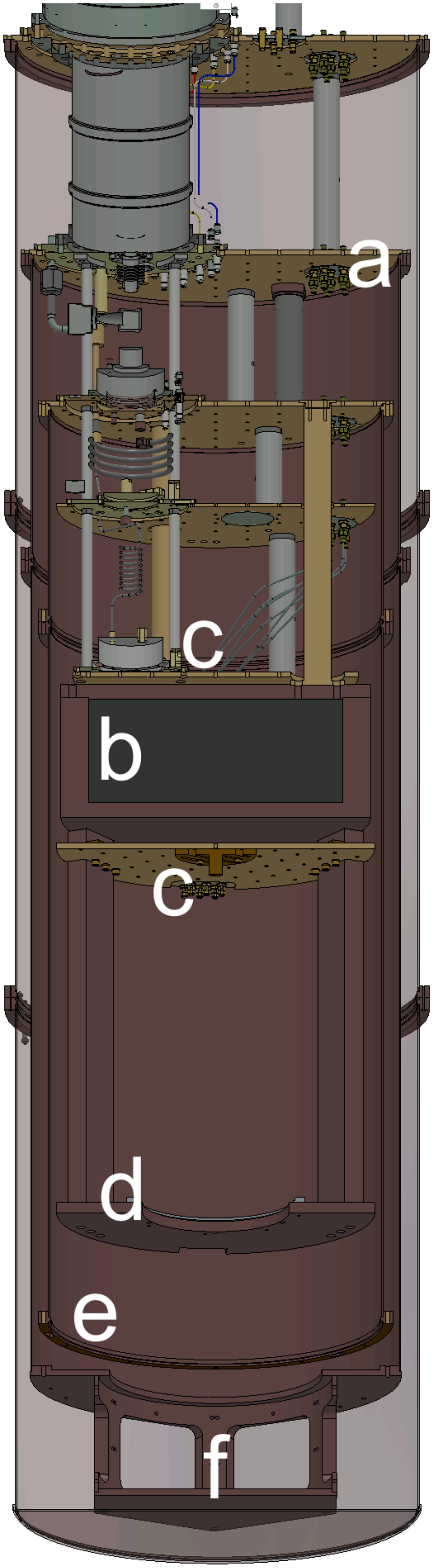


Figure 5.16: NEXUS DR side view, with the 50-K can drawn semi-transparent and outer structures suppressed for visual clarity. a) Upper-refrigerator 4-K plate. b) 15" lead shield, heat-sunk to 1-K plate via the attached copper columns through the cutouts in the MC-temperature plate above. c) MC-temperature nominal (upper) and RF-payload (lower) stages. d) SuperCDMS HVeV and BB photon study detector stage, extended from c) through the cutouts on the side of the 1 K lead shield so to avoid thermal contact. e) Lid-less 1-K can bottom. f) Electronic feedthrough cap attached to 4-K can bottom. The non-transparent region in this figure represents the photon-hermetic space for our BB photon study.

closed space, so we only need to construct this space in our simulation. We should note, this assumption is only valid when the radiation onto the detector is reliably dominated by sources inside the 4-K space, e.g., by our BB radiation calibration source. Since in practice there will likely be small unconverted regions in the copper ground planes of the feedthrough PCBs, e.g., VIAs, or imperfect tape covering for the 4 K refrigerator plate, the outside radiation could still penetrate the assumed isolated 4 K space at a suppressed probability. Based on the T^4 BB radiation power scaling (Eq. (5.1)) and knowing that the surface areas of the 4 K and the 50 K mechanics are comparable, in the absence of the BB radiation calibration source, these radiation-permitting holes only need to permit $\sim 10^{-5}$ of the outside radiation to invalidate the hermetic 4 K space assumption. If found necessary in the future, it would be helpful for future works to update the aforementioned details, such as the not yet finalized feedthrough PCB design and the 4 K payload mounting/screw hole blocking situation, into the simulation for a more accurate result.

5.4.3 TES BB photon detector

We have introduced the following physics components in our BB-photon leakage current generation model:[11] 1) BB photon generation and particle-like propagation in the cryogenic chambers, simulated by Geant4, 2) bounded narrow structure EM wave-like propagation, simulated by HFSS, and 3) photoionization in the detector substrates, modeled by the $1/r^4$ atomic-polarization binding potential theory. To validate such a complex model, especially when the photoionization part is partially phenomenological and has internal degeneracies that need to be tuned, we propose the following plan to independently validate the model components:

1. Isolate the particle-like propagation from the full nominal photon propagation, by first testing a mechanical structure with intentionally created large apertures, e.g., remove detector housing coverage or chamber cans. The test configuration should ensure that the bounded structure effect (Tab. 5.1) is subdominant to the detector radiation exposure, so one can fully attribute the result to particle-like propagation and fine-tune the Geant4 simulation accordingly. We note that, in this test, the radiation exposure is expected to be too high for the DM detectors and therefore requires a dedicated BB radiation detector, which will be discussed next.

2. Incrementally include the previously removed sealing and bounded mechan-

[11] Corresponding to the same bullet points below for the validation plan.

ical structures. The difference in the measured radiation load is then due to the wave-like propagation and can be used for adjusting the HFSS model.

3. Deploy DM detectors alongside the BB radiation detector, so they are exposed to the same radiation environment for a direct comparison. As the radiation environment is well calibrated and understood in prior steps, the leakage current rates obtained in the DM detectors provide the information for tuning the photoionization model based on independently constrained photon flux inputs.

For the last step, a candidate leakage current detector is the HVeV detector(s) already deployed at NEXUS, which is immediately compatible with our BB radiation calibration system. It is also possible to perform the experiment with future SNOLAB towers, since our calibration system is designed to be compatible with the full-tower mechanical design. We should note that, to robustly attribute the leakage current signal to sub/mm photons, further developments for experimental and analysis techniques are needed but not included in this thesis. For example, it is not yet clear if operating the HVeV detector at a high voltage or 0 V is the optimal choice, given that under high voltage it might be difficult to distinguish backgrounds due to other sources from BB photon events. One might also need to consider the threshold photon energy for the proposed photoionization mechanism and control the test radiation spectrum accordingly, which is also not yet designed, except that we do reserve the space for installing IR-filtering materials around the BB radiation source.

We approach the design of the dedicated BB radiation detector by minimally modifying existing SuperCDMS TESs, so the following modeling and fabrication works can be benefited by the already established knowledge as much as possible. Contrasting with SuperCDMS' QET phonon sensor design, which are equipped with large-area Al phonon absorbers that are highly reflective in the BB radiation range, we plan to use the W TES film to directly absorb the BB radiation photons. Under the anticipated high photon flux, the W TES is expected to be driven to an equilibrium exhibiting a parasitic bias power higher than normal SuperCDMS detectors. In this case, the photon flux is manifested by a fully DC signal, in other words a trace baseline variation. The benefits of this technique include: First, we can suppress the noise and therefore improve the energy resolution by long averaging, in practice up to the system drift limit. Note that it does not guarantee a better energy resolution then pulse detectors, due to the generally higher noise at low frequencies than in the

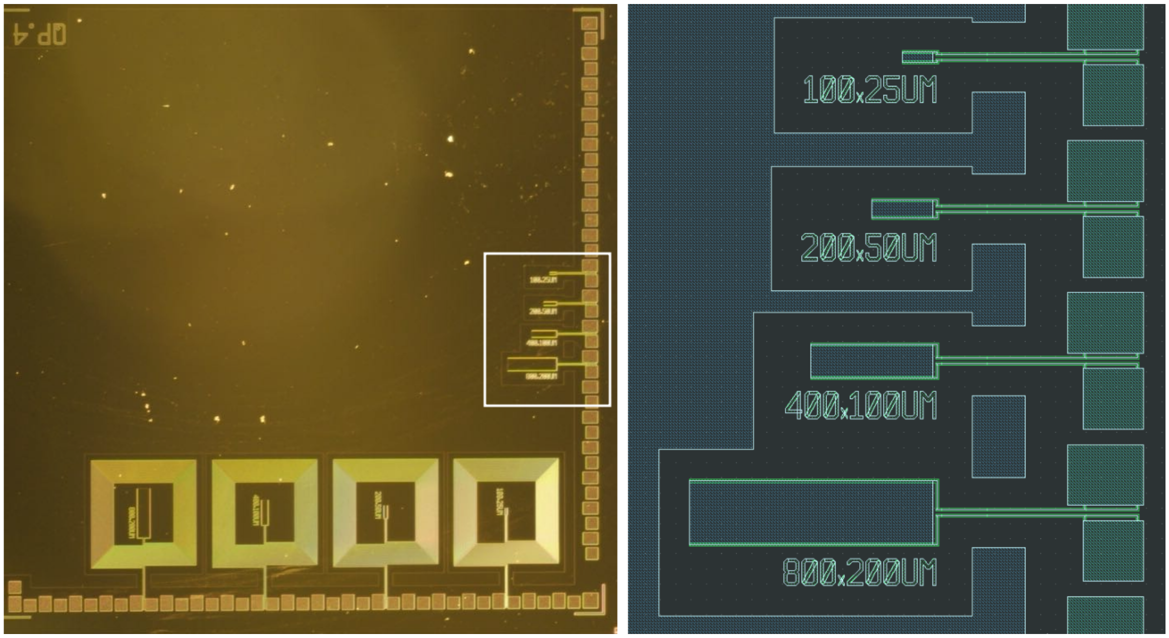


Figure 5.17: The photo (left) for the SuperCDMS bias power study R&D TESs, with the detailed layout for the highlighted region in the photo shown on the right. The black, green, and blue regions in the layout represent substrate surface, Al, and W, respectively. The corresponding dimensions of the TESs are also noted in the layout. Detailed description may be found in the corresponding text. Photo provided by M. Pyle.

pulse band. Second, we can take advantage of the already established TES modeling and fabrication knowledge, existing hardware and software infrastructures, and possibly share substrates in SuperCDMS detector fabrications.

After surveying the available SuperCDMS detector designs, we identify the bias power R&D TESs as our photon detector design baseline. Shown in Fig. 5.17, the design consists of a series of W rectangles at the same aspect ratio of 1:4 that are connected along the long edges with simple Al traces. These TESs are designed in the simplest possible form, while the aspect ratio is chosen to be close to optimal based on previously obtained normal-state W film resistivity and the theoretical bias power model in (Pyle, 2012). Due to the same aspect radio, therefore the same resistance, these rectangular TESs of different sizes are expected to yield the same energy resolution in the absence of parasitic bias power. Therefore, by comparing the sensitivities of them, one can infer the parasitic power as a function of TES size. Since we expect the BB radiation load to scale with the TES area, the bias power R&D TES is the best design for our purpose. In the following subsections, we first develop an analytical model for the TES' thin-film sub/mm-wave radiation

Parameter	Value	Unit	Note
Material	W		
Thickness	20	nm	
Aspect ratio	1:4		Biased along the short side
Geometry	25×100 50×200 100×400 200×800	μm^2	
T_c	50–52	mK	Varies between devices, will measure.
R_n	0.25–0.75	Ω	Varies between devices, will measure.

Table 5.4: The TES design parameters for the SuperCDMS bias power R&D detector. R_n stands for normal-state resistance.

response, and then based on the model, we propose a "meshed" photon detector design modified from the bias power R&D TES, which is optimized for the radiation absorptance in our frequency range of interest.

5.4.3.1 EM thin-film regime

Tab. 5.4 lists the nominal values for the critical parameters of our W TES, while Tab. 5.5 provides other CDMS device design and fabrication parameters that are useful for below discussions.

We first determine the bulk conductivity σ for the W film to be used in later derivations. We relate the nominal normal-state TES resistance R_n in Tab. 5.4 to the conductivity by

$$R_n = R_s \frac{l}{w} = \rho \frac{l}{wd} = \frac{l}{\sigma wd}, \tag{5.93}$$

where R_s is the sheet resistance of the film, l, w, and d are the length (parallel to the bias current), width (transverse to the bias current), and thickness of the rectangle TES, respectively. Using the aspect ratio l:w in Tab. 5.4 and the film thickness in Tab. 5.5, the above relation gives

$$\begin{aligned} \sigma &= \frac{1}{0.25\text{–}0.75\ [\Omega]} \times \frac{1}{4} \times \frac{1}{40\ [\text{nm}]} \\ &\approx (0.83 - 2.50) \times 10^7 \quad [1/\Omega\text{m}]. \end{aligned} \tag{5.94}$$

We typically can only control the fabricated conductivity to within a factor of a few, therefore we assume

$$\sigma = 1 \times 10^7\ [1/\Omega\text{m}] \tag{5.95}$$

for the following derivations. However, once fabricated, a TES film's conductivity is fixed and can be measured accurately, so we remind future readers to update their calculations when the data are available.

Next, we determine whether the W film should be modeled as an infinitely thick, impenetrable material or a finite-thickness film for the BB photons. This can be determined by comparing the film thickness and the classical metal EM attenuation "skin depth," which is given by (Jackson, 1998)

$$\delta = \sqrt{\frac{2\rho}{\omega\mu}} \times \sqrt{\sqrt{1 + (\rho\omega\epsilon)^2} + \rho\omega\epsilon} \tag{5.96}$$

$$\rightarrow \begin{cases} \sqrt{\dfrac{2\rho}{\omega\mu}} & , \quad \rho\omega\epsilon \ll 1 \\ 2\rho\sqrt{\dfrac{\epsilon}{\mu}} & , \quad \rho\omega\epsilon \gg 1. \end{cases} \tag{5.97}$$

The first equation is the general formula, and the second equation gives the approximations for the low (top)- and the high (bottom)-frequency regimes.

For W and all other materials relevant to our TES model, we assume $\mu = \mu_0$. Typically for good conductors at low enough frequencies, one has $\rho = 1/\sigma$ and approximates the material's permittivity by the vacuum permittivity (Jackson, 1998). However, we can substitute $\epsilon \approx \epsilon_0$ and Eq. (5.95) into the skin depth formula above and find the low frequency regime holds up to 10^{17} Hz, which corresponds to an energy (10^{17} Hz $\approx$ 400 eV) much exceeding the energy needed to liberate lattice-bound electrons. This result contradicts the wave model and indicates we can not simply adopt the approximation. Empirically, the approximation is known to be reliable up to

$$\omega \ll \omega_d < \omega_p \sim O(10^2) \quad \text{[THz]} \tag{5.98}$$

for good conductors (Moroz et al., 2019), where ω_d and ω_p are the Drude-model damping and plasma frequencies, respectively. It is worth noting that, for good conductors, ω_d is typically two orders-of-magnitude lower than ω_p, thus applying a more stringent limit at a few THz for the low-frequency regime.

Our trans-mm (sub- to a few-THz) range of interest is close to typical ω_d and suggests that, to reliably estimate the skin depth, we should consider a full frequency-dependent complex permittivity for W rather than adopt the simple low-frequency approximation. Before a measurement is available, we find the best alternative from

Parameter	Value	Unit	Note
Substrate material	Si, Ge		[111], high resistivity
Substrate thickness	525	μm	
Process recipe	"Standard CDMS HV"		
Thickness (top to bottom):			
Al	600	nm	May very depending on
W	40	nm	up-to-date adjustments
Amorphous Si	40	nm	
Si, Ge substrate	525	μm	

Table 5.5: The design and fabrication parameters for SuperCDMS HV detectors, which are also adapted by our BB photon TES detector.

(Ordal et al., 1985), where the Drude model permittivity

$$\epsilon_r = \epsilon_1 + i\epsilon_2 = n^2 = 1 - \frac{\omega_p^2}{\omega^2 + i\omega\omega_d} \tag{5.99}$$

was utilized to extract

$$\begin{cases} \omega_p = & 2\pi \cdot 1551 \quad [\mathrm{THz}] \\ \omega_d = & 2\pi \cdot 14.61 \quad [\mathrm{THz}], \end{cases} \tag{5.100}$$

where ϵ_1 and ϵ_2 are the real and the imaginary parts of the Drude permittivity, respectively, and n is the generalized refractive index.

Fig. 5.18 shows the W permittivity from (Ordal et al., 1985). For frequencies much lower than ω_d, Eq. (5.99) and Eq. (5.100) suggest

$$-\epsilon_1 \longrightarrow \frac{\omega_p^2}{\omega_d^2} \cancel{-1} \approx 1.13 \times 10^4 \quad , \ \omega \ll \omega_d, \tag{5.101}$$

while the full complex permittivity is predominately imaginary. As expected, Fig. 5.18 shows our main frequency range of interest, 10^2–10^4 GHz, is located just below the kinks of the curves. The result supports our choice of including the full frequency dependence of the permittivity and indicates that measurements are indeed necessary for determining the kink locations that can significantly affect later model building.

We substitute the W permittivity from (Ordal et al., 1985) into Eg. (5.96) to determine the skin depth as a function of the BB radiation frequency, shown in Fig. 5.19. Comparing the result to our nominal TES film thickness, we find the 40 nm W film is only truly "thick" at $> \mathcal{O}(10)$ THz, corresponding to >100s K. We therefore conclude, in order to correctly model the trans-mm wave response for our

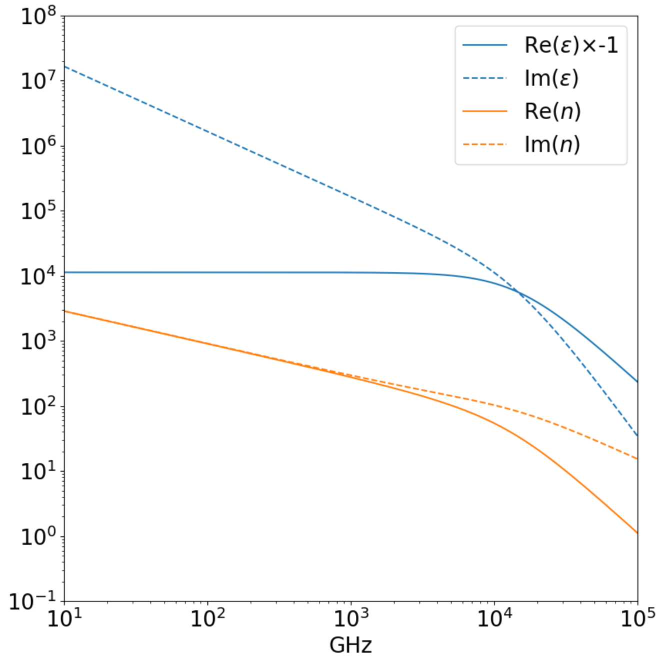


Figure 5.18: The Drude model permittivity (ϵ_r, blue) and its corresponding refractive index (n, orange) calculated using the plasma and damping frequencies obtained by (Ordal et al., 1985). The complex values are presented by their real (Re, solid) and imaginary (Im, dashed) parts as labeled in the legend, with the real part of the permittivity, which is negative in the selected frequency range, multiplied by -1 for presentation. More details are found in the corresponding text.

W TES, the film thickness and the substrate-W interface behind the TES need to be considered.

While we have introduced the conductivity and the permittivity as independent quantities, (Ordal et al., 1985) points out that the quantities are in fact related and can be inferred from each other at DC by

$$\epsilon_{r,\mathrm{DC}} \approx 1 + \frac{i\sigma}{\omega\epsilon_0}. \tag{5.102}$$

We take Eq. (5.99) for

$$\lim_{\omega\to 0}\left(1 - \frac{\omega_p^2}{\omega^2 + i\omega\omega_d}\right) = \epsilon_{r,\mathrm{DC}} \tag{5.103}$$

and find

$$\sigma_{\mathrm{opt}} = \frac{\omega_p^2}{\omega_d}\epsilon_0, \tag{5.104}$$

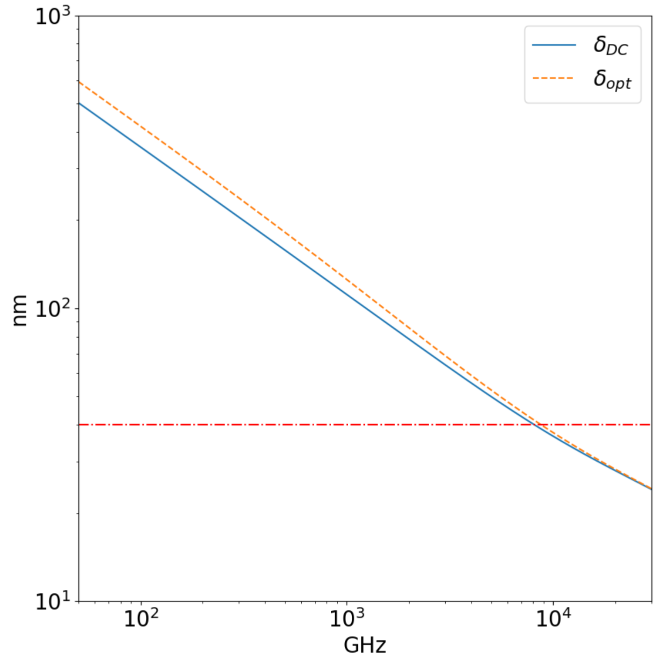


Figure 5.19: The skin depths determined with the damping frequency from (Ordal et al., 1985) (orange dashed) and the assumed DC conductivity (blue solid). The horizontal red dash-dotted line marks the W film thickness for reference, indicating at most lower frequencies in this figure, roughly corresponding to 1–100-K BB radiation, the film is thinner than the skin depths. More details are found in the corresponding text.

where we use the subscript "opt" following (Ordal et al., 1985)'s nomenclature for the high-frequency "optical" conductivity. Substituting (Ordal et al., 1985)'s data (Eq. (5.100)) into Eq. (5.104), we obtain

$$\sigma_{\text{opt}} = 0.92 \times 10^7 \quad [1/\Omega\text{m}], \tag{5.105}$$

which is consistent with the typical conductivities measured for SuperCDMS W films (Eq. (5.94)).

Since it is the DC conductivity that we plan to measure, we decide to reconcile the deduced σ_{opt} value above and σ_{DC}, whether an assumption for now or an actual measurement later, by adjusting the damping frequency while keeping the plasma frequency (Ordal et al., 1985) obtained. We reassign

$$\omega_d = \frac{\omega_p^2}{\sigma_{\text{DC}}} \epsilon_0 \tag{5.106}$$

for our W films. The corresponding adjustment to the skin depth is also provided in Fig. 5.19. We choose to adjust ω_d and keep ω_p based on the fact that the damping frequency is directly related to the electron scattering rate, which is manifested by the DC conductivity. This scattering is generally dominated by the material quality that varies appreciably sample-to-sample, while the plasma frequency describes the release of the electrons from their potential wells, which is predominately due to the much less variable intrinsic (atomic) properties of the material. In summary, we adopt (Ordal et al., 1985)'s methodology, plasma frequency data, and attribute the DC conductivity difference to having independent W samples.

5.4.3.2 HFSS Hagen-Rubens solution

We provide in the following a short analysis for the thick-film model, even though we have just argued it is incomplete for our application. We find it instrumental for building an intuition for the complete model and, as it turns out, understanding HFSS' algorithm for simulating thin films.

We start from the general Fresnel's equations of reflectivity (Jackson, 1998)

$$\begin{cases} r_s = \dfrac{n_t \cos\theta_i - n_i \cos\theta_t}{n_t \cos\theta_i + n_i \cos\theta_t} \\[2ex] r_p = \dfrac{n_t \cos\theta_t - n_i \cos\theta_i}{n_t \cos\theta_t + n_i \cos\theta_i}, \end{cases} \tag{5.107}$$

where r is the reflectivity, the subscripts s and p denote S- and P-polarizations, and the subscripts i and t denote the incident and the transmitted sides, which are related by the Snell's law

$$n_i \sin\theta_i = n_t \sin\theta_t. \tag{5.108}$$

For normal incidence, Eq. (5.107) becomes angular- and polarization-independent and reduces to

$$r = \frac{n-1}{n+1}, \tag{5.109}$$

where we define $n = n_t/n_i$ to simplify the writing. One can also regard the simplification as assuming $n_i = 1$ for subjects in vacuum, which generally holds for our interests. We are interested in the transmitted power, so we calculate (algebra in Appx. A)

$$R = \left|\frac{n-1}{n+1}\right|^2 = \frac{\xi + 1 - \sqrt{2(\xi + \epsilon_1)}}{\xi + 1 + \sqrt{2(\xi + \epsilon_1)}}, \tag{5.110}$$

where we use $\epsilon_r = \epsilon_1 + i\epsilon_2 = n^2$ and define

$$\xi = |\epsilon_r| = \sqrt{\epsilon_1^2 + \epsilon_2^2} \tag{5.111}$$

to simplify the writing.

For a good conductor well below its plasma frequency, we have $|\epsilon_2| \gg |\epsilon_1| \gg 1$ (Fig. 5.18) and can further reduce Eq. (5.110) to

$$R \approx 1 - \frac{2\sqrt{2\epsilon_2}}{\epsilon_2 + 1 + \sqrt{2\epsilon_2}} \approx 1 - 2\sqrt{\frac{2}{\epsilon_2}}. \tag{5.112}$$

We have explained previously that, in this regime, one can adopt the Drude model conductivity relation

$$\sigma \approx \epsilon_2 \epsilon_0 \omega, \tag{5.113}$$

so we arrive at

$$R = 1 - 2\sqrt{\frac{2\epsilon_0 \omega}{\sigma}}. \tag{5.114}$$

Eq. (5.114) is commonly known as the Hagen-Rubens equation that describes metals' reflectance at normal incidence. The adoption of the Drude model conductivity relation indicates that the Hagen-Rubens model assumes the incident metal exhibits a conductivity high enough to allow ignoring the skin depth, effectively setting the metal thickness to infinity.

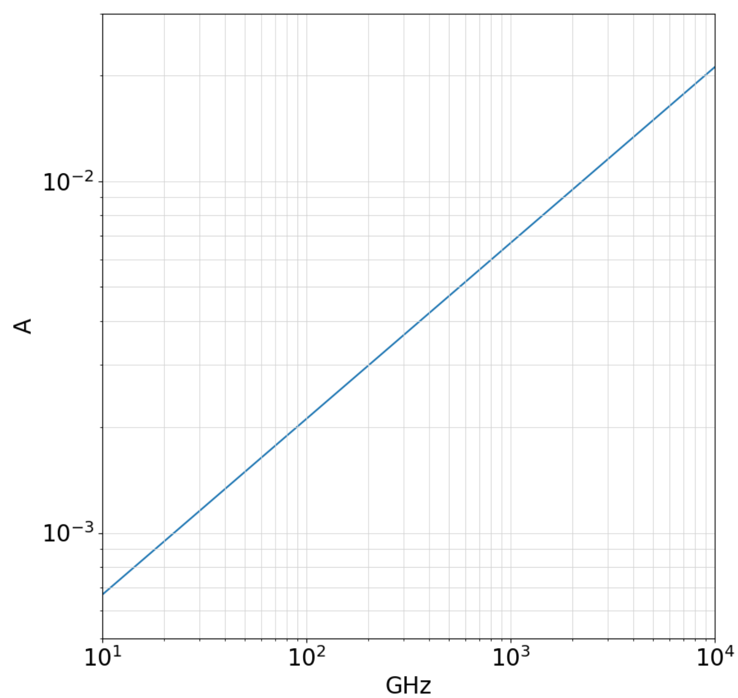


Figure 5.20: The Hagen-Rubens absorptance for thick Tungsten assuming material parameters obtained from SuperCDMS W films. More detailed explaination is found in the corresponding text.

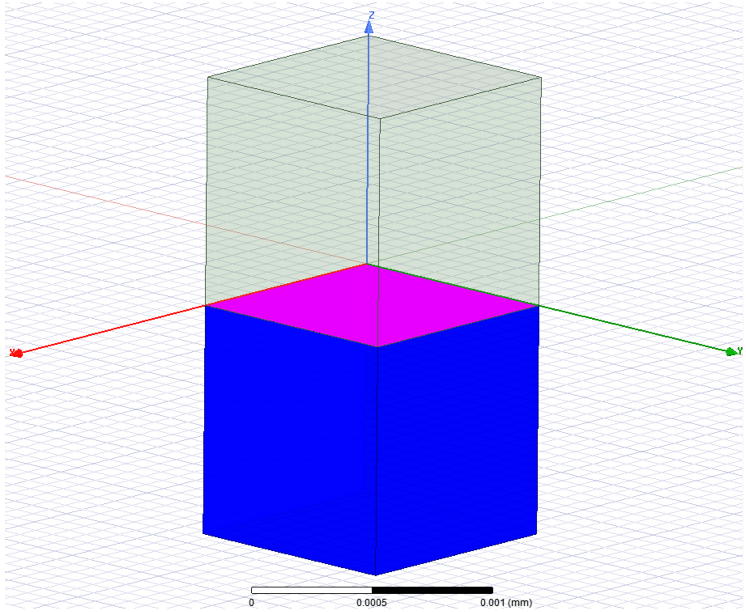


Figure 5.21: The HFSS model for the finite-thickness film loss study. The green transparent and blue cubes are vacuum and silicon, respectively, and the pink surface is the finite-conductivity finite-thickness W film.

We are most interested in the radiation loss in the film that is our photon signal, so we consider the absorptance

$$A = 1 - R = 2\sqrt{\frac{2\epsilon_0\omega}{\sigma}}, \tag{5.115}$$

which does not include a transmission term due to the infinite thickness assumption. Fig. 5.20 shows the result; notice that the prediction falls in the sub-percent range. It is concerning because the copper mechanical structure surrounding the TES photon detector exhibits a much higher absorptance (Fig. 5.4) and therefore is expected to make the detection inefficient. Furthermore, this absorptance is likely to be an overestimate for most frequencies due to unrealistically allowing the penetrating photons to be absorbed in the infinite thickness. In general, Eq. (5.115) quantifies the fact that the low absorptance is primarily a manifestation of the high metal conductivity, which is also the only parameter in the formula that appears to be adjustable in practice. We will introduce our detection efficiency improvement technique based on this conclusion after deriving full thin-film model.

We simulate the thin-film absorptance with the "DC thickness" option for object interfaces in HFSS. It is for constructing thin "coatings" on objects with specified DC conductivity and film thickness without actually constructing finite-thickness films, which would otherwise require very fine meshing that can lead to computation challenges. While this approach is later proven insufficient, it offers useful insights and can be a practical tool for detector design, so we briefly introduce this effort below.

We establish a general technique for simulating radiation losses in thin films and bulk volumes with HFSS, which is worth reporting for future references. We begin by constructing a simple model as shown in Fig. 5.11, consisting of a 1-μm^3 vacuum cube[12] on top, another identical Si cube below, with the cubes' interface, i.e., the zero-thickness pink square in Fig. 5.11, to be a finite-conductivity boundary with a DC thickness as our W film. Since HFSS adopts the convention of specifying material characteristics with real-number permittivity and conductivity, as opposed to a complex-number permittivity, we specify the Tungsten ϵ_1 and $\sigma_{\rm DC}$ derived earlier for the W plane. In addition to the material parameters, we select the "set DC thickness" option to approximate the interface and specify the actual film thickness, 40 nm. We also select "two-sided" under the DC thickness option, so both the W-vacuum and Si-W interfaces would be included according to HFSS technical manual. To complete the model as an infinitely extended object in the xy plane, master/slave-boundary pairs are assigned to the xz and yz sidewalls of the model, effectively extending the vacuum and Si cubes and their interface to infinity in $\hat{x}$ and $\hat{y}$. The top surface is assigned a "Floquet port," emitting at 1 W for a simple normalization, and the bottom surface carries a standard HFSS "radiation boundary" to terminate the wave propagation.

The main technique we would like to report, discovered by mixing HFSS tricks found on YouTube, is in the calculation and extraction for the radiation loss in the modeled finite-thickness W surface. We construct the loss variable as a HFSS "calculator expression" through the "calculator" interface under "field overlays." For the calculator setting, we choose to integrate the built-in "surface loss density" variable over the W plane and then add it to the variable library. The HFSS syntax for the calculator expression is

`Integrate(Surface([TES_plane_name]), Surface_Loss_Density).`

In order to employ the loss variable in a frequency-dependent fashion, the "sweep type" for the analyses needs to be selected as "discrete" with "save fields (all frequencies)." One may then create frequency-dependent field reports using the

[12] Same reason for adapting small spaces as in the waveguide simulation, namely suppressing unphysical resonances between boundaries due to numerical unbalances. The suppression is effective up to where the distance is equal to $\lambda/4$, i.e., 75 THz for the 1-μm^3 box.

predefined loss variable found under the “calculator expression” variable category. Similarly this technique may be applied to obtain bulk volume losses by choosing the volumes of interest during the variable construction.

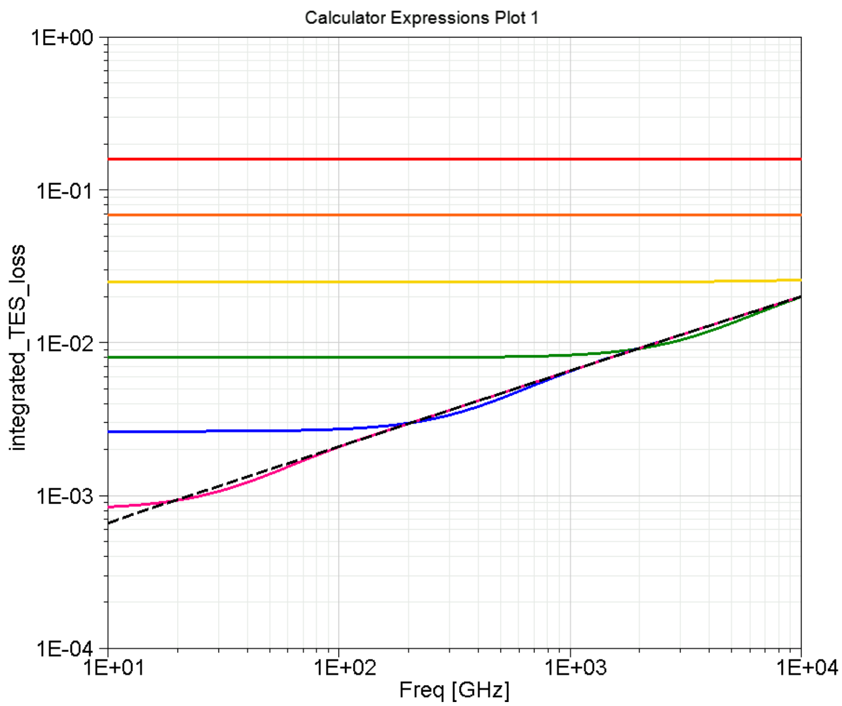


Figure 5.22: The HFSS-simulated finite-thickness absorptances for DC thicknesses of, from top to bottom, 4 nm, 13 nm, 40 nm, 130 nm, 400 nm, 1300 nm, and 4000 nm, colored in red, orange, yellow, green, blue, purple, and black dashed.

Utilizing the technique developed above, Fig. 5.22 shows the HFSS-simulated film loss for the Fig. 5.21 model. By comparing to the Hagen-Rubens equation curve given in Fig. 5.20, we recognize HFSS provides the identical result at high frequencies. Nevertheless, Fig. 5.22 also shows, for different film thicknesses, HFSS results stop following the Hagen-Rubens model and transition to constant losses at different frequencies. Note that the y-axis of Fig. 5.22 is normalized to the fractional absorptance identical to Fig. 5.20 by the designated Floquet-port power; we confirmed the numerical result is independent of the selected power after normalization. We further compare these transition frequencies to the skin depth curve in Fig. 5.19, where it becomes clear that the transitions occur at where the frequency-dependent skin depths exceed the assumed film thicknesses, i.e., violating Hagen-Rubens model's thick film condition. The result cross-checks our analytical calculations for the skin depth and the Hagen-Rubens loss and shows that, at high frequencies where the films may be treated thick with respect to the skin depths, HFSS' DC-thickness feature does reproduce the anticipated result. We will prove later in the complete theory that the asymptotic constant absorptance at low frequencies is also an accurate prediction.

Despite exhibiting these promising features, we find it is rather difficult to adapt HFSS for a bottom-up design process, as opposed to using it for simulating uncalculable but nevertheless semi-determined prototypes. The main practical inconvenience/constraint with HFSS is that it produces computationally consuming artifacts whenever it identifies model boundaries that are far enough to allow standing waves in the simulated frequency range. We find this issue is almost inevitable when simulating sub-mm-sized TES geometries. Due to the recursive FEA error propagation and HFSS algorithm's requirement of finding a stable convergence, slight computational errors at the boundaries would keep generating artifacts to prevent the simulation to complete. The situation is even more challenging concerning our particular application, in which we target multiple orders-of-magnitude frequencies, where we find we need to constantly divide and readjust the simulated sub-frequency ranges, isolate possible artifact-generating structures, but still produce dubious results that cannot be justified based on physics intuition or robust verification methods. From a more fundamental perspective, we attribute the struggle to not preparing ourselves with sufficient physical insights after the Hagen-Rubens model, hence not bringing the trial design to a close-to-optimal condition that suits numerical simulation's trial-and-error approach. We therefore decide to pursue a full analytical theory for thin-film Tungsten's response to mm-wave radiation.

5.4.3.3 Thin-film radiation loss model

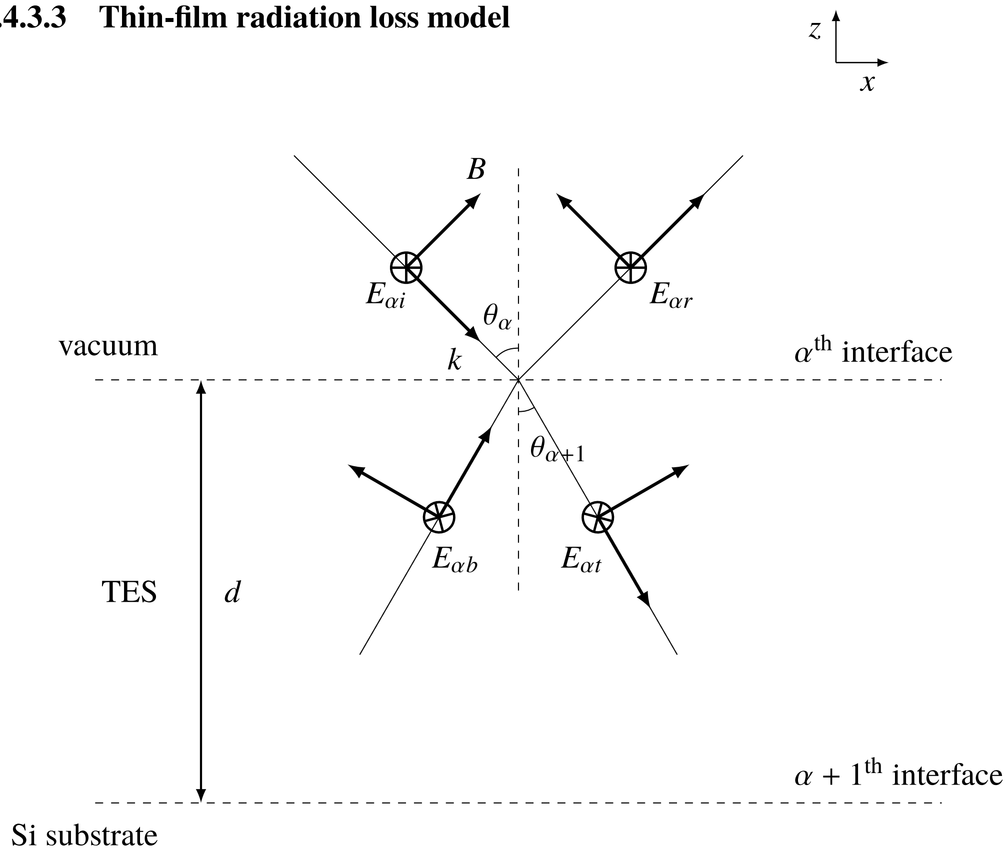


Figure 5.23: The side-view parameter construction for the S-polarization thin-film radiation loss model. Separated by the horizontal dashed lines from top to bottom are the semi-unbounded vacuum, thin-film W, and semi-unbounded substrate (Si) spaces. More explanation is provided in the corresponding text.

Fig. 5.23 presents the thin-film radiation loss model schematic for the TES, consisting of an one-side infinite vacuum space on top, a layer of sandwiched W of thickness d, and another one-side infinite substrate (Si) space below. We make the semi-infinite assumption based on the fact that the vacuum space above the TES and the millimeters-thick substrates utilized by SuperCDMS detectors are both much larger than the trans-mm wavelengths of interest. We also argue that the backside refection at the uninstrumented vacuum-Si interface produces randomized reflections and subsequently a random-phase interference in the substrate. Together with a wide range of incident angles, the impact of the bottom Si-vacuum interface is suppressed. Nevertheless, we acknowledge, while the photons being reflected back into the vacuum space are tracked by the particle-like simulation and are unlikely to return in practice, if the substrate for the detector is sufficiently lossless, while its backside is sufficiently reflective, e.g., due to Si-vacuum interface or close to a highly reflective copper surface, the photons transmitted through the W film can be reflected and form a nonneglegible constant backward light to the TES. In this case, the finite thickness of the substrate and the material below it should be accounted.

A quick estimate to see the level of importance for such an effect: For 10-K, ≈1 THz, BB radiation traveling in Si, $n \approx 3.5$, the wavelength is ≈100 μm. If the substrate is 1 cm in thickness, the total number of wavelengths for a "round trip" is

$$\approx 2 \times \frac{1\ \text{cm}}{100\ \mu\text{m}} \approx 200. \tag{5.116}$$

Assuming a loss tangent of $\approx 1 \times 10^{-5}$ for the substrate and a reflectance of $R = 0.9$ for the backside copper support, the fraction of the transmitted radiation coming back to the W film is

$$R \cdot (1 - \tan\delta)^k \approx 0.9 \times (1 - 10^{-5})^{200} \approx 0.9 \cdot (1 - 2 \times 10^{-3}). \tag{5.117}$$

We find the value is dominated by the reflectance of the backside reflector and could be significant depending on the materials. In this case, if the transmittance of the W film is also not small, although luckily the Hagen-Rubens model seems to suggest otherwise, the backward radiation in the substrate should be modeled. As shown in Fig. 5.23, however, it is currently not considered, but the effect may be rigorously included following the same procedure we will present in a moment. In the meantime, we will also show, our model indicates the transmittance for the W film, even modified to lower the effective conductivity as we will introduce later, is indeed small and undetectable. In terms of practical fabrication, we will dice the substrates to minimize the bare Si surfaces that are not covered by photon sensors, which also reduces the possibility for photons to enter the substrate through the TES as illustrated above. Finally, we will limit our final TES designs to be much larger than the wavelengths in the film's xy plane (coordinate defined in Fig. 5.23), so the radiation can not form resonances via matching the wavelengths to the TES size from edge to edge, therefore justifying the assumption that the layers are extended in the xy plane as in Fig. 5.23.

The mathematical derivation is as follows. We use the subscripts $\alpha = 0, 1, 2, ...,$ to denote the interfaces from top to bottom and i, r, t, b for the incident, reflected, transmitted, and backward waves at the interfaces, respectively, e.g., E_{0t} stands for the transmitted wave at the top interface. For the S-polarization configuration shown in Fig. 5.23, i.e., E field is in the W film plane, while B field is rotated depending on the incident angle, we match the in-plane $\vec{E}$ and $\vec{H}$ continuity boundary conditions at each interface and write (Jackson, 1998)

$$\begin{cases} X_\alpha(E_{\alpha i}+E_{\alpha r}) = X_{\alpha+1}(E_{\alpha t}+E_{\alpha b}) & (E_\parallel = \vec{E}\times\hat{z} \text{ continuity}) \\ Y_\alpha(E_{\alpha i}-E_{\alpha r}) = Y_{\alpha+1}(E_{\alpha t}-E_{\alpha b}) & (H_\parallel = \dfrac{\vec{k}\times\vec{E}}{\mu}\times\hat{z} \text{ continuity}), \end{cases} \tag{5.118}$$

where we use X_α and Y_α to represent the prefactors for the general independent additive $(E+E)$ and subtractive $(E-E)$ boundary-condition equations. For S-polarization in the coordinate system defined in Fig. 5.23, we have $E_\parallel$ in $\hat{x}$ and $H_\parallel$ in $\hat{y}$ with

$$\begin{cases} X_\alpha = 1 \\ Y_\alpha = \dfrac{n_\alpha}{\mu_\alpha}\cos\theta_\alpha, \end{cases} \tag{5.119}$$

where n_α, μ_α, and θ_α are the refractive index, permeability, and incident angle *on the incident side* of the α^{th} interface. Although the trivial X_α is fully expected by the in-plane E-field in S-polarization, we will continue the following derivation with X_α, so we may reuse the general results later for P-polarization, which has a nontrivial X_α. For Y_α, the refractive index is introduced by $|\vec{k}_\alpha| = n_\alpha\omega/c$, where ω and c are shared thus canceled for all α, and the $\cos\theta_\alpha$ factor is due to the cross product with $\hat{z}$ that projects $\vec{H}$ into the film plane. Clarifying these somewhat apparent algebraic details would be helpful for later comparison with P-polarization as well as understating the TES modification.

We also have the relations

$$\begin{cases} \delta_{(\alpha+1)}E_{\alpha t} = E_{(\alpha+1)i} \\ \delta_{(\alpha+1)}E_{(\alpha+1)r} = E_{\alpha b} \end{cases} \tag{5.120}$$

that propagate the waves from (to) the previous (next) interfaces, which promote Eq. (5.118) to become an infinite series, where the phase shift in Eq. (5.120) is

$$\delta_{\alpha+1} = e^{ik_{\alpha+1}d_{\alpha+1}/\cos\theta_{\alpha+1}}, \tag{5.121}$$

where $k_{\alpha+1}$ and $d_{\alpha+1}$ are the wavevector and layer thickness, respectively, and we emphasize again the quantities are for the *incident side* of the $(\alpha+1)^{\text{th}}$ interface.[13] In principle, one may then combine Eq. (5.118) and Eq. (5.120) to form a formal operator treatment for a stack of many layers of media (Born and Wolf, 1999), which

[13]It is important to have the correct sign in Eq. (5.121). For a forward-propagating wave toward $+\vec{x}$, the wave function is $e^{i(\vec{k}\cdot\vec{x}-\omega t)}$. At any given moment ($t=0$), the position $+d$ in front of the current position ($x=0$) experiences the wave that was emitted kd/ω *ahead of time*, i.e., at $t=-kd/\omega$. Substituting the lagged time into the wave equation, one obtains the propagator in Eq. (5.121) with the correct sign.

is useful for future works to include the backside reflection mentioned previously. For the simpler two-interfaces case we focus in this thesis, we set

$$d_0 = 0$$

as the start of the stack,

$$d_2 \to \infty$$

for the semi-opened space, and keep only

$$d_1 = d \tag{5.122}$$

as in Fig. (5.23). We also remove the subscript for δ_1, since it is the only sandwiched layer in our model. Eq. (5.118) becomes

$$\begin{cases} X_0(E_{0i} + E_{0r}) = X_1(E_{0t} + E_{1r}\delta) \\ Y_0(E_{0i} - E_{0r}) = Y_1(E_{0t} - E_{1r}\delta) \\ X_1(E_{0t}\delta + E_{1r}) = X_2 E_{1t} \\ Y_1(E_{0t}\delta - E_{1r}) = Y_2 E_{1t} \ , \end{cases} \tag{5.123}$$

where the upper and the lower two equations are for $\alpha = 0$ and $\alpha = 1$ interfaces, respectively. There are five unknowns, E_{0i}, E_{0r}, E_{0t}, E_{1r}, E_{1t}, with four constraints in Eq. (5.123), so we will determine all the unknowns as ratios relative to the incident E_{0i}. In particular, we are interested in the transmission through the W film,

$$\frac{E_{1t}}{E_{0i}} = 4\left\{\frac{X_1}{X_0}\left[\left(\frac{X_2}{X_1}+\frac{Y_2}{Y_1}\right)\delta^{-1} + \left(\frac{X_2}{X_1}-\frac{Y_2}{Y_1}\right)\delta\right] + \frac{Y_1}{Y_0}\left[\left(\frac{X_2}{X_1}+\frac{Y_2}{Y_1}\right)\delta^{-1} - \left(\frac{X_2}{X_1}-\frac{Y_2}{Y_1}\right)\delta\right]\right\}^{-1} \ , \tag{5.124}$$

and the reflection by it,

$$\frac{E_{0r}}{E_{0i}} = \frac{\frac{X_1}{X_0}\left[\left(\frac{X_2}{X_1}+\frac{Y_2}{Y_1}\right)\delta^{-1} + \left(\frac{X_2}{X_1}-\frac{Y_2}{Y_1}\right)\delta\right] - \frac{Y_1}{Y_0}\left[\left(\frac{X_2}{X_1}+\frac{Y_2}{Y_1}\right)\delta^{-1} - \left(\frac{X_2}{X_1}-\frac{Y_2}{Y_1}\right)\delta\right]}{\frac{X_1}{X_0}\left[\left(\frac{X_2}{X_1}+\frac{Y_2}{Y_1}\right)\delta^{-1} + \left(\frac{X_2}{X_1}-\frac{Y_2}{Y_1}\right)\delta\right] + \frac{Y_1}{Y_0}\left[\left(\frac{X_2}{X_1}+\frac{Y_2}{Y_1}\right)\delta^{-1} - \left(\frac{X_2}{X_1}-\frac{Y_2}{Y_1}\right)\delta\right]} \ ; \tag{5.125}$$

the algebra for deriving Eq. (5.124) and Eq. (5.125) may be found in Appx. A. We confirm Eq. (5.125) reduces to Fresnel's law in Eq. (5.107) at $d \to \infty$.

Despite being completely general, Eq. (5.124) and Eq. (5.125) provide little insight to aid the TES design. Luckily in our case, we may leverage the drastic difference between metals and dielectrics to simplify the results. Specifically, we know

$$n_1 = \sqrt{\epsilon_1} \sim O(10^3)(1+i) \tag{5.126}$$

for Tungsten and

$$n_{0,2} \sim O(10^0) \tag{5.127}$$

for vacuum and silicon, which gives

$$Y_1 \gg Y_{0,2}, \tag{5.128}$$

except for very large angles where one should use Snell's law to really calculate the $\cos\theta_\alpha$ term in Y_α. The above condition further simplifies Eq. (5.124) and Eq. (5.125) to

$$\frac{E_{1t}}{E_{0i}} = 2\left\{\cos(k_1 d/\cos\theta_1) - n_1 \frac{\cos\theta_1}{\cos\theta_0}\sin(k_1 d/\cos\theta_1)\right\}^{-1} \tag{5.129}$$

and

$$\frac{E_{0r}}{E_{0i}} = \frac{\cos(k_1 d/\cos\theta_1) + n_1 \dfrac{\cos\theta_1}{\cos\theta_0}\sin(k_1 d/\cos\theta_1)}{\cos(k_1 d/\cos\theta_1) - n_1 \dfrac{\cos\theta_1}{\cos\theta_0}\sin(k_1 d/\cos\theta_1)}, \tag{5.130}$$

for which we also use the fact that $\mu_\alpha \approx \mu_0$ (non-ferromagnetic) and $n_0 = 1$ (vacuum) to simplify the formulae. With the large W permittivity, or equivalently a large conductivity, the condition effectively decouples the substrate from the system, which is manifested by the removal of all $\alpha = 2$ terms from Eq. (5.124/5.125) to Eq. (5.129/5.130).

We further inspect the $(k_1 d/\cos\theta_1)$ argument for the sine and cosine terms in Eq. (5.129) and Eq. (5.130), which stems from the phase delay term in Eq. (5.121) due to the finite film thickness. In the thin-film limit $(k_1 d/\cos\theta_1) \ll 1$, or more meaningfully

$$\frac{\lambda_1}{2\pi} \gg \frac{d}{\cos\theta_1}, \tag{5.131}$$

i.e., the wavelength is much longer than the path it traverses in the film (up to a factor of 2π), Eq. (5.129) and Eq. (5.130) reduce to

$$t = \frac{E_{1t}}{E_{0i}} = \frac{2}{1 - n_1 k_1 d/\cos\theta_0} \tag{5.132}$$

$$r = \frac{E_{0r}}{E_{0i}} = \frac{1 + n_1 k_1 d/\cos\theta_0}{1 - n_1 k_1 d/\cos\theta_0} = -1 + t. \tag{5.133}$$

We find, for two dielectric materials separated by a thin layer of metal, the transmittance and reflectance have

$$T = |t|^2 \ll 1 \quad \text{(empirical)}, \tag{5.134}$$

$$R = 1 - t - t^* + \cancel{|t|^2} \approx 1 - 2\mathbf{Re}(t), \tag{5.135}$$

and the absorptance is then

$$\begin{aligned} A = |a|^2 &= 1 - |r|^2 - |t|^2 \\ &\approx 2\mathbf{Re}(t). \end{aligned} \tag{5.136}$$

The result suggests, in the reflection-dominated regime, the radiation energy that is not reflected should be predominately absorbed by the film, while the transmitted component is quadratically suppressed to be negligible when Y_1 (W) $\gg$ $Y_{0,2}$ (vacuum, Si).

The point may be further elaborated by recalling we have shown, for W at much below its ≈15 THz damping frequency,

$$\epsilon_r = \cancel{O(10^4)} + O(10^{>6})i \approx \frac{\sigma}{\omega\epsilon_0}i, \tag{5.137}$$

i.e., approximately purely imaginary for a good conductor, and using

$$n_1 k_1 = \sqrt{\epsilon_r \mu_r} \cdot \omega\sqrt{\epsilon\mu} = \frac{\omega\epsilon_r}{c} \tag{5.138}$$

$$= \frac{\sigma}{\epsilon_0 c}i, \tag{5.139}$$

where we substitute Eq. (5.137) for the last step and again letting $\mu_r \approx 1$, the metallic thin-film transmission we derived earlier (Eq. (5.132)) becomes

$$t = \frac{2}{1 - (\sigma d/\epsilon_0 c\cos\theta_0)i}. \tag{5.140}$$

As Fig. 5.22 already hinted, this transmittance formula, as well as the corresponding absorptance per Eq. (5.136), are now frequency-independent constants. If we continue applying the low-frequency skin depth from Eq. (5.97), for normal incidence

($\theta_0 = 0°$), we find (algebra found in Appx. A)

$$A = 2\mathbf{Re}(t) \approx \frac{\delta}{d} \times \left(2\sqrt{\frac{2\epsilon_0\omega}{\sigma}}\right). \tag{5.141}$$

We note again this expression is a frequency-independent constant as explained. Even though Eq. (5.141) appears to carry an ω in the square-root, it is canceled by the implicit $\omega^{-1/2}$ dependence in δ, but since we have purposely rearranged the expression, we may easily compare Eq. (5.141) to the Hagen-Rubens expression in Eq. (5.115) and find the two equations coincide when δ approaches d. We therefore have proven the low-frequency asymptotic behavior given by HFSS is correct, and the high-/low-frequency regime transition is indeed around $\delta \sim d$, i.e., thick/thin-film regimes. Our result is, however, more general than the HFSS normal incidence simulation shown in Fig. 5.21 and Fig. 5.22. In particular, it provides the full angular-dependent description for the film response (Eq. (5.124/5.125)) at all frequencies, which will be useful for modeling our "meshed metal" film design in the next section, including incident angle and polarization dependences.

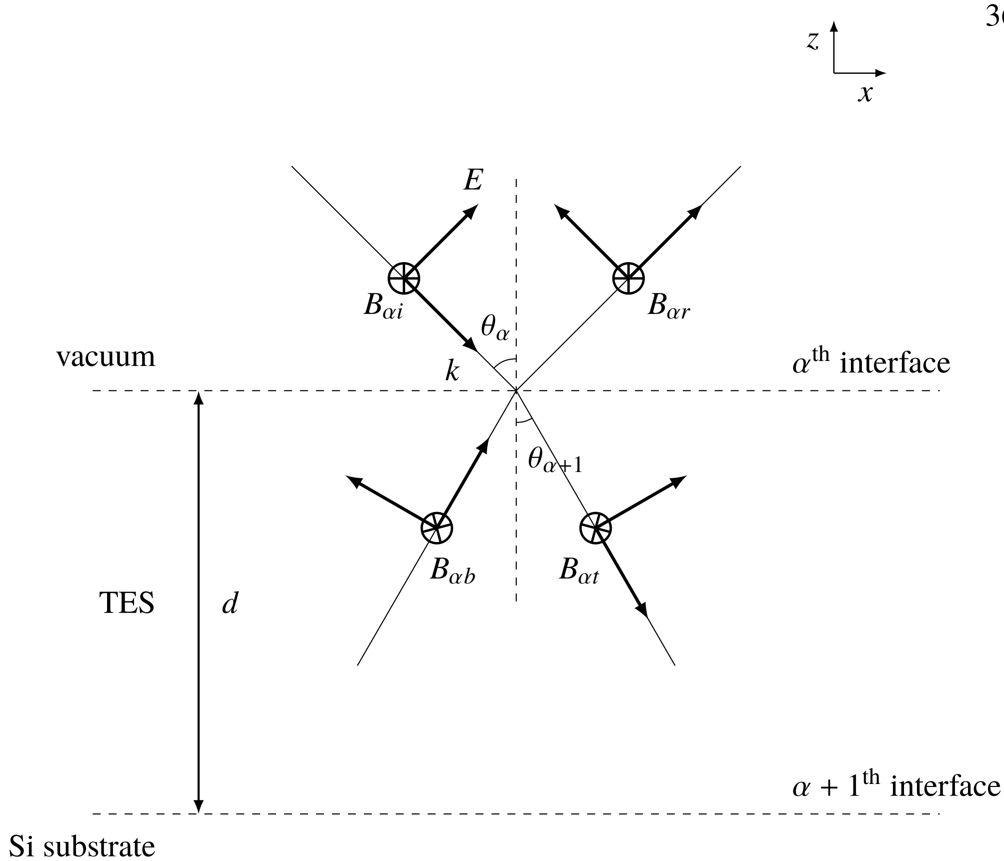


Figure 5.24: The parameter construction for the P-polarization thin-film radiation loss model. More explanations are provided in the corresponding text and Fig. 5.23.

After the detailed analysis for the S-polarization presented above, we may now straightforwardly repeat the procedure for the P-polarization, i.e., B field in the TES plane; Fig. 5.24 shows the model setup. We examine the same independent boundary conditions as in Eq. (5.118):

$$\begin{cases} X_\alpha(E_{\alpha i} + E_{\alpha r}) = X_{\alpha+1}(E_{\alpha t} + E_{\alpha b}) & (H_\parallel \text{ continuity}) \\ Y_\alpha(E_{\alpha i} - E_{\alpha r}) = Y_{\alpha+1}(E_{\alpha t} - E_{\alpha b}) & (E_\parallel \text{ continuity}), \end{cases} \tag{5.142}$$

which, due to the symmetric exchange of E- and B-field orientations, now yield the additive X equation from the $H_\parallel$ continuity and the subtractive Y equation from $E_\parallel$. Owing to the exchange, we have

$$\begin{cases} X_\alpha = \dfrac{n_\alpha}{\mu_\alpha} \\ Y_\alpha = \cos\theta_\alpha \, . \end{cases} \tag{5.143}$$

Compared to Eq. (5.119), here we have (n_α/μ_α) moved to X_α, so it may continue reflect $H_\parallel = (\vec{k} \times \vec{E})/\mu \times \hat{z}$. We find $\cos\theta_\alpha$ remains in Y_α, because in the case of

P-polarization it is the E-field, now corresponding to Y_α, that is orientated relative to the normal direction and needs to be projected depending on the incident angle.

Similarly, we confirm Eq. (5.125) with Eq. (5.143) does reduce to Fresnel's equation for P-polarization at the $d \to \infty$ limit. We may further simplify the formulae with

$$\frac{X_\alpha}{X_\beta} = \frac{n_\alpha}{n_\beta} \tag{5.144}$$

based on the fact that all $\mu_\alpha \approx \mu_0$, and for small angle incidence[14], we may take

$$\frac{Y_\alpha}{Y_\beta} = \sqrt{\frac{1 - \sin^2\theta_\alpha}{1 - \sin^2\theta_\beta}} \approx 1 + \left(\frac{\sin^2\theta_\beta - \sin^2\theta_\alpha}{2} + ...\right). \tag{5.145}$$

Again taking advantage of $|n_1| \gg |n_{0,2}|$ due to the metal-dielectric distinction, Eq. (5.124) and Eq. (5.125) simplify to

$$t = \frac{E_{1t}}{E_{0i}} = 2\left\{-n_1 \sin(k_1 d) + \cos(k_1 d)\right\}^{-1} \tag{5.146}$$

$$r = \frac{E_{1t}}{E_{0i}} = \frac{-n_1 \sin(k_1 d) - \cos(k_1 d)}{-n_1 \sin(k_1 d) + \cos(k_1 d)} = 1 - t. \tag{5.147}$$

Despite these equations' distinct angular dependence compared to the S-polarization (Eq (5.129), Eq. (5.130)), at the thin-film limit, the above transmission reduces to

$$t \approx \frac{2}{-n_1 k_1 d + 1} = \frac{2}{1 - i(\sigma d/c\epsilon_0)}, \tag{5.148}$$

which is exactly identical to S-polarization, and the reflection is also identical, except with a phase difference of π that does not affect the energy transmission. We therefore conclude, the physics conclusions we obtained for the S-polarization are identically applicable for the P-polarization.

Fig. 5.25 summarizes the result of our thin-film mm-wave radiation TES detector model, assuming the nominal W conductivity of $\sigma = 1 \times 10^7$ 1/Ω·cm and thickness of 40 nm. Overall, it is consistent with our expectation for the physics characteristics of the system and earlier order-of-magnitude estimates (Eq. (A.98) & Eq. (A.100)): The film exhibits a close-to-unity reflectance, a negligible 10^{-4} transmittance, and a percent-level absorptance. Phenomenologically, the resulting absorptance varies in correlation with the transmittance but is largely decoupled from the reflectance, which reflects our earlier conclusion that, in the reflectance/conductivity-dominated regime, the thin-film absorptance is effectively proportional to the trasmittivity[15] t

[14] Not a general condition; this is only to compare with HFSS normal incidence simulation.
[15] Notice it is not the transmittance T.

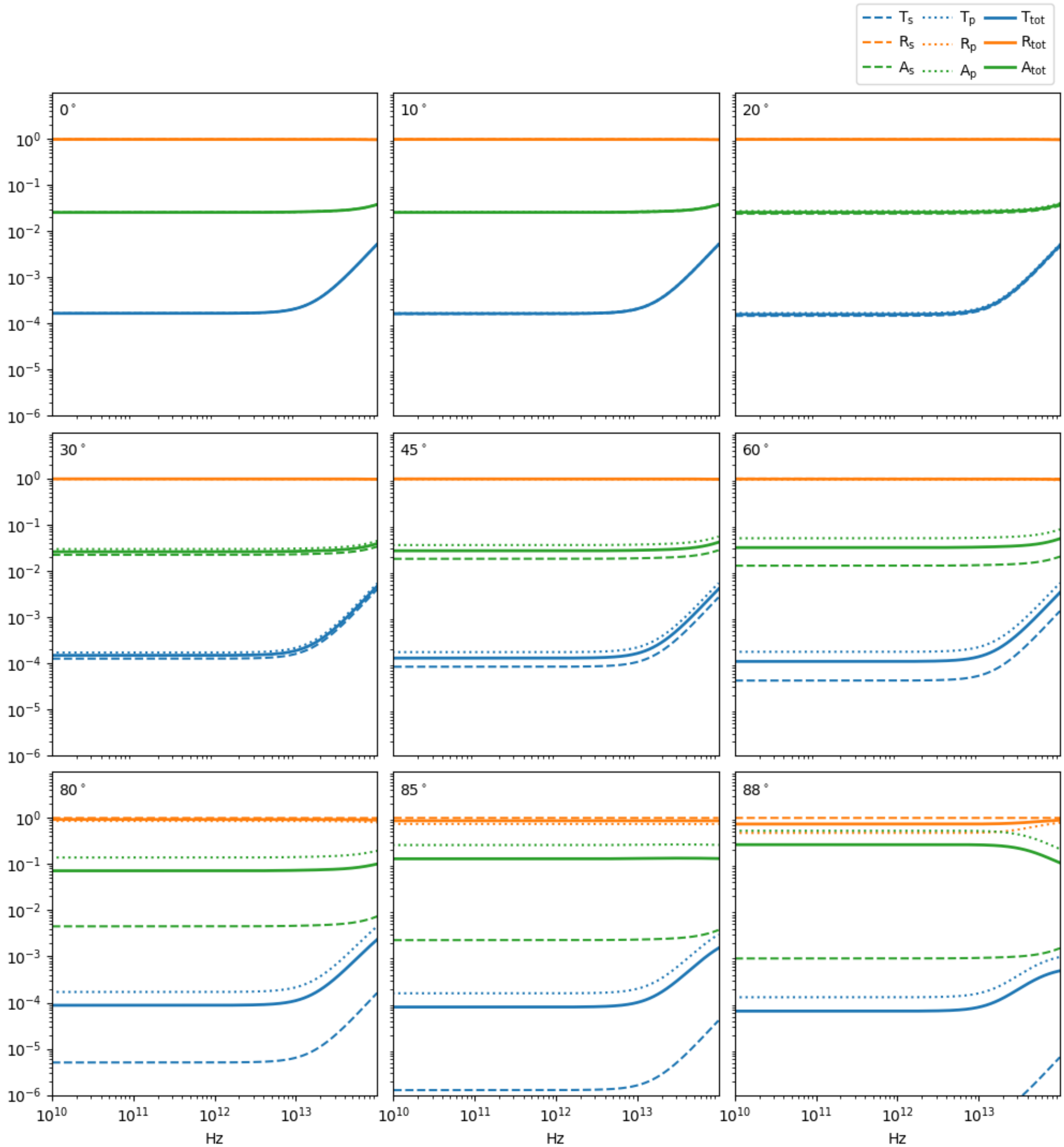


Figure 5.25: The W film transmittance T (blue), reflectance R (orange), and absorptance A (green) versus frequency, assuming nominal conductivity and film thickness. The S- and P-polarizations are represented by dashed and dotted curves, respectively, while their combination, assuming the mean of the two, is represented by the solid curves. The assumed angle of incidence is noted in the upper-left of each panel. Detailed discussion may be found in the corresponding text.

but is insensitive to the reflectance (Eq. (5.136)). At the same time, since the transmittance is modulated by the effective thickness of the thin film that the photons traverse, i.e., $d/\cos\theta$, the transmittance and absorptance expectedly depend on the incident angle, weakly for small angles but nontrivially at near-grazing angles as shown by the bottom panels of Fig. 5.25. Fortunately, for our nominal material parameters, the angular and frequency dependences are only significant at roughly >60° and >10 THz (>50 K), respectively. As the incident angle increases, the parallel E-field impinging on the film in S-polarization does not change in amplitude, but the effective film thickness for the field to penetrate increases and leads to less power transmitted and also absorbed in the conductivity-dominated regime. On the contrary, for P-polarizatioin, large angels imply more E-field is projected to the normal direction of the conductive film to avoid skin-effect attenuation, so the transmittance and the correspondingly absorptance are enhanced for P-polarization at large angles. The angular dependence of the absorptance suppression (enhancement) for S- (P-)polarization results in a factor of <3 variation below 60°, while the two cancel each in a mixed general incident to ≈25% at 60°.

We may examine the model in the conditions that are more closely related to practical experimental situations. We scan the anticipated ranges of fabricated conductivities and film thickness options as shown in Fig. 5.26. In practice, the conductivity may vary by order-of-magnitude as considered in the figure due to uncontrollable factors in fabrication but nevertheless may be accurately measured once fabricated. On the contrary, the film thickness is much more accurately controlled in fabrication to within ±1 nm but could be chosen in the simulated range. Fig. 5.26 concludes our analysis for employing unmodified thin-film W TES as mm-wave photon detectors via simple in-film radiation loss. We find the absorptance, the "detection efficiency," is generally low due to the high conductivity of Tungsten. Demonstrated by the analytical derivation as well as the figures, if one can obtain an order-of-magnitude reduction in conductivity, still in the high-reflectivity regime, the absorptance is expected to be enhanced by roughly also an order-of-magnitude. With realistic material parameters, we find the absorptance exhibits small frequency dependence in the BB radiation frequency range up to 50 K and starts deviating appreciably from the low-frequency absorptance close to 10s THz, i.e., near room temperature. This high-frequency deviation may be further suppressed by reducing the conductivity or thickening the film.

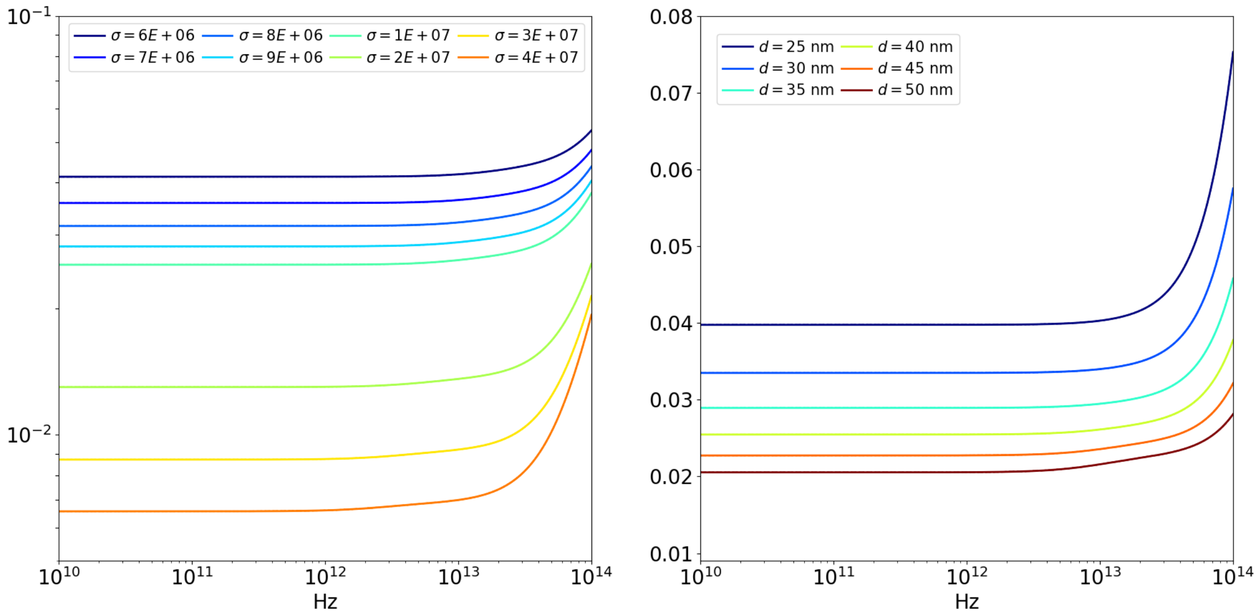


Figure 5.26: The W film absorptance for different conductivities (left) and film thicknesses (right), color-coded according to the legends. The curves from bottom to top correspond to high-to-low conductivities or thick-to-thin films, respectively. As an example, this figure shows the normal incident case, where the S- and P-polarizations are degenerate thus not distinguished in the figure. For both figures, when varying the conductivity or the thickness, the other parameter is fixed to the nominal value, 1×10^7 $1/\Omega{\cdot}\mathrm{cm}$ or 40 nm.

5.4.3.4 Meshed TES photon detector

We have demonstrated with Fig. 5.26 that the BB photon absorptance for the original bias power R&D TES is much smaller than copper (Fig. 5.4), which is inefficient in general and can be worsened by the completion with surrounding mechanical structures. It is also an issue for the simulation, since in this case the prolonged photon propagation caused by the increased reflections becomes more sensitive to the uncertainties affecting the simulation, such as the assumed material properties and geometries. We have also shown in Eq. (5.141) that, in the reflectivity-dominated regime, the thin-film absorptance is inversely proportional to the DC conductivity at frequencies much below the damping frequency. Intuitively, these results suggest that we should simply reduce the conductivity of the film, therefore suppressing the reflectivity and in turn improving the BB photon absorptance. However, while such a solution is achievable by adjusting the W film fabrication parameters (deposition speed, background pressure, etc.), it is not preferred because the TES is designed to operate at its optimal bias condition based on the nominal conductivity. Nevertheless, we can circumvent this constraint by recognizing that the BB radiation predominately impinges on the TES film from near-normal directions and at relatively high frequencies, contrasting with the electro-thermal bias current flowing in the TES plane at near-DC. This distinction inspires us to seek for a conductivity modification scheme that depends on the direction and the frequency of the incoming excitation, so that we can increase the normal-direction BB radiation absorptance while maintain the in-plane electro-thermal bias. Fig. 5.27 gives a simple example illustrating the idea. It shows that, as a resistor biased vertically in the page plane, one can "mesh" the feature horizontally as in the figure without altering the TES' DC resistance. At the same time, the radiation shining perpendicularly onto the page would encounter the meshed pattern and therefore a modified radiation impedance. The designated 1:4 aspect ratio of the rectangular TES requires that we maintain for later modifications a sheet resistance of 1/4 square (□), biased along the short edge as in Fig. 5.27. We therefore seek for patterns modified from the simple rectangle that exhibit reduced effective conductivities in the normal direction while maintain the 1/4 □ resistance.

We reference the metal mesh radiation absorber commonly utilized in sub/mm astronomy research and industry for the new design (Bock et al., 1995). Considering the wavelength/photon size, material properties varying within a trans-mm region is unresolvable and thus smeared for the distributed wave. Taking Fig. (5.27) for example, as long as the unit cell, i.e., one black rectangle with one neighboring white rectangle, is much smaller than the wavelength in discussion, the in-cell discontinuities can not impose localized boundary conditions to the field for generating narrow-frequency responses. On the contrary, the area appears to the incident photons as a smooth composite medium, averaging the properties of the black and white materials roughly by their frontal area ratio. In practice, (W. Jones, 2005) showed with numerical simulations as well as real measurements that, for a similar "spider web" radiation absorber for the BOOMERANG focal plane bolometer, metal films meshed by gap widths[16] smaller than 1/5 of the wavelengths appear as homogeneous materials that average the film metal and vacuum (cutout in the web) by their relative frontal areas.

Since we are most interested in the conductivity reduction, the most direct way to rigorously quantify the meshing effect on the absorptance is through radiation impedance adjustment, i.e., the inverse of conductivity. With the EM-wave

[16] the white regions in Fig. 5.27.

Figure 5.27: TES meshing illustration. The black rectangular and red line features represent the resistive TES and SC bias-line features viewed from atop. Owing to the bias current direction, the original bias-power R&D TES, illustrated by the left-hand side figure, may be meshed into the right-hand side design without altering its DC resistance, while the radiation-sensing pattern is visibly modified for near-normal incidence onto the page. More explanation is provided in the corresponding text.

impedance

$$Z = \sqrt{\frac{\mu}{\epsilon}}, \tag{5.149}$$

we may rewrite Eq. (5.119) as

$$\frac{Y_\alpha}{Y_\beta} = \frac{Z_\beta \cos\theta_\alpha}{Z_\alpha \cos\theta_\beta}, \tag{5.150}$$

for which we find in Eq. (5.124) and Eq. (5.125) this *ratio* of impedances fully determines the reflection and transmission—this is the concept of impedance *matching*. To connect the TES meshing to impedance, we again use the fact that the conductive film has a permittivity many orders-of-magnitude larger than vacuum or the substrate, so instead of averaging the black (W) and white (substrate) regions by their realistic permittivities, we may simply neglect the substrate contribution and adapt a fill fraction

$$0 \le f \le 1, \tag{5.151}$$

i.e., the fractional area of W in the unit cell, to represent the modified effective permittivity as

$$\epsilon_{\text{eff}} = f\epsilon_{\text{W}} = f\frac{i\sigma}{\omega}. \tag{5.152}$$

Eq. (5.152) makes it clear that the meshing is equivalent to introducing a “dilution” factor f for the film to take on a reduced effective conductivity $f\sigma$ for normal-direction incidence.

We note that the conductivity dilution interpretation is not precise in that, considering bulk conductivity, the meshing technique is only easily achievable in the film plane but not fully 3D, e.g., not laminating the W film perpendicularly with dielectrics. Therefore, it is expected that the planar modification would become polarization-dependent, which we will revisit next. If for now we continue with the

$$\sigma \to f\sigma \tag{5.153}$$

interpretation sloppily using the bulk conductivity for a simple explanation, we immediately find in Eq. (5.140/5.141) that we may enhance the absorptance by $1/f$[17], suggesting easily an order-of-magnitude boost with $f < 0.1$. Similar mathematical treatment applies to P-polarization, i.e., Eq. (5.143), where we may rewrite

$$\frac{X_\alpha}{X_\beta} = \frac{Z_\beta^2 \sin\theta_\alpha}{Z_\alpha^2 \sin\theta_\beta} \tag{5.154}$$

for intermediate steps and arrive at an identical conclusion that the absorptance depends linearly on $1/f$. Nevertheless, we will explain later that, due to incident angle and polarization projection to the meshed direction, the f dependence applies differently for S- and P-polarizations. It results in a variable improvement that depends on the orientations of events and is twice more effective for S-polarization on average.

In order to accommodate the in-plane anisotropy shown in the Fig. 5.27 example or the fact that the film can not be laminated perpendicularly, we promote the constant permittivity to a general tensor. Due to the fact that the meshed TES is a passive device that obeys energy conservation and correspondingly time reversal symmetry, its permittivity tensor must be symmetric (Jackson, 1998), i.e.,

$$\epsilon_{ij} = \bar{\epsilon}_{ji}. \tag{5.155}$$

According to the principle axis theorem in linear algebra, we can then choose the principle axes to represent the symmetric tensor, which for Fig. 5.27 is just the Cartesian coordinate, i.e.,

$$\epsilon_{ij} = \begin{cases} 0 & , \ i \neq j \\ f\epsilon_{\mathrm{W}} & , \ i = j = x \\ \epsilon_{\mathrm{W}} & , \ i = j = y, z \, , \end{cases} \tag{5.156}$$

where we choose the only meshed axis to be $\hat{x}$; one can also apply different fill fractions to other diagonal elements based on the design.

The consequence of Eq. (5.156) for S-polarization is that the $H_{||}$ continuity condition in Eq. (5.118) now splits into two: One has the projected E field in $\hat{y}$ and follows the original derivation, and the other projects into $\hat{x}$ and takes the "$\epsilon_{\mathrm{W}} \to f\epsilon_{\mathrm{W}}$" modification. It is in fact easy enough to see from the asymmetric pattern in Fig. 5.27,

[17]This relation holds for small adjustments that does not invalidate the reflection-dominated condition. The modification is otherwise not as effective but still analytically calculable with the modified conductivity.

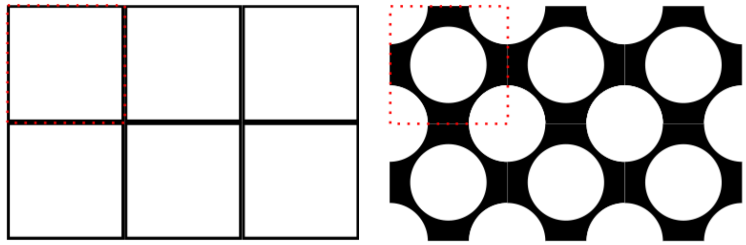


Figure 5.28: Examples of commonly adopted isotropic "photonic crystal" patterns. The red-dotted boxes mark the unit cells with 4-fold rotational symmetries.

but we have nevertheless proven it is indeed the case with the permittivity tensor. If assuming random polar angles for the S-polarized incidence (Fig. 5.23), we realize that the asymmetric ϵ_{xx} and ϵ_{yy} imply on average[18] 1/2 of the incident energy would experience the modified absorptance. Note that the size of modification is a function of the variable E_x projection but not equal for every event/polar angle. To see the consequence for P-polarization, we first rewrite the general $D_\perp$ continuity condition

$$\left[\sum_i \epsilon_i E_i\right] \cdot \hat{z} = 0. \tag{5.157}$$

It shows, unless we modify the W film in the normal direction $\hat{z}$, ϵ is likely to remain unadjustable in $\hat{z}$. Therefore, in order to absorb a P-polarized incidence in the mesh-enhanced direction (Fig. 5.24), the conjugate E field needs to project first into the film plane and then the meshed axis.[19] Assuming random azimuthal and polar angles, the projection implies on average 1/4 of the P-polarized events would encounter the mesh-enhanced absorptance. We emphasize again, this statement should not be confused with that 1/4 of the total absorptance is multiplied by $1/f$. Due to the angular-dependent E_x projection as well as the nonlinear angular and ϵ dependences in the absorptance formulae, the absorptance at every incident angle responds to the modified $f\epsilon_{\mathrm{W}}$ differently. Hence, we need to specify the incident angle with polarization for guiding the following meshed TES design.

We now understand the meshing scheme in Fig. 5.27 would yield an anisotropic absorptance since it is only meshed single-axis. While there is always the vertical modification limit, if we mesh the TES isotropically in the film plane, e.g., adopting the patterns in Fig. 5.28, the film would at least respond isotropically to the in-plane component of the incoming radiation. Adopting the isotropic pattern effectively

[18] Due to the in-plane $E_x = E\cos\phi$ projection, where ϕ is the polar angle. The azimuthal angle is irrelevant due to S-polarization's in-plane definition.

[19] i.e., $(\sin\theta\cos\phi)$, where θ is the azimuthal angle.

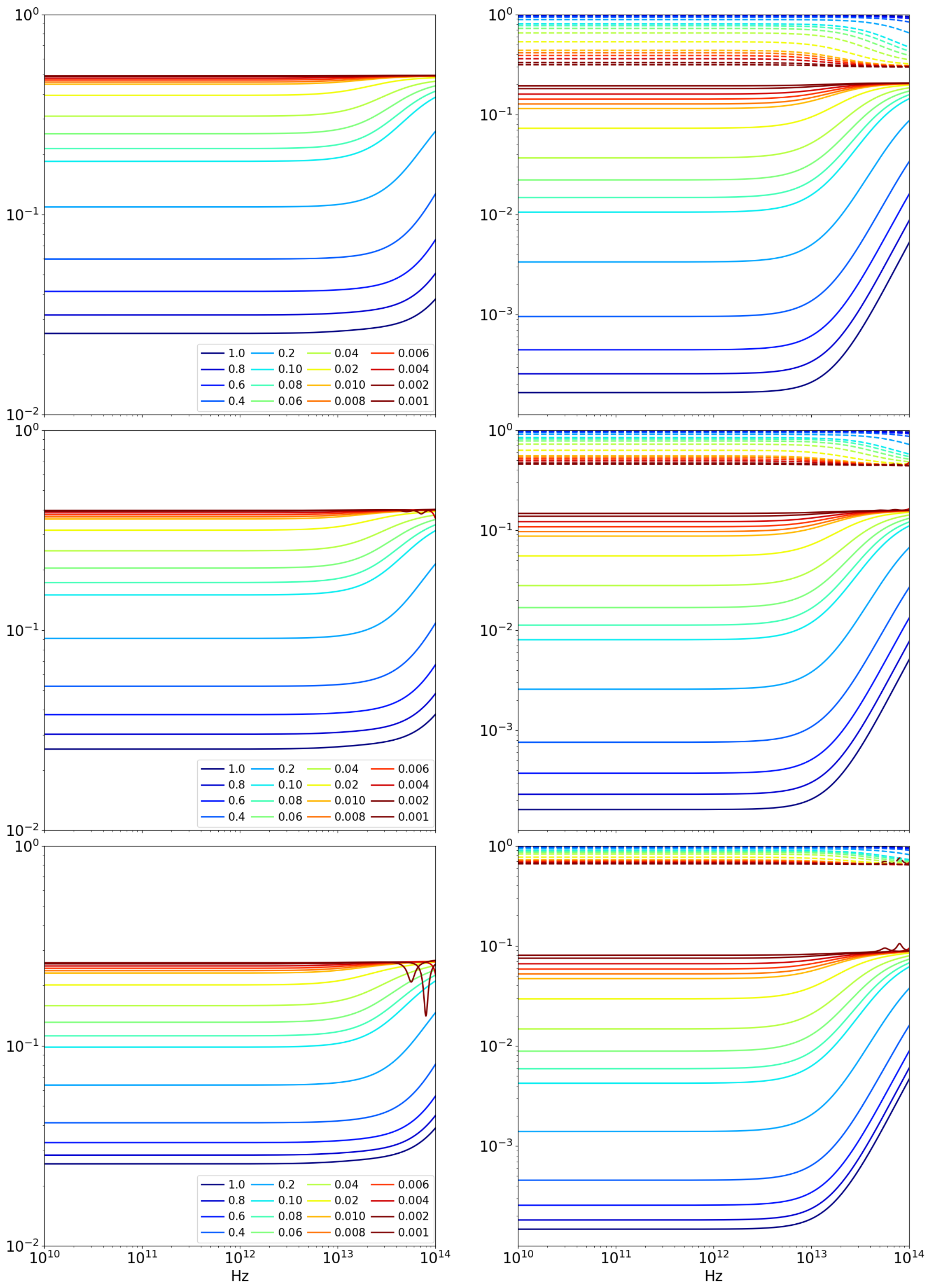


Figure 5.29: The W film absorptance (left), transmittance (solid, right), and reflectance (dashed, right) given by our analytical model for different fill fractions, color-coded as in the legend. The top, middle, and bottom panels are for 0°, 15°, 30° incidents, respectively. For all results shown in the figure, we assume the incoming radiation is randomly polarized, the fill fraction applies isotropically to the film plane but not perpendicularly, and combine the S- and P-polarization results based on their field projections to the f-modified and -unmodified directions. More detailed discussion may be found in the corresponding text.

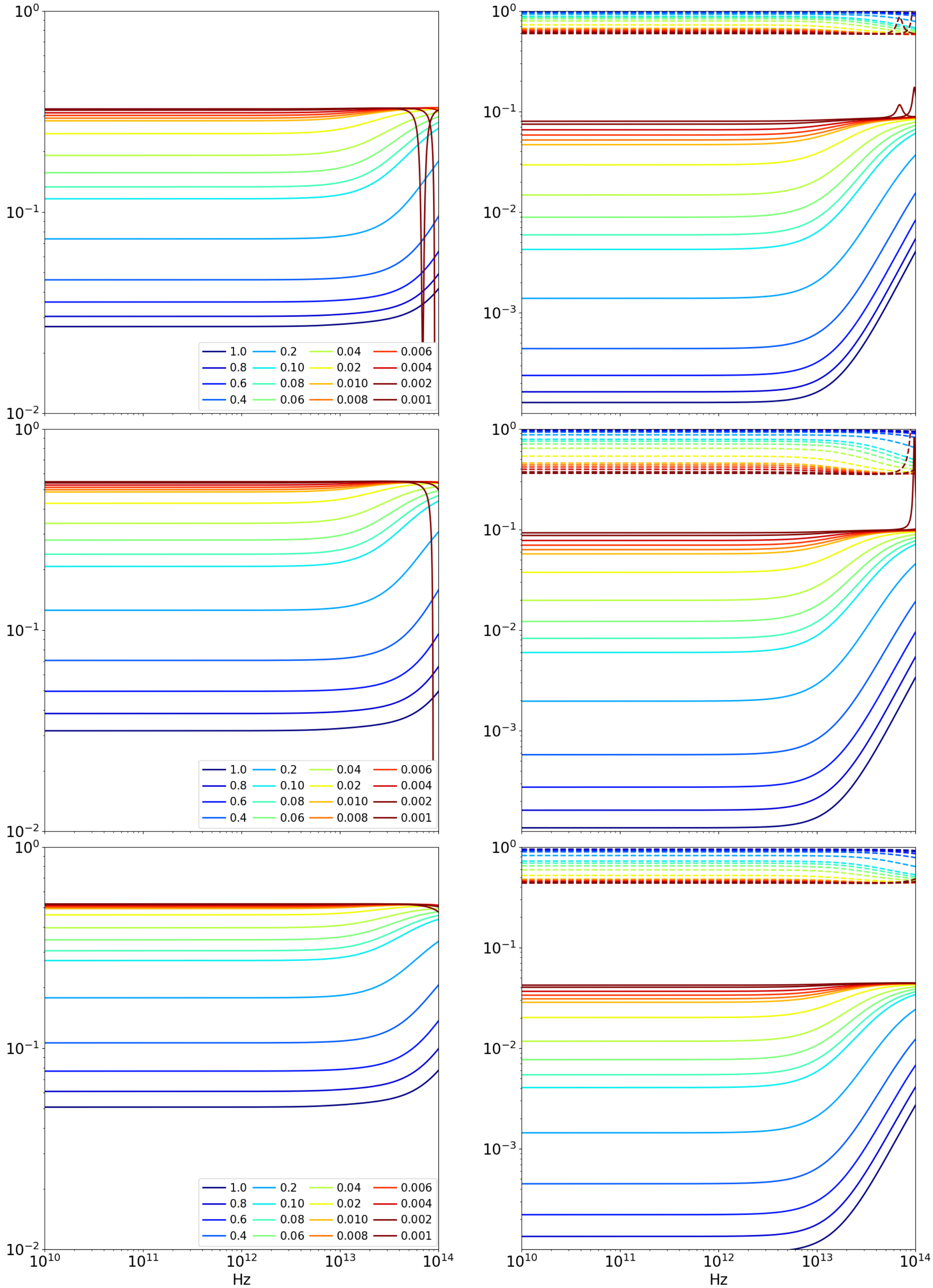


Figure 5.30: The same results of Fig. 5.29 but for 45° (top), 60° (middle), 75° (bottom) incidents, respectively.

removes the polar-angle dependence and further leverages the meshing enhancement relative to unidirectional meshes. Note that this approach does not remove the azimuthal-angle dependence for P-polarization due to variable in-plane projection, therefore also not removing the dependence for general incidence mixing S- and P-polarizations. For the DC resistance, although we might not be able to straightforwardly judge if the new patters preserve the original resistance as in Fig. 5.27, they are easy enough to calculate analytically or numerically based on the unit-cell geometries. Fig. 5.29/5.30 presents a phenomenological scan for the response of an isotropically meshed film for fill fractions in $f = 10^{-3}$–1.0. The calculation assumes the fill fraction applies equally to both ϵ_{xx} and ϵ_{yy} due to isotropic meshing, so $f\epsilon_{\mathrm{W}}$ applies to S-polarization regardless of the incident angle. With the incident azimuthal angles θ specified, particularly for P-polarization, we can determine the fractional energy $\cos^2\theta$ that has the E field projected into the meshed plane thus experiences $\epsilon = f\epsilon_{\mathrm{W}}$, while the rest remains with $\epsilon = \epsilon_{\mathrm{W}}$. Finally, assuming random polarization that introduces equal energies to the film via S- and P-polarizarions, we take the mean of the two to be the results presented in Fig. 5.29/5.30.

Fig. 5.29/5.30 shows our model predicts that a reasonably meshed film, e.g., $1/f = 10$–20, exhibits a 10–40% absorptance, which is an order-of-magnitude enhancement relative to the original value, echoing our earlier explanation for the $1/f$ enhancement. Within most incident angles, the angular variation of the absorptance is below a factor of ≈3, less variable if the film is meshed more aggressively. We again find a weak frequency dependence within our frequency range of interest following the results of Fig. 5.25 and Fig. 5.26. In addition to the above features that are mostly expected by our earlier analyses, Fig. 5.29/5.30 hints, at certain incident angles, there appears to be resonance-like structures close to 10^2 THz that suppress the absorptance. While in practice this frequency range is too high and therefore is only marginally relevant to our cryogenic application focus, we may still examine our model in this range as shown in Fig. 5.31. There are a few features to highlight. First, while we have concluded that the combined angular dependence is modest, Fig. 5.31 shows the meshing enhancement applies unequally to the S-polarization by order-of-magnitude hence relatively insignifies the P-polarization detection. We choose to present the result at $\theta = 30°$, so the difference, which we have explained is caused by the ineffectiveness in enhancing the P-polarization absorptance on the reduced in-plane component, may be visualized. We confirm the two polarizations are degenerate at normal incident, and the difference is consistently more severe at larger angles. Notice that this effect, namely the angular splitting for

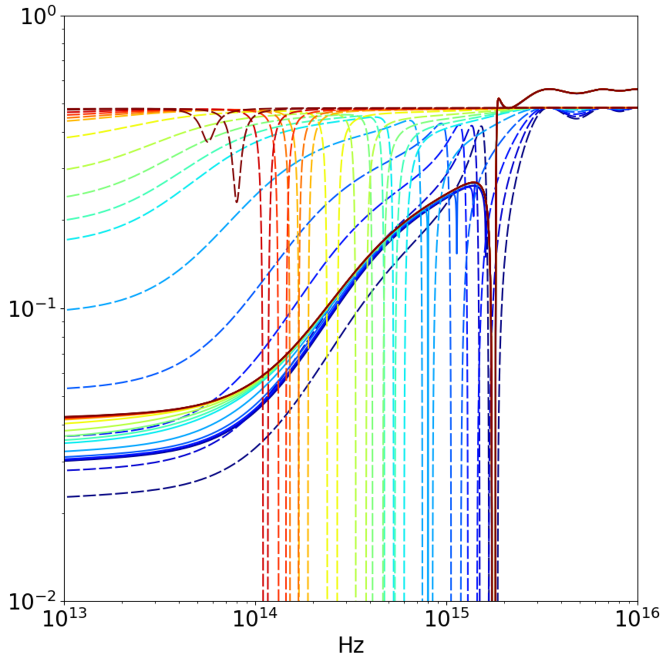


Figure 5.31: The same absorptance results of Fig. 5.29/5.30, based on the same polarization assumptions and color-coded identically, but presented separately for S (dashed)- and P (solid)-polarizations. We choose $\theta = 30°$ for this figure as an example. Detailed discussion is found in the corresponding text.

S-/P-polarization detection, extends into our frequency range of interest. Second, by comparing Fig. 5.31 to Fig. 5.29, we realize the resonance structures in Fig. 5.29 are the results of the resonantly suppressed S-polarization absorption.[20] These S-polarized resonances are sharper and exhibit higher resonant frequencies for less meshed films. If we examine their corresponding transmittances and reflectances as in Fig. 5.32, we realize the transmittances and reflectances undertake resonant enhancements at these frequencies that in turn suppress the absorptance. We believe the phenomenon is due to forming resonances between the front and back sides of the film, which constructively enhances the transmitted and/or reflected waves, while the bulk effect of the film is suppressed therefore the reduced absorption. Further studies are needed to understand the phenomenon if the meshed photon detector is to be adapted to near IR–visible light applications.[21]

[20] Which makes sense since we just argued the S-polarization dominates the absorptance at angles.

[21] In practice it is also more demanding in fabricating micro-meshes that satisfy the $< \lambda/5$ unit-cell requirement for shorter wavelengths.

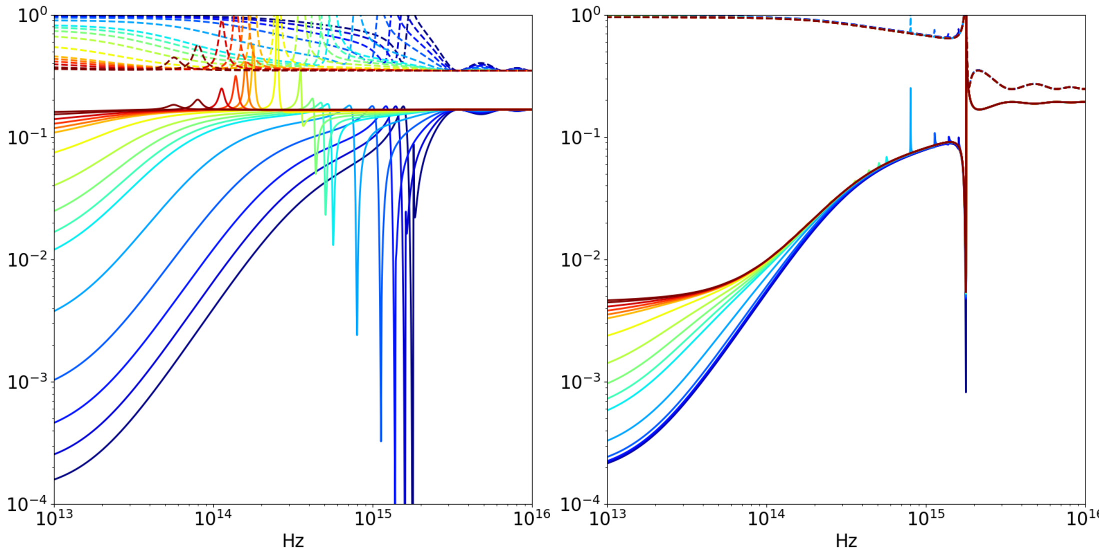


Figure 5.32: The transmittance (solid) and reflectance (dashed) for S-/P-polarizations (left/right), based on the same polarization assumptions and color-coded identically to Fig. 5.31. $\theta = 30°$ is chosen to be comparable with Fig. 5.31. Detailed discussion is found in the corresponding text.

We can now take a bottom-up approach to *redesign* the layout based on the analytical model. We first establish the unit cell pattern that exhibits the highest possible absorptance, calculate its equivalent resistance per square, and then arrange the cells to achieve the desired total resistance. We first list the hard constraints to narrow the design landscape:

1. Linewidth $w \geq 2\ \mu$m: To ensure high fabrication yield for the contact mask photolithography SuperCDMS uses, feature widths must not be smaller than 2 μm.
2. Unit cell size $\leq \lambda/5$: Based on (W. Jones, 2005)'s validation, we require the unit cell size be equal or smaller than 1/5 of the target wavelengths to reliably utilize the conductivity-dilution effect.

We apply the above constraints to a Cartesian-grid arrangement to obtain the most optimized design, i.e., the least filled design at a given unit-cell resistance, which turns out to be the left-hand side pattern in Fig. 5.28. Since we would like to apply the $\lambda/5$ constraint to both $\hat{x}$ and $\hat{y}$ directions to achieve the highest enhancement, the rectangular unit cell becomes a square of the side length of $\lambda/5$ with the minimal linewidth, yielding

$$f = \frac{w}{(\lambda/5)}. \tag{5.158}$$

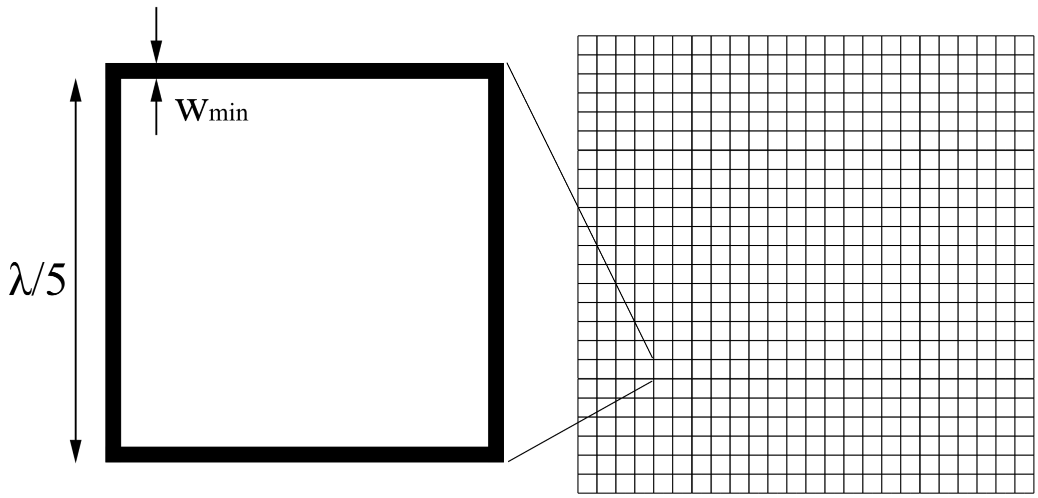


Figure 5.33: The optimal Cartesian-mesh design for the BB-photon TES detector, derived from the linewidth and cell size constraints. More extrapolation is provided in the corresponding text.

When these optimized single cells are combined, we obtain a $\lambda/5$ Cartesian grid with the minimal linewidth as shown in Fig. 5.33. Eq. (5.158) suggests that, when fixing the linewidth to the allowed minimum for an optimized (minimized) fill fraction, the $\lambda/5$ periodicity requirement makes the optimization frequency-dependent. Since the BB spectrum is a function of temperature, we find we need to optimize the photon sensor independently for different BB temperatures. For a design that is optimized for a target temperate, the pattern would be denser than the optimal $\lambda/5$ density for a lower temperature/longer wavelength and thus becomes under-optimized. At the same time, the pattern would be too sparse for higher temperatures and violates the effective homogeneous material criteria. In order to address this frequency-dependent optimization issue, we select three target BB temperatures, 4 K, 9 K, and 20 K, logarithmically spaced in the nominal BB temperature range, and design an array of meshed TESs optimized for each temperature. It is worth noting that, while a TES mesh may be under-optimized at temperatures lower than its design target, due to the strong T^4 BB radiation power scaling, the practical signal size is likely still large compared to that in the TES meshes optimized for lower temperatures. Therefore, in addition to be more effective than a single mesh design, this "color bank" design also provides the information for us to cross-compare the data to improve our simulation and mesh TES model in a frequency-/temperature-dependent fashion.

There are more practical design considerations: (cont'd previous list)

3. $R_{\text{total}} = 1/4\ \square$: To preserve the designed performance of the bias-power R&D TES (in the absence of parasitic power).
4. Chip size <2 cm for convenient mounting.

We realize that, in terms of the BB photon detection performance, once the unit cell design is determined and the pattern extends much wider than the wavelengths, there is little difference in how the cells are joined as long as the constraints listed above are satisfied. So, to standardize the process for a programmable design layout generation, we follow the methodology illustrated in Fig. 5.34 as our first attempt to bootstrap the cell arrangement and the bias scheme. We first select a reasonable size (q) for the chip and align the square cells into one row (n). Based on the geometric resistance of the square unit cell and in turn the row, we then determine the ratio for the numbers of these single rows to be connected in-series versus in-parallel ($m : (q/m)$), so the designated total resistance is obtained. In addition to the ratio, we further require the chip to have equal height and width,[22] i.e., roughly a square chip as shown in Fig. 5.34, so the absolute numbers for stacking the single rows in-series and then biasing the stacks in-parallel are also fixed. One can calculate the corresponding TES resistance as a function of a given w_{min} and the targeted λ. For a realistic design, we take $q = 1$ cm, $w = 2.5\ \mu$m, and target 4-K, 9-K, 20-K BB radiation peak λ, 725 μm, 324 μm, 145 μm, respectively. Also rounding the calculated dimensions to practical values, the design methodology yields total resistances of 0.258 $\square$, 0.260 $\square$, and 0.238 $\square$, respective; the detector layouts are shown in Fig. 5.35. We find, due to the rounding for integer numbers of cells and rows, the resulting resistances deviate from the targeted $R_{\text{TES}} = 0.25\ \square$ by about 3–5%, which is well within the uncertainties due to photolithographic resolution, film T_c, and σ. We therefor deem the result sufficiently accurate and the general design methodology viable.

However, it was pointed out by Kurinsky, who has focused on the design of the latest versions of SuperCDMS Tungsten TESs, that the design may be unstable for thermal-electrical biasing. The reason is that the resistive W traces between the isothermal superconducting Al nodes are atypically long, in some cases exceeding 1 mm (highlighted in Fig. 5.35) as opposed to the typical $O(10^2)$-μm lengths. When driven into transition on Joule heating, these long traces would exhibit large node-to-node resistivity gradients that may lead to normal/SC-phase separation, resulting in

[22] Or other specified aspect ratios; see later designs.

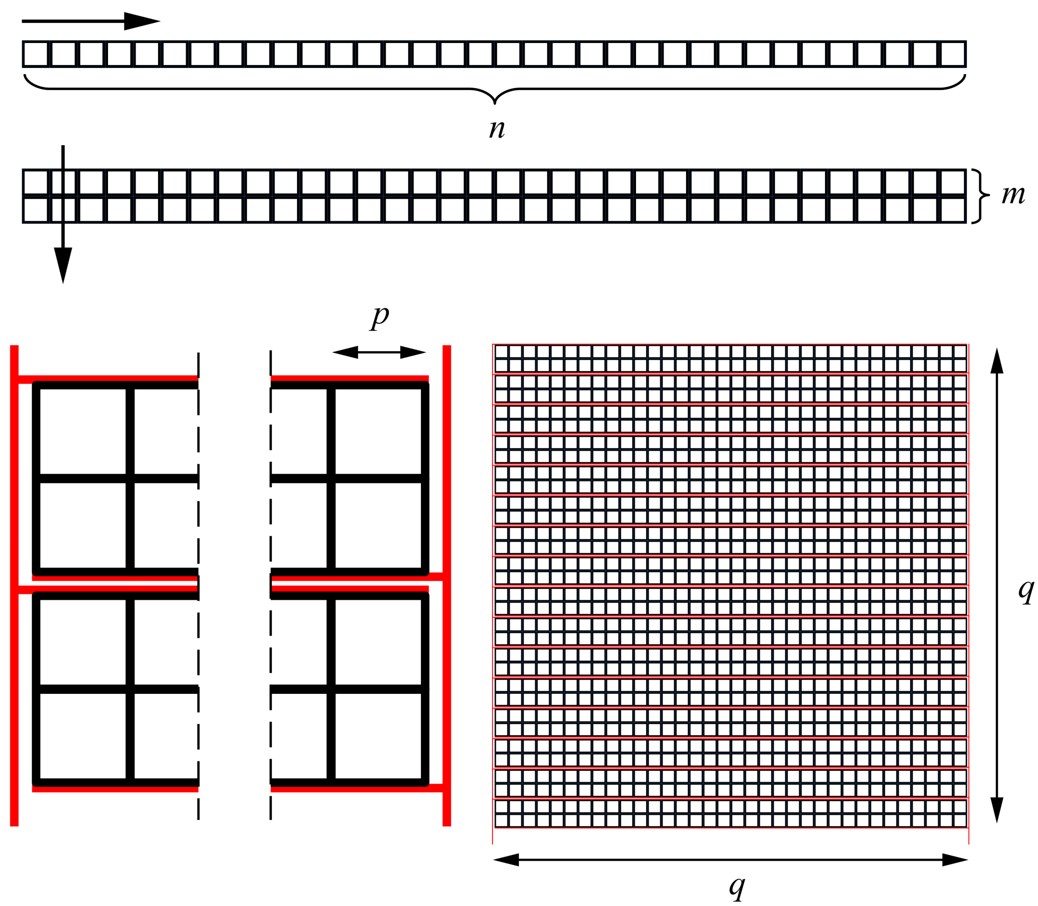


Figure 5.34: An illustration for the square unit cell arrangement bootstrapping process. From top, middle, bottom-left, to bottom-right, the cartoon illustrates first lining the cells to be a single row, stacking the rows, combining the stacked units with SC bias lines (red), and finally completing the full-chip design. A step-by-step description is given in the corresponding text.

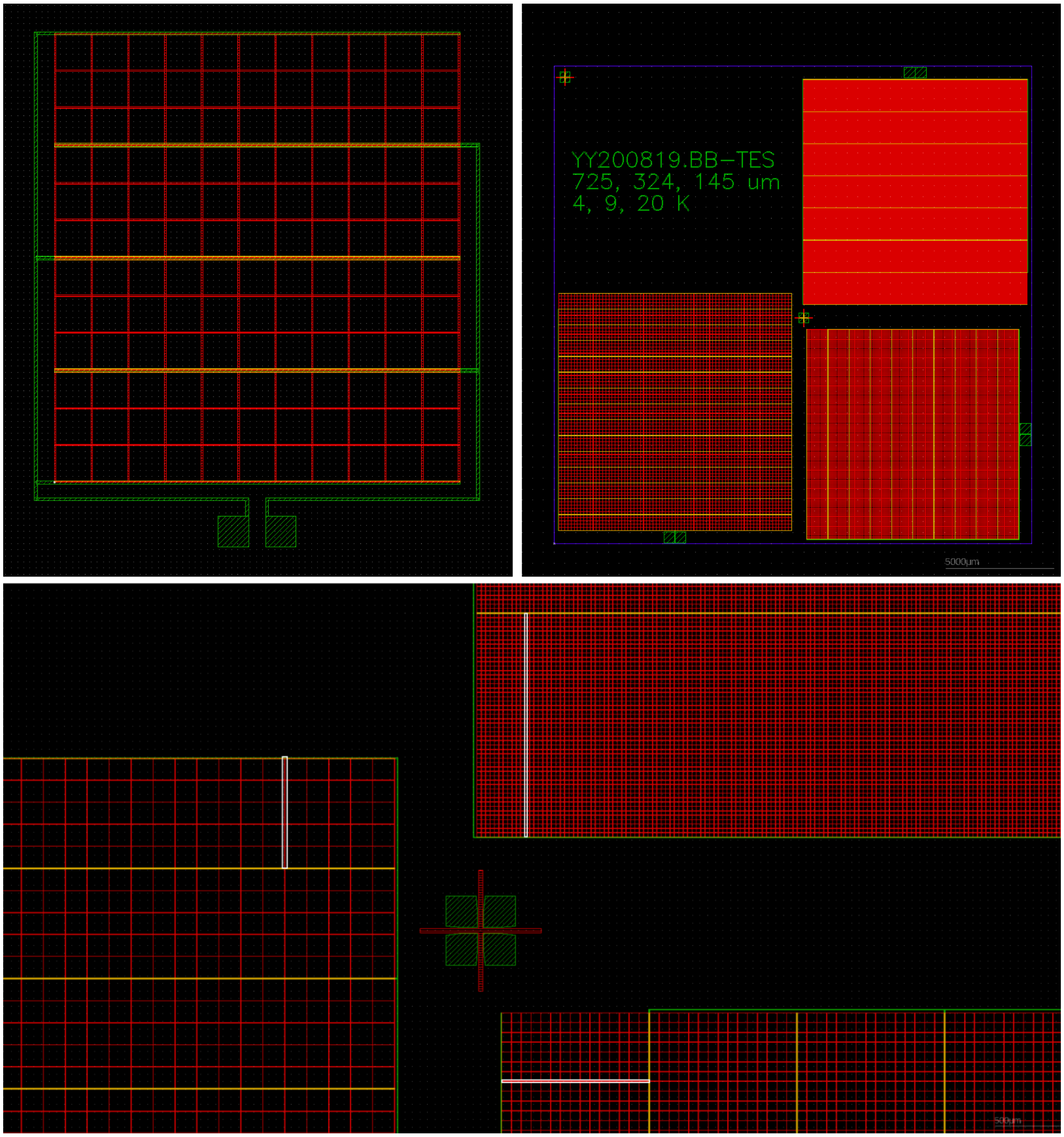


Figure 5.35: Our first meshed BB-photon TES chip design layout. The top-left figure is only for illustration purpose, unrealistically sparsened for visual clarity, and not optimized for any resistance or frequency. The top-right figure is the full **BB TES mk.1** layout targeting 4 K, 9 K, and 20 K, and the bottom figure is the zoom-in to the layout's central area for the details of the mesh design. The green and red regions represent the W and Al films, respectively, and the white boxes highlight the traces that should satisfy the maximal length constraint set by the thermal-electrical biasing phase separation condition.

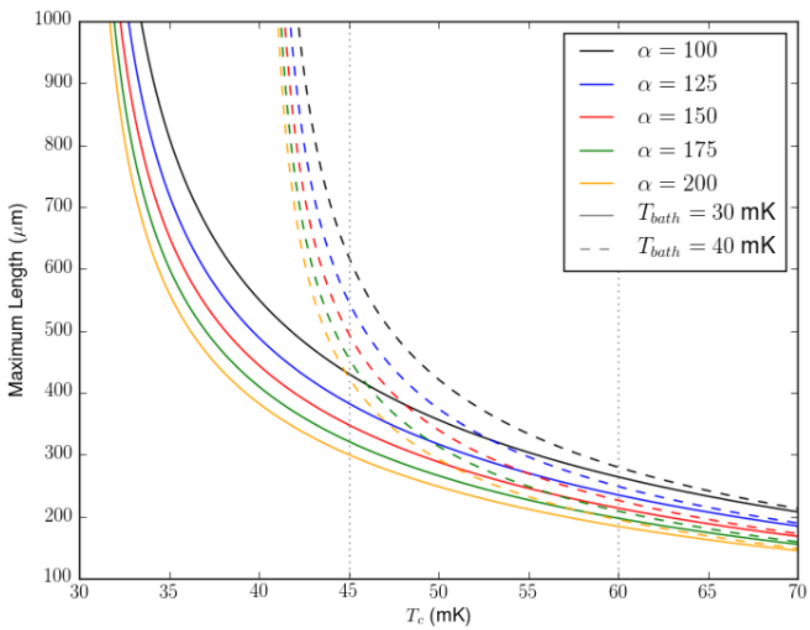


Figure 5.36: Critical trace lengths distinguishing stable/unstable equillibrium for W TES thermal-electrical bias, plotted for different logarithmic temperature sensitivities defined as $\alpha = \partial \log R / \partial \log T$. More discussions are provided in the corresponding text and (Kurinsky, 2018). Figure reproduced from (Kurinsky, 2018).

unstable in-transition resistivity equilibria that diverge on small perturbations. Fig. 5.36 shows the theoretical critical trace length for preventing such phase separation, as a function of T_c and the bath temperature T_{bath} assuming typical SuperCDMS W film. Fig. 5.36 shows, for typical higher-T_c W TESs, e.g., 60–70 mK, the preferred trace length is $\lesssim 150$ μm. If one brings T_c and T_{bath} closer by raising the bath temperature and/or lowering T_c, e.g., to 40–50 mK as in more recent SuperCDMS devices, the constraint becomes less stringent. We therefore further include the following design guidelines: (cont'd previous list)

5. Isothermal trace length <600 μm, an empirical[23] reliable criterion recommended by N. Kurinsky based on NEXUS DR performance and recent SuperCDMS W fabrication.
6. Removing the cm-long W traces perpendicular to the isothermal traces highlighted in Fig. 5.35, i.e., all horizontal black lines in Fig. 5.34.

While adopting the supposedly equipotential and therefore current-less long W traces per 6. makes the mesh pattern symmetric in both directions, they exceed our isothermal trace length criterion; they also contribute trivial resistances as being perpendicular to the bias current. Although in principle there should be no temperature or resistance gradient along these long traces, they are likely to create strong instability due to local thermal inhomogeneity and make the device overly sensitive to small environmental or bias condition disturbances.

With the horizontal lines removed, we are left with a design that is only meshed in one direction for the enhanced photon absorption. In order to maintain a more symmetric response to both x- and y-polarizations, we connect *in-parallel* two TES meshes rotated by 90° relative to one another, so the full design is insensitive to the polarization when the length scale of the field variation is wider than the TES area. The revision adds the design guidelines: (cont'd previous list)

7. Connecting the two polarized detectors in parallel at 90° relative to one another.
8. $R_{\mathrm{TES}} = 1/2\ \square$ for a single TES mesh, so to obtain the desired total resistance of 1/4 $\square$ in a parallel connection.

[23] Based on the consistent successes in fabricating $T_c < 50$ mK W films and the T_{bath} controllability to be close to T_c. Note that Fig. 5.36 only shows $T_{\mathrm{bath}} = 40$ mK as an example thus might appear to some readers suggesting a contradiction to the 600 μm criterion.

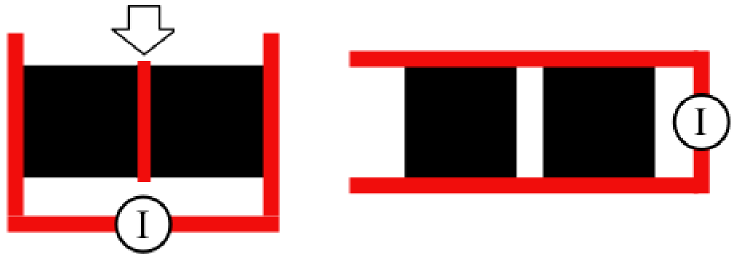


Figure 5.37: An illustration for the disconnected isothermal island in the series (left) mesh architecture, with the parallel architecture provided on the right for comparison. The black rectangles represent the meshes; the red lines represent the SC Al bias lines; the circled **I**s represent the biasing current sources, and the discounted island is highlighted by the arrow. More description is provided in the corresponding text.

While in principle the two polarized meshes may be connected in parallel or in series, the motivation favoring the parallel architecture is again the electrothermal biasing stability due to isothermal trace length. Since one can not guarantee that all the SC Al bias lines are constantly maintained at the same temperature, it is possible for disconnected "islands," shown in Fig. 5.37, to exhibit different temperatures. If so, the maximal W trace length between two isothermal nodes in the series architecture should be accounted for across the two meshes (Fig. 5.37).

Lastly, to take advantage of the exiting detector mounting and electronics infrastructure at NEXUS, we were instructed by the NEXUS team to confine the complete chip layout in a 1×1 cm^2 square area, for which we choose

9. Single square mesh width $q < 3$ mm,

so all the six square meshes (3 colors $\times$ 2 polarizations) may fit into the cm^2 area. One may argue that, since the radiation power per unit area is roughly a constant across the small detector area, while the bias power balancing the conductive cooling of the film also scales with the area[24], the sensitivity of the meshed photon detector, which is set by the relative power of the incident radiation to the electro-thermal bias, should only weakly depend on the absolute length scale (area) of the mesh design.

[24] We require the film feature variation length scale to be much smaller than the size of the entire TES, so the "film" area scales roughly with the size of the entire radiation-sensing area, including the cutouts in the mesh. Combining with the fact that the film thickness varies negligibly, the film volume then scales with the radiation-sensing area. Finally, the bias power scales with the "film" volume (Pyle, 2012), so the bias power consumed per unit "radiation-sensing" area is roughly a constant.

We then realize that the mesh resistance is also invariant[25] under this rescaling, if the number of the stacked-row units (the figure marked with m on the side in Fig. 5.34) connected in parallel is kept unchanged. It is due to the equal scaling of the length and the number of the isothermal W traces, where the former is proportional to the resistance along the bias current, the latter is the number of these W traces put in parallel, therefore the two cancel each other in the limit that the integer number of W traces is large hence scales roughly continuously with the mesh size. Based on the above arguments, we may flexibly resale the overall sizes of the meshes with little impact to the expected detector performance, as long as they remain much wider than the wavelengths of interest and meshed much more densely.

[25] Up to the rounding variation for integer numbers of cells and rows.

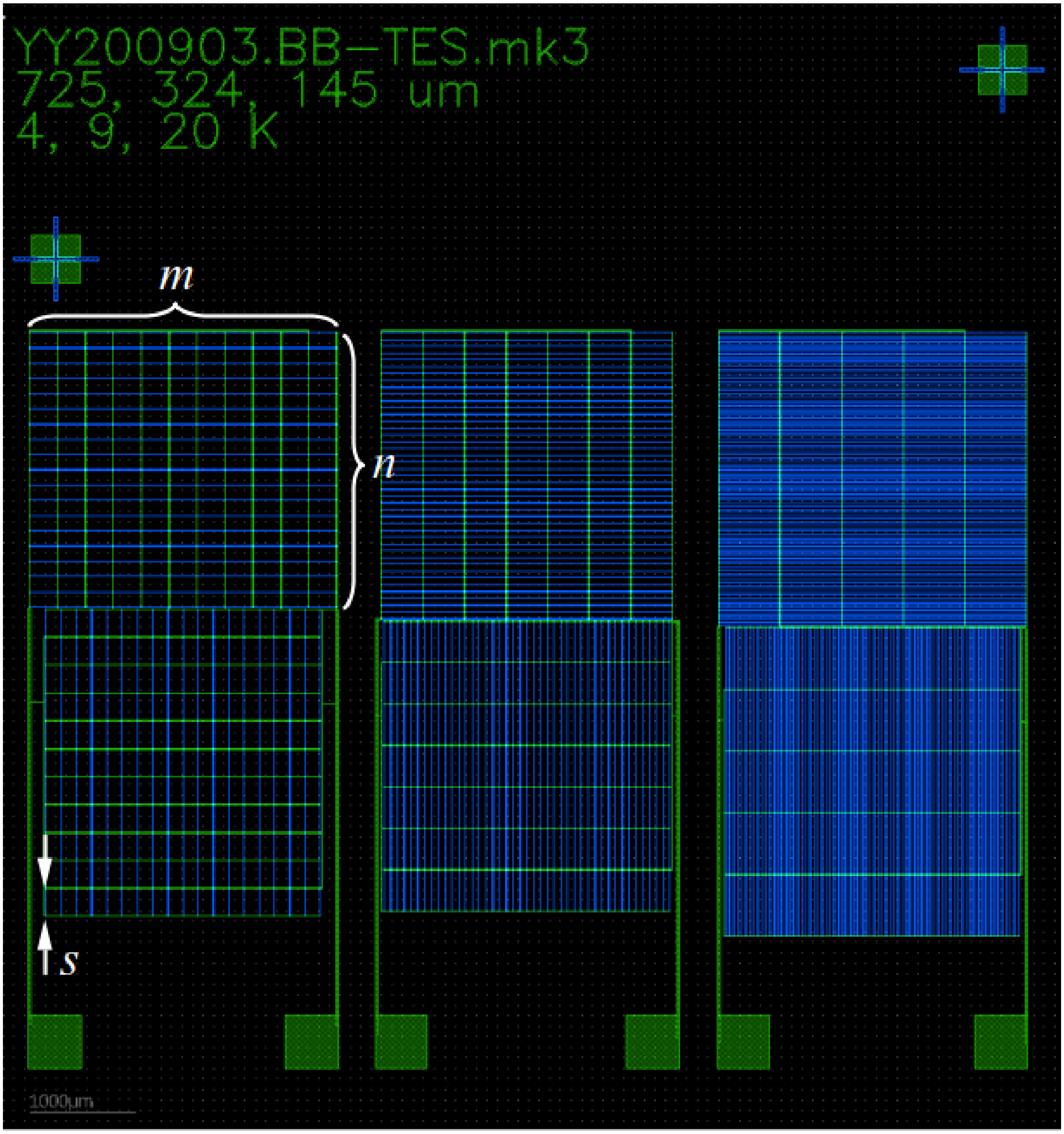


Figure 5.38: The 1×1 cm^2 **BB TES mk.3** layout, adapting the dual-polarization parallel-mesh connection for optimal (left to right) 4-K, 9-K, and 20-K peak BB emission detection. The blue and green regions represent the resistive W and SC Al films, respectively.

We arrive at our final mesh design with the aforementioned design rules:

$R_{\text{TES}} = 1/2\,\square$	4 K	9 K	20 K
s [μm]	278.54	416.66	622.84
n	20.6	46.3	103.4
m	10.8	7.2	4.8
$\hat{n}$	21	46	103
$\hat{m}$	11	7	5
$R_{\text{TES}} \parallel R_{\text{TES}}$	0.241	0.259	0.242

where s is the length of the isothermal W trace, n is the number of parallel W traces in one stacked-row unit – notice the long horizontal lines are now removed to make the stacked rows a "ladder" form (Fig. 5.38), and m is the number of stacked-row units connected in parallel. Note that the values are for single polarized meshes, i.e., $R_{\text{TES}} = 1/2\,\square$, and the hatted values are rounded from their un-hatted counterparts for practical layouts. Fig. 5.38 shows the final layout with the `.gds` file provided on (Chang, n.d.). The final resistances for the photon detectors are within 3–4% deviations from the targeted $1/4\,\square$, an expected result by the comparable variation in the previous symmetric mesh design (Fig. 5.35) and the fact that the removed long W traces contribute trivially to the resistance.

Based on the final design, Kurinsky raised the concern that, compared to typical SuperCDMS detectors, our meshed BB-photon TESs have atypically large amounts of bare W film, i.e., not covering or covered by SC Al. According to the bias power formula (Pyle, 2012)

$$P \propto \Sigma V_{\text{TES}}(T_c^n - T_{\text{bath}}^n), \tag{5.159}$$

where V_{TES} is the volume of bare W, we anticipate our meshed BB TESs to require higher bias powers to be driven into transition. To quantify the statement, the HVeV R2 detector in Fig. 2.1 only contains 0.01 mm^2 of bare W for each distributed TES channel, wheres for our final design in Fig. 5.38, the individual detectors have 1.2 mm^2 (4 K), 0.8 mm^2 (9 K), and 1.6 mm^2 (20 K) of bare W. In an appropriate TES operation, where one would always attempt to have the incoming signal to be about 1/10 of the bias power, so the signal does not dominate and subsequently alter the bias condition nor be trivialized by noise, Eq. (5.159) suggests our meshed BB TES detectors are designed for signal sizes that are orders-of-magnitude larger than HVeV R2. We calculate the expected bias power by knowing that the HVeV R2 NF-C design, which aims for $T_c \approx 75$ mK, empirically consumes 7 pW to be driven into transition on a 10 mK substrate. Assuming all the same factors from NF-C, e.g., film parameters, substrate temperature, etc., yet *tuning the transition temperature to 40 mK*[26] following more recent SuperCDMS fabrication results, the expected bias power for our BB TES is

$$\begin{aligned} P &= 7\ [\mathrm{pW}] \times \frac{(0.8 \sim 1.6)\ [\mathrm{mm}^2]}{0.01\ [\mathrm{mm}^2]} \times \frac{40^5 - 10^5\ [\mathrm{mK}^5]}{75^5 - 10^5\ [\mathrm{mK}^5]} \\ &= 24 \sim 48\ [\mathrm{pW}], \end{aligned} \tag{5.160}$$

in which we adopt $n \approx 5$ for the typical SuperCDMS W film electron-phonon coupling based on (Pyle, 2012). We find, based on the 1/10 stable signal-to-bias principle, our meshed detector is sensitive to an *absorbed* radiation power of a few pW. We should note, however, the simple scaling above is not fully valid in that we will be performing a DC measurement, as opposed to SuperCDMS detectors' nominal pulse detection. On one hand, the difference allows us to leverage the averaging to further suppress the noise relative to pulse detection, hence suggesting the estimate is conservative. On the other hand, it is possible that the low-frequency noise situation may be significantly worse than the OF-filtered pulse detection, which suggests the estimate might be optimistic.

Since we have explained earlier that the unit-area radiation power sensitivity does not scale with the detector area, by know the area is $\approx$16 mm^2 for each BB photon detector,[27] we can convert the sensitivity range derived from the bias power to 0.2–0.3 pW/mm^2. We can put this sensitivity into perspective by substituting it into the Stefan-Boltzmann law, which indicates a BB of 1.4–1.5 K for the same unit-area emission. The correspondence may be interpreted as: The detector is suitable for detecting the radiation from a perfect BB of an equal surface area at 1.4–1.5 K, *if the detector collects all the photon energy*. Our BB radiation source provides a surface area of $\approx$1000 times of each TES mesh detector. Assuming the TES has an absorptance of about 5‰ and using $\xi \geq 1$ for the photon loss rate between emission and sensor incidence, we then have the criterion

$$\left(\frac{T_{\mathrm{BB}}}{1.4\text{–}1.5\ \mathrm{K}}\right)^4 > \frac{1000 \cdot 5‰}{\xi} \tag{5.161}$$

for proper detection based on Stefan-Boltzmann law. For a typical range of ξ representing from poorly to heavily shielded by copper mechanical structures (no

[26] Right before this thesis was submitted, the BB TES mk.3 detectors were fabricated and obtained $T_c \approx 38$ mK.

[27] The "sensitive/sensor" area includes mesh cutouts.

dedicated sub/mm-wave absorbers), we obtain

ξ	10^2	10^3	10^4	10^5	10^6	10^7
T_{BB} [K]	0.3	0.5	0.9	1.6	2.8	5.1

.

We design the BB radiation source to be able to run a 20 K, so the resulting temperature range demonstrates the detector sensitivity is fully compatible with the BB radiation source with ample margin to accommodate unexpected signal/sensitivity degradation.

5.5 Future work

On July 22nd, 2021, all three BB TES mk.3 detectors were fabricated with the HVeV Run 3 detectors on a shared 4 mm-thick Si substrate, shown in Fig. 5.39. The obtained preliminary TES T_c was ≈38 mK. On the same wafer, multiple W film resistance characterizers were also fabricated for an explicit resistivity (conductivity) measurement. The obtained conductivity should be updated to the model detailed in Sec. 5.4.

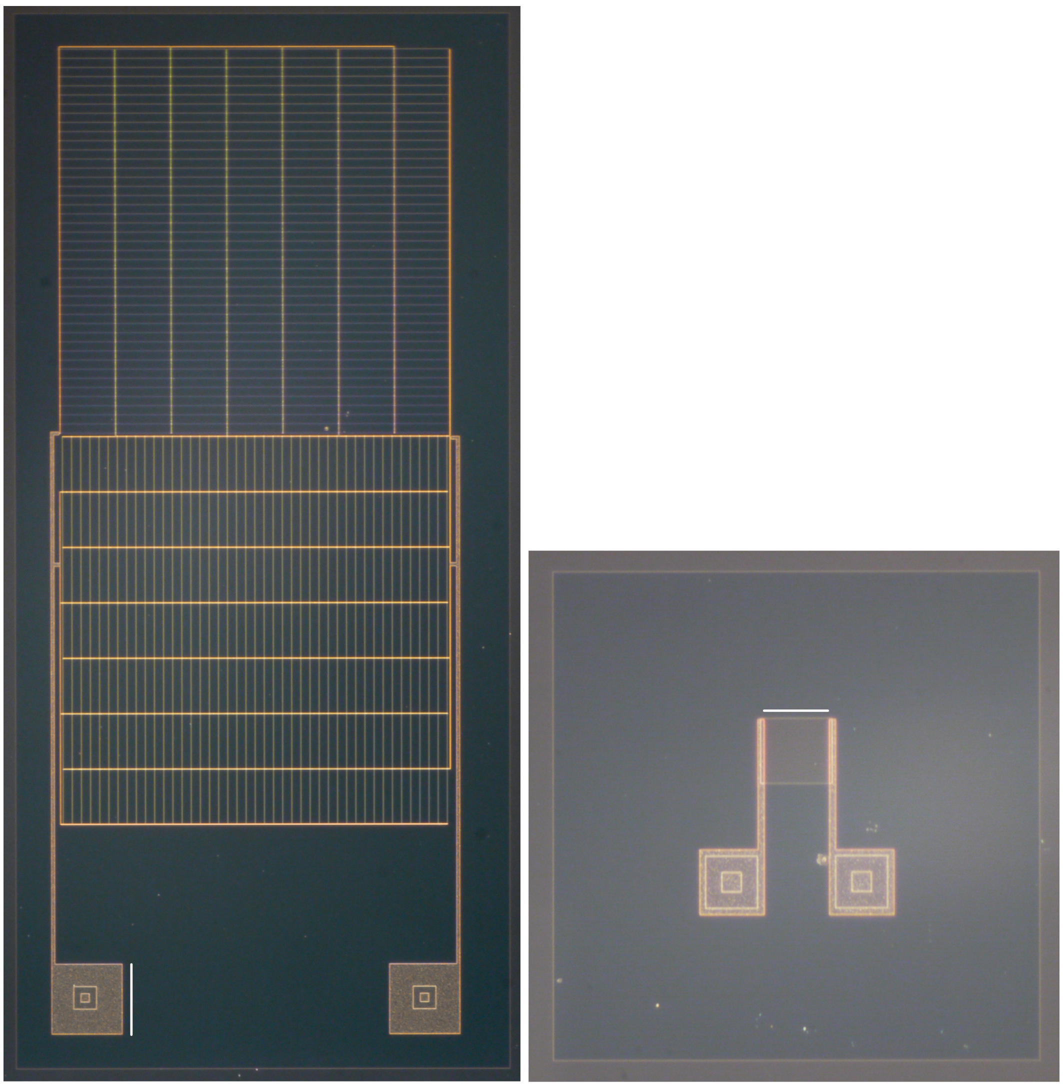


Figure 5.39: The photos for a 9-K BB TES mk.3 detector and a 2-lead resistance characterizer fabricated on the same substrate. The white scale bars represent 500 μm and 200 μm in the left- and right-hand side photos, respectively. In the photos, the brighter and darker features are the W and Al, respectively.

BIBLIOGRAPHY

[1] Y. Sofue and V. Rubin. “Rotation Curves of Spiral Galaxies”. In: *Annu. Revi. Astron. Astrop.* 39 (2001), p. 137.

[2] S. M. Bucklin. “A History of Dark Matter”. In: (2017). eprint: https://arstechnica.com/science/2017/02/a-history-of-dark-matter/.

[3] F. Zwicky. “Die Rotverschiebung von Extragalaktischen Nebeln”. In: *Helv. Phys. Acta.* 6 (1933).

[4] V. Rubin et al. “Rotational Properties of 21 SC Galaxies with a Large Range of Luminosities and Radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc)”. In: *Astrop. Jour.* 238 (1980), p. 471.

[5] M. Tanabashi et al. “Review of Particle Physics”. In: *Phys. Rev. D* 98 (2018), p. 030001.

[6] M. Markevitch et al. “A Textbook Example of a Bow Shock in the Merging Galaxy Cluster 1E 0657–56”. In: *Astrop. Jour.* 567.1 (2002).

[7] D. Clowe et al. “Weak-Lensing Mass Reconstruction of the Interacting Cluster 1E 0657–558: Direct Evidence for the Existence of Dark Matter”. In: *Astrop. Jour.* 604.2 (2004), p. 596.

[8] M. Markevitch et al. “Direct Constraints on the Dark Matter Self-Interaction Cross Section from the Merging Galaxy Cluster 1E 0657–56”. In: *Astrop. Jour.* 606.2 (2004), p. 819.

[9] S. Hong et al. “Revealing the Local Cosmic Web from Galaxies by Deep Learning”. In: *Astrop. Jour.* 913.1 (2021), p. 76.

[10] G. Gamow. “The Evolution of the Universe”. In: *Nature* 162 (1948).

[11] R. A. Alpher and R. Herman. “Evolution of the Universe”. In: *Nature* 162 (1948).

[12] A. A. Penzias and R. W. Wilson. “A Measurement of Excess Antenna Temperature at 4080 Mc/s”. In: *Astrop. Jour.* 142 (1965), p. 419.

[13] N. Aghanim et al. “Planck 2018 Results”. In: *Astron. Astrop.* 641.A1–12 (2020).

[14] A. Friedmann. “On the Possibility of a World with Constant Negative Curvature of Space (English ver.)” In: *Gene. Rela. Grav.* 31 (1999).

[15] The Daily Galaxy. 2015. eprint: https://dailygalaxy.com/.

[16] M. Battaglieri et al. *US Cosmic Visions: New Ideas in Dark Matter 2017: Community Report*. 2017. arXiv: 1707.04591.

[17] R. D. Pecci and H. R. Quinn. “CP Conservation in the Presence of Pseudoparticles”. In: *Phys. Rev. Lett.* 38 (1977), p. 1440.

[18] S. Weinberg. “A New Light Boson?” In: *Phys. Rev. Lett.* 40 (1978), p. 223.

[19] A. Ringwald. *Axions and Axion-Like Particles*. 2014. arXiv: 1407.0546.

[20] S. Sikivie and Q. Yang. “Bose-Einstein Condensation of Dark Matter Axions”. In: *Phys. Rev. Lett.* 103 (2009), p. 111301.

[21] D. N. Spergel and P. J. Steinhardt. “Observational Evidence for Self-Interacting Cold Dark Matter”. In: *Phys. Rev. Lett.* 84 (2000), p. 3760.

[22] D. E. Kaplan et al. “Asymmetric Dark Matter”. In: *Phys. Rev. D* 79 (2009), p. 115016.

[23] K. Zurek. “Asymmetric Dark Matter: Theories, signatures, and constraints”. In: *Phys. Repo.* 537 (2014), p. 91.

[24] Y. Hochberg et al. “Mechanism for Thermal Relic Dark Matter of Strongly Interacting Massive Particles”. In: *Phys. Rev. Lett.* 113 (2014), p. 171301.

[25] G. Jungman et al. “Supersymmetric Dark Matter”. In: *Phys. Repo.* 267 (1996), p. 195.

[26] S. Weinberg. “Varieties of Baryon and Lepton Nonconservation”. In: *Phys. Rev. D* 22 (1980), p. 1694.

[27] S. C. Chulia et al. “Seesaw Roadmap to Neutrino Mass and Dark Matter”. In: *Phys. Lett. B* 781 (2018), p. 122.

[28] K. Griest and M. Kamionkowski. “Unitarity Limits on the Mass and Radius of Dark-matter Particles”. In: *Phys. Rev. Lett.* 64 (1990), p. 615.

[29] E. Calabrese and D. N. Spergel. “Ultra-light Dark Matter in Ultra-faint Dwarf Galaxies”. In: *Mont. Noti. Roya. Astr. Soci.* 460 (4 2016), p. 4397.

[30] B. W. Lee and S. Weinberg. “Cosmological Lower Bound on Heavy-Neutrino Masses”. In: *Phys. Rev. Lett.* 39 (1977), p. 165.

[31] A. Garzilli et al. “How to Constrain Warm Dark Matter with the Lyman-α Forest”. In: *Mont. Noti. Roya. Astr. Soci.* 502 (2 2021), p. 2356.

[32] L. J. Feng et al. *Planning the Future of U.S. Particle Physics (Snowmass 2013): Chapter 4: Cosmic Frontier*. 2014. arXiv: 1401.6085.

[33] W. J. G. deBlok. “The Core-Cusp Problem”. In: *Adva. Astr.* 2010 (2010), p. 1.

[34] A. D. Rivero and C. Dvorkin. “Direct Detection of Dark Matter Substructure in Strong Lens Images with Convolutional Neural Networks”. In: *Phys. Rev. D* 101 (2020), p. 023515.

[35] D. Abercrombie et al. “Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum”. In: *Phys. Dark Univ.* 27 (2020), p. 100371.

[36] F. Tenchini et al. “Performance of High-Level Reconstruction at Belle II”. In: 18^{th} *Inte. Conf. B-Phys. Fron. Mach. - Beau. 2019*. 2019.

[37] M. Campajola et al. “Dark Sector First Results at Belle II”. In: *Phys. Scri.* 96.8 (2021).

[38] S. K. Lee et al. “Evidence for Unresolved γ-Ray Point Sources in the Inner Galaxy”. In: *Phys. Rev. Lett.* 116 (2016), p. 051103.

[39] M. D. Mauro et al. “Search for γ-ray Emission From Dark Matter Particle Interactions From the Andromeda and Triangulum Galaxies with the Fermi Large Area Telescope”. In: *Phys. Rev. D* 99 (2019), p. 123027.

[40] M. G. Aartsen et al. “Search for Dark Matter Annihilations in the Sun with the 79-String IceCube Detector”. In: *Phys. Rev. Lett.* 110 (2013), p. 131302.

[41] K. Abe et al. “Indirect Search for Dark Matter From the Galactic Center and Halo with the Super-Kamiokande Detector”. In: *Phys. Rev. D* 102 (2020), p. 072002.

[42] S. Murgia. “The Fermi–LAT Galactic Center Excess: Evidence of Annihilating Dark Matter?” In: *Annu. Rev. Nucl. Part. Scie.* 70 (2020), p. 455.

[43] M. Boudaud et al. “AMS-02 Antiprotons’ Consistency with a Secondary Astrophysical Origin”. In: *Phys. Rev. Rese.* 2 (2020), p. 023022.

[44] CAST Collaboration. “New CAST Limit on the Axion–photon Interaction”. In: *Natu. Phys.* 13 (2017).

[45] R. Ballou et al. “New Exclusion Limits on Scalar and Pseudoscalar Axionlike Particles from Light Shining Through a Wall”. In: *Phys. Rev. D* 92 (2015), p. 092002.

[46] L. Posti and A. Helmi. “Mass and Shape of the Milky Way’s Dark Matter Halo with Globular Clusters From Gaia and Hubble”. In: *Astron. Astrop.* 621.A56 (2019).

[47] J. D. Lewin and P. F. Smith. “Review of Mathematics, Numerical Factors, and Corrections Dark Matter Experiments Based on Elastic Nuclear Recoil”. In: *Astropar. Phys.* 6 (1996), p. 87.

[48] D. S. Akerib et al. “The LUX-ZEPLIN (LZ) Experiment”. In: *Nucl. Inst. Meth. Phys. Rese. Sec. A: Acce., Spec., Dete. Asso. Equi.* 953 (2020).

[49] R. Agnese et al. “Projected Sensitivity of the SuperCDMS SNOLAB Experiment”. In: *Phys. Rev. D* 95 (2017), p. 082002.

[50] A. Aguilar-Arevalo et al. “Constraints on Light Dark Matter Particles Interacting with Electrons From DAMIC at SNOLAB”. In: *Phys. Rev. Lett.* 123 (2020), p. 181802.

[51] R. Essig et al. “Direct Detection of Sub-GeV Dark Matter with Semiconductor Targets”. In: *Jour. High Ener. Phys.* 46 (2016).

[52] J. Billard et al. “Implication of Neutrino Backgrounds on the Reach of Next Generation Dark Matter Direct Detection Experiments”. In: *Phys. Rev. D* 89 (2014), p. 023524.

[53] Y. Hochberg et al. “Directional Detection of Dark Matter with Two-dimensional Targets”. In: *Phys. Lett. B* 772 (2017), p. 239.

[54] D. Akimov et al. “Observation of Coherent Elastic Neutrino-nucleus Scattering”. In: *Science* 357 (6356 2017).

[55] C. A. J O’Hare. *AxionLimits*. eprint: https://cajohare.github.io/AxionLimits/.

[56] A. Caputo et al. *Dark Photon Limits: A Cookbook*. 2021. arXiv: 2105.04565.

[57] B. Holdom. “Two U(1)’s and ϵ Charge Shifts”. In: *Phys. Lett. B* 166 (2 1986), p. 196.

[58] A. Derevianko et al. “Axio-electric Effect”. In: *Phys. Rev. D* 82 (2010), p. 065006.

[59] R. Agnese et al. “First Dark Matter Constraints from a SuperCDMS Single-Charge Sensitive Detector”. In: *Phys. Rev. Lett.* 122 (2019), p. 069901.

[60] R. Ren et al. *Design and Characterization of a Phonon-Mediated Cryogenic Particle Detector with an eV-Scale Threshold and 100 keV-Scale Dynamic Range*. 2020. arXiv: 2012.12430.

[61] D. W. Amaral et al. “Constraints on Low-mass, Relic Dark Matter Candidates From a Surface-operated SuperCDMS Single-charge Sensitive Detector”. In: *Phys. Rev. D* 102 (2020), p. 091101.

[62] F. Ponce et al. “Measuring the Impact Ionization and Charge Trapping Probabilities in SuperCDMS HVeV Phonon Sensing Detectors”. In: *Phys. Rev. D* 101 (2020), p. 031101.

[63] R. Agnese et al. “Low-mass Dark Matter Search with CDMSlite”. In: *Phys. Rev. D* 97 (2018), p. 022002.

[64] P. L. Brink. “Non-Equilibrium Superconductivity Induced by X-ray Photons”. PhD thesis. University of Oxford, Mar. 1995.

[65] B. Hines et al. “Flux-Coupled Direct Feedback in a SQUID Amplifier”. In: *IEEE Trans. App. Supercond.* 21 (3 2011), p. 262.

[66] L. A. Zadeh and J. R. Ragazzini. “Optimum Filters for the Detection of Signals in Noise”. In: *Proceedings of the IRE* 40.10 (1952), pp. 1223–1231. DOI: 10.1109/JRPROC.1952.274117.

[67] S. Golwala. "Exclusion Limits on the WIMP-Nucleon Elastic-Scattering Cross Section from the Cryogenic Dark Matter Search". PhD thesis. University of California at Berkeley, June 2000.

[68] M. C. Pyle. "Optimizing the Design and Analysis of Cryogenic Semiconductor Dark Matter Detectors for Maximum Sensitivty". PhD thesis. Stanford University, June 2012.

[69] N. Kurinsky. "The Low-mass Limit: Dark Matter Detectors with eV-scale Energy Resolution". PhD thesis. Stanford University, Aug. 2018.

[70] Z. Hong et al. "NEXUS@FNAL". In: 18^{th} *Inte. Work. Low Temp. Dete.* 2019.

[71] M. Wilson. "A New Search for Low-mass Dark Matter and an Examination and Reduction of the Uncertainty Due to the Photoelectric Absorption Cross Section using a Cryogenic Silicon Detector with Single-charge Sensitivity". PhD thesis. University of Toronto, July 2021.

[72] P. Shukla et al. *Energy and Angular Distributions of Atmospheric Muons at the Earth*. 2016. arXiv: 1606.06907.

[73] M. A. Thomason. "Lecture Notes on Statistics". In: *University of Cambridge* https://www.hep.phy.cam.ac.uk/ thomson/lectures/statistics (2015).

[74] Y.-Y. Chang. "HVeV Run 2 Final Spectrum & Resolution". In: *SuperCDMS Confluence Home Page (internal)* (2019). eprint: https://confluence.slac.stanford.edu/display/CDMS/Home.

[75] F. Ponce et al. "Modeling of Impact Ionization and Charge Trapping in SuperCDMS HVeV Detectors". In: *Jour. Low Temp. Phys.* 199 (2020).

[76] D. W. Amaral et al. *Constraints on Low-mass, Relic Dark Matter Candidates from a Surface-operated SuperCDMS Single-charge Sensitive Detector*. 2020. arXiv: 2005.14067.

[77] R. Essig et al. "Direct Detection of sub-GeV Dark Matter". In: (). eprint: http://ddldm.physics.sunysb.edu/ddlDM/.

[78] Y. Hochberg et al. "Absorption of Light Dark Matter in Semiconductors". In: *Phys. Rev. D* 95 (2017), p. 023013.

[79] I. M. Bloch et al. "Searching for Dark Absorption with Direct Detection Experiments". In: *Jour. High Ener. Phys.* 87 (2017).

[80] J. L. Jr. Hodges. "The Significance Probability of the Smirnov Two-Sample Test". In: *Arki. Mate.* 3.43 (1958).

[81] Y.-Y. Chang. "HVeV Run 2 Stage-1 Unblinding Test". In: *SuperCDMS Confluence Home Page (internal)* (2019). eprint: https://confluence.slac.stanford.edu/display/CDMS/Home.

[82] Y.-Y. Chang. "HVeV Run 2 stage-2 Unblinding Test". In: *SuperCDMS Confluence Home Page (internal)* (2019). eprint: https://confluence.slac.stanford.edu/display/CDMS/Home.

[83] S. Yellin. "Finding an Upper Limit in the Presence of an Unknown Background". In: *Phys. Rev. D* 66 (2002), p. 032005.

[84] S. Yellin. *Extending the Optimum Interval Method*. 2007. arXiv: 0709.2701.

[85] HVeV Run 2 team. "HVeV Run 2 Summary Documentation v2". In: *SuperCDMS Confluence Home Page (internal)* (2020). eprint: https://confluence.slac.stanford.edu/display/CDMS/Home.

[86] HVeV Run 2 team. "SuperCDMS Public Documentation - HVeV Run 2". In: (2020). eprint: https://www.slac.stanford.edu/$\sim$zqhong06/v5/HVeVR2_PublicDocumentation.html.

[87] O. Abramoff et al. "SENSEI: Direct-Detection Constraints on Sub-GeV Dark Matter from a Shallow Underground Run Using a Prototype Skipper CCD". In: *Phys. Rev. Lett.* 122 (2020), p. 161801.

[88] R. Essig et al. "New Constraints and Prospects for Sub-GeV Dark Matter Scattering Off Electrons in Xenon". In: *Phys. Rev. D* 96 (2017), p. 043017.

[89] R. Essig et al. "First Direct Detection Limits on Sub-GeV Dark Matter from XENON10". In: *Phys. Rev. Lett.* 109 (2012), p. 021301.

[90] E. Aprile et al. "Light Dark Matter Search with Ionization Signals in XENON1T". In: *Phys. Rev. Lett.* 123 (2019), p. 251801.

[91] T. Aralis et al. "Constraints on Dark Photons and Axionlike Particles from the SuperCDMS Soudan Experiment". In: *Phys. Rev. D* 101 (2020), p. 052008.

[92] I. M. Bloch et al. "Searching for Dark Absorption with Direct Detection Experiments". In: *Jour. High Ener. Phys.* 87 (2017).

[93] H. An et al. "Direct Detection Constraints on Dark Photon Dark Matter". In: *Phys. Lett. B* 747 (2020), p. 331.

[94] V. Viaux et al. "Neutrino and Axion Bounds from the Globular Cluster M5 (NGC 5904)". In: *Phys. Rev. Lett.* 111 (2013), p. 231301.

[95] M. M. M. Bertolami et al. "Revisiting the Axion Bounds from the Galactic White Dwarf Luminosity Function". In: *Jour. Cosm. Astrop. Phys.* 2014.10 (2014).

[96] P. Du et al. *Sources of Low-Energy Events in Low-Threshold Dark Matter Detectors*. 2020. arXiv: 2011.13939.

[97] N. Kurinsky et al. "Dark Matter Interpretation of Excesses in Multiple Direct Detection Experiments". In: *Phys. Rev. D* 102 (2020), p. 015017.

[98] P. K. Day et al. "A Broadband Superconducting Detector Suitable for Use in Large Arrays". In: *Nature* 425 (2003).

[99] S. Doyle. “Lumped Element Kinetic Inductance Detectors”. PhD thesis. Cardiff University, Apr. 2008.

[100] D. C. Moore et al. “Position and Energy-resolved Particle Detection Using Phonon-mediated Microwave Kinetic Inductance Detectors”. In: *Appl. Phys. Lett.* 100 (2012), p. 232601.

[101] J. D. Jackson. *Classical Electrodynamics*. 3rd. Wiley, 1998.

[102] SuperCDMS Collaboration. “SuperCDMS SNOLAB Technical Design Report”. In: (2018).

[103] W. T. Doughty. “SuperCDMS Soudan High Threshold WIMP Search”. PhD thesis. University of California, Berkeley, June 2016.

[104] M. H. Abitbol et al. *CMB-S4 Technology Book, First Edition*. 2017. arXiv: 1706.02464.

[105] J. Gao. “The Physics of Superconducting Microwave Resonators”. PhD thesis. California Institute of Technology, May 2008.

[106] D. C. Mattis and J. Bardeen. “Theory of the Anomalous Skin Effect in Normal and Superconducting Metals”. In: *Phys. Rev.* 111.2 (1958), p. 412.

[107] J. Zmuidzinas. “Superconducting Microresonators: Physics and Applications”. In: *Annu. Rev. Cond. Matt. Phys.* 3 (2012), p. 169.

[108] J. Gao et al. “Experimental Evidence for a Surface Distribution of Two-level Systems in Superconducting Lithographed Microwave Resonators”. In: *Appl. Phys. Lett.* 92 (2008), p. 152505.

[109] W. A. Phillips. “Tunneling States in Amorphous Solids”. In: *Jour. Low Temp. Phys.* 7 (1972), p. 351.

[110] J. Zmuidzinus. “Thermal Noise and Correlations in Photon Detection”. In: *Appl. Opti.* 42.25 (2003), p. 351.

[111] P. D. Mauskopf et al. “Photon-noise Limited Performance in Aluminum LEKIDs”. In: *Jour. Low Temp. Phys.* 176 (2014), p. 545.

[112] A. K. Wandui et al. “Antenna-coupled Thermal Kinetic Inductance Detectors for Ground-based Millimeter-wave Cosmology”. In: *Astr. Tele. + Inst., 2020, Proc. of SPIE* 11453 (2020).

[113] P. J. de Visser et al. “Number Fluctuations of Sparse Quasiparticles in a Superconductor”. In: *Phys. Rev. Lett.* 106 (2011), p. 167004.

[114] J. Baselmans et al. “Ultra Low Background Cryogenic Test Facility for Far-Infrared Radiation Detectors”. In: *Jour. Low Temp. Phys.* 167 (2012), p. 360.

[115] P. J. de Visser et al. “Fluctuations in the Electron System of a Superconductor Exposed to a Photon Flux”. In: *Natu. Comm* 5 (2014), p. 3130.

[116] O. Noroozian et al. "Crosstalk Reduction for Superconducting Microwave Resonator Arrays". In: *IEEE Tran. Micr. Theo. Tech.* 60.5 (2012), p. 1235.

[117] B. A. Mazin et al. "A Superconducting Focal Plane Array for Ultraviolet, Optical, and Near-infrared Astrophysics". In: *Opti. Expr.* 20 (5 2012), p. 1503.

[118] M. S. Khalil et al. "An Analysis Method for Asymmetric Resonator Transmission Applied to Superconducting Devices". In: *Jour. Appl. Phys.* 111 (2012), p. 054510.

[119] B. D. Cornell. "A Dark Matter Search Using the Final SuperCDMS Soudan Dataset and the Development of a Large-Format, Highly Multiplexed, Athermal-Phonon-Mediated Particle Detector". PhD thesis. California Institute of Technology, May 2018.

[120] D. C. Moore. "A Search for Low-Mass Dark Matter with the Cryogenic Dark Matter Search and the Development of Highly Multiplexed Phonon-Mediated Particle Detectors". PhD thesis. California Institute of Technology, May 2012.

[121] E. S. Battistelli et al. "CALDER - Neutrinoless Double-beta Decay Identification in TeO_2 Bolometers with Kinetic Inductance Detectors". In: *Euro. Phys. Jour. C* 75 (2015), p. 353.

[122] L. Cardani et al. "Al/Ti/Al Phonon-mediated KIDs for UV–vis Light Detection Over Large Areas". In: *Supercon. Scie. Tech.* 31.7 (2018), p. 075002.

[123] Y.-Y. Chang et al. "Development of a Massive, Highly Multiplexible, Phonon-Mediated Particle Detector Using Kinetic Inductance Detectors". In: *Jour. Low Temp. Phys.* 193 (2018), p. 1199.

[124] G. Jones et al. "High Quality Factor Manganese-doped Aluminum Lumped-element Kinetic Inductance Detectors Sensitive to Frequencies Below 100 GHz". In: *Appl. Phys. Lett.* 120 (2017), p. 222601.

[125] A. B. Walter et al. "The MKID Exoplanet Camera for Subaru SCExAO". In: *Inst. Meth. Astrphys.* 132.1018 (2020), p. 125005.

[126] B. A. Mazin et al. *Superconducting Materials for Microwave Kinetic Inductance Detectors*. 2020. arXiv: 2004.14576.

[127] L. Grunhaupt et al. "Loss Mechanisms and Quasiparticle Dynamics in Superconducting Microwave Resonators Made of Thin-Film Granular Aluminum". In: *Phys. Rev. Lett.* 121 (2018), p. 117001.

[128] F. Valenti et al. "Interplay Between Kinetic Inductance, Nonlinearity, and Quasiparticle Dynamics in Granular Aluminum Microwave Kinetic Inductance Detectors". In: *Phys. Rev. Appl.* 11 (2018), p. 054087.

[129] L. Cardani et al. *Final Results of CALDER: Kinetic Inductance Light Detectors to Search for Rare Events*. 2021. arXiv: 2104.06850.

[130] L. Minutolo et al. “A Flexible GPU-Accelerated Radio-frequency Readout for Superconducting Detectors”. In: *IEEE Tran. Appl. Supercon.* 29.5 (2019).

[131] S. McHugh et al. “A Readout for Large Arrays of Microwave Kinetic Inductance Detectors”. In: *Revi. Scie. Inst.* 83 (2012).

[132] S. R. Siegel. “A Multiwavelength Study of the Intracluster Medium and the Characterization of the Multiwavelength Sub/millimeter Inductance Camera”. PhD thesis. California Institute of Technology, June 2016.

[133] M. A. Castellanos-Beltran. “Development of a Josephson Parametric Amplifier for the Preparation and Detection of Nonclassical States of Microwave Fields”. PhD thesis. University of Colorado at Boulder, June 2010.

[134] C. Macklin et al. “A Near–quantum-limited Josephson Traveling-wave Parametric Amplifier”. In: *Scie.* 350 (6258 2016).

[135] N. Klimovich et al. “Superconducting Kinetic Inductance Parametric Amplifiers”. In: 18^{th} *Inte. Work. Low Temp. Dete.* 2019.

[136] X. Liu et al. “Superconducting Micro-resonator Arrays with Ideal Frequency Spacing”. In: *Appl. Phys. Lett.* 111 (2017), p. 252601.

[137] G. Coiffard et al. “Parallel Plate Microwave Kinetic Inductance Detectors”. In: 17^{th} *Inte. Work. Low Temp. Dete.* 2017.

[138] S. Beldi et al. “High Q-factor Near Infrared and Sisible Al_2O_3-based Parallel-plate Capacitor Kinetic Inductance Detectors”. In: *Opti. Expr.* 27 (9 2019), p. 13319.

[139] S. R. Golwala et al. “RF Loss Tangent and Two-Level-System Noise of Amorphous Silicon and Crystalline Silicon Dielectrics for Sub mm Astronomy Applications”. In: 18^{th} *Inte. Work. Low Temp. Dete.* 2019.

[140] R. N. Simons. *Coplanar Waveguide Circuits, Components, and Systems*. Wiley, 2001.

[141] B. C. Wadell. *Transmission Line Design Handbook*. Artech House, 1991.

[142] G. Ghione et al. “Analytical Formulas for Coplanar Lines in Hybrid and Monolithic MICs”. In: *Elec. Lett.* 20 (4 1984), p. 179.

[143] G. Ghione et al. “Coplanar Waveguides for MMIC Applications: Effect of Upper Shielding, Conductor Backing, Finite-Extent Ground Planes, and Line-to-Line Coupling”. In: *IEEE Tran. Micr. Theo. Tech.* 35.3 (1987), p. 260.

[144] Microwaves101.com. *Coplanar Waveguide Calculator*. 2001. eprint: https://www.microwaves101.com.

[145] F. Darwis et al. “10 GHz Standing-Wave Coplanar Stripline on $LiNbO_3$ Crystal for Radio to Optical-Wave Conversion”. In: *Jour. Phys.: Conf. Seri.* 101 (2018), p. 012050.

[146] I-Laboratory. *Coplanar Strips Calculator*. 2000. eprint: www17.plala.or.jp/i-lab/.

[147] SONNET. eprint: https://www.sonnetsoftware.com/.

[148] Photo Science Inc. eprint: https://www.photo-sciences.com/.

[149] Ansys. eprint: https://www.ansys.com/products/electronics/ansys-hfss.

[150] Y.-Y. Chang. eprint: https://www.github.com/Yen-Yung.

[151] gdsCAD. eprint: https://pythonhosted.org/gdsCAD.

[152] D. S. Wisbey et al. "New Method for Determining the Quality Factor and Resonance Frequency of Superconducting Micro-Resonators from Sonnet Simulations". In: *Jour. Low Temp. Phys.* 176 (2014).

[153] K. D. Irwin. "Phonon-mediated Particle Detection using Superconduting Tungsten Transition-edge Sensors". PhD thesis. Stanford University, Apr. 1995.

[154] N. W. Ashcroft and N. D. Mermin. *Solid State Physics*. Cengage Learning, 2021.

[155] Geant4. eprint: https://geant4.web.cern.ch.

[156] S. B. Kaplan et al. "Quasiparticle and Phonon Lifetimes in Superconductors". In: *Phys. Rev. B* 14.11 (1976), p. 4854.

[157] Eccosorb. eprint: https://www.laird.com/.

[158] Wikipedia contributors. *Black Silicon*. Wikipedia, The Free Encyclopedia, 2021. eprint: https://en.wikipedia.org/wiki/Black_silicon.

[159] Y.-Y. Chang et al. "A Time- and Amplitude-controllable Technique for Measuring Energy Resolution and other Properties of KID-based Phonon-mediated Particle Detectors". In: 18^{th} *Inte. Work. Low Temp. Dete.* 2019.

[160] S. Knapen et al. "Detection of Light Dark Matter with Optical Phonons in Polar Materials". In: *Phys. Lett. B* 785.10 (2017), p. 386.

[161] N. Kurinsky et al. "Diamond Detectors for Direct Detection of Sub-GeV Dark Matter". In: *Phys. Rev. D* 99 (2019), p. 123005.

[162] S. R. Golwala. "Expected Energy Resolution of a Phonon-Mediated Detector using TiN MKIDs". In: *Unpublished* (2010).

[163] J. Zmuidzinas. "Amplifier-limited MKID Energy Resolution: Prospects for Ultransensitive Photon-Counting Detectors". In: *Unpublished* (2009).

[164] Cosmic Microwave Technology Inc. eprint: https://www.cosmicmicrowavetechnology.com/.

[165] Ettus Research. eprint: https://files.ettus.com/schematics/cbx/CBX.pdf.

[166] W. A. Phillips. "Two-level States in Glasses". In: *Repo. Prog. Phys.* 50 (1987), p. 1657.

[167] M. Virginia et al. *Localization and Reduction of Superconducting Quantum Coherent Circuit Losses*. 2020. arXiv: 2012.07604.

[168] K. Ramanathan et al. "Performance of a Multi-KID Phonon-mediated Particle Detector". In: 19^{th} *Inte. Work. Low Temp. Dete.* 2021.

[169] N. Zobrist et al. "Wide-band Parametric Amplifier Readout and Resolution of Optical Microwave Kinetic Inductance Detectors". In: *Appl. Phys. Lett.* 115 (2019), p. 042601.

[170] B. H. Eom et al. "A Wideband, Low-noise Superconducting Amplifier with High Dynamic Range". In: *Natu. Phys.* 8 (2012).

[171] M. C. Diez et al. "Reflectance Measurements on Submillimeter Absorbing Coatings for HIFI". In: *UV, Opti., IR Spac. Tele. Inst.* 4013 (2000).

[172] P. C. Hargrave et al. "A Low-background ^{3}He Bolometer Array Test Facility and its Use in Evaluating Prototype Arrays for FIRST-SPIRE". In: *Nucl. Inst. Meth. Phys. Rese. Sec. A: Acce., Spec., Dete. Asso. Equi.* 444 (2000).

[173] J. W. Lamb. "Miscellaneous Data on Materials for Millimetre and Submillimetre Optics". In: *Inte. Jour. Infr. Mill. Wave* 17 (1996).

[174] J. Domange et al. "Thermally-Stimulated Current Investigation of Dopant-Related D^- and A^+ Trap Centers in Germanium for Cryogenic Detector Applications". In: *Jour. Low Temp. Phys.* 167 (2012).

[175] D. Bergstrom. "The Absorption of Laser Light by Rough Metal Surfaces". PhD thesis. Lulea University of Technology, Aug. 2008.

[176] M Lax. "Cascade Capture of Electrons in Solids". In: *Phys. Rev.* 119 (1960), p. 1502.

[177] R. Shankar. *Principles of Quantum Mechanics*. 2nd. Plenum Press.

[178] J. J. Sakurai. *Advanced Quantum Mechanics*. 1st. Addison-Wesley.

[179] Aspnes. "Local-field Effects and Effective-medium Theory: A Microscopic Perspective". In: *Appl. Phys. Lett.* 50 (1982), p. 704.

[180] B. J Van Zeghbroeck. "Principles of Semiconductor Devices". In: *University of Colorado at Boulder* (1997). eprint: http://ecee.colorado.edu/$\sim$bart/book/titleold.htm.

[181] A. T. J. Phipps. "Ionization Collection in Detectors of the Cryogenic Dark Matter Search". PhD thesis. University of California at Berkeley, Apr. 2016.

[182] P. J. W. Debye. *Polar Molecules*. Chem. Cata. Comp., Inc., 1929.

[183] L. Onsager. "Eletric Moments of Molecules in Liquids". In: *Jour. Amer. Chem. Soci.* 58 (8 1936), p. 1486.

[184] J. G. Kirkwood. "The Dielectric Polarization of Polar Liquids". In: *Jour. Chem. Phys.* 7 (1939), p. 911.

[185] J. Jones. "On the Determination of Molecular Fields—I. From the Variation of the Viscosity of a Gas with Temperature". In: *Proc. Roya. Soci. A: Math. Phys. Engi. Scie.* 106 (738 1924).

[186] J. Jones. "On the Determination of Molecular Fields—I. From the Equation of States of a Gas". In: *Proc. Roya. Soci. A: Math. Phys. Engi. Scie.* 106 (738 1924).

[187] J. Lennard-Jones. "Cohesion". In: *Proc. Phys. Soci.* 43.5 (1931).

[188] J. Krupka et al. "Measurements of Permittivity and Dielectric Loss Tangent of High Resistivity Float Zone Silicon at Microwave Frequencies". In: *2006 Inte. Conf. Microw., Radar, Wirel. Comm.* (2006).

[189] R. Datta et al. "Large-aperture Wide-bandwidth Antireflection-coated Silicon Lenses for Millimeter Wavelengths". In: *Appl. Opti.* 52.36 (2013), p. 8747.

[190] C. J. Kopas et al. *Low Microwave Loss in Deposited Si and Ge Thin-film Dielectrics at Single-photon Power and Low Temperatures*. 2020. arXiv: 2011.10155.

[191] S. Agostinelli et al. "GEANT4—a simulation Toolkit". In: *Nucl. Inst. Meth. Phys. Rese. Sec. A: Acce., Spec., Dete. Asso. Equi.* 506 (3 2003).

[192] SOLIDWORKS. In: (). eprint: https://www.solidworks.com/.

[193] C. M. Poole et al. "A CAD Interface for GEANT4". In: *Aust. Phys. Engi. Scie. Medi* (2012).

[194] C. M. Poole et al. "Fast Tessellated Solid Navigation in GEANT4". In: *IEEE Trans. Nucl. Scie.* 99 (2012), p. 1.

[195] M. A. Ordal et al. "Optical Properties of Fourteen Metals in the Infrared and Far Infrared: Al, Co, Cu, Au, Fe, Pb, Mo, Ni, Pd, Pt, Ag, Ti, V, and W." In: *Appl. Opti.* 24 (24 1985), p. 4493.

[196] M. Kelsey et al. "SuperSim Reference Manual". In: *SuperCDMS Confluence Home Page (internal)* (2021). eprint: https://confluence.slac.stanford.edu/display/CDMS/SuperSim+Reference+Manual.

[197] R. Brun and F. Rademakers. "ROOT - An Object Oriented Data Analysis Framework". In: *Nucl. Inst. Meth. Phys. Rese. Sec. A: Acce., Spec., Dete. Asso. Equi.* 389 (1997).

[198] Simulia CST Studio Suite. eprint: https://www.3ds.com/products-services/simulia/products/cst-studio-suite/.

[199] COMSOL Multiphysics. eprint: https://www.comsol.com/.

[200] Terahertz. eprint: http://www.terahertz.co.uk/index.php?option=com_content\&view=article\&id=145\&Itemid=448.

[201] Stycast. eprint: https://www.lakeshore.com/products/categories/overview/temperature-products/cryogenic-accessories/epoxy.

[202] G. Hartwig et al. "Fibre-epoxy Composites at Low Temperatures". In: *Cryo.* 24 (11 1984).

[203] N. Kellaris et al. "Sub-Kelvin Thermal Conductivity and Radioactivity of Some Useful Materials in Low Background Cryogenic Experimentss". In: *Jour. Low Temp. Phys.* 176 (2014).

[204] M. C. Runyan et al. "Thermal Conductivity of Thermally-isolating Polymeric and Composite Structural Support Materials Between 0.3 and 4 K". In: *Cryo.* 48 (9 2008).

[205] N. J. Simon et al. "Properties of Copper and Copper Alloys at Cryogenic Temperature". In: *NIST Mono.* 117 (1992).

[206] A. Moroz et al. "A Drude Model Fit Parameters to the Dielectric Function of Free Electron Metals". In: *Wave Scattering* (2019). eprint: http://www.wave-scattering.com/.

[207] M. Born and E. Wolf. *Principle of Optics: Electromegnetic Theory of Propagation, Interference and Diffraction of Light*. 7th. Cambridge Unit. Pres., 1999.

[208] J. J. Bock et al. "A Novel Bolometer for Infrared and Millimeter-wave Astrophysics". In: *Spce. Sci. Rev.* 74 (1995), p. 229.

[209] W. Jones. "A Measurement of the Temperature and Polarization Anisotropies in the Comic Microwave Background Radiation". PhD thesis. California Institute of Technology, May 2005.

[210] E. Gatti et al. "Processing the Signals from Solid-state Detectors in Elementary-particle Physics". In: *Rivi. Nuov. Cime.* 9 (1986).

[211] M. V. P. Altoé et al. *Localization and Reduction of Superconducting Quantum Coherent Circuit Losses*. 2020. arXiv: 2012.07604.

Appendix A

TEDIOUS MATHEMATICAL DERIVATIONS

A.1 KID-based detector RF characterization

- **For Eq. (4.11):**

We may rearrange Eq. (4.10) to obtain

$$2iQ_r\frac{\Delta f}{f_r} = \frac{1-e^{\xi}}{1+e^{\xi}} \tag{A.1}$$

and substitute the result into Eq. (4.9), which then yield

$$S_{21} = 1 - \frac{Q_r}{|\hat{Q}_c|}\left(\frac{1}{1+\dfrac{1-e^{\xi}}{1+e^{\xi}}}\right)e^{i\phi_c} = 1 - \frac{Q_r}{|\hat{Q}_c|}\left(\frac{1+e^{\xi}}{2}\right)e^{i\phi_c} \tag{A.2}$$

$$= 1 + \frac{Q_r}{2|\hat{Q}_c|}\left(1+e^{\xi}\right)e^{i(\phi_c+\pi)}. \tag{A.3}$$

A.2 KID-based detector responsivity calibration

- **For Eq. (4.36):**

We use $\hat{f} = \Delta f/f_r$ to simplify the writing. Substituting Eq. (4.35) into Eq. (4.33),

$$dS_{21} = \left(\partial_{Q_i^{-1}} + \partial_{\hat{f}}\right)\left(1 - \frac{Q_r/Q_c}{1+2iQ_r\hat{f}}\right) \tag{A.4}$$

$$= \left[\frac{Q_c^{-1}}{1+2iQ_r\hat{f}} + \frac{2i\hat{f}(Q_r/Q_c)}{(1+2iQ_r\hat{f})^2}\right]\cdot\partial_{Q_i^{-1}}Q_r\cdot dQ_i^{-1} + \frac{Q_r/Q_c}{(1+2iQ_r\hat{f})^2}\cdot 2iQ_r\cdot d\hat{f}. \tag{A.5}$$

Taking

$$\lim_{f\to f_r}\hat{f} = 0 \tag{A.6}$$

for the near-resonance readout condition and

$$Q_r = (Q_c^{-1} + Q_i^{-1})^{-1} \;\Rightarrow\; \partial_{Q_i^{-1}} Q_r = -Q_r^2 \tag{A.7}$$

according to Eq. (4.2), we may further rewrite Eq. (A.5) by

$$dS_{21} = \left[\frac{Q_c^{-1}}{1 + \cancel{2iQ_r \hat{f}}} + \cancel{\frac{2i\hat{f}(Q_r/Q_c)}{(1+2iQ_r\hat{f})^2}} \right] \cdot -Q_r^2 \cdot dQ_i^{-1} + \frac{Q_r/Q_c}{(1+\cancel{2iQ_r\hat{f}})^2} \cdot 2iQ_r \cdot d\hat{f} \tag{A.8}$$

$$= \frac{-Q_r^2}{Q_c} d\left(\frac{1}{Q_i}\right) + 2i\frac{Q_r^2}{Q_c} d\left(\frac{\Delta f}{f_r}\right). \tag{A.9}$$

- **For Eq. (4.45):**

We start by rearranging Eq. (4.37) to

$$\left.\frac{\Delta f}{f_r}\right|_0^T = \left(\frac{f - f_r(T)}{f_r(T)}\right) - \left(\frac{f - f_{r,0}}{f_{r,0}}\right) = \frac{\alpha\kappa_2 n_{\mathrm{qp,th}}}{2}, \tag{A.10}$$

and by choosing $f = f_{r,0}$, we have

$$\left(\frac{f_{r,0} - f_r(T)}{f_r(T)}\right) - \cancel{\left(\frac{f_{r,0} - f_{r,0}}{f_{r,0}}\right)} = \frac{\alpha\kappa_2 n_{\mathrm{qp,th}}}{2}, \tag{A.11}$$

which may be rearranged to

$$\frac{f_r(T)}{f_{r,0}} = \left(\frac{\alpha\kappa_2 n_{\mathrm{qp,th}}}{2} + 1\right)^{-1}. \tag{A.12}$$

Knowing for typical α, κ_2, and $n_{\mathrm{qp,th}}$ within the controlled temperature ranges,

$$\frac{\alpha\kappa_2 n_{\mathrm{qp,th}}}{2} \sim O(10^{-5} - 10^{-4}) \ll 1, \tag{A.13}$$

Eq. (A.12) may be approximated by

$$\frac{f_r(T)}{f_{r,0}} \approx \left(\frac{-\,\alpha\kappa_2 n_{\mathrm{qp,th}}}{2} + 1\right). \tag{A.14}$$

A.3 KID-based detector pulse response

- **For Eq. (4.66):**

Starting from Eq. (4.65),

$$\frac{dn_{\text{qp}}}{dt} = -R[n_{\text{qp}}^2 - n_{\text{qp,tot}}^2] \tag{A.15}$$

$$= -R(n_{\text{qp}} + n_{\text{qp,tot}})(n_{\text{qp}} - n_{\text{qp,tot}}), \tag{A.16}$$

$$-Rdt = \frac{1}{(n_{\text{qp}} + n_{\text{qp,tot}})(n_{\text{qp}} - n_{\text{qp,tot}})} dn_{\text{qp}} \tag{A.17}$$

$$= \frac{1}{2n_{\text{qp,tot}}} \left[\frac{1}{n_{\text{qp}} - n_{\text{qp,tot}}} - \frac{1}{n_{\text{qp}} + n_{\text{qp,tot}}} \right] dn_{\text{qp}}. \tag{A.18}$$

Letting $n_{\text{qp}} = n_{\text{qp,max}}$ at $t = 0$ and integrating both sides,

$$-2n_{\text{qp,tot}} Rt = \ln\left(\frac{n_{\text{qp}} - n_{\text{qp,tot}}}{n_{\text{qp,max}} - n_{\text{qp,tot}}}\right) - \ln\left(\frac{n_{\text{qp}} + n_{\text{qp,tot}}}{n_{\text{qp,max}} + n_{\text{qp,tot}}}\right) \tag{A.19}$$

$$= \ln\left(\frac{n_{\text{qp}} - n_{\text{qp,tot}}}{n_{\text{qp}} + n_{\text{qp,tot}}}\right) - \ln\left(\frac{n_{\text{qp,max}} - n_{\text{qp,tot}}}{n_{\text{qp,max}} + n_{\text{qp,tot}}}\right). \tag{A.20}$$

Eliminating the logarithms with some rearrangement and defining χ to simplify the writing,

$$\left(\frac{n_{\text{qp,max}} - n_{\text{qp,tot}}}{n_{\text{qp,max}} + n_{\text{qp,tot}}}\right) e^{-2n_{\text{qp,tot}}Rt} = \left(\frac{n_{\text{qp}} - n_{\text{qp,tot}}}{n_{\text{qp}} + n_{\text{qp,tot}}}\right) = \chi, \tag{A.21}$$

we have

$$n_{\text{qp}}(t) = \left(\frac{1+\chi}{1-\chi}\right) n_{\text{qp,tot}}, \tag{A.22}$$

where

$$\chi(t) = \left(\frac{n_{\text{qp,max}} - n_{\text{qp,tot}}}{n_{\text{qp,max}} + n_{\text{qp,tot}}}\right) e^{-2Rn_{\text{qp,tot}}t}. \tag{A.23}$$

- **For Eq. (4.69):**

Considering

$$\chi(t) = e^{-2Rn_{\text{qp,tot}}t} \quad , \; n_{\text{qp,max}} \gg n_{\text{qp,tot}} \tag{A.24}$$

for large pulses, at $t \to 0$, i.e. near the $\chi = 1$ singularity for n_{qp} in Eq. (4.66),

$$\begin{aligned}
n_{qp}(t) &= \left(\frac{1-\chi}{1+\chi}\right)^{-1} n_{qp,tot} && \text{(A.25)}\\
&\approx \left(\left.\frac{dn_{qp}^{-1}}{d\chi}\right|_{\chi=1} \left.\frac{d\chi}{dt}\right|_{t=0} t + O(t^2)\right)^{-1} n_{qp,tot} && \text{(A.26)}\\
&= \left(\left.\frac{-2}{(1+\chi)^2}\right|_{\chi=1} \left.(-2Rn_{qp,tot})\right|_{t=0} t + O(t^2)\right)^{-1} n_{qp,tot} && \text{(A.27)}\\
&= \frac{1}{Rt} + \text{H.C.} && \text{(A.28)}
\end{aligned}$$

A.4 KID-based detector noise characterization

- **For Eq. (4.119):**

From Eq. (4.5), assuming the transmission line attenuation α and the coupling phase ϕ_c are both small, when reading on-resonance, the transmission is

$$\begin{aligned}
S_{21}(f_r) &= 1 - \frac{Q_r}{Q_c} && \text{(A.29)}\\
&= 1 - \frac{\left(Q_i^{-1} + Q_c^{-1}\right)^{-1}}{Q_c} && \text{(A.30)}\\
&= 1 - \frac{Q_i}{Q_i + Q_c} && \text{(A.31)}\\
&= \frac{Q_c}{Q_i + Q_c} && \text{(A.32)}\\
&= \frac{Q_r}{Q_i}. && \text{(A.33)}
\end{aligned}$$

Perturbing the above and substituting Eq. (4.22) for n_{qp},

$$\begin{aligned}
\delta S_{21} &= \delta\left(\frac{Q_c}{Q_i + Q_c}\right) && \text{(A.34)}\\
&= \frac{-Q_c}{(Q_i + Q_c)^2}\delta Q_i \cdot \frac{Q_i}{Q_i}, && \text{(A.35)}
\end{aligned}$$

notice here we adopt the smart $\delta Q_i \to \delta n_{\text{qp}}$ conversion, shown below, by Zmuidzinas,

$$= \frac{-Q_r^2}{Q_i Q_c}\left(\frac{-\chi_{\text{qp}}\delta n_{\text{qp}}}{n_{\text{qp}}}\right) \tag{A.36}$$

$$= \frac{Q_r^2}{Q_i Q_c}\chi_{\text{qp}} \cdot \frac{\alpha\kappa|\gamma|}{S_{21}}\frac{Q_r^2}{Q_c} \cdot \delta n_{\text{qp}} \tag{A.37}$$

$$= \frac{Q_r^3}{Q_c^2}\chi_{\text{qp}} \cdot \alpha\kappa|\gamma| \cdot \delta n_{\text{qp}}. \tag{A.38}$$

Zmuidzinas' (2009) $\delta Q_i \to \delta n_{\text{qp}}$ conversion:
Using

$$\frac{\delta Q_{i,\text{qp}}}{Q_{i,\text{qp}}} = \left(\frac{\delta n_{\text{qp}}}{n_{\text{qp}}}\right)^{-1} \Leftrightarrow \delta \ln Q_{i,\text{qp}} = -\delta \ln n_{\text{qp}}, \tag{A.39}$$

$$\frac{\delta Q_i}{Q_i} = \delta \ln Q_i = \frac{Q_i}{Q_{i,\text{qp}}}\delta \ln Q_{i,\text{qp}} = -\chi_{\text{qp}}\delta \ln n_{\text{qp}} = \frac{-\chi_{\text{qp}}\delta n_{\text{qp}}}{n_{\text{qp}}}. \tag{A.40}$$

Note that since our design is purposely over-coupled it is making the

$$\chi_c = \frac{4Q_r^2}{Q_i Q_c} = \frac{4}{(1 + Q_i/Q_c)(1 + Q_c/Q_i)} \tag{A.41}$$

coupling factor defined in (Zmuidzinas, 2009) far from unity.

A.5 KID-based detector energy resolution

- **For Eq. (B.20):**

Substituting Eq. (B.15) into Eq. (B.8),

$$\sigma_{A'} = \sqrt{\overline{\left(\frac{\displaystyle\sum\!\!\!\!\!\!\int e^{2\pi i f\tau}\tilde{\phi}^*\tilde{S}\,\delta f}{\displaystyle\sum\!\!\!\!\!\!\int \tilde{\phi}^*\tilde{\bar{S}}\,\delta f}\right)^2} - \overline{\left(\frac{\displaystyle\sum\!\!\!\!\!\!\int e^{2\pi i f\tau}\tilde{\phi}^*\tilde{S}\,\delta f}{\displaystyle\sum\!\!\!\!\!\!\int \tilde{\phi}^*\tilde{\bar{S}}\,\delta f}\right)}^2}, \tag{A.42}$$

and using selected pure noise samples for the calibration, i.e. $\tilde{S} = \tilde{N}$, while recognising the noise fluctuates around zero threfore imposing a trivial expectation value, i.e. $\overline{\tilde{N}} = 0$,

$$= \sqrt{\overline{\left(\frac{\displaystyle\sum\!\!\!\!\!\!\int e^{2\pi i f\tau}\tilde{\phi}^*\tilde{N}\,\delta f}{\displaystyle\sum\!\!\!\!\!\!\int \tilde{\phi}^*\tilde{\bar{S}}\,\delta f}\right)^2} - \overline{\left(\frac{\cancel{\displaystyle\sum\!\!\!\!\!\!\int e^{2\pi i f\tau}\tilde{\phi}^*\tilde{N}\,\delta f}}{\displaystyle\sum\!\!\!\!\!\!\int \tilde{\phi}^*\tilde{\bar{S}}\,\delta f}\right)}^2}, \tag{A.43}$$

and then using the definition of OF from Eq. (B.16) and eliminating common powers,

$$= \sqrt{\overline{\left(\frac{\displaystyle\sum\!\!\!\!\!\!\int \left|\tilde{\tilde{S}}\right|^2 \left|\tilde{N}\right|^{-2} \delta f}{\displaystyle\sum\!\!\!\!\!\!\int \left|\tilde{\tilde{S}}\right|^4 \left|\tilde{N}\right|^{-4} \delta f} \right)}}, \tag{A.44}$$

and finally using the fact that $\overline{\left|\tilde{N}\right|^{2\mathbb{Z}}} = \overline{\left|\tilde{N}\right|^{2}}^{\mathbb{Z}}$ and removing common divisors for the numerator and the denominator,

$$= \sqrt{\left(\frac{\displaystyle\sum\!\!\!\!\!\!\int \left|\tilde{\tilde{S}}\right|^2 \overline{\left|\tilde{N}\right|^{-2}} \delta f}{\displaystyle\sum\!\!\!\!\!\!\int \left|\tilde{\tilde{S}}\right|^4 \overline{\left|\tilde{N}\right|^{-4}} \delta f} \right)} \tag{A.45}$$

$$= \sqrt{\left(\sum\!\!\!\!\!\!\int \left|\tilde{\tilde{S}}\right|^2 \overline{\left|\tilde{N}\right|^{-2}} \delta f \right)^{-1}}. \tag{A.46}$$

- **For Eq. (4.196) and Eq. (4.202):**

Fourier-transforming Eq. (4.196) with Eq. (4.191),

$$i\omega\delta\tilde{n}_{\mathrm{qp}}(f) = -\frac{\delta\tilde{n}_{\mathrm{qp}}(f)}{\tau_{\mathrm{qp}}} + \frac{\eta_{\mathrm{ph}} E_{\mathrm{ph}}}{\Delta V \tau_{\mathrm{abs}}} \cdot \frac{\tau_{\mathrm{abs}}}{i\omega\tau_{\mathrm{abs}} + 1} \tag{A.47}$$

$$\delta\tilde{n}_{\mathrm{qp}}(f) = \left(i\omega + \frac{1}{\tau_{\mathrm{qp}}} \right)^{-1} \cdot \frac{\eta_{\mathrm{ph}} E_{\mathrm{ph}}}{\Delta V \tau_{\mathrm{abs}}} \cdot \frac{\tau_{\mathrm{abs}}}{i\omega\tau_{\mathrm{abs}} + 1} \tag{A.48}$$

$$= \frac{\eta_{\mathrm{ph}} E_{\mathrm{ph}}}{\Delta V} \underbrace{\frac{1}{\tau_{\mathrm{abs}}} \left(\frac{\tau_{\mathrm{qp}}}{i\omega\tau_{\mathrm{qp}} + 1} \cdot \frac{\tau_{\mathrm{abs}}}{i\omega\tau_{\mathrm{abs}} + 1} \right)}_{\mathbf{Y}} \tag{A.49}$$

$$= \frac{\eta_{\mathrm{ph}} E_{\mathrm{ph}}}{\Delta V} \tau_{\mathrm{qp}} \left(\frac{1/\tau_{\mathrm{abs}}}{i\omega\tau_{\mathrm{qp}} + 1} - \frac{1/\tau_{\mathrm{qp}}}{i\omega\tau_{\mathrm{abs}} + 1} \right) \cdot \frac{1}{1/\tau_{\mathrm{abs}} - 1/\tau_{\mathrm{qp}}} \tag{A.50}$$

$$= \frac{\eta_{\mathrm{ph}} E_{\mathrm{ph}}}{\Delta V} \frac{1/\tau_{\mathrm{abs}}}{1/\tau_{\mathrm{abs}} - 1/\tau_{\mathrm{qp}}} \left(\frac{\tau_{\mathrm{qp}}}{i\omega\tau_{\mathrm{qp}} + 1} - \frac{\tau_{\mathrm{abs}}}{i\omega\tau_{\mathrm{abs}} + 1} \right) \tag{A.51}$$

$$= \frac{\eta_{\mathrm{ph}} E_{\mathrm{ph}}}{\Delta V} \frac{\tau_{\mathrm{qp}}}{\tau_{\mathrm{qp}} - \tau_{\mathrm{abs}}} \left(\frac{\tau_{\mathrm{qp}}}{i\omega\tau_{\mathrm{qp}} + 1} - \frac{\tau_{\mathrm{abs}}}{i\omega\tau_{\mathrm{abs}} + 1} \right) \tag{A.52}$$

$$= \frac{\eta_{\mathrm{ph}} E_{\mathrm{ph}}}{\Delta V} \frac{\tau_{\mathrm{qp}}}{\tau_{\mathrm{qp}} - \tau_{\mathrm{abs}}} \left(\mathcal{F}\left[e^{-t/\tau_{\mathrm{qp}}} \right] - \mathcal{F}\left[e^{-t/\tau_{\mathrm{abs}}} \right] \right) \tag{A.53}$$

$$\delta n_{\mathrm{qp}}(t) = \frac{\eta_{\mathrm{ph}} E_{\mathrm{ph}}}{\Delta V} \frac{\tau_{\mathrm{qp}}}{\tau_{\mathrm{qp}} - \tau_{\mathrm{abs}}} \left(e^{-t/\tau_{\mathrm{qp}}} - e^{-t/\tau_{\mathrm{abs}}} \right). \tag{A.54}$$

For solving the $|\mathrm{X}|^2$ and $|\mathrm{Y}|^2$ integrals, the $1/|i\omega\tau + 1|^2$ type of integrants, see below, all contribute one positive and one negative imaginary poles symmetrically to the real ω axis, so it does not matter if one chooses the $\mathbb{Z}^+$ or $\mathbb{Z}^-$ contour for the integration, as long as the residue for one of the poles in each pair is included and summed, giving

$$\begin{aligned}
&\int_{-\infty}^{\infty} |\mathbf{Y}|^2 \, df \\
&= \oint_{\mathbb{Z}^+} \left[\frac{1}{\tau_{\text{abs}}^2} \left(\frac{\tau_{\text{qp}}}{-i\omega\tau_{\text{qp}} + 1} \cdot \frac{\tau_{\text{abs}}}{-i\omega\tau_{\text{abs}} + 1} \cdot \frac{\tau_{\text{qp}}}{i\omega\tau_{\text{qp}} + 1} \cdot \frac{\tau_{\text{abs}}}{i\omega\tau_{\text{abs}} + 1} \right) \right] \frac{d\omega}{2\pi} && \text{(A.55)} \\
&= \tau_{\text{qp}}^2 2\pi i \left(\frac{1}{2i\tau_{\text{qp}}} \cdot \frac{1}{(\tau_{\text{abs}}/\tau_{\text{qp}})^2 \cancel{+1}} + \frac{1}{2i\tau_{\text{abs}}} \cdot \frac{1}{\cancel{(\tau_{\text{qp}}/\tau_{\text{abs}})^2} + 1} \right) \frac{1}{2\pi} && \text{(A.56)} \\
&\approx \frac{\tau_{\text{qp}}^2}{2\tau_{\text{abs}}} \quad \text{for } \tau_{\text{qp}} \ll \tau_{\text{abs}} && \text{(A.57)}
\end{aligned}$$

and

$$\begin{aligned}
&\int_{-\infty}^{\infty} |\mathbf{X}|^2 \, df \\
&= \oint_{\mathbb{Z}^+} \left[\left(\frac{\tau_{\text{qp}}}{-i\omega\tau_{\text{qp}} + 1} \cdot \frac{\tau_{\text{qp}}}{i\omega\tau_{\text{qp}} + 1} \right) \right] \frac{d\omega}{2\pi} && \text{(A.58)} \\
&= \tau_{\text{qp}}^2 2\pi i \left(\frac{1}{2i\tau_{\text{qp}}} \right) \frac{1}{2\pi} && \text{(A.59)} \\
&= \frac{\tau_{\text{qp}}}{2}. && \text{(A.60)}
\end{aligned}$$

- **For Eq. (4.196):**

Starting from Eq. (4.224), we rewrite

$$\begin{aligned}
\frac{\eta_{\text{read}}}{V\Delta R} \frac{\chi_{\text{c}}\chi_{\text{qp}}}{2} dP_g &= 2n_{\text{qp}} dn_{\text{qp}} && \text{(A.61)} \\
\frac{dn_{\text{qp}}}{dP_g} &= \frac{\eta_{\text{read}}}{V\Delta R} \frac{\chi_{\text{c}}\chi_{\text{qp}}}{4n_{\text{qp}}} \approx \frac{n_{\text{qp}}}{2p_g}, && \text{(A.62)}
\end{aligned}$$

which may be used in

$$\begin{aligned}
\frac{d\sigma_E}{dP_g} = 0 &= \frac{d}{dP_g} \left[\frac{n_{\text{qp}}}{\sqrt{P_g}} \cdot 2R \left(n_{\text{qp}} + n_{\text{qp,min}} \right) \right] && \text{(A.63)} \\
&= \frac{-1}{2\sqrt{P_g}^3} \cdot 2R \left(n_{\text{qp}}^2 + n_{\text{qp}} n_{\text{qp,min}} \right) \\
&\qquad + \frac{1}{\sqrt{P_g}} \cdot 2R \left(2n_{\text{qp}} + n_{\text{qp,min}} \right) \cdot \frac{dn_{\text{qp}}}{dP_g}, && \text{(A.64)}
\end{aligned}$$

using the assumption that $n_{\rm qp}$ is dominated by P_g,

$$\Rightarrow \frac{1}{2} \cdot 2R \left(n_{\rm qp}^2 + n_{\rm qp} n_{\rm qp,min}\right) \approx P_g \cdot 2R \left(2n_{\rm qp} + n_{\rm qp,min}\right) \cdot \frac{n_{\rm qp}}{2p_g} \tag{A.65}$$

$$\left(n_{\rm qp}^2 + n_{\rm qp} n_{\rm qp,min}\right) \approx \left(2n_{\rm qp}^2 + n_{\rm qp} n_{\rm qp,min}\right), \tag{A.66}$$

so we find

$$\Rightarrow \frac{n_{\rm qp}}{n_{\rm qp,min}} \ll 1. \tag{A.67}$$

A.6 Cryogenic blackbody radiation physics

- **For Eq. (5.28):**

$$\left|\left\langle f \left|\hat{\mathrm{H}}_{\mathrm{i}}\right| i\right\rangle\right| = \frac{eA_0}{2m_e c} \cdot \frac{1}{(\pi r_0^3)^{1/2}} \cdot \frac{1}{(2\pi\hbar)^{3/2}} \int e^{-i\vec{k}_f \cdot \vec{r}} \hat{p} e^{-r/r_0} d^3 r \tag{A.68}$$

$$= \frac{eA_0}{2m_e c} \cdot \frac{1}{(\pi r_0^3)^{1/2}} \cdot \frac{1}{(2\pi\hbar)^{3/2}} \int e^{-r/r_0} \hat{p} e^{-i\vec{k}_f \cdot \vec{r}} d^3 r \tag{A.69}$$

$$= \frac{eA_0}{2m_e c} \cdot \frac{1}{(\pi r_0^3)^{1/2}} \cdot \frac{1}{(2\pi\hbar)^{3/2}} \int e^{-r/r_0} (\hbar\vec{k}_f) e^{-i\vec{k}_f \cdot \vec{r}} d^3 r \tag{A.70}$$

$$= \frac{eA_0 \cdot \hbar\vec{k}_f}{2m_e c} \cdot \frac{1}{(\pi r_0^3)^{1/2}} \cdot \frac{2\pi}{(2\pi\hbar)^{3/2}} \iint e^{-r/r_0} e^{-ik_f r \cos\theta} r^2 dr\, d\cos\theta \tag{A.71}$$

$$= \frac{eA_0 \cdot \hbar\vec{k}_f}{2m_e c} \cdot \frac{1}{(\pi r_0^3)^{1/2}} \cdot \frac{2\pi}{(2\pi\hbar)^{3/2}} \int \frac{e^{-r/r_0}(e^{ik_f r} - e^{-ik_f r})}{-ik_f r} r^2 dr \tag{A.72}$$

$$= \frac{eA_0 \cdot \hbar\vec{k}_f}{2m_e c} \cdot \frac{1}{(\pi r_0^3)^{1/2}} \cdot \frac{2\pi}{(2\pi\hbar)^{3/2}} \left[\frac{1}{ik_f} \partial_{r_0^{-1}} \int e^{-r/r_0}(e^{ik_f r} - e^{-ik_f r}) dr\right] \tag{A.73}$$

$$= \frac{eA_0 \cdot \hbar\vec{k}_f}{2m_e c} \cdot \frac{1}{(\pi r_0^3)^{1/2}} \cdot \frac{2\pi}{(2\pi\hbar)^{3/2}} \cdot \frac{4r_0^3}{(k_f^2 r_0^2 + 1)^2}. \tag{A.74}$$

- **For Eq. (5.39):**

$$\sigma(\omega) = \frac{2\pi}{\hbar}\left(\frac{e\hbar}{2m_e c}\right)^2 \frac{1}{\pi r_0^3(2\pi\hbar)^3} \cdot \frac{64\pi^2 r_0^6 \sqrt{2m_e(\hbar\omega - U)/\hbar^2}^3}{(\sqrt{2m_e(\hbar\omega - U)/\hbar^2}^2 r_0^2 + 1)^4} \cdot m_e\hbar \cdot \frac{4\pi}{3} \cdot \frac{2Z_0c^2\hbar}{\omega} \tag{A.75}$$

$$= \frac{32Z_0e^2r_0^3\hbar^5}{3m_e\omega} \frac{[2m_e(\hbar\omega - U)]^{3/2}}{(2m_e(\hbar\omega - U)r_0^2 + \hbar^2)^4} \tag{A.76}$$

$$= \frac{32Z_0e^2r_0^3\hbar^6}{3m_eE_\gamma} \frac{[2m_e(\hbar\omega - U)]^{3/2}}{(2m_e(\hbar\omega - U)r_0^2 + \hbar^2)^4}. \tag{A.77}$$

A.7 Cryogenic blackbody radiation calibration plan

- **For Eq. (5.110):**

Using

$$n = r_n e^{i\theta}, \tag{A.78}$$

$$\epsilon = \epsilon_1 + i\epsilon_2 = n^2 = {r_n}^2 e^{2i\theta}, \tag{A.79}$$

$$|\epsilon| = \sqrt{{\epsilon_1}^2 + {\epsilon_2}^2} = {r_n}^2 = \xi, \tag{A.80}$$

$$\cos\theta = \sqrt{\frac{\cos 2\theta + 1}{2}} = \sqrt{\frac{\epsilon_1/\xi + 1}{2}}, \tag{A.81}$$

we may rewrite

$$\left|\frac{n-1}{n+1}\right|^2 = \frac{(r_n\cos\theta - 1)^2 + {r_n}^2\sin^2\theta}{(r_n\cos\theta + 1)^2 + {r_n}^2\sin^2\theta} \tag{A.82}$$

$$= \frac{{r_n}^2 + 1 - 2r_n\cos\theta}{{r_n}^2 + 1 + 2r_n\cos\theta} \tag{A.83}$$

$$= \frac{\xi + 1 - 2\sqrt{\xi}\sqrt{(\epsilon_1/\xi - 1)/2}}{\xi + 1 + 2\sqrt{\xi}\sqrt{(\epsilon_1/\xi - 1)/2}} \tag{A.84}$$

$$= \frac{\xi + 1 - \sqrt{2(\epsilon_1 + \xi)}}{\xi + 1 + \sqrt{2(\epsilon_1 + \xi)}}. \tag{A.85}$$

- **For Eq. (5.124) and Eq. (5.125):**

To solve Eq. (5.123) for the analytical forms, we first combine the two $\alpha = 0$ equations by eliminating E_{0t} and get

$$(1 - \frac{Y_0}{Y_1})E_{0i} + (1 + \frac{Y_0}{Y_1})E_{0r} = 2E_{1r}\delta, \tag{A.86}$$

and then, by combining the two $\alpha = 1$ equations, we obtain

$$E_{1r} = \frac{1}{2}(1 - \frac{Y_2}{Y_1})E_{1t}, \tag{A.87}$$

which we supply to the previous equation to have

$$(1 - \frac{Y_0}{Y_1})E_{0i} + (1 + \frac{Y_0}{Y_1})E_{0r} = (1 - \frac{Y_2}{Y_1})E_{1t}\delta. \tag{A.88}$$

Repeating the same prescription but this time eliminating E_{1r} in the first step, we have

$$(1 + \frac{Y_0}{Y_1})E_{0i} + (1 - \frac{Y_0}{Y_1})E_{0r} = 2E_{0t}, \tag{A.89}$$

so we also want to compute

$$E_{0t} = \frac{1}{2}(1 + \frac{Y_2}{Y_1})E_{1t}\delta^{-1} \tag{A.90}$$

from the second step (notice the additional δ^{-1} compared to Eq. (A.87)) and arrive at

$$(1 + \frac{Y_0}{Y_1})E_{0i} + (1 - \frac{Y_0}{Y_1})E_{0r} = (1 + \frac{Y_2}{Y_1})E_{1t}\delta^{-1}. \tag{A.91}$$

One way to solve the algebraically symmetric Eq. (A.88) and Eq. (A.91) is to calculate their common and differential parts, which read

$$\begin{cases} E_{0i} + E_{0r} = \dfrac{1}{2}\left[(1 + \dfrac{Y_2}{Y_1})\delta^{-1} + (1 - \dfrac{Y_2}{Y_1})\delta\right] E_{1t} \\ E_{0i} - E_{0r} = \dfrac{Y_1}{2Y_0}\left[(1 + \dfrac{Y_2}{Y_1})\delta^{-1} - (1 - \dfrac{Y_2}{Y_1})\delta\right] E_{1t} \end{cases}, \tag{A.92}$$

and from these equations we finally obtain

$$\frac{E_{1t}}{E_{0i}} = 4\left\{\left[(1 + \frac{Y_2}{Y_1})\delta^{-1} + (1 - \frac{Y_2}{Y_1})\delta\right] + \frac{Y_1}{Y_0}\left[(1 + \frac{Y_2}{Y_1})\delta^{-1} - (1 - \frac{Y_2}{Y_1})\delta\right]\right\}^{-1} \tag{A.93}$$

and

$$\frac{E_{0r}}{E_{0i}} = \frac{\left[(1 + \frac{Y_2}{Y_1})\delta^{-1} + (1 - \frac{Y_2}{Y_1})\delta\right] - \frac{Y_1}{Y_0}\left[(1 + \frac{Y_2}{Y_1})\delta^{-1} - (1 - \frac{Y_2}{Y_1})\delta\right]}{\left[(1 + \frac{Y_2}{Y_1})\delta^{-1} + (1 - \frac{Y_2}{Y_1})\delta\right] + \frac{Y_1}{Y_0}\left[(1 + \frac{Y_2}{Y_1})\delta^{-1} - (1 - \frac{Y_2}{Y_1})\delta\right]}. \tag{A.94}$$

- **For Eq. (5.141):**

Empirically,

$$\frac{\sigma d}{\epsilon_0 c} \approx 151, \tag{A.95}$$

so

$$t = \frac{2}{1 - (\sigma d/\epsilon_0 c)i} \tag{A.96}$$

$$\approx \frac{2\epsilon_0 c}{\sigma d} i \approx 0.013i \tag{A.97}$$

$$T = |t|^2 \approx 1.8 \times 10^{-4}, \tag{A.98}$$

and

$$A \approx 2\mathbf{Re}(t) \approx 2 \times \frac{2\epsilon_0 c}{\sigma d} \tag{A.99}$$

$$\approx 0.026. \tag{A.100}$$

Using the low-frequency skin depth relation

$$\delta \approx \sqrt{\frac{2}{\omega\sigma}}, \tag{A.101}$$

Eq. (A.99) becomes

$$A \approx \frac{\delta}{d} \times \left(2\sqrt{\frac{2\omega\epsilon_0}{\sigma}}\right), \tag{A.102}$$

where we also use $\epsilon_0 c \approx \sqrt{\epsilon_0}$.

Appendix B

THE OPTIMAL FILTER FORMALISM

B.1 Optimal filter: A time-domain primer

A finite detection resolution is the consequence of detecting the signal in the presence of an unpredictable fluctuation that introduces uncertainty to the obtained signal size. Whether the fluctuation is due to the intrinsic instability of the detector, the added noise by the readout equipment, the external/environmental excitations, or a combination of all the above, in practice the resolution always boils down to comparing the size of the anticipated signal to the undesirable noise, which subsequently determines a relative significance for such a detection not being a product of pure noise. The concept comprises two key ingredients: First and foremost, a *subjective* definition for the anticipated signal that distinguishes the noise. And second, a well-defined methodology for comparing the observation to the signal definition that provides a signal size estimate with a quantitative significance. This significance is then generally taken as the "resolution" for the detection. Historically, the SuperCDMS collaboration adapted and has been actively improving the optimal filer (OF) formalism (Gatti et al., 1986; S. Golwala, 2000), which is more widely known as the matched filter technique in the field of signal processing science. As the name suggests, given the assumptions detailed below, one can mathematically prove that the OF formalism provides the most certain signal size estimate with the lowest "optimal" corresponding resolution. We therefore follow SuperCDMS' conversion and adapt the OF formalism to model our DMKID detector resolution. In the following discussion, we will first assume that the subjective signal "template" can be unambiguously defined, so we can proceed to introduce the general OF formalism. After the general introduction, we will then elaborate our choice for the signal template and the experimental implementation for the OF resolution modeling and calibration.

We start by assuming a general signal timestream $S(t)$ that consists of an idealized, noise-less, signal template $\hat{S}(t)$ and the rest that is not described by $\hat{S}(t)$, which we regard as the noise $N(t)$, i.e.,

$$S(t) = A\hat{S}(t) + N(t). \tag{B.1}$$

A in Eq. (B.1) is the physical amplitude for the fully constrained signal template $\hat{S}(t)$

and is the only variable in the equation. It indicates the energy deposition of the event, therefore we aim to estimate A from the noisy timestream $S(t)$ as the main goal of this section. While the signal size A is to be estimated hence also leaves the noise component $N(t)$ undetermined, we have an unambiguous measurement for the total signal $S(t)$, and at the same time, a fully constrained $\hat{S}(t)$ by definition. Therefore, we search for the most possible signal in the noisy data by rescaling the signal template with A and shifting it in time to achieve the best match. We will first illustrate the procedure in the seemingly more straightforward time domain, which will lead to the conclusion that, despite the mathematical equivalence, it is a worse choice to perform the OF calculation in the time-domain than in the frequency-domain. We will then introduce in the next section our standard frequency-domain OF. Letting A' and τ be the amplitude and temporal offset for the template, respectively, we compare the noiseless condition to the noisy signal and define the squared deviation between the data and the estimate by

$$\chi^2 = \sum\!\!\!\!\!\!\int \left(S(t) - A'\hat{S}(t-\tau)\right)^2 \delta t; \tag{B.2}$$

note that this χ^2 is only used to denote the squared deviation sum but not a formally normalized Chi-squared statistic. We continue to find its minimum by

$$d\chi^2 = \partial_{A'}\chi^2 dA' + \partial_\tau \chi^2 d\tau = 0, \tag{B.3}$$

which, substituting Eq. (B.2), yields

$$\begin{aligned} 0 = \sum\!\!\!\!\!\!\int \Big[&\left(-2S(t)\hat{S}(t-\tau) + 2A'\hat{S}^2(t-\tau)\right) dA' \\ &+ \left(2S(t)A'\dot{\hat{S}}(t-\tau) - 2A'^2\hat{S}(t-\tau)\dot{\hat{S}}(t-\tau)\right) d\tau \Big] \delta t, \end{aligned} \tag{B.4}$$

where we shorthand

$$\dot{\hat{S}}(t-\tau) = \frac{d\hat{S}(t-\tau)}{d\tau} \tag{B.5}$$

for readability.

The principle of separation of variables allows us to equate the $\partial_{A'}$ and ∂_τ parentheses in Eq. (B.4) to zero and find

$$A' = \frac{\left\langle S(t)\middle|\hat{S}(t-\tau)\right\rangle}{\left\langle \hat{S}\middle|\hat{S}\right\rangle} = \frac{\left\langle S(t)\middle|\dot{\hat{S}}(t-\tau)\right\rangle}{\left\langle \hat{S}\middle|\dot{\hat{S}}\right\rangle}. \tag{B.6}$$

For readability, we adopt the notation for the temporal correlation defined in Eq. (4.176) and suppress the time dependence for terms with equal temporal offsets;

these terms are offset-invariant under unbounded integrals. The first equation in Eq. (B.6) shows that the extraction for the optimal amplitude estimate corresponding to the least χ^2 is essentially cross-correlating the data with the assumed signal template and then normalizing the cross-correlation with the template's auto-correlation. The result is a relative amplitude with respect to the template amplitude. One may understand the treatment as comparing the noisy data against the pre-defined signal template at every time sample and weighting the resemblance between the observation and the template, which effectively rejects every less resemble data point by its weight as a time-dependent filter. To normalize this weighted sum to be the amplitude, the referenced signal size needs to be filtered identically, therefore the auto-correlation in the denominator. In addition to filtering with the template for the best signal size estimate, we can also search for the best signal time estimate identically utilizing the template's small variation in time as the filter, which becomes the second equation in Eq. (B.6) on $\dot{\hat{S}}(t-\tau)$. The two equations combined then determine an unique pair of the optimal A' and τ. Combining Eq. (B.1) and Eq. (B.6), we find that the OF expects a "true" signal embedded in the noisy data that is given by

$$A' = A \times \frac{\langle \hat{S}(t) | \hat{S}(t-\tau) \rangle}{\langle \hat{S} | \hat{S} \rangle} + \frac{\langle N(t) | \hat{S}(t-\tau) \rangle}{\langle \hat{S} | \hat{S} \rangle}. \tag{B.7}$$

If τ is estimated correctly for the first term of Eq. (B.7), the estimated signal amplitude A' should be unbiased from the expected result A up to the following filtered noise term. One may then understand the result as multiplying the true amplitude with a fractional bias factor due to signal time estimate, and then the smeared amplitude is combined with the noise contribution that survives the template filter.

Eq. (B.7) suggests that, since $\hat{S}$ is fully defined in shape and size while N fluctuates unpredictably in time, the uncertainty of A' is predominately due to the second term of Eq. (B.7), which is independent of the observed signal or the signal template. More importantly, this uncertainty does not vanish even in the absence of signals. It indicates that there exists a threshold above the non-detection that prevents us from claiming a reliable signal detection for smaller signals. Although the signal excursion from non-detection overcomes this obscuration of pure noise continuously thus does not provide a precise threshold for detection versus non-detection, a standard definition for the threshold size, the so-called "baseline resolution," is with the standard deviation of A' in the absence of a real signal, i.e.,

$$\sigma_x[x = A'(A = 0)] = \sqrt{\overline{(x - \overline{x})^2}} = \sqrt{\overline{x^2} - \overline{x}^2}\,, \tag{B.8}$$

(We write x instead of A′, so Eq. (B.8) may be recycled for later uses. Note the specific arguments for the context of the baseline resolution.)
where the overlines denote the expectation values for the decorated quantities. When we proceed to really calculate this quantity, we find that we face the products of infinite series in time and, similar to our treatment in the electronic noise removal (Eq. (4.177)), the simplest method to reduce the complexity is expending each term into its Fourier integral and arguing that only the non-oscillating products survive the infinite-time integral. The result indicates that the problem exhibits a clear one-to-one Fourier mode correspondence in the frequency domain. Specifically for Eq. (B.7), each $e^{i\omega t}$ mode in the signal only survives the integrals when it multiplies with the $e^{i\omega'(t-\tau)}$ mode from the template, which has

$$\omega t = \omega'(t - \tau) \Rightarrow \Delta\omega = \omega' - \omega = \omega\left(\frac{\tau}{t}\right). \tag{B.9}$$

We have now demonstrated that working in the time domain unnecessarily complicates the OF calculation due to mixing independent frequency modes, so we state without a proof that, if one is willing to carry out the full calculation as prescribed above, one can arrive at the identical results that we derive in the frequency domain in the next section. The true value of this time domain analysis is to show the physical meaning of the OF, which is utilizing the cross-correlation with the assumed signal template to filter the noise and then identify the most possible signal component in the data.

B.2 Standard optimal filter formalism

We now provide a formal introduction to the frequency-domain OF that we utilize to acquire signal sizes and model the resolution. We being by converting the identical time-domain signal into frequency space using Fourier integral,

$$\begin{aligned} S(t) &= A\hat{S}(t) + N(t) \\ &= A \sum\!\!\!\!\!\!\int \left[\tilde{S}(f)e^{2\pi i f t}\right]\delta f + \sum\!\!\!\!\!\!\int \left[\tilde{N}(f)e^{2\pi i f t}\right]\delta f \end{aligned} \tag{B.10}$$

$$\rightarrow \tilde{S}(f) = A\tilde{\hat{S}}(f) + \tilde{N}(f), \tag{B.11}$$

where the tilde symbols denote the Fourier amplitudes (transforms) for the corresponding time-domain quantities. Each mode in the Fourier expansion is linearly independent, so we suppress the summations for the frequency in the following for simplicity. We redefine the squared deviation in frequency domain by

$$\begin{aligned}\chi^2 &= \sum\!\!\!\!\!\!\int \left(S(t) - A'\hat{S}(t-\tau)\right)^2 \delta t \\ \rightarrow \chi^2 &= \sum\!\!\!\!\!\!\int \left[\frac{\left|\tilde{S}(f) - A'\tilde{\hat{S}}(f)e^{-2\pi i f\tau}\right|^2}{\left|\tilde{N}(f)\right|^2}\right] \delta f. \end{aligned} \tag{B.12}$$

There are two modifications added to the frequency-domain χ^2. First, instead of carrying the τ-offset argument everywhere as in the time-domain derivation, the offset is equivalent to multiplying a $e^{-2\pi i f\tau}$ factor to every Fourier mode. It allows us to use the signal template $\tilde{\hat{S}}$ and the data $\tilde{S}$ on the identical argument f, while having a phase offset $e^{-2\pi i f\tau}$ to make up the waveform's offset in time. Second, unlike the loose χ^2 definition we used previously for illustration, we carefully define an actual "Chi-square statistic" χ^2 with a proper normalization. Comparing Eq. (B.12) to Eq. (B.11), one can now easily see without a proof that, if A' and τ are estimated accurately, the true signal component in the noisy data should be exactly canceled in the χ^2, leaving only the noise contribution $|\tilde{N}|^2$ in the numerator of Eq. (B.12). This is the identical conclusion that we provided in the time-domain analysis, which also justifies the $|\tilde{N}|^2$ in the denominator for a proper normalization.

We again search for the χ^2 minimum and find

$$0 = \sum\!\!\!\!\!\!\int \left[\frac{-2\tilde{S}^*\tilde{\hat{S}}e^{-2\pi i f\tau} + 2A'\left|\tilde{\hat{S}}\right|^2}{\left|\tilde{N}\right|^2} dA' + \frac{2\tilde{S}^* A'\tilde{\hat{S}}e^{-2\pi i f\tau}\cdot 2\pi i f}{\left|\tilde{N}\right|^2} d\tau\right]\delta f. \tag{B.13}$$

Providing the cross-correlation theorem

$$\langle p(t)|q(t)\rangle = \sum\!\!\!\!\!\!\int \left(\tilde{p^*}(f)\tilde{q}(f)\right)\delta f, \tag{B.14}$$

Eq. (B.13) is identical to Eq. (B.4) except the $|\tilde{N}|^2$ normalization. We are also allowed to suppress all the frequency argument "(f)" in the above result for simplicity, since they are all identical now. However, there are a few subtle differences worth highlighting. First, due to multiplying $\left(\tilde{\hat{S}}e^{-2\pi i f\tau}\right)$ to its own complex conjugate, the $A'\left|\tilde{\hat{S}}\right|^2$ term has the τ dependence canceled mode-by-mode, which echos our previous statement that only nonoscillating terms survive and contribute to the amplitude estimate. We have now explicitly shown that the contribution is independently from each mode and does not mix. Second, a similar phase cancellation applies to squaring the $\left(A'\tilde{\hat{S}}e^{-2\pi i f\tau}\right)$ term in the χ^2 and leads to a quantity that is independent of τ and therefore does not survive the ∂_τ in Eq. (B.3). The interesting consequence is an apparent "missing" term in the ∂_τ section of Eq. (B.13) compared to Eq. (B.4), which we will explain in a moment that encodes the physics obscured in the time-domain derivation. Now, we again separately equate the $\partial_{A'}$ and ∂_τ parts of Eq. (B.13) to zero and, from the $\partial_{A'}$ equation, find the expression for the amplitude

$$A' = \frac{\displaystyle\sum\!\!\!\!\!\!\!\int \frac{\tilde{S}^* \tilde{\hat{S}} e^{-2\pi i f\tau}}{\left|\tilde{N}\right|^2}\delta f}{\displaystyle\sum\!\!\!\!\!\!\!\int \frac{\left|\tilde{\hat{S}}\right|^2}{\left|\tilde{N}\right|^2}\delta f} = \frac{\displaystyle\sum\!\!\!\!\!\!\!\int e^{2\pi i f\tau}\tilde{\phi}^* \tilde{S}\,\delta f}{\displaystyle\sum\!\!\!\!\!\!\!\int \tilde{\phi}^* \tilde{\hat{S}}\,\delta f}. \tag{B.15}$$

The result is essentially identical to Eq. (B.6) but is expressed with the phase shift by τ, the $\left|\tilde{N}\right|^2$ normalization for χ^2 , the cross-correlation theorem, and most importantly for this section, the explicit definition for the "optimal filter"

$$\tilde{\phi}^*(f) = \frac{\tilde{\hat{S}}^*(f)}{\left|\tilde{N}(f)\right|^2}. \tag{B.16}$$

Note that we swapped the orders of the OFs and their preceding $\tilde{S}^*$ or $\tilde{\hat{S}}^*$ terms by changing complex conjugations, so the OF reads more like an operator that decorates the quantities to be filtered.

Recalling our earlier interpretation of filtering the noise by cross-correlating the noisy timestream with the expected signal template, also considering the cross-correlation theorem that links the cross-correlation to Fourier mode multiplication, we have now explicitly proven the interpretation and the equivalence of Eq. (B.15) and Eq. (B.6). We demonstrate that the OF formalism "optimally," in terms of minimizing χ^2, filters the noise in the data by comparing the assumed signal template, and then the signal size is obtained by comparing the filtered data to the also filtered signal template. The statement applies to both time- and frequency-domain OFs, but in the frequency domain, we are able to obtain a well-defined filter strength for each frequency as in Eq. (B.16). The filter strength is uniquely determined only by the expected signal versus the noise of the same frequency and does not mix with other frequencies. Note that one should not confuse this signal-and-noise comparison for the OF as a SNR; notice that the OF carries the unit of inverse amplitude. Since in practice every pulse shape and noise source elaborated previously exhibits its characteristic frequency range, Eq. (B.16) is a powerful concept that immediately shows the signal-to-noise contribution at each frequency for the optimal detection. However, we note again that the OF formalism works only on predefined signal templates, and we will examine the practicality of this prerequisite when we introduce the realistic considerations modeling the energy resolution.

We still need a prescription for calculating τ for Eq. (B.15). We equate the ∂_τ part in Eq. (B.13) to zero to constrain τ, which simply yields after eliminating all the non-zero constants

$$\tilde{S}^* \tilde{\hat{S}} = 0 \text{ if } f \neq 0. \tag{B.17}$$

We rearrange the result into a more interpretable form (cont'd)

$$\Rightarrow \partial_\tau \left[\sum\!\!\!\!\!\!\int \left(\tilde{S}^* \tilde{\hat{S}} \right) \delta f \right]_\tau = \partial_\tau \left[\langle S | \hat{S} \rangle \right]_\tau = 0. \tag{B.18}$$

Eq. (B.18) indicates that if $\hat{S}$ is offset by the most appropriate τ estimate, the product of the data timestream and the signal template should be a constant that is independent of time, or in the presence of noise, a minimized integral of $\left(\tilde{S}^* \tilde{\hat{S}} \right)$ due to pure noise contribution. This result is exactly what we expect if the signal template perfectly aligns with the actual signal embedded in the data. Interested readers may continue using the true signal information, namely Eq. (B.11), to derive the theoretical timing resolution ($\Delta\tau$), but it is easy enough to see from Eq. (B.13) that

$$\Delta\tau \sim \begin{cases} A^{-2} & \text{, large } A \\ \dfrac{\sum\!\!\!\!\!\!\int |\tilde{N}| \, \delta f}{A} & \text{, small } A, \end{cases} \tag{B.19}$$

which is consistent with the expectation that the timing resolution is dominated by the signal size alone or the relative noise-to-signal fluctuation in the signal- or noise-dominated regime, respectively. However, as long as the pulses trigger the DAQ at a conventional significance level, in practice the intrinsic timing resolution of the OF is most likely subdominant to the timing/sampling uncertainty given by the DAQ electronics, e.g., Fig. 2.16, therefore oftentimes neglected.

We are now fully prepared to calculate the signal size A' and its corresponding resolution $\sigma_{A'}$. In particular, we are most interested in the baseline resolution that is defined for a signal-less condition ($A = 0$), which provides a comparable resolution/threshold to be compared to the expect signals as well as other detection methods. Since the baseline resolution only concerns the noise and the predefined signal template, we can in fact predict it before observing the anticipated signal, e.g., the yet observed DM, and regard it as the idealized minimal resolution forecast. Having derived the formula for the estimated signal size in Eq. (B.15), it is only a matter of substituting the formula into the definition of the standard deviation (Eq. (B.8)) to obtain (algebra in Appx. A)

$$\sigma_{A'} = \left[\sum\!\!\!\!\!\!\int \frac{\left|\tilde{\hat{S}}\right|^2}{\left|\tilde{N}\right|^2} \, \delta f \right]^{\frac{-1}{2}} . \tag{B.20}$$

The above expression indeed only depends on the signal template assumption and the pure noise data, both available prior to an empirical resolution calibration with real pulse signals. In the more detailed derivation in Appx. A, we start from substituting the signal expression in Eq. (B.15) into the definition of σ (Eq. (B.8)), and then by letting either $\tilde{S} = \tilde{N}$ for a pure noise calibration scenario or $\tilde{S} = A\tilde{\hat{S}} + \tilde{N}$ for examining arbitrary signals, where the true signal component is exactly canceled as expected, the equation always reduces to calculating the standard deviation for the random signal sizes A' estimated from pure noise samples. Such a baseline calibration process is physical and therefore can be realized by really collecting a set of signal sizes estimated from many different random noise samples using Eq. (B.15). The resulting A' should exhibit a near-Gaussian[1] distribution that centers at zero with a real standard divination given by Eq. (B.20). The centering at zero indicates that the OF signal size estimate is unbiased, or otherwise indicating an observable systematic bias in the noise data. We realize that the physical meaning of the baseline resolution is the standard deviation of the OF-reconstructed signal sizes A' for random noise samples due to an incompletely filtering, which is nevertheless optimal in terms of yielding the lowest possible χ^2 for the anticipated signal.

[1]Exactly Gaussian if the noise is perfectly stochastic.

Appendix C

PRACTICAL TECHNIQUES FOR DMKID DEVICE FABRICATION

C.1 Substrate preparation

- Always keep in mind the pattern to be fabricated and only tweeze/clamp at the *same* spots that are as far as possible from critical features. Since one would immediately start handling the wafer with tools, these designated spots should be chosen before the entire fabrication process. Unfortunately, it is in practice difficult to constantly reclean the tweezers and tools, so these designated tweezing spots would almost always end up being coated with some particles or drying stains, which are also micro dusts captured by droplets and then dried. Specifically for the DMKID devices presented in this thesis, we typically use the 1–2 and 10–11 o'clock edges for tweezing, where the wafer flat defines the 6 o'clock edge. One may see the two chosen edge sections are the furthest from the meandering feedline patterns.
- Based on our experience, roughly one in ten of Sil'tronix's Si wafers would have visible defects, dents, spikes failed to be polished, or scratches that may be identified with bare eyes out of the box. The technique to locate these defects is by slightly turning the 3-points wafer dipper and use the reflection to contrast the marks. Since we always order double-side polished wafers, do not use the defected surface for instrumentation[1].
- Given the cleanroom's constant particle-carrying downward airflow, one should always avoid exposing sensitive surfaces, such as the wafers or the photomasks, upward toward the air without containers or temporary covers. Note that the surface for the temporary coverage facing the covered object should also be cleaned to avoid particle-shedding.
- In most situations, if the handling sequences are thought through in advance, one actually does not need a tweezer to transfer the wafers/masks. As we have mentioned, we identify tweezing as one of the main causes for dust particle contamination, so one should always avoid using a tweezer or any close-

[1]Actually, just pick another wafer if available.

contact tools whenever possible. Examples for handling the wafers without tools include:

- Transferring the wafers onto other fixtures with wafer holders and lids, e.g. opening a wafer holder upside-down while the wafer sits in the lid on the instrumented side, covering the lid with the surface (e.q. chuck) the wafer is to be transferred to, and then flipping both together so the wafer drops onto the final surface due to gravity.
- Transferring the wafer with a cleanroom wipe *if it is already on one*, e.g. after hot-plate drying. Note that it does not suggest cleanroom wipes are particle-free and may be used for clean touches.
- Handling from the backsides of the wafers if accessible, e.g. when the wafer is on a spin coater suction pole or the MA6 contact mask aligner, which has a middle cut-out in the chuck for users to reach the wafers from the backside.
- When the wafer is attached to a larger backing wafer, one should obviously hold the backing wafer but not the under-construction device.

• When loaded onto the 3-points wafer dipper, face the instrumented surface away from the handle to avoid particle shedding and scratching.

• Per MDL safety guideline/SC group wet bench convention, one should always use beaker lids labeled with chemical names, especially for clear liquids. It is nevertheless allowed and we also strongly recommend putting these labeled lids on the side of the beakers but not on top. Especially for those frequently used Acetone, IPA, and DI-water lids piling on the bench hood, we noticed they have never been cleaned, very dirty, and may easily shed particles into the beakers.

• Pour the first 50–100 cc of any liquids from any containers, faucets, or spray guns into the sink for cleaning the instruments as well as the liquids themselves. In particular, we found the liquids and instruments in MDL H6 are noticeably dirtier in the morning when people just start the daily fabrication cycle.

• Do not finish rinsing with squeeze bottle liquids; they are generally very dirty. Instead, one should try to finish with faucet (the cleanest) DI-water whenever possible.

• After sonication, first pour the solvent into the sink before removing the wafer/wafer dipper from the beaker. Most particles detached from the sonicated wafer would float on the surface due to surface tension, so one may reintroduce the particles onto the wafer when pulling it through the surface before removing this layer.

- Dry the wafer on a hot plate using N_2 blow gun as soon as it leaves the liquid for the last dipping step. If any droplet is let dried by natural evaporation, even in a highest-class cleanroom[2], one would always obtain a circular drying stain, which is an indication of micro-dusts captured and marked onto the wafer by the liquid. It is because normal cleanrooms for human workers are, limited by practicality, defined only by the density of particles $\gtrsim \mu$m.
- Following the last bullet point, however, we would purposely create small drying stains at the designated edge tweezing locations on the contact masks, after cleaning the masks and before hard-contact alignments, to ensure the wafers and the masks do not attach to each other due to vacuum suction when both surfaces are as clean as required; detaching a wafer from a mask held together by atmospheric pressure usually results in disastrous consequence to the device.
- One should always use a layer of cleanroom wipe between a wet wafer and the drying platform, such as the hot plate or photoresist (PR) developer bench, to prevent capturing particles from the surface with liquid, hence cleanroom wipes are typically provided besides/atop the hoods. Meanwhile, similar to the chemical labels and squeeze bottles prepared for the bench users, these wipes have been collecting dusts thus are dirty on the front surface. We recommend avoiding the top draw, i.e., drawing directly from the lower part of the stack, and definitely flip to use the backside of the wipe.
- Whenever possible, finish blow-drying or blow-gun cleaning with a few last blows, before immediately enclose the workpiece with a clean container, while the blown-cleaned surface faces downward or sideways, so the removed particles do not re-plate back onto the surface.
- Avoid mix-use of tools (tweezers, beakers, funnels, etc.) with others. We found the bast practice is to clean a set of sharp-tip, flat-head, round carbon-tip, and wafer-edge tweezers together with a small beaker, using the same solvent-cleaning procedure for the wafers, and carry the tools in the beaker for personal use.

[2]Unless it is a vacuum chamber not suitable for human workers but robots.

- It is always helpful and likely turns out saving rather than spending more fabrication time, when using a dummy wafer for uncertain/under-development processes, especially those requiring controls of doses by timing, while the dosage rates are known to drift over time, e.g. PR UV exposure intensity ages with the bulb, or long-stored chemicals that may be depleted, e.g. HF vapor. It is definitely less favorable to use the main under-construction devices to test out potentially irreversible yet new or variable processes.
- The chuck for HF vapor deoxidation is carefully covered by Kapton tapes, so only non-metal, dull-tip, plastic tweezers are allowed when handling wafers on the HF chuck.
- Recent studies have confirmed HF-deoxidized Si surfaces can only remain oxide-free for a few minutes and would be fully oxidized, i.e., $SiO_{(2)}$, within 15 minutes (Altoé et al., 2020). While currently there is no direct evidence for partial Si oxidation to contribute to the TLS noise, the degree of SiO_2-oxidation has been unambiguously linked to the TLS noise level. So for minimizing the time for potentially reverting the fully deoxidized wafer, one should confirm in advance the availability of the sputtering chamber before HF deoxidation and ideally transfer the wafer into the vacuum chamber within a few to 10 min. after the HF vapor deoxidation.
- Bolting wafers too tightly on the sputtering chuck could damage the crystal lattice, or even crush the wafer, immediately or during the sputtering thermal cycle, so we have produced a pressureless ring tool for fixing $\varnothing 3" \times 1$ mm wafers in place on the chuck, with tiny teeth extending onto the wafer area to prevent it from falling. Not only does the tool apply zero stress to the wafer, but it also reduces assembling tool contacts, unsputtered areas, and is fast to assemble thus also minimizing the post-HF air exposure time. We highly recommend future DMKID fabricators to adapt similar techniques.
- It is totally acceptable to reuse old silicon wafer carriers with the "backing spiders[3]" stored in cleanrooms, such as in the storage bin in the entrance corridor of MDL H6. It is, however, apparently unacceptable to use a wafer carrier, even a new one, before a thorough upside-down[4] blow-gun cleaning, leaving it open, or uncovered. It is also a safety precaution for the device,

[3]Usually too tight for 1 mm wafers in the carriers.
[4]c.f. previous bullet point, for avoiding particle replating.

as well as courtesy to other colleagues, that one correctly labels the carried device as soon as the wafer carrier is taken.

C.2 Photoresist patterning

- Use a dummy wafer to test the vacuum suction and the stability of the spin-coater. Due to the accumulation of uncleaned PRs, it is not uncommon for a spin-coater to show an uneven weight distribution, which would significantly affect the balance hence the quality of spin-coating, or even lose vacuum suction in the middle of spinning due to clogging, which would obviously be disastrous.
- As noted in Tab. 4.3, we always utilize a dynamic dispense, where we spin the wafer at 75 rpm while steadily dropping the liquid onto the center of the wafer, and then we release the throttled spinning to reach the coating speed. We find the technique significantly improves the human handling-related PR application quality. On the other hand, the procedure requires a rather complex coordination of smooth apparatus setting/control, wafer and dropper handling, and arm movements that minimize particle shedding. We recommend future DMKID fabricators to particle the technique with dummy wafers until confident and adapt the technique into their processes.
- Only use fresh dispensable plastic droppers, blow-clean the insides and outsides of the droppers, and label the droppers to avoid mix-use[5].
- Gently blow-clean the wafer on the spin-coater *right before* PMMA/PR application, since any particles plated and subsequently trapped under the PMMA/PR would eventually be printed to the lithographic pattern. Note that strong blows may agitate the dusts in the dirty spin-coater sink therefore is also hazardous when the wafer is facing upward unprotected.
- To achieve the smoothest process where the PR/PMMA is immediately bake-cured at its perfect distribution and cleanness, while there are multiple hot plates preset to different temperatures, located side-by-side, potentially being scheduled/used simultaneously by many people, it is generally helpful to first identify the exact baking plate at the intended temperature, confirm its availability, and announce for a few-minutes exclusive use before starting the spin-coating.

[5] We usually place the droppers on a fresh cleanroom wipe and write on the wipe.

- Squeeze the dropper before entering the bottled liquids so to minimize carrying dirty air with dusts to contaminate the entire bottles. Take the liquids from significantly below the surfaces, since particles are mostly held on surfaces by surface tension.
- To clean the dropper, we usually squeeze the first 20–30% of the liquid directly into the sink before applying it to the wafer.
- When using a dropper, one naturally pushes his/her thumb “toward” the index finger, which would lead to a net force toward one side, bend the light plastic dropper, and splash across the wafer, generally leading to unacceptable results. To achieve a high-quality, fringe-less spin-coat, the correct technique is first squeeze out any bubbles to ensure a continuous flow (do not tap because it might introduce particles), and then mindfully squeeze thumb, index finger, and middle finger, forming a rough triangle, evenly toward the central axis of the dropper, while the tip of dropper should point vertically and steadily toward the center of the wafer. One should try to extract the correct amount of the liquid, so he/she may comfortably push out all the liquid onto the wafer and then draw back his/her arm while the dropper maintains squeezed; it prevents random drops to fall on the wafer by accident when the dropper is released. The hand/arm movements should also be choreographed to avoid mindless waving atop the wafer that sheds particles. Again, we identify the application of PR as one of the critical steps to the success of the fabrication, so it should be done as carefully as possible.
- PMMA is sensitive to latex, so one should not touch the wafer with latex gloves after the PMMA is applied.
- According to our extensive trial-and-error tuning, we find 3 min. baking time is minimally required to fully cure the selected PMMA, with sometimes up to 5 min. being fully reliable, which is critical for avoiding mixing with the PR and subsequently leading to poorly controlled RIEs. Meanwhile, unnecessary baking may also potentially alter the stress thus the superconductivities of the SC films. Since we have found, if maintaining the 115 °C soft baking temperature, the baking time may vary significantly in 3–5 min., we recommend, if the exact sources for the PR and PMMA have not been used for exactly the process in hand for a week or longer, it would be very helpful to first test with a dummy wafer the appropriate combination for the PR/PMMA application conditions, baking times, UV doses (power $\times$ time), post UV-baking

conditions, and the development conditions, until one establishes a certain combination of all the above that guarantees a high quality result. According to our experience, it typically takes 3–5 iterations of stripping/reapplying the PR, re-exposure/baking, and re-developing, to identify a precise procedure, which is certainly undesirable to find out with the main devices.

C.3 Photo-mask alignment

- Before using the MA6 aligner, make sure its exposure mode is set to hard contact (hard-ct/wec) but not flood exposure, which would immediately start the exposure without an aligning step and ruin the wafer.
- To prevent the deeply cleaned mask and wafer from attaching to each other by vacuum suction/atmospheric pressure, one may purposely create a small drying stain on the mask at the designated wafer edge tweezing location.
- Install the UV filter, the mask, and the wafer in a top-to-bottom order, i.e., in this written order, to avoid particle shedding to the lower instruments while the above ones are being worked on, and only remove each instrument from its enclosure and clean before installation.
- The mask holder plate and the wafer chuck, although less critical due to their appreciable separations from sensitive regions, should also be blow-cleaned before use.
- To achieve a more precise hard-contact mask-wafer co-leveling, which we find significantly improves the uniformity of the UV dosing, subsequently the PR developments, and eventually the RIE results, one should always choose the most appropriate wafer chuck according to the wafer dimensions. Specifically for our 3"×1 mm Si wafers in MA6, we find the chuck with the largest smaller-than-3" backing platform gives the best result. In general, one may typically start with the backing platform that is slightly smaller than the wafer, but the best choice may vary depending on the stiffness, thickness, and shape of the wafers.
- In principle, a minimal UV dose that exactly disintegrates the exposed pattern-edge regions should give the sharpest photolithography, thus favoring the shortest sufficient exposure time. However, under-exposed PR would typically result in undeveloped/unetchable patterns that lead to simply unusable final features. While we find the suggested 120 sec. being the typical minimal appropriate time for the MDL-nominal 25 mW/cm^2 UV power, we do notice,

likely due to the aging of the UV source, the indicated UV power might be inaccurate and in some cases require exposure times as long as 150 sec. Since it is practically impossible to realign, after development, a mask to an under-exposed pattern without introducing further smearing, the experience motivates testing the variable lithographic process with a dummy wafer in advance, and if needed, slightly/incrementally extend the UV exposure by 5–10 sec/ea.

- If all the PR, PMMA baking, exposure, and post-baking are executed correctly, the UV-printed pattern should appear at about 40 sec. into the post-UV bake.
- Per MDL SC group convention, the nominal UV filter to be kept in the MA6 aligner should be the i-line filter, so do remove our deep-UV filter and reinstall the i-line one after use, and also label the "UV300 in-use" status during operation.

C.4 Cl RIE for Al KID definition

(a) Initialization: LL evacuation and automatic wafer transfer.
Focus the end-point laser detector to where the Al film is to be removed, preferably also with PR-protected features in the monitor field for visual comparison. Since we anticipate a 50% decrease of reflection, when the Al is removed to reveal bare Si, as the end-of-RIE signal, adjust the laser aperture so the initial end-point signal is higher than 90%. If the post-RIE cleaning facilities, i.e., ashing and wet bench, are not yet available, one may choose to wait at this step indefinitely.

(b) O_2 flow stablization: Injecting ashing and assisting gases.
Manually proceed when pressure stabilizes.

(c) O_2 plasma activation: Turning on RF power to initialize ashing plasma.
The chamber pressure, RF power, and RF bias are temporarily adjusted to 12 mTorr, 100 W, and 40 W, respectively, to activate the Oxygen plasma. The step automatically proceeds in 10 sec.

(d) O_2 RIE: Main ashing.
This step uses parameters in Tab. 4.5 to indefinitely etch the device until manually proceeded, so do click "next step" in 3 sec. if it is unneeded. Since this ashing recipe is mainly for an overall cleaning for the wafers before the main RIE, it is chosen to have less distinction between PR and PMMA etch rates, so, together with the reasons detailed in the main content, it is less preferable to be

utilized as the main PMMA removing process. Since we have applied extensive dedicated cleaning processes in advance, one may choose to immediately proceed or allow the O_2 cleaning for a few sec. just to clean the surface. Note that, if the PR region starts to show a color change on the monitor, it means the PR is being significantly reduced by the ashing and should definitely be avoided.

(e) BCl_3/Cl_2 flow stablization: Purging ashing plasma and injecting Chlorine gases. One needs to manually proceed from this step when the chamber pressure stabilizes. Be very careful, once proceeded, the RF power is turned on and the Al etching begins immediately, so do confirm the process to be expected before continuing.

(f) BCl_3/Cl_2 RIE: Main Al etching.
Once started, the Cl plasma would instantly penetrate the thin TiN and start removing the Al film at the noted rate (Tab. 4.5). Based on our experience, the end-point signal should start decreasing at 25 sec. and plateau at 28 sec., with a total decrease of about 50%. The timing may vary by ±1 sec., for which we typically also let the plateauing lasts for another 1 sec. to accommodate the etch rate variation across the wafer, as well as confirming the signal indeed marks the completion of the process but not a slowly-updated noise fluctuation (the signal refreshes as sub-to-1 Hz) . As the end-point laser signal would sometimes fluctuate due to mechanical vibration to falsely generate plateau-like changes during the stressful dynamic process, it is helpful to also cross-compare the camera image, which, if adjusted to contain KID features as recommended previously, typically starts to show unmistakable large area, dramatic color/contrast changes 1–2 sec. *before* the end-point laser signal drops. We should note, however, depending the exact materials or viewing angles, not all the RIE processes would show easily identifiable contrast changes through the camera, and it is certainly less accurate for end-of-process determination, which is particularly crucial for preventing over-etch and sidewall corrosion. We again emphasize, our fabrication does suffer from potential corrosion if the residual Cl atoms are not immediately purged from the etched film, letting along exposed to active Cl plasma for any unnecessary periods of time. In addition to damaging the Al features, we have also discovered by accident that, if the already-exposed Si surface is further etched by our Cl recipe for 10–20 sec., we may create the "black silicon" surface structure, which indicates a drastic surface dehomogenization that is most likely detrimental to both the substrate phonon energy distribution and the device's radiation/thermal properties. So for all the reasons above, one

should immediately terminate the etching process once completed (set to manual control). Finally, it is worth to note, while BCl_3 provides a relatively small etching power to the Al than $Cl_{(2)}$, it was realized by R. LeDuc that the Boron plasma may assist the penetration of the surface Al_xO_y, which always presents but at uncontrolled levels, and lead to much more uniform (horizontally) and vertical etches, hence subsequently higher quality superconducting devices.

(g) Chamber purging, end of process.
Note that currently this step is set to automatically vent the LL (with N_2, not air) once the device exits the main chamber. Although the current choice, if any, probably introduces minimal negative impacts to the devices, future uses may consider separating the device extraction and venting for more flexibility.

In addition to the detailed step-by-step RIE instruction, we provide a few more handling tips:

- First and foremost, make sure the asher is available for an immediate post-RIE Cl purge before initiating the RIE.

- Instead of sending the 3" device wafer as is into the etcher, which will be detached from the chuck and held by three narrow pins from the back, we find carrying the device with a 6" large backing wafer prevents the issue, where sometimes smaller/lighter 3" devices may be dropped from the pins due to the cooling He flow or during transfers. So far since we have been producing devices mainly for R&D purposes, we simply use Kapton tapes to hold the device at the center of a 6" backing wafer coated with thick PR. Because the tape-covered areas would result in unetched SC film patches, we always tape the tweezing spots at the edges of the wafers, covering areas as narrow as we can, typically smaller than 1 mm across.

- As mentioned previously, we expect the end-point signal to decrease by about 50% and then plateau. Since it is crucial that one can accurately determine the end of the process from the signal, it is important to adjust the aperture, usually in the range of f/8–f/4, so that the initial signal is high enough, but not saturating, to avoid bottoming before the signal plateaus.

- Etching is a completely irreversible process but generally stoppable/resumable, so it is important that the fabricator has a clear expectation to the anticipated signal progression before executing the process, and if the process progresses otherwise, immediately shut the RF/plasma (or leaves the wet etcher), retrieve the device to understand the issue, and may resume if allowed.

- Although we recommend removing the PMMA in a separated ashing step beforehand, in principle one may still choose to do so by the pre-RIE O_2 cleaning in the RIE system. In this case, providing the wavelength of the laser (green), the thickness of the PMMA, and its refractive index, the signal should start in a close-to-constructive film interference, and as the PMMA is being removed, the signal should show exactly one sinusoidal-like decrease and recover to a different level due to PMMA and Si's different reflectances, and the time should be 60–70 sec. proving the etch rate in Tab. 4.5. Again, due to multiple practical factors, the readers may understand this is a less controllable option for PMMA removal.

C.5 F RIE for feedline definition

(a) Initialization: LL evacuation and automatic wafer transfer.
Focus the end-point laser detector to where the Al film is to be revealed after the Nb on top is etched through, e.g. KID inductors, preferably also with regions that do not have Al features underneath, and those covered by PR, e.g. CPW, in the monitor field, so bare Si surface will be revealed for high-contrast visual comparison, while one may also confirm the PR-protected features still stand after the long etch. We aim directly to the Al in order to minimize over-etches/damages once the Al is exposed. Since we are anticipating a 40–50% *steep* increase of reflection by Al, adjust the laser aperture so the initial end-point signal is around 20–30%, and unlike in the Al feature-defining RIE where we identify unambiguous plateaus, here we shut off the RF immediately when the rising slope shows a distinguishable slowdown.

(b) O_2 flow stablization: (Same as Cl RIE)

(c) O_2 RIE: (Same as Cl RIE)
May skip if unnecessary.

(d) CHF_3/O_2 flow stablization: Purging ashing plasma and injecting main RIE gases.
One needs to manually proceed from this step when the chamber pressure stabilizes. The purpose of Oxygen is that is has been shown, due to promoting temporary oxide formation on sidewalls, which is easier than the metallic film

for the main etcher(s) to react with, adding the small amount of Oxygen improves the verticalness of the RIE results.

(e) CHF_3/O_2 RIE: Main Nb etching.
As noted in (a), we look for a steep signal rise as the end-of-process Al exposure. While we suggest immediately terminate the process when the signal is relatively dominated by Al over Nb, i.e., the transitioning of the reflection slows down as it approaches the higher value, depending on the location of the laser spot and the size of the wafer, the full Nb instrumentation might not be completed across the wafer due to known plasma nonuniformity, e.g. might still have a few nm of Nb film at high radius. So in order to achieve the most precise possible result, one should retrieve the wafer for a visual inspection and, if needed, continue the RIE at 1–2-sec. increments until the remaining Nb is visibly removed. Similar to our Cl-RIE recipe, we also found, besides damaging the Al features, our F-RIE recipe may also produce black Silicon if applied to bare substrate for extended periods of time.

(f) Chamber purging, end of process. (Same as Cl RIE)

Ideally after the Nb feature definition, we would also apply the same deep cleaning treatment as for the Cl etch, but it is worth noting that, empirically, it is not as time critical for our pure-F recipe to be high-pressure Oxygen RIE "immediately." On the other hand, since at this point the device is basically completed, all the remaining PR, or any carbon-oxide, hydroxide, organics formed during the fabrication, should all be completely stripped from the device, so we typically interleaving strong ashing, solvent cleaning, and sometimes strong PR stripping agents (Toluene), starting with 2 min. of the high-pressure O_2 RIE, until all the organic compounds are removed. If the design requires wafer dicing, we would first inspect for the final result after the full cleaning, and then coat the device with an easily-removable i-line PR for dust protection, perform the dicing, and strip the PR with solvents. Finally, we transfer completed devices from JPL MDL H6 cleanroom, using Silicon wafer carriers in N_2-inflated sealed bags, to a normal non-cleanroom laboratory environment at Caltech for subsequent wire-bonding and characterization/operation, typically done in a few days to a week.