

PULSE SHAPE MODELING AND CHARACTERIZATION OF THE SUPERCDMS SUB-GEV
DARK MATTER A THERMAL PHONON MEDIATED DETECTORS

By

DAVID SADEK

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2026

© 2026 David Sadek

To Mom and Dad

ACKNOWLEDGEMENTS

I would like to begin by remembering my mother and dedicating this dissertation to her soul. I wish you were here to share this moment with me, but I know you would have looked at me with the same eyes full of love and pride that you always did. To my father and my siblings, thank you for your support through all these years. To my sister Sarah, who effectively raised me and became my second mother after my mom passed away, I have an immense amount of love and respect for you. Thank you for everything.

I would also like to thank Dr. Ahmed Zewail for creating Zewail City, which gave me the chance to get to where I am today. Thank you to all of my undergraduate professors and friends, the community that shaped a large part of who I am. Our pursuit of knowledge and our professional goals separated us, but I hope that fate brings us all back to Egypt one day to benefit our home. To Moaaz, Saadawy, Robair, Madonna, Mariz, and Mohamed Hasan, thank you for being such wonderful friends and for being like brothers and sisters to me.

To Matthew Reinhard, Brooke Emison, Jay Motka, Dan George, and Regan Bhatta, thank you for being such good friends, and for all the discussions about physics, geopolitics, and the human condition. In you, I found an intellectual fulfillment I had previously only known with my friends from undergrad.

To my advisor, Tarek Saab, thank you for your patience, support, and understanding throughout my training, and for giving me the chance to take part in such cutting-edge research. From the SuperCDMS collaboration, I would like to thank Ziqing Hong for the in-depth discussions, the jokes, and the laughs. I extend my thanks to the members of the DMC group, in particular Mike Kelsey, Stefan and Birgit Zatschler, Miriam Diamond, Emile Michaud and Warren Perry. I would also like to thank Yan Liu and Aditi Pradeep and Sukee Dharani for the interesting discussions and analysis work we did together in the CUTE tower testing run as well as SNOLAB detector characterization during the commissioning phase

In the summer of 2024, I had a wonderful experience at the SuperCDMS shifters' house. I am grateful to Ruslan Podviianiuk, Muad Ghaith, Shubham Pandey, Joel Sander and Emanuele Michielin for making it what it was.

I would also like to extend my gratitude to the organizers of SUSi 2024 especially Stephen Sekula and Jodi Cooley for the rich academic and professional experience and advice they provided me with.

I want to close by thanking everyone who played a role in supporting and helping me reach the end of this journey .

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGEMENTS	4
LIST OF TABLES.....	9
LIST OF FIGURES.....	12
LIST OF ABBREVIATIONS	13
ABSTRACT.....	14
CHAPTER	
1 THE DARK MATTER PROBLEM	16
1.1 Baryonic Mass Inference	16
1.2 Gravitational Evidence for the Mass Discrepancy	21
1.2.1 Mass from Gravitational Dynamics.....	21
1.2.2 Mass from Gravitational Lensing.....	24
1.2.3 Mass from the Cosmic Microwave Background	25
1.3 The Case for Particle Dark Matter	27
1.3.1 Modified Gravity	28
1.3.2 Dark Matter as a New Particle.....	30
1.4 Dark Matter Cosmology and Particle Physics.....	32
1.4.1 Properties and Distribution	32
1.4.2 Elastic Scattering Kinematics	34
1.4.3 Interaction Models	37
1.5 The Experimental Landscape	43
1.5.1 Nuclear Recoil Searches	43
1.5.2 Electron Recoil Searches.....	45
2 SOLID-STATE DETECTOR RESPONSE	48
2.1 The High-Voltage Detectors	48
2.2 Energy Deposition and Ionization Yield.....	49
2.2.1 The Semiconductor Crystal	50
2.2.2 Recoil Types	51
2.2.3 Ionization Yield.....	53
2.3 Charge Transport	55
2.3.1 Conduction Valleys and Intervalley Scattering	55
2.3.2 Neganov-Trofimov-Luke Phonon Emission.....	57
2.3.3 Charge Transport in Simulation	60
2.4 Phonon Dynamics	61
2.4.1 Initial Population, Propagation, and Modes	61
2.4.2 Bulk Scattering and the NTL Spectrum.....	63
2.4.3 The QET Sensor and the Phonon Cycle.....	65
2.4.4 Limitations of the Phonon Model.....	68

2.5	Position Dependence of the Phonon Response	70
2.5.1	Response of the Phonon Channels	70
2.5.2	Energy Asymmetry Between the Faces	71
3	TES ELECTROTHERMAL DYNAMICS	76
3.1	The Resistive Phase Transition	76
3.1.1	The Phase Boundary	78
3.1.2	Transition Broadening and Sensitivity Parameters	79
3.2	Electrothermal Feedback Architecture	85
3.2.1	The Voltage-Biased TES Circuit	85
3.2.2	The QET Thermal Architecture	93
3.3	Linearization and the Dynamics Matrix	102
3.3.1	Linearization	102
3.3.2	The Dynamics Matrix	104
3.4	Time-Domain Solution	107
3.4.1	Characteristic Polynomial and Eigenvalues	108
3.4.2	Eigenvectors, Mode Weights, and the Impulse Response	111
3.4.3	Stability and Electrothermal Oscillations	117
4	HVEV PULSE SHAPE TUNING AND SPECTRUM RECONSTRUCTION	120
4.1	The HVeV Detector at CUTE	120
4.1.1	The HVeV Detector Design	121
4.1.2	Experimental Setup and Data Acquisition	122
4.2	Templates and the LED Spectra	124
4.3	Phonon Power Input	127
4.3.1	CrystalSim Parameters	127
4.3.2	TES Sub-Gap Absorption	129
4.3.3	Rising Edge Slow Onset	131
4.4	Pulse Shape Tuning of TES Parameters	133
4.4.1	TESSim Model and Parameter Space	133
4.4.2	MCMC Framework	134
4.4.3	Global Fit and Parameter Degeneracies	135
4.4.4	Fixing Measured Parameters and the Final Fit	140
4.5	Charge Transport Tuning Against the LED Spectrum	144
4.5.1	Simulation Geometry and the CTII Model	144
4.5.2	Mean Photon Number and Comparison with CUTE Data	148
4.6	Reproducing the Long Tail with a Substrate Thermal Block	150
4.6.1	The Two-Thermal-Block Model	150
4.6.2	Heat Capacities	151
4.6.3	Fit Results	152
4.6.4	Discussion and Limitations	153
5	SUPERCDCMS DETECTOR TESTING AND OPERATION	155
5.1	The SuperCDMS Experiment and the CUTE Facility	155

5.2	Noise in the Phonon Channels	159
5.2.1	A Bias-Dependent Feature in the Noise Spectrum	159
5.2.2	The Noise Model	159
5.2.3	Comparison with the Measured Spectrum	161
5.3	Electrothermal Oscillations.....	164
5.3.1	Standard Single Block Electrothermal Oscillations Solution	164
5.3.2	Oscillation Region and Frequency.....	165
5.3.3	Application to the Detector	166
5.4	Performance of the SuperCDMS SNOLAB Detectors.....	168
5.4.1	Event Selection and the Noise Sample	169
5.4.2	Optimal Filtering and Baseline Resolution	170
5.5	Thermal Parameter Extraction	173
5.5.1	Method	175
5.5.2	Extracted Coefficients	176
5.5.3	Substrate Temperature and Relaxation Time at the Operating Point.....	177
6	SUMMARY AND OUTLOOK.....	180
6.1	Summary.....	180
6.2	The Substrate as a Dynamical Element.....	182
6.3	Relevance to the Dark Matter Search.....	183
6.4	Limitations	185
6.5	Outlook	186
6.5.1	A Semi-Analytical Model for Joint Parameter Extraction	186
6.6	Closing.....	187
APPENDIX		
A	HIGHER-ORDER RESISTANCE EXPANSION	189
B	EIGENVALUE ALGEBRA FOR THE LINEARIZED TES	191
B.1	Closed-Form Eigenvalues	191
B.2	Eigenvectors and Mode Weights	192
C	NUMERICAL INTEGRATION OF THE TES ARRAY EQUATIONS.....	195
C.1	The coupled equations for the sensor array	195
C.2	Critical-current suppression	197
C.3	Reparameterization and analytic inversion.....	197
C.4	A dynamical substrate temperature.....	199
C.5	The explicit system	201
LIST OF REFERENCES.....		202
BIOGRAPHICAL SKETCH		207

LIST OF TABLES

<u>Tables</u>	<u>page</u>
3-1 Impulse response parameters for different SuperCDMS detectors.....	119
4-1 Geometric parameters of the NFC QET.....	122
4-2 TES simulation parameters for the two NFC SO channels.....	144
4-3 Charge trapping and impact ionization mean free paths used in the simulation.....	148
4-4 One- and two-block fit parameters for the HVeV pulse.....	153
5-1 Parameters of the PES1 channel.....	163
5-2 Extracted thermal parameters for the two detectors.....	177

LIST OF FIGURES

<u>Figures</u>	<u>page</u>
1-1 The stellar IMF	19
1-2 Rotation curves of spiral galaxies	22
1-3 The Bullet Cluster	25
1-4 CMB angular power spectrum	26
1-5 Energy transfer fraction	35
1-6 Spin-independent differential recoil spectra	37
1-7 Spin-independent dark-matter–nucleon exclusion limits	44
1-8 Constraints on dark-matter–electron interactions	46
2-1 The high-voltage detectors	50
2-2 Ionization yield in silicon versus nuclear recoil energy	54
2-3 Simulated electron arrival positions in silicon	60
2-4 Simulated phonon emission spectrum	62
2-5 Phonon mean free path in silicon	64
2-6 Simulated NTL phonon emission spectra	65
2-7 Schematic of the QET layout	66
2-8 Schematic of the full phonon cycle	69
2-9 Phonon channel layout on the HV detector	70
2-10 Simulated phonon power in Si HV detector channels	72
2-11 Simulated trace in Si HV detector channels	73
2-12 Phonon collection asymmetry versus event depth	74
3-1 Resistance surface $R(T, I)$	80
3-2 Sensitivity parameters $\alpha(T, I)$ and $\beta(T, I)$	84
3-3 TES bias circuit	86
3-4 I_b – I_s characteristic	90
3-5 Derived IV quantities	91
3-6 QET thermal circuit	94
3-7 $I_b I_s$ family for HV Z2 at multiple bath temperatures	98

3-8	Joule power and resistance vs bias for HV Z2.....	99
3-9	Joule power vs bath temperature, fits with and without parasitic power.....	100
3-10	HV Z2 per-channel thermal parameter summary.....	101
3-11	Operating current and substrate temperature versus K_{s-b}	113
3-12	HVeV impulse response.....	114
3-13	Ge HV impulse response.....	115
3-14	Si HV impulse response.....	116
4-1	NFC mask and single QET layout	121
4-2	Detector payload and top-box housing	123
4-3	NFC-SO average pulse template per channel	126
4-4	HVeV LED optimum-filter amplitude spectra.....	126
4-5	Phonon power vs TESSubgapAbsorption.....	130
4-6	Simulated TES traces vs TESSubgapAbsorption.....	130
4-7	Shifted average phonon power.....	132
4-8	Reproducing the onset fo the template	133
4-9	MCMC walker chains.....	136
4-10	Corner plot of the nine-parameter pulse shape fit	137
4-11	Parameter correlation matrix from the nine-parameter fit.....	138
4-12	Nine-parameter best fit—MAP	138
4-13	Nine-parameter best fit—Median	139
4-14	Transition Curve from $I_b I_s$	141
4-15	Corner plot for a two-parameter fit.....	142
4-16	Final two-parameter best fit for both channels—MAP.....	143
4-17	Charge creation and combination sites	145
4-18	Charge carrier path lengths under -35 V bias	147
4-19	Maximum amplitude spectra at -35 V bias.....	149
4-20	Inner–outer asymmetry parameter	149
4-21	HVeV pulse template, one- and two-block fits	152

5-1	The SuperCDMS SNOLAB apparatus	156
5-2	A SuperCDMS detector tower	157
5-3	The CUTE facility	158
5-4	Current noise spectrum at five bias points	160
5-5	Measured current noise spectrum and the noise model with its components at four bias points.	162
5-6	Loop-gain bounds of the oscillating region	166
5-7	Electrothermal oscillation frequency versus operating parameters	167
5-8	Simulated electrothermal oscillations vs bias point in Z2.....	168
5-9	Event distribution in min-to-max amplitude	169
5-10	Pulse shapes rejected by the event selection	170
5-11	Raw traces of Tower 3 detectors	171
5-12	Current noise spectra of the Tower 3 channels	172
5-13	Arbitrarily scaled baseline resolution of the tower channels and detectors	174

LIST OF ABBREVIATIONS

ALP	Axion-like particle
BBN	Big Bang nucleosynthesis
BCS	Bardeen-Cooper-Schrieffer
CMB	Cosmic microwave background
CUTE	Cryogenic Underground Test facility
ETF	Electrothermal feedback
G4CMP	Geant4 Condensed Matter Physics
GL	Ginzburg-Landau
HV	High voltage
HVeV	High-voltage electron-volt resolution
IMF	Initial mass function
iZIP	Interleaved Z-sensitive Ionization and Phonon
Λ CDM	Lambda cold dark matter
LED	Light-emitting diode
MACHO	Massive astrophysical compact halo object
MAP	Maximum a posteriori
MCMC	Markov chain Monte Carlo
MOND	Modified Newtonian Dynamics
NTL	Neganov-Trofimov-Luke
ODE	Ordinary differential equation
QCD	Quantum chromodynamics
QET	Quasiparticle-trap-assisted electrothermal-feedback transition-edge sensor
SD	Spin-dependent
SI	Spin-independent
SQUID	Superconducting quantum interference device
SuperCDMS	Super Cryogenic Dark Matter Search
TES	Transition-edge sensor
WIMP	Weakly interacting massive particle

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

PULSE SHAPE MODELING AND CHARACTERIZATION OF THE SUPERCDMS SUB-GEV
DARK MATTER ATHERMAL PHONON MEDIATED DETECTORS

By

David Sadek

August 2026

Chair: Tarek Saab

Major: Physics

The Super Cryogenic Dark Matter Search (SuperCDMS) experiment aims to detect dark matter with masses below the GeV scale, using Si and Ge crystal targets read out by transition edge sensors biased within their superconducting transition. An interaction in the crystal creates electron-hole pairs and athermal phonons, and under a High Voltage (HV) bias the drifting charges release the work done by the bias field as additional athermal phonons through the Neganov-Trofimov-Luke effect (NTL), amplifying the signal from very low-energy depositions. The charges and phonons propagate through the crystal, modeled in G4CMP, delivering the deposited energy as a phonon flux into the readout sensor.

The response of the transition edge sensor is treated in the small-signal approximation, where the coupled electrical and thermal equations of a two-body thermal model are solved in the time domain to establish the fundamental decay times of the pulse and examine the regime of electrothermal oscillations. The same treatment enables the extraction of the electrical and thermal parameters of the readout circuit.

Combining the crystal transport and sensor models reproduces the pulse shapes from the LED calibration data of the latest generation of the HVeV detector operated at the Cryogenic Underground TESt facility (CUTE). The simulation parameters are tuned to the data using a Markov chain Monte Carlo (MCMC), and the resulting parameter degeneracies are examined. The long tail of the pulse is attributed to the thermal relaxation of the substrate.

The operation, stability, and energy resolution of the SuperCDMS SNOLAB high-voltage detectors are then evaluated, both at CUTE and during the commissioning of the experiment at SNOLAB. This includes a measurement of the Kapitza conductance between two of the detectors and the heat bath, together with a procedure developed, following similar approaches established for other previous detector generations, for their full characterization in future studies.

CHAPTER 1

THE DARK MATTER PROBLEM

The determination of mass in stellar systems presents a fundamental challenge in astrophysics. Unlike laboratory settings, where mass can be measured directly, astronomical masses must be inferred indirectly from observable quantities within a theoretical framework. Two conceptually independent methodologies have emerged for this inference. The first derives mass from electromagnetic observations combined with microphysical models of radiating matter. The second infers mass from gravitational phenomena using a theory of gravity to relate observables to the underlying mass distribution.

These two approaches need not, in principle, yield the same answer. Electromagnetic inference is sensitive only to matter that radiates or absorbs light, while gravitational inference responds to all sources of the gravitational field. A systematic discrepancy between the two would signal either the presence of non-luminous gravitating matter or a failure of the gravitational theory employed in the dynamical analysis. Such a discrepancy is observed across an enormous range of astrophysical scales, from individual galaxies to the large-scale structure of the Universe itself.

1.1 Baryonic Mass Inference

The mass in luminous matter can be inferred from electromagnetic observations by combining measured fluxes with models of the underlying emission processes. For stellar systems, which constitute the dominant baryonic component in most galaxies, the key quantity linking light to mass is the stellar mass-to-light ratio $\Upsilon_{\star}$. Because $\Upsilon_{\star}$ depends sensitively on how stellar mass is distributed among stars of different luminosities, it is fundamentally controlled by the stellar initial mass function (IMF), which specifies the relative number of stars formed as a function of their initial mass. The IMF therefore provides the statistical link between observed stellar light and total stellar mass in a system.

Consider a stellar system composed of N individual stars, where the i -th star has mass m_i and luminosity $\ell_{i,\lambda}$ in photometric band λ . The total stellar mass and luminosity are

$$M_\star \equiv \sum_{i=1}^N m_i, \quad (1-1)$$

$$L_\lambda \equiv \sum_{i=1}^N \ell_{i,\lambda}, \quad (1-2)$$

and the stellar mass-to-light ratio is defined as

$$\Upsilon_{\star,\lambda} \equiv \frac{M_\star}{L_\lambda}. \quad (1-3)$$

The value of $\Upsilon_{\star,\lambda}$ is not determined by this definition alone; it depends on how stellar luminosity relates to stellar mass across the population. One might attempt to define an individual stellar mass-to-light ratio $v_{i,\lambda} \equiv m_i/\ell_{i,\lambda}$ and regard the system-level ratio as some average of these individual values. However,

$$\Upsilon_{\star,\lambda} = \frac{\sum_i m_i}{\sum_i \ell_{i,\lambda}} \neq \frac{1}{N} \sum_i \frac{m_i}{\ell_{i,\lambda}}, \quad (1-4)$$

since a ratio of sums is not generally equal to an arithmetic mean of ratios. Rewriting the denominator as $\sum_i m_i/v_{i,\lambda}$ yields

$$\Upsilon_{\star,\lambda} = \frac{\sum_i m_i}{\sum_i m_i/v_{i,\lambda}}, \quad (1-5)$$

which shows that $\Upsilon_{\star,\lambda}$ is the harmonic mean of the individual ratios, weighted by mass. Unlike an arithmetic mean, the harmonic mean is pulled toward smaller values, foreshadowing a key result, that rare massive stars produce most of the luminosity while the abundant low-mass population contains most of the mass.

For large stellar populations, the discrete sum can be replaced by a continuum description. Let $n(m) dm$ denote the number of stars with masses in the interval $[m, m + dm]$. The total stellar

mass and luminosity become

$$M_{\star} = \sum_i m_i \rightarrow \int m n(m) dm, \quad (1-6)$$

$$L_{\lambda} = \sum_i \ell_{i,\lambda} \rightarrow \int \bar{\ell}_{\lambda}(m) n(m) dm, \quad (1-7)$$

where $\bar{\ell}_{\lambda}(m)$ is the mean luminosity in band λ for stars of mass m , averaged over the distribution of ages and metallicities present in the population. It is conventional to factor the number density as $n(m) = N_{\star}\phi(m)$, where $N_{\star}$ is the total number of stars and $\phi(m)$ is the stellar IMF. The IMF specifies the relative probability of forming a star of mass m , and is normalized such that $\int \phi(m) dm = 1$. Defining the expectation value of any function $f(m)$ with respect to the IMF as

$$\langle f \rangle_{\phi} \equiv \int f(m) \phi(m) dm, \quad (1-8)$$

the total mass and luminosity become $M_{\star} = N_{\star}\langle m \rangle_{\phi}$ and $L_{\lambda} = N_{\star}\langle \bar{\ell}_{\lambda} \rangle_{\phi}$. The stellar mass-to-light ratio is therefore

$$\Upsilon_{\star,\lambda} = \frac{\langle m \rangle_{\phi}}{\langle \bar{\ell}_{\lambda}(m) \rangle_{\phi}}. \quad (1-9)$$

This result shows that the system's mass-to-light ratio is the ratio of two population averages, not the average of individual ratios. Whether $\Upsilon_{\star,\lambda}$ is large or small depends on how differently mass and luminosity weight the stellar population.

The IMF was first measured by Salpeter [1] in the solar neighborhood and found to follow a power-law form over the mass range $0.4 \lesssim m/M_{\odot} \lesssim 10$. Subsequent measurements have established that this power law does not extend to the lowest masses but flattens below about half a solar mass. The general shape of the IMF is illustrated in Figure 1-1 ¹.

The key parameters characterizing the IMF are labeled in Figure 1-1. The lower mass limit $m_l \approx 0.08 M_{\odot}$ corresponds to the hydrogen-burning limit, below which the core of a contracting

¹ The text defines $\phi(m) = dN/dm \propto m^{\alpha}$, while Figure 1-1 plots $dN/d \log m \propto m^{\Gamma}$. Since $d \log m = dm/(m \ln 10)$, these are related by $\Gamma = \alpha + 1$. The Salpeter slope is $\alpha = -2.35$ or equivalently $\Gamma = -1.35$.

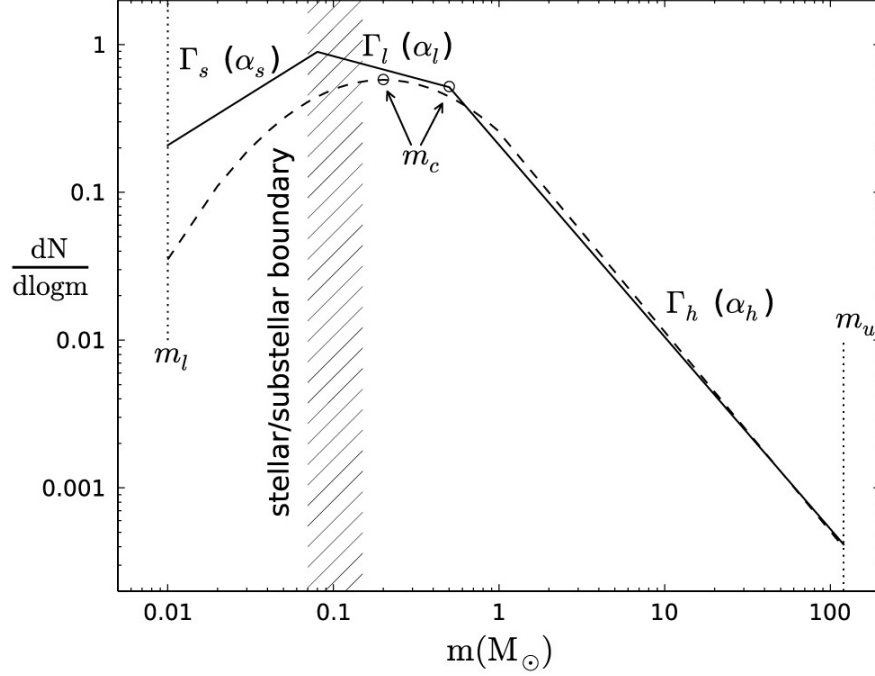


Figure 1-1. The stellar IMF as $dN/d \log m$ versus mass. Solid and dashed curves show the piecewise power-law and log-normal parameterizations, with m_l , m_u , m_c , and the slopes Γ marked. Adapted from Hopkins [2].

protostar never reaches the temperature required to sustain proton-proton fusion, and the object becomes a brown dwarf rather than a true star. The upper mass limit $m_u \approx 100\text{--}150 M_\odot$ arises from radiation pressure. The luminosity of massive stars scales steeply with mass ($L \propto M^3$ to M^4), and above $\sim 100 M_\odot$ the outward radiation pressure on infalling material exceeds the gravitational binding force, halting further accretion. The characteristic mass $m_c \approx 0.2\text{--}0.5 M_\odot$ marks the transition between the steep high-mass slope and the flatter low-mass behavior. Following the notation of Hopkins [2], the slopes in logarithmic mass bins are denoted Γ_s (substellar, $m < m_l$), Γ_l (low-mass, $m_l < m < m_c$), and Γ_h (high-mass, $m > m_c$), with corresponding linear-bin slopes $\alpha = \Gamma - 1$. The original Salpeter slope corresponds to $\Gamma_h = -1.35$ (or $\alpha_h = -2.35$).

Consider simplified power-law forms for both the IMF and the mass-luminosity relation in linear mass bins, $\phi(m) \propto m^\alpha$ and $\ell(m) \propto m^\beta$. Observations indicate $\alpha_h \approx -2.35$ for $m \gtrsim M_\odot$ and stellar structure theory gives $\beta \approx 3\text{--}4$ for main-sequence stars. The numerator of Eq. (1-9) then

scales as

$$\langle m \rangle_\phi \propto \int m^{1+\alpha} dm \propto \int m^{-1.35} dm, \quad (1-10)$$

which is dominated by the lower mass limit. Conversely, the denominator scales as

$$\langle \ell \rangle_\phi \propto \int m^{\beta+\alpha} dm \propto \int m^{0.65-1.65} dm, \quad (1-11)$$

which is dominated by the upper mass limit. This analysis reveals a fundamental asymmetry. The stellar mass budget is controlled by numerous low-mass stars, while the luminosity is dominated by rare high-mass stars. Consequently, characterizing the full IMF requires distinct observational techniques for its high-mass and low-mass ends.

This separation has profound implications for the time evolution of stellar populations. Because high-mass stars evolve rapidly off the main sequence, the luminosity of a stellar population fades with time as its most massive members exhaust their nuclear fuel and become compact remnants. The mass, by contrast, remains largely unchanged. Consequently, $Y_{\star,\lambda}$ increases monotonically with the age of the stellar population. For young, actively star-forming systems $Y_{\star,\lambda} \sim 0.1\text{--}1 M_\odot/L_\odot$ in optical bands, while for old stellar populations like those in elliptical galaxies $Y_{\star,\lambda} \sim 3\text{--}10 M_\odot/L_\odot$.

Stellar mass constitutes the dominant baryonic component in most galaxies, but interstellar gas contributes significantly in many systems and dominates in others. The total baryonic mass may be written as

$$M_{\text{bar}} = Y_{\star,\lambda} L_\lambda + M_{\text{gas}}, \quad (1-12)$$

where the stellar contribution enters through the mass-to-light ratio and M_{gas} encompasses all gas phases. The uncertainties in M_{bar} propagate primarily from the IMF, which remains perhaps the greatest source of systematic uncertainty in stellar mass estimates. These uncertainties are typically factors of ~ 2 , whereas the discrepancies with gravitationally inferred masses can exceed an order of magnitude.

1.2 Gravitational Evidence for the Mass Discrepancy

Mass announces itself gravitationally whether or not it shines. It bends the orbits of nearby objects, deflects the paths of light rays, and shapes the evolution of the universe as a whole. Three methods exploit these effects, independently of one another and of any assumption about what the mass is made of. Orbital dynamics, gravitational lensing, and the cosmic microwave background each return a total mass that exceeds the baryonic mass inferred from electromagnetic observations, and each does so at a different scale and a different epoch.

1.2.1 Mass from Gravitational Dynamics

The most direct route to gravitational mass is through Newton's laws. A test particle orbiting within a spherically symmetric mass distribution $M(< r)$ experiences a gravitational acceleration

$$g(r) = \frac{GM(< r)}{r^2}. \quad (1-13)$$

For circular motion, this acceleration provides the centripetal force, so

$$\frac{v_c^2}{r} = \frac{GM(< r)}{r^2}, \quad (1-14)$$

which can be inverted to give the enclosed mass,

$$M(< r) = \frac{v_c^2(r) r}{G}. \quad (1-15)$$

This equation is the foundation of rotation curve analysis. Measure the circular velocity $v_c(r)$ as a function of radius, and Eq. (1-15) returns the enclosed mass profile directly.

The critical prediction concerns the behavior at large radii. If the mass is concentrated within some characteristic radius $r_\star$, as the stellar light suggests, then for $r \gg r_\star$ the enclosed mass becomes approximately constant, $M(< r) \approx M_{\text{total}}$, and Eq. (1-15) implies

$$v_c(r) = \sqrt{\frac{GM_{\text{total}}}{r}} \propto r^{-1/2}. \quad (1-16)$$

This is the Keplerian expectation, that rotation velocities should decline as $r^{-1/2}$ beyond the luminous disk, just as planetary velocities decline with distance from the Sun.

Observations contradict this prediction. Beginning with the work of Rubin and Ford in the 1970s [3], rotation curves of spiral galaxies have been measured to large radii using optical emission lines in the inner disk and the 21-cm hyperfine transition of neutral hydrogen in the outer regions. The result, illustrated in Figure 1-2, is striking. Rotation curves remain approximately flat, $v_c(r) \approx \text{constant}$, far beyond the edge of the stellar disk.

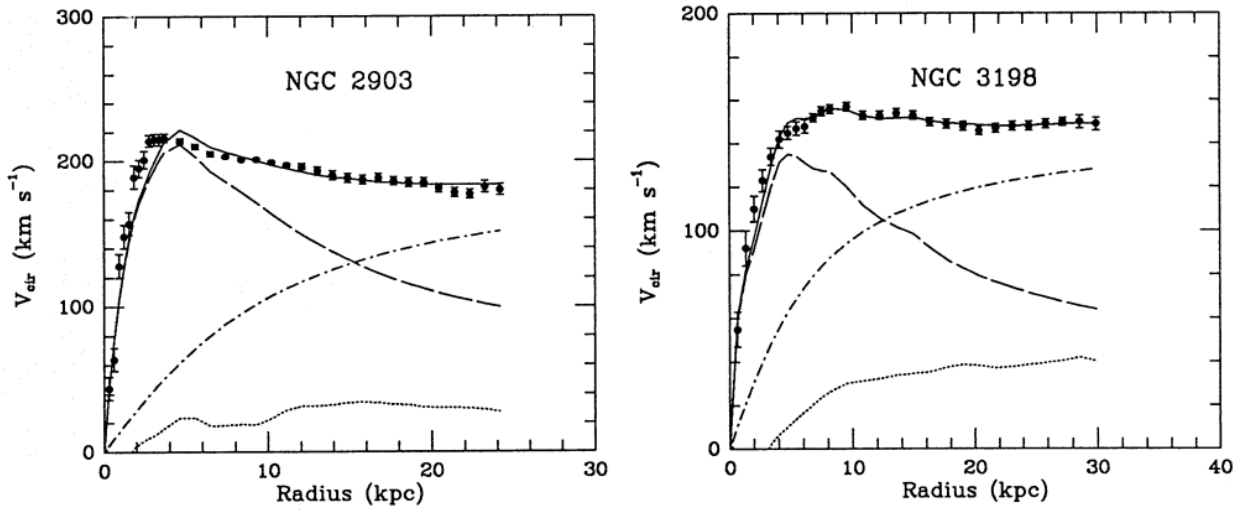


Figure 1-2. Rotation curves of NGC 2903 (left) and NGC 3198 (right). Observed velocities (points) with the stellar disk (dashed), gas (dotted), dark matter halo (dash-dotted), and total (solid). Adapted from Albada et al. [4].

Flat rotation curves have immediate implications for the mass distribution. From Eq. (1-15), constant v_c requires

$$M(< r) \propto r, \quad (1-17)$$

so the enclosed mass continues to grow linearly with radius even where the light has fallen off exponentially. Differentiating to obtain the density,

$$\rho(r) = \frac{1}{4\pi r^2} \frac{dM}{dr} \propto r^{-2}. \quad (1-18)$$

The mass distribution required by the kinematics, an extended halo with $\rho \propto r^{-2}$, bears no resemblance to the sharply falling stellar density profile.

To quantify the discrepancy, we can decompose the rotation curve into contributions from the known baryonic components. The stellar disk, with surface density $\Sigma_\star(R)$ inferred from the surface brightness and mass-to-light ratio, contributes a rotation velocity $v_\star(r)$. The gas disk, observed directly in 21-cm emission, contributes $v_{\text{gas}}(r)$. For a spherical mass distribution these contributions add in quadrature,

$$v_{\text{bar}}^2(r) = v_\star^2(r) + v_{\text{gas}}^2(r). \quad (1-19)$$

The observed velocity $v_{\text{obs}}(r)$ systematically exceeds $v_{\text{bar}}(r)$ at large radii, requiring an additional component,

$$v_{\text{obs}}^2(r) = v_{\text{bar}}^2(r) + v_{\text{DM}}^2(r). \quad (1-20)$$

Figure 1-2 shows this decomposition for two well-studied spiral galaxies. The dark matter contribution dominates at large radii, where it accounts for 80–90% of the enclosed mass.

For systems without ordered rotation, such as elliptical galaxies, dwarf spheroidals, and galaxy clusters, velocity dispersions replace rotation velocities as the kinematic observable. The virial theorem relates the total kinetic energy $K = \frac{1}{2}M\langle v^2 \rangle$ to the gravitational potential energy $W \sim -GM^2/R$ through

$$2K + W = 0, \quad (1-21)$$

yielding the virial mass estimator

$$M_{\text{vir}} \sim \frac{\sigma^2 R}{G}, \quad (1-22)$$

where σ is the velocity dispersion and R is a characteristic size. Zwicky first applied this method to the Coma cluster in 1933, finding a mass far exceeding that attributable to the visible galaxies [5]. Modern measurements confirm his result. Galaxy clusters require mass-to-light ratios $Y \sim 200\text{--}400 M_\odot/L_\odot$, compared to $Y_\star \sim 3\text{--}10 M_\odot/L_\odot$ for the stellar populations themselves.

1.2.2 Mass from Gravitational Lensing

Gravitational lensing measures mass without any assumption about dynamical equilibrium. In general relativity, mass curves spacetime and light follows geodesics through that curved geometry. A light ray passing a mass M at impact parameter b is deflected by an angle

$$\hat{\alpha} = \frac{4GM}{c^2 b}, \quad (1-23)$$

twice the value a Newtonian calculation would give, a discrepancy whose measurement was among the first confirmations of general relativity. Because the deflection responds to all the mass along the line of sight, it traces the total gravitating mass directly, regardless of whether that mass radiates.

When a foreground mass lies close to the line of sight to a distant source, the deflection produces dramatic distortions such as multiple images, luminous arcs, and Einstein rings, whose geometry encodes the enclosed mass. Far more common is the weak regime, in which the alignment is loose and each background galaxy is only slightly stretched tangentially around the lens. The stretching of any single galaxy is buried under its intrinsic ellipticity, but intrinsic ellipticities point in random directions while the lensing distortion is coherent. Averaging the shapes of thousands of background galaxies recovers the systematic alignment and, from it, the projected mass. Lensing masses measured this way consistently exceed the baryonic masses by the same factor found in the dynamical studies.

The most compelling lensing evidence comes from the Bullet Cluster (1E 0657–558), a pair of galaxy clusters whose cores passed through each other roughly 100 Myr ago. The collision sorted the matter by interaction cross-section. The galaxies, effectively collisionless, passed through unimpeded, while the hot intracluster gas, which dominates the baryonic mass by a factor of five to ten, felt ram pressure and lagged behind. The X-ray emission tracing that gas ends up spatially offset from the weak lensing mass reconstruction, as shown in Figure 1-3.

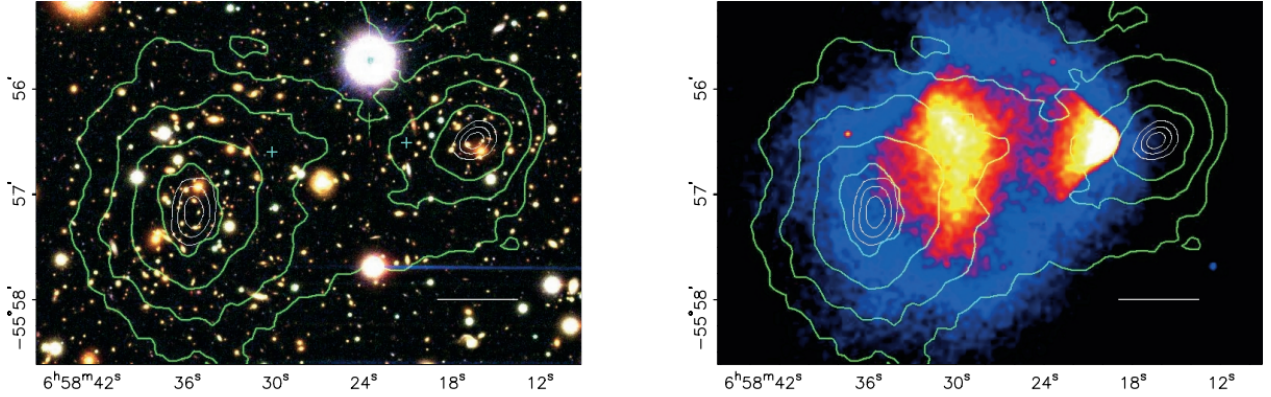


Figure 1-3. The Bullet Cluster (1E 0657–558). Optical image of galaxies (left) and a *Chandra* X-ray image of the intracluster plasma (right). Green contours show the weak lensing reconstruction; white contours mark the mass peaks. Adapted from Clowe et al. [6].

This separation is decisive. In any theory where gravity is sourced by baryonic matter alone, the gravitational potential must trace the baryons, and the dominant baryon component is the gas. Yet the lensing mass follows the galaxies, with the peak offset from the gas detected at the 8σ level [6]. The mass must therefore be both dark, since it emits no X-rays, and effectively collisionless, since it passed through the merger with the galaxies rather than lagging with the gas. The Bullet Cluster does not merely reveal a mass discrepancy. It spatially separates the missing mass from the baryons.

1.2.3 Mass from the Cosmic Microwave Background

The cosmic microwave background offers a completely independent determination of the matter content, probing an epoch long before galaxies existed. The CMB photons last scattered at redshift $z \approx 1100$, roughly 380,000 years after the Big Bang, when the primordial plasma cooled enough for electrons and protons to combine into neutral hydrogen. The temperature fluctuations imprinted on this radiation encode the physical conditions at that epoch.

Prior to recombination, baryonic matter was tightly coupled to photons through Thomson scattering. Gravity pulled this baryon-photon fluid toward overdense regions while radiation pressure pushed back against the compression. The competition drove acoustic oscillations, sound waves in the primordial plasma. At recombination the oscillations froze in, leaving a pattern of

hot and cold spots whose characteristic angular scale is set by the distance a sound wave could travel before that moment.

The two densities leave distinct fingerprints on the spectrum, which is what lets the data fix them separately. Baryons add inertia to the oscillating fluid. They deepen the compressions that gravity drives but barely affect the rarefactions that follow, so raising the baryon density lifts the odd-numbered acoustic peaks, set by compression, relative to the even-numbered ones, set by rarefaction. The ratio of peak heights is therefore the lever on Ω_b .

The total matter density acts through a different observable. Matter and radiation contributed equally to the energy density at a redshift fixed by Ω_m , and perturbations that entered the horizon while radiation still dominated were driven differently from those that entered later. This imprints the epoch of matter-radiation equality on the spectrum, most visibly in the height of the third peak and the overall amplitude of the first few peaks. Because this signature is largely separate from the odd-even pattern that tracks the baryons, the two densities can be extracted independently from a single spectrum.

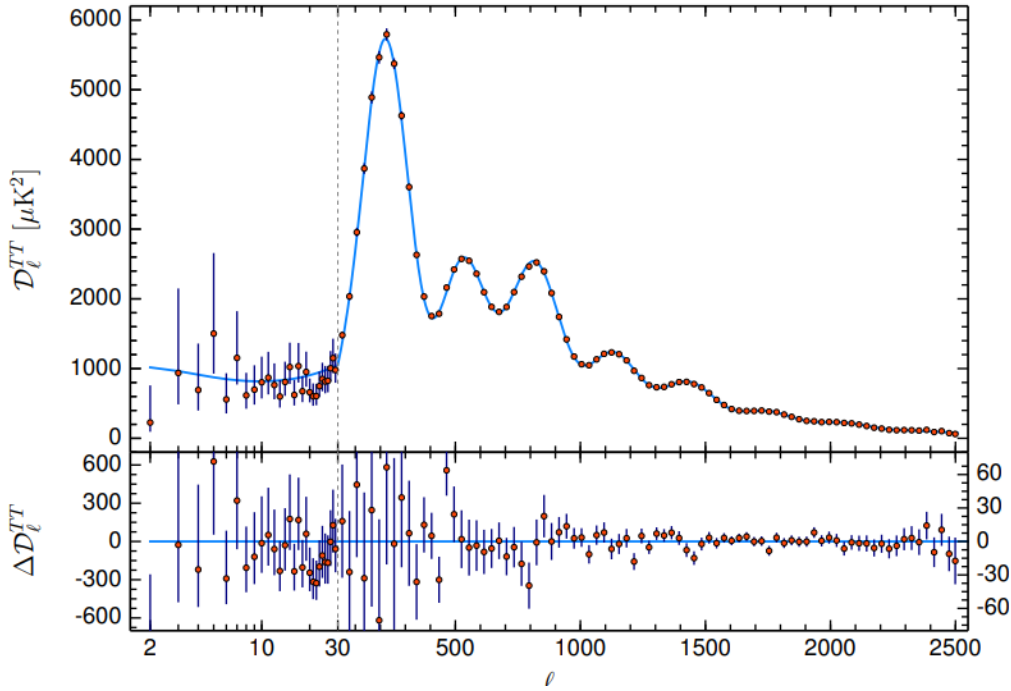


Figure 1-4. Angular power spectrum of CMB temperature anisotropies measured by *Planck*. From Planck Collaboration [7].

Fitting the six-parameter Λ CDM model to the observed power spectrum (Figure 1-4) pins down both densities,

$$\Omega_b h^2 = 0.0224 \pm 0.0001, \quad (1-24)$$

$$\Omega_m h^2 = 0.143 \pm 0.001, \quad (1-25)$$

whose ratio is

$$\frac{\Omega_b}{\Omega_m} = 0.157 \pm 0.002 \approx \frac{1}{6}. \quad (1-26)$$

Baryons make up only about a sixth of the total matter. The remaining five-sixths leaves no imprint on the odd-even peak pattern, so it cannot have been coupled to the photons, yet it is needed to source the potential wells that drove the oscillations in the first place. The CMB thus calls for matter that gravitates but is not baryonic.

This conclusion comes from the universe at $z \approx 1100$, long before any galaxy had formed, and it rests on linear perturbation theory in a plasma whose physics is well understood, free of the modeling uncertainties that attend nonlinear gravitational collapse. The factor of roughly six it gives between total and baryonic matter is the same factor seen in the rotation curves, the cluster velocity dispersions, and the lensing masses. Three measurements at three scales and three epochs converge on one number.

1.3 The Case for Particle Dark Matter

Two broad categories of explanation have been proposed for the mass discrepancy. The first holds that our theory of gravity is incomplete, that the apparent missing mass is an artifact of applying Newtonian gravity or general relativity in regimes where they fail. The second accepts the observations at face value and posits that the mass is real but composed of matter that does not emit or absorb light, what Zwicky called “dunkle Materie” when he first identified the discrepancy in galaxy clusters. The totality of the evidence favors the latter interpretation and further implies that the dark matter consists of a new species of elementary particle.

1.3.1 Modified Gravity

Modified gravity theories attempt to explain the mass discrepancy by altering the laws of gravity rather than introducing new matter. The most influential such proposal is Modified Newtonian Dynamics (MOND), introduced by Milgrom in 1983 [8].

In Newtonian mechanics, the gravitational acceleration of a test particle follows from equating the gravitational force $F = m_g g$, where $g = GM/r^2$, with Newton's second law $F = m_i a$. The equivalence principle asserts that gravitational and inertial mass are identical, $m_g = m_i$, so that all objects fall with the same acceleration,

$$a = g = \frac{GM}{r^2}. \quad (1-27)$$

For circular orbits, the centripetal acceleration $a = v^2/r$ must equal the gravitational field, yielding the Keplerian prediction $v \propto r^{-1/2}$ beyond the extent of the mass distribution.

MOND proposes that this relation breaks down at low accelerations. The modified equation of motion is

$$F = m a \mu\left(\frac{a}{a_0}\right), \quad (1-28)$$

where $a_0 \approx 1.2 \times 10^{-10} \text{ m/s}^2$ is a new fundamental constant and the interpolating function satisfies $\mu(x) \rightarrow 1$ for $x \gg 1$ while $\mu(x) \rightarrow x$ for $x \ll 1$. In the low-acceleration regime ($a \ll a_0$), equating gravitational and modified inertial forces gives

$$\frac{a^2}{a_0} = \frac{GM}{r^2}, \quad (1-29)$$

so that

$$a = \sqrt{\frac{GM a_0}{r^2}} = \frac{\sqrt{GM a_0}}{r}. \quad (1-30)$$

For circular orbits, $v^2/r = a$ then implies

$$v = (GM a_0)^{1/4}, \quad (1-31)$$

which is independent of radius, precisely the flat rotation curves observed. This relation also predicts that rotational velocity scales as the fourth root of the baryonic mass, recovering the empirical Tully-Fisher relation without invoking dark matter.

Despite this phenomenological success, MOND carries a significant theoretical cost. The modification in Equation 1-28 can be rewritten in terms of an effective inertial mass $m_i^{\text{eff}} = m \mu(a/a_0)$ that depends on acceleration. In the deep MOND regime, $m_i^{\text{eff}} \rightarrow 0$ as $a \rightarrow 0$, and objects offer vanishing resistance to acceleration at sufficiently low accelerations. The equivalence of gravitational and inertial mass, a foundational principle of general relativity confirmed experimentally to extraordinary precision, is thereby broken.

Relativistic extensions of MOND preserve $F = ma$ while modifying the gravitational field equations, but these theories encounter a different obstacle, nonlocality. General relativity is a local theory in which the curvature of spacetime at any point is determined entirely by the stress-energy at that point. Any modification that preserves locality and avoids dynamical instabilities cannot reproduce the phenomenology attributed to dark matter. To generate the gravitational effects of dark matter halos through modifications to gravity alone, the gravitational field at one location must depend on the field configuration arbitrarily far away, a radical departure from the structure of known physical theories.

Einstein constructed general relativity on two pillars, the equivalence principle and locality. Modified gravity theories must sacrifice one or the other. Dark matter preserves both. It is simply additional matter that sources gravity in the standard way, requiring no modification to the theoretical framework that has passed every other experimental test.

Beyond these theoretical considerations, modified gravity confronts direct empirical challenges. The Bullet Cluster exhibits a spatial offset between the dominant baryonic component (the X-ray-emitting gas) and the gravitational potential (mapped via weak lensing). In any theory where gravity is sourced exclusively by baryonic matter, the gravitational potential must trace the baryons. The observed separation, with the mass concentration lying with the nearly collisionless galaxies rather than the collisional gas that dominates the baryonic mass, cannot be

accommodated without a distinct, non-baryonic mass component that passed through the collision largely unimpeded.

Structure formation reinforces this conclusion. The amplitude of baryon-photon acoustic oscillations in the CMB constrains the gravitational potential wells at recombination. Without a pressureless component that decoupled from photons well before recombination, the potential wells would be too shallow and the observed large-scale structure of galaxies and clusters could not have formed by the present epoch. Modified gravity theories that attempt to match rotation curves systematically fail to reproduce the CMB power spectrum without introducing additional degrees of freedom that, upon closer examination, behave as dark matter in disguise.

1.3.2 Dark Matter as a New Particle

If the laws of gravity are correct, the mass discrepancy demands the existence of non-luminous matter. The question then becomes what it is made of. The simplest possibility would be ordinary baryonic matter in forms that evade detection, such as cold molecular gas, brown dwarfs, stellar remnants, or black holes distributed throughout galactic halos. Two independent lines of evidence exclude this hypothesis.

Big Bang nucleosynthesis provides an absolute ceiling on the baryon density of the universe. In the first few minutes after the Big Bang, nuclear reactions assembled protons and neutrons into light nuclei (deuterium, helium-3, helium-4, and lithium-7) in proportions that depend sensitively on the baryon-to-photon ratio [9]. Deuterium is particularly constraining because it is only destroyed, never created, in stellar interiors; its primordial abundance thus provides a direct measure of the baryon density at nucleosynthesis. Observations of deuterium in pristine, high-redshift gas clouds yield

$$\Omega_b h^2 = 0.0224 \pm 0.0009 \quad (\text{BBN}), \quad (1-32)$$

in precise agreement with the independent determination from CMB anisotropies. Since the total matter density inferred from gravitational dynamics and cosmological observations is

$\Omega_m h^2 \approx 0.14$, baryons constitute only about one-sixth of the matter in the universe. The remaining five-sixths cannot be baryonic regardless of its luminosity, visibility, or spatial distribution.

Microlensing surveys independently constrain compact baryonic objects in galactic halos. If the Milky Way's dark halo consisted of massive compact objects (brown dwarfs, white dwarfs, neutron stars, or stellar-mass black holes), these would occasionally pass near the line of sight to background stars in the Magellanic Clouds, producing temporary gravitational magnification. The EROS and MACHO collaborations monitored tens of millions of stars over more than a decade, searching for such events [10]. The observed rate of microlensing events falls far below what would be expected if compact objects constituted a significant fraction of the halo mass, excluding MACHOs over the mass range $10^{-7} \lesssim M/M_\odot \lesssim 10$ as the primary dark matter component. Primordial black holes outside this range face complementary constraints from gravitational wave observations, stellar dynamics in ultra-faint dwarf galaxies, and accretion signatures.

The dark matter must therefore be non-baryonic, not composed of protons, neutrons, or any bound states thereof. Combined with the requirements established by gravitational observations, a coherent picture emerges. Dark matter must couple to gravity, or it could not produce the observed dynamical and lensing effects. It must be electrically neutral, or nearly so, since it neither emits nor absorbs light at detectable levels. The Bullet Cluster demonstrates that dark matter self-interactions are weak, since the dark matter halos of the merging clusters passed through one another without significant deceleration, implying that dark matter is effectively collisionless. And it must be stable on cosmological timescales, having persisted from the early universe to the present day.

The Standard Model of particle physics contains no particle satisfying these requirements. Neutrinos are the sole Standard Model particles that are electrically neutral, stable, and weakly interacting. However, neutrino masses are constrained by oscillation experiments to satisfy $\sum m_\nu \lesssim 0.1$ eV, far too light to constitute the dark matter. At such low masses, neutrinos remained relativistic until late cosmological times, and their large thermal velocities would have erased density perturbations through free-streaming on scales corresponding to galaxy clusters and

below. The observed abundance of dwarf galaxies and the small-scale structure encoded in the Lyman- α forest exclude any dark matter candidate that was relativistic at matter-radiation equality. Furthermore, the cosmological neutrino abundance, fixed by their thermal decoupling in the early universe, falls short of the required dark matter density by more than an order of magnitude.

The existence of dark matter thus constitutes direct evidence for physics beyond the Standard Model. The universe contains a particle species not yet produced or identified in terrestrial experiments, one that shaped the formation of cosmic structure and continues to dominate the gravitational dynamics of galaxies and clusters today. Identifying the nature of this particle, constraining its properties, and ultimately detecting it directly have become central objectives of both particle physics and cosmology.

1.4 Dark Matter Cosmology and Particle Physics

Dark matter must be gravitationally coupled, electrically neutral, stable over cosmological timescales, cold (non-relativistic at matter-radiation equality), and weakly self-interacting. These requirements follow directly from observation. Gravitational coupling produces the dynamical and lensing signatures; electrical neutrality explains the absence of electromagnetic emission; stability ensures persistence from the early universe to present; coldness permits the formation of small-scale structure; and weak self-interaction is demanded by the spatial offset in merging clusters. Beyond these constraints, the particle nature of dark matter remains remarkably open, with viable candidates spanning some 70 orders of magnitude in mass and coupling to ordinary matter through a variety of interaction channels.

1.4.1 Properties and Distribution

For direct detection experiments, the critical quantities are the local dark matter density and velocity distribution. The local density ρ_0 at the Sun's position is inferred from the Galactic rotation curve, stellar kinematics, and dynamical modeling [11]. The community-standard value adopted for direct detection is [12]

$$\rho_0 = 0.3 \text{ GeV/cm}^3, \tag{1-33}$$

with systematic uncertainties at the 20–30% level. For a dark matter particle of mass m_χ , the local number density is $n_\chi = \rho_0/m_\chi$, so a 100 GeV particle has $n_\chi \sim 3 \times 10^{-3} \text{ cm}^{-3}$, while a 1 GeV particle has $n_\chi \sim 0.3 \text{ cm}^{-3}$.

The velocity distribution follows the Standard Halo Model (SHM), a Maxwell-Boltzmann distribution in the Galactic rest frame,

$$f(\vec{v}) \propto \exp\left(-\frac{|\vec{v}|^2}{v_0^2}\right) \Theta(v_{\text{esc}} - |\vec{v}|), \quad (1-34)$$

truncated at the Galactic escape velocity $v_{\text{esc}} = 544 \text{ km/s}$, with characteristic speed $v_0 = 238 \text{ km/s}$ set by the local circular velocity [12, 13]. The Sun’s motion through the halo at $\sim 250 \text{ km/s}$ produces a “dark matter wind” from the direction of Cygnus, and Earth’s orbital motion adds a small ($\sim 2\text{--}6\%$) annual modulation to the detection rate.

The cosmological abundance is precisely measured. Planck observations of CMB anisotropies yield [7]

$$\Omega_\chi h^2 = 0.120 \pm 0.001, \quad (1-35)$$

corresponding to 26% of the total energy density. How dark matter acquired this abundance depends on its early-universe interactions, and different production mechanisms point toward different mass scales and couplings.

Thermal freeze-out provides the classic framework. Dark matter begins in thermal equilibrium with the Standard Model plasma, annihilating and being produced at equal rates. As the universe expands and cools, the annihilation rate drops below the Hubble rate, and the comoving abundance freezes in. The resulting density scales inversely with the annihilation cross-section, and remarkably, weak-scale cross-sections ($\langle\sigma v\rangle \sim 10^{-26} \text{ cm}^3/\text{s}$) yield the observed abundance for masses in the GeV–TeV range, the so-called “WIMP miracle” that motivated decades of direct detection searches targeting this mass window. Asymmetric dark matter offers a different perspective. Since $\Omega_\chi \approx 5\Omega_b$, perhaps both abundances originate from related

asymmetries in the early universe. If the number densities of dark matter and baryons are comparable, then $m_\chi \sim 5m_p \sim 5 \text{ GeV}$, pointing toward the light WIMP regime.

Axions and axion-like particles arise from a distinct mechanism. The QCD axion emerges from the Peccei-Quinn solution to the strong CP problem, with its mass and couplings determined by the symmetry-breaking scale f_a . In the early universe, the axion field begins displaced from its potential minimum; when the Hubble rate drops below the axion mass, the field oscillates coherently, and these oscillations behave as cold dark matter. This misalignment mechanism naturally produces the observed abundance for axion masses in the μeV – meV range. Axion-like particles extend this framework to broader mass and coupling ranges. Freeze-in production offers yet another possibility, dark matter that never thermalizes but is gradually produced through feeble interactions with the Standard Model bath. This permits extremely small couplings across a wide mass range. Each production history populates different regions of the mass–coupling parameter space, and no single mechanism is uniquely favored by current data.

1.4.2 Elastic Scattering Kinematics

Direct detection relies on observing the recoil of a target particle, a nucleus or electron, struck by an incoming dark matter particle. Consider a dark matter particle of mass m_χ with velocity $\vec{v}$ in the laboratory frame, incident on a target of mass m_T initially at rest. For elastic scattering, kinetic energy is conserved. In the center-of-mass frame, where the total momentum vanishes, this implies that the particles emerge with the same speeds but rotated through some scattering angle θ_{CM} . Transforming back to the laboratory frame, the target recoil energy is

$$E_R = \frac{\mu^2 v^2}{m_T} (1 - \cos \theta_{\text{CM}}), \quad (1-36)$$

The scattering angle ranges from 0 (forward, no energy transfer) to π (backscatter, maximum transfer), bounding the recoil energy,

$$E_R^{\text{max}} = \frac{2\mu^2 v^2}{m_T} = \frac{4m_\chi m_T}{(m_\chi + m_T)^2} E_\chi, \quad (1-37)$$

where $E_\chi = \frac{1}{2}m_\chi v^2$ is the incident kinetic energy. The energy transfer fraction $4m_\chi m_T / (m_\chi + m_T)^2$ is maximized when the masses are matched and falls off for mismatched masses, as shown in Figure 1-5. A 10 GeV particle scattering off silicon ($m_N \approx 26$ GeV) transfers up to 80% of its kinetic energy, while xenon ($m_N \approx 122$ GeV) captures only 28%. For lighter dark matter the mismatch worsens, with a 1 GeV particle on silicon transferring at most 14%. This kinematic suppression is the fundamental obstacle to detecting sub-GeV dark matter with heavy nuclear targets. Lighter targets like helium offer significantly better kinematic matching in this regime.

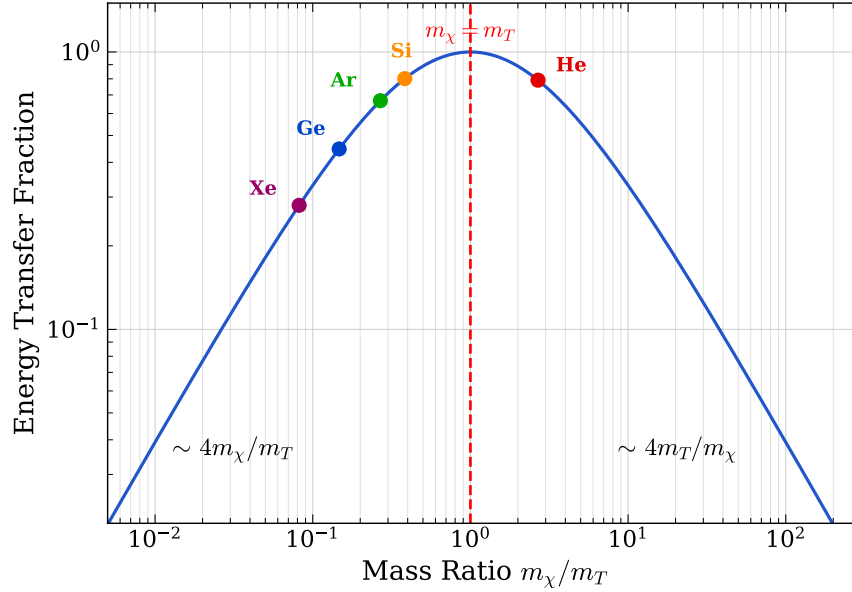


Figure 1-5. Energy transfer fraction versus mass ratio for elastic scattering. Colored points mark different nuclear targets for $m_\chi = 10$ GeV.

Inverting Eq. (1-36), a recoil of energy E_R requires a minimum incident velocity

$$v_{\min}(E_R) = \sqrt{\frac{m_T E_R}{2\mu^2}}. \quad (1-38)$$

For light dark matter ($m_\chi \ll m_T$), $v_{\min} \propto 1/m_\chi$, so lighter particles must move faster to produce the same recoil. Since the halo velocity distribution cuts off at $v_{\text{esc}} \approx 544$ km/s, there exists a minimum detectable dark matter mass for any given energy threshold. Conversely, for a given

dark matter mass, lighter targets yield larger recoils and thus require lower $v_{\min}$, accessing more of the velocity distribution.

These kinematic relations connect to experimental observables through the differential interaction rate. A detector of mass M_{det} contains $N_T = M_{\text{det}}/m_T$ target particles, immersed in a dark matter number density $n_\chi = \rho_0/m_\chi$. Dark matter particles with velocity $\vec{v}$ contribute a differential flux $n_\chi v f(\vec{v}) d^3v$, where $f(\vec{v})$ is the normalized velocity distribution. Multiplying by the differential cross-section $d\sigma/dE_R$ and integrating over all velocities capable of producing a recoil E_R ,

$$\frac{dR}{dE_R} = N_T \frac{\rho_0}{m_\chi} \int_{v>v_{\min}} v f(\vec{v}) \frac{d\sigma}{dE_R} d^3v. \quad (1-39)$$

For a contact interaction, valid when the mediator mass far exceeds the momentum transfer, the scattering is isotropic in the center-of-mass frame. Since $E_R \propto (1 - \cos \theta_{\text{CM}})$, the recoil energy is uniformly distributed between 0 and $E_R^{\max}$, giving a differential cross-section

$$\frac{d\sigma}{dE_R} = \frac{\sigma_0}{E_R^{\max}} = \frac{m_T \sigma_0}{2\mu^2 v^2}, \quad (1-40)$$

where σ_0 is the total cross-section at zero momentum transfer. Substituting and defining the velocity integral

$$g(v_{\min}) \equiv \int_{v>v_{\min}} \frac{f(\vec{v})}{v} d^3v, \quad (1-41)$$

the differential rate becomes

$$\frac{dR}{dE_R} = \frac{N_T \rho_0 \sigma_0 m_T}{2m_\chi \mu^2} g(v_{\min}). \quad (1-42)$$

The rate scales inversely with dark matter mass, since lighter particles are more numerous at fixed mass density, and linearly with the cross-section σ_0 . The velocity integral $g(v_{\min})$ decreases as $v_{\min}$ increases, so the recoil spectrum falls with increasing E_R . When $v_{\min}$ exceeds v_{esc} , the spectrum terminates.

For nuclear targets, finite nuclear size introduces an additional suppression. When the momentum transfer $q = \sqrt{2m_T E_R}$ becomes comparable to the inverse nuclear radius, the dark

matter no longer scatters coherently off the entire nucleus. This is encoded in a form factor $F(q)$,

$$\frac{dR}{dE_R} = \frac{N_T \rho_0 \sigma_0 m_T}{2m_\chi \mu^2} F^2(E_R) g(v_{\min}). \quad (1-43)$$

The form factor and total cross-section σ_0 depend on the specific interaction model.

Figure 1-6 shows how the differential recoil spectrum depends on dark matter mass for a silicon target. Lighter dark matter produces softer spectra concentrated at low energies, while heavier dark matter extends to higher recoils. The spectrum falls approximately exponentially with recoil energy, reflecting the velocity distribution cutoff at v_{esc} , since higher recoils require higher velocities, which become increasingly rare. For sub-GeV masses, the spectrum is compressed to very low energies, demanding sub-keV thresholds for detection.

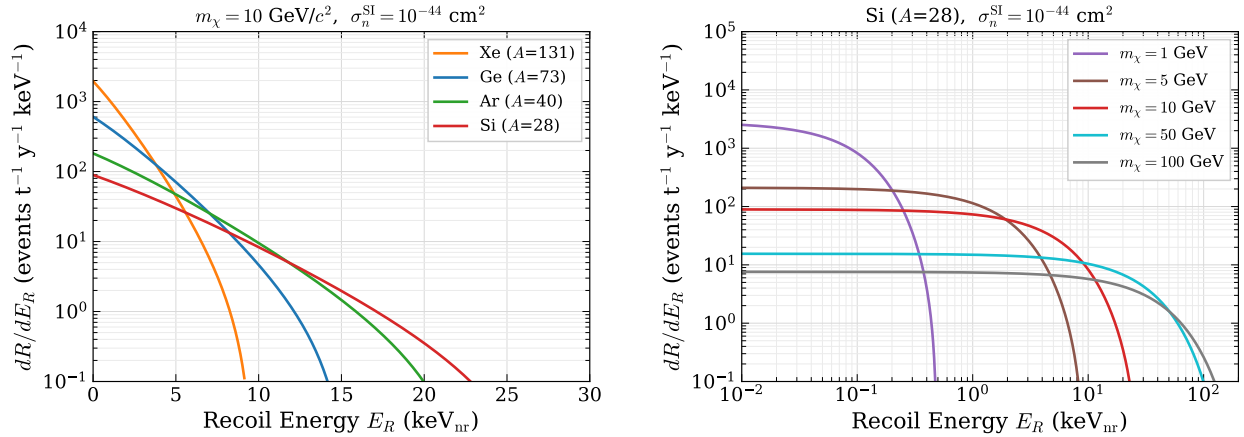


Figure 1-6. Spin-independent differential recoil spectra for $\sigma_n^{\text{SI}} = 10^{-44} \text{ cm}^2$, $m_\chi = 10 \text{ GeV}$ on xenon, germanium, argon, and silicon (left); silicon across dark matter masses from 1 to 100 GeV (right).

1.4.3 Interaction Models

The cross-section σ_0 in Eq. (1-43) encodes the particle physics of the dark matter interaction. Different coupling structures lead to distinct spectral shapes, optimal target choices, and mass ranges. Comprehensive reviews of the direct detection landscape can be found in Ref. [14].

1.4.3.1 Nuclear recoil models

Spin-independent interactions. Spin-independent (SI) interactions arise when dark matter couples to nucleons through scalar or vector mediators. When the de Broglie wavelength of the exchanged momentum is much larger than the nucleus, the dark matter scatters coherently off all nucleons. The amplitudes add constructively, and for equal couplings to protons and neutrons, the cross-section on a nucleus of mass number A relates to the dark matter–nucleon cross-section by

$$\sigma_0 = \sigma_{\chi N}^{\text{SI}} = \frac{\mu_{\chi N}^2}{\mu_{\chi n}^2} A^2 \sigma_{\chi n}^{\text{SI}}, \quad (1-44)$$

where $\mu_{\chi N}$ and $\mu_{\chi n}$ are the reduced masses for the dark matter–nucleus and dark matter–nucleon systems respectively. This A^2 enhancement strongly favors heavy targets. Silicon ($A = 28$) provides an enhancement of ~ 780 relative to hydrogen; germanium ($A = 73$) provides ~ 5300 .

As the momentum transfer increases and the dark matter wavelength becomes comparable to the nuclear radius, the scattering resolves the spatial distribution of nucleons and coherence is partially lost. This suppression is encoded in the nuclear form factor $F(q)$, the Fourier transform of the nucleon density. The Helm parameterization [15, 16] provides the standard form,

$$F(q) = \frac{3j_1(qR_1)}{qR_1} e^{-(qs)^2/2}, \quad (1-45)$$

where j_1 is the spherical Bessel function, $s \approx 0.9$ fm is the nuclear skin thickness, and the effective nuclear radius R_1 follows the parameterization of Lewin and Smith [16].

The SI differential rate, combining Eqs. (1-43) and (1-44), is

$$\left. \frac{dR}{dE_R} \right|_{\text{SI}} = \frac{N_T \rho_0 A^2 \sigma_{\chi n}^{\text{SI}} m_N}{2m_\chi \mu_{\chi n}^2} F^2(E_R) g(v_{\min}). \quad (1-46)$$

This is the baseline spectral shape shown in Figure 1-6.

Spin-dependent interactions. Spin-dependent (SD) interactions arise when dark matter couples to nucleon spin through axial-vector mediators. Unlike SI scattering, only unpaired

nucleons contribute, since paired nucleons have opposite spins that cancel. The cross-section depends on the nuclear spin J and the expectation values of proton and neutron spin content,

$$\sigma_{\chi N}^{\text{SD}} = \frac{32\mu_{\chi N}^2 G_F^2}{\pi} \frac{J+1}{J} [a_p \langle S_p \rangle + a_n \langle S_n \rangle]^2, \quad (1-47)$$

where $a_{p,n}$ are the effective couplings to protons and neutrons, and $\langle S_{p,n} \rangle$ are the expectation values of the proton and neutron spin operators in the nucleus. Without coherent A^2 enhancement, SD cross-sections are typically subdominant to SI for heavy nuclei unless the SI coupling is suppressed by the underlying particle physics. Nuclei with odd mass numbers, such as ^{29}Si , ^{73}Ge , ^{129}Xe , and ^{131}Xe , are sensitive to SD interactions.

Light mediator interactions. The SI and SD cross-sections above assume a contact interaction, valid when the mediator mass m_ϕ far exceeds the momentum transfer $q \sim \mu v \sim \text{few} \times 10 \text{ MeV}$ for typical halo velocities. When m_ϕ becomes comparable to or smaller than q , the mediator propagator introduces additional momentum dependence. The matrix element acquires a factor $(q^2 + m_\phi^2)^{-1}$, modifying the differential cross-section,

$$\frac{d\sigma}{dE_R} \propto \frac{1}{(2m_N E_R + m_\phi^2)^2}. \quad (1-48)$$

In the massless mediator limit ($m_\phi \rightarrow 0$), this gives $d\sigma/dE_R \propto 1/E_R^2$, dramatically enhancing the rate at low recoil energies compared to the contact interaction. The differential rate becomes

$$\left. \frac{dR}{dE_R} \right|_{\text{light}} = \left. \frac{dR}{dE_R} \right|_{\text{SI}} \times \left(\frac{q_{\text{ref}}^2}{2m_N E_R + m_\phi^2} \right)^2, \quad (1-49)$$

where q_{ref} is a reference momentum at which the cross-section is normalized. For a given total event rate, the light mediator spectrum is redistributed relative to the contact case, enhanced at low recoil energies where $q \lesssim m_\phi$, crossing over near $q \sim m_\phi$, and suppressed at high recoil energies where $q \gg m_\phi$. The net effect is a steeper spectral shape that concentrates events toward threshold, favoring detectors with lower energy thresholds. Light mediators are theoretically

motivated for sub-GeV dark matter, where they can enable thermal freeze-out production while evading constraints that would exclude direct weak-scale couplings to the Standard Model.

The Migdal effect. The Migdal effect provides a mechanism to extend nuclear recoil sensitivity to lighter dark matter masses by exploiting the electronic response to sudden nuclear motion [17, 18]. When a nucleus recoils, its electron cloud does not instantaneously follow; the sudden perturbation can excite or ionize atomic electrons. The probability of ionization accompanying a nuclear recoil scales as

$$\frac{dP}{dE_e} \propto \frac{1}{E_R} \left| \langle f | e^{i\vec{q}_e \cdot \vec{r}} | i \rangle \right|^2, \quad (1-50)$$

where $\vec{q}_e = m_e \vec{v}_N$ is the momentum imparted to the electrons by the recoiling nucleus and the matrix element connects initial bound and final continuum electronic states.

The total detected energy is $E_{\text{det}} = E_R + E_e + E_b$, where E_e is the kinetic energy of the ejected electron and E_b is its binding energy. For light dark matter, the ionization energy E_e can substantially exceed the nuclear recoil E_R , bringing the total signal above threshold even when the nuclear recoil alone would be undetectable. The differential rate including the Migdal effect is

$$\frac{dR}{dE_{\text{det}}} = \int dE_R \int dE_e \frac{dR_{\text{NR}}}{dE_R} \frac{dP}{dE_e} \delta(E_{\text{det}} - E_R - E_e - E_b). \quad (1-51)$$

The Migdal effect extends nuclear recoil searches from ~ 1 GeV down to ~ 10 – 100 MeV, bridging the gap between traditional nuclear recoil and electron recoil regimes. The ionization signal also benefits from typically better energy resolution and lower thresholds than nuclear recoil detection in many technologies.

1.4.3.2 Electron recoil models

As the transfer fraction (Figure 1-5) illustrates, matching between projectile and target masses maximizes energy transfer. For dark matter lighter than ~ 1 GeV scattering off nuclei with $A \gtrsim 28$, the transfer fraction falls below $\sim 10\%$, and the resulting recoil energies drop below practical detector thresholds. Scattering off atomic electrons provides an alternative detection

channel, where the light electron mass ($m_e = 511$ keV) offers favorable kinematics for sub-GeV dark matter.

Dark matter–electron scattering. The kinematics of electron scattering differ from nuclear recoils in an important way. The electron is bound in an atom with binding energy E_b . In nuclear recoils, the nucleus is both the scatterer and the recoiling body, locking momentum transfer and recoil energy together through $q = \sqrt{2m_N E_R}$. For a bound electron, the atom as a whole absorbs recoil momentum, so the momentum transfer q and the deposited energy become independent variables. The scattering must supply at least E_b to liberate the electron, and the total deposited energy is

$$E_{\text{dep}} = E_e + E_b, \quad (1-52)$$

where E_e is the kinetic energy of the outgoing electron.

The minimum velocity for electron scattering reflects this independence,

$$v_{\min} = \frac{E_{\text{dep}}}{q} + \frac{q}{2m_\chi}, \quad (1-53)$$

where q is the momentum transfer. Unlike nuclear recoils where $v_{\min}$ depends only on recoil energy, here $v_{\min}$ depends on both deposited energy and momentum transfer, requiring integration over allowed q values. For a given deposited energy, the optimal momentum transfer that minimizes $v_{\min}$ is $q_{\text{opt}} = \sqrt{2m_\chi E_{\text{dep}}}$, giving $v_{\min}^{\text{opt}} = \sqrt{2E_{\text{dep}}/m_\chi}$.

The differential rate for electron scattering involves the ionization form factor $|f_{\text{ion}}(k', q)|^2$, which encodes the probability of transitioning from a bound state to a continuum state with final electron momentum $k' = \sqrt{2m_e E_e}$ given momentum transfer q [19],

$$\frac{dR}{d \ln E_{\text{dep}}} = N_T \frac{\rho_0}{m_\chi} \frac{\bar{\sigma}_e}{8\mu_{\chi e}^2} \int d \ln q \, q^2 |F_{\text{DM}}(q)|^2 |f_{\text{ion}}(k', q)|^2 \eta(v_{\min}), \quad (1-54)$$

where $\bar{\sigma}_e$ is a reference cross-section defined at momentum transfer $q_0 = \alpha m_e$, $F_{\text{DM}}(q)$ encodes the momentum dependence of the dark matter interaction ($F_{\text{DM}} = 1$ for a heavy mediator,

$F_{\text{DM}} = (\alpha m_e / q)^2$ for an ultralight mediator), and $\eta(v_{\text{min}}) = \int_{v > v_{\text{min}}} f(\vec{v}) / v \, d^3v$ is the velocity integral. The additional integration over q , absent in the nuclear recoil rate, is a direct consequence of the decoupled kinematics.

The ionization form factor $|f_{\text{ion}}|^2$ must be computed from atomic or crystal wavefunctions and depends on the target material. Physically, it measures the overlap between the bound-state momentum distribution and the momentum kick $\vec{q}$ delivered by the dark matter. The form factor peaks when q matches the characteristic momentum scale of the bound electron, $q \sim \alpha m_e Z_{\text{eff}}$, and falls off at higher momentum transfers where the bound-state wavefunction lacks high-momentum components. For crystalline targets such as silicon, the form factor must be computed numerically using density functional theory and band-structure methods [20]. The optimal mass range for electron scattering is $m_\chi \sim 1 \text{ MeV}$ to $\sim 1 \text{ GeV}$, where kinematic matching to electrons is favorable and semiconductor targets with eV-scale band gaps provide the necessary low thresholds.

Dark photon absorption. Bosonic dark matter candidates, such as dark photons, axion-like particles, and scalar dark matter, can be absorbed by atoms rather than scattered. In absorption, the dark matter particle is destroyed and its entire rest mass energy is deposited,

$$E_{\text{dep}} = m_{A'} - E_b, \quad (1-55)$$

where $m_{A'}$ is the dark photon mass and E_b is the binding energy of the ejected electron. This produces a monoenergetic signal, in contrast to the continuous spectrum from scattering.

The dark photon couples to the electromagnetic current through kinetic mixing with the ordinary photon, parameterized by ϵ . The absorption cross-section is related to the photoelectric cross-section,

$$\sigma_{\text{abs}} = \epsilon^2 \sigma_{\text{PE}}(m_{A'}), \quad (1-56)$$

where $\sigma_{\text{PE}}(\omega)$ is the photoelectric cross-section at photon energy $\omega = m_{A'}$. The absorption rate is

$$R = N_T \frac{\rho_0}{m_{A'}} \epsilon^2 \sigma_{\text{PE}}(m_{A'}) \cdot v, \quad (1-57)$$

where the factor $\rho_0/m_{A'}$ gives the dark photon number density and $v \approx 10^{-3}c$ is the characteristic velocity.

The monoenergetic signature provides powerful background discrimination. A signal appears as a sharp peak at $E_{\text{dep}} = m_{A'}$ rather than a continuous spectrum, set apart from the falling backgrounds that dominate at low energy. The sensitivity decreases at high masses because the number density $\rho_0/m_{A'}$ falls, and at low masses because the photoelectric cross-section decreases. The optimal mass range is $m_{A'} \sim 1$ eV to ~ 100 eV for semiconductor targets, extending to keV scales for noble liquid detectors.

Axion-like particles (ALPs) can similarly be absorbed through their coupling to electrons, $\mathcal{L} \supset g_{ae} a \bar{e} \gamma^5 e$, producing the axioelectric effect analogous to the photoelectric effect. The absorption rate scales as g_{ae}^2 rather than ϵ^2 , but the phenomenology is similar, with monoenergetic deposition at the ALP mass.

No single detector technology covers the full parameter space. The 70 orders of magnitude in viable dark matter mass, combined with the diversity of possible interaction types, necessitates a broad experimental program employing complementary detection strategies.

1.5 The Experimental Landscape

The energy a dark matter particle deposits in a detector emerges through three measurable channels (ionization, scintillation, and phonons). A detector built to read two of them at once gains event-by-event background discrimination, since electron and nuclear recoils divide their energy between the channels in different proportions. A detector built around a single channel gives up that discrimination but reaches lower in energy, down to the eV scale and to single quanta of charge. This choice of channel, set against the energy threshold and target mass a technology can practically achieve, is what places each experiment within the broad parameter space of viable dark matter.

1.5.1 Nuclear Recoil Searches

A dark matter particle that scatters elastically from an atomic nucleus deposits energy as a low-energy nuclear recoil. For weak-scale masses the coherent enhancement of the cross section

with nuclear mass favors heavy targets, and liquid xenon time projection chambers exploit this with multi-tonne active masses, reading scintillation and ionization to reject electron-recoil backgrounds. Experiments such as LZ and XENONnT set the strongest spin-independent limits above roughly $10 \text{ GeV}/c^2$. Below a few GeV/c^2 the recoil energy imparted to a heavy nucleus falls below these thresholds, and the search passes to cryogenic calorimeters. CRESST instruments CaWO_4 crystals, DarkSide-50 reads ionization in liquid argon, and SuperCDMS operates germanium and silicon detectors that read athermal phonons to reach nuclear-recoil energies inaccessible to the noble liquids. Figure 1-7 places these constraints in context, together with the projected reach of SuperCDMS SNOLAB across the low-mass range targeted by this work.

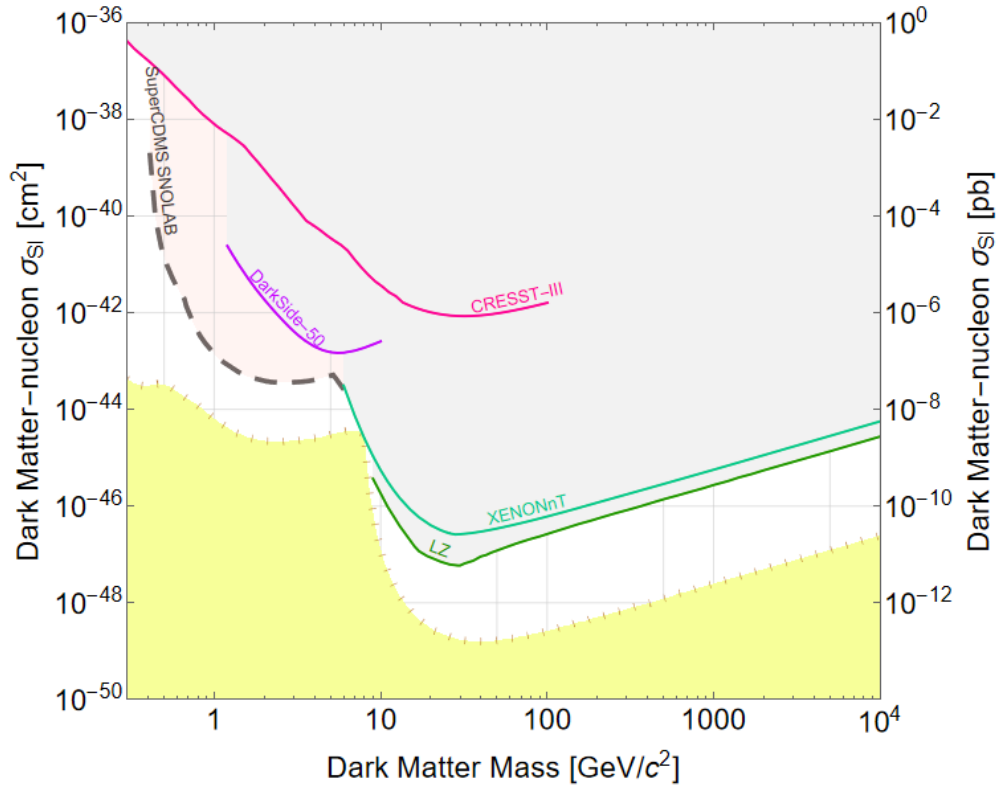


Figure 1-7. Exclusion limits on the spin-independent dark-matter-nucleon cross section versus dark matter mass. Solid curves are measured 90% C.L. limits. The SuperCDMS SNOLAB curve is a projected sensitivity. The yellow region marks the neutrino floor. Generated with the SuperCDMS dark matter limit plotter [21].

1.5.2 Electron Recoil Searches

Below about $1 \text{ GeV}/c^2$ the kinematics of elastic nuclear scattering become prohibitive, and the most sensitive searches instead look for dark matter scattering from electrons. Because an electron is light and loosely bound, even a sub-MeV particle can liberate a measurable charge. Detectors for this channel are built around ionization readout with single-quantum sensitivity. Skipper charge-coupled devices, used by SENSEI and DAMIC-M, resolve individual electron-hole pairs through repeated non-destructive charge measurement. The SuperCDMS HVeV detectors take the phonon route instead, amplifying each electron-hole pair into a phonon signal that sits well above the noise. Operated underground with eV-scale resolution, these detectors constrain dark-matter–electron scattering for masses down to the MeV scale, as well as the absorption of bosonic dark matter such as dark photons and axion-like particles. Figure 1-8 shows the current constraints across these channels. Across both channels, no experiment has reported a statistically significant excess above background. Each search instead places an upper bound on the interaction cross section, and the steady descent of these bounds over the past decade traces the falling thresholds and growing exposures of the field. That descent does not continue without limit. At sufficiently small cross sections, coherent elastic scattering of solar, atmospheric, and diffuse supernova neutrinos produces nuclear recoils indistinguishable from a dark matter signal in a non-directional detector [23]. This neutrino fog is no longer a projection; LZ has reported the observation of coherent elastic scattering of boron-8 solar neutrinos [24], the first direct encounter of a dark matter experiment with this irreducible background. Reaching beyond it will require directional sensitivity or channels that can statistically separate the two populations.

LZ met the fog at the sensitivity its multi-tonne exposure has reached, while the cryogenic phonon detectors pushing into the sub-GeV and electron-recoil region are limited instead by how low a threshold they can reach and how faithfully they can model a signal of only a few quanta. This is the region that motivates reading the phonon channel, and it is where SuperCDMS operates germanium and silicon crystals instrumented with athermal phonon sensors. What such a

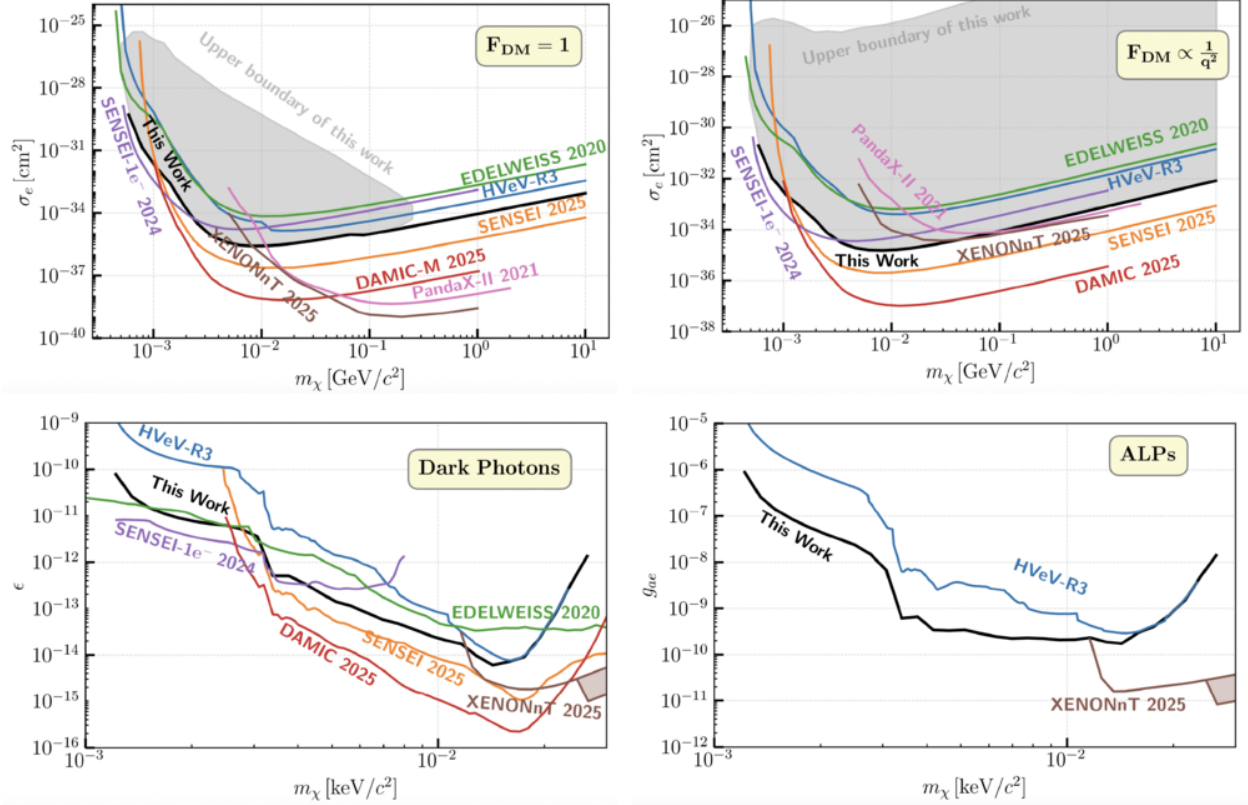


Figure 1-8. Constraints on dark-matter–electron interactions, with the SuperCDMS HVeV result [22] labeled “This Work.” The top panels show the scattering cross section for a heavy and a light mediator, and the bottom panels show dark photon and axion-like particle absorption limits.

detector can exclude, or one day claim, rests on a quantitative understanding of how a recoil inside the crystal becomes the phonon signal that reaches its sensors.

CHAPTER 2

SOLID-STATE DETECTOR RESPONSE

SuperCDMS uses solid-state targets of germanium or silicon held at a few tens of millikelvin. The cryogenic operating temperature does two things at once. It suppresses the thermal background of the crystal, since the band gap is far larger than the thermal energy available at these temperatures and almost no carriers are excited across it on their own, so an excitation that does appear has been produced by an interaction. It also places the crystal below the transition temperature of the superconducting sensors that read it, which is what allows them to be biased into their transition and operated at all.

A recoil excites the crystal in two ways at once, promoting electrons across the band gap into mobile carriers and setting the lattice vibrating, and it divides its energy between the two. A large voltage bias then drives the carriers from the interaction site to the faces of the detector, and as they drift they radiate further phonons through the Neganov-Trofimov-Luke effect, drawing that energy from the bias supply rather than from the recoil. What arrives at the instrumented surface is carried entirely by phonons, and it holds both the energy the recoil deposited and the work the field performed on the charges it freed. Superconducting sensors patterned onto that surface absorb them.

The number of carriers a deposit creates, the paths those carriers follow under the bias and the phonons they radiate as they drift, the propagation of those phonons through the crystal, and the fraction of them a sensor absorbs all contribute to the recorded pulse. Each carries its own uncertainty, and together they set how faithfully that pulse reflects the recoil that produced it.

2.1 The High-Voltage Detectors

Germanium and silicon are both used as targets. The lighter silicon nucleus is better matched kinematically to sub-GeV dark matter and yields a larger recoil for a given particle mass, while germanium offers the greater target mass per detector. Running the two together extends the reach across the mass range and lets a candidate signal be checked against a response that should differ between the targets.

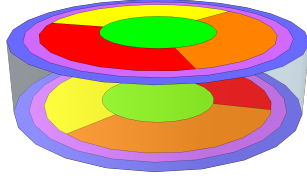
The SuperCDMS SNOLAB detectors are cylindrical crystals 10 cm in diameter and 3.3 cm thick, 1.4 kg for germanium and 0.6 kg for silicon. Both faces are instrumented with Quasiparticle-trap-assisted Electrothermal-feedback Transition-edge Sensors, or QETs, which respond to phonons. The QETs on each face are grouped into six channels, an inner region surrounded by wedge-shaped and annular outer channels (Figure 2-1a), and the relative signal across them indicates where in the crystal a recoil occurred. The crystal is biased at up to roughly 100 V, and with no charge electrodes to read, nothing but the phonon sensors sets the threshold. The alternative iZIP design reads ionization and phonons separately at a bias of a few volts, buying the ability to tell an electron recoil from a nuclear one at the cost of a threshold limited by the charge measurement. The high-voltage design gives that discrimination up and reaches lower in energy, and it is the design that carries the search for sub-GeV dark matter.

The HVeV detector sits at the other end of the scale (Figure 2-1b). It is a prototype from the same program, a silicon crystal of $1\text{ cm} \times 1\text{ cm} \times 4\text{ mm}$ carrying a QET array on one face and a bias electrode on the other, the QETs divided into two channels, an inner region and the annulus around it. It is biased and read out on the same principle, a high voltage across the crystal and phonons alone at the sensors. What the small volume buys is resolution. With little crystal to fill and few sensors to add noise, the HVeV resolves phonon energies at the electronvolt scale, enough to separate the discrete peaks produced by one electron-hole pair, by two, by three, and so on. A model of what a recoil does inside the crystal can be confronted with data of that sharpness in a way no kilogram-scale detector allows.

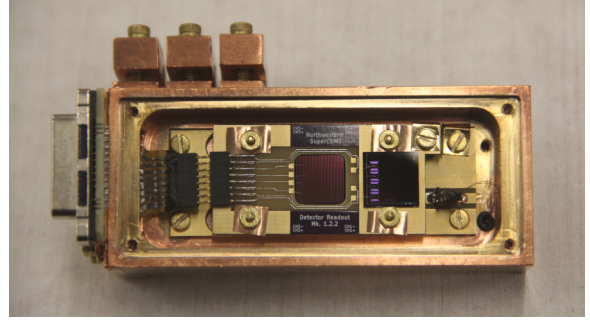
The two differ in size and in geometry rather than in physics. Both reduce a recoil to a single number, the phonon energy delivered to an instrumented face. What that number records, and how faithfully it reflects the recoil that produced it, is fixed by everything that happens in between.

2.2 Energy Deposition and Ionization Yield

A particle that scatters inside a detector crystal deposits its energy into two subsystems. Part of it frees electrons from their bonds and creates mobile charge carriers, and the rest sets the lattice vibrating and is carried away as phonons. Every recoil populates both, and the fraction that



(a) A SuperCDMS SNOLAB HV detector with six phonon channels on each face [25].



(b) The HVeV run 2 detector in its copper housing [26].

Figure 2-1. The two high-voltage detectors studied in this work.

appears as charge rather than phonons is what the experiment must know to convert a measured signal back into a recoil energy. That fraction is not universal. It depends on whether the recoiling particle struck an electron or a nucleus, and for nuclear recoils it falls steeply at the low energies where sub-GeV dark matter would deposit its signal.

2.2.1 The Semiconductor Crystal

Silicon and germanium are semiconductors. The periodic potential of the crystal lattice organizes the allowed electronic energies into bands, and a forbidden gap of width E_g separates the filled valence band from the empty conduction band. This gap is 1.17 eV in silicon and 0.74 eV in germanium at the operating temperature. It is the minimum energy needed to lift an electron from the valence band into the conduction band.

The gap is what makes these crystals usable as detectors. In thermal equilibrium the density of carriers excited across it is

$$n_i \propto \exp\left(-\frac{E_g}{2k_B T}\right), \quad (2-1)$$

and at the operating temperature of a few tens of millikelvin the exponent is so large that this density corresponds to far less than a single carrier in the entire crystal. The crystal is a near-perfect insulator in the dark, and any electron-hole pair that appears is the signature of an energy deposit. At the same time the gap is small enough that a recoil of a few electronvolts creates such a pair.

Silicon and germanium are both indirect-gap semiconductors, meaning the conduction-band minimum and the valence-band maximum occur at different crystal momenta. An electron crossing the gap must therefore change its momentum as well as its energy, and since a photon carries negligible momentum, a phonon must take part in the transition to conserve crystal momentum. Pair creation in these materials is intrinsically coupled to the lattice.

This coupling is one reason the energy required to create a pair exceeds the gap, though not the largest. The carriers freed by an energetic deposit are born well above the band edges, and as they relax toward equilibrium they shed their excess kinetic energy through a long succession of phonon emissions. The average energy consumed per electron-hole pair, including all of these losses, is the pair-creation energy ε , a material constant several times larger than the gap, $\varepsilon = 3.82$ eV in silicon and 2.96 eV in germanium.

A deposit of energy E_r does not in general spend all of itself on ionization. Only a fraction $Y(E_r)$, the ionization yield, goes into freeing electron-hole pairs; the remainder is delivered directly to the lattice as prompt phonons. The yield depends strongly on the type of recoil, taking very different values for electronic and nuclear recoils. The average number of pairs produced by a deposit of energy E_r is therefore

$$\langle N \rangle = \frac{Y(E_r) E_r}{\varepsilon}, \quad (2-2)$$

with the remaining energy $[1 - Y(E_r)]E_r$ flowing into the lattice as phonons. The gap sets whether a pair can be created at all; the pair-creation energy and the yield together set how many are created, and with them how the deposited energy divides between the electronic and vibrational systems.

2.2.2 Recoil Types

How the deposited energy divides between charge and phonons depends on what the incoming particle struck. A collision with an atomic electron and a collision with an atomic nucleus deposit the same energy in very different ways.

When the incident particle transfers its energy to an atomic electron, that electron is ejected with a kinetic energy that can reach the full recoil energy. While its energy stays high it drives a

cascade of secondary ionizations, each collision freeing further electrons and dividing the available energy among the carriers. Once the carrier energies fall below the threshold for further ionization, the remaining kinetic energy is shed as phonons. The cascade ends with a population of electron-hole pairs whose mean number is given by Eq. (2-2). That number fluctuates less than a Poisson process would, because the total energy is fixed and successive ionizations are therefore anticorrelated. The Fano factor F quantifies this suppression through $\text{Var}(N) = F\langle N \rangle$ with $F < 1$, and it sets the intrinsic resolution of an ionization measurement,

$$\sigma_E = \sqrt{F \varepsilon E}. \quad (2-3)$$

Because an electron recoil converts essentially all of its energy into ionization, it defines the reference scale for the detector response.

A nuclear recoil instead sets an entire atom in motion. The recoiling nucleus travels through the lattice and loses energy by two mechanisms that act at once. Electronic stopping is the excitation and ionization of the surrounding electrons, and it produces the electron-hole pairs that form the measurable signal. Nuclear stopping is elastic scattering from other lattice nuclei, which transfers energy directly into atomic motion and, where the transfer is large enough, displaces atoms from their sites. The energy routed through nuclear stopping ends up as phonons and lattice damage rather than as ionization. The balance between the two depends on the recoil energy. A fast recoil couples efficiently to the electrons and loses most of its energy to ionization, while a slow recoil loses an increasing share to atomic motion. The struck atoms are themselves set in motion and initiate their own sub-cascades, each again dividing its energy between the two mechanisms, so a single primary recoil generates a branching tree of collisions whose ionization is summed over every branch. The outcome is that a nuclear recoil produces fewer electron-hole pairs than an electron recoil of the same energy, and the deficit grows as the energy falls.

2.2.3 Ionization Yield

The quantity that captures the difference between the two recoil types is the *ionization yield* Y , defined as the fraction of the recoil energy that emerges as ionization, normalized to the response of an electron recoil of the same energy. An electron recoil deposits nearly all of its energy into the electronic system and is taken as the reference, $Y \equiv 1$. A nuclear recoil diverts part of its energy into atomic motion that never produces charge, so it yields fewer electron-hole pairs and $Y < 1$. The ionization yield is also called the quenching factor and is conventionally quoted as a percentage.

The standard description of the nuclear-recoil ionization yield is due to Lindhard, Nielsen, Scharff, and Thomsen [27]. Their result is written in terms of a dimensionless recoil energy

$$\epsilon = 11.5 Z^{-7/3} E_R [\text{keV}], \quad (2-4)$$

where E_R is the recoil energy and Z the atomic number of the target. The factor $Z^{-7/3}$ originates in the Thomas-Fermi screening of the nuclear charge, so ϵ expresses the recoil energy in units natural to the screened collision. In terms of ϵ the ionization yield is

$$Y(\epsilon) = \frac{k g(\epsilon)}{1 + k g(\epsilon)}, \quad (2-5)$$

with

$$g(\epsilon) = 3 \epsilon^{0.15} + 0.7 \epsilon^{0.6} + \epsilon. \quad (2-6)$$

The single parameter k sets the strength of the electronic stopping relative to the nuclear stopping and depends on the target through

$$k = 0.133 Z^{2/3} A^{-1/2}, \quad (2-7)$$

where A is the atomic mass number, giving $k \approx 0.146$ for silicon and $k \approx 0.157$ for germanium. Lindhard model reproduce the measured yield in silicon and germanium well above a few keV.

The Lindhard model treats the target electrons as a free-electron gas and imposes no lower limit on the velocity at which a recoiling ion can still excite them. This assumption fails at low energy, where a slow ion should not be able to lift an electron across the band gap, and the model consequently overpredicts the yield in the sub-keV region. Sarkis [28] addressed this by restoring the atomic binding energy to the calculation. In their treatment a recoiling atom must overcome the energy binding it to its lattice site before it can pass energy into a sub-cascade, which introduces a cutoff and suppresses the yield below a recoil energy of roughly twice the binding energy. The Sarkis model bends the predicted yield below the Lindhard curve at low energy, in the direction indicated by measurement.

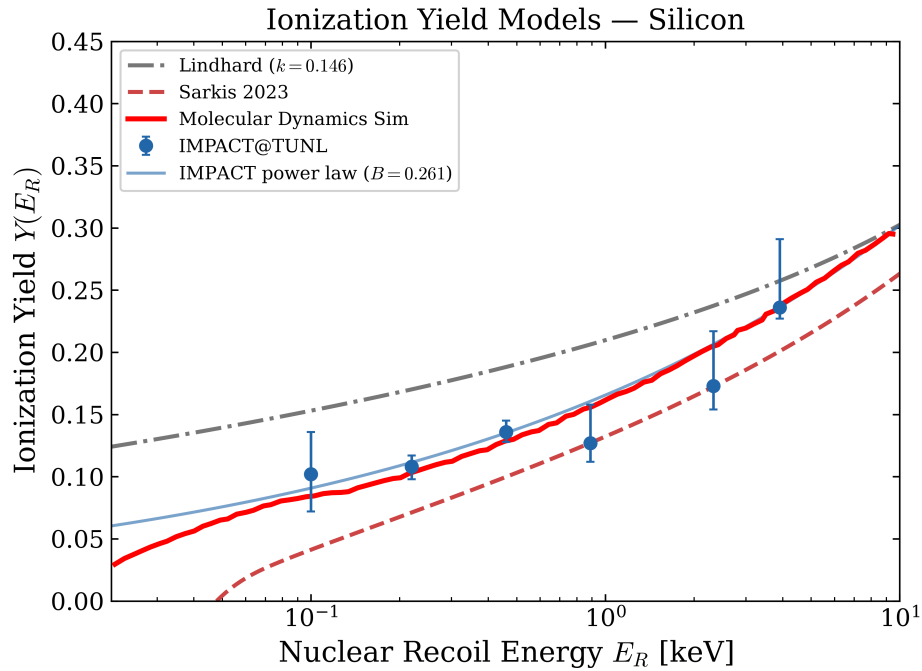


Figure 2-2. Ionization yield in silicon as a function of nuclear recoil energy, expressed as a quenching factor. The Lindhard and Sarkis models are compared with the molecular-dynamics calculation of Fang et al. [29] and with the IMPACT@TUNL measurement [30].

Figure 2-2 compares these models with measurement. The IMPACT@TUNL experiment [30] measured the silicon ionization yield down to a recoil energy of 100 eV, the lowest reached to date, using a cryogenic phonon-sensitive detector exposed to a monoenergetic neutron beam. The measurement confirms the deviation from Lindhard theory and shows no sign

of an ionization threshold down to 100 eV. A molecular-dynamics calculation, which follows the cascade atom by atom rather than treating it statistically [29], is shown alongside. The spread among the models at sub-keV energies is large, and it falls precisely in the recoil-energy range relevant to sub-GeV dark matter, where the ionization yield directly determines how a deposited energy maps onto a detectable signal and so sets the sensitivity of the search.

2.3 Charge Transport

A recoil leaves behind a population of electron-hole pairs, and what becomes of those carriers depends on whether a voltage bias is applied across the crystal. Without a bias the only forces on a carrier are its random scattering from lattice vibrations and impurities. The electrons and holes diffuse outward from the interaction point without a preferred direction, and most never reach a surface. An electron and a hole that meet recombine, returning their energy to the lattice as phonons, and carriers that encounter a charged impurity are trapped on it. Little charge is collected, and the energy that began as ionization is largely returned to the lattice as vibrations. The high-voltage design exists to change this. With a bias applied the carriers are driven across the crystal, and in crossing it they radiate the phonons that carry the amplified signal. The simulations shown throughout use the G4CMP package [31], which models each of the transport processes that govern this drift.

2.3.1 Conduction Valleys and Intervalley Scattering

The way a carrier drifts under the bias is set by the band structure it moves through, and holes and electrons differ sharply because their effective masses differ. A hole sits at the valence-band maximum, which in silicon lies at the zone center, and its effective mass is well approximated by an isotropic scalar, so a hole responds to the field much as a free particle would and drifts along the field direction.

Near the bottom of a valley the electron energy is quadratic in the crystal momentum measured from the valley minimum $\mathbf{k}_0$,

$$E(\mathbf{k}) = E_c + \frac{\hbar^2}{2} \sum_{i,j} (k - k_0)_i (M^{-1})_{ij} (k - k_0)_j, \quad (2-8)$$

where E_c is the valley minimum energy and M_{ij} the effective-mass tensor. In a frame aligned with the valley, M is diagonal, with a longitudinal mass $m_{\parallel}$ along the valley axis and a degenerate pair of transverse masses $m_{\perp}$ across it. In silicon these are $m_{\parallel} = 0.91 m_e$ and $m_{\perp} = 0.19 m_e$, so the valley is strongly elongated, while the hole mass is $m_h = 0.50 m_e$.

The anisotropy of the mass tensor makes the electron dynamics awkward. In the valley frame the equation of motion under a field $\mathbf{E}$ is

$$\frac{dv_i}{dt} = \frac{eE_i}{m_i}, \quad (2-9)$$

with m_i the mass along axis i . Because the longitudinal and transverse masses differ, the acceleration is not parallel to the field, and the electron drifts *obliquely*, along its valley axis rather than along $\mathbf{E}$ [32]. An electron population spread over the six valleys therefore splits into spatially distinct streams as it crosses the crystal.

The Herring-Vogt transformation [33] removes the anisotropy by rescaling each coordinate axis. It is the diagonal map

$$T_{\text{HV}} = \text{diag}\left(\sqrt{\frac{m^*}{m_x}}, \sqrt{\frac{m^*}{m_y}}, \sqrt{\frac{m^*}{m_z}}\right), \quad (2-10)$$

with the conductivity effective mass m^* defined by

$$\frac{3}{m^*} = \frac{1}{m_x} + \frac{1}{m_y} + \frac{1}{m_z}. \quad (2-11)$$

Transformed coordinates and velocities are obtained by applying T_{HV} to the originals. Carrying the rescaling through the kinetic term of Eq. (2-8) leaves it isotropic,

$$E = E_c + \frac{\hbar^2 k^{*2}}{2m^*}, \quad (2-12)$$

and the equation of motion follows the same simplification,

$$\frac{dv_i^*}{dt} = \frac{eE_i^*}{m^*}, \quad (2-13)$$

the equation of an isotropic carrier. Once an electron is carried into the transformed frame it can be propagated, and its phonon emission computed, with the same recipe used for a hole. The tensor character of the electron survives only as the oblique drift direction and the velocity rescaling.

A drifting electron does not remain in one valley. Intervalley scattering moves it from one conduction valley to another, and since each valley carries a different oblique drift direction, a valley change redirects the electron. Repeated intervalley scattering drives the electron population away from its separated oblique streams and back toward the field direction. The process occurs by scattering off the lattice or off impurities, and its rate depends on the field strength, with lattice scattering dominant above a few volts per centimeter and impurity scattering dominant below. G4CMP provides two parameterizations of this dependence. A quadratic form [34],

$$\nu_{IV} = A \left(|\mathbf{E}|^2 - E_0^2 \right)^{\alpha/2}, \quad (2-14)$$

and a linear form [35],

$$\nu_{IV} = b + m |\mathbf{E}|^\alpha. \quad (2-15)$$

The coefficients are fixed by fits to test devices. For silicon the quadratic form takes

$A = 3.5 \times 10^{-20}$ Hz, $E_0 = 3.4 \times 10^3$ V/m, and $\alpha = 7.47$, while the linear form takes

$b = 1.5 \times 10^6$ Hz, $m = 1.5$ Hz, and $\alpha = 4.0$. Holes, having no valley structure, do not undergo intervalley scattering.

2.3.2 Neganov-Trofimov-Luke Phonon Emission

Under an applied bias the field continuously does work on a drifting carrier, feeding it kinetic energy. The carrier cannot accelerate without limit. It sheds the energy it gains through charge-phonon scattering, an intravalley process in which the carrier stays in its valley but emits

an acoustic phonon. Treating the emission with Fermi's golden rule gives a scattering rate¹

$$\frac{1}{\tau} = \frac{v_L}{3 l_0} \left(\frac{k}{k_L} \right)^2 \left(1 - \frac{k_L}{k} \right)^3, \quad (2-16)$$

where k is the carrier wavevector, v_L the longitudinal phonon phase velocity, l_0 a characteristic scattering length fixed by the crystal density and the deformation potential, and k_L the wavevector set by the longitudinal sound speed,

$$k_L = \frac{m v_L}{\hbar}. \quad (2-17)$$

The phonons emitted by this process are the Neganov-Trofimov-Luke (NTL) phonons. The emission is a charge-phonon scattering analogous to Cherenkov radiation, in which a carrier driven faster than the lattice can quietly accommodate radiates the excess away. The phonon is emitted at an angle θ to the carrier's instantaneous wavevector $\mathbf{k}$, and the same golden-rule treatment that gives the rate fixes the distribution of that angle,

$$P(\theta) d\theta = \frac{v_L}{l_0} \left(\frac{k}{k_L} \right)^2 \left(\cos \theta - \frac{k_L}{k} \right)^2 \sin \theta d\theta, \quad (2-18)$$

defined over the range $0 \leq \theta \leq \arccos(k_L/k)$.

The shape of this distribution carries the physics of the emission. The two factors compete. The first, $\sin \theta$, vanishes in the forward direction and favors wide angles, while the second, $(\cos \theta - k_L/k)^2$, vanishes at the kinematic edge $\theta = \arccos(k_L/k)$ and favors the forward direction. The most probable angle lies between the two. Setting $dP/d\theta = 0$ and writing $c = k_L/k$ for the ratio of the threshold wavevector to the carrier wavevector, the maximum satisfies $3 \cos^2 \theta - c \cos \theta - 2 = 0$, so

$$\cos \theta^* = \frac{c + \sqrt{c^2 + 24}}{6}. \quad (2-19)$$

¹ The quadratic $(k/k_L)^2$ in Eq. (2-16) follows from the standard derivation [36, 37] and from integrating Eq. (2-18) over its kinematic range. A linear (k/k_L) appears in Leman [38] and is carried into the G4CMP review [31], but the G4CMP code itself implements the rate correctly.

The behavior in the two limits is instructive. A fast carrier, with $k \gg k_L$ and $c \rightarrow 0$, emits most probably at $\theta^* = \arccos\sqrt{2/3} \approx 35^\circ$, well off the forward direction. As the carrier slows toward $k \rightarrow k_L$ and $c \rightarrow 1$, the allowed angular range $0 \leq \theta \leq \arccos(k_L/k)$ shrinks to zero and the emission rate of Eq. (2-16), which carries a factor $(1 - k_L/k)^3$, vanishes with it. The wavevector $k_L = mv_L/\hbar$ is therefore a hard emission threshold. It corresponds to a carrier moving at the longitudinal sound speed, and a carrier at or below this speed cannot emit an NTL phonon at all, exactly as a charge slower than the phase velocity of light emits no Cherenkov radiation. Emission occurs only for $k > k_L$, and the cone, narrow just above threshold, opens toward 35° as the carrier becomes increasingly supersonic. Because the carrier wavevector is continually redirected by scattering, the cone follows the instantaneous direction of motion and the emitted phonons are spread over a correspondingly broad range of laboratory angles. A carrier quickly reaches a terminal drift speed at which the energy it draws from the field is emitted as NTL phonons as fast as it is supplied.

The consequence is an exact bookkeeping of energy. A carrier that drifts across a potential difference V converts the full electrostatic work eV into NTL phonons. For a recoil that creates $\langle N \rangle$ electron-hole pairs, all of which drift the bias V , the phonon energy collected is

$$E_{\text{tot}} = E_r + \langle N \rangle eV, \quad (2-20)$$

the sum of the original recoil energy and the NTL contribution. With $\langle N \rangle$ set by the pair-creation energy of Eq. (2-2), the NTL term grows linearly with the bias, and at the high voltages of the HV design it dominates the phonon signal. This amplification is what lifts a recoil of a few electronvolts above the sensor threshold, and it is the reason the HV detector reaches the lowest energies. The NTL phonons are emitted continuously along the drift path rather than at a point, so a recoil deposits phonon energy spread across the carrier trajectories, a feature that shapes the phonon population reaching the sensors.

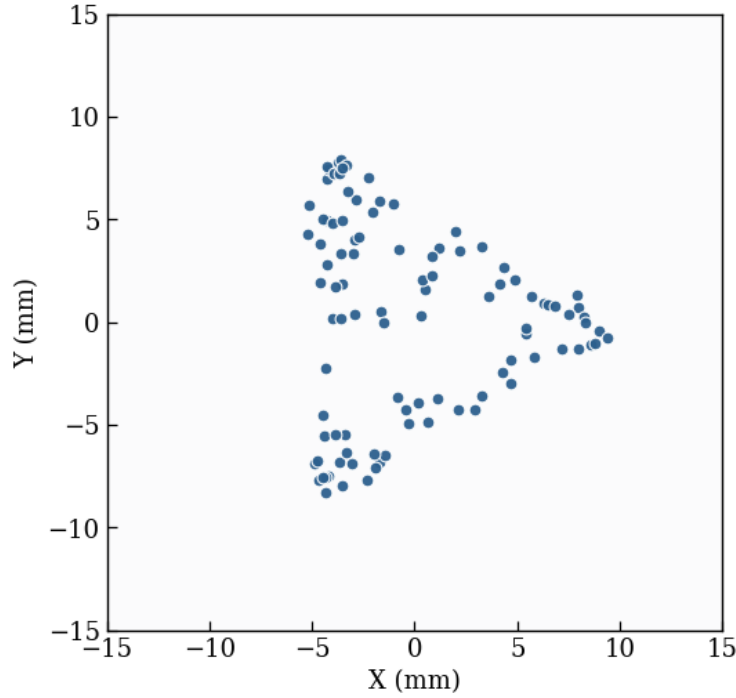


Figure 2-3. Simulated electron arrival positions at the face of a silicon detector, for events at the crystal center under a 1 V bias along $\langle 111 \rangle$. The arrivals spread into a three-fold pattern set by the conduction valleys.

2.3.3 Charge Transport in Simulation

Figure 2-3 shows where electrons arrive at the face when these transport processes are simulated together. The arrivals do not collect at the event position. Instead they spread into a distinct three-fold pattern, each lobe fed by valleys whose oblique drift carries the electrons away from the bias axis [39, 40]. The bias is deliberately low; were it much higher, intervalley scattering would redirect the electrons toward the field on a length shorter than the crystal, washing the pattern out before the arrivals were recorded. The figure therefore makes the valley structure visible in a way that the tracks alone cannot. Holes, which carry no valley structure, drift compactly along the field and arrive at the opposite face clustered near the event position. This electron-hole asymmetry in the transport carries through to an asymmetry in the phonon energy collected at the two faces.

The transport model carries known limitations. G4CMP does not explicitly simulate optical phonons, so in an intervalley scattering event crystal momentum is not conserved. Rather than

emitting an optical phonon and rotating the electron wavevector, the model relocates the electron to the new valley directly and lets it re-radiate acoustic phonons. The spectrum of phonons produced by intervalley scattering at high field is therefore only approximate. The anisotropy of hole propagation, a small effect measured in dedicated transport studies, is likewise not included, the hole being treated as a strictly isotropic carrier. Neither approximation is significant for the low-field behavior that governs the detectors studied here, but both bound the accuracy of the simulated phonon spectra that follow.

2.4 Phonon Dynamics

A recoil leaves a population of phonons in the crystal. Some are prompt phonons released as the recoil carriers thermalize toward the band edges. Some come from recombination of carriers that fail to reach a face. The largest contribution in a biased detector is the NTL phonons emitted along the drift paths. What the sensor records is not this initial population but what it becomes, after the phonons have crossed millimetres of crystal, scattered, decayed in frequency, and interacted repeatedly with the superconducting films on the detector face.

2.4.1 Initial Population, Propagation, and Modes

When a recoil creates an electron-hole pair, the carriers are born with kinetic energies well above the band edges. They cool to the band bottoms in picoseconds, emitting phonons across a broad spectrum that reaches the Debye cutoff, $\nu_D = 15$ THz in silicon and 8.6 THz in germanium. Once the carriers have settled to the band edges and begin to drift under the applied bias, NTL emission adds a much softer population, governed by the charge-phonon scattering of Eq. (2-16) and emitted into the angular distribution of Eq. (2-18). The NTL spectrum peaks below 1 THz and dominates the population whenever the bias is more than a few volts. The same charge-phonon scattering produces a residual NTL population even at zero bias, from the motion of the carriers between creation and recombination. Recombination itself releases an energy of order the bandgap as phonons, a contribution that is suppressed by the bias and dominant only at zero bias. The initial population therefore extends from sub-THz frequencies up to the Debye

cutoff, and the highest-frequency phonons in it are short-lived, decaying anharmonically before they propagate any appreciable distance.

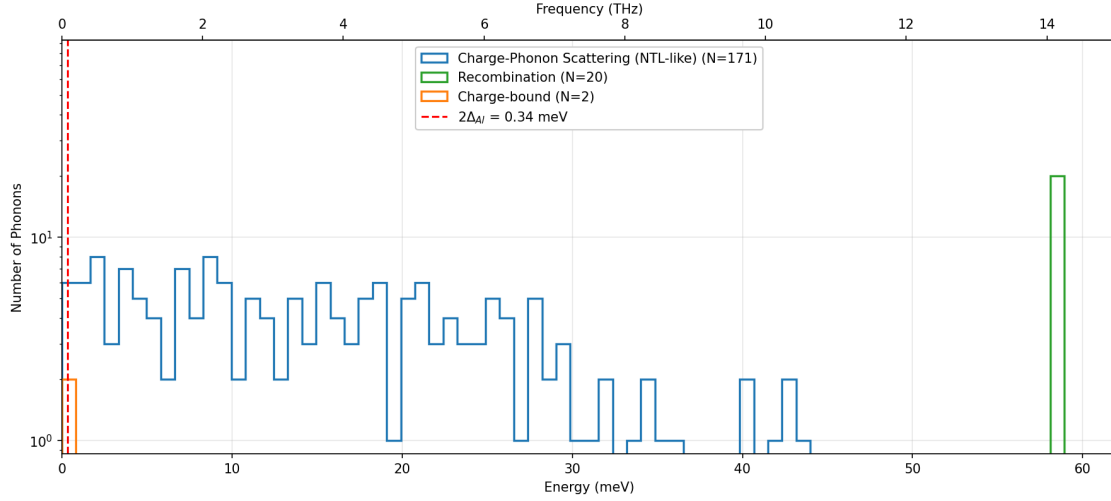


Figure 2-4. Simulated phonon emission spectrum at zero bias in a silicon HVeV detector (Chapter 4), from G4CMP. The charge-phonon scattering (NTL-like), recombination, and charge-bound contributions are shown separately. The charge-bound term being the residual kinetic energy above the band gap released when carriers recombine at the boundary. The aluminum pair-breaking threshold $2\Delta_{Al}$ is marked.

Figure 2-4 shows the simulated emission spectrum at zero bias. The charge-phonon scattering of §2.3.2, here driven only by the residual motion of the carriers since no bias is applied, produces phonons of the same type as the NTL phonons. The label recombination covers the bandgap energy released as phonons when a carrier recombines. The label charge-bound covers the residual kinetic energy that a carrier carries above the band edge at the moment of recombination, released as phonons in the same event.

The phonons that emerge from this initial population do not propagate isotropically. In an anisotropic elastic medium the relation between frequency and wavevector follows from the Christoffel equation,

$$\rho \omega^2 e_i = C_{ijml} k_j k_m e_l, \quad (2-21)$$

in which ρ is the crystal density, C_{ijml} the elasticity tensor, and $\mathbf{e}$ the polarization vector. For each wavevector $\mathbf{k}$ the equation gives three eigenvalues corresponding to three acoustic modes, one

longitudinal and two transverse. The two transverse branches are distinguished along generic directions in the cubic Brillouin zone and are called fast-transverse and slow-transverse, while they remain degenerate along the high-symmetry axes. The group velocity, the velocity at which phonon energy moves, is $\mathbf{v}_g = \nabla_{\mathbf{k}} \omega(\mathbf{k})$. Because C_{ijkl} is anisotropic, $\mathbf{v}_g$ is in general not parallel to $\mathbf{k}$, and the directions of largest mismatch concentrate phonon flux into caustic channels [41].

2.4.2 Bulk Scattering and the NTL Spectrum

Two processes act on the initial population as it crosses the crystal. The first is anharmonic decay. Anharmonic terms in the elastic coupling between atoms allow a longitudinal phonon to decay into two phonons of lower frequency, conserving energy and momentum, at a rate

$$\Gamma_A = A \nu^5, \quad (2-22)$$

where ν is the phonon frequency and A is a decay constant fixed by the crystal, $A = 7.41 \times 10^{-56} \text{ s}^4$ in silicon and $6.43 \times 10^{-55} \text{ s}^4$ in germanium [42]. The fifth-power scaling makes the decay catastrophic near the Debye cutoff and harmless far below it. A phonon near ν_D decays in picoseconds, while one at a tenth of ν_D lives indefinitely on the timescale of the experiment. The transverse decay rate is negligible and is not implemented in G4CMP, so the cascade is driven entirely by longitudinal decays. A transverse phonon can only decay if it first scatters into a longitudinal one through the isotopic mode mixing described below. Each decay produces either a longitudinal and a transverse phonon ($L \rightarrow L'T$) or two transverse phonons ($L \rightarrow TT$), and the branching ratio between the two channels is a property of the crystal, taking the value 0.20 for the $L \rightarrow L'T$ fraction in silicon and 0.26 in germanium [42]. The energies of the outgoing phonons within each channel are drawn from probability distributions set by the elastic constants.

The second process is isotopic scattering. Mass defects from naturally occurring isotopes act as elastic scatterers, with a rate

$$\Gamma_I = B \nu^4, \quad (2-23)$$

where $B = 2.43 \times 10^{-42} \text{ s}^3$ in silicon and $3.67 \times 10^{-41} \text{ s}^3$ in germanium [43]. Isotopic scattering preserves the phonon energy, randomizes the propagation direction, and can mix the polarization mode. The probability of each polarization outcome is set by the overlap of the incoming and outgoing polarization vectors weighted by the density of states in the destination branch, so over many scatterings the populations relax toward density-of-states ratios. In silicon those ratios are approximately 9.3% longitudinal, 37.6% fast-transverse, and 53.1% slow-transverse.

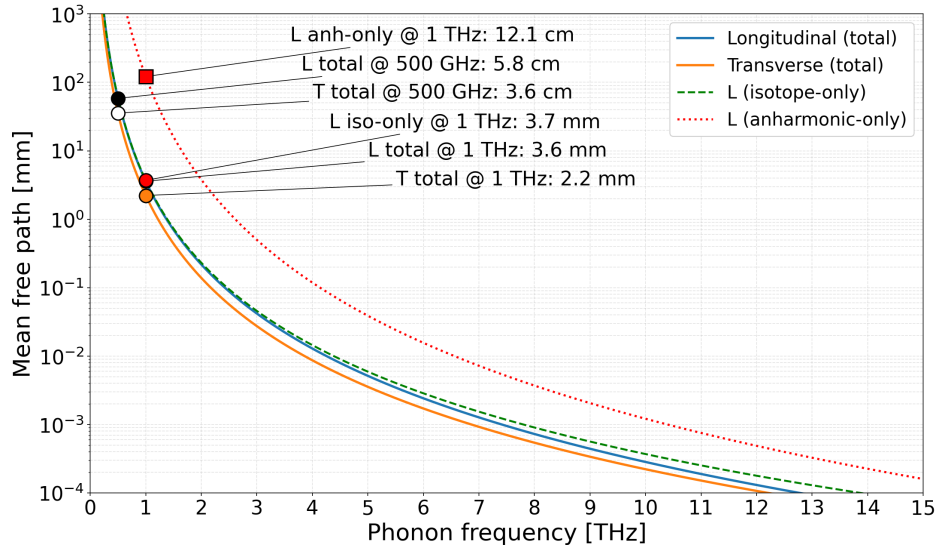


Figure 2-5. Phonon mean free path in silicon versus frequency, for longitudinal and transverse phonons, with isotopic and anharmonic contributions shown separately. Anharmonic decay sets the high-frequency limit, isotopic scattering the low-frequency one.

The two processes together set the phonon mean free path, shown in Figure 2-5. Anharmonic decay dominates at high frequency, isotopic scattering at low frequency, and the mean free path falls steeply between them. At 1 THz, representative of the bulk of the NTL emission, the mean free path is 3.6 mm for longitudinal phonons and 2.2 mm for transverse phonons in silicon, a small fraction of the crystal thickness. The consequence is a change in transport character with frequency. A high-frequency phonon scatters and decays many times over a short distance, moving quasi-diffusively while cascading into softer phonons. Once the population has softened into the sub-THz range the mean free path grows to millimetres, and the remaining propagation is ballistic and shaped by phonon focusing.

The same density-of-states weighting that drives isotopic mode mixing also governs the polarization of the NTL emission. When a drifting carrier emits an NTL phonon, the probability that the phonon is longitudinal, fast-transverse, or slow-transverse is drawn from the same mode densities, so the emitted population is already in the equilibrium ratios. Figure 2-6 shows the NTL emission spectrum in a SuperCDMS HV silicon detector, separated into the longitudinal and transverse branches.

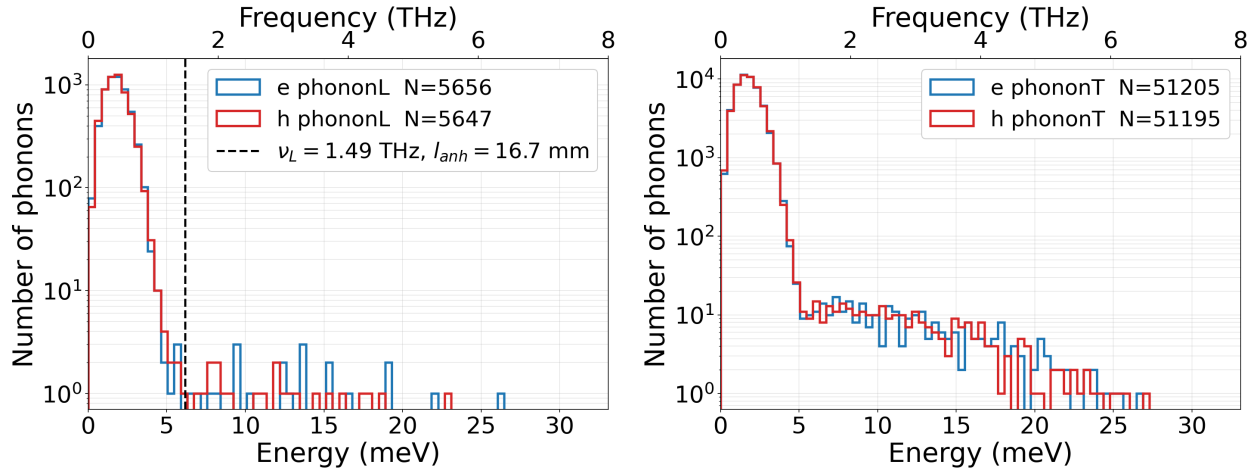


Figure 2-6. Simulated NTL phonon emission spectra in a silicon SuperCDMS HV detector, from G4CMP, separated into longitudinal and transverse modes. The dashed lines mark where the anharmonic mean free path falls to 16.65 and 1.0 mm, both well above the bulk of the emission.

The transverse population dominates by roughly an order of magnitude, the ratio implied by the density-of-states weighting. The dashed vertical lines mark the frequencies at which the anharmonic mean free path equals half the crystal thickness, 16.65 mm, the farthest a phonon born at the centre travels to reach a face, and at which it falls to 1 mm. Almost the entire NTL spectrum lies below both lines, so most NTL phonons reach a face without anharmonically decaying.

2.4.3 The QET Sensor and the Phonon Cycle

The phonon-sensitive faces are instrumented with Quasiparticle-trap-assisted Electrothermal-feedback Transition-edge Sensors, or QETs. A QET pairs a narrow tungsten transition-edge sensor wire with a set of aluminum collector fins, sketched in Figure 2-7. The tungsten wire runs along the centre of the QET, contacted on each side by short stubs that connect

to the aluminum fins in the overlap region where the trap lives. The fins extend outward from the stubs along the wire and in two end fins at the ends of the wire. Both metals are superconducting at the operating temperature, but their energy gaps differ by more than an order of magnitude. The aluminum gap is $\Delta_{\text{Al}} = 180 \mu\text{eV}$ and the tungsten gap is $\Delta_{\text{W}} \approx 10 \mu\text{eV}$. The fins cover far more area than the wire, so the QET acts as an antenna, collecting phonon energy over the fin area and funnelling it into the tungsten in the centre.

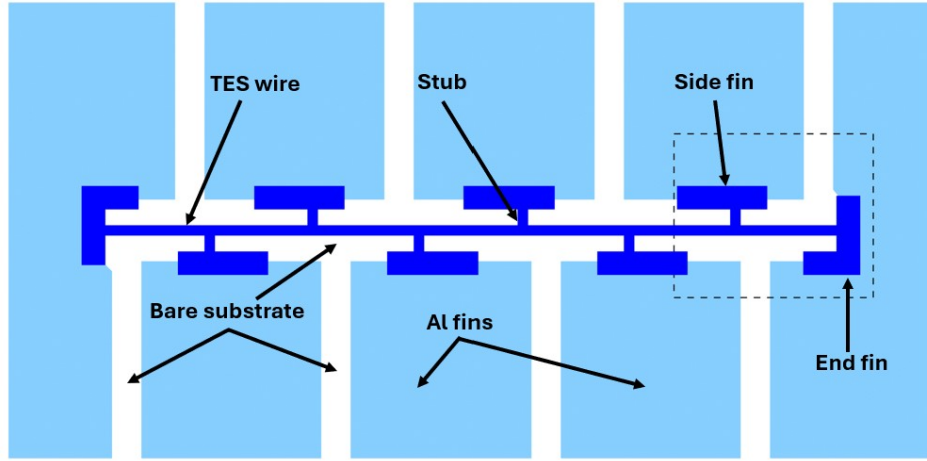


Figure 2-7. Schematic of the QET layout, showing the central tungsten wire, the stubs connecting it to the aluminum fins, and the surrounding side and end fins. The dashed box marks one QET cell; bare silicon occupies the regions between fins [26].

A phonon arriving at the crystal surface does not always make it to a sensor. The QETs cover part of the phonon-sensitive face; the rest is bare silicon between them, and the side walls are bare too. A phonon striking a bare surface reflects, either specularly with the angle of incidence preserved or diffusely with the reflection direction randomized. The choice between the two is made by a probability that G4CMP carries as a parameter of the surface. Reflected phonons re-enter the bulk and continue to propagate. Only phonons that strike a QET have a chance of contributing to the signal.

When a phonon hits a QET, the probability that it enters the QET footprint at all rather than reflecting back into the substrate is set by the QET phonon absorption probability, the parameter `PhononAbsQETs` in G4CMP. A phonon that is reflected returns to the bulk. G4CMP does not resolve the internal geometry of the QET; it treats the QET as a single rectangular footprint of the

same area as the actual device, with the aluminum and the tungsten lumped within it. After a phonon enters that footprint, a second roll determines whether it is absorbed directly into the tungsten transition-edge sensor without first producing quasiparticles in the aluminum. The probability of that direct absorption is the parameter `TESSubgapAbsorption`. Although the name suggests a sub-gap role, the same parameter is rolled for every phonon that enters the QET, regardless of whether its energy lies above or below $2\Delta_{\text{Al}}$. Its value is generally of order a few percent, and it is the dominant control on the slope of the long tail of the pulse, an effect developed in Chapter 4 Section 4.3.2.

If the phonon energy exceeds $2\Delta_{\text{Al}}$, the phonon breaks a Cooper pair into two quasiparticles in the aluminum, sharing its energy between them according to the Kaplan distribution,

$$f(E_{qp} | E_{ph}) \propto \frac{E_{qp} (E_{ph} - E_{qp}) + \Delta_{\text{Al}}^2}{\sqrt{(E_{qp}^2 - \Delta_{\text{Al}}^2)((E_{ph} - E_{qp})^2 - \Delta_{\text{Al}}^2)}}, \quad (2-24)$$

where E_{ph} is the incident phonon energy and E_{qp} the energy of one quasiparticle, the other carrying $E_{ph} - E_{qp}$ [44]. The quasiparticles diffuse through the aluminum and shed energy by emitting further phonons. The emitted phonons take a fraction of the parent quasiparticle's energy and the quasiparticle continues with the remainder. Some of these emitted phonons themselves carry enough energy to break another Cooper pair, extending the cascade inside the film. Once a quasiparticle falls below $3\Delta_{\text{Al}}$ it can no longer emit a pair-breaking phonon and is collected at the tungsten transition-edge sensor, where its energy enters the signal that Chapter 3 develops.

Once a quasiparticle is collected at the tungsten it thermalizes there by emitting subgap phonons that carry away its remaining kinetic energy. These phonons emerge from the tungsten and re-enter either the substrate or the aluminum, with most going to the substrate because of the geometry. The phonons emitted from the tungsten in this process are not implemented in G4CMP, which treats the tungsten as a perfect collector and stops following quasiparticles once they cross into it.

What G4CMP does follow is the phonon emission from the aluminum side of the cascade. The phonons emitted by quasiparticles in the aluminum can escape the fin and return to the substrate. The escape probability is approximated by

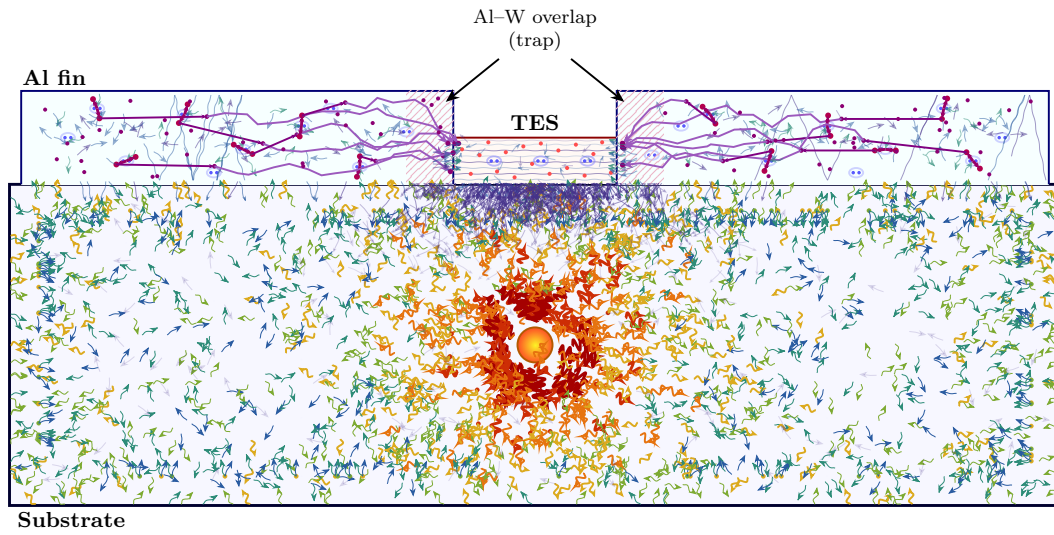
$$P_{\text{esc}}(E_{ph}) = \exp\left(-\frac{4d}{\lambda(E_{ph})}\right), \quad (2-25)$$

where d is the film thickness and $\lambda(E_{ph})$ the phonon mean free path in the film, set by the pair-breaking process. The factor of four accounts for the round-trip path across the film and for averaging over emission angle. The phonons that escape are assigned random directions and polarizations drawn from the equilibrium density-of-states ratios. They re-enter the bulk and propagate ballistically, scattering off bare surfaces, striking other QETs on either face, and feeding additional rounds of pair-breaking and quasiparticle production. The energy from a single recoil therefore makes many passes through the crystal before it is collected at a tungsten sensor or absorbed at a surface. What the detector measures is the accumulation of these passes.

Figure 2-8 illustrates the full cycle, from the recoil to the substrate to the QET and back.

2.4.4 Limitations of the Phonon Model

The simulated phonon population rests on several approximations that bound its accuracy. Optical phonons are not propagated. They are taken to downconvert instantaneously to acoustic modes, an assumption sound at millikelvin temperatures where the optical branches are unoccupied but inaccurate in the moments after high-energy decay. The quasiparticle dynamics in the aluminum are treated as lumped, with a single energy assigned to each quasiparticle and a single escape probability for the phonons it emits, rather than tracking the spatial diffusion of quasiparticles within the film, which remains an active area of development. Energy delocalization through that diffusion, on timescales relevant to the rise of the phonon signal, is therefore not captured. Anharmonic decay is treated as isotropic in its angular distribution after the initial mode mixing, which is accurate for the bulk-averaged spectrum but underestimates the caustic structure of the phonons born high in frequency. None of these approximations changes



••
 Cooper pair
 quasiparticle
 athermal phonon
 subgap phonon

•
 particle hit

Figure 2-8. Schematic of the full phonon cycle. Phonons born from the recoil propagate through the substrate to a QET, where above-gap phonons break Cooper pairs into quasiparticles in the aluminum. The quasiparticles diffuse to the tungsten and contribute to the signal; the phonons emitted during the cascade may escape to the substrate and feed further QET interactions. The TES then cools by emitting thermal phonons into the crystal.

the qualitative phonon cycle, but each bounds the precision with which the simulated phonon power, timing, and spatial distribution can be matched to data.

2.5 Position Dependence of the Phonon Response

The charge transport and phonon dynamics combine to make the collected signal depend on where in the crystal an event occurred. This dependence is what allows a detector to reconstruct event position and to distinguish bulk interactions from those near a surface.

2.5.1 Response of the Phonon Channels

The HV detector phonon-sensitive faces are divided into six channels per face as illustrated in Figure 2-9 labelled A through F on each face. An event deposits phonon energy across these channels in a pattern that depends on where in the crystal the event occurred.

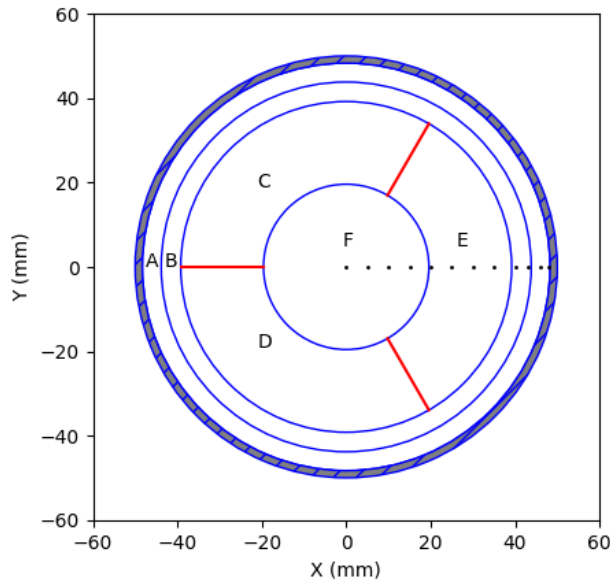


Figure 2-9. Phonon channels layout on HV detector, labelled A through F. The dots mark the event positions scanned radially outward along the x axis in the crystal central plane.

To isolate the position dependence from the depth dependence, the events of the scan are placed in the central plane of the crystal, the cross section midway between the two faces and perpendicular to the z axis, and stepped radially outward from the centre along the x axis. The positions are marked on Figure 2-9. Figure 2-10 shows the simulated phonon power reaching each channel as a function of time for this scan, in a silicon detector under a bias of 90 V. The channels

nearest the event collect the earliest and largest signal, while those farther away collect a smaller signal that arrives later, smeared by the longer propagation. As the event moves outward the balance of power shifts systematically from one channel to another.

The power reaching the channels is converted by the transition-edge sensors into the recorded pulses. Figure 2-11 shows the resulting traces, normalized, for the same set of radial positions. The dynamics of this conversion, the electrothermal response of the sensors, are developed in Chapter 3; here the traces serve to show that the position dependence of the phonon signal survives the sensor response and appears in the simulated pulse shapes. While the TES parameters vary per channel as well, the position dependence of the phonon arrival mainly affects the rising edge of the TES pulse, most visibly in the channels closest to and farthest from the event.

2.5.2 Energy Asymmetry Between the Faces

The traces fold together the crystal physics and the sensor response. The crystal physics can be isolated by asking how much phonon energy the bare simulation delivers to each face, with no sensor involved. The natural variable is the difference between the energy collected at the two faces as a function of the depth z of the event.

For a fixed deposit at the center of the detector, the total phonon energy emitted into phonons by the electron and by the hole is identical, matching the expected value set by the deposited energy and the bias. This follows from energy conservation by construction. A carrier that drifts to an electrode converts the full electrostatic work into NTL phonons regardless of the path it takes, so the two species deliver the same total energy. Any asymmetry between the faces in the simulation outcome is therefore not a matter of one carrier emitting more than the other. It is a matter of where along the trajectory the phonons are emitted and which face the ballistic phonons reach.

With the energy budget symmetric, the depth dependence of the face asymmetry comes entirely from the geometry of transport. The depth of an event sets not only how far each carrier travels but where along its path the NTL phonons are emitted, and that emission profile determines how the phonon energy partitions between the two faces once the phonons propagate ballistically.

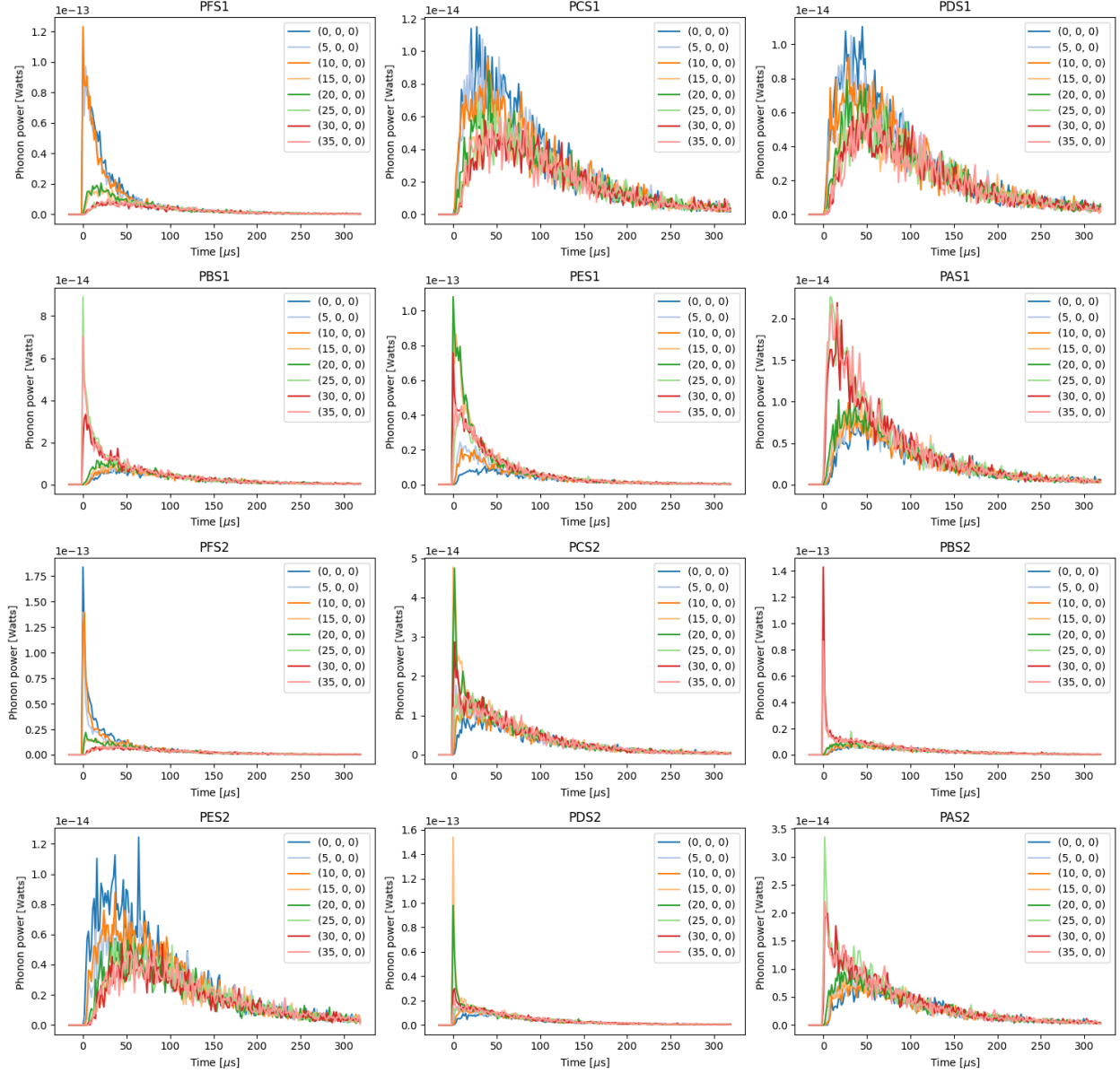


Figure 2-10. Simulated phonon power in each channel as a function of time, for the central-plane radial scan of Figure 2-9 in a Si HV detector under a 90 V bias. The distribution of power among the channels shifts with event position.

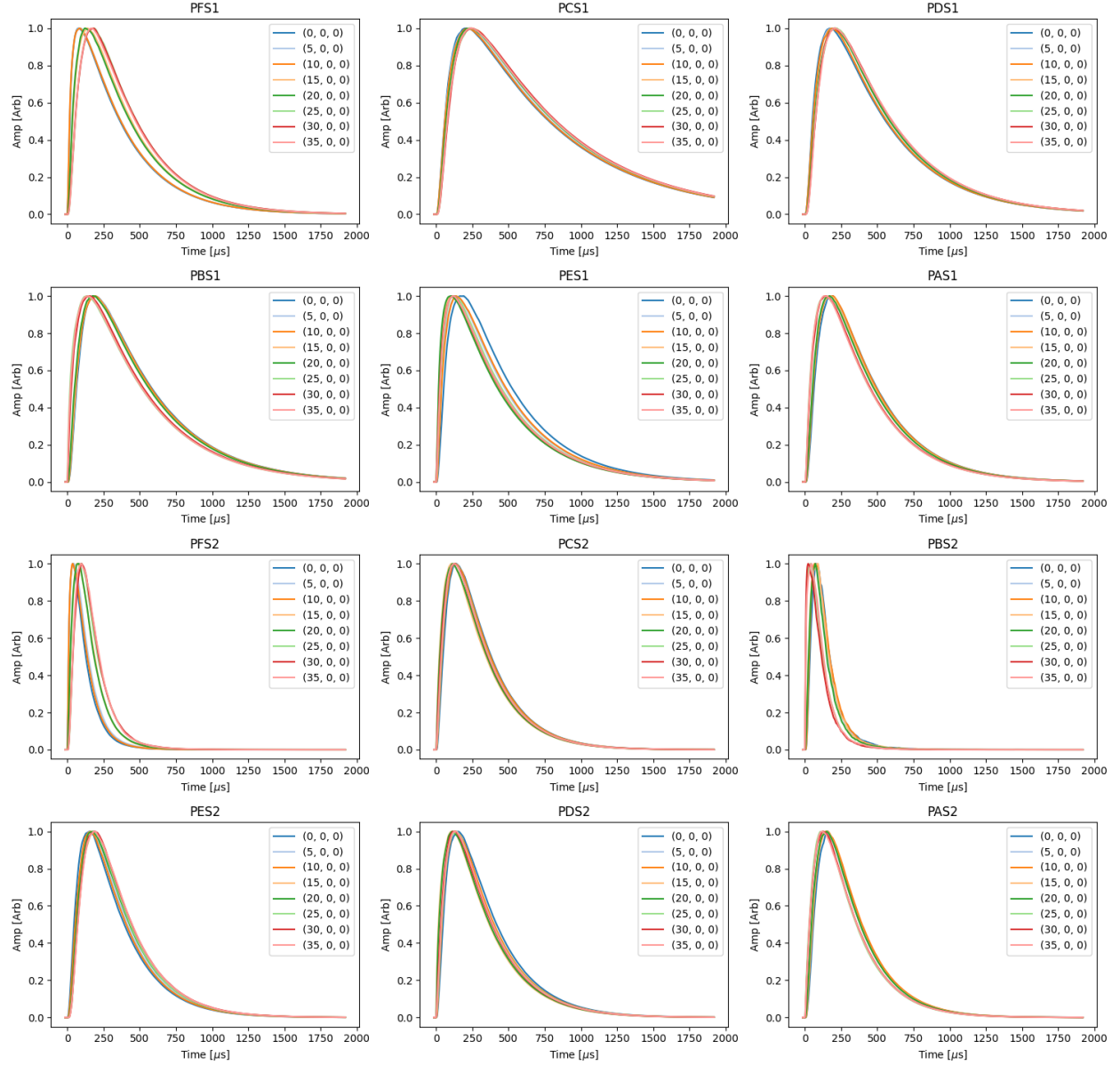


Figure 2-11. Simulated detector traces for each channel, normalized, for the same radial positions as Figure 2-10. The pulse shape carries the position dependence of the phonon signal through the sensor response.

Were the carriers to drift in straight lines along the field, the partition would be a simple linear function of z , identical in any crystal. The valley structure makes it otherwise, and it does so differently in silicon and germanium. Figure 2-12 shows the difference between the energy collected at the two faces against event depth, for 10,000 events of 500 eV each distributed uniformly through the crystal under a 100 V bias, in silicon and in germanium, both cut with the field along $\langle 001 \rangle$. The bulk events form the tight central band, where the difference is at most a few percent of the per-face energy. The points scattered around the band are events near the sidewalls, where a carrier reaches a wall and recombines after picking up only partial NTL amplification, so the two faces receive less energy and in unequal measure. Setting those aside, the bands themselves are not the same in the two materials. In germanium the asymmetry runs nearly linearly through zero at the crystal centre. In silicon it is curved, peaking near the centre rather than vanishing there.

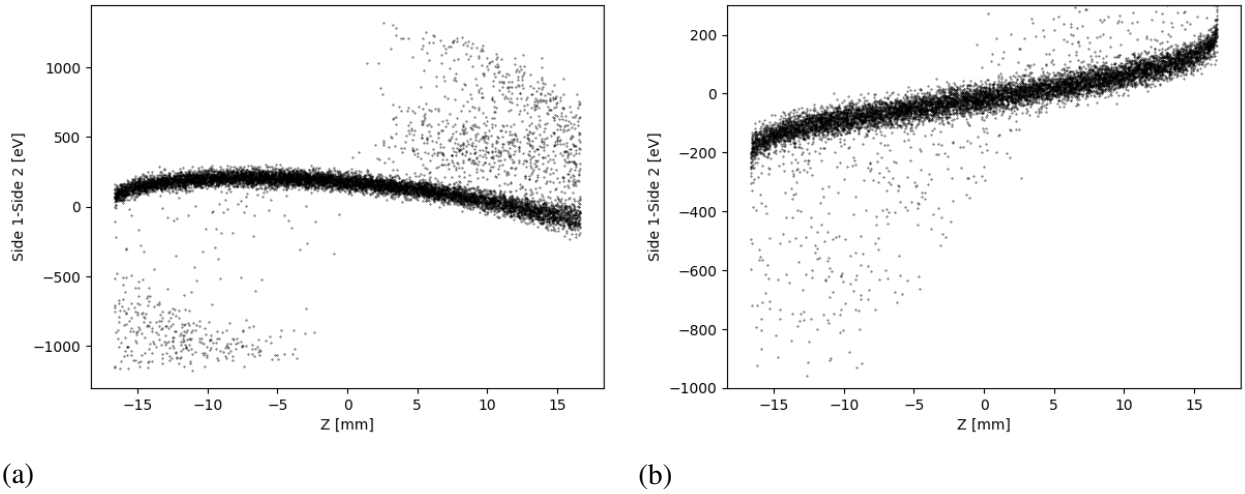


Figure 2-12. Difference between the phonon energy collected at the two faces as a function of event depth, with the field along $\langle 001 \rangle$. (a) Silicon, where the band is curved and does not vanish at the centre. (b) Germanium, where it runs nearly linearly through zero.

The difference is rooted in the contrast between the two carriers. A hole has approximately a scalar effective mass and drifts straight along the field, delivering its ballistic phonons to the near face. An electron carries the anisotropic mass tensor of its valley and drifts obliquely, so its NTL emission is laid down along a slanted path rather than straight down the field. The face asymmetry

is therefore set by the electrons, and how it depends on depth is fixed by how the valleys are oriented relative to the field. With the field along $\langle 001 \rangle$, silicon's six $\langle 100 \rangle$ valleys split into two along the field and four perpendicular to it, so the electrons do not all drift the same way. Germanium's four $\langle 111 \rangle$ valleys are instead all oblique to the field at the same angle, $\arccos(1/\sqrt{3}) \approx 54.7^\circ$, and every electron drifts identically. The curvature in silicon against the near linearity in germanium follows from that difference. The magnitude is modest, a few percent of the per-face energy at most, largest for events near a face and smallest near the centre.

CHAPTER 3

TES ELECTROTHERMAL DYNAMICS

A transition-edge sensor is a superconducting film operated within its resistive phase transition, where the resistance depends steeply on temperature and bias current. The resistance surface $R(T, I)$ couples an electrical circuit to a thermal circuit, and the interplay between them, mediated by the Joule power that heats the film and the voltage bias that regulates it, produces the negative electrothermal feedback that gives the TES its speed, linearity, and self-regulating operating point.

3.1 The Resistive Phase Transition

The steep temperature dependence of resistance near the superconducting critical temperature T_c makes superconducting thin films natural candidates for sensitive thermometry. A film biased within its transition, where the resistance varies rapidly with temperature, converts a small thermal excursion into a measurable change in current, with a sensitivity that depends on both the temperature T and the bias current I . These two state variables parameterize a resistance surface $R(T, I)$ whose shape encodes the physics of the transition and determines the performance of the sensor.

In the ideal case for a non-thin superconducting material, the transition is infinitely sharp. Two quantities define the boundary of the superconducting state. The critical temperature T_c is the temperature at which the film transitions at zero bias current, and the critical current $I_c(0)$ is the current at which the film is driven normal at zero temperature. At finite current the critical temperature is suppressed below T_c to an effective value $T_c^{\text{eff}}(I) < T_c$; equivalently, at finite temperature the critical current is reduced below $I_c(0)$ to $I_c(T) < I_c(0)$. These are two readings of the same phase boundary in the (T, I) plane. The resistance surface is the two-dimensional step function

$$R(T, I) = R_n \Theta\left[T - T_c^{\text{eff}}(I)\right], \quad (3-1)$$

where Θ is the Heaviside function. Below $T_c^{\text{eff}}(I)$ the resistance vanishes; above it the resistance equals the normal-state resistance R_n . The transition is a vertical cliff separating two flat plateaus (Figure 3-1a).

The sensitivity of the sensor at any point on the resistance surface is characterized by the logarithmic derivatives

$$\begin{aligned}\alpha(T, I) &\equiv \left. \frac{\partial \log R}{\partial \log T} \right|_I = \frac{T}{R} \left. \frac{\partial R}{\partial T} \right|_I, \\ \beta(T, I) &\equiv \left. \frac{\partial \log R}{\partial \log I} \right|_T = \frac{I}{R} \left. \frac{\partial R}{\partial I} \right|_T,\end{aligned}\tag{3-2}$$

which measure the fractional change in resistance per fractional change in temperature (at fixed current) and per fractional change in current (at fixed temperature), respectively. The convention of treating α and β as separate partial derivatives follows Irwin and Hilton [45], and is natural because temperature and current couple to the TES through distinct physical pathways, temperature through the thermal circuit and current through the electrical circuit. Under electrothermal feedback these pathways are coupled. A temperature change modifies the resistance, which changes the current, which changes the Joule power, which feeds back on the temperature. The feedback loop determines the trajectory on that surface, not the surface's shape.

Evaluating on the ideal surface of Equation 3-1, $\partial R / \partial T = R_n \delta[T - T_c^{\text{eff}}(I)]$, so

$$\alpha_{\text{ideal}} = \frac{T}{R_n \Theta[T - T_c^{\text{eff}}]} \cdot R_n \delta[T - T_c^{\text{eff}}] = T \frac{\delta[T - T_c^{\text{eff}}]}{\Theta[T - T_c^{\text{eff}}]},\tag{3-3}$$

and similarly for β . The expression is dimensionally consistent because the delta function $\delta[T - T_c^{\text{eff}}]$ carries units of K^{-1} (the inverse of its argument), which the prefactor T cancels, leaving α dimensionless as required. On the normal plateau ($T > T_c^{\text{eff}}$) the delta vanishes and $\alpha = \beta = 0$, so the resistance is flat and the sensor is blind. On the superconducting plateau ($T < T_c^{\text{eff}}$) the denominator $R = 0$ makes α and β undefined. At the phase boundary itself the ratio δ/Θ is formally ill-defined because the delta is nonzero precisely where the step is discontinuous.

The sensitivity is either zero, undefined, or singular; nowhere on the ideal surface does it take a finite, useful value. A calorimetric sensor requires the cliff to soften into a finite-width ramp where α is large but finite, providing the steep temperature-to-resistance conversion that underlies the sensor's sensitivity, and where β is finite so that the operating point can be chosen continuously.

3.1.1 The Phase Boundary

The broadened surface inherits its shape from the ideal boundary it replaces. On the ideal surface, the delta function in Equation 3-3 lives on the curve $T = T_c^{\text{eff}}(I)$, the locus of points where superconductivity is destroyed. The functional form of this curve determines the geometry of the cliff, and when that cliff is broadened into a ramp, it determines the geometry of the ramp as well. Every property of the transition surface traces back to this single curve.

The boundary can be written in two equivalent forms. Defining the reduced variables $t \equiv T/T_c$ and $j \equiv I/I_c(0)$, normalized by the zero-current critical temperature and zero-temperature critical current respectively,

$$\frac{I_c(T)}{I_c(0)} = g(t), \quad \frac{T_c^{\text{eff}}(I)}{T_c} = g^{-1}(j), \quad (3-4)$$

where $g(t)$ gives the effective critical current at finite temperature as a fraction of $I_c(0)$. The exact form of $g(t)$ depends on the underlying microscopic model, yet, all models satisfy $g(0) = 1$ and $g(1) = 0$; they differ in the shape of the curve between these endpoints. Two routes lead to predictions for $g(t)$. In the thermodynamic route, the transport current destroys superconductivity when its self-field reaches the thermodynamic critical field $H_c(T)$ at the conductor surface (the Silsbee hypothesis), and in the kinetic route, the superfluid momentum suppresses the order parameter to zero at the depairing limit. Three models span these routes [46, 47, 48, 49, 50],

$$\text{Gorter-Casimir:} \quad g(t) = 1 - t^2, \quad g^{-1}(j) = \sqrt{1 - j}, \quad (3-5)$$

$$\text{BCS:} \quad g(t) = H_c^{\text{BCS}}(t)/H_c(0), \quad g^{-1}(j) \text{ (numerical)}, \quad (3-6)$$

$$\text{Ginzburg-Landau:} \quad g(t) = (1 - t)^{3/2}, \quad g^{-1}(j) = 1 - j^{2/3}. \quad (3-7)$$

The Gorter–Casimir and BCS boundaries are concave and vanish linearly as $t \rightarrow 1$, since $1 - t^2 \approx 2(1 - t)$ near the critical temperature. The Ginzburg–Landau boundary is convex and vanishes as $(1 - t)^{3/2}$, lying below both across the full range. It is strictly a near- T_c expansion, so carrying it across the entire boundary extrapolates it beyond its formal range of validity, and we adopt it as a parameterization of the boundary rather than as a quantitative description over the full range.

3.1.2 Transition Broadening and Sensitivity Parameters

SuperCDMS TES films are tungsten, approximately 40 nm thick, $2.4 \mu\text{m}$ wide, and $208 \mu\text{m}$ long [51]. The films crystallize in a mixture of α -W (bcc, $T_c \approx 15 \text{ mK}$) and metastable β -W (A15, $T_c \sim 1\text{--}4 \text{ K}$), with the critical temperature controlled by the relative phase fractions and tuned into the 40–80 mK operating range.

The transition in these films extends over a finite temperature width $\Delta T_c \equiv T_w$. Cabrera [52] introduced the ideal resistance surface and its broadening through a smooth interpolation. Irwin et al. [53] proposed a two-fluid model for the TES current, separating the supercurrent and quasiparticle contributions, and Bennett et al. [54] derived the resulting $R(T, I)$ surface using the Ginzburg–Landau critical current, obtaining predictions for α and β that agree with measured data from x-ray and gamma-ray TES devices. Here we follow the Cabrera parameterization used in the SuperCDMS simulation framework TESSim [55], replacing the step function in Equation 3-1 with

$$\frac{R(T, I)}{R_n} = \frac{1}{2} [1 + \tanh \mathcal{A}] , \quad (3-8)$$

where

$$\mathcal{A}(T, I) \equiv \frac{T - T_c g^{-1}(j)}{T_w} \quad (3-9)$$

is the signed distance of the operating point from the phase boundary in units of the transition width, with $j = I/I_c(0)$. The 10%–90% width is $\Delta T_{10-90} \approx 2.2 T_w$, and the ideal step function is recovered as $T_w \rightarrow 0$.

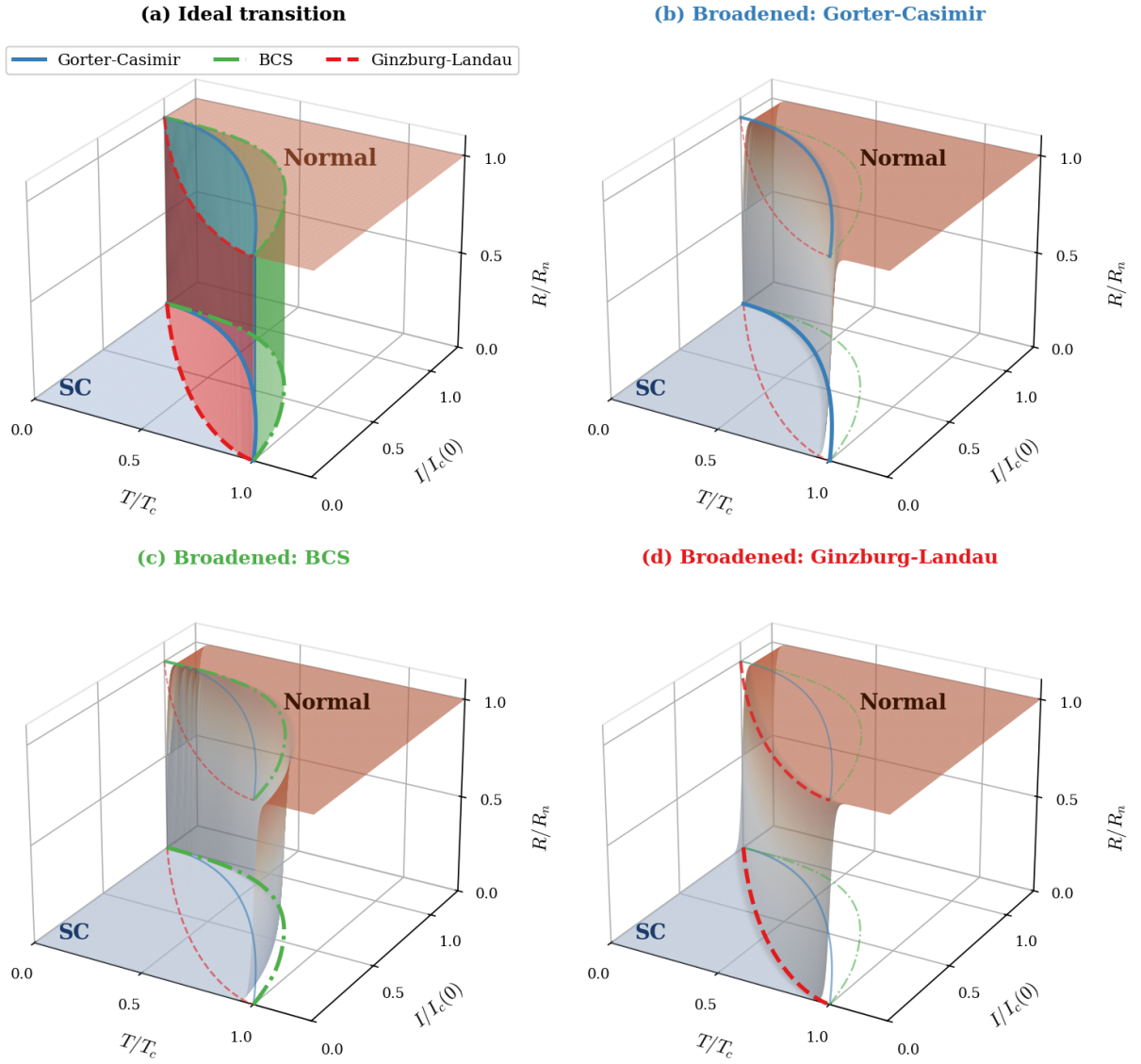


Figure 3-1. (a) Ideal transition with GC (blue), BCS (green), and GL (red) phase boundaries. (b)–(d) Broadened surfaces for each model with $T_w = 0.02 T_c$.

Figure 3-1(a) shows the three phase boundaries on the ideal surface. The concave GC and BCS boundaries differ in curvature, with BCS enclosing a larger region along the current axis and GC extending further along the temperature axis. The convex GL boundary lies inside both.

Figures 3-1(b)–(d) show how these differences propagate into the broadened surfaces.

The sensitivities α and β (Equation 3-2) are now finite throughout the transition. Irwin and Hilton note that some treatments define a single sensitivity parameter rather than the separate α and β [45]. The reason is that R depends on T and I only through the signed distance $\mathcal{A}$ from the phase boundary, so its response to any perturbation is governed by a single quantity, how fast the resistance changes with that distance. Differentiating Equation 3-8,

$$\frac{dR}{d\mathcal{A}} = \frac{R_n}{2} \operatorname{sech}^2 \mathcal{A}. \quad (3-10)$$

This can be rewritten entirely in terms of the fractional resistance $f_{R_n} \equiv R/R_n$. From

$f_{R_n} = \frac{1}{2}(1 + \tanh \mathcal{A})$, inverting gives $\tanh \mathcal{A} = 2f_{R_n} - 1$, so

$$1 + \tanh \mathcal{A} = 2 f_{R_n}, \quad 1 - \tanh \mathcal{A} = 2(1 - f_{R_n}). \quad (3-11)$$

Since $\operatorname{sech}^2 \mathcal{A} = 1 - \tanh^2 \mathcal{A} = (1 + \tanh \mathcal{A})(1 - \tanh \mathcal{A})$,

$$\operatorname{sech}^2 \mathcal{A} = 4 f_{R_n} (1 - f_{R_n}), \quad (3-12)$$

and

$$\frac{dR}{d\mathcal{A}} = 2 R (1 - f_{R_n}). \quad (3-13)$$

The derivative $dR/d\mathcal{A}$ measures how sensitive the resistance is to motion toward or away from the phase boundary. It is a universal function of the bias point alone, independent of the phase boundary model or the transition width. It peaks at $f_{R_n} = 1/2$ (the center of the transition, where $R = R_n/2$) and vanishes at both endpoints $f_{R_n} = 0$ and $f_{R_n} = 1$, reflecting the fact that the ramp is steepest at its midpoint and flattens into the superconducting and normal plateaus.

The two sensitivity parameters then follow from the chain rule as projections of this common response onto the temperature and current axes,

$$\alpha = \frac{T}{R} \frac{dR}{d\mathcal{A}} \frac{\partial \mathcal{A}}{\partial T} \Big|_I, \quad \beta = \frac{I}{R} \frac{dR}{d\mathcal{A}} \frac{\partial \mathcal{A}}{\partial I} \Big|_T. \quad (3-14)$$

Each has three factors. The ratio T/R (or I/R) is the logarithmic weight from the definition. The derivative $dR/d\mathcal{A}$ is the common ramp projection onto the (T, I) plane, the single parameter that Irwin and Hilton refer to. The partial derivative $\partial \mathcal{A}/\partial T$ (or $\partial \mathcal{A}/\partial I$) projects the distance $\mathcal{A}$ onto the corresponding axis, selecting how much of a temperature or current perturbation translates into motion across the ramp.

Substituting Equation 3-13 into Equation 3-14, the factor R cancels the $1/R$ in the logarithmic weight,

$$\alpha = 2 T (1 - f_{R_n}) \frac{\partial \mathcal{A}}{\partial T} \Big|_I, \quad \beta = 2 I (1 - f_{R_n}) \frac{\partial \mathcal{A}}{\partial I} \Big|_T. \quad (3-15)$$

The factor $(1 - f_{R_n})$ is the fractional distance from the normal state. It equals unity at the bottom of the transition ($R = 0$) and vanishes at the top ($R = R_n$). All of the phase boundary geometry and the transition width enter through the partial derivatives of $\mathcal{A}$.

Assuming a constant transition width T_w along the boundary, the partial derivatives of $\mathcal{A}(T, I) = (T - T_c g^{-1}(j))/T_w$ are

$$\frac{\partial \mathcal{A}}{\partial T} \Big|_I = \frac{1}{T_w}, \quad \frac{\partial \mathcal{A}}{\partial I} \Big|_T = -\frac{T_c}{T_w I_c(0)} (g^{-1})'(j), \quad (3-16)$$

where the minus sign arises because increasing the current lowers T_c^{eff} . Substituting into Equation 3-15,

$$\alpha(T, I) = \frac{2T}{T_w} (1 - f_{R_n}), \quad (3-17)$$

$$\beta(T, I) = -\frac{2T_c}{T_w} j (g^{-1})'(j) (1 - f_{R_n}). \quad (3-18)$$

In a real film the transition width is not constant. Proximity effects and nonuniform current flow produce a T_w that varies both along the phase boundary and across the ramp at each point. With $T_w = T_w(T, I)$, the variable $\mathcal{A} = (T - T_c g^{-1}(j))/T_w(T, I)$ is a ratio of two functions of (T, I) . Applying the quotient rule with numerator $N = T - T_c g^{-1}(j)$ and denominator T_w , and recognizing that $N = \mathcal{A} T_w$, the partial derivatives acquire corrections from differentiating the denominator,

$$\left. \frac{\partial \mathcal{A}}{\partial T} \right|_I = \frac{1}{T_w} - \frac{\mathcal{A}}{T_w} \left. \frac{\partial T_w}{\partial T} \right|_I, \quad (3-19)$$

$$\left. \frac{\partial \mathcal{A}}{\partial I} \right|_T = -\frac{T_c}{T_w I_c(0)} (g^{-1})'(j) - \frac{\mathcal{A}}{T_w} \left. \frac{\partial T_w}{\partial I} \right|_T. \quad (3-20)$$

The first term in each line is the constant- T_w result (Equation 3-16). The corrections are proportional to $\mathcal{A}$ and vanish at the center of the transition ($\mathcal{A} = 0$, where $f_{R_n} = 1/2$), regardless of how T_w varies. Away from the center, a temperature- or current-dependent T_w tilts the ramp and produces asymmetric sensitivity even though the underlying tanh model is symmetric. The constant- T_w approximation is adopted for the remainder of this thesis.

The current sensitivity β has an explicit dependence on the phase boundary model through $(g^{-1})'(j)$,

$$\text{Gorter-Casimir:} \quad (g^{-1})'(j) = -\frac{1}{2\sqrt{1-j}}, \quad (3-21)$$

$$\text{BCS:} \quad (g^{-1})'(j) \text{ (numerical)}, \quad (3-22)$$

$$\text{Ginzburg-Landau:} \quad (g^{-1})'(j) = -\frac{2}{3} j^{-1/3}. \quad (3-23)$$

The concave GC and BCS boundaries produce $(g^{-1})'$ that diverges as $j \rightarrow 1$; the convex GL boundary produces a bounded derivative.

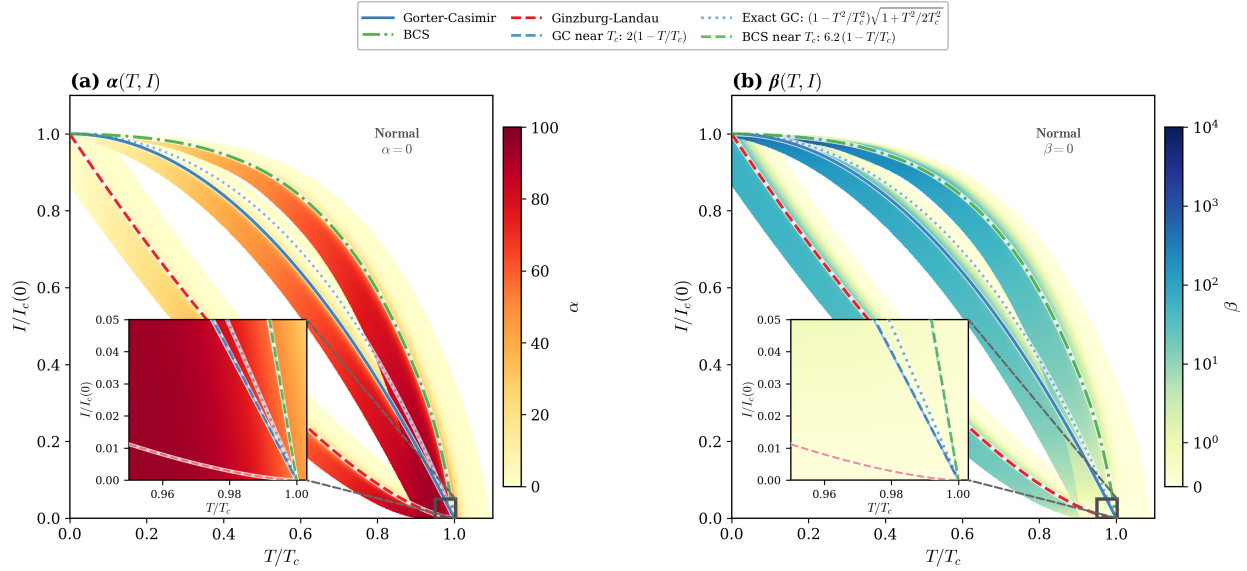


Figure 3-2. (a) $\alpha(T, I)$, linear scale. (b) $\beta(T, I)$, symlog scale. All three models overlaid with the exact GC boundary (dotted blue).

Figure 3-2 reveals a complementary structure that is visible directly in the phase boundary geometry. Near T_c (right side of the plot), the GC and BCS boundaries are nearly vertical and a small temperature fluctuation at fixed current sweeps the operating point across the narrow ramp, producing a large α . Near $I_c(0)$ (top of the plot), the boundaries flatten and $(g^{-1})'(j)$ diverges. Since β is proportional to $(g^{-1})'$, it inherits this divergence, concentrating its weight at the low-temperature, high-current end of the boundary. The two parameters partition the transition

band. α dominates where the boundary is steep, β dominates where it is flat, and together they are complementary projections of the gradient of $R(T, I)$ onto the temperature and current axes.

3.2 Electrothermal Feedback Architecture

The resistance surface $R(T, I)$ is a static property of the film. It describes which states are accessible, not how the device moves between them. The dynamics arise when the film is embedded in two circuits, one electrical and one thermal, each introducing its own inertia. The input inductor L in the readout circuit resists changes in current, giving the electrical branch a characteristic response time. The thermal boundary between the film and the cold bath resists changes in temperature, giving the thermal branch its own. These two inertial timescales, together with the resistance surface that couples them, determine the complete electrothermal dynamics of the sensor [45].

3.2.1 The Voltage-Biased TES Circuit

A TES biased with a constant current source is unstable. If a thermal fluctuation raises the film temperature, the resistance increases, and since the current is fixed, the Joule power $P_J = I^2 R$ increases with it. The additional heating drives the temperature higher still, producing a positive feedback loop that ends with the film latched in the normal state. Under voltage bias the situation reverses. A temperature increase raises the resistance, but now the current decreases as $I = V/R$, so the Joule power $P_J = V^2/R$ decreases. The reduced heating allows the film to cool back toward its equilibrium temperature. This negative electrothermal feedback is the essential operating principle of the TES calorimeter [56].

Voltage biasing a TES channel requires maintaining an approximately constant voltage across the sensor while it operates within its transition. A small shunt resistor R_{sh} connected in parallel with the TES branch achieves this. Because $R_{\text{sh}} \ll R_p + R_{\text{ch}}$, the shunt carries most of the bias current and only a small fraction flows through the sensor. The voltage at the junction is then set by the current through the shunt, $V_{\text{node}} \approx I_b R_{\text{sh}}$, and is nearly independent of the TES resistance. Changes in R_{ch} redistribute a small amount of current between the two branches but barely perturb the node voltage. The voltage across the TES departs from the ideal value $I_b R_{\text{sh}}$ by

a fractional correction R_l/R_{ch} ,

$$V_{\text{ch}} = I_b R_{\text{sh}} \left(1 - \frac{R_l}{R_{\text{ch}}} + \dots \right), \quad (3-24)$$

so the voltage bias approximation holds to the extent that $R_l \ll R_{\text{ch}}$. In the SuperCDMS readout, $R_{\text{sh}} \approx 5 - 20 \text{ m}\Omega$, while $R_{\text{ch}} \sim 100\text{--}300 \text{ m}\Omega$.

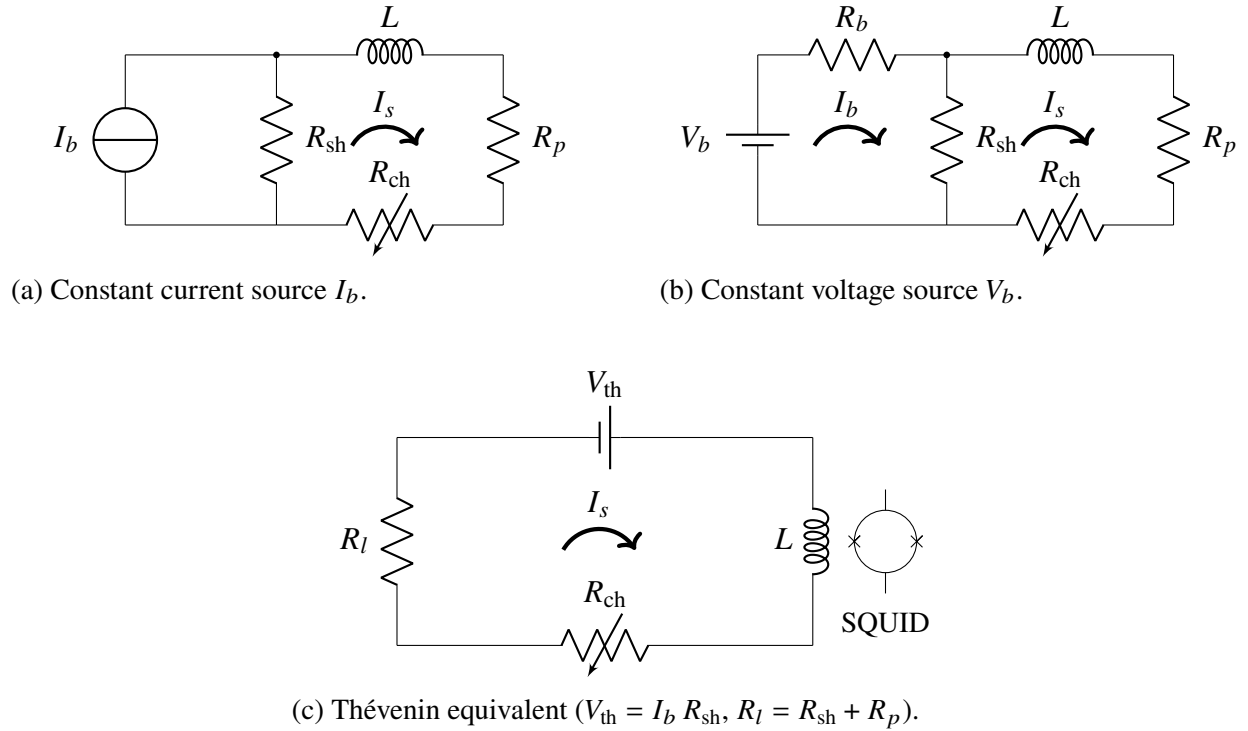


Figure 3-3. (a) Voltage bias achieved with a constant current source I_b . (b) Voltage bias achieved with a constant voltage source V_b to (a) when R_b is very large. (c) Thévenin equivalent with $V_{\text{th}} \approx I_b R_{\text{sh}}$ and $R_l \approx R_{\text{sh}} + R_p$ [55].

To derive Equation 3-24, consider the current-source circuit (Figure 3-3a). The channel resistance $R_{\text{ch}} \equiv R(T, I_s)$ depends on the film temperature and the sensor current through the resistance surface developed in Section 3.1, and is the parallel combination of the N individual TES that make up the channel. All N TES sit between the same two nodes, so current divides inversely with resistance,

$$I_i = \frac{R_{\text{ch}}}{R_i} I_s. \quad (3-25)$$

For N identical TES at the same operating point this reduces to $I_i = I_s/N$. Localized energy depositions can drive different TES within a channel to different operating points; the resulting intra-channel inhomogeneity broadens the effective channel-level $R(T, I)$. In the equations that follow, R denotes $R(T, I_s)$ with the arguments suppressed for brevity. The current source fixes I_b , and the shunt carries $I_b - I_s$. Kirchhoff's voltage law around the single TES loop gives

$$L \frac{dI_s}{dt} = (I_b - I_s) R_{\text{sh}} - I_s (R_p + R). \quad (3-26)$$

The voltage-source circuit (Figure 3-3b) has two current loops. The sensor loop equation is identical to Equation 3-26, while the bias loop gives

$$V_b = I_b R_b + (I_b - I_s) R_{\text{sh}}. \quad (3-27)$$

Solving for I_b ,

$$I_b = \frac{V_b + I_s R_{\text{sh}}}{R_b + R_{\text{sh}}}, \quad (3-28)$$

and substituting into the sensor loop eliminates I_b . The shunt voltage becomes

$$(I_b - I_s) R_{\text{sh}} = \frac{(V_b - I_s R_b) R_{\text{sh}}}{R_b + R_{\text{sh}}} = \frac{V_b R_{\text{sh}}}{R_b + R_{\text{sh}}} - I_s \frac{R_b R_{\text{sh}}}{R_b + R_{\text{sh}}}, \quad (3-29)$$

so the sensor current equation for the voltage-source circuit is

$$L \frac{dI_s}{dt} = \underbrace{\frac{V_b R_{\text{sh}}}{R_b + R_{\text{sh}}}}_{V_{\text{th}}} - I_s \left(\underbrace{\frac{R_b R_{\text{sh}}}{R_b + R_{\text{sh}}}}_{R_l} + R_p + R \right). \quad (3-30)$$

This defines the Thévenin equivalent quantities exactly,

$$V_{\text{th}} = \frac{V_b R_{\text{sh}}}{R_b + R_{\text{sh}}}, \quad R_l = \frac{R_b R_{\text{sh}}}{R_b + R_{\text{sh}}} + R_p. \quad (3-31)$$

When $R_b \gg R_{\text{sh}}$, these reduce to $V_{\text{th}} \approx I_b R_{\text{sh}}$ and $R_l \approx R_{\text{sh}} + R_p$, recovering the current-source result with the identification $I_b = V_b/R_b$.

Both circuits therefore share the same Thévenin equivalent (Figure 3-3c),

$$L \frac{dI_s}{dt} = V_{\text{th}} - I_s (R_l + R), \quad (3-32)$$

the electrical equation of motion for the sensor current.

Returning to the current-source form (Equation 3-26) to verify the voltage bias condition, in steady state ($dI_s/dt = 0$), the voltage across the TES is

$$I_s R = (I_b - I_s) R_{\text{sh}} - I_s R_p = I_b R_{\text{sh}} - I_s R_l. \quad (3-33)$$

Solving for I_s ,

$$I_s = \frac{I_b R_{\text{sh}}}{R_l + R}, \quad (3-34)$$

which maps the externally controlled bias current I_b to the sensor current I_s through the TES resistance. Substituting back into $V_{\text{ch}} = I_s R$,

$$V_{\text{ch}} = I_b R_{\text{sh}} \frac{R}{R_l + R} = \frac{I_b R_{\text{sh}}}{1 + R_l/R} = I_b R_{\text{sh}} \left(1 - \frac{R_l}{R} + \dots \right), \quad (3-35)$$

confirming Equation 3-24. The steady-state operating point current is

$$I_0 = \frac{V_{\text{th}}}{R_l + R_0}, \quad (3-36)$$

where $R_0 \equiv R(T_0, I_0)$, and the Joule power dissipated in the TES is

$$P_{J,0} = I_0^2 R_0 = \frac{V_{\text{th}}^2 R_0}{(R_l + R_0)^2}. \quad (3-37)$$

3.2.1.1 Electrical Parameter Extraction from $I_b I_s$

The inductance L in the Thévenin equivalent (Figure 3-3c) introduces inertia into the electrical branch. It resists instantaneous changes in current and imposes a finite response time. To isolate this timescale, hold the resistance at its steady-state value R_0 , decoupling the electrical dynamics from the thermal. Writing $I_s = I_0 + \delta I$ in the Thévenin ODE (Equation 3-32) and noting that the constant terms cancel against the steady-state condition $V_{th} = I_0(R_l + R_0)$,

$$L \frac{d(\delta I)}{dt} = -\delta I (R_l + R_0). \quad (3-38)$$

The solution is $\delta I(t) = \delta I(0) e^{-t/\tau_{el}^{(0)}}$, where

$$\tau_{el}^{(0)} = \frac{L}{R_l + R_0} \quad (3-39)$$

is the bare electrical time constant. This is the frozen- R limit, the case in which the thermal state responds infinitely slowly compared to the current so that R is effectively constant on the timescale of the electrical relaxation. In a real TES the resistance changes through $R(T, I_s)$ as the current and temperature evolve together, and the coupled dynamics produce effective time constants that differ from the bare $\tau_{el}^{(0)}$.

The value of $\tau_{el}^{(0)}$ sets the bandwidth of the current response. For the signal information encoded in $I_s(t)$ to pass through the readout without distortion, the corresponding electrical bandwidth $1/\tau_{el}^{(0)}$ must fall within the passband of the SQUID amplifier.

The steady-state relation $I_s(I_b)$ (Equation 3-34) is directly measurable by sweeping I_b and recording I_s with the SQUID. For this measurement to trace out the equilibrium curve, the sweep must be slow enough that $dI_s/dt \approx 0$ at each step, not only on the electrical timescale $\tau_{el}^{(0)}$, but also on the other timescales in the system associated with the temperature dependence of $R(T, I_s)$. The resulting I_b - I_s characteristic has three branches, a superconducting branch ($R = 0$) where $I_s = I_b R_{sh}/R_l$, a normal branch ($R = R_n$) where $I_s = I_b R_{sh}/(R_l + R_n)$, and a transition region

connecting them. The slopes of the superconducting and normal branches determine R_p and R_n directly given R_{sh} , and at each point in the transition $R_0 = R_{sh}(I_b/I_s - 1) - R_p$.

The I_b-I_s characteristic is measured by stepping the QET bias current through a sequence of values spanning the superconducting, transition, and normal regions. At each bias point, the detector is held at a fixed I_b while a set of triggered events is acquired. Events containing energy depositions are rejected, and the mean baseline current $\langle I_s \rangle$ over the remaining traces gives the quiescent sensor current at that operating point. This procedure ensures that each point on the I_b-I_s curve represents a true steady-state operating point, free of signal contamination. A small current offset arising from the SQUID feedback loop is removed by enforcing $I_s \rightarrow 0$ as $I_b \rightarrow 0$ separately on the superconducting and normal branches.

Figure 3-4 shows the corrected I_b-I_s characteristic for a single HVeV channel. The superconducting branch (blue) and normal branch (red) are each fit to a line through the origin; the slopes give $R_p = 11.5 \text{ m}\Omega$ and $R_n = 325 \text{ m}\Omega$ with $R_{sh} = 18.6 \text{ m}\Omega$. The transition region (green) connects the two branches over a narrow range in I_b .

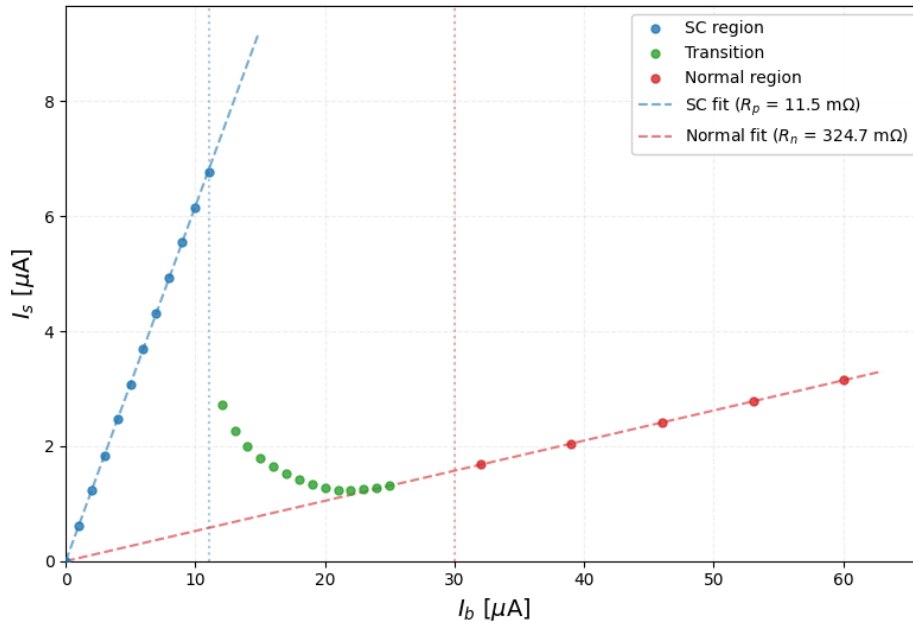


Figure 3-4. I_b-I_s characteristic for a single HVeV channel. Dashed lines are linear fits to the superconducting (blue) and normal (red) branches, yielding R_p and R_n .

From this measurement, the TES resistance and Joule power at each bias point follow directly from Equations 3-34 and 3-37. Figure 3-5 shows these derived quantities as functions of both I_b (the controlled variable) and I_s (the measured variable). The resistance rises steeply through the transition and saturates at R_n . The Joule power vanishes in the superconducting state, remains nearly constant through the transition, a consequence of the electrothermal feedback regulating the operating point, and grows quadratically in the normal state where $P_J \approx V_{th}^2/R_n$.

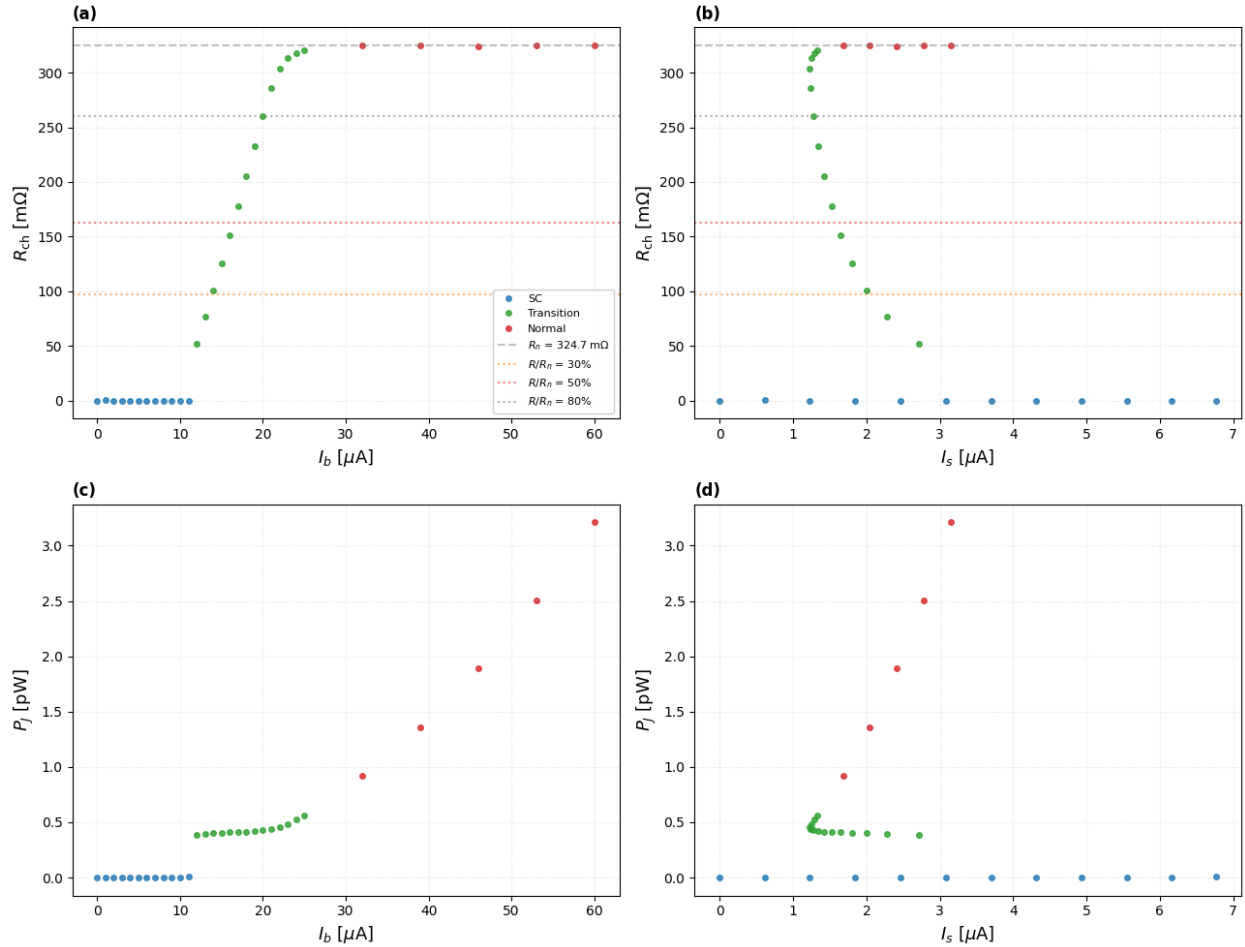


Figure 3-5. Quantities derived from the I_b – I_s characteristic. (a) R_{TES} vs I_b . (b) R_{TES} vs I_s . (c) P_J vs I_b . (d) P_J vs I_s . Horizontal lines in (a, b) mark $R/R_n = 30\%$, 50% , and 80% .

The four panels of Figure 3-5 are projections of a single trajectory on a surface in the three-dimensional space (I_b, I_s, R) or (I_b, I_s, P_J) .

3.2.1.2 Electrothermal Feedback Stability

The sign of the electrothermal feedback follows from the circuit alone, independent of any thermal model. The Joule power (Equation 3-37) depends on the operating resistance through $P_J = V_{th}^2 R / (R_l + R)^2$. Differentiating with respect to R ,

$$\frac{dP_J}{dR} = V_{th}^2 \frac{R_l - R}{(R_l + R)^3}. \quad (3-40)$$

The numerator changes sign at $R = R_l$. When $R < R_l$, an increase in resistance increases the Joule power, producing positive feedback that drives the film further from equilibrium. When $R > R_l$, an increase in resistance decreases the Joule power, producing the negative feedback that restores the operating point. The condition for stable voltage bias operation is therefore

$$R_0 > R_l. \quad (3-41)$$

The second derivative determines how the feedback strength varies across the stable branch. Differentiating Equation 3-40,

$$\frac{d^2 P_J}{dR^2} = V_{th}^2 \frac{2R - 4R_l}{(R_l + R)^4}. \quad (3-42)$$

This expression is negative for $R < 2R_l$ and positive for $R > 2R_l$. At $R = R_l$ it evaluates to $-V_{th}^2 / (8R_l^3) < 0$, establishing that P_J is maximized there. On the stable branch ($R > R_l$), the sign of the second derivative has a direct physical consequence. Where $d^2 P_J / dR^2 < 0$, between R_l and $2R_l$, the first derivative is becoming more negative with increasing R , so the negative feedback is strengthening as the operating resistance moves away from the turnover. Where $d^2 P_J / dR^2 > 0$, above $2R_l$, the first derivative is becoming less negative, so the feedback is weakening. At the boundary $R = 2R_l$ the second derivative vanishes and the first derivative reaches its most negative value,

$$\left. \frac{dP_J}{dR} \right|_{R=2R_l} = -\frac{V_{th}^2}{27 R_l^2}, \quad (3-43)$$

the strongest negative feedback the circuit can produce at any operating point. A TES biased between R_l and $2R_l$ sits in the strengthening regime, where the restoring force grows with displacement from equilibrium. A TES biased above $2R_l$ sits in the weakening regime, where the restoring force diminishes with further displacement. In the strong voltage bias limit $R_l \ll R_0$ the operating point is deep in the weakening regime, and R_l becomes negligible in the Thévenin equivalent (Figure 3-3c); the TES branch reduces to the inductance L in series with R driven by a fixed voltage V_{th} .

The first two derivatives establish the feedback sign and its variation across the transition. Higher derivatives of the Joule power reveal additional structure in $P_J(R)$, a pattern of zeros at $R = nR_l$ and turnovers at $R = (n + 1)R_l$, that becomes relevant when a large energy deposition drives the resistance across a significant fraction of the transition.

3.2.2 The QET Thermal Architecture

The operating resistance R_0 is not a free parameter. It is fixed self-consistently by the power balance between Joule heating in the W film and the thermal link to the bath, and small excursions of R around R_0 are governed by the second inertial timescale of the coupled system, the thermal relaxation time of the W electrons. Setting up the thermal circuit requires tracking energy from its deposition in the substrate, through the Al fins, into the W electrons, and back out to the bath.

An energy deposition in the substrate launches a population of athermal phonons that propagate to the instrumented faces, where they are collected by the QETs, the quasiparticle-trap-assisted electrothermal-feedback transition-edge sensors that read out the crystal [45]. A QET pairs a tungsten TES with superconducting aluminum fins that extend the collection area well beyond the TES footprint. The phonon transport across the crystal, together with the Cooper-pair breaking, quasiparticle cascade, and trapping that turn an absorbed phonon into heat in the tungsten electron system, is developed in Section 2.4. For the electrothermal dynamics the relevant input is the end product of that chain, a time-dependent power $P_{signal}(t)$ delivered to the W electron system.

The heated W electron system cools through a chain of thermal links visible in the thermal circuit (Figure 3-6). The electrons couple to the W phonon system through the electron-phonon interaction, characterized by the coupling constant Σ_{e-ph}^W ; the W phonons couple across the Kapitza boundary to the substrate; and the substrate cools to the thermal bath. In parallel, the Al fin has its own internal electron-phonon coupling Σ_{e-ph}^{Al} and its own Kapitza link to the substrate. Energy also flows between the Al and W systems through the overlap. Altogether, the circuit contains six thermal bodies, the Al electrons (T_e^{Al}), Al phonons (T_{ph}^{Al}), W electrons (T_e^W), W phonons (T_{ph}^W), substrate (T_{subst}), and the thermal bath (T_b), connected by six thermal conductances, the electron-phonon couplings Σ_{e-ph}^{Al} and Σ_{e-ph}^W within each film, the Kapitza conductances G_{s-Al} and G_{s-W} at the Al-substrate and W-substrate interfaces, the Al-W overlap link G_{Al-W} , and the substrate-to-bath conductance G_{s-b} . In addition to the signal power entering through the Al fins, the W electron system receives the Joule power $P_J = I_s^2 R$ from the bias circuit, the same dissipation whose derivative structure was analyzed in Section 3.2.1.2.

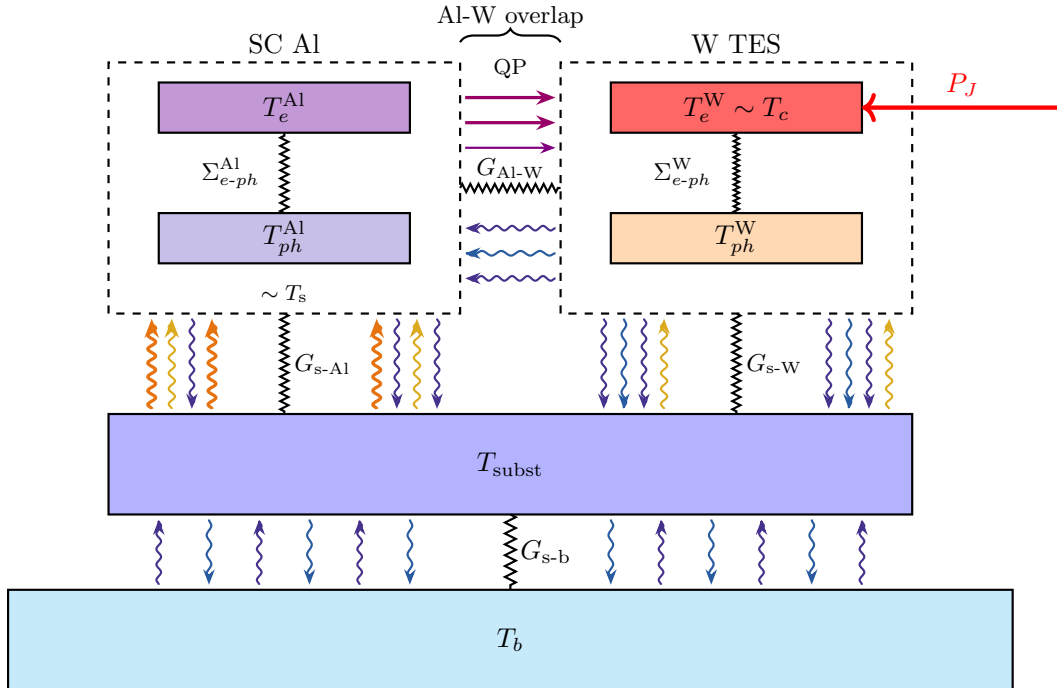


Figure 3-6. Thermal circuit of a single QET, with the bias Joule power P_J deposited in the W electron system.

3.2.2.1 Realistic Thermal Model and ODEs

The thermal circuit of Figure 3-6(b) contains six thermal bodies and six conductance links. Not all of these degrees of freedom are independent on the timescales relevant to the TES response. Several can be eliminated or absorbed on physical grounds, reducing the system to a minimal set of coupled ODEs without discarding any essential dynamics.

The aluminum fins are superconducting well below T_c^{Al} at the operating temperature. The electron and phonon systems in a superconductor are not independent reservoirs. They are bound together through the Cooper pair condensate. There is no meaningful sense in which the Al electrons and Al phonons carry separate temperatures.

Phonon exchange between the QET structure and the substrate, athermal phonons entering the Al, sub-gap phonons leaking back, thermal phonons emitted by the TES, is not a separate dynamical degree of freedom. These phonon flows are part of the signal power profile $P_{\text{signal}}(t)$ as seen by the TES, the time-dependent power delivered to the W electron system after the phonon collection, Cooper pair breaking, quasiparticle cascade, diffusion, and trapping processes have played out. The shape of $P_{\text{signal}}(t)$ encodes the phonon transport physics of the crystal and the QET collection efficiency, but once that profile is specified, from simulation or an analytical model, it enters the TES equation as a source term.

After these reductions, three dynamical variables remain, the sensor current I_s , the W electron temperature T , and the substrate temperature T_{subst} . The coupled ODEs are,

$$L \frac{dI_s}{dt} = V_{\text{th}} - I_s (R_l + R(T, I_s)), \quad (3-44)$$

$$C \frac{dT}{dt} = -\Sigma_W \mathcal{V}_W (T^n - T_{\text{subst}}^n) + I_s^2 R(T, I_s) + P_{\text{signal}}(t) + P_{\text{parasitic}}, \quad (3-45)$$

$$C_{\text{subst}} \frac{dT_{\text{subst}}}{dt} = +\Sigma_W \mathcal{V}_W (T^n - T_{\text{subst}}^n) - \mathcal{K} A_c (T_{\text{subst}}^m - T_b^m), \quad (3-46)$$

$$\frac{dT_b}{dt} = f_b(t). \quad (3-47)$$

Here $C \equiv C_e^W$ is the W electronic heat capacity, $T \equiv T_e^W$ is the W electron temperature, and T_{ph} is the W phonon temperature. The W phonon system tracks T_{subst} through the nearly transparent Kapitza boundary ($T_{\text{ph}} \approx T_{\text{subst}}$) and is absorbed into the substrate degree of freedom. If phonon trapping inside the W film is significant, T_{ph} would need to be promoted to its own dynamical variable, extending the model to three thermal bodies. The term $P_{\text{parasitic}}$ accounts for stray power deposited in the TES electron system that is not captured by the Joule or signal terms, such as electrical pickup from the bias and readout wiring, infrared loading from warm surfaces, and phonon radiation from nearby hot components. This power is always present, is generally not well characterized, and obscures the extraction of both the thermal conductance and the transition temperature from the I_b – I_s characteristic. The bath is usually taken to be at a constant base temperature. However, the fourth equation allows T_b to evolve according to a measured function $f_b(t)$, the mixing chamber temperature oscillations seen in SuperCDMS SNOLAB commissioning, for example.

The thermal conductance from the W electron system to the substrate, evaluated at the operating point, is

$$G = \left. \frac{dP_{\text{cool}}}{dT} \right|_{T_0} = n \Sigma_W \mathcal{V}_W T_0^{n-1}, \quad (3-48)$$

and the natural thermal relaxation time is

$$\tau_{\text{e-ph}} = \frac{C}{G}, \quad (3-49)$$

the counterpart to $\tau_{\text{el}}^{(0)} = L/(R_l + R_0)$ from Section 3.2.1.1. The product $\Sigma_W \mathcal{V}_W$ can be absorbed into a single cooling coefficient K , but we retain the electron-phonon coupling constant Σ_W dependence explicitly as the fundamental quantity governing the TES electron system cooling.

In steady state with no signal and constant bath temperature ($dT/dt = 0$, $dI_s/dt = 0$, $dT_{\text{subst}}/dt = 0$, $f_b = 0$),

$$I_0 = \frac{V_{\text{th}}}{R_l + R_0}, \quad (3-50)$$

$$\Sigma_W \mathcal{V}_W (T_0^n - T_{\text{subst}}^n) = I_0^2 R_0. \quad (3-51)$$

These conditions, together with $R_0 = R(T_0, I_0)$, determine the operating point (T_0, I_0, R_0) self-consistently. The resistance surface couples the electrical and thermal ODEs. A temperature excursion changes R , which changes I_s , which changes P_J , which feeds back into T . This is the electrothermal (ETF) feedback loop.

3.2.2.2 Thermal Parameter Extraction from $I_b I_s$

A single $I_b I_s$ characteristic at one bath temperature constrains the operating resistance and Joule power at every bias point through Equations 3-34 and 3-37, but it leaves the thermal parameters underdetermined. Defining the cooling coefficient $K_{\text{e-ph}} \equiv \Sigma_W \mathcal{V}_W$, the steady-state power balance at fixed fractional resistance $f_{R_n} \equiv R_0/R_n$,

$$P_{J,0} + P_{\text{parasitic}} = K_{\text{e-ph}} (T_0^n - T_{\text{subst}}^n), \quad (3-52)$$

contains four unknowns, T_0 , T_{subst} , $K_{\text{e-ph}}$, and the exponent n , against a single measured power. Sweeping the bath temperature breaks some of this degeneracy. At each T_b the operating point shifts, and the locus of $(T_b, P_{J,0})$ traced out at fixed f_{R_n} has a shape set by the thermal model.

Four simplifications make the extraction tractable. The substrate is assumed to thermalize to the bath through its mounting, $T_{\text{subst}} \approx T_b$; the parasitic loading is set to zero; the exponent n is fixed at the theoretical value $n = 5$ for clean-limit electron-phonon coupling in a normal metal at low temperatures [57]; and the operating point is fixed at $f_{R_n} = 1/2$, where the TES temperature equals the current-suppressed critical temperature, $T_0 = T_c^{\text{eff}}(I_0)$. The detectors are designed so that the operating current at this bias point lies well below the critical current, $I_0 \ll I_{c0}$, so the

current suppression of the transition temperature is small and $T_c^{\text{eff}} \approx T_c$. Equation 3-52 reduces to

$$P_{J,0}(T_b)\big|_{f_{R_n}=1/2} = K_{\text{e-ph}} \left(T_c^5 - T_b^5 \right), \quad (3-53)$$

and a two-parameter fit of $P_{J,0}(T_b)$ at this fractional resistance to Equation 3-53 determines T_c and $K_{\text{e-ph}}$. The thermal conductance evaluated at the operating point follows from Equation 3-48,

$$G(T_c) = 5 K_{\text{e-ph}} T_c^4. \quad (3-54)$$

The procedure is illustrated for the Si HV detector Z2 from the T3 testing at CUTE, which is the same detector (T3P2) deployed in the SuperCDMS SNOLAB experiment. The mixing chamber is swept across bath temperatures from 14 to 44 mK; at each temperature, the system is held until the detector substrate thermalizes to the bath, after which an $I_b I_s$ characteristic is acquired in parallel for every channel. Figure 3-7 shows the resulting family for a single channel.

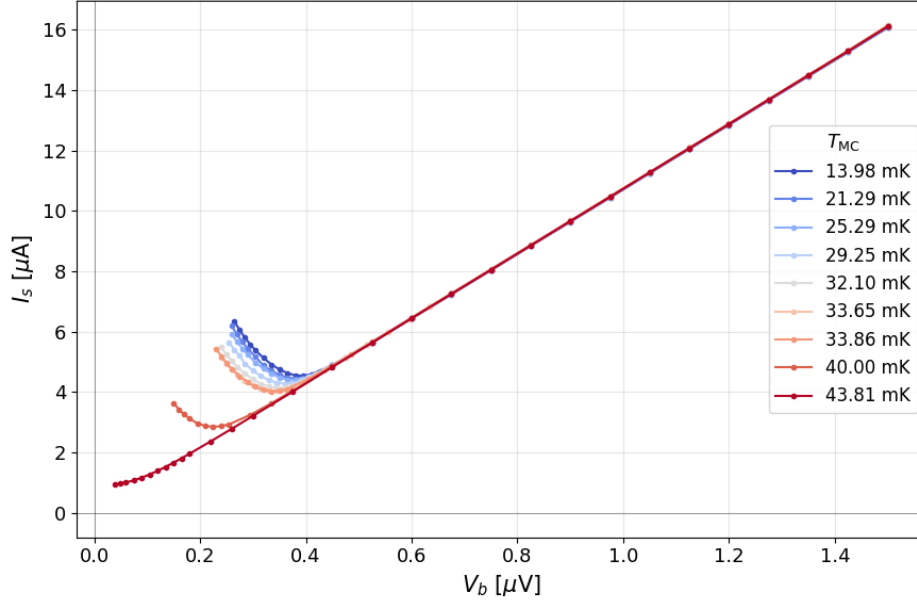


Figure 3-7. $I_b I_s$ characteristics for a single channel of the HV detector Z2 at nine bath temperatures from 14 to 44 mK.

The transition shrinks as T_b rises, since less Joule power is required to hold the channel at the operating temperature. At the highest bath temperature the transition foot has nearly collapsed

onto the normal branch, indicating T_b approaches T_c . The corresponding Joule power and channel resistance are shown in Figure 3-8. The Joule power curves overlap in the normal branch where the channel behaves as a fixed resistor and $P_J \approx V_{th}^2/R_n$ independent of T_b , but separate cleanly in the transition where electrothermal feedback regulates the operating point against the bath load. The resistance traces show the same behavior, with the transition shifting to lower bias voltage as T_b rises.

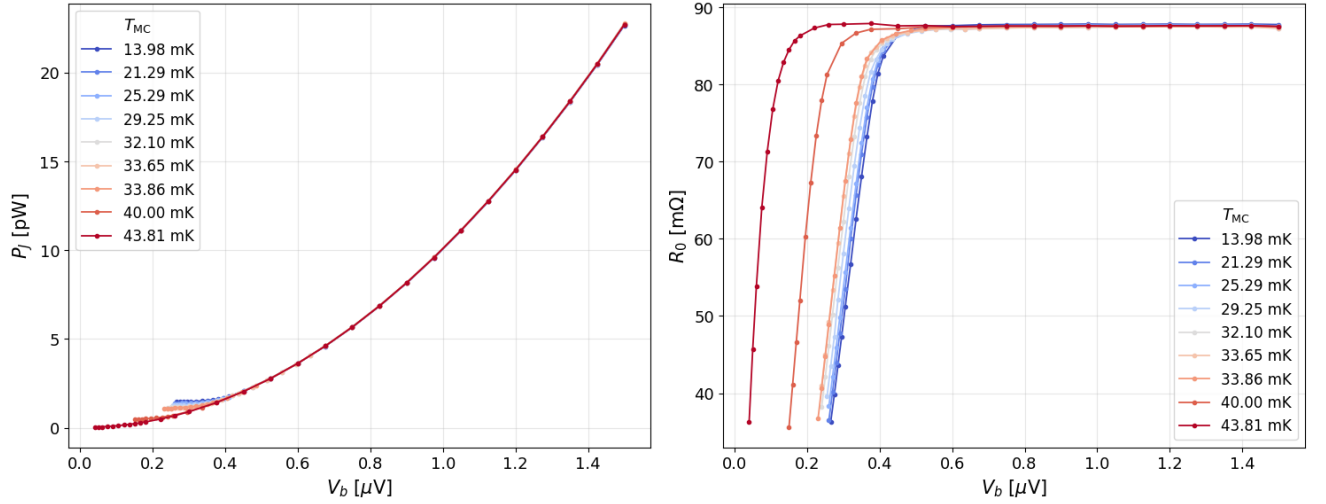


Figure 3-8. Joule power (left) and channel resistance (right) for the same channel and bath temperatures as Figure 3-7.

The Joule power evaluated at $f_{R_n} = 1/2$ is plotted against the bath temperature in Figure 3-9, with the fit to Equation 3-53 overlaid. The left panel uses $P_{parasitic} = 0$ and returns $T_c = 43.98 \pm 0.31$ mK and $K_{e-ph} = (8.87 \pm 0.43) \times 10^{-6}$ W/K⁵, yielding $G(T_c) = 166$ pW/K. The right panel allows $P_{parasitic}$ to float; the fitted parasitic power is 1.0 ± 1.2 pW, the central T_c shifts to 48.9 mK with an uncertainty of 53 mK, and K_{e-ph} is unchanged. The blow-up of the T_c and $P_{parasitic}$ uncertainties reflects a strong degeneracy between the two. A finite parasitic load shifts the entire $P_{J,0}(T_b)$ curve uniformly, which the fit can absorb either by raising T_c or by adding $P_{parasitic}$, with no purely-thermal handle to choose between them. Resolving this degeneracy requires an external constraint on $P_{parasitic}$.

Repeating the $P_{parasitic} = 0$ fit independently for each of the eleven channels of Z2 yields the per-channel parameters in Figure 3-10. The channels divide into two families. The lower (S1)

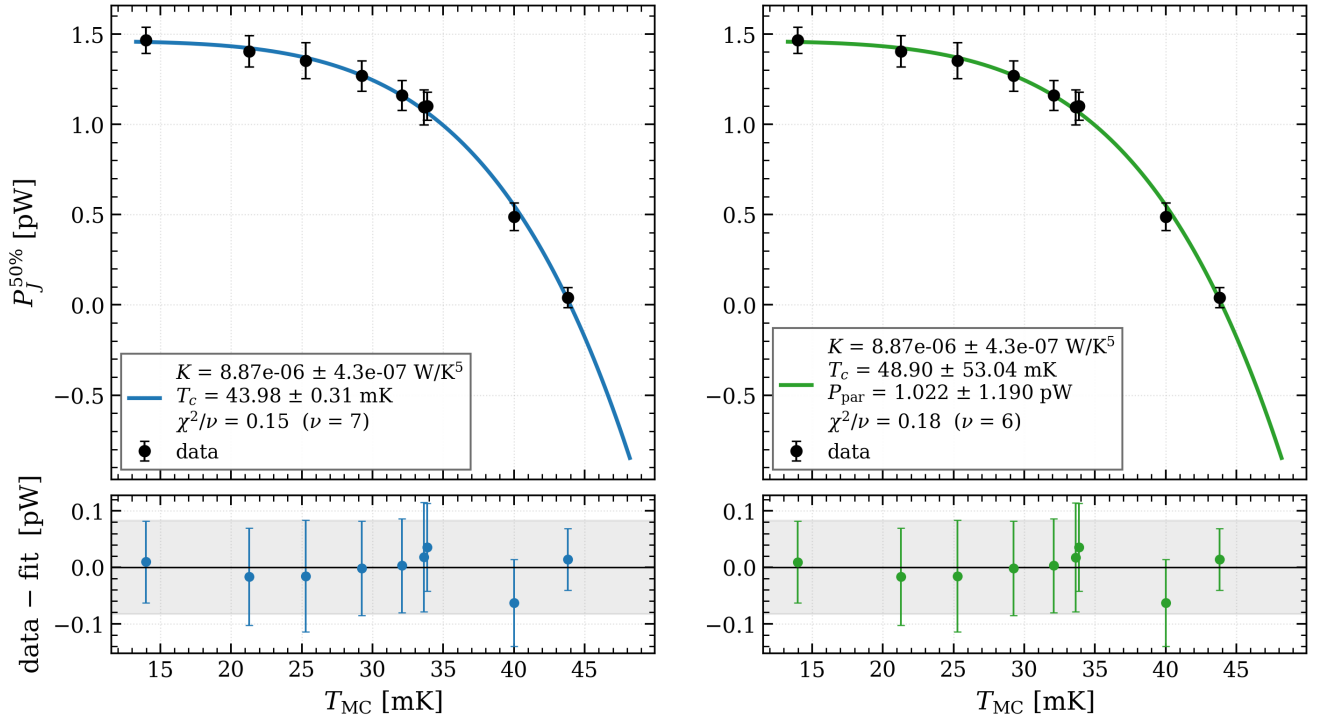


Figure 3-9. Joule power at $f_{R_n} = 1/2$ versus bath temperature for a single HV Z2 channel, fit to Equation 3-53 with $P_{\text{parasitic}} = 0$ (left) and with $P_{\text{parasitic}}$ free (right), residuals below. Error bars are the standard deviation of $P_{J,0}$ across the $I_b I_s$ transition at each bath temperature.

wafer side groups around $T_c \approx 43\text{--}46$ mK with $K_{e\text{-ph}} \approx 6\text{--}9 \times 10^{-6}$ W/K⁵, while the upper (S2) wafer side groups tightly around $T_c \approx 49$ mK with similar $K_{e\text{-ph}}$. The 5 mK T_c split between the two halves is consistent with a known thickness gradient in the W deposition across the wafer, and the within-side spread of 1–2 mK reflects the residual non-uniformity of the W film at the patterning scale. The uncertainty bars are dominated by the same $T_c\text{--}P_{\text{parasitic}}$ correlation seen in Figure 3-9; channels with smaller transition powers (PCS1, PAS1, PAS2) carry larger fit uncertainties because their bath sweeps reach further into the regime where the assumed-zero parasitic loading is least defensible.

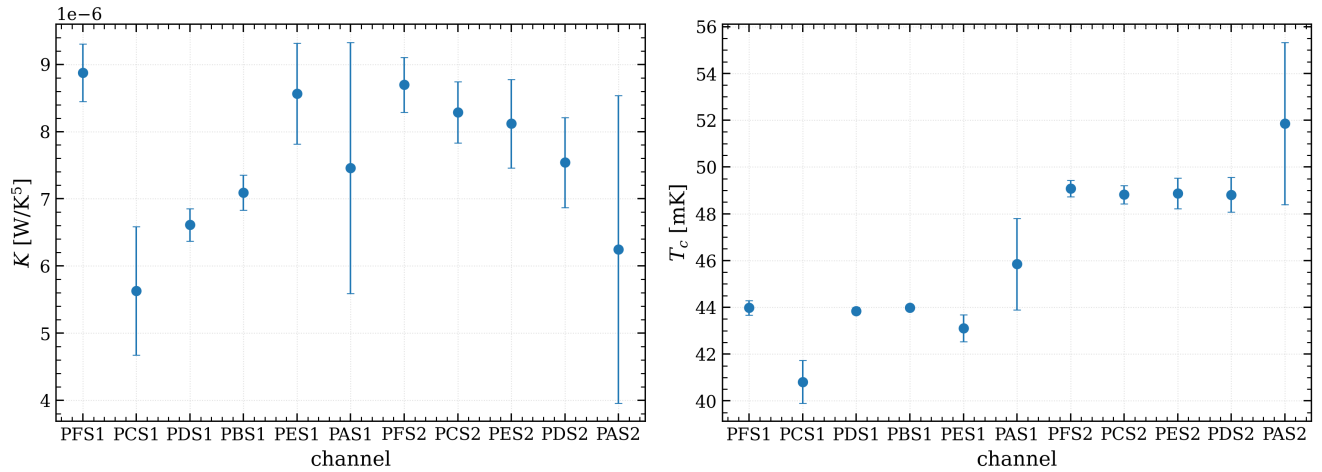


Figure 3-10. Per-channel fit results across the HV detector Z2, with $P_{\text{parasitic}} = 0$. **Left:** cooling coefficient $K_{e\text{-ph}}$ for each channel. **Right:** critical temperature T_c for each channel.

Two systematics dominate this extraction. The first is the $T_c\text{--}P_{\text{parasitic}}$ degeneracy of Figure 3-9, which inflates per-channel uncertainties and biases the central T_c values low under the $P_{\text{parasitic}} = 0$ assumption. The second is the channel-to-channel variation of Figure 3-10, which carries a real fabrication signature, the ~ 5 mK T_c split between the two wafer halves and the few-mK within-side spread, but cannot be cleanly separated from the parasitic-power systematic without per-channel constraints on $P_{\text{parasitic}}$. Lifting both requires either an external constraint on the parasitic loading, as used in the HVeV analysis, or a joint fit that brings in the dynamical response and independent substrate thermometry; both are left to future work.

3.3 Linearization and the Dynamics Matrix

The coupled ODEs of Section 3.2.2.1 describe the full nonlinear dynamics of the TES but cannot be solved in closed form. Connecting the model to data requires linearizing around the operating point, which converts the nonlinear system into a matrix equation whose elements have direct physical meaning and whose transfer functions can be compared to measurements.

3.3.1 Linearization

A SuperCDMS detector has M readout channels patterned across the crystal surface, each containing N TES wired in parallel. The N TES within a channel are treated as a single effective TES with parallel resistance R/N , total heat capacity NC , and total electron-phonon coupling $N\Sigma_W\mathcal{V}_W$; spread in T_c or R_n across the N TES broadens the effective $R(T, I)$ of the parallel combination. The M channels share the same crystal substrate at temperature T_{subst} . When all channels are identical and biased at the same operating point, each dumps the same Joule power into the substrate, and the substrate equation picks up a factor of M on the electron-phonon cooling term. During a pulse the channels are no longer at the same operating point. The phonon flux from a localized energy deposition reaches each channel differently, so $P_{\text{signal},k}(t)$ differs from channel to channel and so does the cooling power each channel delivers to the substrate.

Consider the system at a stable operating point $(I_0, T_0, T_{\text{subst},0}, T_{b,0})$ satisfying the steady-state conditions. A small perturbation displaces the state variables,

$$I_s = I_0 + \delta I, \quad T = T_0 + \delta T, \quad T_{\text{subst}} = T_{\text{subst},0} + \delta T_{\text{subst}}, \quad T_b = T_{b,0} + \delta T_b. \quad (3-55)$$

Each nonlinear term in the ODEs is expanded to first order. The resistance coupling is

$$\delta R = \left. \frac{\partial R}{\partial T} \right|_0 \delta T + \left. \frac{\partial R}{\partial I_s} \right|_0 \delta I = R_0 \left[\frac{\alpha}{T_0} \delta T + \frac{\beta}{I_0} \delta I \right], \quad (3-56)$$

where α and β are the sensitivity parameters defined in Section 3.1. Higher-order terms in the expansion generate the nonlinear sensitivities α_n, β_n , and the mixed coefficients γ_{mn} developed in appendix A.

The Joule power $P_J = I_s^2 R(T, I_s)$ depends on I_s both explicitly through the prefactor and implicitly through the resistance surface,

$$\delta P_J = \left. \frac{\partial P_J}{\partial T} \right|_0 \delta T + \left. \frac{\partial P_J}{\partial I_s} \right|_0 \delta I. \quad (3-57)$$

The temperature partial acts only through R ,

$$\left. \frac{\partial P_J}{\partial T} \right|_0 = I_0^2 \left. \frac{\partial R}{\partial T} \right|_0 = \frac{P_{J,0} \alpha}{T_0}. \quad (3-58)$$

The current partial picks up contributions from both the explicit I_s^2 and the implicit $R(T, I_s)$,

$$\left. \frac{\partial P_J}{\partial I_s} \right|_0 = 2 I_0 R_0 + I_0^2 \left. \frac{\partial R}{\partial I_s} \right|_0 = I_0 R_0 (2 + \beta). \quad (3-59)$$

The factor $(2 + \beta)$ combines the explicit current dependence (2) with the implicit resistance change (β). The linearized Joule power is

$$\delta P_J = \frac{P_{J,0} \alpha}{T_0} \delta T + I_0 R_0 (2 + \beta) \delta I. \quad (3-60)$$

The electron-phonon power $P_{\text{e-ph}} = \Sigma_W \mathcal{V}_W (T^n - T_{\text{subst}}^n)$ linearizes to

$$\delta P_{\text{e-ph}} = G \delta T - \eta G \delta T_{\text{subst}}, \quad (3-61)$$

where $G = n \Sigma_W \mathcal{V}_W T_0^{n-1}$ is the differential thermal conductance at the TES electron temperature and

$$\eta = \frac{T_{\text{subst},0}^{n-1}}{T_0^{n-1}} \quad (3-62)$$

accounts for the asymmetry of the power-law link when the two ends are at different temperatures ($\eta < 1$ because the substrate is colder than the TES).

The substrate-bath power $P_K = \mathcal{K} A_c (T_{\text{subst}}^m - T_b^m)$ linearizes to

$$\delta P_K = G_K \delta T_{\text{subst}} - G_{K,b} \delta T_b, \quad (3-63)$$

where $G_K = m \mathcal{K} A_c T_{\text{subst},0}^{m-1}$ is the differential Kapitza conductance at the substrate temperature and $G_{K,b} = m \mathcal{K} A_c T_b^{m-1}$ is the corresponding quantity at the bath temperature.

Substituting the linearized terms into the ODEs, the steady-state contributions cancel exactly. In the electrical equation, $V_{\text{th}} - I_0(R_l + R_0) = 0$. In the thermal equation, $P_{J,0} + P_{\text{par}} - P_{\text{e-ph},0} = 0$, so P_{par} drops out of the linearized system entirely. In the substrate equation, $P_{\text{e-ph},0} - P_{K,0} = 0$. The linearized system is

$$L \frac{d(\delta I)}{dt} = -[R_l + R_0(1 + \beta)] \delta I - \frac{I_0 R_0 \alpha}{T_0} \delta T, \quad (3-64)$$

$$C \frac{d(\delta T)}{dt} = I_0 R_0 (2 + \beta) \delta I - \left(G - \frac{P_{J,0} \alpha}{T_0} \right) \delta T + \eta G \delta T_{\text{subst}} + \delta P_{\text{signal}}, \quad (3-65)$$

$$C_{\text{subst}} \frac{d(\delta T_{\text{subst}})}{dt} = G \delta T - (\eta G + G_K) \delta T_{\text{subst}} + G_{K,b} \delta T_b. \quad (3-66)$$

Three state variables, δI , δT , δT_{subst} , coupled through the resistance surface (α , β), the Joule heating ($P_{J,0}$, $I_0 R_0$), the electron-phonon link (G , η), and the Kapitza boundary (G_K). In the multi-channel common mode, G in the substrate equation (Equation 3-66) is replaced by $M G$, and ηG by $M \eta G$.

3.3.2 The Dynamics Matrix

Defining the state vector $\mathbf{x} = (\delta I, \delta T, \delta T_{\text{subst}})^T$, the linearized system takes the form

$$\frac{d\mathbf{x}}{dt} = \mathbf{A} \mathbf{x} + \begin{pmatrix} \delta V/L \\ \delta P_{\text{signal}}/C \\ G_{K,b} \delta T_b/C_{\text{subst}} \end{pmatrix}, \quad (3-67)$$

where $\mathbf{A}$ is the dynamics matrix,

$$\mathbf{A} = \begin{pmatrix} -\frac{1}{\tau_{\text{el}}} & -\frac{\mathcal{L} G}{I_0 L} & 0 \\ \frac{I_0 R_0 (2 + \beta)}{C} & -\frac{1 - \mathcal{L}}{\tau_{\text{e-ph}}} & \frac{\eta}{\tau_{\text{e-ph}}} \\ 0 & \frac{1}{\tau_{\text{load}}} & -\frac{1}{\tau_{\text{subst}}} \end{pmatrix}. \quad (3-68)$$

Every element traces back to a specific perturbation acting on a specific subsystem.

$A_{11} = -1/\tau_{\text{el}}$, where

$$\tau_{\text{el}} = \frac{L}{R_l + R_0(1 + \beta)} \quad (3-69)$$

is the electrical time constant. The factor $(1 + \beta)$ enters because even at fixed temperature, a change in current shifts the resistance through β , adding to the effective impedance. This generalizes the frozen- R result $\tau_{\text{el}} = L/(R_l + R_0)$ from Section 3.2.1.1.

$A_{22} = -(1 - \mathcal{L})/\tau_{\text{e-ph}}$, where

$$\tau_{\text{e-ph}} = \frac{C}{G} \quad (3-70)$$

is the bare electron-phonon relaxation time and

$$\mathcal{L} = \frac{P_{J,0} \alpha}{G T_0} \quad (3-71)$$

is the loop gain. The sign of A_{22} determines the thermal stability at frozen current. When $\mathcal{L} < 1$, A_{22} is negative and the thermal system is self-restoring. When $\mathcal{L} > 1$, A_{22} is positive and the system is thermally unstable at frozen current, surviving only because the electrical branch responds fast enough to regulate the Joule power.

$A_{33} = -1/\tau_{\text{subst}}$, where

$$\frac{1}{\tau_{\text{subst}}} = \frac{\eta}{\tau_{\text{load}}} + \frac{1}{\tau_{\text{K}}}, \quad (3-72)$$

with

$$\tau_{\text{load}} = \frac{C_{\text{subst}}}{G}, \quad \tau_{\text{K}} = \frac{C_{\text{subst}}}{G_{\text{K}}}. \quad (3-73)$$

The substrate relaxes through two parallel channels, reduced inflow from the TES (rate η/τ_{load}) and increased outflow to the bath (rate $1/\tau_K$). In the multi-channel common mode,

$$\tau_{\text{load}} = C_{\text{subst}}/(M G).$$

$A_{12} = -\mathcal{L} G/(I_0 L)$ is the thermal-to-electrical coupling. A temperature perturbation changes the resistance, which modifies the voltage drop across the TES, and the inductor converts this into a current rate of change. The sign is negative, since heating increases the resistance, which opposes current flow.

$A_{21} = I_0 R_0 (2 + \beta)/C$ is the electrical-to-thermal coupling. A current perturbation changes the Joule power, and the heat capacity converts this into a temperature rate of change. These two couplings form the electrothermal feedback loop, and their product $A_{12} \cdot A_{21}$ is the round-trip gain.

$A_{23} = \eta/\tau_{\text{e-ph}}$ is the substrate-to-TES coupling. A warmer substrate reduces the temperature gradient across the electron-phonon link, decreasing the power flowing out of the TES.

$A_{32} = 1/\tau_{\text{load}}$ is the TES-to-substrate coupling. A warmer TES pushes more power into the substrate.

$A_{13} = A_{31} = 0$ because the electrical and substrate subsystems do not interact directly. Any influence between them must pass through the TES temperature.

The off-diagonal products define two coupling frequencies that appear throughout the eigenvalue problem. Each is the product of two rates from opposite legs of a feedback loop, and the two loops have a parallel structure.

The electrothermal coupling frequency is

$$\Omega_{\text{etc}}^2 = -A_{12} A_{21} = \frac{\mathcal{L}}{\tau_{\text{e-ph}}} \cdot \frac{1}{\tau_{\text{dyn}}}, \quad (3-74)$$

where $\tau_{\text{dyn}} = L/R_{\text{dyn}}$ is the dynamic electrical time constant with $R_{\text{dyn}} = R_0(2 + \beta)$ [45]. The first factor $\mathcal{L}/\tau_{\text{e-ph}}$ is the thermal leg of the loop, how fast a temperature perturbation produces a resistance change, amplified by the loop gain $\mathcal{L}$. The second factor $1/\tau_{\text{dyn}}$ is the electrical leg, how fast the circuit converts that resistance change into a Joule power response, limited by the

inductance L and the dynamic resistance $R_{\text{dyn}} = R_0(2 + \beta)$ that governs the sensitivity of the Joule power to current. The minus sign in the definition compensates for the opposite signs of $A_{12} < 0$ and $A_{21} > 0$, making Ω_{etc}^2 positive.

The TES-substrate thermal coupling frequency is

$$\Omega_{\text{tsc}}^2 = A_{23} A_{32} = \frac{\eta}{\tau_{\text{e-ph}}} \cdot \frac{M}{\tau_{\text{load}}}, \quad (3-75)$$

with an analogous two-leg structure. The first factor $\eta/\tau_{\text{e-ph}}$ is the feedback leg, how fast a substrate temperature perturbation reduces the TES cooling rate, controlled by the temperature ratio $\eta = (T_{\text{subst},0}/T_0)^{n-1}$ which plays the role that $\mathcal{L}$ plays in the electrothermal loop. The second factor $M/\tau_{\text{load}} = MG/C_{\text{subst}}$ is the inertial leg, how fast the TES drives the substrate temperature, limited by the substrate heat capacity C_{subst} and the total thermal conductance MG from all M channels. In the round trip, the TES heats the substrate (A_{32}), and the warmer substrate reduces the temperature gradient across the electron-phonon link, weakening the TES cooling (A_{23}).

The two loops both reduce power when the TES heats up, but they act on different parts of the thermal budget. The electrothermal loop reduces the Joule power, which is the heat source, actively restoring the operating point. The thermal-substrate loop reduces the cooling power, which is the heat sink the restoring force relies on. The first helps stability by weakening the source; the second hurts stability by weakening the sink. This is reflected in the characteristic polynomial, where $+\Omega_{\text{etc}}^2$ strengthens the restoring dynamics and $-\Omega_{\text{tsc}}^2$ weakens them.

The matrix has a block structure reflecting the physical architecture. The upper-left 2×2 block is the one-block model of Irwin and Hilton, a single thermal body coupled to the electrical circuit. The lower-right 2×2 block is the thermal block, the TES coupled to the substrate through the electron-phonon link. The full 3×3 matrix is a two-block model in the classification of Maasilta [58].

3.4 Time-Domain Solution

The dynamics matrix encodes the full linear response of the TES, but its content becomes physical only when the eigenvalue problem is solved. The eigenvalues set the rates at which

perturbations decay or oscillate, and the eigenvectors set the amplitudes with which each mode appears in the current. Together they give the impulse current response as a closed-form sum of three exponentials, computable directly from the physical parameters

$\{\tau_{\text{el}}, \tau_{\text{e-ph}}, \tau_{\text{subst}}, \mathcal{L}, \Omega_{\text{etc}}^2, \Omega_{\text{tsc}}^2\}$ without numerical integration.

3.4.1 Characteristic Polynomial and Eigenvalues

The eigenvalues of $\mathbf{A}$ are the roots of the characteristic polynomial

$$P(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}) = \lambda^3 + a_2 \lambda^2 + a_1 \lambda + a_0 = 0, \quad (3-76)$$

whose coefficients are the three independent matrix invariants of $\mathbf{A}$. They can be expressed either as traces of powers of $\mathbf{A}$,

$$a_2 = -\text{tr}(\mathbf{A}), \quad (3-77)$$

$$a_1 = \frac{1}{2} [(\text{tr} \mathbf{A})^2 - \text{tr}(\mathbf{A}^2)], \quad (3-78)$$

$$a_0 = -\det(\mathbf{A}), \quad (3-79)$$

or, by Viète's relations, as the elementary symmetric polynomials of the eigenvalues,

$$\lambda_1 + \lambda_2 + \lambda_3 = -a_2, \quad (3-80)$$

$$\lambda_1 \lambda_2 + \lambda_1 \lambda_3 + \lambda_2 \lambda_3 = a_1, \quad (3-81)$$

$$\lambda_1 \lambda_2 \lambda_3 = -a_0. \quad (3-82)$$

The trace and determinant identities are theorems, not definitions. Both quantities are basis-invariant, computable either from the matrix elements or from the eigenvalues. The eigenvalues themselves differ from the diagonal elements of $\mathbf{A}$ once the off-diagonal couplings are nonzero. The off-diagonal couplings $A_{12}, A_{21}, A_{23}, A_{32}$ are precisely the electrothermal feedback that mixes the bare subsystem rates into the dressed eigenvalue rates of the linearized system. The

diagonal sums and the polynomial coefficients constrain the eigenvalues but do not equal them individually.

In terms of the time constants and coupling frequencies of Section 3.3.2,

$$a_2 = \frac{1}{\tau_{\text{el}}} + \frac{1 - \mathcal{L}}{\tau_{\text{e-ph}}} + \frac{1}{\tau_{\text{subst}}}, \quad (3-83)$$

$$a_1 = \frac{1 - \mathcal{L}}{\tau_{\text{e-ph}} \tau_{\text{el}}} + \frac{1}{\tau_{\text{el}} \tau_{\text{subst}}} + \frac{1 - \mathcal{L}}{\tau_{\text{e-ph}} \tau_{\text{subst}}} + \Omega_{\text{etc}}^2 - \Omega_{\text{tsc}}^2, \quad (3-84)$$

$$a_0 = \frac{1 - \mathcal{L}}{\tau_{\text{e-ph}} \tau_{\text{el}} \tau_{\text{subst}}} + \frac{\Omega_{\text{etc}}^2}{\tau_{\text{subst}}} - \frac{\Omega_{\text{tsc}}^2}{\tau_{\text{el}}}. \quad (3-85)$$

The polynomial has real coefficients, so the eigenvalues are either three real values or one real value and a complex-conjugate pair. The substitution $\lambda = \mu - a_2/3$ shifts the roots to their mean and removes the quadratic term, giving the depressed cubic

$$\mu^3 + p\mu + q = 0, \quad p = a_1 - \frac{a_2^2}{3}, \quad q = a_0 - \frac{a_1 a_2}{3} + \frac{2a_2^3}{27}. \quad (3-86)$$

The discriminant

$$\Delta = (\mu_1 - \mu_2)^2(\mu_1 - \mu_3)^2(\mu_2 - \mu_3)^2 \quad (3-87)$$

is positive when all three roots are real (each squared factor is positive), zero when two roots coincide, and negative when two roots form a complex-conjugate pair $\mu = a \pm ib$, since one factor becomes $(2ib)^2 = -4b^2$. In terms of the depressed-cubic coefficients,

$$\Delta = -4p^3 - 27q^2. \quad (3-88)$$

The two regimes have distinct physical character. When $\Delta > 0$, all three eigenvalues are real and negative, and the system relaxes through three monotonic decay modes. When $\Delta < 0$, one eigenvalue is real and the other two form a complex-conjugate pair, producing one monotonic mode and one damped electrothermal oscillation.

When $\Delta > 0$, the three real eigenvalues take the trigonometric form

$$\lambda_k = -\frac{a_2}{3} + 2\sqrt{-\frac{p}{3}} \cos \left[\frac{1}{3} \arccos \left(\frac{3q}{2p} \sqrt{-\frac{3}{p}} \right) - \frac{2\pi k}{3} \right], \quad k = 0, 1, 2. \quad (3-89)$$

When $\Delta < 0$, the real eigenvalue takes the radical form

$$\lambda_r = -\frac{a_2}{3} + \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}, \quad (3-90)$$

where the cube roots are taken as real (Appendix B.1). The complex pair follows from the quadratic factor of the original polynomial. Dividing $P(\lambda)$ by $(\lambda - \lambda_r)$ gives $\lambda^2 + b_1\lambda + b_0$ with $b_1 = a_2 + \lambda_r$ and $b_0 = -a_0/\lambda_r$, and the complex eigenvalues are

$$\lambda_{\pm} = -\frac{b_1}{2} \pm i \sqrt{b_0 - \frac{b_1^2}{4}} \equiv -\frac{1}{\tau_{\text{osc}}} \pm i \omega_{\text{ETO}}, \quad (3-91)$$

with damping rate $1/\tau_{\text{osc}} = b_1/2$ and oscillation frequency $\omega_{\text{ETO}} = \sqrt{b_0 - b_1^2/4}$.

The chain $(\tau_{\text{el}}, \tau_{\text{e-ph}}, \tau_{\text{subst}}, \mathcal{L}, \Omega_{\text{etc}}^2, \Omega_{\text{tsc}}^2) \rightarrow (a_2, a_1, a_0) \rightarrow (p, q) \rightarrow \Delta \rightarrow \{\lambda_k\}$ is closed-form at every step. The eigenvalues are computable directly from the physical parameters of the operating point.

The eigenvalue structure simplifies in the one-block limit $\Omega_{\text{tsc}}^2 \rightarrow 0$, where the substrate decouples and the cubic factors as $(\lambda - A_{33})(\lambda^2 + b_1\lambda + b_0) = 0$ with

$$b_1 = \frac{1}{\tau_{\text{el}}} + \frac{1 - \mathcal{L}}{\tau_{\text{e-ph}}}, \quad b_0 = \frac{1 - \mathcal{L}}{\tau_{\text{e-ph}} \tau_{\text{el}}} + \Omega_{\text{etc}}^2. \quad (3-92)$$

The substrate eigenvalue $\lambda_3 = A_{33} = -1/\tau_{\text{subst}}$ separates out, and the remaining two eigenvalues are

$$\lambda_{\pm}^{(1b)} = -\frac{b_1}{2} \pm \frac{1}{2} \sqrt{b_1^2 - 4b_0}, \quad (3-93)$$

the standard rise/fall pair of Irwin and Hilton [45]. The full three-block result reduces to this pair in the limit of vanishing substrate coupling, with the substrate pole at $-1/\tau_{\text{subst}}$ as a third decoupled mode that does not participate in the current response.

3.4.2 Eigenvectors, Mode Weights, and the Impulse Response

The eigenvalues fix the rates of the three exponential modes; the eigenvectors fix their amplitudes. The sparsity $A_{13} = A_{31} = 0$ in the dynamics matrix gives the right eigenvector components in closed form. Normalizing the temperature component to unity, the right eigenvector for eigenvalue λ_k is

$$\mathbf{v}_k = \begin{pmatrix} v_{I,k} \\ 1 \\ v_{S,k} \end{pmatrix} = \begin{pmatrix} \frac{A_{12}}{\lambda_k - A_{11}} \\ 1 \\ \frac{A_{32}}{\lambda_k - A_{33}} \end{pmatrix}. \quad (3-94)$$

Each component has the structure of a coupling element divided by the distance between the eigenvalue and the bare self-relaxation rate of the corresponding subsystem. A mode participates strongly in a given channel only when its eigenvalue is close to that channel's bare rate.

The mode weight for the current response to a thermal-channel perturbation, derived in Appendix B.2, is

$$v_{I,k} w_{T,k} = \frac{A_{12} (\lambda_k - A_{33})}{P'(\lambda_k)}, \quad (3-95)$$

where $P'(\lambda_k) = 3\lambda_k^2 + 2a_2\lambda_k + a_1 = \prod_{j \neq k} (\lambda_k - \lambda_j)$ is the derivative of the characteristic polynomial, measuring the separation between eigenvalues. The mode weights satisfy the sum rule

$$\sum_k v_{I,k} w_{T,k} = 0, \quad (3-96)$$

which enforces $\delta I(0^+) = 0$ after an impulsive energy deposition, reflecting the inductor's resistance to instantaneous current changes.

An impulsive energy deposition E_0 delivered to the TES electrons appears as $\delta P_{\text{signal}}(t) = E_0 \delta(t)$ and produces an initial condition $\delta T(0^+) = E_0/C$ in the thermal channel. The current response is the sum of three exponentials weighted by the mode weights,

$$\delta I(t) = \frac{E_0}{C} \sum_{k=1}^3 v_{I,k} w_{T,k} e^{\lambda_k t}, \quad t \geq 0. \quad (3-97)$$

In the overdamped regime $\Delta > 0$, all three eigenvalues are real and the response is a sum of three monotonic exponentials. Each term corresponds to a physical mode, a fast electrical-thermal pair set by the upper-left 2×2 dynamics and a slow substrate mode whose rate is close to $-1/\tau_{\text{subst}}$ but shifted by the substrate coupling.

In the ETO regime $\Delta < 0$, one eigenvalue is real and the other two form a complex-conjugate pair $\lambda_{\pm} = -1/\tau_{\text{osc}} \pm i \omega_{\text{ETO}}$ from Equation 3-91. The corresponding mode weights $c_r \equiv v_{I,r} w_{T,r}$ and $c_{\pm} \equiv v_{I,\pm} w_{T,\pm}$ follow from the same closed form (Equation 3-95) evaluated at the relevant eigenvalues; c_r is real and $c_+ = \overline{c_-}$ is complex. Their contributions combine into a real damped oscillation,

$$\delta I(t) = \frac{E_0}{C} \left[c_r e^{\lambda_r t} + 2 |c_{\pm}| e^{-t/\tau_{\text{osc}}} \cos(\omega_{\text{ETO}} t + \phi) \right], \quad (3-98)$$

where $|c_{\pm}|$ and $\phi = \arg(c_+)$ are the modulus and argument of the complex mode weight, evaluated numerically from the physical parameters of the operating point. The response combines a slow exponential decay at rate $|\lambda_r|$ with a damped oscillation at frequency ω_{eto} , both computable in closed form from the chain $\{\tau, \mathcal{L}, \Omega^2\} \rightarrow \{\lambda_k\} \rightarrow \{c_k\} \rightarrow \delta I(t)$.

The eigenvalues, eigenvectors, and mode weights all depend on the steady-state operating point through I_0 , $P_{J,0}$, $T_{\text{subst},0}$, and the dimensionless parameters $\mathcal{L}$ and η . Solving the steady-state balance with $P_{\text{par}} = 0$ gives T_{subst} and I_0 as functions of the substrate-bath Kapitza coefficient K_{s-b} for a given bias point. Figure 3-11 shows this dependence for three SuperCDMS devices, an HVeV detector, a Ge HV detector, and a Si HV detector, with parameters listed in Table 3-1.

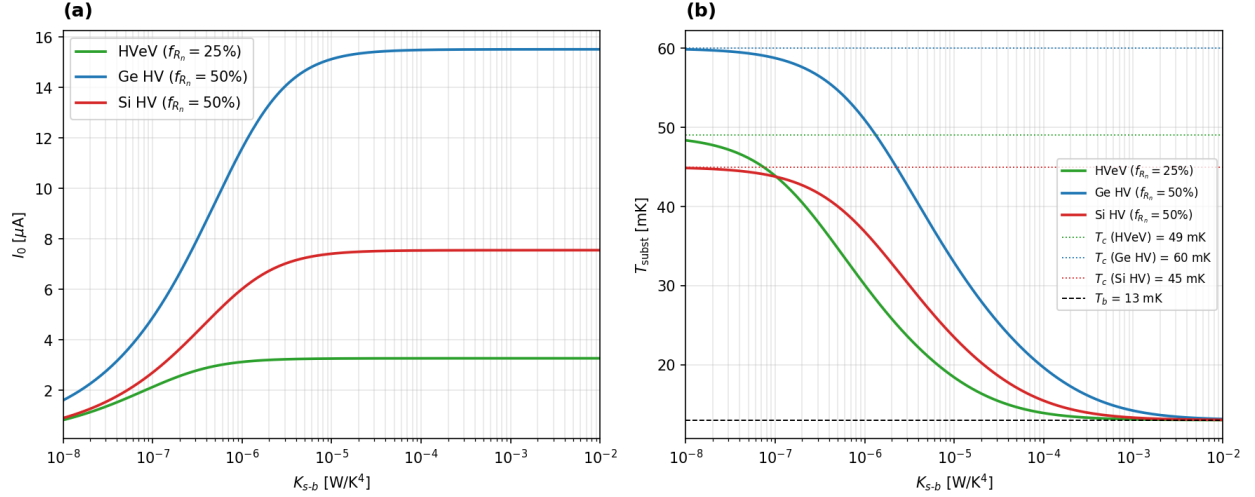


Figure 3-11. Steady-state operating current I_0 (a) and substrate temperature T_{subst} (b) versus K_{s-b} for the three devices of Table 3-1. Reference lines mark T_c for each device (dotted) and the shared bath temperature $T_b = 13$ mK (dashed).

The impulse response computed from Equations 3-97 and 3-98 is shown for each device in Figures 3-12, 3-13, and 3-14. Panels (a) and (b) sweep K_{s-b} at the operating inductance on linear and logarithmic scales; panel (c) sweeps the substrate heat capacity C_{subst} at fixed K_{s-b} ; panel (d) shows the ETO regime at three increased inductance values, with the one-block result overlaid.

The substrate pole's contribution to the pulse shape is governed by the factor $\lambda_k - A_{33}$ in the numerator of the mode weight (Equation 3-95). In the one-block limit $\Omega_{\text{tsc}}^2 \rightarrow 0$, the substrate eigenvalue coincides with $A_{33} = -1/\tau_{\text{subst}}$, and the substrate mode weight vanishes identically, and the third exponential drops out and the response collapses to the two-exponential pulse of Irwin and Hilton [45]. Restoring the substrate coupling shifts λ_3 away from $-1/\tau_{\text{subst}}$ by an amount that scales with Ω_{tsc}^2 , and the third exponential reappears with weight proportional to this shift. The substrate pole therefore manifests as a slow tail whose amplitude reflects the strength of the TES-substrate coupling, while leaving the fast rising edge dominated by the electrical-thermal pair largely unchanged. The trend is visible in panel (b) of each impulse response figure, where the tail plateau in the two-block model rises as K_{s-b} decreases, while the one-block result decays cleanly. Panel (c) shows the complementary effect of the substrate heat capacity, since C_{subst} does

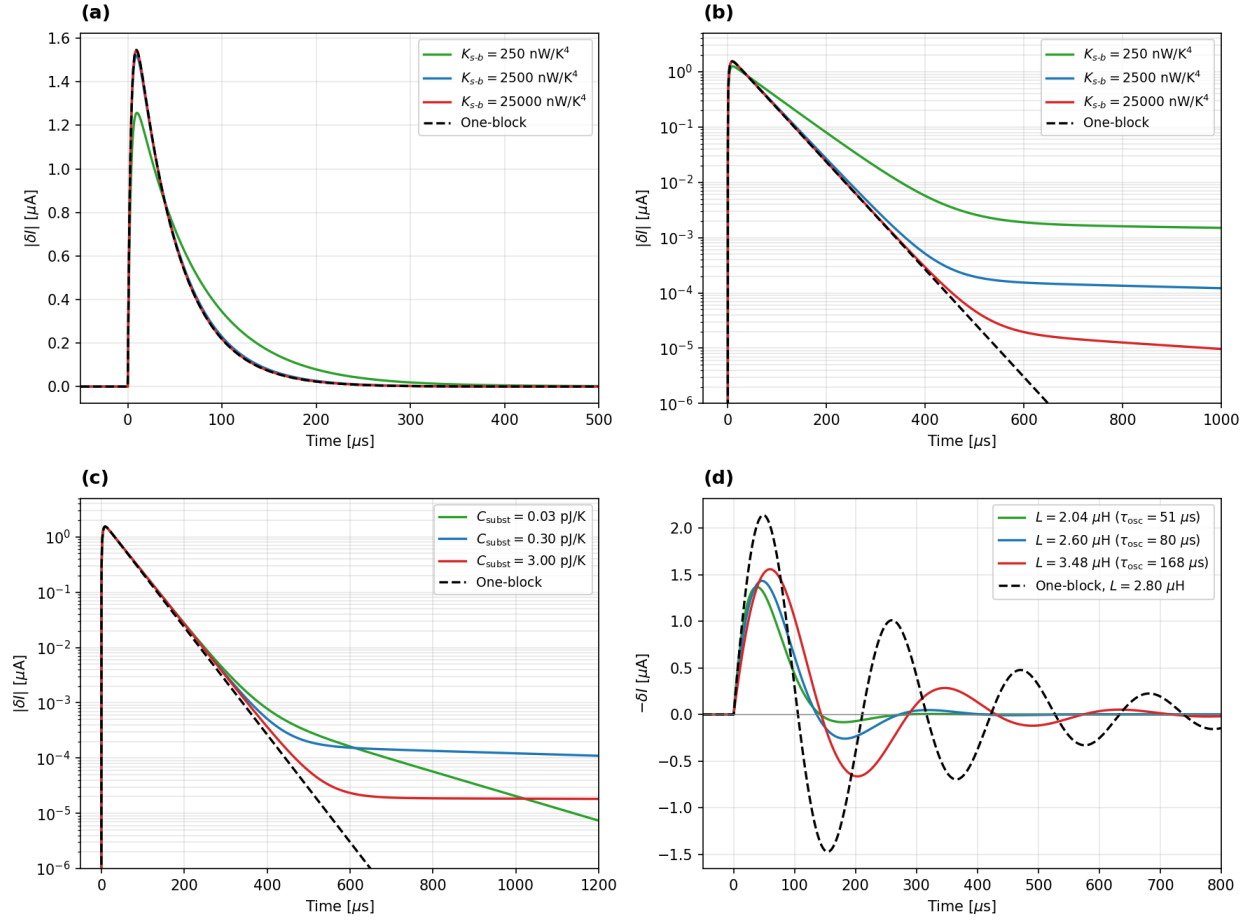


Figure 3-12. HVeV impulse response from the closed-form chain. Panel (c) is at $K_{s-b} = 250 \text{ nW/K}^4$. Panel (d) is in the ETO regime at $L \sim 2\text{--}3.5 \mu\text{H}$.

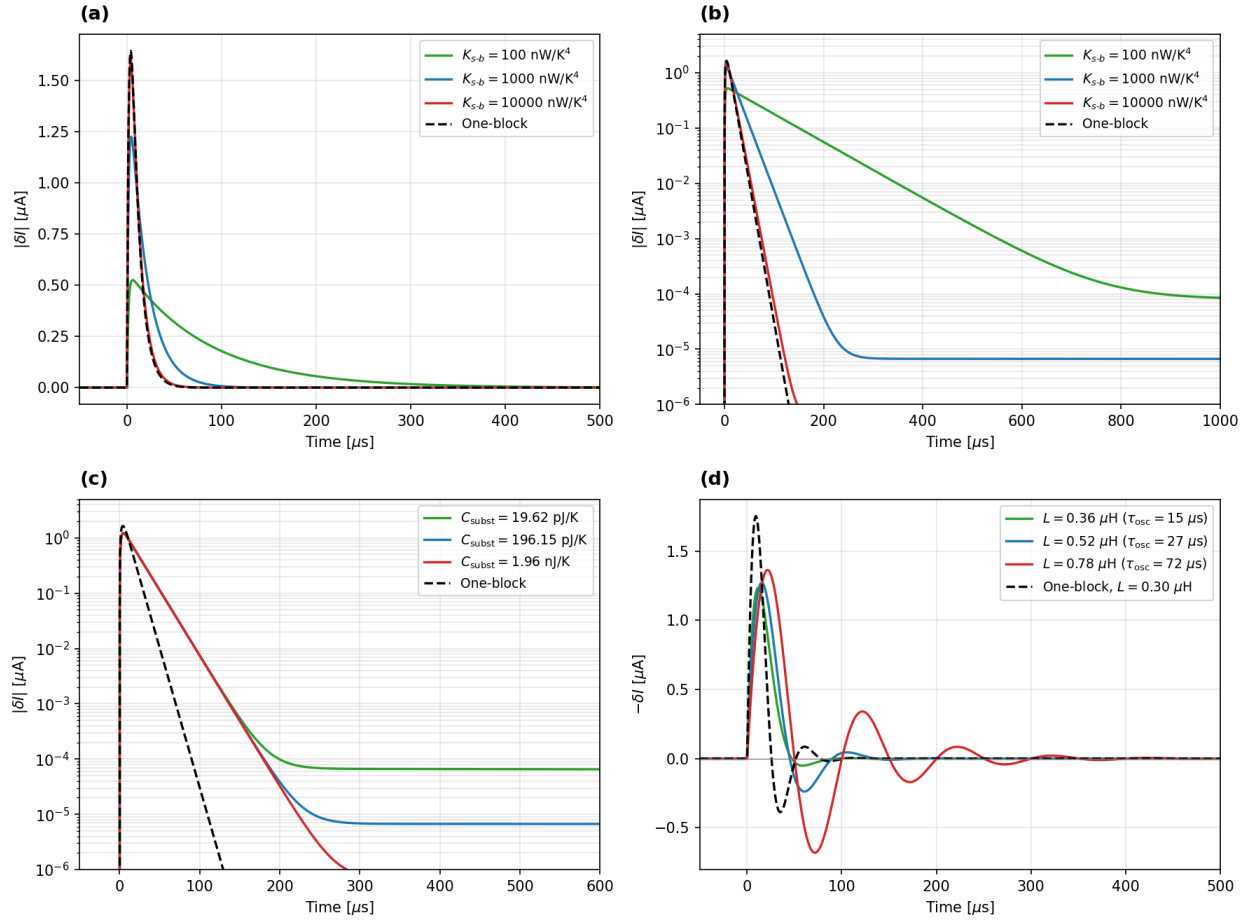


Figure 3-13. Ge HV impulse response, theoretical channel parameters. Panel (c) is at $K_{s-b} = 1000$ nW/K⁴. The ETO window in panel (d) is at smaller inductances (~ 0.4 – 0.8 μH) than the HVeV case.

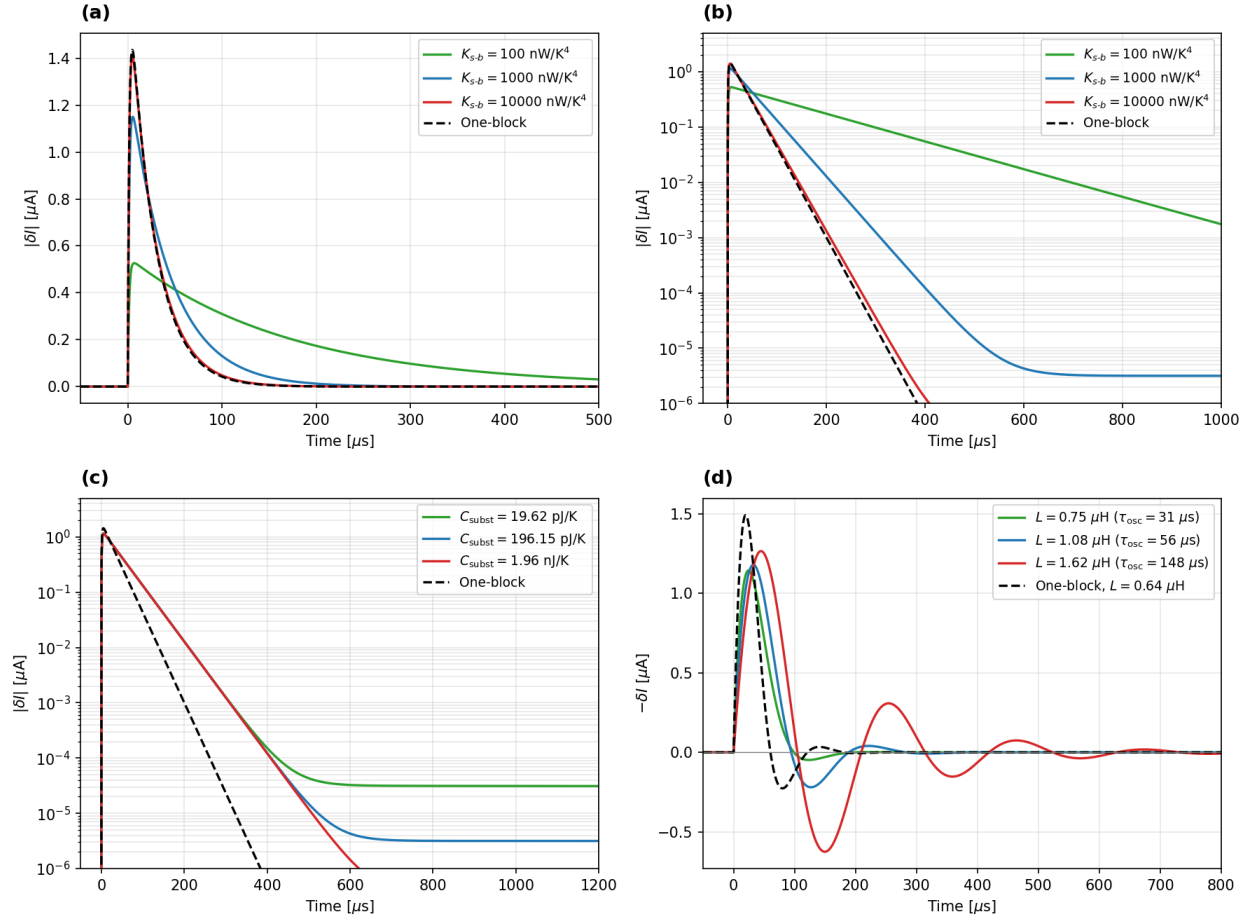


Figure 3-14. Si HV impulse response, theoretical channel parameters. The lower T_c relative to Ge HV pushes the ETO window in panel (d) to larger inductances (~ 0.7 – 1.6 μH).

not enter the steady-state operating point but sets the substrate relaxation timescale τ_{subst} , controlling how quickly the substrate-pole tail decays.

3.4.3 Stability and Electrothermal Oscillations

The linearized system is stable when all eigenvalues have negative real parts. The Routh-Hurwitz criterion gives this condition directly from the polynomial coefficients without computing the eigenvalues, and can be motivated from the Viète relations.

For all three eigenvalues to be negative, or to have negative real parts, three conditions must hold simultaneously. The sum of eigenvalues must be negative, $\lambda_1 + \lambda_2 + \lambda_3 = -a_2 < 0$, requiring

$$a_2 > 0. \quad (3-99)$$

With $a_2 = 1/\tau_{\text{el}} + (1 - \mathcal{L})/\tau_{\text{e-ph}} + 1/\tau_{\text{subst}}$, this requires the total damping from the electrical and substrate branches to exceed the thermal instability driven by $\mathcal{L}$. When $\mathcal{L} \leq 1$ the condition is automatically satisfied; when $\mathcal{L} > 1$, it places an upper bound on the inductance.

The product of eigenvalues must be negative, $\lambda_1 \lambda_2 \lambda_3 = -a_0 < 0$, requiring

$$a_0 > 0. \quad (3-100)$$

This is the DC stability condition, ensuring the system returns to equilibrium after a static perturbation.

These two conditions are necessary but not sufficient. The sum and product can both be satisfied while the real part of a complex pair is positive, with a sufficiently negative real eigenvalue making the sum and product come out right. The additional requirement is

$$a_2 a_1 > a_0, \quad (3-101)$$

which follows from the Routh array construction. This coupled condition involves all three subsystems and determines the critical inductance L_{crit} above which the electrothermal oscillation

damping rate $1/\tau_{\text{osc}}$ passes through zero. Below L_{crit} the oscillations are damped; at L_{crit} the system sustains steady oscillations; above it the linear theory predicts exponential growth whose amplitude is limited by the nonlinear terms.

The ETO boundary takes a transparent form in the one-block limit. The one-block discriminant of Equation 3-93 is negative when

$$2\Omega_{\text{etc}} > \left| \frac{1}{\tau_{\text{el}}} - \frac{1 - \mathcal{L}}{\tau_{\text{e-ph}}} \right|, \quad (3-102)$$

the condition for the electrothermal coupling frequency to exceed half the mismatch between the bare electrical and thermal rates. ETO sets in when these two rates approach each other. When τ_{el} becomes comparable to $\tau_{\text{e-ph}}/(1 - \mathcal{L})$, the right-hand side of Equation 3-102 shrinks below the coupling threshold $2\Omega_{\text{etc}}$, and the response transitions from monotonic to oscillatory. Below this threshold the response is overdamped; above it, oscillatory. In the full three-block system the third Routh-Hurwitz condition $a_2 a_1 > a_0$ generalizes this boundary by including the substrate's contribution. The substrate provides an additional thermal reservoir that stores energy and returns it with a delay, modifying the phase relationships that determine whether the oscillation grows or decays. The one-block analysis is therefore more conservative, predicting the onset of growing oscillations at a lower inductance than the full three-block theory, because it does not account for the additional damping the substrate provides through $1/\tau_{\text{subst}}$. Panel (d) of each impulse response figure illustrates this effect, where the one-block result shows a longer-lived oscillation at the same inductance as the two-block curves, because the substrate's damping contribution is missing.

Table 3-1. Impulse response parameters for the three SuperCDMS detectors. HVeV values are measured at the channel level; Ge HV and Si HV are theoretical.

Parameter	HVeV (measured)	Ge HV (theoretical)	Si HV (theoretical)
<i>TES geometry</i>			
W thickness h [nm]	40	40	40
TES length ℓ [μm]	158	220	220
TES width w [μm]	2.4	2.4	2.4
Number of TESs / channel	506	1833	1833
<i>Material</i>			
T_c [mK]	49	60	45
T_b [mK]	13	13	13
Σ_W [$10^8 \text{ W/K}^5/\text{m}^3$]	2.0-4.0	4.0	4.0
c_w [$\text{J/K}^2/\text{m}^3$]	108	108	108
n (e-ph power law)	5	5	5
m (Kapitza power law)	4	4	4
C_{subst} [pJ/K]	0.30	196 (of Si)	196 (scaled)
<i>Circuit</i>			
L [nH]	363	100	100
R_n [m Ω]	324.7	100	100
R_p [m Ω]	11.5	1	1
R_{sh} [m Ω]	18.6	5	5
<i>Operating point</i>			
$f_{R_n} = R_0/R_n$	0.25	0.50	0.50
α	100	150	150
β	0.3	0.3	0.3

CHAPTER 4

HVEV PULSE SHAPE TUNING AND SPECTRUM RECONSTRUCTION

A real energy deposition is not an impulse. Energy released in the crystal partitions into electron-hole pairs and prompt phonons, the charges drift under the bias field and emit NTL phonons along their trajectories, and the resulting phonon population propagates through the substrate before reaching the QETs. Quasiparticles generated in the Al fins traverse the trapping geometry and dump their energy into the W TES over a finite interval set by the quasiparticle collection physics. The driving term $P_{\text{signal}}(t)$ in the TES thermal equation (Equation 3-45) is an extended flux with structure on multiple timescales, and the pulse shape carries the imprint of both the phonon and quasiparticle dynamics in the crystal and the response of the TES.

Disentangling these two contributions is the central difficulty. The phonon and quasiparticle dynamics are simulated with G4CMP, which generates charge transport, phonon emission and propagation, and quasiparticle collection from a set of material and geometry parameters. The TES response is described by the coupled ODEs of Section 3.2.2.1, integrated numerically without linearization. The HVeV LED calibration data supplies two observables. The discrete quanta peaks in the energy spectrum constrain the charge-transport parameters, fixing the integrated energy delivered by $P_{\text{signal}}(t)$ but saying little about its shape in time. The average pulse template constrains both the temporal shape of $P_{\text{signal}}(t)$ and the TES electrothermal parameters that govern the response to it, and these two sets of parameters are degenerate.

4.1 The HVeV Detector at CUTE

The High-Voltage eV-resolution (HVeV) detector is a prototype that is part of the SuperCDMS R&D program for sub-GeV dark matter and rare-event searches [59]. Each generation has improved on the sensor design, packaging, and operating conditions, with the program now reaching sub-eV phonon resolution and single-charge sensitivity.

The latest generation, deployed at the Cryogenic Underground TESt (CUTE) facility at SNOLAB, was run as a dedicated campaign to study the low-energy excess (LEE), a persistent rising rate of unknown origin observed below a few hundred eV across most low-threshold

rare-event search experiments and currently overwhelming the projected dark matter signal in the sub-GeV mass range [60].

4.1.1 The HVeV Detector Design

The NFC-SO device is a $1\text{ cm} \times 1\text{ cm} \times 4\text{ mm}$ silicon crystal. Both 1 cm^2 faces are thermally oxidized to a 40 nm SiO_2 layer, suppressing charge leakage from the instrumentation into the substrate [59]. A 30 nm Al grid patterned on the back face provides the high-voltage bias electrode, while the QET array is patterned on the front face on top of the SiO_2 .

The QET array uses the NFC mask, a two-channel design with concentric inner and outer regions. The inner channel comprises 506 QETs covering the central portion of the face, and the outer channel comprises 538 QETs covering the surrounding annulus. The Al coverage for the HVeV device is $\sim 35\%$. The mask layout and the geometry of a single QET are shown in Figure 4-1.

The QET is the elementary unit of the array. Each QET consists of a 40 nm thick tungsten TES strip with two side fins and two end fins of 600 nm thick superconducting aluminum overlapping the W at the connection regions. Geometric parameters are summarized in Table 4-1.

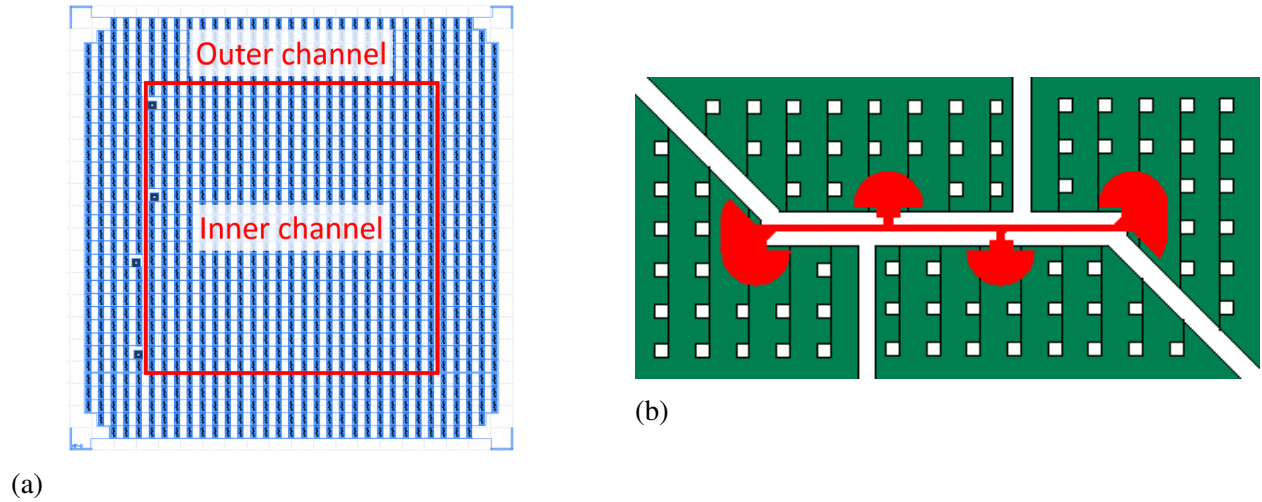


Figure 4-1. (a) NFC mask layout. (b) Single QET layout with the W TES wire and its Al-overlap regions in red, and the Al fins in green [26].

The W TES film is tuned during deposition to a critical temperature of approximately 40 mK, lower than the $\sim 65\text{ mK}$ of earlier HVeV generations [59]. The lower T_c improves the

Table 4-1. Geometric parameters of the NFC QET.

W film thickness	40 nm
Al film thickness	600 nm
TES length	158 μm
TES width	2.4 μm
Al-W overlap (stub) width	3.4 μm
Al-W overlap (stub) length	4.0 μm
Side fin width	10 μm
Side fins per QET	2
End fin width	4.8 μm
End fin height	11.2 μm
End fins per QET	2

athermal phonon resolution at the cost of stricter requirements on the operating environment, particularly on infrared photon density.

The Al-W overlap stubs are proximitized. The bilayer adopts an intermediate transition temperature lower than that of pure Al, and at the operating point the overlap is fully superconducting. The boundary between the active TES volume and the proximitized stub is therefore not sharp at the geometric edge of the W film. This ambiguity is absorbed into a dimensionless fin efficiency ε_{fin} , defined as the fraction of the fin volume contributing to the TES heat capacity. For NFC QETs the fins are taken to be fully proximitized and $\varepsilon_{\text{fin}} = 0$, so the active heat capacity is set by the W TES volume alone.

4.1.2 Experimental Setup and Data Acquisition

The NFC-SO device was operated at CUTE [61], a low-background test platform built around a dilution refrigerator that reaches a base temperature near 11 mK. The facility's overburden suppresses the cosmogenic background, and the cryostat is mounted on a vibration-isolated platform that decouples the payload from environmental and pulse-tube noise. The detector volume sits inside a light-tight cavity at the mixing-chamber stage, with cable feedthroughs designed to block infrared photons from leaking into the cold stages.

The CUTE payload for the LEE campaign comprises four detector boxes arranged in a vertical stack along the cold-finger axis. The lower two boxes carry the older detectors from the NEXUS R13 and R14 runs in NF-C, NF-H, and NF-E mask variants. A new top box, added for

this campaign, houses the NFC-SO and NFH-SO devices in a dedicated optical cavity that is decoupled from the lower stack by an intervening blank lid. An LED lid is mounted above the top box, and infrared filters are inserted between the boxes to suppress thermal radiation traveling along the stack.

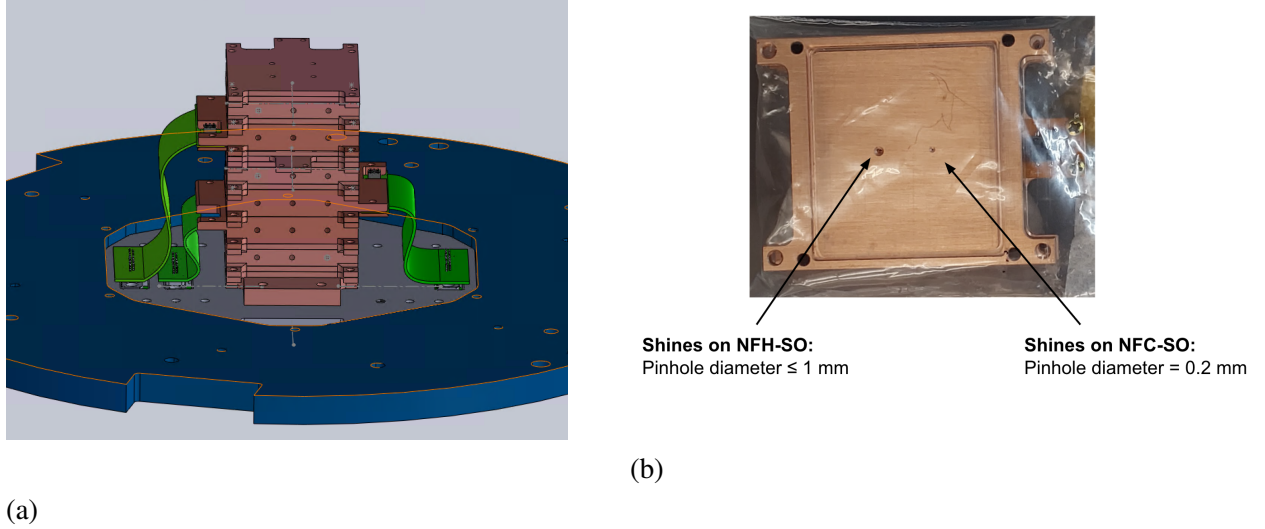


Figure 4-2. (a) Stacked detector payload mounted on the mixing-chamber plate. (b) Top-box housing for the two SO devices, with the LED pinholes drilled through the lid [26].

The top-box housing clamps the NFC-SO and NFH-SO devices side by side under a copper lid that defines the optical aperture for each. Two pinholes drilled through the lid couple the LED illumination to the detectors below, a 0.2 mm aperture above NFC-SO and a ≤ 1 mm aperture above NFH-SO. The aperture diameters set the photon spot size on each device face, which feeds into the G4CMP simulations of Section 4.3.

Calibration is provided by cryogenic LEDs installed directly inside the payload, emitting at 2.0468 ± 0.0002 eV (605 nm) as measured in liquid helium. Each LED is paired with an infrared filter that blocks thermal radiation from the warm LED body. Short voltage pulses drive the LEDs at a 10 Hz rate, releasing a Poisson-distributed number of photons per pulse with a mean controlled by the drive amplitude; a synchronized external trigger marks the rising edge of each pulse for offline alignment.

The TES is read out by a single-stage SQUID amplifier modeled on the SuperCDMS design, with the SQUID and shunt resistor placed on the 100 mK cold plate to keep parasitic resistance in the bias loop low. The CUTE configuration uses a $\sim 20\text{ m}\Omega$ shunt and a $2\text{ k}\Omega$ SQUID feedback resistor at the 4 K stage. SQUID feedback and signal digitization are handled by a SuperCDMS Detector Control and Readout Card. The high-voltage substrate bias is supplied by an isolated programmable source. Wiring between the mixing chamber and the warm electronics is split across four shielded looms, with the LED drive lines routed separately from the TES and SQUID lines to suppress crosstalk from the fast pulse edges into the readout.

Two calibration runs were taken back-to-back, labeled R39 and R40, with both inner and outer channels biased at 25% of R_n . Continuous time streams were digitized at 157 kHz through the DCRC. Calibration events were tagged by an external trigger synchronized to the LED drive pulse, and background events were identified offline by a threshold trigger applied to the gaussian-derivative-filtered time stream, with a 2048-sample ($\sim 13\text{ ms}$) window centered on each trigger saved for analysis.

The same bias point did not produce the same energy spectrum across the two runs. The discrete quanta peaks in R40 sit at systematically lower amplitudes than those in R39, despite identical base temperature. The most natural explanation is that the substrate temperature relaxed slowly between the two runs, leaving the operating point shifted at the start of R40 and drifting back over the course of the run. This observation foreshadows the substrate dynamics revealed in the long pulse tail of Section 3.4 and 4.6, where the same slow thermal channel produces the residual decay component the two-pole TES model cannot reproduce.

4.2 Templates and the LED Spectra

The calibration data feed two products that anchor the rest of the chapter. A pulse template, built from the average of clean single e-h pair events, sets the shape that the simulation must reproduce in Section 4.4.2. An LED energy spectrum, reconstructed from optimum-filter amplitudes, sets the energy peaks that the charge-transport simulation must reproduce in Section 4.5.

Both products rely on optimum filtering [62], the matched-filter estimator that returns the maximum-likelihood pulse amplitude under stationary Gaussian noise. Given a noise power spectral density and a signal template, the filter weights each frequency bin by the template-to-noise ratio and integrates the result, producing an amplitude estimate $\hat{A}$ for each trace. The full derivation is presented in the context of parameter extraction in Section 5.4.2. Two complementary modes are computed for each trace. The floating-time estimator A_{OFL} lets the trigger time vary within a window around the initial trigger, while the fixed-time estimator A_{OF0} pins the time at the trigger sample.

Constructing the template is a bootstrap. The first pass through the data uses a generic pre-template based on the rising-edge and decay timescales of a few raw averaged pulses. The optimum filter applied with this pre-template produces an amplitude estimate for every trigger, and the resulting histogram resolves the discrete e-h pair structure of the LED data. A clean sample of single e-h pair events is then selected by a sequence of quality cuts. A frequency-domain χ^2 cut to the pre-template rejects pulses with anomalous shape, suppressing pile-up and noise glitches. A pre-trigger baseline mean and slope cut removes events with drifting baselines that would bias the optimum-filter amplitude. A Δt cut between consecutive triggers removes pile-up not caught by the χ^2 cut. After these cuts, an OF amplitude window centered on the first peak in the histogram, defined by a Gaussian fit, isolates the single e-h pair events. The cuts and the peak window are computed on the channel sum and applied uniformly to the inner and outer readouts.

The accepted pulses are aligned on the LED trigger and averaged sample by sample. The Gaussian noise from the SQUID readout is uncorrelated between traces and averages down as $1/\sqrt{N}$ with the number of accepted events, while the underlying pulse shape adds coherently. The resulting trace is the average pulse template for the channel.

A second pass of the optimum filter, now with the average pulse template from each channel, yields the calibrated amplitudes for the LED energy spectrum. No quality cuts are applied to this spectrum. The LED firing time is recorded on a dedicated LEMO trigger channel during DAQ, and only events identified by this trigger are processed, so the spectrum contains

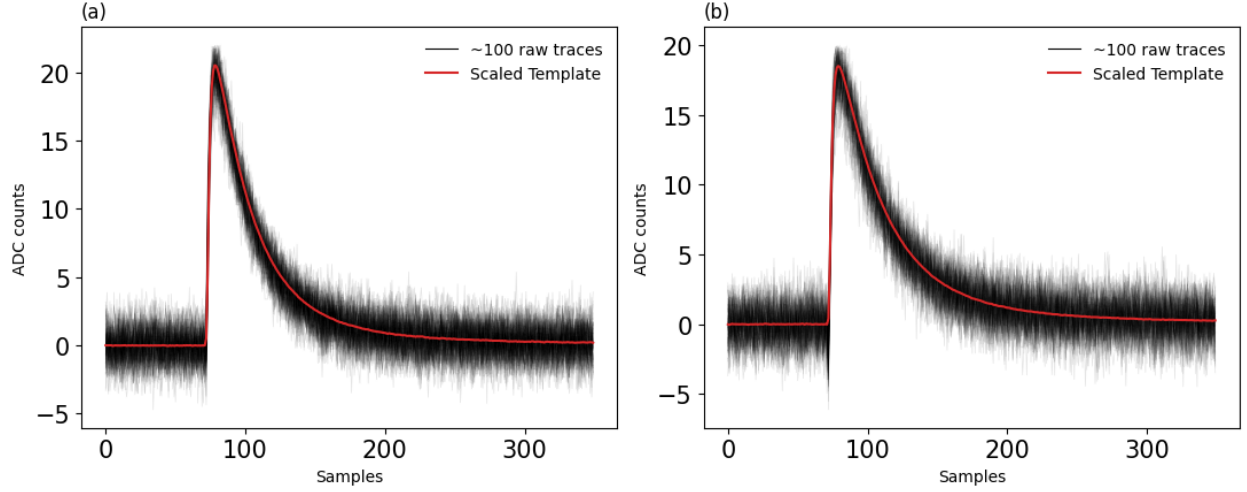


Figure 4-3. Single e-h pair pulses from the (a) inner and (b) outer channels of the NFC-SO detector. Individual traces (black) are overlaid with the average pulse template (red), scaled to the trace amplitude.

LED-correlated events by construction. The discrete peaks correspond to integer numbers of LED photons absorbed in the crystal, each producing a fixed number of e-h pairs that are accelerated across the bias to deposit $n_e e V_b$ of NTL phonon energy in addition to the recovered fraction of the photon energy itself.

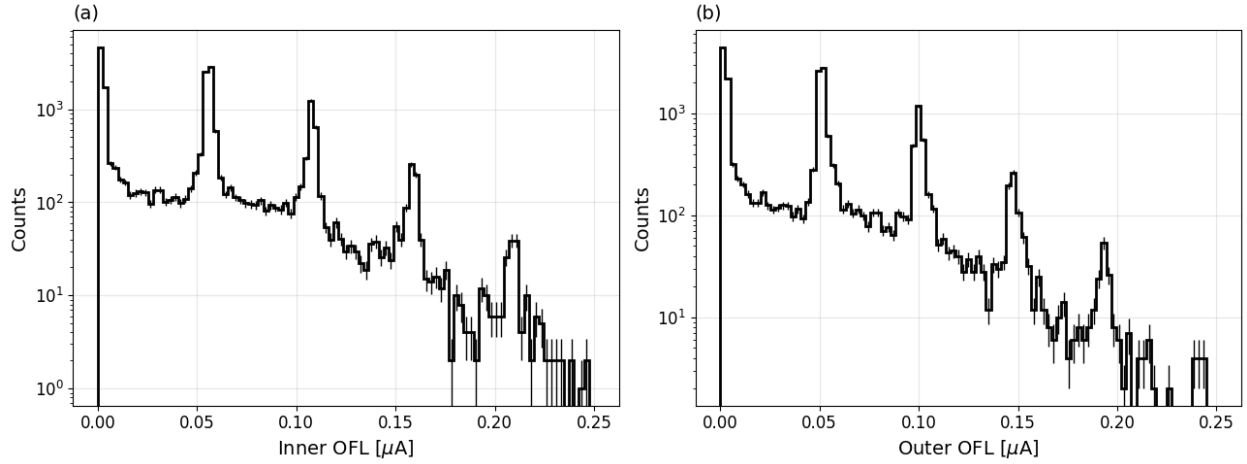


Figure 4-4. Optimum-filter amplitude spectra for the (a) inner and (b) outer channels of the NFC-SO detector under LED illumination operated at -35 V bias voltage.

The calibration and resolution of this detector have been characterized in detail in Ref. [63]. The baseline energy resolution reaches 612 ± 4 meV, well below the 1.14 eV silicon band gap, so

the discrete peaks in Figure 4-4 are fully resolved with negligible overlap between adjacent quantizations. From a fit to the LED spectrum, the bulk charge-trapping and impact-ionization probabilities per charge carrier are measured to be $p_{CT} = 27.2 \pm 0.2\%$ and $p_{II} = 1.1 \pm 0.4\%$. The same fit returns a fitted photon energy ~ 0.9 eV below the LED photon energy measured independently in liquid Helium, an offset that is consistent across two facilities and has no counterpart in the simulation, which assumes that every absorbed photon deposits its full energy as e-h pairs. The implication for the simulation comparison is twofold. The charge trapping and impact-ionization probabilities are the dominant detector-response physics that the charge-transport model must reproduce. The unaccounted-for 0.9 eV loss per charge excitation is a known discrepancy that the simulation is not built to explain, but its size is small compared to the e-h pair energy scale at the bias voltages used here, and it can be absorbed into an effective per-charge energy without distorting the peak structure that drives the fit.

4.3 Phonon Power Input

The TES ODE is driven by a phonon power $P_{ph}(t)$ that represents the energy delivered to the QET-instrumented surface as a function of time. This input is generated by a separate Geant4-based simulation, G4CMP [31], which propagates phonons and charge carriers through the crystal until they are absorbed at the sensor surface or lost. Two phonon-side parameters of the crystal simulation must be set before the input can be passed to the TES model. These are the QET absorption efficiency at the Si/Al interface and the TES sub-gap absorption probability that controls the late-time component of the phonon population. A third feature of the data, the slow onset of the rising edge in the average pulse template, must also be reproduced before any TES parameter can be inferred.

4.3.1 CrystalSim Parameters

The crystal simulation is configured by a set of phonon-side and charge-side parameters that together set the timing and amplitude of the spectrum delivered to the TES. The charge-transport parameters, which control trapping and impact ionization in the bias field, are tuned in Section 4.5

against the LED energy spectrum and are not relevant to the pulse shape fit. The phonon-side parameters are summarized below.

Phonon transport in the bulk. Athermal phonons emitted by the initial energy undergo isotropic scattering or anharmonic down-conversion. Both rates scale as a power of the phonon frequency,

$$\Gamma_{\text{scatter}} = B \nu^4, \quad \Gamma_{\text{anh}} = A \nu^5, \quad (4-1)$$

with material-specific coefficients A and B [42, 64]. For silicon, the default G4CMP values are used. The high-frequency phonons produced in the initial deposit start in a quasi-diffusive regime with mean free paths of order microns, then transition to ballistic propagation as anharmonic decay carries them down to the meV scale where the mean free path exceeds the crystal dimensions.

Surface interactions. When a phonon reaches a substrate surface, three outcomes are possible. The phonon may be absorbed by an aluminum QET fin, with a probability set by the AMM transmission discussed below; it may scatter specularly or diffusely off a bare silicon surface, with the diffuse fraction controlled by the surface-roughness parameter; or, if it has frequency below twice the aluminum gap and arrives at a sensor location, it may be absorbed directly by the tungsten film with a probability set by the TES sub-gap absorption parameter discussed at the end of this subsection. The default specular fraction at bare Si surfaces, the diffuse fraction, and the geometry of the sensor coverage are all carried over from the G4CMP defaults documented in Ref. [31].

Si/Al transmission. For a phonon incident on the Si/Al interface from the silicon side, the AMM transmission probability for a normally-incident longitudinal phonon is

$$\eta_{\text{AMM}} = \frac{4 Z_{\text{Si}} Z_{\text{Al}}}{(Z_{\text{Si}} + Z_{\text{Al}})^2}, \quad (4-2)$$

where $Z_i = \rho_i c_{s,i}$ is the acoustic impedance of material i and the densities and longitudinal sound speeds are temperature-independent at the sub-percent level between room temperature and 15 mK. With $\rho_{\text{Si}} = 2329 \text{ kg/m}^3$, $c_{\text{Si}} = 8433 \text{ m/s}$, $\rho_{\text{Al}} = 2700 \text{ kg/m}^3$, and $c_{\text{Al}} = 6420 \text{ m/s}$, the

impedances differ by about 12% and Equation 4-2 evaluates to $\eta_{\text{AMM}} \approx 99.6\%$ for longitudinal phonons. The corresponding value for transverse phonons is approximately 94%, and a density-of-states-weighted mode average is $\approx 94\%$. The Si and Al substrates are nearly impedance-matched, so the AMM is not the dominant loss mechanism for phonon collection. The total phonon collection efficiency in the device is much lower than this; the geometric coverage of the QET fins on the bottom face is well below unity, off-normal phonons see additional reflection from the Al-fin surfaces, and quasiparticle generation and diffusion in the Al fins introduce further losses that fall outside the AMM. The simulation is configured with a QET absorption efficiency of 54% and the value is held fixed, since it is degenerate with the TES parameters that affect the pulse amplitude such as the heat capacity and the transition width. This degeneracy can be seen directly from the TES electron equation 3-45, where the input power and the heat capacity together set the amplitude of the temperature response.

4.3.2 TES Sub-Gap Absorption

Phonons that reach the QET-instrumented surface with frequency below twice the aluminum superconducting gap cannot break Cooper pairs in the Al fins and are otherwise lost back to the substrate. The `TESSubgapAbsorption` parameter, which corresponds to the `subgapAbsorption` property of the G4CMP phonon electrode [31], controls the probability that such a phonon is absorbed directly into the tungsten film instead of reflecting back into the crystal. Increasing `TESSubgapAbsorption` steepens the slope of the late-time tail of the phonon power delivered to the TES, as shown in Figure 4-5. A small fraction of the early phonon population also arrives below the gap and is captured by the same mechanism, so the peak amplitude shifts slightly with the parameter, while the shape of the rising edge and the dominant early decay are essentially unchanged.

The slope inherited by the late-time tail of the simulated TES pulse traces the slope of the input phonon power directly. The TES electrothermal feedback time scale, set by $\tau_{\text{e-ph}}$ and $\mathcal{L}$, is much shorter than the late-time tail. Once the ETF region has decayed away, the TES is no longer driving its own response and simply tracks the residual sub-gap phonon power flowing in. The

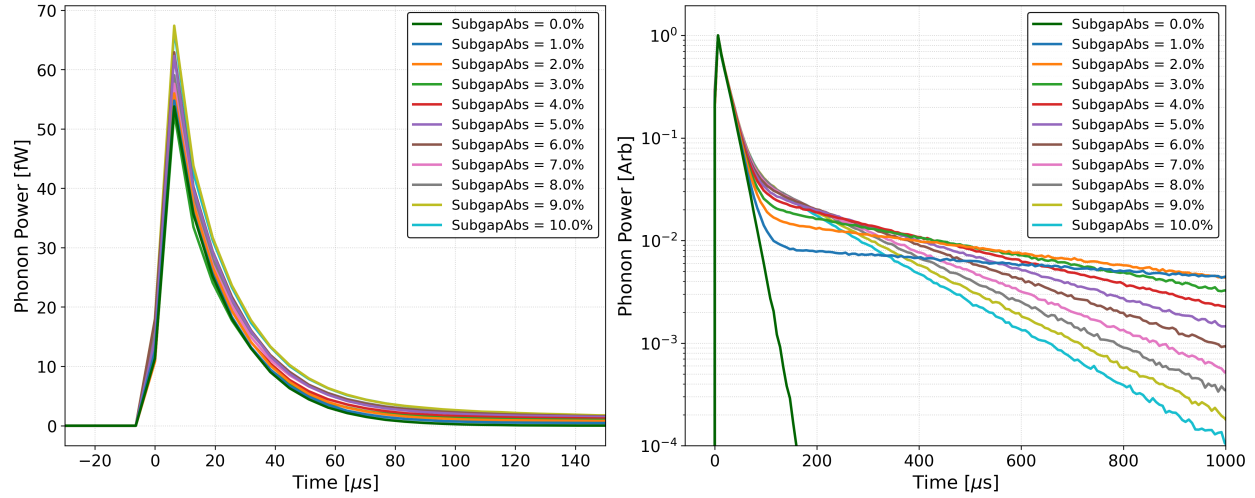


Figure 4-5. Average of 100 simulated phonon powers for the inner channel for TESSubgapAbsorption values from 1% to 10%. **Left:** absolute scale. **Right:** normalized to peak amplitude.

long tail of the simulated pulse is therefore decoupled from the TES parameters and depends only on TESSubgapAbsorption and on the slope of the input phonon population.

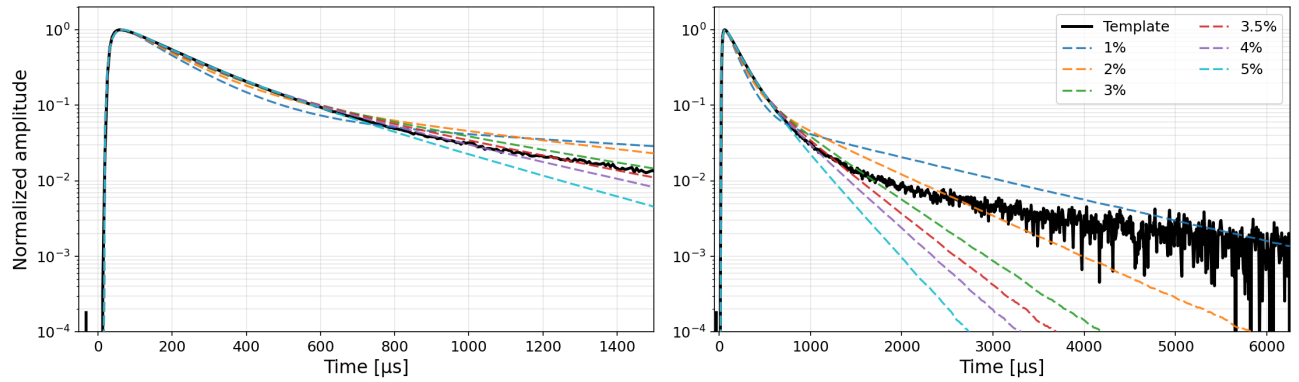


Figure 4-6. Simulated TES traces for TESSubgapAbsorption from 1% to 5% (dashed) overlaid on the average pulse template (solid black) for the inner channel, at fixed TES parameters. **Left:** early-time region. **Right:** extended time axis showing the long tail.

Figure 4-6 compares the simulated traces to the average pulse template at a fixed set of TES parameters, on two time scales. In the early-time region, all values of TESSubgapAbsorption between 1% and 5% reproduce the rising edge and the ETF decay up to about 700 μs. Beyond this point the simulated traces split, with each value producing a single straight slope in log scale set by the corresponding sub-gap phonon population. The data tail, in contrast, shows a curvature that

no single straight line can match. The closest agreement to the data is bracketed between 3% and 5%, and after several iterations of hand fitting the value `TESSubgapAbsorption` = 4.5% was settled on for the MCMC tuning of Section 4.4.4.

The residual tail mismatch persists for every value of `TESSubgapAbsorption` tested, regardless of the TES parameters used to generate the trace, and is not an artifact of poor sub-gap tuning. It is a feature of the data that the two-equation TES model and a single sub-gap source cannot reproduce, and it is addressed by the substrate-pole extension in Section 4.6. The parameter is held fixed at 4.5%.

4.3.3 Rising Edge Slow Onset

The average pulse template exhibits a finite rise time at the leading edge that is slower than the rise predicted by a single model run. The discrepancy is not a TES dynamics effect but an artifact of the template construction. The data acquisition system samples each pulse on a discrete clock, so the trigger is recorded at one of a finite set of sample times, while the actual phonon arrival at the QET-instrumented surface occurs at a continuous time that falls somewhere within one sample period of the trigger. For a position-dependent device, where different events can produce phonon arrivals with somewhat different shapes, the per-event variation contributes additional spread. Averaging many pulses with onset times distributed across the sample window broadens the leading edge of the average pulse template by an amount comparable to the sample period.

To test that this mechanism accounts for the observed slow onset, the simulation deposits a small parcel of energy in a disk of radius 0.5 mm at the bottom center of the detector face opposite the TES, generating a phonon arrival spectrum at the QET-instrumented surface. The phonon arrival spectrum is generated 100 times, each realization offset by a continuous time shift $\Delta t \in [0, \Delta t_{\text{bin}})$ drawn from a uniform distribution, where Δt_{bin} is the digitizer sample period. The 100 shifted realizations are averaged into a single phonon power input. Figure 4-7 compares one such shifted realization against the unshifted version, showing that the two are identical in shape

with only a sub-bin time offset; averaging many such realizations smooths the leading edge of the input.

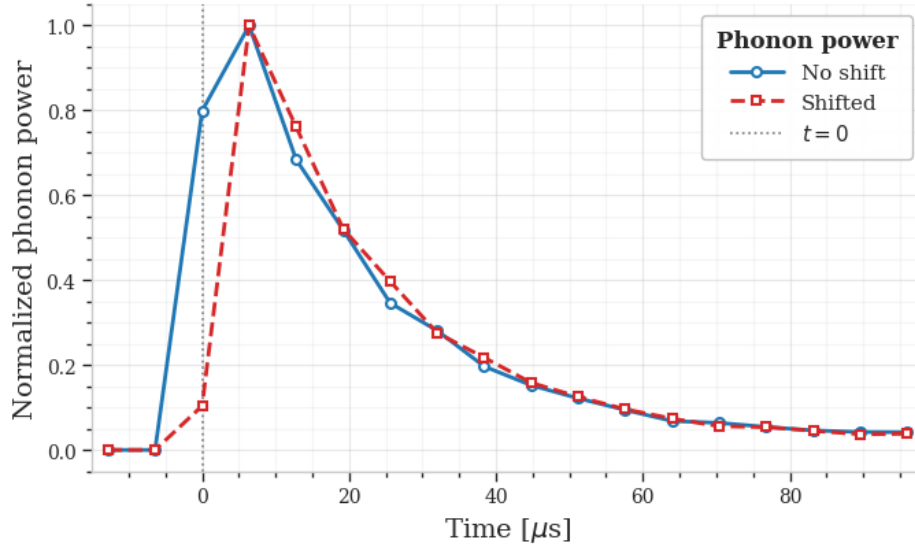


Figure 4-7. Shifted average phonon power (red, dashed) compared to the unshifted version (blue, solid), with the trigger sample at $t = 0$.

The averaged phonon power is then passed to the model, producing a single simulated TES trace that has the trigger jitter built in. Figure 4-8 compares the average pulse template, the simulated TES trace generated from the averaged phonon power, and the averaged phonon power itself, on linear scale with an inset zoomed on the rising edge. The simulated TES trace tracks the template through the leading edge, the peak, and the dominant decay; the phonon power input itself is much faster than either, peaking and decaying on the timescale of tens of microseconds before the TES electrothermal response builds up. The match through the rising edge confirms that the slow onset is a consequence of trigger discretization in the template construction rather than an intrinsic feature of the TES response.

This averaging procedure is applied to every simulated pulse used in the fits that follow, so any further mismatch between simulation and data must be attributed to physics that the TES model has not yet captured.

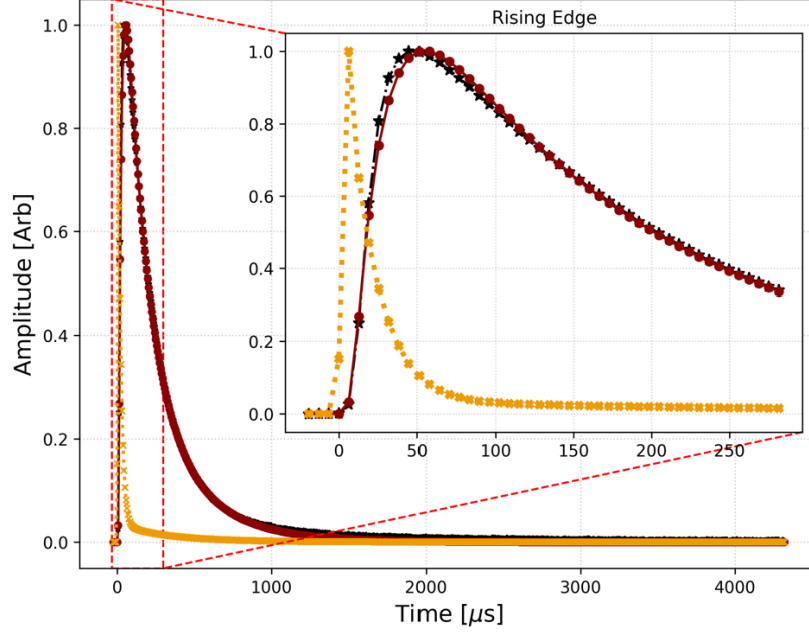


Figure 4-8. Average pulse template (black), the simulated TES trace generated from the averaged phonon power (dark red), and the averaged phonon power itself (orange), on linear scale. The inset zooms on the rising-edge region.

4.4 Pulse Shape Tuning of TES Parameters

With the phonon power input characterized in Section 4.3, the TES model can be exercised against the average pulse template to extract the parameters that govern the time response of the TES electron system. The fit is performed inside an MCMC framework that identifies which combinations of TES parameters are consistent with the measurement and that exposes the structural degeneracies of the single-pulse fit.

4.4.1 TESSim Model and Parameter Space

The model is a Python port of the SuperCDMS TESSim code that integrates the electrical and TES electron-temperature equations of Section 3.2.2.1, namely Equations 3-44 and 3-45, with the substrate temperature held fixed at the bath value. The work in this chapter is what motivated the extension to a dynamical substrate temperature, treated analytically in Chapter 3. The integration uses `scipy.integrate.ode` with the VODE method; the manipulation needed to put the equations into a numerically tractable form is given in Appendix C.

A pulse simulation begins from a phonon arrival spectrum $P_{\text{ph}}(t)$ generated as described in Section 4.3. The model deposits this power into the TES through the QET absorption efficiency, integrates the equations from a steady-state initial condition, and returns the resulting current trace $I_s(t)$.

The initial condition is set by either fixing the fractional resistance R/R_n or fixing the bias voltage $I_b R_{\text{sh}}$. In the fractional-resistance mode, the steady-state thermal balance reduces to

$$I_0^2 R(T_0, I_0) = \Sigma_W \mathcal{V}_W (T_0^n - T_b^n), \quad (4-3)$$

with $R(T_0, I_0) = (R/R_n) R_n$ and the substrate temperature taken equal to the bath. Solving this equation simultaneously with the resistance surface $R(T, I)$ yields the operating-point current I_0 and electron temperature T_0 that achieve the requested fractional resistance. In the fixed-bias mode, the electrical and thermal equations are solved simultaneously in steady state for I_0 , T_0 , and the resulting fractional resistance.

The TES parameters fall into three groups for the fit. The crystal-side parameters that shape $P_{\text{ph}}(t)$ are held fixed at the values determined in Section 4.3. The critical temperature T_c is measured from the $I_b I_s$ similar to Section 3.2.2.2 and is held fixed throughout. The remaining nine parameters (the four thermal parameters T_w , c_W , Σ_W , T_{subst} , the fin efficiency, and the four electrical parameters L , f_{R_n} , R_{sh} , R_p) are in principle either pinned to measured values or treated as free.

4.4.2 MCMC Framework

Parameter inference is performed with the affine-invariant ensemble sampler `emcee` [65]. At each step, $N_{\text{walk}} = 250$ walkers propose new positions in the parameter space; for each proposal the model is run with the proposed parameters and the resulting pulse is compared to the average pulse template through the log-likelihood

$$\ln \mathcal{L}(\theta) = -\frac{1}{2\sigma^2} \sum_{i \in W} [I_{\text{data}}(t_i) - I_{\text{sim}}(t_i; \theta)]^2, \quad (4-4)$$

where W is the fit window, θ is the parameter vector, and σ is the standard deviation of the pre-trigger baseline of the template, used as a single noise scale. The fit is performed on the absolute amplitude pulse template in physical units of current, so the pulse amplitude enters the likelihood directly. The fit window covers 110 samples spanning the rising edge through approximately 700 μs after the trigger; this upper bound excludes the long-tail region where Section 4.3.2 has already established that no parameter combination of the two-equation model reproduces the tail curvature apparent in the data. Priors on the free parameters are uniform on intervals wide enough to span the physically reasonable range for each. Walkers are initialized near literature values for tungsten and run for $N_{\text{prod}} = 2500$ production steps with no separate burn-in phase. Posterior summaries are reported using both the maximum a posteriori (MAP) sample, defined as the single sample with the highest posterior density, and the marginal median computed component-wise across the chain.

4.4.3 Global Fit and Parameter Degeneracies

The first attempt treats all nine non-crystal parameters as free. The walker chains in Figure 4-9 settle around their median values within the first few hundred steps, but several parameters develop secondary structure that persists for the duration of the run. The T_w and c_w chains separate into two distinct bands; the f_{R_n} , R_{sh} , and L chains show late-time drift toward higher values; and the fin efficiency develops a sub-population pulled toward zero.

These features carry through to the marginal posteriors in the corner plot of Figure 4-10. The T_w marginal is bimodal with a narrow peak near 0.2 mK and a broader secondary peak near 1.5–2 mK; the c_w marginal is bimodal with peaks at approximately 50 and 280 J/(m³K²). The 2D marginal for T_w – c_w shows a banana-shaped contour connecting the two modes, and the parameter correlation matrix in Figure 4-11 quantifies the structure with two distinct blocks, a thermal block dominated by the T_w – c_w anti-correlation at $\rho = -0.75$, and an electrical block in which L , f_{R_n} , and R_{sh} are positively correlated at $\rho = 0.66$ – 0.69 and all three are negatively correlated with the fin efficiency.

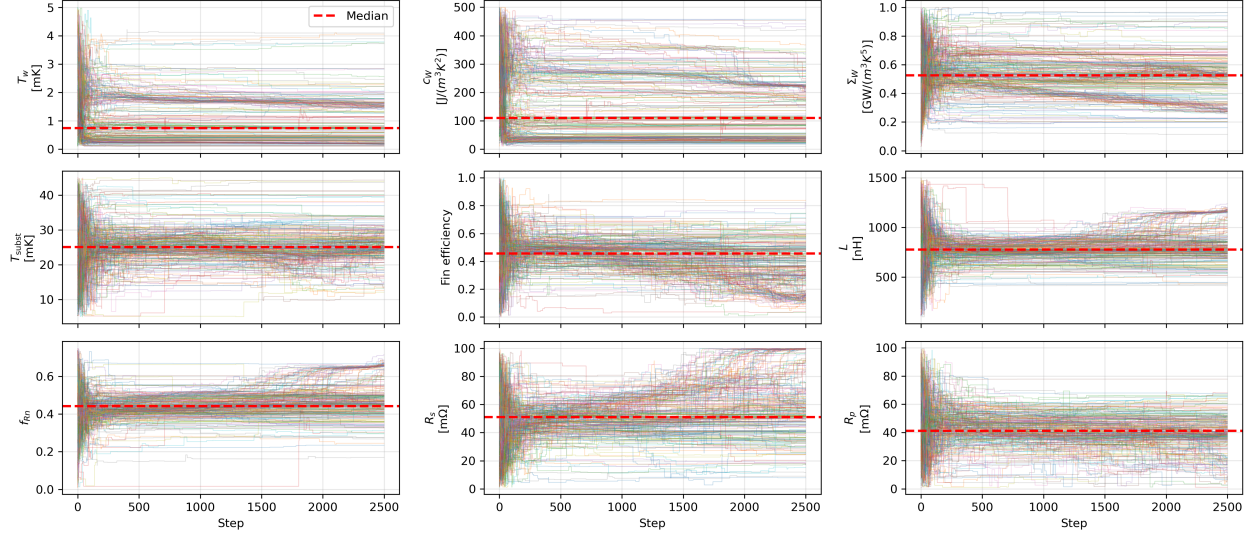


Figure 4-9. MCMC walker chains for the nine-parameter pulse shape fit. Red dashed lines mark the marginal medians.

The bimodality drives a quantitative disagreement between the two point estimates. The MAP and median pulses are compared to the average pulse template in Figures 4-12 and 4-13. The MAP pulse, with $T_w = 0.13$ mK and $c_w = 278$ J/(m³K²), captures the rising edge, peak, and dominant decay within the fit window and gives $\chi^2/\text{dof} = 263$. The median pulse, with $T_w = 0.75$ mK and $c_w = 110$ J/(m³K²), undershoots the peak amplitude and decays below the data through the early part of the pulse, with $\chi^2/\text{dof} = 1166$. The MAP sits inside the smaller, narrower mode of the joint posterior, while the marginal median lands in the low-density valley between the two modes, where the joint posterior density is small. The structure of the correlations and the bimodality both follow from the analytical form of the pulse time constants.

Thermal degeneracy. The ETF-shortened thermal time constant in the small-signal limit is $\tau_{\text{eff}} = \tau_{\text{e-ph}}/(1 + \mathcal{L})$, with $\tau_{\text{e-ph}} = C/G$ and loop gain $\mathcal{L} = \alpha P_{J,0}/(G T_c)$ from Section 3.3.1. At the operating point $f_{R_n} \approx 0.5$ on the broadened transition, the sensitivity parameter takes the simple form $\alpha \approx T_c/T_w$. For the operating points relevant here the loop gain is large, $\mathcal{L} \gg 1$, and

$$\tau_{\text{eff}} \approx \frac{C}{G \mathcal{L}} = \frac{C T_w}{P_{J,0}}. \quad (4-5)$$

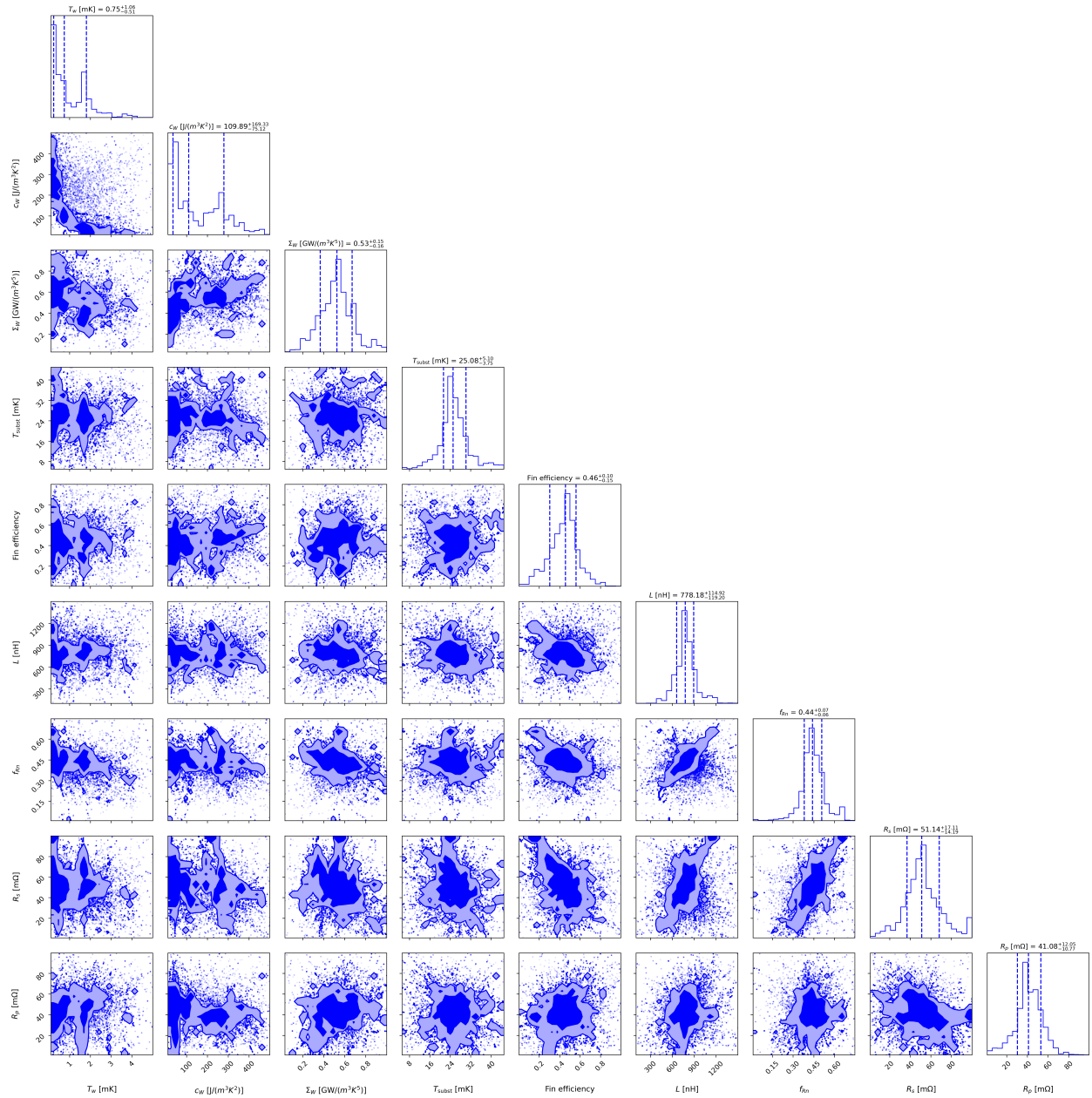


Figure 4-10. Corner plot of the nine-parameter pulse shape fit. Diagonal panels show 1D marginal posteriors; off-diagonal panels show 2D marginals with 1σ and 2σ enclosed-probability contours.

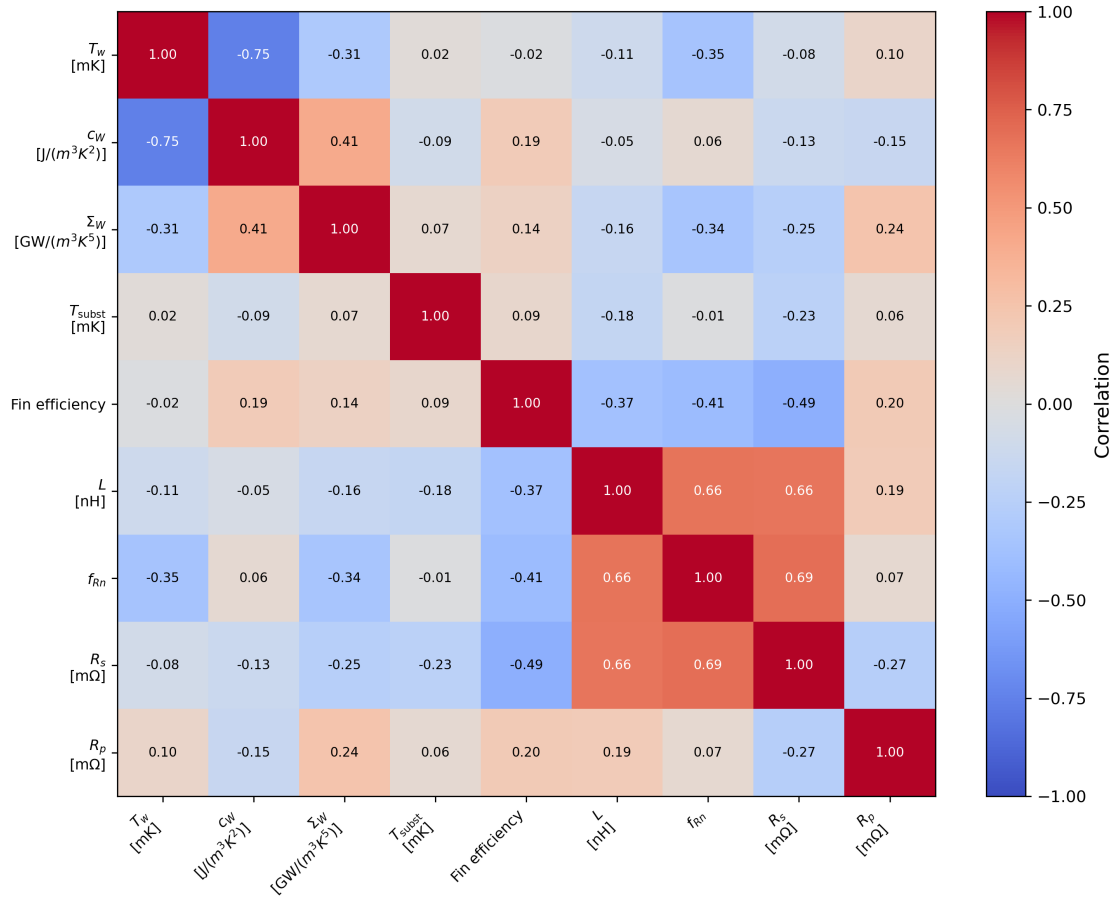


Figure 4-11. Parameter correlation matrix from the nine-parameter fit.

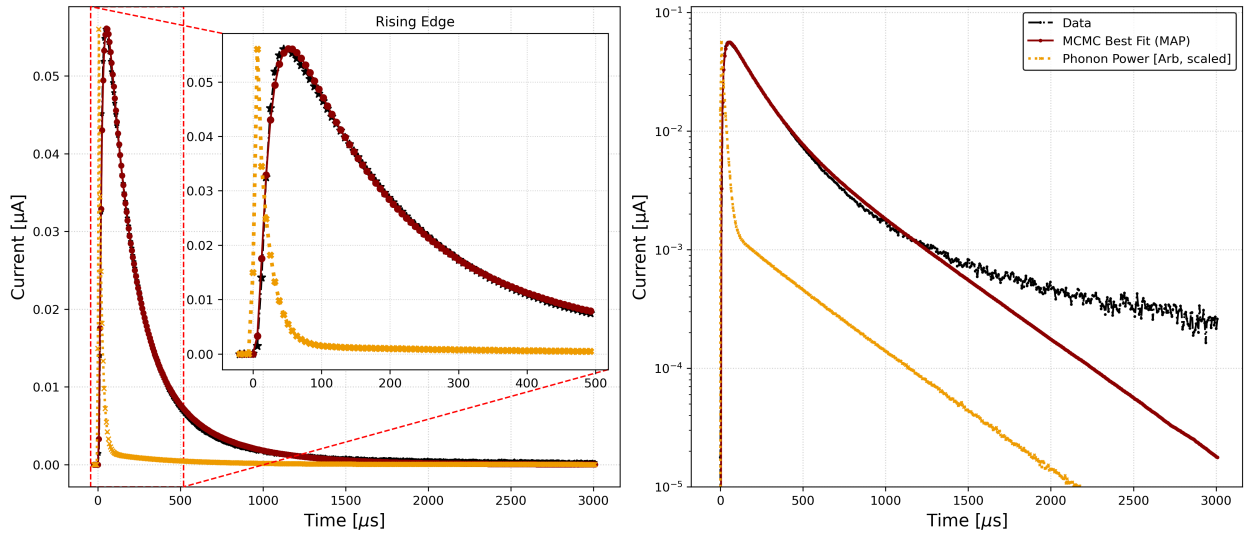


Figure 4-12. Average pulse template (black), MAP pulse (dark red), and phonon power input (orange) for the nine-parameter fit.

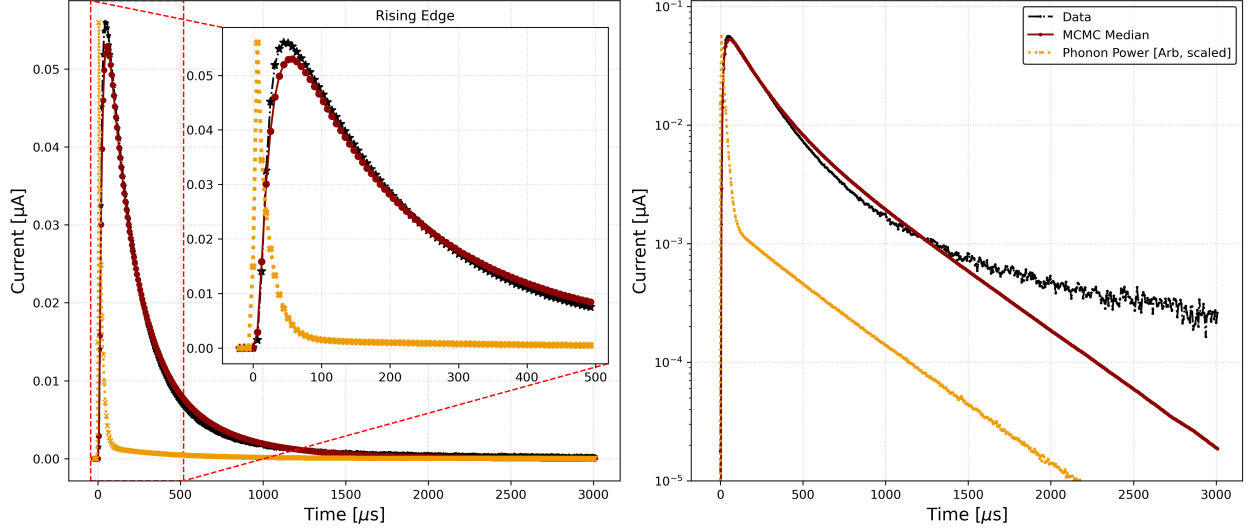


Figure 4-13. Same as Figure 4-12 but with the median pulse in place of the MAP.

Expanding the heat capacity $C = c_W \mathcal{V}_W T_c$ and the steady-state Joule power

$P_{J,0} = \Sigma_W \mathcal{V}_W (T_c^n - T_b^n)$ for a TES held at T_c on a bath at T_b ,

$$\tau_{\text{eff}} \approx \frac{c_W T_c T_w}{\Sigma_W (T_c^n - T_b^n)}. \quad (4-6)$$

The volume $\mathcal{V}_W$ drops out entirely, and τ_{eff} depends on c_W and T_w only through their product.

Trading c_W against T_w along the locus $c_W T_w = \text{const}$ leaves the dominant decay timescale unchanged. The anti-correlation $\rho(T_w, c_W) = -0.75$ in Figure 4-11 and the banana-shaped 2D contour in Figure 4-10 trace this locus directly. The electron-phonon coupling Σ_W enters as $1/\Sigma_W$ in the denominator, which aliases part of the T_w - c_W degeneracy onto Σ_W and gives the secondary correlation $\rho(c_W, \Sigma_W) = 0.41$.

This combination has a clean physical reading. The numerator can be rearranged as $\tau_{\text{eff}} \sim C T_w / (G T_c)$, with $C T_w \sim C \Delta T_{\text{sat}}$ the saturation energy of the TES electron system (the amount of heat the TES can absorb before the transition is driven through its full width) and $G T_c$ the maximum cooling power scale the electron-phonon link can sustain when the bath is held far below the transition. The effective thermal time constant is therefore the ratio of how much energy the TES can absorb to the maximum rate at which the link can dump it. Two parameter choices

that keep this ratio fixed produce indistinguishable decay timescales, regardless of how the saturation energy is partitioned between heat capacity and transition width.

Electrical degeneracy. The bare electrical time constant of the readout circuit is

$$\tau_{\text{el}} = \frac{L}{R_{\text{sh}} + R_p + f_{R_n} R_n} . \quad (4-7)$$

The MAP fit returns τ_{el} of order a few microseconds, shorter than the timescale on which the phonon power input is delivered to the TES. The circuit is fast enough to track the phonon source, so the rising edge of the pulse is set by the arrival of the phonon power rather than by the electrical response itself. This is a departure from the impulse-response analysis of Chapter 3, which assumed a delta-function deposition; the finite duration of the phonon arrival shifts control of the rising edge from the circuit to the input. Coordinated changes in L , f_{R_n} , and R_{sh} that preserve τ_{el} produce indistinguishable pulses, and the strong positive correlations between these three parameters reflect this. The fin efficiency enters through the absolute amplitude of the phonon power delivered to the TES; reducing it requires compensating increases in the operating-point resistance (f_{R_n}) and the load impedance (R_{sh}) to preserve the pulse amplitude, which gives the negative correlations between the fin efficiency and the electrical parameters.

Both degeneracies are structural. They follow from the analytical form of the time constants and from the duration of the phonon arrival, and they cannot be removed by collecting more data, running the chains longer, or tightening the priors. The way forward is to constrain the parameters that enter the degenerate combinations using measurements outside the pulse shape fit.

4.4.4 Fixing Measured Parameters and the Final Fit

The structural degeneracies are broken by pinning every parameter that can be determined outside the pulse shape fit. The circuit elements (L , R_{sh} , R_p , R_n) come from $I_b I_s$ and complex-impedance measurements, the substrate temperature T_{subst} from the mixing-chamber thermometer, and the critical temperature T_c from $I_b I_s$ at different bath temperatures. The volumetric heat-capacity coefficient is fixed at the tabulated tungsten value $c_W = 108 \text{ J}/(\text{m}^3 \text{K}^2)$,

the electron-phonon coupling at $\Sigma_W = 0.36 \text{ GW}/(\text{m}^3\text{K}^5)$ computed from the conductance through $G = n \Sigma_W \mathcal{V}_W T_c^{n-1}$, and the fin efficiency at $f_{\text{fin}} = 0$.

The transition width T_w does not admit a clean independent extraction. The $I_b I_s$ data of Section 3.2.2.2 can be converted to an $R(T)$ curve through the steady-state thermal balance

$$T = \left[T_{\text{subst}}^5 + \frac{I_s^2 R(I_b, I_s)}{\Sigma_W \mathcal{V}_W} \right]^{1/5}, \quad (4-8)$$

with $R(I_b, I_s) = (I_b - I_s) R_{\text{sh}}/I_s - R_p$, and the resulting curve fit to the broadened transition surface to recover T_w . The conversion requires Σ_W and T_{subst} as inputs, both of which are also inputs to the pulse model. Holding them at their best-measured values, $\Sigma_W = 0.36 \text{ GW}/(\text{m}^3\text{K}^5)$ and $T_{\text{subst}} = 15 \text{ mK}$, the tanh surface smears across the interval between T_{subst} and T_c and returns $T_w \sim 20 \text{ mK}$; letting them float to recover a tight fit drives Σ_W to half the measured value and T_{subst} to 30 mK (Figure 4-14). Either way, T_w is set by what is assumed about Σ_W and T_{subst} rather than by the data, and would be fed back into the pulse model that depends on the same two parameters. The extraction is circular.

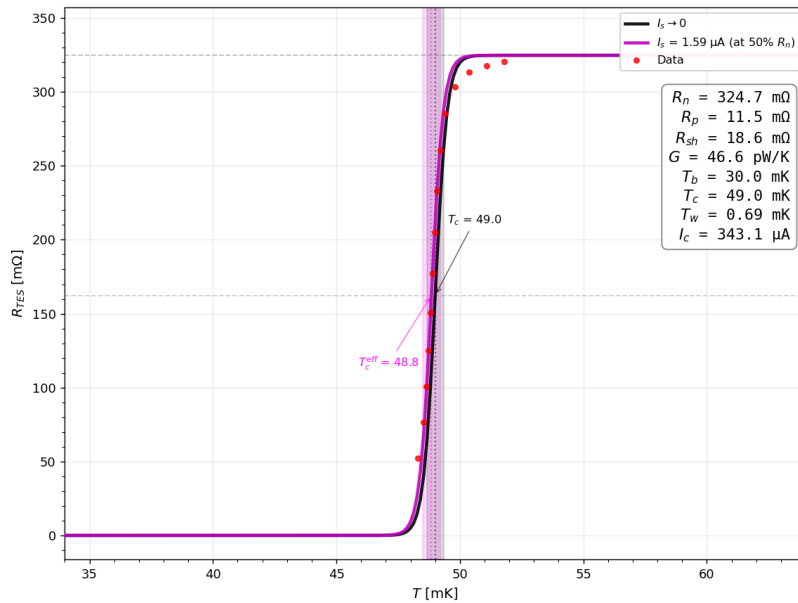


Figure 4-14. Tanh fits to the $R(T)$ curve derived from the $I_b I_s$ data through Eq. 4-8.

Independent of the circularity, the tanh surface only tracks the data in the linear central portion of the transition and misses the curvature at the upper and lower shoulders, indicating that it is at best a local description of $R(T, I)$ near $f_{R_n} = 0.5$. The LED pulses analyzed here remain inside this region throughout their excursion, so the tanh form is adequate for the pulse model.

The bias point f_{R_n} at the time of data acquisition is also not reliably accessible from outside the pulse fit, since the nominal value used to set up the bias is known to drift from the operating point recovered from the pulse shape. Both T_w and f_{R_n} are therefore retained as free parameters in the MCMC, with the fit window extended to cover the full pulse from the trigger.

The marginal posteriors for the two channels are shown in Figures 4-15 and pulse comparison in Figure 4-16. The fit converges and the marginal posteriors are unimodal with sub-percent fractional widths, so the MAP and median estimates coincide.

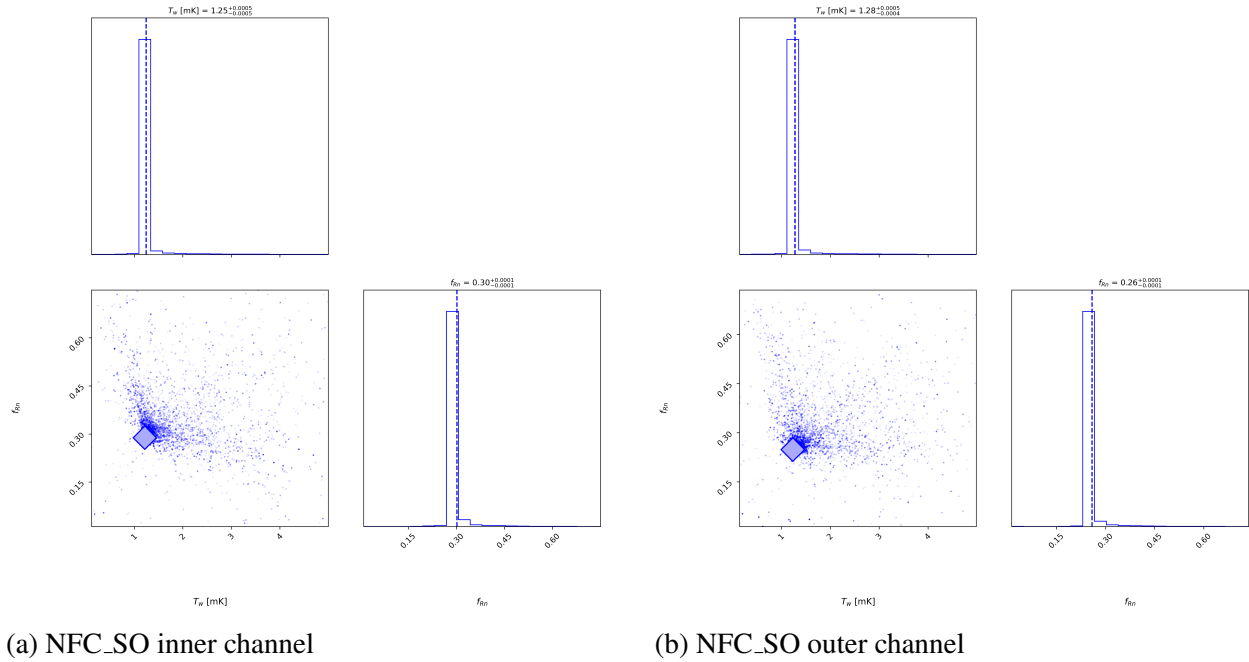
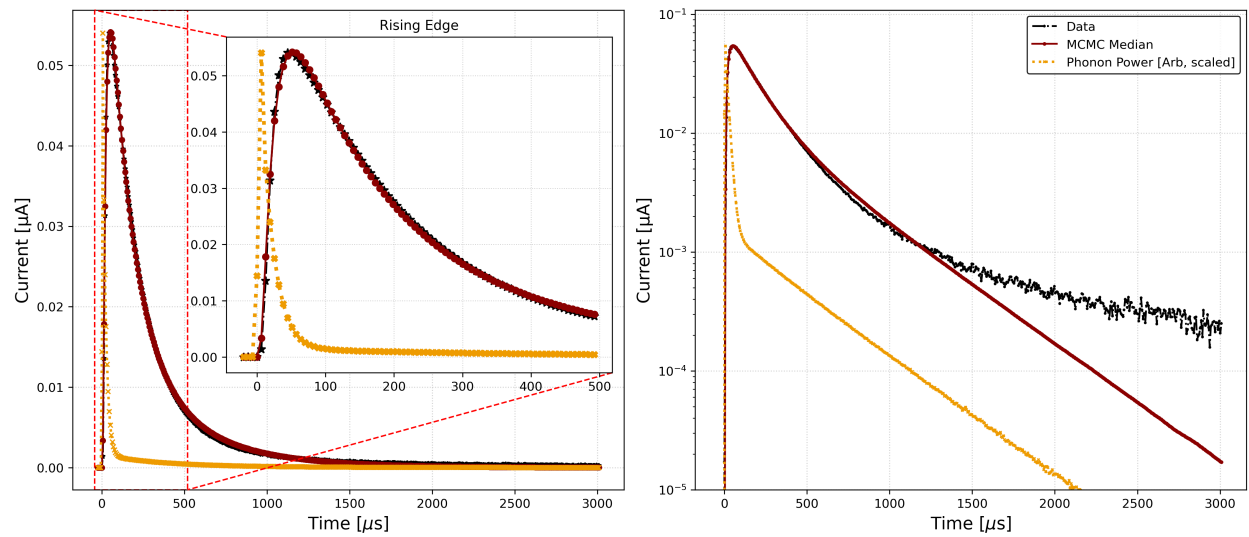
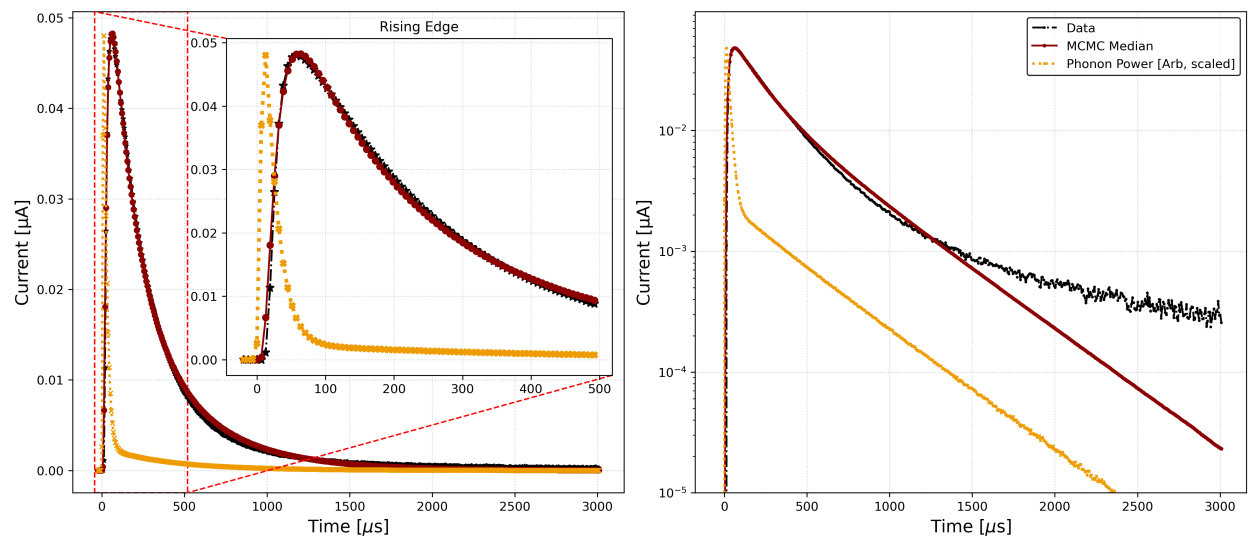


Figure 4-15. Marginal posteriors for T_w and f_{R_n} from the two-parameter pulse shape fit.

Beyond approximately $500 \mu\text{s}$ the logarithmic-scale comparison shows the model failing to match the data in both channels. The mismatch is the long-tail behavior identified in Section 4.3.2 and is addressed by the substrate-pole extension of Section 4.6. The two channels return T_w values



(a) NFC SO inner channel.



(b) NFC SO outer channel.

Figure 4-16. Average pulse template (black) and posterior-median pulse (dark red) on linear and logarithmic scales.

consistent at the 2% level and f_{R_n} values within 15% of one another. The full set of parameter values used in the rest of this work is collected in Table 4-2.

Table 4-2. Parameter values used in the TES simulation for the two NFC SO channels. Fixed parameters are held at their measured or tabulated values; fitted parameters are posterior medians from the two-parameter MCMC fit.

Parameter	Symbol	Inner channel	Outer channel	Source / Notes
<i>Fixed parameters</i>				
Tungsten heat-capacity coefficient [J m ⁻³ K ⁻²]	c_W	108	108	literature
Tungsten electron–phonon coupling [GW m ⁻³ K ⁻⁵]	Σ_W	0.36	0.36	from [63]
Substrate / bath temperature [mK]	T_{subst}/T_b	15	15	cryostat base
Fin collection efficiency	f_{fin}	0	0	model assumption
Readout inductance [nH]	L	363.3	363.3	SQUID readout
Shunt resistance [mΩ]	R_{sh}	18.6	18.6	$I_b I_s$
Parasitic resistance [mΩ]	R_p	13.2	13.2	$I_b I_s$
Critical temperature [mK]	T_c	49	49	$R(T)$ measurement
<i>Fitted parameters</i>				
Transition width	T_w [mK]	$1.25^{+0.0005}_{-0.0005}$	$1.28^{+0.0005}_{-0.0004}$	this work
Bias point fraction	f_{R_n}	$0.30^{+0.0001}_{-0.0001}$	$0.26^{+0.0001}_{-0.0001}$	this work

4.5 Charge Transport Tuning Against the LED Spectrum

4.5.1 Simulation Geometry and the CTII Model

The G4CMP framework [31] handles the full charge transport chain in the simulated detector. Photons are injected over a disk of radius 0.5 mm centered on the bottom face of the crystal, with depth distributed as $\exp(-z/\delta)$ with $\delta = 5 \mu\text{m}$ to model the absorption length of the LED wavelength in silicon. A bias of -35 V is applied at the bottom face with the top face grounded, so each absorbed photon produces an e - h pair in which the hole drifts toward the bottom face and the electron toward the top.

Figure 4-17 traces this for a simulation run with the default CTII parameters bundled with G4CMP for the HVeV detector. The creation sites (blue) cluster within the injection disk on the bottom face. The recombination sites (orange) populate two regions, a dense band on the top face

where most electrons terminate after drifting across the crystal, and a scattered population throughout the bulk corresponding to electrons captured by charge trapping before reaching the top surface. A smaller number of blue points appear in the bulk, marking hole creation sites from impact ionization. The distribution of recombination positions on the top face determines which fraction of the NTL phonon shower is collected by the inner channel and which by the outer.

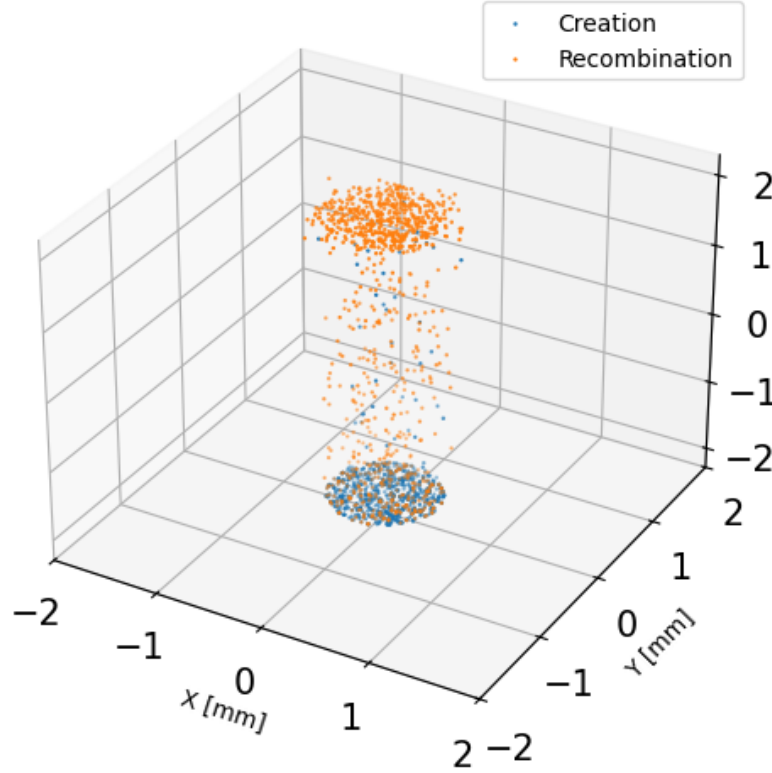


Figure 4-17. Creation sites (blue) and recombination sites (orange) of charge carriers from a G4CMP simulation of LED events at -35 V bias, using the default CTII parameters for the HVeV detector.

Drift through the crystal is not ballistic. Each carrier accelerates between scattering events, emits an NTL phonon at every scatter, and is re-accelerated by the field. The net displacement follows the field at a saturated drift velocity, but the total path length traversed is several times the substrate thickness. In a perfect crystal each $e-h$ pair eventually traverses the full bias and produces exactly eV_{bias} of NTL phonon energy, giving the spectrum its discrete peaks at integer multiples of n_{eh} .

Two processes spoil this quantization [66]. In charge trapping (CT), a propagating carrier is captured at an impurity site and ceases to emit NTL phonons before completing its traversal. In impact ionization (II), a carrier ejects an additional charge from a neutral impurity, adding a sub-thickness contribution to the NTL energy. Both processes fill the regions between the integer peaks. They are encoded in the simulation through species- and process-specific mean free paths ℓ_i , with i running over electron and hole trapping and the four II combinations $\{ee, eh, he, hh\}$.

The default CTII parameters bundled with G4CMP reproduce the qualitative features of the spectrum but are not specific to this detector. The HVeV laser calibration of [63], performed on a detector from the same silicon wafer family analyzed here, reports per-traversal probabilities $p_{CT} = 27.2 \pm 0.2\%$ and $p_{II} = 1.1 \pm 0.4\%$. Translating these probabilities into the mean free paths required by G4CMP requires the path length that a carrier actually traverses, which is set by the random-walk geometry rather than by the substrate thickness.

In silicon, electrons propagate along the L-valley conduction-band minima oblique to the applied field and undergo intervalley scattering, redirecting the trajectory each time the carrier jumps between valleys. The combination of oblique propagation, intervalley scattering, and the lateral excursions from NTL emission produces a wandering path several times longer than the substrate thickness. Holes, in principle, have a smaller but nonzero anisotropic component to their propagation [39, 40]; in practice, G4CMP models holes with an isotropic scalar effective mass [31], an approximation made because the degeneracy of the hole bands at the Γ point precludes the Herring-Vogt transformation used for electrons. Within this approximation, holes drift along the field and accumulate length only from the small-angle redirections of each NTL emission.

The resulting path-length distributions, extracted from a G4CMP run with CT and II disabled so that the unperturbed traversal lengths are exposed, are shown in Figure 4-18. Holes appear as a narrow spike near 14 mm, drift along the field plus modest lateral wander from NTL emission, with no contribution from the unmodeled anisotropy. Electrons spread broadly between 15 and 30 mm with a median of 26.34 mm, reflecting the full oblique-plus-intervalley random

walk. The factor of 3.5 to 6.6 between these path lengths and the substrate thickness sets the conversion between probabilities and mean free paths.

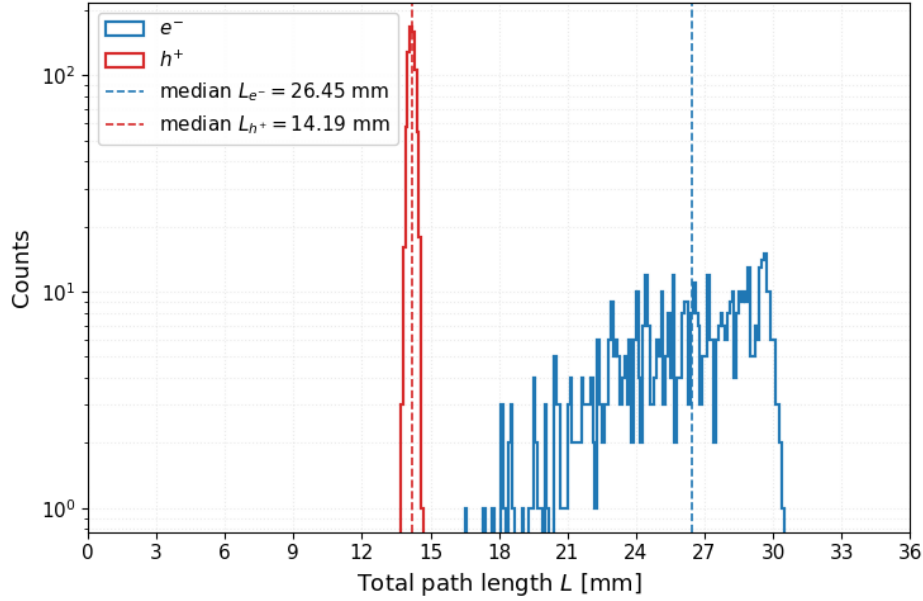


Figure 4-18. Total path length traversed by electrons and holes under -35 V bias, from a G4CMP simulation with charge trapping and impact ionization disabled. Dotted lines mark the median of each distribution.

The conversion follows from the exponential survival law for a carrier exposed over a path length L_{path} ,

$$p_i = 1 - \exp(-L_{\text{path}}/\ell_i) \iff \ell_i = -\frac{L_{\text{path}}}{\ln(1 - p_i)}. \quad (4-9)$$

Electron and hole trapping probabilities are taken equal, and each of the four II channels is assigned the full p_{II} , so that $p_{CTe} = p_{CTh} = p_{CT}$ and $p_{IIee} = p_{IIeh} = p_{IIhe} = p_{IIhh} = p_{II}$.

Applying Eq. 4-9 with the species-specific path lengths from Figure 4-18 gives the starting estimates. Trapping converts to $\ell_{CTe} = 83$ mm and $\ell_{CTh} = 44$ mm, and the II channels convert to $\ell_{IIee} = \ell_{IIeh} = 2380$ mm and $\ell_{IIhe} = \ell_{IIhh} = 1270$ mm.

These starting values were refined against the CUTE LED data, with the channel-level OFL spectrum providing the comparison. The electron trapping MFP was left at the value from the conversion; the hole trapping MFP was nudged from 44 to 46 mm; the II MFPs settled within 10% of their starting estimates. The final values are summarized in Table 4-3.

Table 4-3. Charge trapping and impact ionization mean free paths used in the simulation.

Process	Symbol	MFP [mm]
Electron trapping	ℓ_{CTe}	83
Hole trapping	ℓ_{CTh}	46
II (e creates e)	ℓ_{IIee}	2180
II (e creates h)	ℓ_{IIeh}	2180
II (h creates e)	ℓ_{IIhe}	1200
II (h creates h)	ℓ_{IIhh}	1200

4.5.2 Mean Photon Number and Comparison with CUTE Data

With the CTII parameters fixed at the values of Table 4-3, the relative heights of the integer n_{eh} peaks in the LED spectrum are controlled by the mean number of absorbed photons per trigger, λ . The photon count per pulse follows a Poisson distribution with mean λ , so the peak at k pairs is populated in proportion to $\lambda^k e^{-\lambda}/k!$. Higher λ pushes weight from the low- k peaks toward higher k , flattening the spectrum's exponential envelope. Scanning λ and comparing simulated to measured peak heights converges on $\lambda \sim 1.5$ as the value that best reproduces the channel-level OFL spectrum at -35 V bias.

Figure 4-19 shows the optimal-filter amplitude (OFL) spectra for the inner and outer channels with the final CTII parameters and $\lambda = 1.5$. Peak positions, peak heights, and the between-peak floor are reproduced on both channels through the first five e - h pair peaks. The agreement extends across the full range of the LED signal, with the simulation tracking the data through three orders of magnitude in normalized density.

The spectral agreement validates the CTII parameters and the value of λ . The partition of energy between the two channels is a separate observable, sensitive to the location of the absorption footprint rather than to the total NTL energy produced. Figure 4-20 shows the distribution of the asymmetry parameter $(I_{\text{in}} - I_{\text{out}})/(I_{\text{in}} + I_{\text{out}})$. Both distributions peak near zero with comparable central widths, indicating that the simulation captures the dominant feature, with events near the LED spot deposit their NTL energy on the readout face symmetrically between the inner and outer channels. The data shows a broader tail extending to both sides that the simulation does not reproduce.

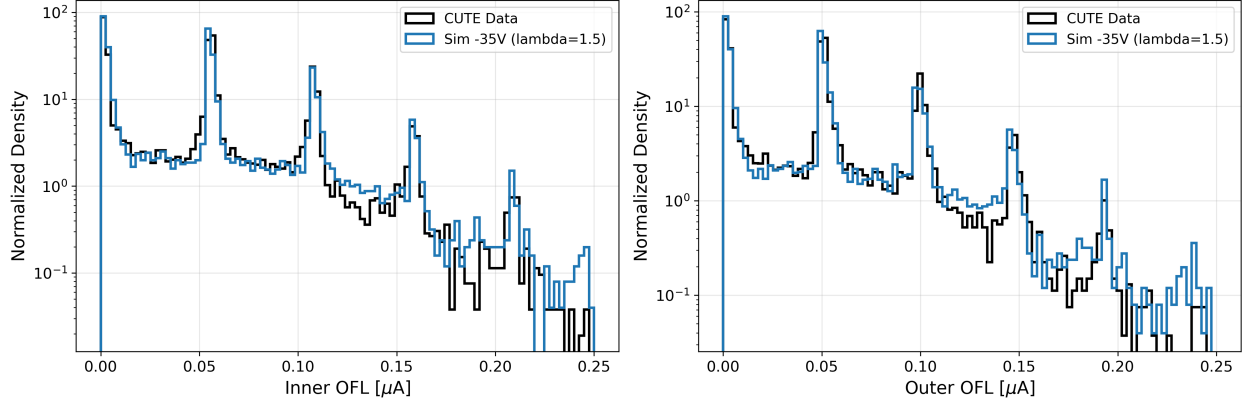


Figure 4-19. Maximum amplitude spectra for the inner (left) and outer (right) channels at -35 V bias. Optimal filter spectra from HVeV@CUTE data (black) and simulation with $\lambda = 1.5$ and the CTII parameters of Table 4-3 (blue).

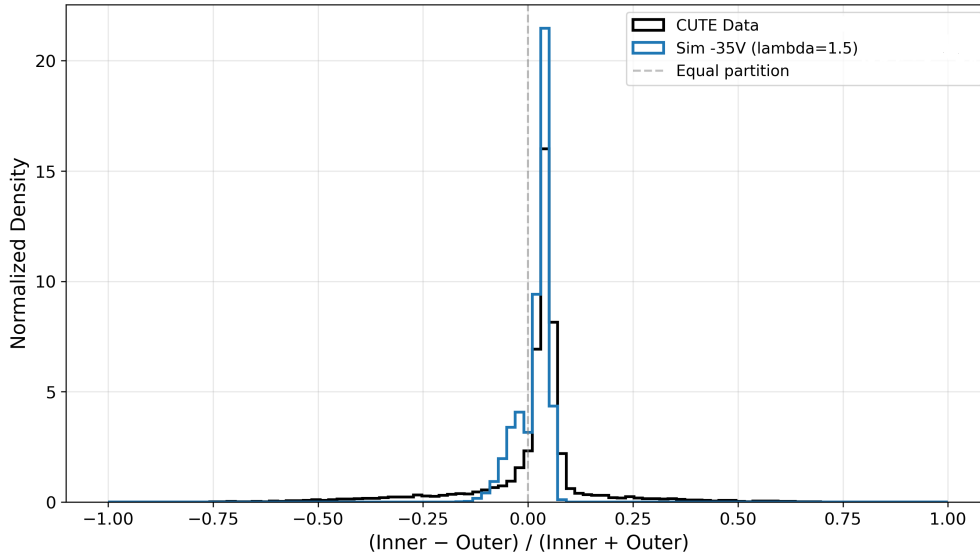


Figure 4-20. Asymmetry parameter $(I_{\text{in}} - I_{\text{out}}) / (I_{\text{in}} + I_{\text{out}})$ between the inner and outer channels, for CUTE data (black) and simulation (blue).

The injection geometry used in the simulation, a 0.5 mm disk centered on the bottom face, is a first-order estimate of the LED spot. In the CUTE setup the LED is housed inside the detector tower together with the crystal, and photons can bounce around in the LED box or the detector box and hit the detector in different places other than the center of the bottom face. A full model of this geometry would distribute the absorption sites over a wider and generally asymmetric region of the bottom face, broadening the partition tails in the simulation to match the data. The

charge-transport physics is unchanged by this; only the input distribution of absorption sites changes. Implementing the full LED-tower geometry is left to future work.

The simulation reproduces the energy spectrum that drives the λ and CTII tuning, and captures the central feature of the channel partition. The remaining parameter set is now fixed.

4.6 Reproducing the Long Tail with a Substrate Thermal Block

4.6.1 The Two-Thermal-Block Model

The single-thermal-block TES model of Section 4.4.4 reproduces the rising edge and the bulk of the exponential decay of the LED pulse but falls below the data past approximately $1500 \mu\text{s}$. The mismatch is a slow pole the model cannot supply. A single thermal block has one thermal timescale, $\tau_{\text{ep}} = C_{\text{TES}}/G_{\text{ep}}$, and no choice of parameters within that block can produce a second decay component. The natural extension is to promote the substrate from a passive boundary at fixed T_b to an active thermal block with its own temperature and heat capacity, coupled to the TES through the electron-phonon link and to the bath through a Kapitza link. This adds the additional pole needed to match the long tail.

The model is known in the TES literature as the two-thermal-body or two-thermal-block model. For brevity we refer to it as 3T2G, indicating three temperatures (W electrons, Si phonons, bath) coupled by two thermal conductances (electron-phonon and substrate-bath). The bath temperature is held fixed and acts as a boundary condition; the three coupled equations governing the dynamics are

$$L \frac{dI}{dt} = V_b - I [R_p + R_{\text{sh}} + R_{\text{TES}}(T, I)], \quad (4-10)$$

$$C_{\text{TES}} \frac{dT}{dt} = I^2 R_{\text{TES}}(T, I) - \Sigma_{\text{W}} \mathcal{V}_{\text{W}} (T^n - T_{\text{subst}}^n) + P_{\text{ph}}(t), \quad (4-11)$$

$$C_{\text{subst}} \frac{dT_{\text{subst}}}{dt} = \Sigma_{\text{W}} \mathcal{V}_{\text{W}} (T^n - T_{\text{subst}}^n) - K_{s-b} (T_{\text{subst}}^m - T_b^m), \quad (4-12)$$

with $n = 5$ for tungsten electron-phonon coupling and $m = 4$ for the Kapitza link. The substrate equation introduces two new parameters, C_{subst} and K_{s-b} , neither of which is constrained by the single-block fit.

4.6.2 Heat Capacities

The substrate heat capacity follows from the Bose-Einstein occupation of lattice phonons combined with the Debye density of states. In the Debye model the crystal supports $3N$ vibrational modes distributed quadratically in frequency up to a cutoff ω_D ,

$$g(\omega) = \frac{9N}{\omega_D^3} \omega^2, \quad 0 \leq \omega \leq \omega_D, \quad (4-13)$$

with each mode contributing the thermal Bose-Einstein energy $\langle E(\omega, T) \rangle = \hbar\omega / (e^{\hbar\omega/k_B T} - 1)$.

Integrating over the spectrum gives the internal energy

$$U(T) = 9Nk_B T \left(\frac{T}{\Theta_D} \right)^3 \int_0^{\Theta_D/T} \frac{x^3}{e^x - 1} dx, \quad (4-14)$$

with $\Theta_D = \hbar\omega_D/k_B$ the Debye temperature. Differentiating with respect to T and taking the $T \ll \Theta_D$ limit, where the upper integration limit goes to infinity and the integral evaluates to $4\pi^4/15$, yields the familiar Debye T^3 law,

$$c_V(T) = \frac{12\pi^4}{5} nk_B \left(\frac{T}{\Theta_D} \right)^3 \equiv \beta T^3, \quad \beta = \frac{12\pi^4 nk_B}{5\Theta_D^3}, \quad (4-15)$$

with n the atomic number density.

For silicon, the calorimetric measurements of [67] give $\Theta_D = 636$ K, leading to $\beta = 6.3 \times 10^{-7} \text{ J}/(\text{cm}^3 \text{K}^4)$. The HVeV substrate is a $1 \times 1 \times 0.4 \text{ cm}^3$ silicon crystal, and the implementation in TESSim treats C_{subst} as temperature-independent. Evaluating βT^3 at the bath temperature $T_b = 15 \text{ mK}$ gives

$$C_{\text{subst}}^{\text{Debye}}(T_b) = \beta V T_b^3 = 0.85 \text{ pJ/K}. \quad (4-16)$$

The TES heat capacity follows from the electronic specific heat of tungsten, $c_e = c_W T$ with $c_W = 108 \text{ J}/(\text{m}^3 \text{K}^2)$. Tungsten is superconducting below T_c , and at T_c the BCS specific-heat jump

enhances the electronic heat capacity by a factor $f_{SC} = 2.43$. With a TES volume per channel of $\mathcal{V}_W = 7.7 \times 10^{-15} \text{ m}^3$ ($40 \text{ nm} \times 158 \text{ } \mu\text{m} \times 2.4 \text{ } \mu\text{m}$, 506 TESs per channel), this gives

$$C_{\text{TES}}(T_c) = f_{SC} c_W T_c \mathcal{V}_W \approx 99 \text{ fJ/K}, \quad (4-17)$$

roughly four orders of magnitude below the substrate heat capacity.

4.6.3 Fit Results

Equations 4-10–4-12 are fit to the average HVeV pulse template with C_{subst} , K_{s-b} , and Σ_W floated; the remaining parameters are fixed at the values of Table 4-2. Figure 4-21 compares the one-block and two-block fits to the data on the rising edge and the decay tail. The single-block model reproduces the rising edge and matches the early decay through approximately the first millisecond, after which it falls below the data. The two-block fit follows the data through the full decay tail.

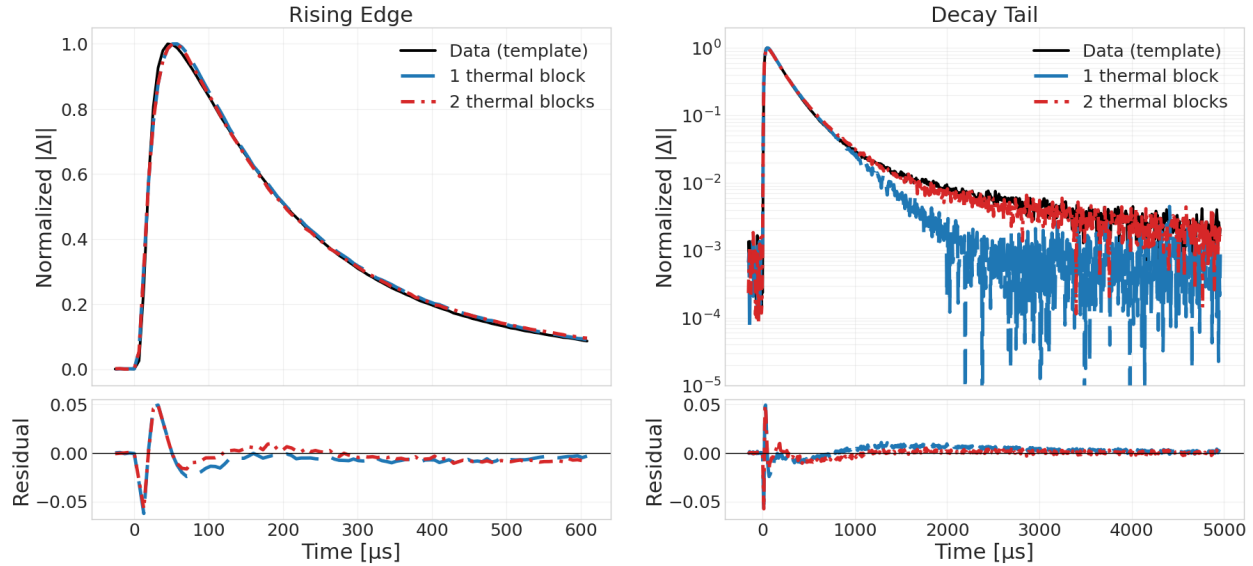


Figure 4-21. HVeV pulse template (black) compared with the one-block (blue) and two-block (red) fits on the rising edge (left) and decay tail (right), with residuals below.

The fit parameters and the derived thermal conductances are collected in Table 4-4. The two-block fit returns $C_{\text{subst}} = 0.30 \text{ pJ/K}$ and $K_{s-b} = 250 \text{ nW/K}^4$, corresponding to a Kapitza conductance $G_{K,b} = 4K_{s-b}T_b^3 = 3.4 \text{ pW/K}$. The electron-phonon coupling shifts from

$\Sigma_W = 3.6 \times 10^8 \text{ W}/(\text{m}^3 \text{K}^5)$ in the one-block fit to $6.0 \times 10^8 \text{ W}/(\text{m}^3 \text{K}^5)$ in the two-block fit, with $G_{\text{eph}}(T_c)$ shifting by the same factor.

Table 4-4. Parameters and derived thermal conductances for the one-block and two-block fits to the HVeV pulse. The substrate heat capacity and Kapitza conductance are new degrees of freedom in the two-block model; the Debye prediction for C_{subst} at T_b is shown for comparison.

Parameter	Symbol	1-block	2-block
Transition width	T_w [mK]	1.27	1.27
TES heat capacity	$C_{\text{TES}}(T_c)$ [fJ/K]	99	99
Electron-phonon coupling	Σ_W [$10^8 \text{ W}/(\text{m}^3 \text{K}^5)$]	3.6	6.0
Electron-phonon conductance	$G_{\text{eph}}(T_c)$ [pW/K]	80	133
Substrate heat capacity (fit)	C_{subst} [pJ/K]	–	0.30
Substrate heat capacity (Debye)	$C_{\text{subst}}^{\text{Debye}}(T_b)$ [pJ/K]	–	0.85
Kapitza coefficient	K_{s-b} [nW/K ⁴]	–	250
Kapitza conductance	$G_{K,b}$ [pW/K]	–	3.4

4.6.4 Discussion and Limitations

The two-block fit reproduces the long tail that the single-block model misses, which points to the substrate phonon system as the missing degree of freedom. What the fit cleanly measures is the substrate relaxation time, not the heat capacity and thermal links that make it up, which are degenerate in the pulse.

The relaxation time is consistent with the fit once the substrate temperature is set correctly. The substrate block relaxes with

$$\tau_{\text{subst}} = \frac{C_{\text{subst}}}{G_K + \eta G_{\text{eph}}}, \quad (4-18)$$

in which both conductances climb steeply with the substrate temperature, $G_{\text{eph}} \propto T^4$ and $G_K \propto T^3$. At the bath temperature the electron-phonon and Kapitza terms are only about 1.2 and 3.4 pW/K, giving $\tau_{\text{subst}} \approx 65$ ms, far longer than the few-millisecond tail, so the substrate must sit well above the bath. Solving the steady state of Eq. 4-12 with the biased TES near T_c ,

$$\Sigma_W \mathcal{V}_W (T_c^n - T_{\text{subst}}^n) = K_{s-b} (T_{\text{subst}}^m - T_b^m), \quad (4-19)$$

gives $T_{\text{subst}}^{\text{op}} \approx 42$ mK. There the electron-phonon and Kapitza conductances rise to about 72 and 74 pW/K and the time constant falls to $\tau_{\text{subst}} \approx 2$ ms, the right order of magnitude for the observed

tail. The fit therefore returns a substrate relaxation time on the scale of the long tail at a sensible operating temperature, rather than an arbitrary slow pole.

Only the combination in Eq. 4-18 enters the pulse shape, so the heat capacity and the thermal links are not pinned on their own. The fitted $C_{\text{subst}} = 0.30$ pJ/K sits more than an order of magnitude below the literature Debye value at the operating temperature, ≈ 19 pJ/K, and the electron-phonon coupling shifts by nearly a factor of two between the one- and two-block fits. Since the steady state fixes only the combined Joule-to-bath heat flow, separating C_{subst} , Σ_W , and K_{s-b} needs an independent measurement, and Section 5.5 outlines one for K_{s-b} . Letting C_{subst} vary with temperature over the pulse as well as including the other channel effects are left for future work.

CHAPTER 5

SUPERCDCMS DETECTOR TESTING AND OPERATION

The high-voltage operation that gives the HV detectors their reach also sets a standard their hardware has to meet. Biasing the crystal at the full operating voltage amplifies each electron-hole pair into a large phonon signal, lowering the effective threshold to the scale of a single pair and opening the sub-GeV and electron-recoil channels that motivate the detector. That reach is only realized if the sensors hold a stable operation at that bias and if their noise stays below the small signals being resolved. Whether the detectors meet that standard is a question of measurement.

This chapter examines that question through the testing operation of a tower of HV detectors at SNOLAB. The studies treat the stability of the detector operation, the noise that limits the energy resolution, and the thermal properties of the detector that the model of the preceding chapters took as given.

5.1 The SuperCDMS Experiment and the CUTE Facility

SuperCDMS SNOLAB is the latest generation of the Cryogenic Dark Matter Search, a direct-detection program that has been refining cryogenic crystal detectors for low-mass dark matter across successive experiments. It operates deep underground at SNOLAB, a laboratory hosted in the Vale Creighton mine near Sudbury, Ontario [25]. Rare-event searches of this kind have to be shielded from cosmic rays, whose secondaries would otherwise produce a continuous background of interactions and, through spallation, generate high-energy neutrons that mimic a dark matter recoil. SNOLAB provides that shielding through depth. At roughly two kilometres of rock, an overburden equivalent to about 6000 metres of water, the cosmic-ray muon flux is reduced by several orders of magnitude relative to the surface, and the hadronic component is removed entirely. The residual muon-induced and radiogenic backgrounds are suppressed further by layered shielding around the detectors.

The experiment deploys two detector types and two target materials. The iZIP detectors read out both ionization and phonons and discriminate electron from nuclear recoils, which serves the search above a few GeV. The HV detectors, the subject of this chapter, read out phonons alone and are biased near 100 V (90 V achieved at CUTE) so that the NTL amplification of the drifting

charge dominates the phonon signal, lowering the effective threshold to the scale of a single electron-hole pair and opening the sub-GeV reach. The detectors are cylindrical crystals 10 cm in diameter and 3.3 cm thick, 1.4 kg for germanium and 0.6 kg for silicon, and are deployed in towers of six. A tower holds its detectors in a stack and provides the thermal stages, wiring, and first-stage amplifier electronics that connect them to the cryostat, as shown in Figure 5-2.

The towers are housed in the SNOBOX, a nest of six copper cans suspended from the dilution refrigerator, each can mapped to a thermal stage of the cooling chain. Surrounding the SNOBOX are concentric layers of shielding, polyethylene to moderate and absorb neutrons and lead to attenuate gamma rays, with outer water tanks against the cavern neutron flux and a radon purge barrier to keep radon out of the inner volume. The whole assembly rests on a seismic platform. Figure 5-1 shows the arrangement.

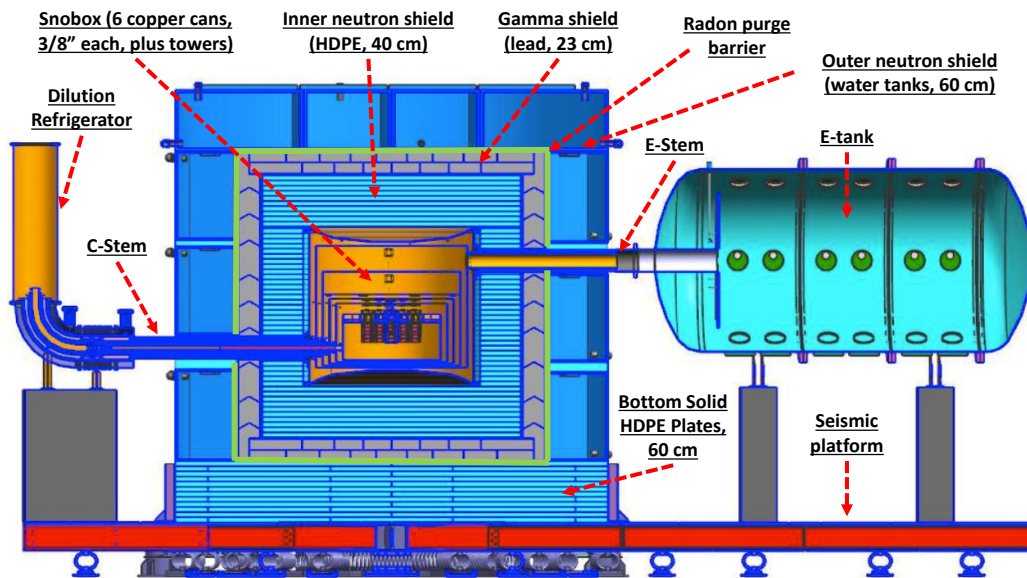


Figure 5-1. The SuperCDMS SNOLAB apparatus. The detector towers sit inside the SNOBOX cryostat, surrounded by concentric layers of polyethylene neutron shielding, a lead gamma shield, and outer water tanks, with the whole assembly mounted on a seismic platform deep underground. Adapted from [25].

Before a tower is committed to the shielded experiment, its detectors are studied at CUTE under conditions close to those they will later see underground. Testing them in a surface laboratory was not an option. The detectors are large and have a low threshold, so an unshielded

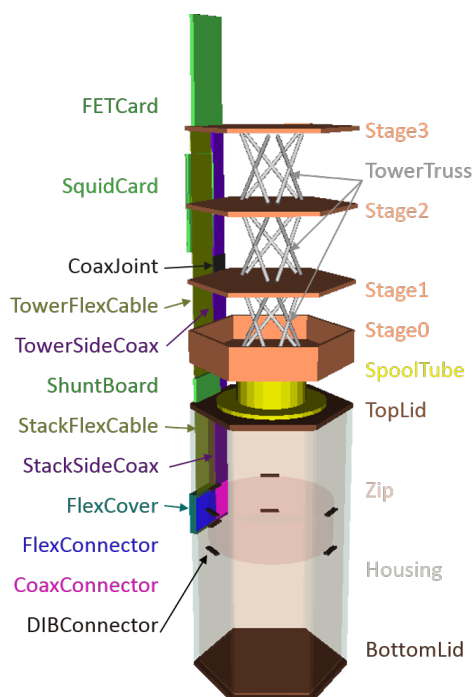


Figure 5-2. A SuperCDMS detector tower, a stack of six detectors with the thermal stages, wiring, and first-stage amplifier electronics that connect them to the cryostat. Adapted from [25].

background rate would swamp their slow response, and the materials they are built from activate under cosmic-ray exposure, which would compromise them for the eventual search. Operating underground in the same mine removes the cosmic-ray flux, and layered water, lead, and polyethylene shielding reduce the residual radiogenic background to a few events per kilogram per keV per day below about 100 keV.

Three of the facility's conditions bear directly on the measurements that follow. The dilution refrigerator reaches a base temperature near 12 mK, well below the roughly 30 mK at which the tungsten sensors operate, so the detectors can be biased and tuned in their transition. The large mass of the HV detectors makes them microphonic, and CUTE suppresses the dominant pulse-tube vibration through a gas-coupled thermal link and a mechanically isolated cryostat suspension, without which vibration would appear directly in the sensor noise. A magnetic shield reduces the ambient field at the payload, as the SQUID-based readout requires.

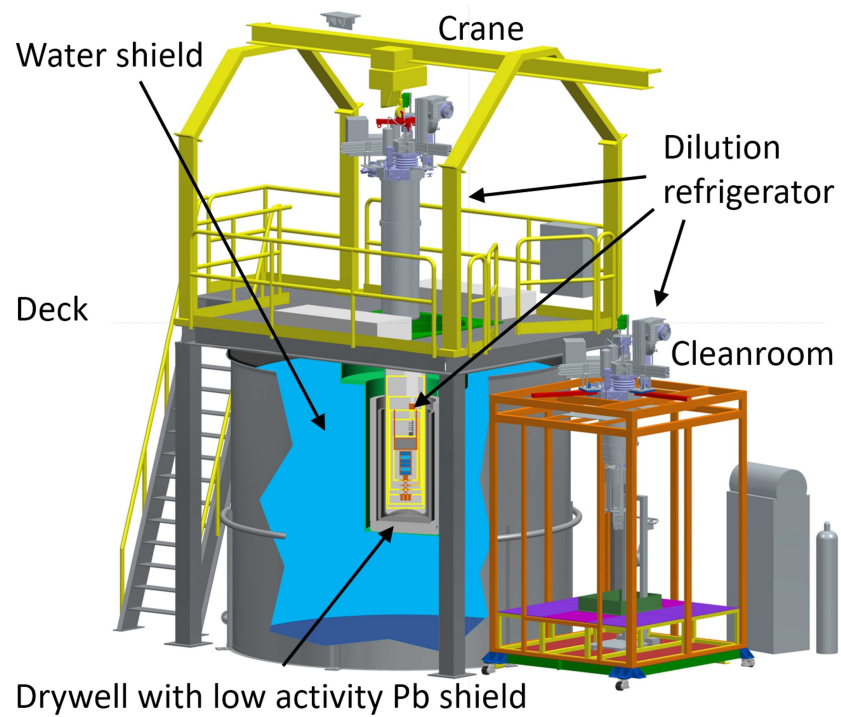


Figure 5-3. The CUTE facility, with the dilution refrigerator and its payload in the drywell at the centre of the water and lead shielding, and the cleanroom alongside. Adapted from [61].

Most of the studies in this chapter draw on Tower 3, a stack of six HV detectors, four germanium and two silicon, labelled T3P1 through T3P6, as the tower was operated and characterized at CUTE. The substrate measurement of the final section instead uses two detectors from Tower 2, T2P1 (Ge) and T2P5 (Si), chosen as a precaution because all of their channels were operational.

5.2 Noise in the Phonon Channels

The energy a phonon sensor can resolve is set by how much its output current fluctuates in the absence of a signal, and the spectrum of that current carries the signature of every process that drives the sensor. Most of those processes are well understood and reproduced by an established model. One feature of the measured spectrum is not, and tracking down its origin is the thread that runs through this section and the next.

5.2.1 A Bias-Dependent Feature in the Noise Spectrum

The noise of a channel is summarized by the power spectral density of its current. Figure 5-4 shows the measured spectrum of a single channel of Ge HV detector T3P1, tested at CUTE, at a sequence of bias points, given as the operating resistance in units of the normal resistance R_n , from $0.25 R_n$ deep in the transition up to $0.50 R_n$ near its midpoint. The broadband floor behaves as one would expect, but near 15 kHz there is a broad bump that grows steadily as the bias point is lowered. At $0.25 R_n$ it stands well clear of the floor; by $0.50 R_n$ it has nearly flattened into it.

That the bump moves with the bias point is itself informative. A fixed external interference would sit at the same frequency and amplitude regardless of where the sensor is biased, so a feature that strengthens as the operating point is lowered is not coming from outside the detector but from its own electrothermal state. The baseline resolution worsens in step with it, degrading toward $0.25 R_n$ where the bump is largest. To see whether the established noise sources can account for the feature, it has to be compared against a model of them.

5.2.2 The Noise Model

The current noise of a voltage-biased TES comes from three established sources, derived in full in [45, 35]. The passive resistances in the bias circuit, the shunt and any parasitic resistance,

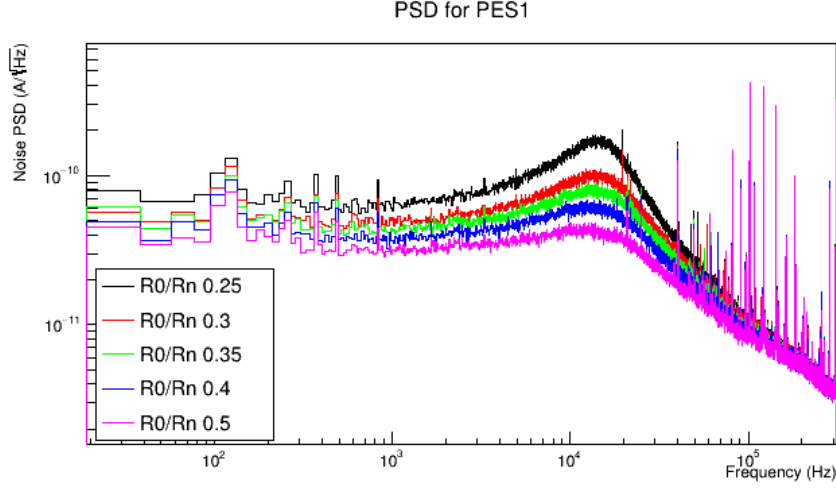


Figure 5-4. Measured current noise spectrum of a phonon channel at five bias points from 0.25 to $0.50 R_n$. A bump near 15 kHz grows as the bias point is lowered.

contribute Johnson noise. The TES itself contributes Johnson noise, shaped by the electrothermal feedback that opposes the temperature excursions a current fluctuation would cause. And the thermal link between the TES and its bath carries thermal fluctuation noise from the random exchange of heat across it. To these is added the noise of the SQUID amplifier that reads out the current. The four are uncorrelated, so they add in quadrature.

Carrying the small-signal response of the previous chapter through to the current, the passive Johnson, TES Johnson, and thermal fluctuation contributions are

$$I_{\text{PJN}}^2(\omega) = \frac{4k_B T^*}{R_0} \left| \frac{(1 - \mathcal{L} + i\omega\tau_{\text{eph}})}{\left(\frac{R_l}{R_0} + (1 + \beta) + i\omega\frac{L}{R_0}\right) (1 - \mathcal{L} + i\omega\tau_{\text{eph}}) + \mathcal{L}(2 + \beta)} \right|^2, \quad (5-1)$$

$$I_{\text{TJN}}^2(\omega) = \frac{4k_B T_0(1 + 2\beta)}{R_0} \left| \frac{(1 + i\omega\tau_{\text{eph}})}{\left(\frac{R_l}{R_0} + (1 + \beta) + i\omega\frac{L}{R_0}\right) (1 - \mathcal{L} + i\omega\tau_{\text{eph}}) + \mathcal{L}(2 + \beta)} \right|^2, \quad (5-2)$$

$$I_{\text{TFN}}^2(\omega) = \frac{10k_B T_0}{R_0} \left| \frac{\mathcal{L}}{\left(\frac{R_l}{R_0} + (1 + \beta) + i\omega\frac{L}{R_0}\right) (1 - \mathcal{L} + i\omega\tau_{\text{eph}}) + \mathcal{L}(2 + \beta)} \right|^2, \quad (5-3)$$

where $R_l = R_{\text{sh}} + R_p$ is the combined load resistance, τ_{eph} is the TES electron-phonon time constant, and $\mathcal{L}$ is the loop gain. The passive Johnson term is set by an effective temperature T^* , the resistance-weighted average of the temperatures of the passive components,

$$T^* = \frac{1}{R_0} \sum_i T_i R_i, \quad (5-4)$$

which collapses the shunt and parasitic resistors, each at its own temperature, into a single equivalent noise temperature referred to the operating resistance. The three terms share the same denominator, the characteristic polynomial of the small-signal system, and it is this shared denominator that will matter when the model is compared to the data.

The SQUID noise sits at the output of the chain and does not propagate through the TES response. It is a white floor with a rising low-frequency tail,

$$I_{\text{SQ}}^2(\omega) = \frac{4k_B T_{\text{SQ}}}{R_0} \left(1 + \left(\frac{\omega_{\text{sq}}}{\omega} \right)^{n_{\text{sq}}} \right)^2, \quad (5-5)$$

with a corner frequency $f_{\text{sq}} \approx 200$ Hz and a shallow low-frequency slope $n_{\text{sq}} \approx 0.45$ for these channels. The form is monotonic in frequency and carries no peak of its own. The total noise is the quadrature sum

$$I_n(\omega) = \sqrt{I_{\text{PJN}}^2 + I_{\text{TJN}}^2 + I_{\text{TFN}}^2 + I_{\text{SQ}}^2}, \quad (5-6)$$

plotted as an amplitude spectral density in the same units as the measured spectrum.

5.2.3 Comparison with the Measured Spectrum

Figure 5-5 overlays the model and its separate components on the measured spectrum at four bias points, using the channel parameters of Table 5-1. Across most of the band the model and data agree. The SQUID term sets the low-frequency rise and the high-frequency floor, the Johnson terms set the plateau in between, and the bias dependence of the floor follows from the loop gain, which rises as the operating point is lowered and lifts the thermal and Johnson contributions with it.

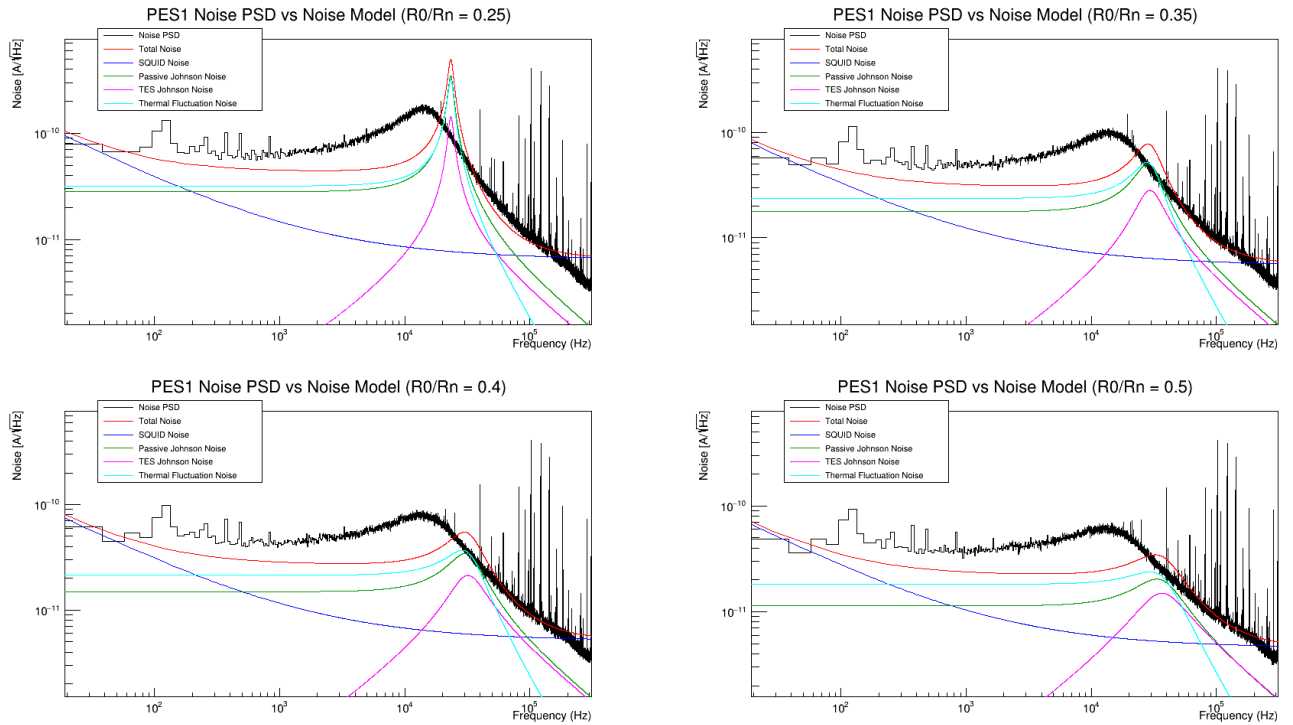


Figure 5-5. Measured current noise spectrum (black) and the noise model (red) with its components at four bias points. The resonance in the model is carried by the TES Johnson and thermal fluctuation terms; the SQUID term (blue) is monotonic. The model resonance shifts in frequency with bias point more strongly than the feature in the data does.

Table 5-1. Parameters of the PES1 channel used in the noise model and the electrothermal oscillation analysis, with measured and derived values listed separately.

Parameter	Value
Critical temperature, T_c	64.2 mK
Bath temperature, T_b	13 mK
Conduction power-law index, n	5
Electron-phonon coupling, Σ_W	$4.7 \times 10^8 \text{ W/K}^5/\text{m}^3$
Effective TES volume, $\mathcal{V}_W$	$6.9 \times 10^{-14} \text{ m}^3$
Tungsten heat-capacity coefficient, c_W	$108 \text{ J/K}^2/\text{m}^3$
Heat-capacity scaling, f_{sc}	1
Temperature sensitivity, α	100
Current sensitivity, β	0.3
Normal resistance, R_n	71.67 m Ω
Parasitic resistance, R_p	1.17 m Ω
Shunt resistance, R_{sh}	5 m Ω
Inductance, L	75 nH
Parasitic-resistor temperature, T_p	300 mK
Shunt-resistor temperature, T_{sh}	400 mK
DC SQUID current noise	5.5 pA/ $\sqrt{\text{Hz}}$
SQUID noise corner frequency, f_{sq}	200 Hz
SQUID noise low-frequency exponent, n_{sq}	0.45

The model also produces a resonance, and this is where it parts ways with the data. The peak comes from the TES Johnson and thermal fluctuation terms, which share the electrothermal denominator of Equations (5-1)–(5-3); when the system is underdamped that denominator has a resonance, and the two terms inherit it. The SQUID term, by contrast, is smooth and contributes no peak, so the bump in the model is unambiguously electrothermal in origin. So far this supports reading the observed feature as the same resonance. The difficulty is in how the resonance moves. In the model it is sharp and sits high in frequency deep in the transition, at $0.25 R_n$, and as the bias point is raised it broadens and slides down to higher frequency. The feature in the data does not march across the spectrum that way. It stays near 15 kHz at every bias point and changes mostly in height, not in frequency.

The model therefore confirms that a bias-dependent electrothermal resonance exists in the right part of the spectrum, but it does not reproduce the near-constant frequency of the observed feature, and on its own it cannot establish that the feature is electrothermal oscillation. The peak

in the noise spectrum is suggestive rather than conclusive. Settling the question calls for the dynamics directly, treating the electrothermal loop as a system that can oscillate and asking whether these detectors, at these bias points, fall in the regime where it does.

5.3 Electrothermal Oscillations

The residual feature in the noise spectrum strengthens as the bias point is lowered, the same direction in which the electrothermal feedback strengthens.

5.3.1 Standard Single Block Electrothermal Oscillations Solution

Restricting the electrothermal model of Chapter 3 to a single thermal body on a fixed bath, the small-signal response of the sensor temperature and current to a perturbation is governed by

$$\frac{d}{dt} \begin{pmatrix} \delta T \\ \delta I \end{pmatrix} = \begin{pmatrix} -\frac{1 - \mathcal{L}}{\tau_{\text{eph}}} & \frac{I_0 R_0 (2 + \beta)}{C} \\ -\frac{\mathcal{L} G_{\text{eph}}}{I_0 L} & -\frac{1}{\tau_{\text{el}}} \end{pmatrix} \begin{pmatrix} \delta T \\ \delta I \end{pmatrix}, \quad (5-7)$$

with the electrical and electron-phonon time constants

$$\tau_{\text{el}} = \frac{L}{R_l + R_0(1 + \beta)}, \quad \tau_{\text{eph}} = \frac{C}{G_{\text{eph}}}, \quad (5-8)$$

the loop gain $\mathcal{L} = P_J \alpha / (G_{\text{eph}} T_0)$, and $R_l = R_{\text{sh}} + R_p$. The sign of the lower-left entry is the heart of the feedback. A rise in temperature raises the resistance and, at fixed bias, lowers the current, so the electrical equation couples to temperature with a negative sign. It is this negative coupling that allows the loop to overshoot and oscillate rather than relax monotonically.

The free response is a sum of two exponential modes, $\delta T, \delta I \propto e^{-t/\tau_{\pm}}$, whose rates are the eigenvalues of the matrix. Writing the trace and determinant of (5-7) as

$$\frac{1}{\tau_+} + \frac{1}{\tau_-} = \frac{1}{\tau_{\text{el}}} + \frac{1 - \mathcal{L}}{\tau_{\text{eph}}}, \quad \frac{1}{\tau_+ \tau_-} = \frac{1}{\tau_{\text{el}}} \frac{1 - \mathcal{L}}{\tau_{\text{eph}}} + \frac{\mathcal{L}(2 + \beta)}{\tau_{\text{dyn}} \tau_{\text{eph}}}, \quad (5-9)$$

with the dynamical circuit time constant $\tau_{\text{dyn}} = L/[R_0(2 + \beta)]$, the two rates are

$$\frac{1}{\tau_{\pm}} = \frac{1}{2} \left(\frac{1}{\tau_{\text{el}}} + \frac{1 - \mathcal{L}}{\tau_{\text{eph}}} \right) \pm \frac{1}{2} \sqrt{\left(\frac{1}{\tau_{\text{el}}} - \frac{1 - \mathcal{L}}{\tau_{\text{eph}}} \right)^2 - \frac{4 \mathcal{L}(2 + \beta)}{\tau_{\text{dyn}} \tau_{\text{eph}}}}. \quad (5-10)$$

The feedback term enters the discriminant with a negative sign, through the negative off-diagonal product, and drives it toward negative values as the loop gain grows.

5.3.2 Oscillation Region and Frequency

The modes are real when the discriminant in (5-10) is positive and the perturbation relaxes without oscillating. When the discriminant turns negative the rates become complex conjugates and the response is an exponentially damped oscillation. The boundary, where the discriminant vanishes, defines the edges of the oscillating region. Solving for the loop gain at which it vanishes gives the two bounds

$$\mathcal{L}_{\pm} = 1 + \frac{\tau_{\text{eph}}}{\tau_{\text{dyn}}} \left[(3 + \beta) - \frac{R_l}{R_0} \pm 2 \sqrt{(2 + \beta) \left(1 + \frac{\tau_{\text{dyn}}}{\tau_{\text{eph}}} - \frac{R_l}{R_0} \cdot \frac{\tau_{\text{dyn}}}{\tau_{\text{eph}}} \right)} \right], \quad (5-11)$$

so that oscillation occurs for $\mathcal{L}_- < \mathcal{L} < \mathcal{L}_+$. The width of the region is governed by the ratio of the dynamical and electron-phonon time constants, which through τ_{dyn} and τ_{eph} depends on the inductance, the operating resistance, and the heat capacity and electron-phonon coupling of the film. Of the parameters that set these, the critical temperature is the one accessible at fabrication.

Within the oscillating region the complex rates are written $1/\tau_{\pm} = 1/\tau_0 \pm i \omega_{\text{eto}}$, with a common decay τ_0 and an oscillation frequency

$$\omega_{\text{eto}} = \left[\frac{\mathcal{L}(2 + \beta)}{\tau_{\text{dyn}} \tau_{\text{eph}}} - \frac{1}{4} \left(\frac{1}{\tau_{\text{el}}} - \frac{1 - \mathcal{L}}{\tau_{\text{eph}}} \right)^2 \right]^{1/2}. \quad (5-12)$$

The frequency is set by the same combination of loop gain, inductance, and time constants that bounds the region, and the oscillation is the time-domain counterpart of a peak in the noise spectrum at ω_{eto} .

5.3.3 Application to the Detector

Evaluating these expressions for the detector uses the same channel parameters as the noise model, listed in Table 5-1. The loop gain at the bias points of Figure 5-4 is of order $\mathcal{L} \approx 20$. Figure 5-6 shows the bounds $\mathcal{L}_{\pm}$ of (5-11) as the operating resistance, inductance, critical temperature, and parasitic resistance are varied, with the shaded bands marking the detector's range in each. Across these parameters the operating loop gain sits just above the lower bound $\mathcal{L}_{-}$, placing the detector inside the oscillating region but near its lower edge.

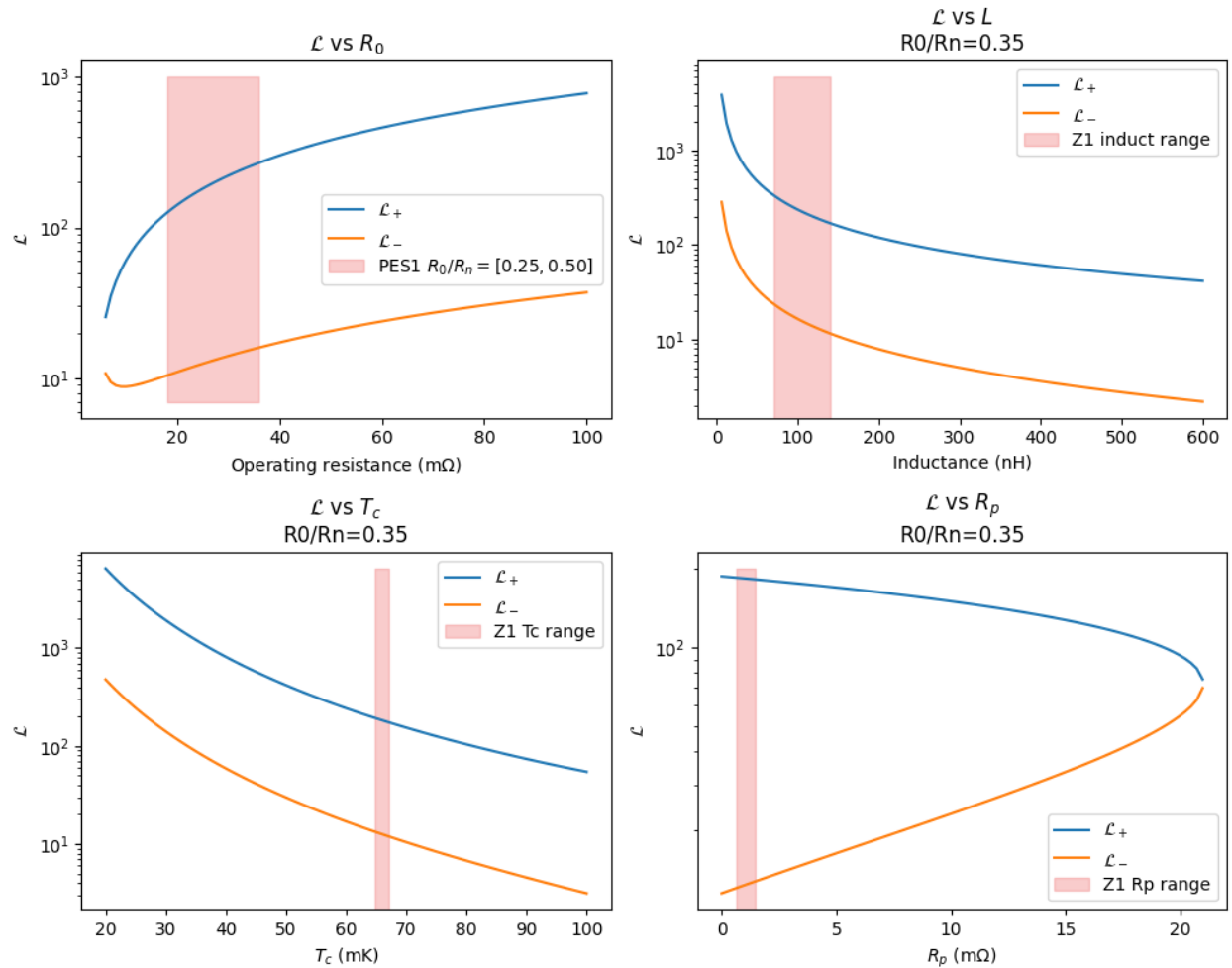


Figure 5-6. Loop-gain bounds $\mathcal{L}_{\pm}$ of the oscillating region against operating resistance, inductance, critical temperature, and parasitic resistance. Shaded bands mark the detector's range; the operating loop gain of order 20 lies just inside the lower edge.

With the detector inside the region, the oscillation frequency of (5-12) follows from the same parameters. Figure 5-7 shows how ω_{eto} varies with operating resistance, inductance, bath temperature, and load resistance over the detector's range, and across that range it falls close to 15 kHz, matching the frequency of the feature in the measured spectrum.

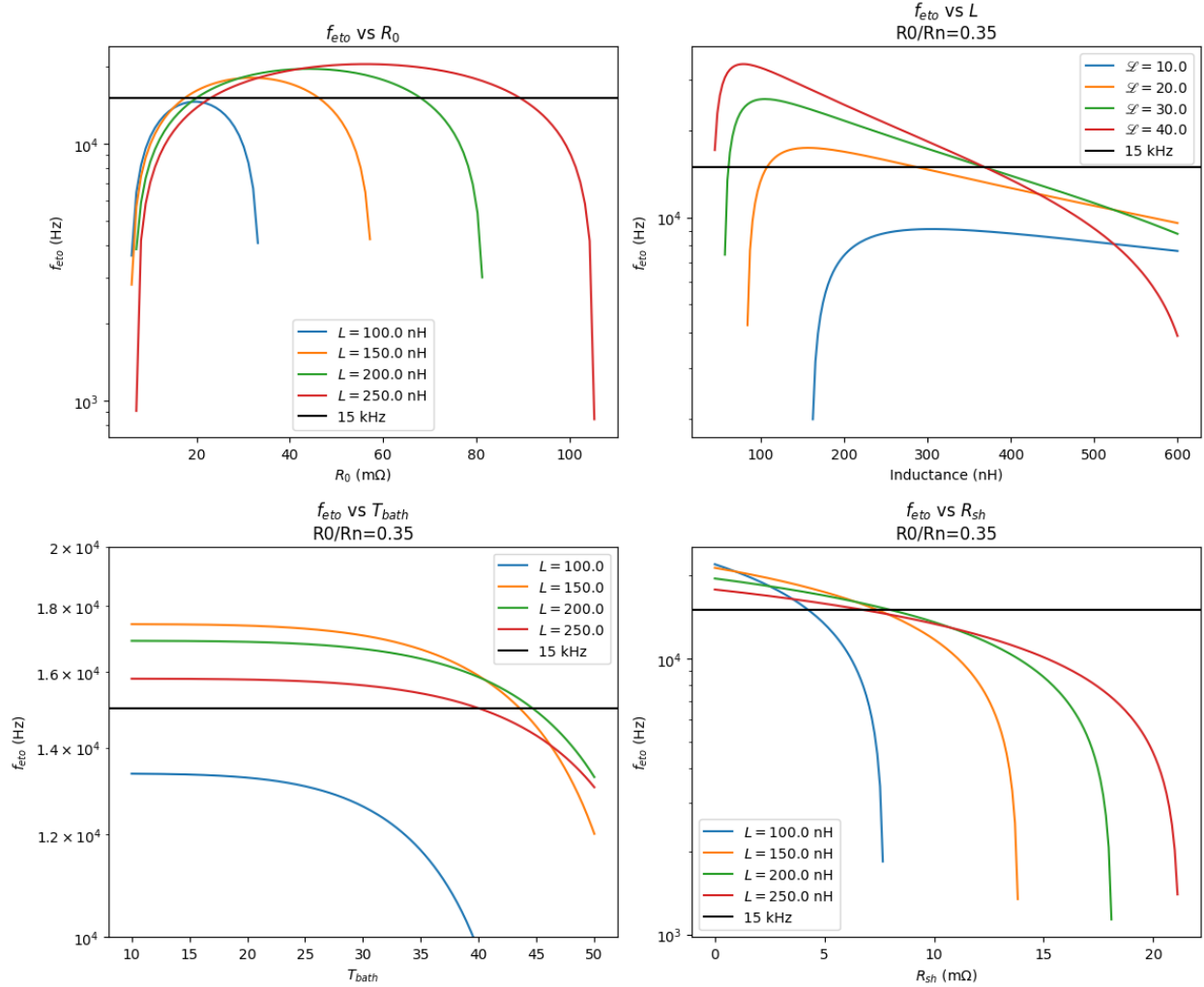


Figure 5-7. Electrothermal oscillation frequency f_{eto} against operating resistance, inductance, bath temperature, and load resistance. Over the detector's expected parameter range the predicted frequency settles near 15 kHz.

A direct check is to simulate the sensor response. The detectors were operated at $0.50 R_n$ during the runs, the bias point at which the feature is weakest, so an oscillating pulse was not recorded directly. Simulating the sensor with its measured parameters at lower bias points instead, Figure 5-8 shows the resulting current pulse, which develops the undershoot and ringing

characteristic of the damped oscillation as the bias point is lowered, growing more pronounced deeper in the transition exactly where the spectral feature is strongest.

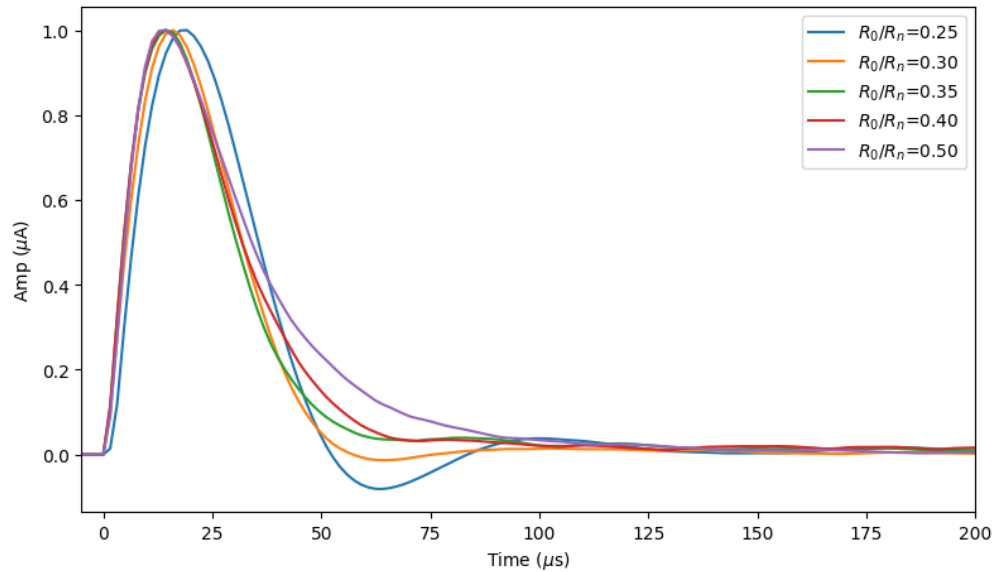


Figure 5-8. Simulated current pulse at a sequence of bias points. The undershoot and ringing of the electrothermal oscillation deepen as the bias point is lowered.

The agreement is in the location of the feature and its trend with bias, not in a fitted amplitude, so the evidence is consistent with an electrothermal oscillation rather than a proof of one. The detector sits inside the oscillating region by the small-signal analysis, the predicted frequency matches the observed feature, and a simulated pulse oscillates at the corresponding bias points. Together these establish the electrothermal oscillation as the origin of the bias-dependent feature in the noise spectrum and of the resolution degradation that accompanies it.

5.4 Performance of the SuperCDMS SNOLAB Detectors

The studies of the preceding sections were carried out at CUTE, where the tower was characterized before installation. Once the detectors are operating in the SuperCDMS SNOLAB experiment, the questions become operational, whether a stream of randomly triggered data can be cleaned of pulses to leave the noise, and what baseline resolution each channel reaches in the configuration it will run in. Both are taken up for the six detectors of Tower 3, using data from the experiment's second run.

5.4.1 Event Selection and the Noise Sample

Measuring the noise of a channel means measuring its current in the absence of a signal, but randomly triggered data contain a mixture of genuine phonon pulses and traces that are pure noise. Isolating the noise therefore begins by removing the pulses. A sequence of selection cuts does this, and the traces that survive form the clean noise sample. The cuts fall into two groups. The template-independent cuts act on simple properties of each trace, its minimum-to-maximum amplitude, its baseline level, and its slope, and require no knowledge of the pulse shape. The template-dependent cuts use the optimal filter of Section 5.4.2, rejecting traces on their fitted amplitude and on the goodness of fit χ^2 to the pulse template. Each cut is set at a threshold of four standard deviations from the noise distribution, and the template-dependent cuts are applied iteratively as the template is refined.

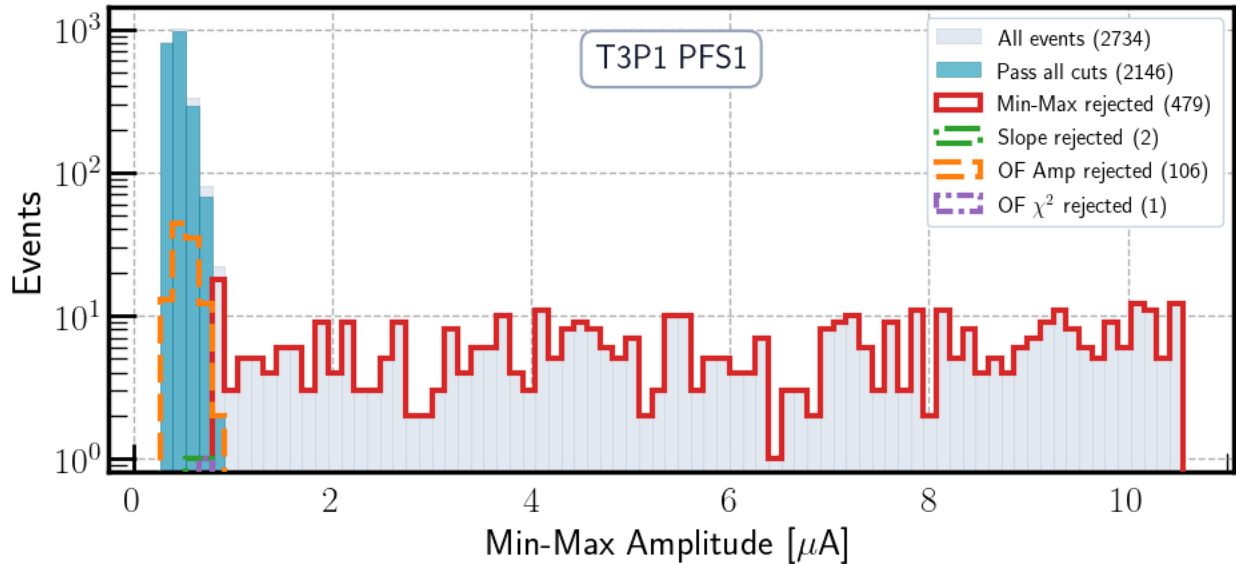


Figure 5-9. Distribution of events in minimum-to-maximum amplitude for one channel, showing the populations removed by each cut and the events surviving all cuts. Genuine pulses extend to large amplitude while noise concentrates at small amplitude.

Figure 5-9 shows the distribution of events in the minimum-to-maximum amplitude, the simplest discriminant. The noise events cluster at small amplitude, while genuine pulses spread out to much larger values, and the cuts separate the two populations. The shapes of the rejected

pulses are collected in Figure 5-10, which spans the range of pulse morphologies the selection removes, from clean large pulses to distorted and pile-up traces that depart from the template.

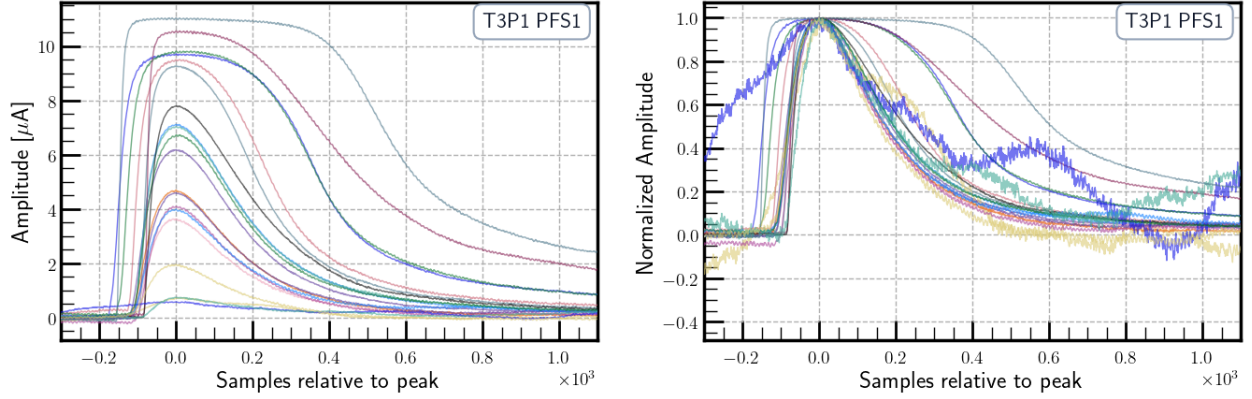


Figure 5-10. Pulse shapes rejected by the event selection for one channel, spanning clean pulses and traces whose shape departs from the template.

The same events seen across the tower are shown in Figure 5-11, which displays the raw traces of several detectors together for a set of randomly triggered events. The display gives a direct view of what the selection acts on, the coincident activity across channels and the distinction between traces that survive and those that are removed.

With the pulses removed, the surviving traces give the noise of each channel. Figure 5-12 shows the resulting current noise spectra across the channels of the tower. These spectra are the input to the resolution estimate that follows, and characterize the noise of the detectors as they run in the experiment.

5.4.2 Optimal Filtering and Baseline Resolution

The estimator that both drives the template-dependent cuts and sets the baseline resolution is the optimal filter [35]. A measured current trace $d(t)$ is a pulse of unknown amplitude a arriving at unknown time t_0 on top of noise $n(t)$,

$$d(t) = a s(t - t_0) + n(t), \quad (5-13)$$

where $s(t)$ is a normalized pulse template, in this work the average pulse template built from many pulses. In the frequency domain $\tilde{d}(\omega) = a e^{-i\omega t_0} \tilde{s}(\omega) + \tilde{n}(\omega)$. The amplitude and arrival time are

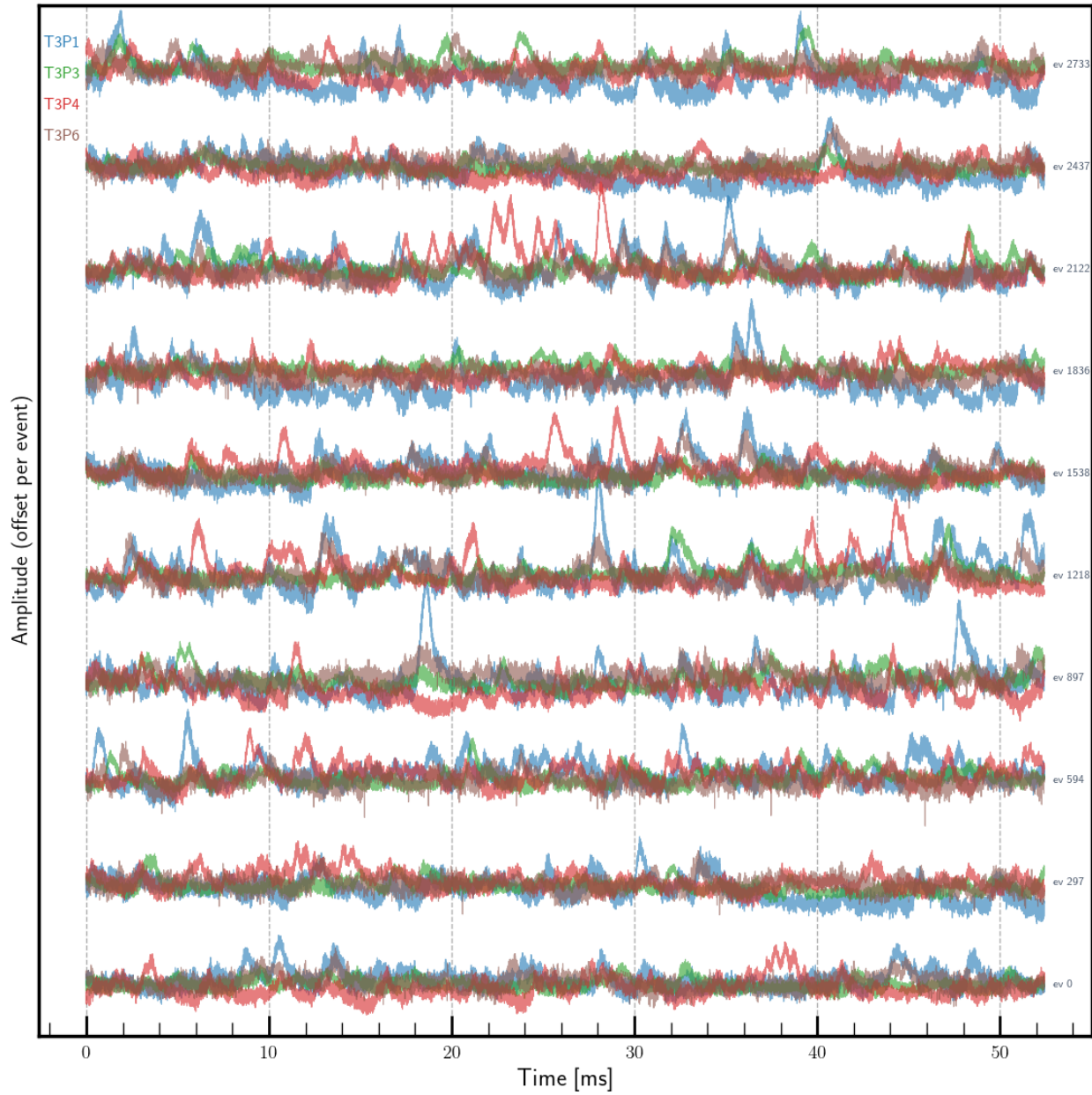


Figure 5-11. Raw traces of several Tower 3 detectors for a set of randomly triggered events, offset vertically per event.

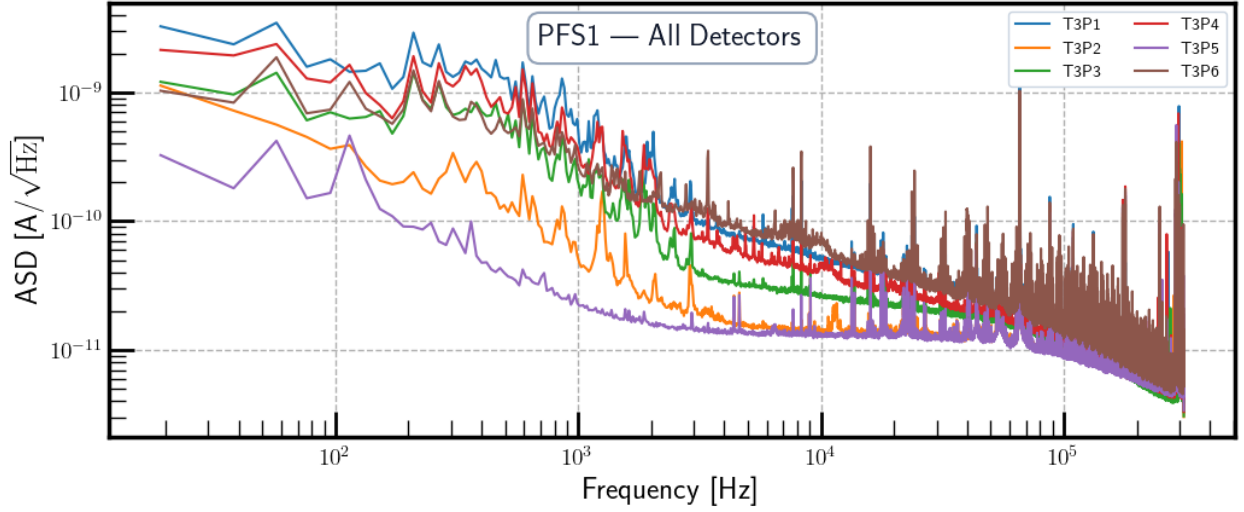


Figure 5-12. Current noise spectra of the Tower 3 channels, obtained from the randomly triggered traces surviving the event selection.

estimated by minimizing the noise-weighted mismatch between the trace and the template,

$$\chi^2(a, t_0) = 4 \int_0^\infty \frac{|\tilde{d}(\omega) - a e^{-i\omega t_0} \tilde{s}(\omega)|^2}{I_n^2(\omega)} \frac{d\omega}{2\pi}, \quad (5-14)$$

with $I_n(\omega)$ the one-sided noise amplitude spectral density measured above. Each frequency is weighted by the inverse of the noise power, so bins with a high signal-to-noise ratio dominate the estimate and noisy bins are suppressed. Minimizing with respect to a at fixed t_0 gives the amplitude estimator

$$\hat{a}(t_0) = \frac{4 \int_0^\infty \frac{\tilde{s}^*(\omega) \tilde{d}(\omega) e^{i\omega t_0}}{I_n^2(\omega)} \frac{d\omega}{2\pi}}{4 \int_0^\infty \frac{|\tilde{s}(\omega)|^2}{I_n^2(\omega)} \frac{d\omega}{2\pi}}, \quad (5-15)$$

a noise-weighted cross-correlation of the trace with the time-shifted template, normalized so that $\langle \hat{a} \rangle = a$ for a noiseless pulse. The arrival time follows from maximizing $|\hat{a}(t_0)|$, and the χ^2 at the best fit provides the per-pulse goodness of fit used in the event selection to reject pile-up and shape-distorted traces.

The variance of the amplitude estimator over noise realizations is

$$\sigma_a^2 = \left[4 \int_0^\infty \frac{|\tilde{s}(\omega)|^2}{I_n^2(\omega)} \frac{d\omega}{2\pi} \right]^{-1}, \quad (5-16)$$

the baseline resolution, the smallest amplitude fluctuation the filter can distinguish for the given template and noise spectrum, independent of the signal size. No linear filter does better for stationary Gaussian noise. The integrand $|\tilde{s}|^2/I_n^2$ is the contribution of each frequency to the measurement, so the resolution is governed by the band where the pulse has the most power relative to the noise.

Figure 5-13 shows the baseline resolution of the individual channels of the tower, evaluated from (5-16) with the measured noise spectrum and average pulse template of each channel. The resolutions vary from channel to channel with the noise and the pulse shape, and the combined resolution of all channels on each detector, shown in the adjacent panels, characterizes the detector as it runs in the experiment. These calculations used a premade template, not necessarily the true per-channel template, to give a rough estimate of the baseline resolution.

Figure 5-13 shows evidence of correlated noise between the phonon channels. The two bottom panels both give the total baseline resolution, from noise estimated two ways. In the bottom left, the individual channel random traces, after event rejection, are summed in the time domain and the total PSD is taken from that summed trace, so it retains the correlations between channels. In the bottom right, the total PSD is instead the sum of the individual channel PSDs added in quadrature, which drops the cross terms and treats the channels as independent. Both panels are on the same arbitrary scale, the difference between the two panels is the cross-spectral contribution itself, and its size measures the correlated noise across the tower at that stage of the experiment.

5.5 Thermal Parameter Extraction

The long tail in the HVeV pulses of Chapter 4 is the substrate cooling back to the bath after an event. How long that takes is the substrate relaxation time, and it depends on the substrate-bath Kapitza conductance K_{s-b} , one of the two thermal parameters the two-block model of Chapter 3

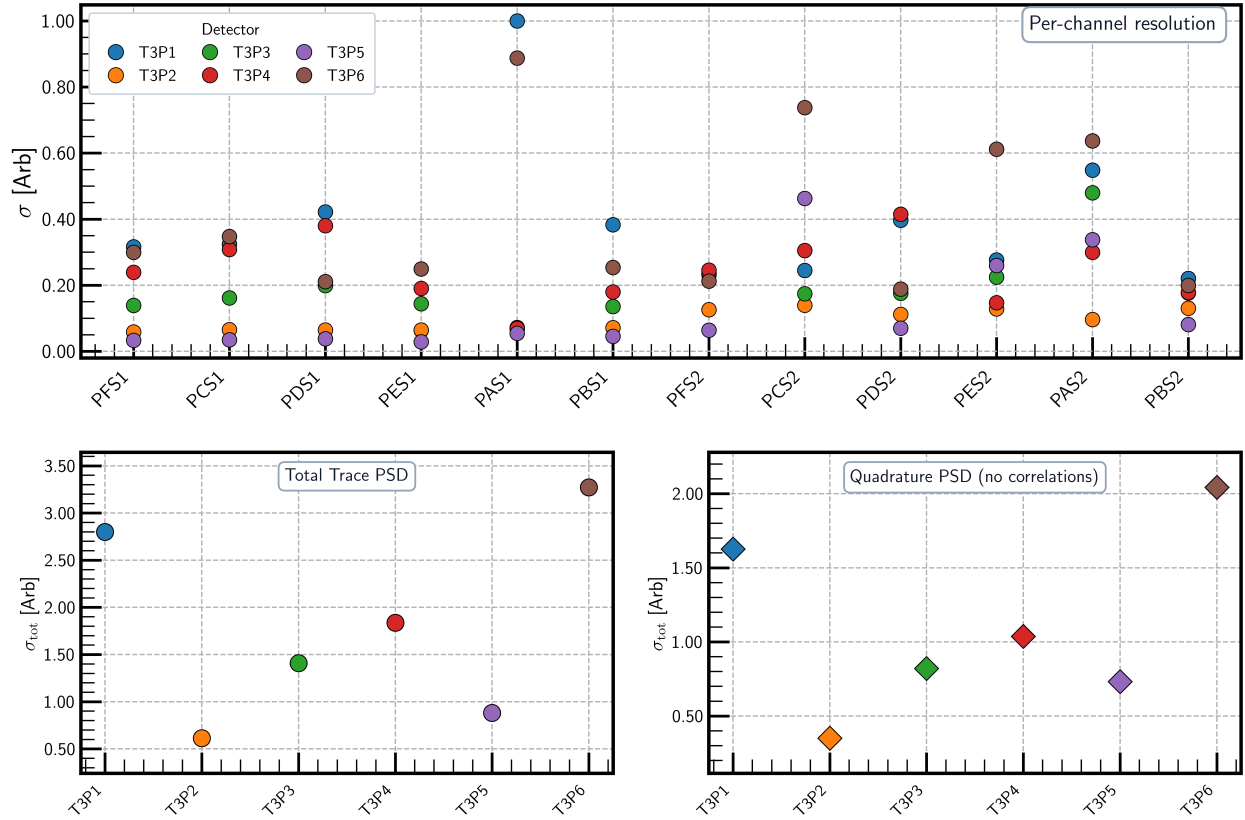


Figure 5-13. Arbitrarily scaled baseline resolution of the individual channels of the tower (top), with the per-detector resolution in the bottom row from the total-trace PSD (left) and the quadrature PSD (right). The per-channel resolution follows from the optimal filter applied to the measured noise and average pulse template of each channel.

needs but cannot supply on its own. The relaxation time also governs how the substrate responds to a sustained event rate. It sets how quickly the substrate sheds the heat each event deposits, and so bounds the rate, from a calibration source or from background, that the crystal can absorb before its steady-state temperature begins to rise; read the other way, an elevated substrate temperature is a measure of the integrated rate the detector is seeing. The tungsten electron-phonon coefficient $\Sigma_W \mathcal{V}_W$ and K_{s-b} are degenerate in the pulse, so a fit to the pulse cannot constrain both at once. Separating them takes a measurement that loads the substrate by a known amount and reads its temperature back without going through the pulse, here with no background loading, the substrate sitting at the bath with the heaters off.

5.5.1 Method

Each detector carries phonon channels on both faces, and the channels on one face can be driven fully normal to act as resistive heaters while a single channel on the opposite face stays in transition and reports the substrate temperature through its operating point. A measured bias current on the heaters sets a known Joule load, and the mixing chamber stays cold because the heat enters the crystal rather than the stage. The measurement ran on T2P1 (Ge) and T2P5 (Si) during a SNOLAB run. On T2P1 the central channel is PFS1 and the heaters are the channels on side 2; on T2P5 the central channel is PFS2 and the heaters are on side 1.

The central channel sits at 50% of its normal resistance, which pins its temperature at the effective critical temperature T_c . Its Joule power leaves the tungsten through electron-phonon coupling into the substrate,

$$P_J = \Sigma_W \mathcal{V}_W (T_c^n - T_{\text{subst}}^n), \quad n = 5, \quad (5-17)$$

and the substrate sheds its heat to the bath across the Kapitza link,

$$\sum_{i=1}^N P_J^{(i)} = K_{s-b} (T_{\text{subst}}^m - T_b^m), \quad m = 4, \quad (5-18)$$

where the sum runs over the N channels that are on, the six normal heaters and the central channel. The exponents differ because the two links are different kinds of coupling. The TES cools to the substrate through electron-phonon coupling, which sets the fifth power, while the substrate cools to the bath through phonon-phonon transmission across the Kapitza link, which sets the fourth. Only the central channel enters Equation (5-17), while the shared crystal enters Equation (5-18).

With the heaters off the substrate equilibrates to the bath, $T_{\text{subst}} = T_b$, and Equation (5-17) returns $\Sigma_W \mathcal{V}_W$ from the central-channel Joule power alone, given an independently measured T_c . Turning the heaters on raises the substrate above the bath, and inverting Equation (5-17) for the central channel at its shifted operating point recovers the substrate temperature,

$$T_{\text{subst}} = \left(T_c^n - \frac{P_J}{\Sigma_W \mathcal{V}_W} \right)^{1/n}. \quad (5-19)$$

The total power crossing the substrate-bath link is the sum over the six normal heater channels and the central channel, and Equation (5-18) then fixes K_{s-b} . Each heater carries a current set by the bias and its own normal and parasitic resistance,

$$I_s^{(i)} = I_b \frac{R_{\text{sh}}}{R_{\text{sh}} + R_p^{(i)} + R_n^{(i)}}, \quad P_J^{(i)} = \left(I_s^{(i)} \right)^2 R_n^{(i)}, \quad (5-20)$$

so the loaded power follows from the measured resistances rather than from any assumption about the heater operating point. The bias current is $I_b = 500 \mu\text{A}$ and the shunt resistance $R_{\text{sh}} = 5 \text{ m}\Omega$.

5.5.2 Extracted Coefficients

With the heaters off the central channel dissipates 15.47 pW on the germanium device and 1.56 pW on the silicon device at a bath temperature of 18.97 mK, fixing the electron-phonon coefficients. Turning the heaters on loads the germanium crystal with 420 pW across its six channels and the silicon crystal with 360 pW, placing the substrate at 29.9 mK and 37.7 mK respectively, from which the total link power fixes the Kapitza coefficients. The error bars are statistical, from the standard deviation of the central-channel Joule power across the 50 to 90% window propagated through the extraction, with T_c , T_b , and the heater power held fixed.

Table 5-2. Extracted thermal parameters for the two detectors, central channel (PFS1/T2P1, PFS2/T2P5). Heater-run values for $\Sigma_W \mathcal{V}_W$, K_{s-b} , and T_{subst} ; the rest at $T_{\text{subst}}^{\text{op}}$. Uncertainties are statistical.

Quantity	T2P1 (Ge)	T2P5 (Si)
P_{central} (heaters off) [pW]	15.47 ± 0.19	1.56 ± 0.04
P_{total} (6 heaters + central) [pW]	~ 434.76	~ 360.58
$\Sigma_W \mathcal{V}_W$ [W/K ⁵]	$(3.72 \pm 0.05) \times 10^{-5}$	$(1.66 \pm 0.04) \times 10^{-5}$
K_{s-b} [W/K ⁴]	$(6.51 \pm 1.86) \times 10^{-4}$	$(1.90 \pm 0.06) \times 10^{-4}$
T_c [mK]	53.0	39.5
T_{subst} (heater run) [mK]	29.9 ± 1.8	37.7 ± 0.3
$T_{\text{subst}}^{\text{op}}$ (12 ch at 50%) [mK]	~ 25.4	~ 21.9
$T_{\text{subst}}^{\text{op}}/T_c$	~ 0.48	~ 0.55
<i>At the operating temperature $T_{\text{subst}}^{\text{op}}$:</i>		
G_K [nW/K]	42.6 ± 12.0	7.9 ± 0.2
G_{eph} per channel [pW/K]	77 ± 1	19.1 ± 0.5
C_{subst} [nJ/K]	~ 12.0	~ 1.5
τ_{subst} [ms]	281 ± 80	191 ± 6

The two devices land in different regimes. By Equation (5-19), how far the substrate sits below T_c depends on the central channel's heaters-on Joule power. On germanium that power is 14.67 pW and the substrate reads at little more than half T_c ; on silicon it is only 0.33 pW and the substrate reads just 1.8 mK below T_c , at 96% of it. The silicon extraction therefore rests on a much smaller temperature lever, and its tighter error bars reflect statistical scatter alone, not the dominant systematic. That systematic is the critical temperature, read from the transition midpoint and entering Equation (5-17) as a fifth power, so an imprecise T_c moves $\Sigma_W \mathcal{V}_W$ and, through the inversion, K_{s-b} by more than the statistical bars suggest. Pinning T_c from a full transition fit would tighten the result, and the bounded ratio T_{subst}/T_c cancels much of this sensitivity.

5.5.3 Substrate Temperature and Relaxation Time at the Operating Point

The detector runs with all twelve channels biased at 50%, not the single channel of the heater run, so the relaxation time is evaluated at that operating condition. Taking the twelve channels to be identical to the measured central one, each dissipating its heaters-off Joule power P_0 , the load on the substrate is $12P_0$, and balancing it against the Kapitza outflow gives the

operating substrate temperature,

$$12 P_0 = K_{s-b} \left(\left(T_{\text{subst}}^{\text{op}} \right)^4 - T_b^4 \right), \quad T_{\text{subst}}^{\text{op}} = \left(\frac{12 P_0}{K_{s-b}} + T_b^4 \right)^{1/4}. \quad (5-21)$$

This evaluates to 25.4 mK on the germanium device and 21.9 mK on the silicon device, both well below T_c , so the full tower does not warm the substrate enough to pull the sensors out of transition.

The relaxation time is the substrate heat capacity over the conductance draining it. Two paths carry heat off the substrate, the Kapitza link to the bath and the electron-phonon return into the channels, with differential conductances

$$G_K = 4 K_{s-b} \left(T_{\text{subst}}^{\text{op}} \right)^3, \quad G_{\text{eph}} = 5 \Sigma_W \mathcal{V}_W \left(T_{\text{subst}}^{\text{op}} \right)^4, \quad (5-22)$$

of 42.6 nW/K against 77 pW/K per channel on germanium and 7.9 nW/K against 19.1 pW/K per channel on silicon. The electron-phonon return, summed over the twelve channels and weighted by $\eta = (T_{\text{subst}}^{\text{op}}/T_c)^{n-1}$, stays well under a percent of the Kapitza conductance, so the substrate drains almost entirely through the Kapitza link, but both enter the relaxation time,

$$\tau_{\text{subst}} = \frac{C_{\text{subst}}}{G_K + N\eta G_{\text{eph}}}, \quad N = 12. \quad (5-23)$$

The heat capacity completes the estimate. The silicon substrate heat capacity scales from the HVeV value by the crystal volume ratio,

$$C_{\text{subst}}^{\text{Si}} = C_{\text{subst}}^{\text{HVeV}} \frac{V_{\text{HV}}}{V_{\text{HVeV}}}, \quad \frac{V_{\text{HV}}}{V_{\text{HVeV}}} = \frac{\pi(50 \text{ mm})^2(33.3 \text{ mm})}{(4 \text{ mm})^2(10 \text{ mm})} \approx 1600, \quad (5-24)$$

the HV detector being a cylinder against the rectangular HVeV substrate. Germanium does not scale from a silicon reference, having the larger Debye heat capacity, and comes from the

germanium Debye coefficient at the detector volume,

$$C_{\text{subst}}^{\text{Ge}} = \beta_{\text{Ge}} V_{\text{HV}} \left(T_{\text{subst}}^{\text{op}} \right)^3, \quad \beta_{\text{Ge}} = 2.8 \times 10^{-6} \text{ J cm}^{-3} \text{ K}^{-4}, \quad (5-25)$$

the coefficient measured below 4 K in [67]. The resulting relaxation times are 281 ms on the germanium device and 191 ms on the silicon device. Their uncertainty propagates from the uncertainty on the central-channel Joule power through K_{s-b} , the heat capacity entering only through its fixed coefficient.

These coefficients fix Σ_{W} $\mathcal{V}_{\text{W}}$ and K_{s-b} separately, from steady-state measurements alone. The line of work began with the HVeV long tail of Chapter 4, which exposed the substrate as a thermal stage the model had to account for and which left the two parameters degenerate. The heater method answers that opening question directly, supplying both for the HV detectors from their steady-state operating points, to inform the model.

CHAPTER 6

SUMMARY AND OUTLOOK

What a dark matter experiment can exclude, and what it could one day claim, rests on how faithfully its detectors convert a recoil into a signal. In a high-voltage detector there is no charge readout to fall back on. The phonon pulse carries everything the experiment will ever know about an event, and every step between the recoil and the sensor is therefore part of the measurement. The work in this dissertation was an attempt to model that chain end to end and to confront the model with data at each stage.

6.1 Summary

The response of the crystal was built from the band structure outward. A recoil divides its energy between charge carriers and lattice vibrations in a proportion set by the ionization yield, and the surviving carriers are driven across the crystal by the bias, radiating the work the field performs on them back into the lattice as Neganov-Trofimov-Luke phonons. That emission is a Cherenkov-like process with a hard threshold at the longitudinal sound speed, and its scattering rate carries a quadratic dependence on the carrier wavevector.

Electrons occupy anisotropic conduction valleys and holes do not, so the two carriers drift differently under the same field. An electron follows its valley axis rather than the field and lays its NTL emission along an oblique path, while a hole drifts compactly along the field. The consequence is an asymmetry in the phonon energy arriving at the two faces, and its dependence on the event depth differs between silicon and germanium because their valleys sit differently with respect to the bias. The six $\langle 100 \rangle$ valleys of silicon split into two along the field and four across it, while the four $\langle 111 \rangle$ valleys of germanium are all oblique at the same angle, and every electron drifts identically. The asymmetry is at most a few percent of the per-face energy, but it follows from the valley structure alone and it separates the two targets even at identical crystal orientation.

The sensor was treated as an electrothermal system rather than as a thermometer. The resistance surface $R(T, I)$ couples the bias circuit to the thermal circuit, and the linearized dynamics of the coupled system fix the pulse shape, the complex impedance, and the propagation of noise. Three results follow from that treatment which a conventional single-block model does

not give. Several channels on one crystal are coupled to one another through that shared substrate, through a coupling frequency Ω_{tsc} that enters the characteristic polynomial directly. And the stability boundary is a joint condition on all three subsystems, which fixes a critical inductance above which the electrothermal loop oscillates rather than damps.

The model met data on the HVeV detector, where an LED calibration supplies two observables that constrain different things. The discrete quanta peaks fix the energy delivered by the charge transport and say almost nothing about how that energy arrives in time. The average pulse constrains both the temporal shape of the phonon power and the electrothermal parameters that respond to it, and those two sets of parameters are degenerate in the pulse. Fixing everything that could be measured on its own, the critical temperature from $R(T)$, the normal and parasitic resistances from the I_b-I_s characteristic, and the inductance from the readout, left the transition width and the bias point to the fit, which returned a transition width near 1.25 mK and bias fractions of 0.30 and 0.26 for the two calibration runs.

The tail of the pulse is where the model first failed. A two-pole TES cannot produce it. A second thermal block (third pole) can, but only if the substrate is allowed to sit well above the bath, and this is the point at which the substrate stopped being a boundary condition and became a part of the detector.

Detectors from the SuperCDMS tower 3 were characterized at CUTE. A bias-dependent feature near 15 kHz appears in the noise of several channels, and a four-source noise model reproduces a resonance in the right part of the spectrum without reproducing the near-constant frequency of the observed feature, so the noise spectrum alone leaves the question open. Treating the electrothermal loop as a dynamical system that can oscillate settles it. The detectors sit inside the oscillating region by the small-signal analysis, the predicted oscillation frequency matches the observed feature and follows its trend with bias, and a simulated pulse oscillates at the corresponding bias points. The evidence is consistent with an electrothermal oscillation and with the resolution degradation that accompanies it, though the agreement is in the features of the bump and its trend rather than numerical matching.

A heater-based measurement extracted the thermal parameters of a germanium and a silicon tower detector directly, using the central channel as a thermometer while the surrounding channels were driven. On germanium the substrate rose to 29.9 mK and on silicon to 37.7 mK against a bath at 18.97 mK, with electron-phonon conductances of 77 and 19.1 pW/K per channel and Kapitza conductances of 42.6 and 7.9 nW/K. These are the quantities the pulse fit could not separate.

6.2 The Substrate as a Dynamical Element

The substrate appears three times in this work, in three settings that had no reason to agree with one another, and each time it refuses to behave as a bath.

The linearized dynamics predicted it. Promoting T_{subst} to a state variable produces a slow pole with no counterpart in the single-block model, gives the substrate its own responsivity, and couples channels that share a crystal. The HVeV pulse then required it. The observed tail relaxes on a few milliseconds, and a substrate held at the bath would relax in roughly 65 ms, far too slowly. Solving the steady state with the biased sensor near T_c places the substrate near 42 mK against a 15 mK bath, where the conductances have climbed steeply enough to bring the relaxation time to about 2 ms. The detector runs with a warm substrate under it, and the tail is what that warmth looks like in the time domain. The same slow channel accounts for the shift in operating point between two calibration runs taken back to back at the same bath temperature. And the tower measurement confirmed it, placing the substrate of a biased germanium detector at 29.9 mK against a 18.97 mK bath, well above the bath and in the same direction the pulse demanded.

The pulse fit could not go further on its own. The heat capacity of the substrate, the electron-phonon coupling of the tungsten, and the substrate-bath Kapitza conductance enter the pulse shape only through the combination

$$\tau_{\text{subst}} = \frac{C_{\text{subst}}}{G_K + \eta G_{\text{eph}}}, \quad (6-1)$$

so the fit constrains their ratio and leaves the individual quantities floating. The consequence is visible in the numbers. The fitted substrate heat capacity came out more than an order of magnitude below the Debye value at the operating temperature, and the electron-phonon coupling

moved by nearly a factor of two between the one-block and two-block fits. A fit alone cannot say whether this means the model is wrong or the parameters are simply unidentifiable from a pulse.

Breaking that degeneracy needs a measurement that touches the substrate without going through the pulse, which is what the heater measurement provides. The two halves of this dissertation therefore meet at exactly this point. Chapter 4 identifies a degeneracy that a pulse cannot resolve, and Chapter 5 supplies the independent constraint that resolves it. Closing that loop, by carrying the measured conductances back into the pulse fit as priors and re-extracting the heat capacity on its own, is the most immediate piece of work this thesis leaves undone. If the substrate heat capacity remains far below the Debye value once the conductances are pinned, the discrepancy is physical rather than a fitting artifact.

There is a second prediction of the same origin that this thesis states but does not test. If several channels on one crystal are coupled through a shared substrate, their noise should be correlated, and the size of the correlation is fixed by the coupling frequency Ω_{tsc} that the linearized model already supplies. Correlated noise is observed across the tower channels. Whether the observed cross-spectrum is the one the substrate coupling predicts is a question that can be asked of data already in hand.

6.3 Relevance to the Dark Matter Search

A dark matter limit is set by two numbers, how low in energy a detector can see and how many background events survive the analysis. A high-voltage detector buys the first by giving up the tool that ordinarily supplies the second. There is no charge channel, so there is no ionization yield to separate a nuclear recoil from an electron recoil and no charge asymmetry to mark an event as having occurred at a surface. Everything the experiment will ever know about an event is carried by the phonon pulses its channels record. Modeling those pulses is not preparatory work for the search. It is the only place where the two levers of the search can be reached at all.

The most direct consequence is fiducialization. The phonon energy arriving at the two faces is asymmetric, and the asymmetry depends on where in the crystal the event occurred, while the partition of energy among the channels on a face locates it laterally. Together these give a map

from event position to observables, computed rather than measured, and a map of that kind is what a fiducial cut is built from. Surface events are where the radiogenic background concentrates and where the field and the charge collection are least well behaved, so removing them removes background and cleans the energy scale at once. A cut derived from simulation carries a second advantage that a cut derived from calibration data does not, since its signal acceptance can be computed across the whole crystal rather than inferred from the few positions a collimated source can illuminate, and that acceptance is a systematic in every exclusion limit ever published. None of it is available unless the simulation is known to reproduce the pulses real events produce, which is what the tuning in Chapter 4 was for.

The second consequence is operational. A detector that oscillates has degraded baseline resolution, degraded resolution raises the threshold, and the threshold is what terminates an exclusion curve at low mass. The electrothermal oscillation identified here is not a nuisance to be discovered after the fact but a region of the operating space whose boundaries follow from parameters that can be measured beforehand. The bias, the bath temperature, and the readout inductance together decide whether a channel sits inside that region. An operating point is ordinarily chosen by scanning the bias and keeping whichever point gives the best resolution, and knowing why a given point is bad turns that scan into a prediction. It also explains a resolution loss that would otherwise be attributed to noise of unknown origin.

The same reasoning applies to the noise itself. A model that reproduces the measured spectrum from four identified sources tells an experiment which term limits its resolution and therefore which one is worth attacking, whether by a colder bath, a different tungsten volume, or a change in the readout. The correlated noise seen across the tower channels is a sharper case. An optimal filter built from per-channel spectra assumes the channels are independent. If they are coupled through the substrate, as the multi-channel model says they are, the correct filter is built from the full cross-spectral density matrix, and whatever resolution that recovers costs nothing in hardware.

The energy scale rests on the same simulation. A detector reports not the recoil energy but the phonon energy that reached its sensors, and the relation between the two runs through the ionization yield, the NTL amplification, and the fraction of the phonon population a sensor absorbs. The discrete quanta peaks in the HVeV spectrum anchor that chain at known numbers of electron-hole pairs, and a simulation tuned to reproduce them is one whose prediction of the collected energy can be trusted. An error in the energy scale does not weaken a limit, it moves it, and it moves it along the axis on which every experiment is compared against every other.

The parameters extracted in this work are physical rather than phenomenological, and that distinction is what carries them beyond the devices measured here. A heat capacity, a Kapitza conductance, an electron-phonon coupling, and a critical temperature are properties of materials and geometries, and they scale to devices that do not yet exist. A model built from them can be asked what a larger crystal, a different sensor coverage, or a different transition temperature would do before any of it is fabricated, which is the difference between a model and a fit. The substrate result is the clearest instance. If the substrate warms under bias and its relaxation shapes the tail of every pulse, then the thermal path from the crystal to the bath, the mounting, and the contact area are design variables that set the pulse shape of the next detector, and they have not generally been treated as such. The stability condition is a design variable in the same sense, since the readout inductance is chosen rather than given. Solid-state detectors are the technology carrying the search below a GeV, and the generation after this one will be designed rather than discovered. Designing it requires a detector model that predicts, and this work is a contribution toward one.

6.4 Limitations

The results above rest on approximations that bound them, and it is worth collecting the boundaries in one place.

The transport and phonon simulation carries the limitations of G4CMP. Optical phonons are not propagated but are taken to downconvert instantaneously, which is sound at millikelvin temperatures and inaccurate in the moments after a high-energy decay. Crystal momentum is not conserved in an intervalley scattering event, since the electron is relocated to the new valley rather

than being rotated by the emission of an optical phonon, so the phonon spectrum produced by intervalley scattering at high field is approximate. Holes are treated as strictly isotropic, and the small measured anisotropy of hole propagation is not included.

The sensor model rests on a phenomenological resistance surface. The tanh parameterization of $R(T, I)$ is fitted to data rather than derived, and the Ginzburg-Landau critical-current relation used to locate the phase boundary is a near- T_c depairing form applied across the full boundary, which is an extrapolation beyond its formal validity. The electron-phonon exponent is fixed at $n = 5$ rather than fitted. The linearized dynamics hold for small signals and say nothing about a deposition large enough to drive the resistance across a significant fraction of the transition, where the nonlinear terms in the resistance surface govern the response.

The ionization yield below a keV is the least certain input in the whole chain. The models on offer disagree with one another by a large margin precisely in the recoil energy range where sub-GeV dark matter would deposit, and that disagreement propagates directly into the energy scale of a high-voltage detector, where the yield multiplies the NTL amplification.

6.5 Outlook

6.5.1 A Semi-Analytical Model for Joint Parameter Extraction

The obstacle to a full parameter extraction is that the forward model is split in two. The crystal response is produced by a Monte Carlo simulation that is expensive, stochastic, and non-differentiable, and the sensor response is a set of ordinary differential equations that is none of those things. A sampler cannot be run through the pair, so the crystal parameters and the electrothermal parameters have never been varied together, and the degeneracy between them has never been mapped.

A surrogate model removes the obstacle. The phonon power arriving at the sensor can be written as a sum of exponentials,

$$P_{\text{signal}}(t; \theta_c) = \sum_k A_k(\theta_c) \exp\left(-\frac{t}{\tau_k(\theta_c)}\right), \quad (6-2)$$

in which the amplitudes and time constants depend on the crystal parameters θ_c . Those dependences are unknown but they are smooth, and they can be learned. Sampling θ_c across its plausible range, running the simulation at each design point, and fitting Eq. (6-2) to the resulting phonon power gives a training set from which a regression model can learn $A_k(\theta_c)$ and $\tau_k(\theta_c)$ directly. With those functions in hand the crystal response becomes an analytic, differentiable expression, the driving term of the TES equations is available in closed form, and the whole forward model can be sampled in a single posterior over the crystal and electrothermal parameters at once. The correlations and degeneracies between the two blocks then fall out of the covariance of that posterior, which is the object this work has needed and never had.

Four things decide whether such a program yields a trustworthy posterior rather than a fast one.

The exponentials must be anchored to physical processes. Fitting a free sum of exponentials is an inverse Laplace problem, and it is badly conditioned. Two similar time constants trade against one another, the fitted coefficients wander, and a surrogate trained on them will learn the wandering rather than the physics. Tying each term to a process that has a name, the quasiparticle collection in the fins, the absorption of athermal phonons from the crystal, and the relaxation of the substrate, makes the decomposition well posed. If it remains unstable, a decomposition into principal components of the simulated pulse family avoids committing to a functional form at all, and the surrogate can learn the component coefficients instead.

The surrogate must be validated where it will be used. The test that matters is not whether it reproduces the training set but whether parameters tuned at one bias voltage predict the pulse at another, and whether parameters tuned on a silicon HVeV device predict a larger HV tower detector. Both are falsifiable, and a failure of either would be as informative as a success.

6.6 Closing

The exclusion curves that opened this dissertation are drawn from events a detector accepted and events it threw away, and the line between the two is set by how well the detector is understood. A tungsten film forty nanometers thick, a substrate that will not stay at the

temperature of its bath, and a population of phonons scattering through a crystal are a long way from a cross section plotted against a mass. They are nonetheless what the cross section is measured with, and the accuracy of the one is the accuracy of the other.

APPENDIX A HIGHER-ORDER RESISTANCE EXPANSION

The two-variable Taylor expansion of the resistance surface $R(T, I_s)$ around the operating point (T_0, I_0) is

$$R(T_0 + \delta T, I_0 + \delta I) = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\delta T \frac{\partial}{\partial T} + \delta I \frac{\partial}{\partial I_s} \right)^k R \Big|_0. \quad (\text{A-1})$$

Applying the binomial theorem to the differential operator at each order, the first three contributions are

$$\delta R = \frac{\partial R}{\partial T} \Big|_0 \delta T + \frac{\partial R}{\partial I_s} \Big|_0 \delta I, \quad (\text{A-2})$$

$$\delta^{(2)} R = \frac{1}{2} \left[\frac{\partial^2 R}{\partial T^2} \Big|_0 (\delta T)^2 + 2 \frac{\partial^2 R}{\partial T \partial I_s} \Big|_0 \delta T \delta I + \frac{\partial^2 R}{\partial I_s^2} \Big|_0 (\delta I)^2 \right], \quad (\text{A-3})$$

$$\delta^{(3)} R = \frac{1}{6} \left[\frac{\partial^3 R}{\partial T^3} \Big|_0 (\delta T)^3 + 3 \frac{\partial^3 R}{\partial T^2 \partial I_s} \Big|_0 (\delta T)^2 \delta I + 3 \frac{\partial^3 R}{\partial T \partial I_s^2} \Big|_0 \delta T (\delta I)^2 + \frac{\partial^3 R}{\partial I_s^3} \Big|_0 (\delta I)^3 \right]. \quad (\text{A-4})$$

Rewriting the displacements as $\delta T = T_0 \epsilon_T$ and $\delta I = I_0 \epsilon_I$ pulls the dimensional scales into the partial derivatives. At every order, the same combination $T_0^m I_0^n R_0^{-1} \partial^{m+n} R / (\partial T^m \partial I_s^n)$ multiplies the corresponding monomial $\epsilon_T^m \epsilon_I^n$, motivating the dimensionless sensitivities

$$\gamma_{mn} = \frac{T_0^m I_0^n}{R} \frac{\partial^{m+n} R}{\partial T^m \partial I_s^n} \Big|_0, \quad (\text{A-5})$$

with the pure-temperature and pure-current sensitivities $\alpha_n \equiv \gamma_{n0}$ and $\beta_n \equiv \gamma_{0n}$:

$$\alpha_n = \frac{T_0^n}{R} \frac{\partial^n R}{\partial T^n} \Big|_0, \quad \beta_n = \frac{I_0^n}{R} \frac{\partial^n R}{\partial I_s^n} \Big|_0, \quad (\text{A-6})$$

satisfying $\gamma_{n0} = \alpha_n$, $\gamma_{0n} = \beta_n$, and $\gamma_{00} = 1$. The standard first-order parameters of Irwin and Hilton [45] are recovered as $\alpha \equiv \alpha_1$ and $\beta \equiv \beta_1$. The full expansion collapses to

$$R(T_0 + \delta T, I_0 + \delta I) = R_0 \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{\gamma_{mn}}{m! n!} \epsilon_T^m \epsilon_I^n. \quad (\text{A-7})$$

The first three orders, written in the compact notation, are

$$\delta R = R_0 [\alpha \epsilon_T + \beta \epsilon_I] , \quad (\text{A-8})$$

$$\delta^{(2)} R = \frac{R_0}{2} [\alpha_2 \epsilon_T^2 + 2 \gamma_{11} \epsilon_T \epsilon_I + \beta_2 \epsilon_I^2] , \quad (\text{A-9})$$

$$\delta^{(3)} R = \frac{R_0}{6} [\alpha_3 \epsilon_T^3 + 3 \gamma_{21} \epsilon_T^2 \epsilon_I + 3 \gamma_{12} \epsilon_T \epsilon_I^2 + \beta_3 \epsilon_I^3] . \quad (\text{A-10})$$

The first-order result is the linearized coupling used in Section 3.3.1. The higher orders and the mixed sensitivities γ_{mn} with $m, n \geq 1$ enter when the perturbation is large enough that the linear approximation breaks down.

APPENDIX B EIGENVALUE ALGEBRA FOR THE LINEARIZED TES

This appendix collects the algebraic results referenced in Section 3.4: the closed-form eigenvalues of the cubic in each discriminant regime, the eigenvector and mode-weight expressions for the impulse response.

B.1 Closed-Form Eigenvalues

The depressed cubic $\mu^3 + p\mu + q = 0$ admits closed-form solutions whose form depends on the sign of the discriminant $\Delta = -4p^3 - 27q^2$.

Cardano's formula ($\Delta < 0$, one real root and a complex pair)

The substitution $\mu = u + v$ with the constraint $uv = -p/3$ converts the cubic into

$$u^3 + v^3 = -q, \quad u^3 v^3 = -\frac{p^3}{27}, \quad (\text{B-1})$$

so u^3 and v^3 are the roots of the quadratic $z^2 + qz - p^3/27 = 0$:

$$u^3, v^3 = -\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}. \quad (\text{B-2})$$

When $\Delta < 0$, the inner square root is real, and the cube roots can be taken as real. Restoring

$\lambda = \mu - a_2/3$,

$$\lambda_r = -\frac{a_2}{3} + \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}. \quad (\text{B-3})$$

Once λ_r is known, the complex pair follows from the quadratic factor of the original polynomial:

dividing $P(\lambda)$ by $(\lambda - \lambda_r)$ gives $\lambda^2 + b_1\lambda + b_0$ with

$$b_1 = a_2 + \lambda_r, \quad b_0 = -\frac{a_0}{\lambda_r}, \quad (\text{B-4})$$

and the complex eigenvalues are

$$\lambda_{\pm} = -\frac{b_1}{2} \pm i \sqrt{b_0 - \frac{b_1^2}{4}} = -\frac{1}{\tau_{\text{osc}}} \pm i \omega_{\text{ETO}}, \quad (\text{B-5})$$

with damping rate $1/\tau_{\text{osc}} = b_1/2$ and oscillation frequency $\omega_{\text{ETO}} = \sqrt{b_0 - b_1^2/4}$.

Viète trigonometric formula ($\Delta > 0$, three real roots)

When $\Delta > 0$, Cardano's formula expresses real roots as sums of complex conjugate cube roots, which is computationally awkward. The substitution $\mu = 2\sqrt{-p/3} \cos \theta$ avoids this. Using the triple-angle identity $\cos 3\theta = 4 \cos^3 \theta - 3 \cos \theta$, the depressed cubic reduces to

$$\cos 3\theta = \frac{3q}{2p} \sqrt{-\frac{3}{p}}, \quad (\text{B-6})$$

which has three solutions θ_k separated by $2\pi/3$. Restoring $\lambda = \mu - a_2/3$,

$$\lambda_k = -\frac{a_2}{3} + \frac{2}{3} \sqrt{a_2^2 - 3a_1} \cos \left[\frac{1}{3} \arccos \left(\frac{2a_2^3 - 9a_2a_1 + 27a_0}{2(a_2^2 - 3a_1)^{3/2}} \right) - \frac{2\pi k}{3} \right], \quad k = 0, 1, 2. \quad (\text{B-7})$$

The leading term $-a_2/3$ is the mean eigenvalue, the prefactor $(2/3)\sqrt{a_2^2 - 3a_1}$ sets the spread around the mean, and the arccosine argument encodes the asymmetry among the three roots.

B.2 Eigenvectors and Mode Weights

The right eigenvector $\mathbf{v}_k$ for eigenvalue λ_k satisfies $(\mathbf{A} - \lambda_k \mathbf{I}) \mathbf{v}_k = 0$. Writing $\mathbf{v}_k = (v_{I,k}, v_{T,k}, v_{S,k})^T$, the three component equations are

$$(A_{11} - \lambda_k) v_{I,k} + A_{12} v_{T,k} = 0, \quad (\text{B-8})$$

$$A_{21} v_{I,k} + (A_{22} - \lambda_k) v_{T,k} + A_{23} v_{S,k} = 0, \quad (\text{B-9})$$

$$A_{32} v_{T,k} + (A_{33} - \lambda_k) v_{S,k} = 0, \quad (\text{B-10})$$

where the sparsity $A_{13} = A_{31} = 0$ has been used. The first and third equations express $v_{I,k}$ and $v_{S,k}$ in terms of $v_{T,k}$ alone:

$$v_{I,k} = \frac{A_{12}}{\lambda_k - A_{11}} v_{T,k}, \quad v_{S,k} = \frac{A_{32}}{\lambda_k - A_{33}} v_{T,k}. \quad (\text{B-11})$$

Normalizing $v_{T,k} = 1$ gives Equation 3-94. The second equation is automatically satisfied because λ_k is a root of the characteristic polynomial.

Mode weights via partial fractions

The impulse response is obtained by Laplace-transforming the linearized equations. For an impulsive thermal drive $\delta P_{\text{signal}}(t) = E_0 \delta(t)$, the current response in the s -domain is

$$\delta \hat{I}(s) = \frac{E_0}{C} \frac{N(s)}{P(s)}, \quad (\text{B-12})$$

where $P(s) = \prod_k (s - \lambda_k)$ is the characteristic polynomial evaluated at s and $N(s)$ is the numerator polynomial set by the Cramer's-rule cofactor for the current channel. Working out the cofactor with the matrix elements of Equation 3-68 gives

$$N(s) = A_{12}(s - A_{33}). \quad (\text{B-13})$$

Partial-fraction expansion in the s -plane,

$$\frac{A_{12}(s - A_{33})}{\prod_k (s - \lambda_k)} = \sum_k \frac{A_{12}(\lambda_k - A_{33})}{P'(\lambda_k)} \frac{1}{s - \lambda_k}, \quad (\text{B-14})$$

gives the residue at each pole, which is the mode weight of Equation 3-95. Inverse-transforming term by term yields the three-exponential impulse response.

Sum rule

The sum rule $\sum_k v_{I,k} w_{T,k} = 0$ follows from a residue argument. The numerator $A_{12}(s - A_{33})$ has degree one in s , and the denominator $P(s) = \prod_k (s - \lambda_k)$ has degree three, so the rational function decays as $1/s^2$ at large $|s|$. The contour integral around a circle of radius R vanishes as $R \rightarrow \infty$, but it also equals the sum of residues at the three poles. Therefore the sum of mode weights is zero, which translates back to $\delta I(0^+) = 0$ in the time domain.

ETO mode weight in amplitude-phase form

In the ETO regime $\Delta < 0$, the complex pair $\lambda_{\pm} = -1/\tau_{\text{osc}} \pm i\omega_{\text{ETO}}$ has complex-conjugate mode weights $c_{\pm} = v_{I,\pm} w_{T,\pm}$. Writing $c_+ = a + ib$ in terms of its real and imaginary parts,

$$c_- = \overline{c_+} = a - ib, \quad |c_{\pm}| = \sqrt{a^2 + b^2}, \quad \phi = \arctan(b/a). \quad (\text{B-15})$$

The contribution of the complex pair to the impulse response is

$$c_+ e^{\lambda_+ t} + c_- e^{\lambda_- t} = 2 e^{-t/\tau_{\text{osc}}} [a \cos(\omega_{\text{ETO}} t) - b \sin(\omega_{\text{ETO}} t)] = 2 |c_{\pm}| e^{-t/\tau_{\text{osc}}} \cos(\omega_{\text{ETO}} t + \phi), \quad (\text{B-16})$$

which is the form used in Equation 3-98.

APPENDIX C NUMERICAL INTEGRATION OF THE TES ARRAY EQUATIONS

The simulation exercised against the pulse data in Chapter 4 advances an array of N transition-edge sensors wired in parallel within a single phonon channel. Integrating the electrothermal equations of Section 3.2.2.1 with a standard explicit solver, rather than the slower and less robust differential-algebraic solver that the equations call for as written, requires a change of variables that removes the algebraic constraint introduced by the critical-current dependence of the resistance. The construction is due to J. S. Wilson [55], and is reproduced here in the notation of Chapter 3. The substrate temperature, held fixed in the original treatment, is afterward promoted to a dynamical variable, which recovers the two-thermal-block system of Section 4.6. The bias circuit is that of Figure 3-3, with the N sensors occupying the branch drawn there as a single variable resistor, and the electrical conductance of a sensor is written $\mathcal{G}_i \equiv 1/R_i$ in calligraphic type to keep it distinct from the thermal conductance G .

C.1 The coupled equations for the sensor array

The N sensors sit in parallel between the same two nodes of the circuit, so their resistances combine into a single channel resistance $R = (\sum_k \mathcal{G}_k)^{-1}$. Writing the two Kirchhoff loops of Figure 3-3 and eliminating the current in the bias branch leaves one equation for the current I_s through the inductor,

$$L \frac{dI_s}{dt} = V_{\text{th}} - I_s \left(R_l + \frac{1}{\sum_k \mathcal{G}_k} \right), \quad (\text{C-1})$$

which is Equation 3-44 with the channel resistance resolved into the parallel sum of the sensor conductances, and with the Thévenin source V_{th} and load resistance R_l carrying the definitions of Chapter 3. The inductor current divides among the sensors in proportion to their conductance,

$$I_i = I_s \frac{\mathcal{G}_i}{\sum_k \mathcal{G}_k}. \quad (\text{C-2})$$

Each sensor exchanges power with its surroundings through resistive Joule heating, cooling to the substrate by way of the tungsten electron-phonon coupling, and the athermal-phonon power

delivered through the aluminum collection fins. Its electron temperature evolves as

$$C \frac{dT_i}{dt} = \frac{I_i^2}{\mathcal{G}_i} - \Sigma_W \mathcal{V}_W (T_i^n - T_{\text{subst}}^n) + P_{\text{signal},i}(t), \quad (\text{C-3})$$

with C the per-sensor electronic heat capacity, held constant in the implementation, $\mathcal{V}_W$ the tungsten volume of one sensor, $n = 5$ for clean-limit electron-phonon coupling, and $P_{\text{signal},i}(t)$ the athermal-phonon power reaching sensor i from the crystal simulation. The substrate temperature T_{subst} enters as a fixed parameter for now. The Joule term is written $I_i^2/\mathcal{G}_i$ in place of $I_i^2 R_i$ for the numerical reason that follows.

The resistance enters Equations C-1 and C-3 through the broadened transition surface of Chapter 3, $R_i = \frac{1}{2} R_n [1 + \tanh((T_i - T_c)/T_w)]$, with T_w the transition width. Deep in the superconducting state $R_i \rightarrow 0$, and the parallel combination $1/\sum_k (1/R_k)$ would divide by a vanishing resistance. The conductance form avoids this,

$$\mathcal{G}_i = \frac{2}{R_n} \cosh\left(\frac{T_i - T_c}{T_w}\right) e^{-(T_i - T_c)/T_w} = \frac{1}{R_n} \left(1 + e^{-2(T_i - T_c)/T_w}\right), \quad (\text{C-4})$$

which is finite for every temperature and grows large but never singular as a sensor turns fully superconducting.

The conductances of Equation C-4 depend on the sensor temperatures alone. Equations C-1, C-2, C-3, and C-4 can therefore be arranged into a purely differential system by substituting the conductances directly,

$$\frac{dI_s}{dt} = \frac{1}{L} \left[V_{\text{th}} - I_s \left(R_l + \frac{1}{\sum_k \mathcal{G}_k} \right) \right], \quad (\text{C-5})$$

$$\frac{dT_i}{dt} = \frac{1}{C} \left[\left(\frac{I_s}{\sum_k \mathcal{G}_k} \right)^2 \mathcal{G}_i - \Sigma_W \mathcal{V}_W (T_i^n - T_{\text{subst}}^n) + P_{\text{signal},i} \right], \quad (\text{C-6})$$

a form that any explicit integrator handles directly. This convenience survives only as long as the conductances are functions of temperature alone.

C.2 Critical-current suppression

A finite sensor current suppresses the transition temperature below its zero-current value. The Ginzburg–Landau depairing boundary sets

$$T_{c,i}^{\text{eff}} = T_c \left[1 - \left(\frac{|I_i|}{I_c} \right)^{2/3} \right], \quad (\text{C-7})$$

with I_c the critical current of one sensor. Replacing T_c by $T_{c,i}^{\text{eff}}$ in the conductance of Equation C-4 makes the conductance carry a dependence on the sensor current I_i alongside its dependence on temperature. Because I_i depends through Equation C-2 on the full set of conductances, the conductances are no longer fixed by the temperatures alone, and the direct substitution that produced Equations C-5 and C-6 is no longer available. The system reverts to a differential-algebraic form, with the conductances determined implicitly at each step.

C.3 Reparameterization and analytic inversion

The algebraic constraint is removed by changing variables from the sensor temperatures to the signed distance of each sensor from its phase boundary,

$$\mathcal{A}_i \equiv \frac{T_i - T_{c,i}^{\text{eff}}}{T_w}, \quad (\text{C-8})$$

the quantity introduced in Equation 3-9. Taken as the primary variables alongside I_s , the set $\{\mathcal{A}_i\}$ recovers the conductance as a function of $\mathcal{A}_i$ alone,

$$\mathcal{G}_i(\mathcal{A}_i) = \frac{2}{R_n} \cosh \mathcal{A}_i e^{-\mathcal{A}_i}, \quad (\text{C-9})$$

the sensor currents and effective transition temperatures as functions of $\{\mathcal{A}_j\}$ and I_s through Equations C-2 and C-7, and the sensor temperatures as

$$T_i = T_w \mathcal{A}_i + T_{c,i}^{\text{eff}}. \quad (\text{C-10})$$

The time derivative of the temperature follows by the chain rule,

$$\dot{T}_i = \sum_j \frac{\partial T_i}{\partial \mathcal{A}_j} \dot{\mathcal{A}}_j + \frac{\partial T_i}{\partial I_s} \dot{I}_s. \quad (\text{C-11})$$

With the critical-current term in place, $T_{c,i}^{\text{eff}}$ depends on the currents and therefore on every $\mathcal{A}_j$ and on I_s , so the derivatives in Equation C-11 no longer collapse. Collecting the heat-flow powers of Equation C-3 into $P_i \equiv I_i^2/\mathcal{G}_i - \Sigma_W \mathcal{V}_W (T_i^n - T_{\text{subst}}^n) + P_{\text{signal},i}$ and using $C\dot{T}_i = P_i$ turns Equation C-11 into a linear system for the $\dot{\mathcal{A}}_j$,

$$\sum_j \frac{\partial T_i}{\partial \mathcal{A}_j} \dot{\mathcal{A}}_j = \frac{P_i}{C} - \frac{\partial T_i}{\partial I_s} \dot{I}_s. \quad (\text{C-12})$$

Recovering an explicit form requires inverting the matrix $\partial T_i/\partial \mathcal{A}_j$.

The partial derivatives factor through the current and the effective transition temperature,

$$\frac{\partial T_i}{\partial \mathcal{A}_j} = T_w \delta_{ij} + \frac{dT_{c,i}^{\text{eff}}}{dI_i} \frac{\partial I_i}{\partial \mathcal{A}_j}, \quad \frac{\partial T_i}{\partial I_s} = \frac{dT_{c,i}^{\text{eff}}}{dI_i} \frac{\partial I_i}{\partial I_s}, \quad (\text{C-13})$$

with the constituent factors

$$\frac{dT_{c,i}^{\text{eff}}}{dI_i} = -\frac{2}{3} \frac{T_c}{I_c} \text{sgn}(I_i) \left(\frac{|I_i|}{I_c} \right)^{-1/3}, \quad (\text{C-14})$$

$$\frac{d\mathcal{G}_j}{d\mathcal{A}_j} = -\frac{2}{R_n} e^{-2\mathcal{A}_j}, \quad (\text{C-15})$$

$$\frac{\partial I_i}{\partial \mathcal{A}_j} = \frac{I_s}{\sum_k \mathcal{G}_k} \frac{d\mathcal{G}_j}{d\mathcal{A}_j} \left(\delta_{ij} - \frac{\mathcal{G}_i}{\sum_k \mathcal{G}_k} \right), \quad (\text{C-16})$$

$$\frac{\partial I_i}{\partial I_s} = \frac{\mathcal{G}_i}{\sum_k \mathcal{G}_k}. \quad (\text{C-17})$$

The derivative of the conductance in Equation C-15 is negative, since $\mathcal{G}_i = R_n^{-1} (1 + e^{-2\mathcal{A}_i})$ falls as a sensor moves toward the normal state. Substituting these factors into Equation C-13 gives a

matrix with the diagonal-plus-rank-one structure $\partial T_i / \partial \mathcal{A}_j = a_i \delta_{ij} + u_i v_j$, where

$$a_i = T_w + B_i I_s \frac{2}{R_n} e^{-2\mathcal{A}_i}, \quad (\text{C-18})$$

$$u_i = B_i I_i, \quad (\text{C-19})$$

$$v_j = -\frac{2}{R_n} e^{-2\mathcal{A}_j}, \quad (\text{C-20})$$

$$B_i \equiv \frac{2 T_c}{3 I_c} \frac{1}{\sum_k \mathcal{G}_k} \text{sgn}(I_i) \left(\frac{|I_i|}{I_c} \right)^{-1/3}. \quad (\text{C-21})$$

The companion derivative is $\partial T_i / \partial I_s = -B_i \mathcal{G}_i$.

A matrix of the form $a_i \delta_{ij} + u_i v_j$ is inverted in closed form by the Sherman-Morrison identity,

$$\left(\frac{\partial T_i}{\partial \mathcal{A}_j} \right)^{-1} = \frac{\delta_{ij}}{a_i} - \frac{(u_i/a_i)(v_j/a_j)}{1 + \sum_k u_k v_k/a_k}, \quad (\text{C-22})$$

which requires only the N diagonal reciprocals $1/a_i$ and a single scalar sum, never the factorization of a dense matrix. Writing the right-hand side of Equation C-12 as

$$D_i \equiv \frac{P_i}{C} + B_i \mathcal{G}_i \dot{I}_s, \quad (\text{C-23})$$

where $\dot{I}_s$ is supplied by Equation C-1, the inverse yields the sensor evolution in explicit form,

$$\dot{\mathcal{A}}_i = \frac{D_i}{a_i} - \frac{u_i/a_i}{1 + \sum_k u_k v_k/a_k} \sum_j \frac{v_j}{a_j} D_j. \quad (\text{C-24})$$

Setting the critical current aside provides a check on the algebra. As $I_c \rightarrow \infty$ the coefficient $B_i \rightarrow 0$, so $u_i \rightarrow 0$, $a_i \rightarrow T_w$, and $D_i \rightarrow P_i/C$, and Equation C-24 collapses to $\dot{\mathcal{A}}_i = P_i/(T_w C)$, the zero-current reduction of Equation C-3.

C.4 A dynamical substrate temperature

The treatment to this point holds the substrate at a fixed temperature, which is the boundary condition used for the single-thermal-block fits of Section 4.4. Reproducing the long tail of the pulse requires the substrate to carry its own temperature and heat capacity, exchanging power with

the sensors through the electron-phonon link and with the bath through a Kapitza link, as in Section 4.6. The substrate is taken as a single lumped thermal node shared by the sensors, with lateral conduction between sensors neglected. Its temperature obeys

$$C_{\text{subst}} \frac{dT_{\text{subst}}}{dt} = \sum_{i=1}^N \Sigma_W \mathcal{V}_W(T_i^n - T_{\text{subst}}^n) - \mathcal{K} A_c (T_{\text{subst}}^m - T_b^m), \quad (\text{C-25})$$

with C_{subst} the substrate heat capacity, $\mathcal{K} A_c$ the substrate-to-bath Kapitza term written K_{s-b} in Section 4.6, $m = 4$, and T_b now the true bath temperature. The sum runs over the sensors that deposit power into the substrate, and generalizes to all sensors of all channels when M channels share the substrate, following the shared-substrate accounting of Section 3.3.

The extension leaves the inversion of Section C.3 intact. The change of variables rests on Equation C-10, and $T_{c,i}^{\text{eff}}$ depends through Equation C-7 on the sensor currents alone, hence on $\{\mathcal{A}_j\}$ and I_s and not on T_{subst} . The matrix $\partial T_i / \partial \mathcal{A}_j$, the companion derivative $\partial T_i / \partial I_s$, and the Sherman-Morrison inverse of Equation C-22 are therefore unchanged. The substrate temperature enters only through the cooling power in P_i , where the fixed value is replaced by the state variable $T_{\text{subst}}(t)$ inside D_i , and Equation C-25 is itself already explicit. The augmented state $(\{\mathcal{A}_i\}, I_s, T_{\text{subst}})$ thus advances under a fully explicit right-hand side, adding one scalar equation without a second matrix inversion.

A fixed bias with no signal sets the operating point through the vanishing of the time derivatives. The substrate balance then reads

$$\sum_{i=1}^N \Sigma_W \mathcal{V}_W((T_i^0)^n - (T_{\text{subst}}^0)^n) = \mathcal{K} A_c ((T_{\text{subst}}^0)^m - T_b^m), \quad (\text{C-26})$$

which places the steady-state substrate temperature T_{subst}^0 above the bath whenever the sensors are biased into the transition, by the amount needed to pass the Joule power through the Kapitza link to the bath. The magnitude of this offset and its dependence on the number of coupled channels are taken up in Section 4.6.

C.5 The explicit system

The array advances under the explicit equations

$$\frac{dI_s}{dt} = \frac{1}{L} \left[V_{\text{th}} - I_s \left(R_l + \frac{1}{\sum_k \mathcal{G}_k} \right) \right], \quad (\text{C-27})$$

$$\frac{d\mathcal{A}_i}{dt} = \frac{D_i}{a_i} - \frac{u_i/a_i}{1 + \sum_k u_k v_k / a_k} \sum_j \frac{v_j}{a_j} D_j, \quad (\text{C-28})$$

$$\frac{dT_{\text{subst}}}{dt} = \frac{1}{C_{\text{subst}}} \left[\sum_i \Sigma_W \mathcal{V}_W (T_i^n - T_{\text{subst}}^n) - \mathcal{K} A_c (T_{\text{subst}}^m - T_b^m) \right], \quad (\text{C-29})$$

in which the auxiliaries follow from the current state in order: the conductances $\mathcal{G}_i$ (Equation C-9) and their sum, the currents I_i (Equation C-2), the effective transition temperatures $T_{c,i}^{\text{eff}}$ (Equation C-7) and the temperatures T_i (Equation C-10), the Sherman-Morrison factors a_i, u_i, v_j , and B_i (Equations C-18–C-21), and finally $\dot{I}_s$ from Equation C-27 and D_i from Equation C-23. No step requires a matrix factorization, and the array integrates with the same explicit solver used for a single sensor.

LIST OF REFERENCES

- [1] E. E. Salpeter. The luminosity function and stellar evolution. *ApJ*, 121:161, 1955.
- [2] A. M. Hopkins. The dawes review 8: Measuring the stellar initial mass function. *Publications of the Astronomical Society of Australia*, 35:e039, 2018.
- [3] Vera C. Rubin, W. Kent Ford, and Norbert Thonnard. Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc). *The Astrophysical Journal*, 238:471–487, 1980.
- [4] T. S. van Albada, J. N. Bahcall, K. Begeman, and R. Sancisi. Distribution of dark matter in the spiral galaxy ngc 3198. *ApJ*, 295:305, 1985.
- [5] F. Zwicky. Die rotverschiebung von extragalaktischen nebeln. *Helvetica Physica Acta*, 6:110–127, 1933.
- [6] D. Clowe et al. A direct empirical proof of the existence of dark matter. *ApJ*, 648:L109, 2006.
- [7] Planck Collaboration, N. Aghanim, et al. Planck 2018 results. VI. Cosmological parameters. *Astronomy & Astrophysics*, 641:A6, 2020.
- [8] M. Milgrom. A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. *The Astrophysical Journal*, 270:365–370, 1983.
- [9] Richard H. Cyburt, Brian D. Fields, Keith A. Olive, and Tsung-Han Yeh. Big bang nucleosynthesis: Present status. *Reviews of Modern Physics*, 88:015004, 2016.
- [10] C. Alcock et al. The macho project: Microlensing results from 5.7 years of lmc observations. *ApJ*, 542:281, 2000.
- [11] J. I. Read. The local dark matter density. *Journal of Physics G: Nuclear and Particle Physics*, 41(6):063101, 2014.
- [12] D. Baxter et al. Recommended conventions for reporting results from direct dark matter searches. *Eur. Phys. J. C*, 81(10):907, 2021.
- [13] M. C. Smith, G. R. Ruchti, A. Helmi, et al. The RAVE survey: constraining the local galactic escape speed. *Monthly Notices of the Royal Astronomical Society*, 379(2):755–772, 2007.
- [14] Julien Billard et al. Direct detection of dark matter—APPEC committee report. *Reports on Progress in Physics*, 85:056201, 2022.
- [15] Richard H. Helm. Inelastic and elastic scattering of 187-MeV electrons from selected even-even nuclei. *Physical Review*, 104(5):1466–1475, 1956.
- [16] J. D. Lewin and P. F. Smith. Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil. *Astroparticle Physics*, 6:87–112, 1996.

- [17] A. Migdal. Ionization of atoms accompanying α - and β -decay. *Journal of Physics (USSR)*, 4:449–453, 1941.
- [18] Masahiro Ibe, Wakutaka Nakano, Yutaro Shoji, and Kazumine Suzuki. Migdal effect in dark matter direct detection experiments. *Journal of High Energy Physics*, 2018(3):194, 2018.
- [19] Rouven Essig, Jeremy Mardon, and Tomer Volansky. Direct detection of sub-GeV dark matter. *Physical Review D*, 85:076007, 2012.
- [20] Rouven Essig, Marivi Fernández-Serra, Jeremy Mardon, Adrián Soto, Tomer Volansky, and Tien-Tien Yu. Direct detection of sub-GeV dark matter with semiconductor targets. *Journal of High Energy Physics*, 2016(5):046, 2016.
- [21] SuperCDMS Collaboration. SuperCDMS Dark Matter Limit Plotter. <https://supercdms.slac.stanford.edu/science-results/dark-matter-limit-plotter>, 2023. Curve compilation last updated 7 April 2023.
- [22] SuperCDMS Collaboration. Search for low-mass electron-recoil dark matter using a single-charge sensitive SuperCDMS-HVeV detector. *Phys. Rev. D*, 113:032001, 2026.
- [23] J. Billard, E. Figueroa-Feliciano, and L. Strigari. Implication of neutrino backgrounds on the reach of next generation dark matter direct detection experiments. *Phys. Rev. D*, 89(2):023524, 2014.
- [24] LZ Collaboration. Searches for light dark matter and evidence of coherent elastic neutrino-nucleus scattering of solar neutrinos with the LUX-ZEPLIN (Lz) experiment. *arXiv e-prints*, 2025. Submitted to Phys. Rev. Lett.
- [25] SuperCDMS Collaboration. Projected sensitivity of the SuperCDMS SNOLAB experiment. *Physical Review D*, 95:082002, 2017.
- [26] SuperCDMS Collaboration. Internal documentation. Technical report.
- [27] J. Lindhard, V. Nielsen, M. Scharff, and P. V. Thomsen. Integral equations governing radiation effects (Notes on atomic collisions, III). *Matematisk-fysiske Meddelelser Kongelige Danske Videnskabernes Selskab*, 33(10):1–42, 1963.
- [28] Y. Sarkis, A. Aguilar-Arevalo, and J. C. D’Olivo. Ionization efficiency for nuclear recoils in silicon from ~ 50 eV to 3 MeV. *Physical Review A*, 107(6):062811, 2023.
- [29] Chang-Hao Fang. Investigating ultra-low energy ionization yield from nuclear recoils in semiconductor detectors via molecular dynamics simulations. In *Proc. 19th Int. Conf. on Topics in Astroparticle and Underground Physics (TAUP2025)*, 2025.
- [30] SuperCDMS Collaboration. First measurement of the nuclear-recoil ionization yield in silicon at 100 eV. *Physical Review Letters*, 131(9):091801, 2023.
- [31] M. H. Kelsey, R. Agnese, Y. F. Alam, et al. G4CMP: Condensed matter physics simulation using the Geant4 toolkit. *Nucl. Instrum. Methods Phys. Res. A*, 1055:168473, 2023.

- [32] B. Cabrera, M. Pyle, R. Moffatt, K. Sundqvist, and B. Sadoulet. Oblique propagation of electrons in crystals of germanium and silicon at sub-kelvin temperature in low electric fields. *arXiv e-prints*, 2010.
- [33] Conyers Herring and Erich Vogt. Transport and deformation-potential theory for many-valley semiconductors with anisotropic scattering. *Physical Review*, 101(3):944–961, 1956.
- [34] A. Broniatowski. Intervalley scattering of hot electrons in germanium at millikelvin temperatures. *Journal of Low Temperature Physics*, 176(5-6):860–869, 2014.
- [35] N. A. Kurinsky. *The Low-Mass Limit: Dark Matter Detectors with eV-Scale Energy Resolution*. PhD thesis, Stanford University, 2018.
- [36] C. Jacoboni and L. Reggiani. The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials. *Reviews of Modern Physics*, 55(3):645–705, 1983.
- [37] Kyle Michael Sundqvist. *Carrier Transport and Related Effects in Detectors of the Cryogenic Dark Matter Search*. PhD thesis, University of California, Berkeley, 2012.
- [38] S. W. Leman. Invited review article: Physics and Monte Carlo techniques as relevant to cryogenic, phonon, and ionization readout of Cryogenic Dark Matter Search radiation detectors. *Review of Scientific Instruments*, 83(9):091101, 2012.
- [39] R. A. Moffatt et al. Spatial imaging of charge transport in silicon at low temperature. *Applied Physics Letters*, 114(3):032104, 2019.
- [40] C. Stanford et al. High-field spatial imaging of charge transport in silicon at low temperature. *AIP Advances*, 10(2):025316, 2020.
- [41] James P. Wolfe. *Imaging Phonons: Acoustic Wave Propagation in Solids*. Cambridge University Press, Cambridge, 1998.
- [42] S. Tamura. Spontaneous decay rates of la phonons in quasi-isotropic solids. *Physical Review B*, 31:2574(R), 1985.
- [43] S. Tamura. Isotope scattering of dispersive phonons in Ge. *Physical Review B*, 27(2):858–866, 1983.
- [44] S. B. Kaplan, C. C. Chi, D. N. Langenberg, J. J. Chang, S. Jafarey, and D. J. Scalapino. Quasiparticle and phonon lifetimes in superconductors. *Physical Review B*, 14(11):4854–4873, 1976.
- [45] K. D. Irwin and G. C. Hilton. Transition-edge sensors. In C. Enss, editor, *Cryogenic Particle Detection*, volume 99 of *Topics in Applied Physics*, pages 63–150. Springer, 2005.
- [46] M. Tinkham. *Introduction to Superconductivity*. McGraw-Hill, 2nd edition, 1996.
- [47] C. J. Gorter and H. Casimir. On supraconductivity I. *Physica*, 1:306–320, 1934.

- [48] J. Bardeen, L. N. Cooper, and J. R. Schrieffer. Theory of superconductivity. *Physical Review*, 108:1175–1204, 1957.
- [49] V. L. Ginzburg and L. D. Landau. On the theory of superconductivity. *Zh. Eksp. Teor. Fiz.*, 20:1064–1082, 1950.
- [50] F. B. Silsbee. A note on electrical conduction in metals at low temperatures. *Journal of the Washington Academy of Sciences*, 6:597–602, 1916.
- [51] J. M. Kreikebaum. *Superconducting Thin Film Properties and Fabrication Techniques for TES Microcalorimeters*. PhD thesis, Stanford University, 2015.
- [52] B. Cabrera. Introduction to TES physics. *Journal of Low Temperature Physics*, 151:82–93, 2008.
- [53] K. D. Irwin, G. C. Hilton, D. A. Wollman, and J. M. Martinis. Thermal-response time of superconducting transition-edge microcalorimeters. *Journal of Applied Physics*, 83(8):3978–3985, 1998.
- [54] D. A. Bennett, D. S. Swetz, R. D. Horansky, D. R. Schmidt, and J. N. Ullom. A two-fluid model for the transition shape in transition-edge sensors. *Journal of Low Temperature Physics*, 167:102–107, 2012.
- [55] J. S. Wilson. TESSim: A transition-edge-sensor simulation framework for SuperCDMS. SuperCDMS Collaboration software, 2018. Internal software package.
- [56] K. D. Irwin. An application of electrothermal feedback for high resolution cryogenic particle detection. *Applied Physics Letters*, 66(15):1998–2000, 1995.
- [57] F. C. Wellstood, C. Urbina, and John Clarke. Hot-electron effects in metals. *Phys. Rev. B*, 49(9):5942–5955, March 1994.
- [58] I. J. Maasilta. Complex impedance, responsivity and noise of transition-edge sensors: Analytical solutions for two- and three-block thermal models. *AIP Advances*, 2(4):042110, 2012.
- [59] 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. *Phys. Rev. D*, 104(3):032010, 2021.
- [60] P. Adari et al. EXCESS workshop: Descriptions of rising low-energy spectra. *SciPost Phys. Proc.*, 9:001, 2022.
- [61] P. Camus et al. CUTE: A cryogenic underground TEst facility at SNOLAB. *Front. Phys.*, 11:1319879, 2024.
- [62] S. R. Golwala. *Exclusion Limits on the WIMP-Nucleon Elastic-Scattering Cross-Section from the Cryogenic Dark Matter Search*. PhD thesis, University of California, Berkeley, 2000.

- [63] Kyle Kennard et al. Performance of a SuperCDMS HVeV detector with sub-eV energy resolution and single charge-sensitivity. 2026.
- [64] S. Tamura. Quasidiffusive propagation of phonons in silicon: Monte carlo calculations. *Physical Review B*, 48:13502, 1993.
- [65] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: The mcmc hammer. *Publications of the Astronomical Society of the Pacific*, 125:306, 2013.
- [66] M. J. Wilson, A. Zaytsev, B. von Krosigk, et al. Improved modelling of detector response effects in phonon-based crystal detectors used for dark matter searches. *arXiv preprint*, 2024.
- [67] P. Flubacher, A. J. Leadbetter, and J. A. Morrison. The heat capacity of pure silicon and germanium and properties of their vibrational frequency spectra. *Philosophical Magazine*, 4(39):273–294, 1959.

BIOGRAPHICAL SKETCH

David Sadek was born and raised in Al Menia, Egypt. He completed his undergraduate studies at Zewail City of Science and Technology, where he earned his Bachelor of Science in physics in 2020. During his time at Zewail City, he developed an interest in astroparticle physics and the problem of dark matter that led him to pursue graduate study in the United States.

He enrolled in the doctoral program in the Department of Physics at the University of Florida in 2021, joining the research group of Professor Tarek Saab. His doctoral research focused on the SuperCDMS direct detection experiment, with an emphasis on modeling phonon transport and the electrothermal response of cryogenic detectors instrumented with transition-edge sensors. He contributed to the characterization of sub-GeV dark matter detectors at the Cryogenic Underground TEST (CUTE) facility, to the commissioning of the SuperCDMS experiment at SNOLAB, and to the development and validation of simulation tools within the G4CMP framework. He received his Doctor of Philosophy from the University of Florida in the summer of 2026.

Soccer was David's first passion, long before physics. He had hoped to pursue the sport professionally, but set that ambition aside under the social expectation that academic success should come first. He has remained an avid player and follower of the game ever since. His broader intellectual interests range across history, geopolitics, and philosophy, and he enjoys long conversations with friends on the questions these subjects raise about the human condition.