

# Estimating the Energy Threshold of Phonon-mediated Superconducting Qubit Detectors Operated in an Energy-Relaxation Sensing Scheme

R. Linehan,<sup>1,\*</sup> I. Hernandez,<sup>2,1</sup> D. J. Temples,<sup>1</sup> S. Q. Dang,<sup>3,1</sup> D. Baxter,<sup>1,4</sup>  
 L. Hsu,<sup>1</sup> E. Figueroa-Feliciano,<sup>4,1</sup> R. Khatiwada,<sup>2,1</sup> K. Anyang,<sup>2,1</sup> D. Bowring,<sup>1</sup>  
 G. Bratrud,<sup>4,1</sup> G. Cancelo,<sup>1</sup> A. Chou,<sup>1</sup> R. Gualtieri,<sup>4,1</sup> K. Stifter,<sup>1,†</sup> and S. Sussman<sup>1</sup>

<sup>1</sup>*Fermi National Accelerator Laboratory, Batavia, IL 60510, USA*

<sup>2</sup>*Department of Physics, Illinois Institute of Technology, Chicago, IL, 60616*

<sup>3</sup>*Department of Physics, Cornell University, Ithaca, NY 14853, USA*

<sup>4</sup>*Department of Physics & Astronomy, Northwestern University, Evanston, IL 60208, USA*

(Dated: April 9, 2024)

In recent years, the lack of a conclusive detection of WIMP dark matter at the 10 GeV/c<sup>2</sup> mass scale and above has encouraged development of low-threshold detector technology aimed at probing lighter dark matter candidates. Detectors based on Cooper-pair-breaking sensors have emerged as a promising avenue for this detection due to the low (meV-scale) energy required for breaking a Cooper pair in most superconductors. Among them, devices based on superconducting qubits are interesting candidates for sensing due to their observed sensitivity to broken Cooper pairs. We have developed an end-to-end G4CMP-based simulation framework and have used it to evaluate performance metrics of qubit-based devices operating in a gate-based “energy relaxation” readout scheme, akin to those used in recent studies of qubit sensitivity to ionizing radiation. We find that for this readout scheme, the qubit acts as a phonon sensor with an energy threshold ranging down to  $\simeq 0.4$  eV for near-term performance parameters.

## I. Introduction

A wealth of astrophysical and cosmological evidence points to the abundance of a massive, cold, non-baryonic “dark” matter (DM) in today’s universe [1]. Despite dark matter’s abundance, experiments designed to directly observe it in the lab have so far only yielded null results [2–4], and as a result many of its fundamental properties such as its mass are not well known. Notably, direct detection experiments searching for particle-like DM candidates continue to exclude WIMP dark matter with masses in the GeV/c<sup>2</sup>-TeV/c<sup>2</sup> mass range, which has spurred interest in searching for other well-motivated particle-like DM candidates at lower masses down to 50 keV/c<sup>2</sup> [5].

While lower-mass dark matter searches benefit from an increased flux of DM passing through the detector relative to traditional WIMP searches, this rate increase comes at the cost of a drop in the average energy deposited during a DM interaction and a corresponding drop in the detector energy threshold required for observing scatters. For example, to observe inelastic scattering of a DM particle of mass  $\simeq 1$  MeV/c<sup>2</sup> on a target, a threshold of  $\simeq 1$  eV or lower is required [5]. This requirement has launched community-wide effort to develop and deploy sensors capable of observing sub-eV energy depositions. “Pairbreaking” sensors, which operate near 10 mK and by sensing broken Cooper pairs (Bogoliubov Quasiparticles/QPs) in superconductors, form

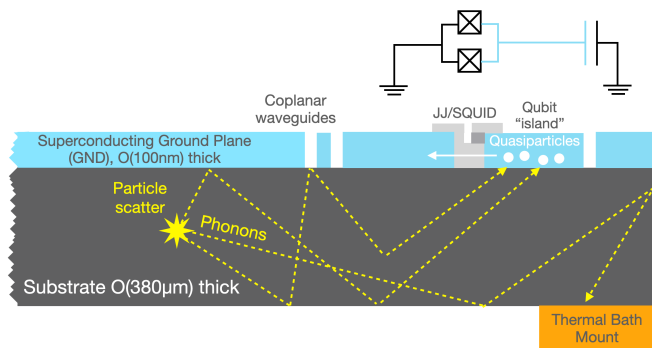


FIG. 1. A cross-sectional diagram of a segment of a qubit chip (not to scale), showing the substrate in gray, superconductor in blue, and thermalization mounts in orange. The qubit island and region around the JJ/SQUID are regions where enhanced QP density may be sensed. A simple circuit representation is shown above to illustrate how the qubit island (blue segment) is formed from a SQUID loop (the “inductor”) and capacitive gap between the island and ground plane.

a popular class of such low-threshold devices. Common among these are microwave kinetic inductance detectors (MKIDs) and devices based on transition-edge-sensors (TESs), both of which consist of thin O(100) nm superconductor patterned on top of a more massive substrate that acts as the dark matter target. Scatters in the substrate produce meV-scale athermal phonons which can break Cooper pairs in the superconducting sensor.

Energy resolutions of such pairbreaking detectors are often quoted in terms of the “in-sensor” resolution on the energy absorbed by a superconducting sensor and the “in-chip” resolution on the energy absorbed by the substrate, only a fraction of which makes it into any given

\* linehan3@fnal.gov

† Presently at SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

superconducting sensor. In-sensor energy resolutions at the 40 meV scale have been achieved in TES-based devices, with best-achieved in-chip energy resolutions at the O(eV) scale [6–9]. MKID in-sensor energy resolutions have been achieved at the 2 eV scale, with in-chip energy resolutions at the O(300 eV) scale [10].

Another candidate for low-threshold sensing uses a superconducting transmon qubit architecture. Commonly used as the fundamental hardware units in a superconducting quantum computer, transmon qubits are anharmonic LC circuits etched into a thin, O(100) nm superconducting film lying on top of a thicker, O(400)  $\mu\text{m}$  substrate as shown in Figure 1. They are composed of a superconducting “island” separated from a ground plane by a gap (the capacitor) spanned by a Josephson junction (JJ) or SQUID (the inductor). The flexibility in this device architecture has enabled their use in a variety of quantum sensing applications including single GHz-photon detection [11], THz photon detection [12, 13], ultra-sensitive magnetometry [14], and detection of in-chip charge and phonon bursts [15]. Many of these sensing modes rely on a demonstrated sensitivity of superconducting qubits to QPs in the leads of the JJ/SQUID [16–19], suggesting that qubits may also act as pairbreaking sensors via QP sensing schemes different from (and potentially competitive with) those in other pairbreaking devices. However, only a few studies have been performed to assess the sensitivity of such sensing schemes to in-substrate energy depositions [20, 21].

The objective of this work is to demonstrate a bottom-up, simulation-based estimate of the energy threshold of a phonon-mediated detector built from transmon qubits operated in an “energy relaxation” readout scheme. We begin in Section II by demonstrating the use of the G4CMP simulation tool in studying phonon transport between an interaction site and a qubit island. Section III then presents the Quantum Device Response (QDR) simulation tool, which we use to estimate the energy threshold for a single qubit island when using a gate-based “energy relaxation” readout scheme [18, 19]. We then explore in Section IV how both chip design and single-qubit response contribute to an overall chip energy threshold. We demonstrate model viability in Section V with a preliminary application of our modeling to data recently published in Ref [19]. We discuss major takeaways from this work in Section VI, and close in Section VII.

## II. Phonon Propagation from an Impact Site to a Qubit Sensor

The energy threshold of a qubit-based device is strongly dependent on how efficiently energy can propagate from an interaction site to the qubit islands. In this study, we use a Geant4-based particle tracking software called G4CMP [22] to model this phonon propagation through an example chip geometry (Figure 2): a silicon

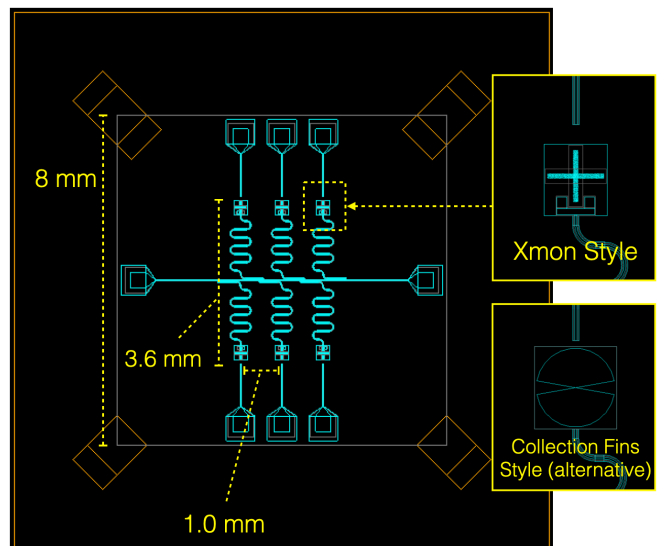


FIG. 2. Illustration of the baseline 6-qubit chip geometry used in this work’s simulation. Each Xmon qubit (top inset) is a small cross (“island”) about  $200\mu\text{m}$  across, and is coupled to a resonator that is itself coupled inductively to the central feedline. A biasing line and wirebonding pad are included for each qubit. The bottom inset is a visualization of an alternative style of qubit island that employs collection fins, as put forth by Ref. [20]. The chip (gray outline) is in contact with copper mounts (orange) on the underside of its four corners.

substrate of dimension  $8\text{ mm} \times 8\text{ mm} \times 380\ \mu\text{m}$  mounted on four copper corners for thermalization and patterned with six aluminum 100-nm-thick 2D flux-tunable qubits. Each qubit is modeled after the Xmon architecture commonly used for quantum computation, where the cross acts as the qubit island (and in our case, as the target for phonons) [23]. While this geometry is based on designs from the quantum computing community such as those in Ref. [15, 17] and is therefore not optimized for particle detection, we use it as a launchpad for discussing more optimal designs for dark matter detection. What follows in this section is functionally an inverse study to Ref. [24] in which we explore chip parameters to maximize the deposition of phonon energy into the qubit islands.

At an interaction site, energy may be injected into the detector medium in a number of ways: ionization by a high energy particle impact followed by electron-hole phonon scattering and recombination into phonons, direct phonon excitation through a particle impact, or phonon production from stress-induced crystal relaxation [15, 25, 26]. For most kinds of interactions that we consider, the majority of the initial energy will be deposited in the device substrate rather than in the much less massive superconducting layer into which the qubits are patterned. Each event type ultimately results in the production of athermal phonons whose total energy is approximately equal to the energy deposited by the initial interaction in the chip.

While the energies of the created athermal phonons

will vary depending on the event type, relaxation processes rapidly downconvert (over ns timescales) high-energy phonons into a larger number of lower-energy phonons that travel ballistically in the substrate [27]. In silicon, the rate of this anharmonic downconversion  $\Gamma_{\text{anh}}$  rises with phonon energy to the fifth power,  $\Gamma_{\text{anh}} \propto E_{\text{ph}}^5$ . As a result, the mean free path of this process is approximately 5  $\mu\text{m}$  for a 30 meV phonon and a much longer 1 cm for a 6 meV phonon. Similarly, the rate at which phonons scatter on isotopic impurities within the crystal lattice scales as  $E_{\text{ph}}^4$ , increasing steeply with phonon energy. When the mean free paths for both of these processes are longer than the device dimensions, the phonons will stream largely unimpeded along the material’s phonon caustics [28], and their momenta and energy spectra are influenced primarily by boundary scatterers. For our baseline detector which has a 380  $\mu\text{m}$  thickness (or for detectors with similar thickness), phonons are effectively ballistic at energies below the 6-meV scale. These low-energy phonons are responsible for conveying the deposited energy from the interaction site to the qubit sensors, and do so quasidiffusively over 10-100  $\mu\text{s}$  timescales for standard 1x1-cm<sup>2</sup>-area wafers [15].

Phonons may be lost either through absorption in the superconducting layer or by absorption in the copper thermalization mounts in the corners on the underside of the chip. In this work, we model an “all-or-none” phonon absorption process at a given surface with an effective “phonon absorption probability” ( $p_a$ ), such that a phonon incident upon an interface has a probability  $p_a$  of being fully absorbed and a probability  $1 - p_a$  of being fully reflected. We will use the notation  $p_{a,s}$  to represent this probability for interfaces between substrate and superconductor, and  $p_{a,c}$  to represent this probability for interfaces between substrate and (copper) corner thermalization mounts. This “all-or-none” model is a necessary oversimplification of the physics involved in phonon absorption, as rigorous first-principles modeling needs to capture complex dependencies on several parameters such as phonon energy and angle, interfacial acoustic mismatch, film crystallinity and thickness, and (in the case of superconductors) the gap parameter  $\Delta$ , not all of which are currently captured in G4CMP [18, 29]. Some of this complexity can be mapped into our simple modeling paradigm using known approximations to  $p_{a,s}$ :

$$p_{a,s} \simeq 1 - \exp \left[ - \frac{2l}{\pi v_s \tau_0^{ph}} \left( \frac{E_{ph}}{\Delta} \right) \right], \quad (1)$$

where  $E_{ph}$  is the phonon energy,  $l$  is the film thickness,  $\tau_0^{ph}$  is a material-dependent characteristic phonon scattering time [30],  $v_s$  is the sound speed in the material, and  $\Delta$  is the superconducting gap energy. This expression holds in the limit  $E_{ph} \gg \Delta$  [18, 24], which is a reasonable assumption for a phonon created in a down-conversion process in a silicon chip and impinging upon a superconducting aluminum film. For a film thickness of 100 nm, Equation 1 gives  $p_{a,s} \gtrsim 0.2$  for  $E_{ph} > 1$  meV.

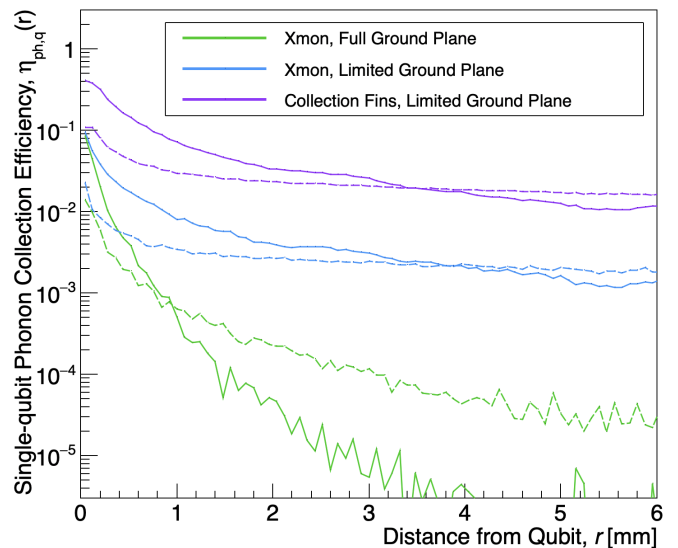


FIG. 3. Single-qubit phonon collection efficiency  $\eta_{\text{ph},q}$  versus radial distance from a given qubit island. Each curve is averaged over the six qubits simulated in our baseline geometry. This simulation is estimated for six different chip designs: full ground plane, limited ground plane, and limited ground plane with collection-fin qubit design. Solid (dashed) lines are for a  $p_{a,s}$  of 1.0 (0.1) at the Si-SC interface.

We will therefore focus our study on  $p_{a,s}$  values in the range from 0.1 to 1.0. A more rigorous handling of superconductor and normal-metal phonon modeling will be the focus of follow-up work.

With G4CMP, we first estimate the probability with which a ballistic phonon emerging from an interaction point will be absorbed into a qubit island. We define a qubit’s spatially-varying phonon collection efficiency  $\eta_{\text{ph},q}(r)$  as the average fraction of a total emitted phonon energy that is absorbed by a qubit a distance  $r$  from the emission point. We launch simulated 20-meV phonons uniformly throughout the chip, explicitly tabulate this fraction as a function of radial distance from each qubit, and average the fraction across the six qubits (Figure 3). We run these simulations for three different potential chip conditions that vary in qubit design and overall superconductor coverage: an Xmon-style island with a full Al ground plane, an Xmon-style island with a nearly absent (“Limited”) ground plane, and a collection-fin style island (Figure 2 lower inset) with a nearly absent ground plane.<sup>1</sup> In the three above-mentioned cases, the ratios of total qubit area to total superconductor area are 0.0011, 0.019, and 0.156, respectively. For each of these, we make

<sup>1</sup> While we present this third design to highlight its advantageous phonon collection, we have not yet performed studies to understand how feasibly this design may function as a transmon from an RF standpoint. For more discussion, see Ref. [20]

TABLE I. Estimates of  $\eta_{ph,sp}$  and  $E_{thr,chip}$  for a set of 12 chip designs. Relative statistical uncertainties on the simulated estimates of  $\eta_{ph,sp}$  are at or below the few-% level. The calculation of thresholds in the rightmost column relies on the discussion in Sections III and IV, but results are tabulated here for conciseness. For threshold estimates, the single-qubit readout parameters used are the near-term “baseline” ones discussed:  $T_{1,base} = 2$  ms,  $SSF = 0.98$  (see Section III text for definitions). Under these conditions,  $\sigma_{E,abs} \simeq 0.088$ eV. We note that for Designs 13 and 14, the calculation of threshold accounts for a few additional effects that modify  $\sigma_{E,abs}$  as discussed at the end of Section III C.

Design	Number of Qubits $N_q$	Qubit Design	Ground Plane	Si-SC phonon absorption prob. $p_{a,s}$	Si-Cu phonon absorption prob. $p_{a,c}$	Spatially-averaged phonon collection eff. $\eta_{ph,sp}$	Chip Threshold $E_{thr,chip}$
1	6	Xmon	Full	1.0	0.1	0.14%	737 eV
2	6	Xmon	Full	0.1	0.1	0.12%	860 eV
3	6	Xmon	Limited	1.0	0.1	2.07%	49 eV
4	6	Xmon	Limited	0.1	0.1	1.44%	71 eV
5	6	Xmon	Full	1.0	1.0	0.14%	737 eV
6	6	Xmon	Full	0.1	1.0	0.12%	860 eV
7	6	Xmon	Limited	1.0	1.0	1.76%	58 eV
8	6	Xmon	Limited	0.1	1.0	0.76%	135 eV
9	2	Xmon	Full	1.0	0.1	0.05%	1157 eV
10	10	Xmon	Full	1.0	0.1	0.24%	574 eV
11	2	Xmon	Limited	1.0	0.1	1.38%	41 eV
12	10	Xmon	Limited	1.0	0.1	2.39%	57 eV
13	6	Collection Fins	Limited	1.0	0.1	17.0%	O(0.1) eV
14	6	Collection Fins	Limited	0.1	0.1	12.6%	O(0.1) eV

estimates for  $p_{a,s} = 0.1$  and  $p_{a,s} = 1.0$  at the Si-SC (silicon-superconductor) interface, and assume  $p_{a,c} = 0.1$  at the Si-Cu interface where the chip’s four corners sit on the copper thermalization mount.

These six scenarios in Figure 3 already provide a wide range of possible phonon collection efficiencies. A chip with a full ground plane may display phonon collection efficiencies that are reduced by orders of magnitude relative to a chip with a spatially limited ground plane. A full ground plane significantly decreases the ratio of “sensing” superconductor (i.e. the qubit island) area to total superconductor area in which phonons may be absorbed, reducing the overall collection efficiency. An additional order of magnitude improvement in phonon collection can be achieved with the use of collecting fins for the same reason. For a given ground plane scenario, a lower probability of phonon absorption at the Si-SC interface will lead to a lower  $\eta_{ph,q}$  close to the qubit, but a larger  $\eta_{ph,q}$  far from the qubit. This is as a result of the increased ability of phonons to travel laterally in the plane of the chip by reflecting multiple times off of its top and bottom.

Another useful metric to quote is the total phonon collection efficiency spatially averaged over the entire chip,  $\eta_{ph,sp}$ , which we define as

$$\eta_{ph,sp} \equiv \left\langle \frac{\sum_i E_{dep,i}}{E_{dep,chip}} \right\rangle, \quad (2)$$

where  $E_{dep,i}$  is the energy absorbed by qubit  $i$ ,  $E_{dep,chip}$  is the total deposited energy, and the brackets denote spatial averaging of the location of the energy deposition  $E_{dep,chip}$  over the chip. This is a single metric that describes the overall conversion of in-chip energy to in-qubit energy. Table I presents estimates of this spatially averaged  $\eta_{ph,sp}$  for an expanded set of simulation conditions.

Beyond the chip conditions shown in Figure 3 (Designs 1-4,13,14), we explore wide variations in the absorption probability for phonons impinging upon the Si-Cu interface at the mount points in the four corners of the chip (Designs 5-8), as this probability is not easy to know *a priori*. We also study variations in the number of qubit-resonator-control-line assemblies included (Designs 9-12), to understand how the overall collection scales as phonon energy is necessarily split between multiple sensors.

An analytical approximation to  $\eta_{ph,sp}$  can be constructed as the ratio of sensitive to total absorptive area weighted by the phonon absorption probability at each interface [20]:

$$\eta_{ph,sp} \approx \frac{p_{a,s} A_s}{\sum_i p_{a,i} A_i}, \quad (3)$$

where  $A_s$  is the sum of areas of all sensitive elements (qubit islands) and  $A_i$  represents a more general interfacial element’s area, whether sensitive or not. We recast this general form slightly to acquire a form explicitly dependent on the number of qubits  $N_q$ :

$$\eta_{ph,sp} \approx \frac{N_q A_q p_{a,s}}{N_q A_t p_{a,s} + A_c p_{a,c} + f(N_q)}, \quad (4)$$

where the  $A_q$  is area of a qubit island,  $A_t$  is the total superconducting interfacial area added per qubit assembly (i.e. qubit, resonators, control lines, etc.),  $A_c$  is the total area of the copper corner thermal mounts, and  $f(N_q)$  is a function representing the average product of absorption probability and area for the remaining superconductor area. The form of  $f$  is dependent on ground plane scenario: for a negligible ground plane,  $f(N_q)$  is just the product of the transmission line area and  $p_{a,s}$ , and is

therefore independent of  $N_q$ . For a full ground plane,  $f(N_q) \simeq p_{a,s}(A_{chip} - N_q A_t)$ , and therefore falls with increasing  $N_q$ . We find that Equation 4 agrees with our simulation-determined  $\eta_{ph,sp}$  values to within approximately 15% across the design variations in Table I.

We can intuit a few key lessons from the combination of Equation 4 and Table I. First, the clearest advantage in overall phonon collection unsurprisingly comes from a limited ground plane, a large  $p_{a,s}$ , and from use of a qubit architecture with collection fins. In this scenario, the overall energy collection rises with the number of qubits but the energy collected per qubit in any given event falls, as the total phonon energy must be shared among a larger number of absorbers. Second, the effect of varying the probability of phonon absorption at the thermal mount interface is strongest in the limited ground plane case with a Si-SC absorption probability of 0.1: if phonon loss is weaker in the superconductor, thermal mount losses have a stronger effect on the overall phonon loss.

### III. Single Transmon Signal Response to Substrate Phonons

The second component responsible for determining an overall detector energy threshold is the signal response of an individual qubit island to incident phonon energy. This depends on the time evolution of the in-qubit QP density, the evolution of the resulting quantum state, and the scheme used to read out that quantum state. The estimate of threshold also depends on how interactions are reconstructed using this readout scheme. In this section we discuss in detail both how we model the system time evolution and how we reconstruct an energy deposition in a qubit. We then present estimates of our single-qubit threshold for a variety of qubit performance parameters.

As the parameter space affecting signal response is very broad, we include Table II as a guide to the single-qubit response parameters that we keep constant throughout this process for the purpose of reducing the dimensionality of our simulation parameter space. Notably, these parameters are for the Xmon-style qubit design, as the collection fin designs (Designs 13 and 14 in Table I) have yet to be experimentally demonstrated. Further discussion of these parameters can be found in the sections that follow.

#### A. Modeling Qubit Signal Time Evolution

As G4CMP does not yet handle the QP/superconductor response rigorously in arbitrary superconducting layers, a different tool is needed to properly capture this physics in this modeling step. To do this, we have developed a simulation tool called QuantumDeviceResponse (QDR). QDR takes as input the timestamps and magnitudes of energies absorbed in a qubit island simulated in G4CMP, and performs three

TABLE II. QDR parameters held constant within the simulation study.

Parameter	Description	Value	Refs.
Material	SC material	Aluminum	
$\delta t$	QDR time step	100 ns	
$r$	Recombination rate	0.005 ns <sup>-1</sup>	[16, 30]
$s_0$	Linear loss rate	10 <sup>-6</sup> ns <sup>-1</sup>	[16]
$g$	QP generation rate	10 <sup>-14</sup> ns <sup>-1</sup>	[16]
$n_{CP}$	Superfluid (CP) density	4 × 10 <sup>24</sup> m <sup>-3</sup>	[16, 18]
$\Delta$	SC gap energy	180 $\mu$ eV	[16, 18]
$V$	Island volume	1000 $\mu$ m <sup>3</sup>	
$\epsilon$	Phonon-to-QP efficiency	0.6	[31]
$F$	QP-creation Fano factor	0.2	[32]
$\omega_q/2\pi$	Qubit frequency	4 GHz	
$N$	Measurement binning	20	

operations on them:

1. It converts the deposited energy into a QP density, and evolves this QP density in discrete time steps  $\delta t$  according to established models of QP dynamics in superconductors.
2. Using a simplified representation of qubit state, it uses the QP density at a given time to stochastically determine qubit state evolution.
3. It simulates qubit control and measurement using simplified gate sequences to mimic a gate-based readout scheme.

In the following discussions we explore these operations in more detail.

#### 1. Quasiparticle Creation and Evolution

Once in the qubit island of volume  $V$ , phonons with energy larger than twice the superconducting gap,  $2\Delta$ , can break Cooper pairs (CPs), producing Bogoliubov QPs. QPs with energies higher than  $\Delta$  lose energy by radiating additional phonons in the superconductor, which if sufficiently energetic can create additional QP pairs. At the end of such a cascade, the final state of the detector involves a set of created QP pairs each with near- $\Delta$  energies and a set of low-energy phonons that may re-enter the substrate. The fraction  $\epsilon$  of an initial energy deposition retained in QPs is dependent on the initial phonon energy and film. For Al, where  $2\Delta \sim 360 \mu$ eV is much smaller than the few-meV energy of an average ballistic phonon,  $\epsilon \simeq 0.6$  [31]. In QDR, an energy deposition  $E_{\text{dep}}$  produces a number of QPs  $N_{qp}$  according to a gaussian distribution with a mean of  $E_{\text{dep}}\epsilon/\Delta$  and a spread  $\sqrt{FE_{\text{dep}}\epsilon}/\Delta$ , where a Fano factor  $F = 0.2$  has been used [32]. For ballistic phonons arriving later at the qubit over approximately 10-100  $\mu$ s, this overall  $N_{qp}$  will increase in discrete steps over this timescale.



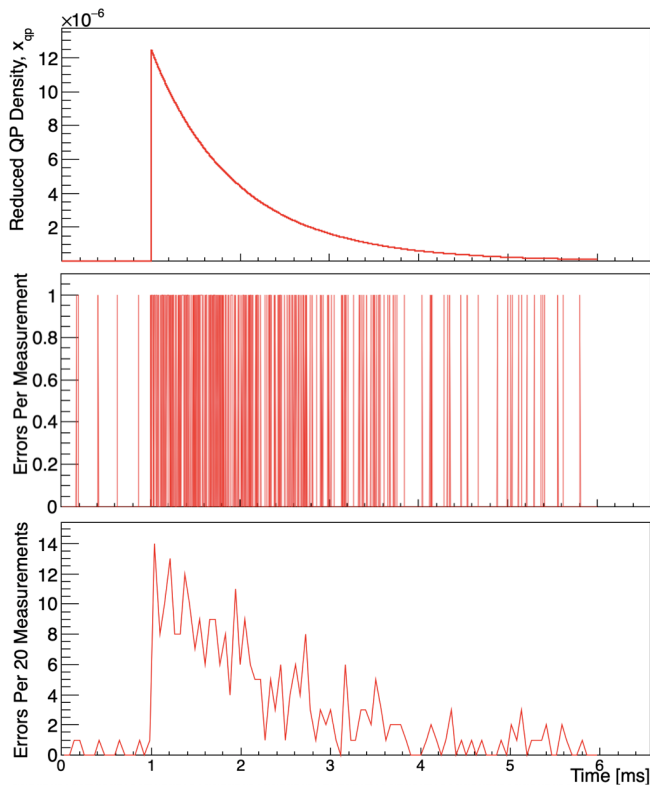


FIG. 4. A simulated event in which 15 eV is deposited directly into a single transmon island at  $t=1$  ms. **Top:** Reduced QP density as a function of time. **Center:** The time series of read-out qubit state for the “energy relaxation” read-out scheme in Figure 5. Here, measurements right after energy injection frequently return an “error”, and measurements many ms after the injection less frequently return an “error” due to the reduced QP density. For this readout scheme, we assume a qubit  $T_{1,\text{base}} = 2$  ms,  $SSF = 98\%$ , and search time  $\Delta t = 2000$  ns. **Bottom:** The errors in the central curve, binned in time such that each “sample” represents the number of errors in 20 measurements of the qubit.

After an energy deposition in the qubit, evolution of QP populations is governed primarily by diffusion, recombination, and trapping. Diffusion throughout islands of order  $200 \mu\text{m}$  happens within  $O(10) \mu\text{s}$  [33–35], allowing us to treat QPs effectively as if they exist with a uniform density  $n_{qp} \equiv N_{qp}/V$  in the island and leads after this time has elapsed. Knowing the superfluid (CP) density  $n_{CP}$  for a given material, it is then possible to model the normalized quasiparticle density  $x_{qp} \equiv n_{qp}/n_{CP}$  with

$$\frac{dx_{qp}}{dt} = -rx_{qp}^2 - s_0x_{qp} + g. \quad (5)$$

where the coefficients  $r$  and  $s_0$ , and the constant  $g$  represent rates of QP “self”-recombination, linear loss due to QP trapping or recombination with “quiescent” QPs, and QP generation, respectively [10, 16, 36]. In each time step  $\delta t$ , QDR computes the corresponding  $dx_{qp}$  using these parameters and uses it to update the total  $x_{qp}$  in a simulated qubit (Figure 4, top).

For our model, we motivate choices of  $r$ ,  $s_0$ , and  $g$  using values measured in literature. The recombination coefficient  $r$  is expected to be as large as its theoretically motivated value,  $r_{\text{Kaplan}} = 0.05 \text{ ns}^{-1}$ , but may be suppressed by around an order of magnitude or more depending on phonons that re-break CPs before leaving the film [16, 24, 30]. While this suppression factor is device-specific, we have also observed that the results that follow are only weakly dependent on  $r$  within the range of values typically measured. In this study we use a value of  $r = 0.005 \text{ ns}^{-1}$ , corresponding to a suppression factor of 10. The values of  $s_0$  and  $g$  are difficult to predict *a priori* and depend heavily on experimental parameters like material impurities, spatial variation in the superconducting gap  $\Delta$ , and IR shielding [37, 38]. In our modeling, these are set to the values in Table II.

## 2. Quantum State Evolution

QDR contains a description of a qubit state that also evolves in time. For this study, we use a very simple model in which the state is in either the first excited energy eigenstate  $|1\rangle$  or the energy ground state  $|0\rangle$ . The state can be changed using simulated gates in a gate sequence: a  $\pi$ -pulse performs a bit flip operation, taking  $|0\rangle \rightarrow |1\rangle$  and  $|1\rangle \rightarrow |0\rangle$ . The state can also be changed if energy relaxation occurs. QPs present in the qubit island may tunnel across the qubit’s Josephson junction and induce relaxation of a qubit from the excited state to ground state. The rate of relaxations due to such tunneling events in a qubit with frequency  $\omega_q$  is known [39] to be

$$\Gamma_{qp} = \sqrt{\frac{2\omega_q\Delta}{\pi^2\hbar}} x_{qp}, \quad (6)$$

which can be recast in terms of a characteristic relaxation coherence time  $T_{1,qp} = 1/\Gamma_{qp}$ . The relaxation rate contributes to an overall energy relaxation rate of a qubit  $\Gamma_{\text{tot}} = \Gamma_{\text{other}} + \Gamma_{qp}$ , where  $\Gamma_{\text{other}}$  represents the energy decay rate contributions from non-QP-related dissipative effects caused by other environmental noise [40].

QDR computes a total relaxation rate from both the instantaneous QP density at a given time and a parameterized “baseline” coherence time,  $T_{1,\text{base}}$ , due to non-QP effects. If the qubit is brought away from the  $|0\rangle$  state, QDR uses the instantaneous characteristic decay time  $T_1$  to randomly sample the next time at which state relaxation back to  $|0\rangle$  is induced. It updates this estimate on the fly as  $T_{1,qp}$  evolves, using techniques similar to those used in Ref. [41] for particle tracking through inhomogeneous materials.

## 3. Quantum State Readout

The final role of QDR is to simulate the readout value extracted from a qubit. We include a projective measure-

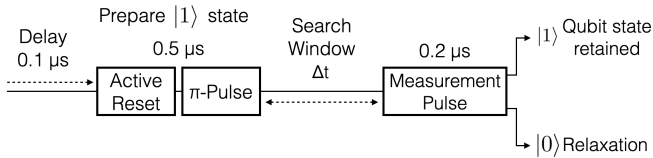


FIG. 5. The qubit gate sequence representing one measurement forming the foundation of the energy relaxation detection scheme.  $N$  such measurements are binned into one “waveform sample.”

ment as a “readout gate” that can map the qubit state  $|0\rangle$  ( $|1\rangle$ ) onto a boolean 0 (1) value. This gate suffers from readout noise, characterized by a probability  $p_1$  of incorrectly measuring a true  $|1\rangle$  state as  $|0\rangle$  and a probability  $p_2$  of incorrectly measuring a true  $|0\rangle$  state as  $|1\rangle$ . The single-shot-fidelity (SSF) of this readout gate is then defined [42] as

$$SSF = 1 - p_1 - p_2. \quad (7)$$

For this work, we make the simplifying assumption that the SSF is symmetric, i.e.  $p_1 = p_2$ .

While for now the quantum state evolution and readout simulated in QDR is limited to the operations and physics listed above, a thrust of future effort is its expansion to handle the modeling of the charge parity state of a qubit,  $\pi/2$ -pulses, and phase noise in addition to relaxation noise.<sup>2</sup>

### B. Reconstruction of an In-Qubit Energy Deposition Using an Energy Relaxation Readout Scheme

Determining the energy threshold achievable by a qubit sensor requires defining a readout scheme and signal reconstruction methodology. Here, we focus on a time-binned “energy relaxation” scheme (Figure 5) used in recent studies of superconducting qubit susceptibility to environmental radiation [18, 19], using QDR to directly simulate indefinitely-repeated calls to the gate sequence. This scheme is functionally equivalent to measuring a change in coherence time  $T_1$  as a function of the time since an energy deposition. We first build a signal template for waveforms read out with this scheme, and then discuss a strategy for energy reconstruction using that signal template.

<sup>2</sup> More generally, QDR is designed to facilitate construction of the readout response for general pairbreaking sensors and modeling of chips with multiple sensor types integrated. To that effect, it also includes a response for phonon-sensing microwave kinetic inductance detectors (MKIDs).

#### 1. Signal Template for Energy Relaxation Readout Scheme

We now construct a template  $S(t)$  for the signal read out using this relaxation scheme after an energy deposition in a qubit island. The output of one cycle of this gate sequence is a boolean value (Figure 4, center) which reflects whether a decoherence has happened within the search window  $\Delta t$  occurring after the  $\pi$ -pulse. These boolean values can be binned in time to produce a “waveform” of qubit errors per  $N$  measurements (Figure 4, bottom). Unless otherwise noted, we take  $\Delta t = 2 \mu\text{s}$  and  $N = 20$  so that we may still maintain sufficient time granularity to make use of the shape of a waveform in energy reconstruction.

First, we assume an initial instantaneous energy deposition  $E_{\text{dep}}$  causes an initial QP density  $x_i$  in an island volume  $V$ . For this scenario, the solution to the Equation 5, given in Ref. [16], is

$$x_{qp} = x_i \frac{1 - r'}{e^{t/\tau_{ss}} - r'} + x_0 \quad (8)$$

$$= \left[ \frac{E_{\text{dep}} \epsilon}{n_{CP} V \Delta} \right] \frac{1 - r'}{e^{t/\tau_{ss}} - r'} + x_0.$$

Here,  $\tau_{ss}$ ,  $r'$ , and  $x_0$  are functions of  $r$ ,  $s_0$ ,  $g$ , and  $x_i$ .<sup>3</sup> Notably,  $\tau_{ss}$  is the “steady-state” decay time for the QP relaxation, found when  $x_{qp}$  becomes small enough that self-recombination becomes subdominant.

The rate of qubit decoherence then follows from Equation 6 with an additional non-QP-induced term:

$$\Gamma(t) = \sqrt{\frac{2\omega_q \Delta}{\pi^2 \hbar}} \left( \left[ \frac{E_{\text{dep}} \epsilon}{n_{CP} V \Delta} \right] \frac{1 - r'}{e^{t/\tau_{ss}} - r'} + x_0 \right) + \Gamma_{\text{other}}. \quad (9)$$

Using our readout scheme, this time-varying relaxation rate can be effectively mapped onto a probability of measuring a readout value of 0 (an “error”) after the search window  $\Delta t$  has elapsed. For any one such iteration of the gate sequence, the probability  $p_r$  of a qubit relaxing in a time  $\Delta t$  between  $\pi$ -pulse and measurement is then given by:

$$p_r = 1 - e^{-\Gamma(t)\Delta t}$$

$$= 1 - \exp \left[ \frac{-\alpha E_{\text{dep}}}{E_{\text{dep}} [\tau_{ss} r (e^{t/\tau_{ss}} - 1)] + \beta e^{t/\tau_{ss}} - \gamma} \right], \quad (10)$$

where for consolidation we also define the variables  $\alpha = \Delta t \sqrt{2\omega_q \Delta / \pi^2 \hbar}$ ,  $\beta = n_{CP} V \Delta / \epsilon$ , and  $\gamma = -\alpha x_0 -$

<sup>3</sup> Though not explored in this work, we note that phonon “recycling” from  $\simeq 2\Delta$  recombination phonons may distort this functional form:  $2\Delta$  phonons can rebreak CPs, which may later recombine to produce more  $2\Delta$  phonons. This cycle may continue as long as the  $2\Delta$  phonons do not exit through thermal contacts. As a result, for chips with minimal contact with their thermal mounts, this may significantly lengthen the time dependence of  $x_{qp}$  and the overall response time of the qubits [18].

$\Gamma_{\text{other}}\Delta t$ . Here,  $\gamma$  represents a ‘‘baseline’’ contribution to relaxation composed of both baseline QP contributions ( $\alpha x_0$ ) and non-QP contributions ( $\Gamma_{\text{other}}\Delta t$ ).

The probability  $p_{\text{obs}}$  of acquiring 0 (an error) in a single measurement depends on  $p_r$  and any imperfect single-shot fidelity in the qubit and readout chain. Using the nomenclature of Section III A 3 and defining  $p_0$  as the probability of correctly measuring a true  $|0\rangle$  state as  $|0\rangle$ , we find

$$\begin{aligned} p_{\text{obs}} &= p_r p_0 + (1 - p_r) p_1 \\ &= p_r (p_0 - p_1) + p_1 \\ &= p_r (1 - p_2 - p_1) + p_1 \\ &= p_r * SSF + p_1 \end{aligned} \quad (11)$$

Summing  $N$  sequential measurements in a time ‘‘bin’’ gives our signal template  $S(t)$ . The value of such a waveform sample follows a binomial distribution described by  $N$  Bernoulli trials and a success probability  $p_{\text{obs}}$ . The expected signal template in this readout scheme is then simply

$$\begin{aligned} S(t) &= N p_{\text{obs}} \\ &= N (p_r SSF + p_1) \\ &= N \left( SSF \left( p_r - \frac{1}{2} \right) + \frac{1}{2} \right), \end{aligned} \quad (12)$$

where in the last expression we have taken advantage of our assumption of a symmetric SSF. This expression is valid for values of  $N$  for which the relaxation observation probability  $p_{\text{obs}}$  stays relatively constant over one  $N$ -measurement sample. In the following development, we limit ourselves to using  $N$  and other parameters such that this  $p_{\text{obs}}$  is constant to well within 10% in such a sample.

## 2. Energy Reconstruction Methodology

Using the signal template in Equation 12, we may estimate an in-qubit energy deposit from a binned time series of qubit measurements. The following reconstruction procedure is based loosely on the optimal filter strategy employed by Ref. [10, 43], though due to the non-gaussianity of the signal under consideration here, we must use explicit maximum likelihood fits in our reconstruction methodology to extract physics information from a waveform. We will first briefly describe this general fitting procedure as it pertains to our binomially-distributed signal template  $S(t)$ .

For any single waveform sample built from  $N$  measurements each with probability  $p_{\text{obs}}$  of yielding an error, the probability of finding  $s$  errors is given by a binomial distribution:

$$p(s) = \binom{N}{s} p_{\text{obs}}^s (1 - p_{\text{obs}})^{(N-s)} \quad (13)$$

Since the expected time-dependent waveform composed of many  $N$ -measurement samples is given by the signal

template  $S(t) = N p_{\text{obs}}(t)$  in Equation 12, the joint probability of acquiring a set of *measured* waveform samples  $\{s_i\}$  at times  $\{t_i\}$  given this expected signal template can be written as:

$$p(\{s_i\}|S(t)) = \prod_i \binom{N}{s_i} p_{\text{obs}}^{s_i}(t_i) (1 - p_{\text{obs}}(t_i))^{(N-s_i)} \quad (14)$$

where the range of times  $\{t_i\}$  should be between the injection time  $t_0$  and a few characteristic decay times after  $t_0$ . The probability in Equation 14 is also equivalent to the likelihood  $\mathcal{L}$  of the signal template given the measured waveform  $\{s_i\}$ . Maximizing the likelihood as a function of a signal template parameter of interest (with all others constrained) gives a best-fit value of that parameter of interest. We will use this maximum-likelihood (ML) fitting process twice in our energy reconstruction methodology, as discussed below.

First, a detector calibration must be performed to acquire the pulse-shape-determining quantities  $\tau_{ss}$ ,  $r$ , and  $\gamma$ . The quantity  $\gamma$  can be extracted from basic qubit  $T_1$  and SSF characterization tests, and so does not require us to use the ML formalism. However, the quantities  $\tau_{ss}$  and  $r$  are less trivial to estimate. These can be determined by injecting a known energy  $E_{\text{dep}}$  into a qubit island, and performing a 2D ML fit over those two variables for a stacked waveform. For a discussion of experimental tools that may enable such a calibration, interested readers are encouraged to read Ref. [44, 45].

With  $\tau_{ss}$ ,  $r$ , and  $\gamma$  determined using a set of calibration pulses, one can then assess the injected energy  $E_{\text{dep}}$  of any other pulses (say, in a physics search period) as long as those parameters don’t change significantly in time. This is done by performing a 1D maximum likelihood fit over  $E_{\text{dep}}$  to determine reconstructed energy of the pulse. We call the resulting best-fit  $E_{\text{dep}}$  the ‘‘reconstructed’’ energy  $E_r$ . The energy determined in this way is manifestly non-negative, in contrast to energies that may be reconstructed according to the optimal filter prescription in Ref [43].

## C. Estimating Energy Resolution and Threshold

Energy resolution for this detection scheme is given by the width of a distribution of reconstructed event energies for a monoenergetic simulated in-qubit energy  $E_{\text{dep}}$ . In the following discussion, we simulate and reconstruct energies for populations of events with an instantaneous energy injection into a single qubit at an arbitrary time  $t = 1$  ms followed by time evolution of the system until  $t = 6$  ms. This time range is used for the maximum likelihood fits. The resulting energy resolution is dependent on a number of parameters:

- $E_{\text{dep}}$  : For a ‘‘baseline’’ near-term qubit performance characterized by  $T_{1,\text{base}} = 2$  ms and  $SSF = 0.98$ , and a  $\Delta t = 2$   $\mu\text{s}$ , Figure 6 demonstrates the spread in reconstructed energy  $E_r$  as a function of



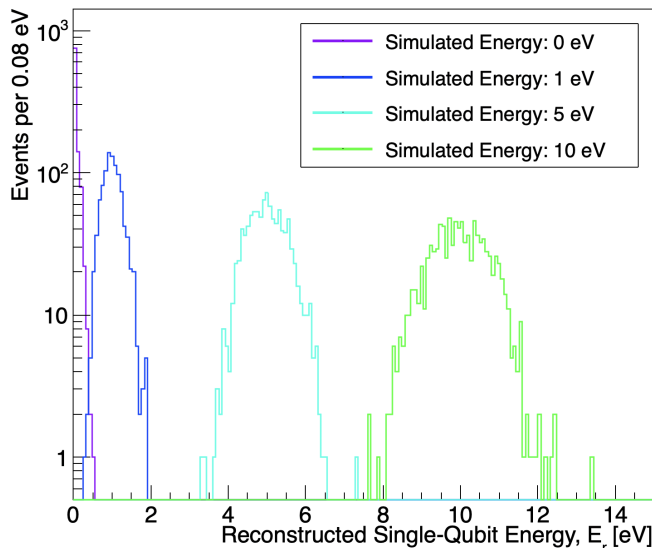


FIG. 6. Distributions of reconstructed energy  $E_r$  for simulated input energies  $E_{\text{dep}}$  of 0 eV, 1 eV, 5 eV, and 10 eV injected into a single transmon island. Each distribution is formed from 1000 events run through QDR with baseline simulation parameters:  $T_{1,\text{base}} = 2$  ms,  $\text{SSF} = 0.98$ .

injected  $E_{\text{dep}}$ . With increasing  $E_{\text{dep}}$ , the  $E_r$  distribution broadens considerably even from 1 eV to 10 eV. We attribute the broadening at these energies to the uncertainty (variance) associated with a binomial distribution defined by  $N$  measurements and success probability  $p_{\text{obs}}$ :  $Np_{\text{obs}}(1 - p_{\text{obs}})$ . For each waveform sample, this variance is largest for  $p_{\text{obs}} = 0.5$ , leading to an increase in the overall uncertainty in  $E_r$  as we progress from near-threshold waveforms to those created by 5 eV or 10 eV energy depositions. However, even for  $E_{\text{dep}}$  large enough that  $p_{\text{obs}}$  is initially above 0.5 (which occurs  $\gtrsim 20$  eV), distributions do not re-narrow according to the binomial distribution variance. At these higher energies, a saturation in waveform shape sets in that makes energy reconstruction highly sensitive to small variations in waveform. Though rigorous exploration of this effect depends on details of phonon arrival times and is beyond the scope of this first model, we generally expect energy resolution to worsen monotonically with increased energy. Overall, we understand the nonlinear response of this readout scheme to be primarily attributed to the binomial nature of the signal read out, and will likely be present in other readout schemes where the signal is composed of discrete probabilistic measurements of a quantum state.

A notable quantity of interest is the “noise-only” energy resolution of a single qubit sensor,  $\sigma_{E,\text{abs}}$ , corresponding to the spread in the 0 eV curve in Figure 6. This is the resolution due to the jitter in a reconstructed energy that is fundamentally at-

tributable to physical noise sources (from imperfect SSF and finite  $T_{1,\text{base}}$ ), and is a metric one can use to compare the low-threshold performance of this detection scheme with low-threshold performance of other pairbreaking sensor architectures. Since energies are strictly non-negative, this curve has an appreciable sub-population of reconstructed events at exactly zero reconstructed energy. This distribution is therefore approximated well by a rectified Gaussian distribution derived from a pure Gaussian described by a mean of 0 and a  $\sigma = \sigma_{E,\text{abs}}$ . We may numerically estimate  $\sigma_{E,\text{abs}}$  from this population by fitting the  $E_r > 0$  region to a gaussian peaked at 0, and find for our baseline simulation that  $\sigma_{E,\text{abs}} = 0.088$  eV. Consistent with the convention used in Refs. [7, 10], we define the sensor energy threshold  $E_{\text{thr}} \equiv 5\sigma_{E,\text{abs}}$ , which for our baseline simulation gives  $E_{\text{thr}} = 0.44$  eV.

- $T_1$  and SSF: Both finite  $T_{1,\text{base}}$  values and imperfect SSF will ultimately contribute noise to this readout scheme. The effect of these parameters on  $\sigma_{E,\text{abs}}$  is explored in Figure 7. Longer  $T_{1,\text{base}}$  values and higher SSF values are better, as they limit false positives from non-QP decoherence and readout error, respectively. Lower SSF values will also lead to a higher leveling-off of  $\sigma_{E,\text{abs}}$  with increasing  $T_{1,\text{base}}$ : for a fixed SSF, even an infinitely long coherence time will not continue to improve the overall signal-to-noise ratio. We also note that the assumptions underlying our QP time evolution will break down in the limit of small numbers of QPs produced by an event – in such few-QP scenarios we cannot treat  $x_{qp}$  in the junction leads as being approximately equal to  $x_{qp}$  in the qubit island. For aluminum, this may cause deviation from the curves in Figure 7 for in-qubit energies below approximately 100 meV in our Xmon design, though we note that this figure will vary depending on relative volumes of the qubit island and junction leads.
- Search window  $\Delta t$ : Though a rigorous development of the dependence of  $\sigma_{E,\text{abs}}$  on  $\Delta t$  is beyond the scope of this work, we present two intuitive limits that demonstrate that an operational “sweet spot” exists in a selection of  $\Delta t$ . First, in the limit  $\Delta t \gg T_{1,\text{base}}$  a relaxation will always be expected, and no useful information about time-dependent QP-induced relaxation can be extracted. In the limit  $\Delta t \rightarrow 0$ , relaxation *within* the measurement pulse dominates over relaxation in the search window, which may prevent faithful measurement of a well-defined qubit energy eigenstate. Though this in-measurement decay is still possible with longer search windows, it is no longer the dominant (non-SSF-induced) effect in the measured value when  $\Delta t$  is much longer than the measurement pulse length,  $\Delta t_m$ . Moreover, shorter search windows increase the overall rate of measurement cycles, which in-

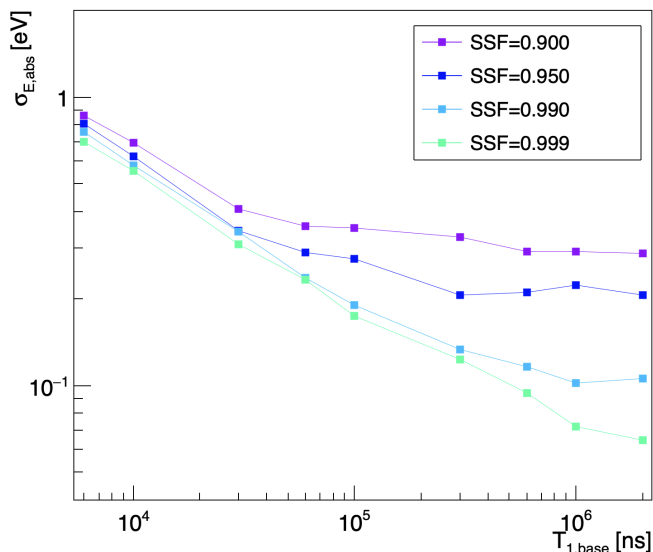


FIG. 7. Noise-only single-qubit sensor energy resolution  $\sigma_{E,\text{abs}}$  for a search window of  $\Delta t = 2 \mu\text{s}$  and varying  $T_{1,\text{base}}$  and SSF. Error bars indicating statistical uncertainty are included, and for most points are on the order of the size of the point itself.

creases the rate at which imperfect SSF may cause misreconstruction and reduce SNR. As a result, there exists some optimal  $\Delta t_m \lesssim \Delta t \lesssim T_{1,\text{base}}$  that can be selected to achieve a minimized  $\sigma_{E,\text{abs}}$  (maximized SNR).

These estimates of  $\sigma_{E,\text{abs}}$  are expected to be slightly better (lower) than practically achieved values, because our current simulations only model smearing from a limited number of effects. We do not account for temporal variations in our experimental quantities like  $T_{1,\text{base}}$  or SSF, which will add additional spread to reconstructed energies beyond what’s shown here. There will also be spread induced from variation in other parameters in Equation 10. For example, if there is experimental noise in the qubit frequency (either due to flux noise in the case of split transmons or charge noise in the case of single-junction transmons), then this will add additional variation to the reconstructed energy around the baselines we have shown. These estimates are also likely to worsen slightly in future estimates as we add into QDR a residual occupation of the qubit excited state and the possibility for nonequilibrium QPs to cause upward energy transitions in the qubit, both of which are effects that have been demonstrated experimentally [46, 47]. Finally, we have made these estimates under the condition that the time of the initial in-qubit energy injection (and therefore the start time for the ML fits) is known. For a more realistic “physics search” scenario where the times of particle impacts are not known, energy resolution of reconstructed pulses may be larger than what we have quoted.

TABLE III. A summary of the single-qubit sensor energy resolution for our baseline Xmon design (top) and our baseline design with collection fins and QP trap assumed.

Simulation Quantity	Value
Input Parameter: $T_{1,\text{base}}$	2 ms
Input Parameter: SSF	0.98
Input Parameter: Search window $\Delta t$	2 $\mu\text{s}$
Performance (Xmon): $\sigma_{E,\text{abs}}$	0.088 eV
Performance (Xmon): $E_{\text{thr}}$	0.44 eV
Performance (collection-fins): $\sigma_{E,\text{abs}}$	O(0.001) eV
Performance (collection-fins): $E_{\text{thr}}$	O(0.005) eV

Given these results, we comment on the single-qubit energy resolution and threshold realistically achievable for near-term detectors based on this readout scheme and demonstrated qubit architectures. Maximum published coherence times for 2D transmons are at the 300-400  $\mu\text{s}$  scale [40]. As a result, while an experimental achievement of our baseline simulated coherence time of  $T_{1,\text{base}} = 2 \text{ ms}$  in a transmon qubit has yet to be published, we expect that such coherence may plausibly be achieved in the near-term. SSF values of 0.995 have been experimentally demonstrated even without the use of a quantum-limited amplifier [48], suggesting that our baseline SSF=0.98 is realistic. Together, these observations suggest that our quoted single-qubit sensor energy resolution  $\sigma_{E,\text{abs}} = 0.088 \text{ eV}$  and threshold  $E_{\text{thr}} = 0.44 \text{ eV}$  are realistic targets in the near term with this sensing scheme. These figures are summarized in Table III. Moreover, Figure 7 shows that for an SSF near 0.98 these estimates may only change by a factor of order unity even if a lower, currently-achieved coherence time is assumed.

Though it is unclear whether the collection fin designs are compatible with the energy relaxation readout scheme we have discussed so far, we will nonetheless entertain a coarse estimate of their potential sensor thresholds if they can be operated as such. Their increased collection area implies an increase in (degradation of)  $\sigma_{E,\text{abs}}$  through the increased volume  $V$ . To counteract this, it is possible to design lower- $\Delta$  QP trapping structures near the junction, as discussed in Ref [20]. This not only reduces the effective volume of the QPs near the junction to the much smaller trapping region volume, but also results in a QP gain process as near-gap QPs in the fins cascade to lower energy in the trap. From the formalism in Ref. [20] and similar design specifications, approximately 4% of the QPs from a energy deposition in a collection fin can fall into the trap, be multiplied with a gain of  $\approx 7$ , and be localized within a volume O(0.02%) that of a full collection fin. When propagated through Equation 6, these effects combine to give a factor of O(100) increase in the decoherence rate for a given  $E_{\text{dep}}$  relative to our baseline Xmon design and readout parameters. As a result, it may be possible to achieve sensor resolutions at the O(1) meV scale and sensor thresholds at the O(5) meV scale with such traps, though such estimates are likely idealized given the fact that these energy scales

represent single- or few-phonon impacts. Moreover, these estimates rely on maintaining the same baseline readout performance ( $T_{1,\text{base}}$ , etc.), which may be significantly more challenging given the propensity of ambient QPs to diffuse into the trap region near the junction.

#### IV. Transmon Array Response to Energy Depositions

Combining the phonon and single-qubit responses, one may now arrive at the response of the entire chip to energy depositions, and estimate an overall in-chip energy threshold. We reconstruct the full energy in the chip as

$$E_{r,\text{chip}} = \frac{\sum_i E_{r,i}}{\eta_{ph,sp}}. \quad (15)$$

Here, we use a sum over single-qubit reconstructed energies  $E_{r,i}$ , rather than a fit to a summed waveform, to get the total in-chip reconstructed energy. As different qubits may have different values of  $r$  and  $\tau_{ss}$  depending on their fabrication, it is important to do individual calibrations and fits to each qubit waveform – a single fit to a summed waveform may wash out the differences in these parameters and degrade energy reconstruction. While calculating  $E_{r,\text{chip}}$ , we continue to reconstruct  $E_{r,i}$  assuming an energy injection into a qubit at exactly the time of energy injection into the substrate. In reality, phonons will take 10-100  $\mu\text{s}$  to propagate around the chip before striking the sensor. This will introduce a slight misreconstruction of our  $E_{r,i}$ , but that misreconstruction should be small as long as the characteristic quasiparticle fall time  $\tau_{ss}$  is much longer than that phonon diffusion timescale.

For a collection of events occurring spatially uniformly within the chip, Equation 15 implies that the mean  $E_{r,\text{chip}}$  should match the true mean deposited in-chip energy. However, it will also introduce a spread in  $E_{r,\text{chip}}$  that is dependent on the spatial nonuniformity of the phonon collection efficiency:  $E_{r,\text{chip}}$  will be overestimated for events close to the qubits and underestimated for events far from the qubits. This can be mitigated somewhat if one uses qubit spatial information with the single-qubit signals to reconstruct a coarse position of the interaction within the qubit. Such a position reconstruction capability, while beyond the scope of this work, could enable a position-based correction of the reconstructed energy, reducing the spread in energies from a monoenergetic source even if the phonon absorption efficiency isn't perfectly uniform. Moreover, this position reconstruction capability might also enable sophisticated coincidence-based and position-based tagging of pathological interactions from nearby high-energy radioactivity and cosmic ray muons, which is invaluable in a low-threshold search.

With an in-chip reconstructed energy defined, we may proceed to study in-chip energy resolution and threshold. Figure 8 shows several distributions of reconstructed energies for events simulated uniformly in XY in our baseline 6-qubit chip (Design 3 from Table I). For each event,

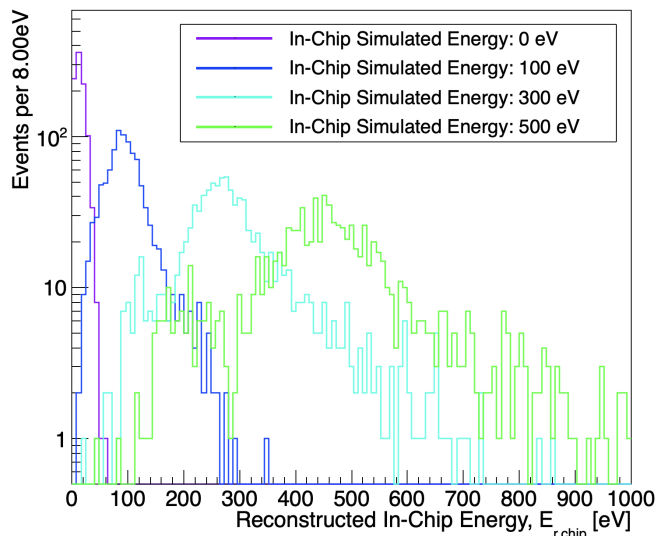


FIG. 8. Distributions of  $E_{r,\text{chip}}$  for various injected energies, for chip Design 3 and our baseline single-qubit parameters. Here, the 0 eV, “noise-only” distribution is peaked at nonzero reconstructed energy. This is due to the explicitly non-negative reconstructed single-qubit energies  $E_{r,i}$  involved in the sum in Equation 15 (consistent with a distribution composed of a sum of several rectified Gaussian variables). The non-Gaussian shape of the 100, 300, and 500 eV cases is due to the nonuniformity in phonon collection efficiency across the chip.

a true total phonon energy was spawned from a random position in the chip as a sum of 20 meV phonons. This allows us to remain agnostic to the initial type of energy deposit but ignores a small degree of smearing in the initial phonon locations (like what would be expected from freed charge traveling a distance through the chip prior to phonon creation). As with the single-qubit case, the spread in each distribution grows with injected energy.

To understand in-chip “noise-only” resolution and threshold, we make use of the 0 eV in-chip-energy curve in Figure 8. The peak of this distribution is offset from  $E_{r,\text{chip}} = 0$ , as expected for the sum of 6 random variables each described by identical rectified Gaussian distributions [49]. As a result, a noise-only energy resolution and threshold cannot be quoted in the exact same way as for the single-qubit case or for other pairbreaking detectors which display Gaussian noise profiles. However, a useful threshold for comparison to other detectors can still be quoted as the energy at which the noise-only distribution reaches the same statistical significance level as a  $5\sigma$  deviation from the mean of a Gaussian. For a scenario in which each qubit’s response is described by the same underlying rectified Gaussian distribution, this threshold is dependent on the number of qubits  $N_q$  and can be reasonably modeled as

$$E_{\text{thr},\text{chip}} \simeq \frac{\sigma_{E,\text{abs}}}{\eta_{ph,sp}} \left[ \frac{N_q}{\sqrt{2\pi}} + aN_q^b \right]. \quad (16)$$

where  $a \simeq 4.27$  and  $b \simeq 0.44$ . The first term in brackets represents the mean of the noise-only distribution, and can be derived using the moments of the rectified Gaussian distribution [49]. The second term captures the distance from that mean to the threshold. Its approximate power-law form is extracted computationally from Monte-Carlo-generated distributions corresponding to sums of  $N_q$  rectified Gaussian distributions, a process which also gives the above-quoted values for  $a$  and  $b$ .

Table I gives this  $E_{\text{thr,chip}}$  value for the various chip parameters tested in Section II. We observe that the most optimal overall in-chip thresholds for our baseline Xmon hardware design are achieved using a limited ground plane, for which thresholds are as low as 41 eV. As this is still well above 1 eV, it is clear that achieving the goal of a sub-eV threshold will require additional iteration on qubit/resonator/control line design beyond the variations we have performed on the “baseline” Xmon architecture in Designs 1-12. Designs 13 and 14 illustrate a potential such iteration: if transmons with collection fins as in Ref. [20] can be achieved and read out in an energy relaxation readout scheme, the benefit of the larger collecting area and QP trap may enable sub-eV in-chip thresholds near the 100 meV scale.

Of some additional utility in thinking forward to larger-scale detector design is the fact that the full-chip “simple” threshold may also be approximated analytically. Using equation 16 and substituting in Equation 4, we arrive at

$$E_{\text{thr,chip}} = \sigma_{E,\text{abs}} \left[ \frac{N_q}{\sqrt{2\pi}} + aN_q^b \right] \times \left[ \frac{N_q A_t p_{a,s} + A_c p_{a,c} + f(N_q)}{N_q A_q p_{a,s}} \right].$$

Figure 9 shows results of this calculation for two sets of scenarios: one in which a common  $p_{a,s}$  is parameterized in a limited ground plane scenario, and one in which a ground plane with a parameterized fill factor is implemented.<sup>4</sup> For low absorption probabilities and/or high ground plane fill factors, the gain in phonon collection efficiency from adding qubits largely outweighs the additional noise for small numbers of qubits  $N_q$ . However, at higher absorption probability (equivalently, lower fill factor), the benefit of gained phonon collection efficiency  $\eta_{ph,sp}$  balances the additional noise, showing an optimum  $N_q$  for minimizing chip threshold.<sup>5</sup>

<sup>4</sup> This fill factor creates a scenario somewhere between our initially-stated “Full” and “Limited” ground plane scenarios.

<sup>5</sup> This optimum assumes no degradation of the individual sensor threshold upon adding more qubits to the chip. As qubit  $T_{1,base}$  values tend to drop with increasing  $N_q$  in a way that is commonly attributed to on-chip high-frequency environmental noise, this estimate may therefore overestimate the optimum number of qubits [40].

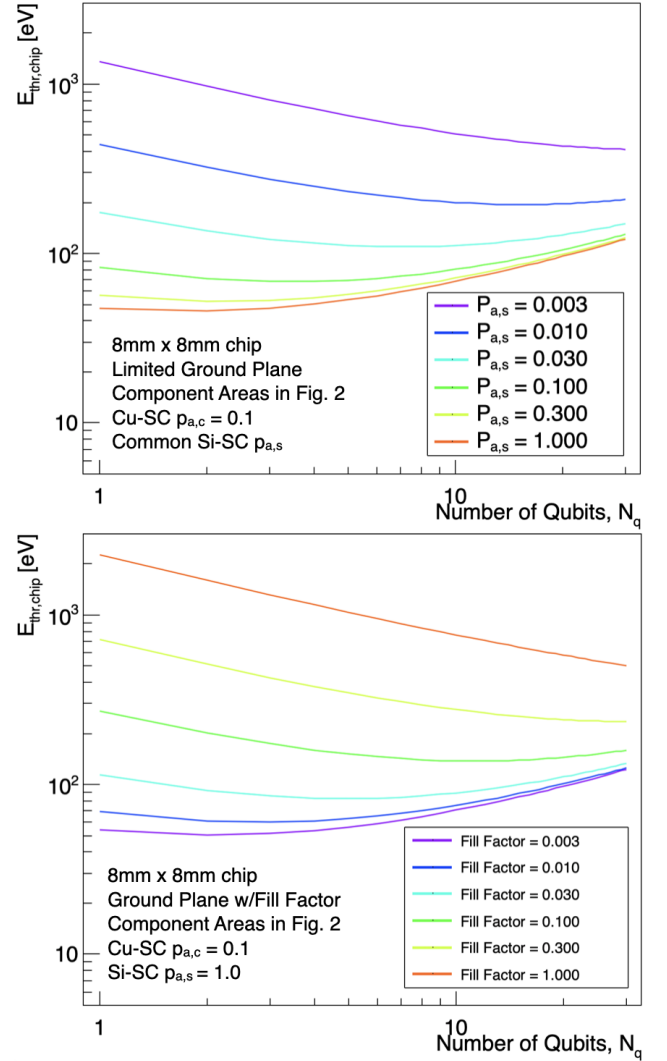


FIG. 9. Chip threshold vs.  $N_q$ , with chip parameters described shown in the bottom left of each plot. In both scenarios we use the Xmon-style island design and assume  $\sigma_{E,\text{abs}} = 0.088\text{eV}$ . Top: variations in a common Si-SC  $p_{a,s}$  in a limited ground plane scenario. Bottom: variations in ground plane fill factor for  $p_{a,s} = 1.0$ .

Though it is often challenging to know the absorption parameters in Equation 4 *a priori*, this model for threshold may provide some intuitive guidance for how to select parameters of a future array of qubits for dark matter detection. In a scenario where a chip mass is fixed, we propose the following general strategy for creating a chip layout that has been optimized for achieving low in-chip energy thresholds. First, one chooses a minimum inter-qubit distance based on RF performance grounds – qubits that are too closely packed can become capacitively coupled and may swap electromagnetic quanta, degrading isolated qubit performance [50]. Next, one can use estimates of the areas of various phonon-absorbing regions on the chip ( $A_q$ ,  $A_t$ , etc. from Equation 4) with esti-



mates of the interfacial absorption probabilities  $p_{a,s}$  and  $p_{a,c}$  to determine an optimal number of qubits per chip. Lastly, those qubits are tiled on the chip at no closer than the minimum inter-qubit distance, and as spatially uniformly as possible so that  $\eta_{ph}$  is as uniform as possible throughout the chip. If more target mass is desired for exposure, additional chips may be added and read out in parallel.

It is worth noting that a more realistic optimization may be driven by a budget-based constraint on the number of readout channels, with the less-expensive chip mass being allowed to vary to increase DM search exposure. While we do not attempt to solve this more general optimization problem here, the several rules of thumb we have discussed in the “budget-unconstrained” scenario (finding an optimal  $N_q$ , uniform qubit spacing, etc.) will still help constrain this problem.

## V. Data-Model Comparison

We now demonstrate a coarse viability of our models and reconstruction by exercising them on a set of published experimental results that use the same energy relaxation readout scheme that we have studied in this work. In Ref. [19], the effects of environmental radiation on a 10-transmon qubit chip (similar to Design 10 in Table I) were studied using waveforms constructed from a binned energy relaxation readout scheme like that described in Section III B. In this section, we demonstrate that for a reasonable set of assumed detector parameters, we can reconstruct a reasonable in-chip energy deposition for one of the events claimed to be from a cosmic ray muon passing through the chip.

The cosmic ray event shown in Figure 2 of Ref. [19] includes a waveform in each of 10 transmon qubits. Here, the relaxation readout scheme uses a  $\Delta t = 3 \mu\text{s}$ , an overall time-per-cycle of  $15.3 \mu\text{s}$ , and a binning  $N$  of 40. Our goal is to fit these waveforms and extract estimated energies from them. To perform this fit, we need four additional “calibration” parameters: the SSF and  $r$ ,  $\tau_{ss}$ , and  $\gamma$  from Equation 10. The most trivial of these to estimate is  $\tau_{ss}$  – this was already measured in Ref. [19] to be either 6 ms or 0.7 ms, depending on the junction of a given qubit. For this study, we only attempt to study those waveforms with  $\tau_{ss} = 6$  ms, which corresponds to the JJ design that is more susceptible to QP tunneling out of the qubit island.

To estimate SSF, we can make our standard assumption of a symmetric fidelity and establish a lower bound on the value of the SSF using pre-pulse baseline errors. For Ref. [19], we use this pre-pulse period to estimate a minimum SSF at the 80% scale, but since a non-infinite  $T_{1,\text{base}}$  also realistically contributes to this error baseline, we assume for our calculations a symmetric SSF of 95%, commonly achieved in the field. Using that SSF, taking  $E_{\text{dep}} = 0$ , and measuring an average pre-pulse baseline error probability for a pulse, we can fully invert Equat-

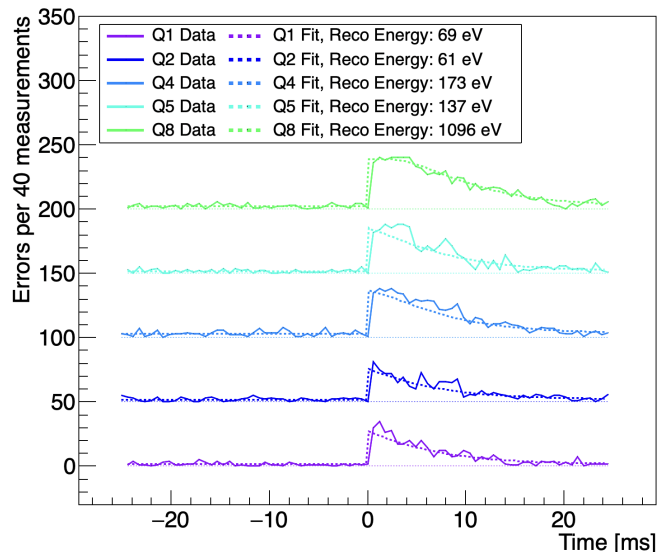


FIG. 10. Event waveforms digitized from Figure 2a of Ref. [19], with fits employing the energy reconstruction technique developed in this work. Waveforms are offset vertically at intervals of 50 for clarity, with the “zero-point” of each plot shown as a dotted line. Fit functions are shown as dashed lines. Reconstructed best-fit energies are included in the legend.

tion 12 for  $\gamma$ .

The recombination constant  $r$  is the hardest quantity to estimate with the data provided. This difficulty is in large part due to uncertainty of the “suppression factor”  $F$  that should be divided into  $r$  to account for recombination phonons that re-break Cooper pairs before leaving the superconducting film. From Refs [16, 29], this suppression factor should be in the range of 5-10 for interfaces of Sapphire with 80-nm-thick Al film.<sup>6</sup> In the case of Ref. [19] where a Si chip with a 250 nm Al film is used, the change in substrate-film acoustic mismatch approximately cancels the thicker film, and as a result we assume in this calculation  $F \simeq 10$ .

Figure 10 shows reproduced waveforms for Qubits 1, 2, 4, 5, and 8 in the 10-transmon chip, along with maximum-likelihood fits to the overall in-qubit energies. Waveforms were extracted by digitizing the plots in Ref. [19]. The estimated summed energy within all 5 qubits is 1.537 keV. If we compute a per-qubit average from these, each qubit has absorbed on average 307 eV in this event.

Given the qubit dimensions and overall chip dimen-

<sup>6</sup> It is interesting to note that the measurements of this suppression factor mentioned in Ref. [29] seem to be somewhat inconsistent with the theoretical estimates of the phonon absorption length projected from Ref. [18]. While in this case we assume the measured values for our approximation, it is worth mentioning that there is a substantial acknowledged degree of uncertainty in this parameter.

sions quoted in Ref. [19], we can continue and estimate an in-chip overall energy. Using the transmon and chip dimensions shown in their work and our own assumption of  $p_{a,s} = 1.0$  for the superconductor (which is not unreasonable due to their thick film), we estimate an  $\eta_{ph,sp} \sim 0.7\%$ . Together with our average in-qubit energy, we then estimate a total in-chip energy of about 440 keV. For a minimum-ionizing particle traversing a Si chip of thickness 0.35 mm, this estimated energy deposition implies an incident angle of about  $14^\circ$  away from the plane of the chip [51]. Since the zenith is approximately in the plane of the chip according to Ref. [19], this suggests that the muon also arrived at approximately  $14^\circ$  from the zenith. Given the displayed chip orientation and positioning relative to the other scintillation detectors, an incident particle track angled at  $14^\circ$  with respect to the zenith is reasonably consistent with the observation of a qubit-scintillator coincidence in this event and the conclusion that the event was caused by a cosmic muon.

Though we have deliberately averaged out position information for simplicity, this exercise also demonstrates that reconstruction of position information in an interaction is realistic using our energy reconstruction formalism. In Figure 10, Qubit 8 clearly has a more saturated waveform than Qubit 1, indicating that the bulk of the phonon energy emitted occurred closer to Qubit 8 (i.e. on the right side of the chip as presented in Ref [19]). This suggests that even at lower energies than this muon event, position information may indeed be applied to remove pathological background events and perform corrections on reconstructed energy to improve in-chip energy resolution.

Sources of uncertainty in this exercise include slight errors in plot digitization, the aforementioned worsening energy resolution for in-qubit energy depositions above  $\gtrsim 20$  eV, our assumptions about phonon absorption probability  $p_{a,s}$  at the Si-SC interface, our averaging of the qubits' energy depositions (ignoring spatial variation), our reconstruction assuming negligible phonon propagation time from the energy deposition to the qubit islands, and our previously-addressed assumptions about the recombination constant  $r$ . However, even though these many uncertainties are challenging to characterize given the limited information available, we are able to construct a picture of this event as having an in-chip deposited energy consistent with a cosmic muon passing through the chip, in agreement with the conclusions of Ref. [19] claiming that the event's origin is a cosmic muon. This ability is a promising first step for better understanding the tools and formalism for using this superconducting qubit technology for detection purposes.

## VI. Discussion

Given the goal of using superconducting qubits as particle detectors, we must discuss how the results of the

last three sections fit into the larger landscape of low-mass dark matter searches. For a near-term ‘‘baseline’’ chip design with Xmon qubits, our estimated chip energy threshold is 41 eV for demonstrated qubit architectures (though for yet-to-be-demonstrated ground plane designs), which is only a factor of a few higher than state-of-the-art TES-based detectors [8] but competitive with state-of-the-art MKID detectors [10]. To compete with these technologies, additional iteration on our baseline design is needed. There are three plausible avenues for improving detector performance: readout scheme, materials choices, and layout optimization.

First, the readout scheme we use plays a role in limiting our projected threshold. Here, the link between  $x_{qp}$  and  $\Gamma_{qp}$  in Equation 6 ultimately determines our quoted sensor energy resolution  $\sigma_{E,abs}$ . However, different readout schemes, such as those enabling mapping of the qubit island charge parity to the readout value, may offer increased sensitivity for nearly identical hardware. If such readout schemes are also composed of gate operations, they will also inherit much of the mathematical formalism discussed and many of the trends we have observed for the energy relaxation scheme presented here. Studying alternative readout schemes will be a focus of follow-up studies.

Second, materials choices may also have a major impact on the overall in-chip threshold through both  $\sigma_{E,abs}$  and  $\eta_{ph,sp}$ . While our studies assume purely aluminum qubits, use of a lower-gap material may yield at least a  $1/\sqrt{\Delta}$  improvement of  $\Gamma_{qp}$  in Equation 6 [52]. Lower-gap materials such as Hf, AlMn, or IrPt are still an area of exploration within the qubit community, but they are regularly exploited by the TES and MKID communities for lowering threshold [53–55]. The advantage of exploiting such lower gap materials is also evident in the coarsely estimated thresholds for the collection-fins Designs 13 and 14 if a QP trap is assumed to be present as in Ref. [20]. Moreover, materials design may also improve our phonon collection efficiency  $\eta_{ph,sp}$ : if significant areas of on-chip superconducting structures are strictly needed, say, for RF performance (i.e. ground plane, transmission line, etc.), it is advantageous to make these structures from higher-gap material such as Nb ( $2\Delta \simeq 3$  meV) [56]. Here, the few-meV phonons from the in-substrate downconversion cascade are much closer to the superconducting gap, and by Equation 1 are therefore much more likely to stream back out of the film without breaking CPs. This enables better phonon collection in the target (Al) structures sensitive to QPs.

Finally, iteration on chip layout to reduce the phonon-insensitive on-chip superconductor is also a strategy for improving  $\eta_{ph,sp}$  and lowering in-chip threshold. While we nominally reach up to 17% with Designs 13 and 14, there is a clear benefit to further studying methods by which to reduce the area of the readout resonators, transmission line, and qubit control lines. Some potential strategies for increasing the overall sensitive fraction include minimizing the total number of transmission lines



through frequency multiplexing, removing the readout resonator [20], and/or coupling multiple qubits to each resonator [57]. While thoroughly exploring the details of such additional designs is beyond the scope of this paper, it will likely be required to compete with leading TES-based technologies.

It is also instructive to consider impacts of this study on design of qubit chips for the quantum computing community. Correlated errors are undesirable as they impair quantum error correction (QEC) with the surface code, and as a result one generally wants to invert many of the decisions we have discussed for how to design a low-threshold detector. In particular, in-qubit phonon collection efficiencies should be minimized through use of a full ground plane and, if possible, high- $\Delta$  qubit superconductor material and normal metal baths for phonon down-conversion [17]. Junction gap engineering strategies can further suppress errors by restricting the diffusion of QPs out of the qubit islands, as demonstrated in Refs. [19, 58]. Beyond this, a detailed optimization of the chip layout for minimizing correlated errors depends on the specific algorithm and physical layout of the chip, and is a topic for future study with G4CMP.

## VII. Conclusions

In this work, we have performed an end-to-end modeling of the physics relevant to understanding the energy threshold of a detector based on superconducting qubits operated in a relaxation-detection readout scheme. We have explored the properties of phonon propagation from an interaction site to a qubit sensor using G4CMP for a variety of different chip design parameter values, and have estimated the corresponding phonon collection efficiencies  $\eta_{ph,sp}$ . We've also explored the single-qubit response to phonon energy depositions using a custom-built simulation software called Quantum Device Response (QDR), and found that for an example energy relaxation readout scheme using Xmon qubits, we estimate a qubit sensor energy threshold of 0.44 eV for near-term qubit readout parameters. We then studied the combined effect of phonon transport and single-qubit readout to estimate thresholds of in-chip energy deposited, in which we found a 41 eV in-chip threshold if one can significantly limit the ground plane present in the chip. We also found that if a qubit design with collection fins and QP traps near the junction can still be operated with this readout scheme, the in-chip threshold may plausibly be lowered to the O(0.1 eV) scale. Finally, we have also demonstrated the viability of the formalism that we've used to develop these estimates via comparison to a real cosmic

ray muon event in Ref. [19], though further validation is needed.

While this estimated in-chip 41 eV threshold is still higher than other leading sensor technologies, it sets a benchmark from which qubit-based detector technology can be further improved. Moreover, the work laid out here may act as a guide for the development of other low-threshold qubit-based detectors: much of the formalism in Section III may also be applied to more sensitive gate-based readout schemes, and the strategies in Section IV may enable coarse optimization of chip parameters given any new qubit/resonator/control line design and some knowledge of phonon absorption probabilities. Together, these strategies lay the groundwork for designing and understanding the sensitivity of superconducting-qubit-based detectors with gate-based readout schemes in the field of sub-GeV/ $c^2$  dark matter direct detection.

## Acknowledgments

The authors would like to thank Karthik Ramanathan, Sunil Golwala, Osmond Wen, Noah Kurinsky, and Mike Kelsey for useful discussions of phonon and quasiparticle microphysics in superconducting devices. We also thank Robert McDermott, Sohair Abdullah, and Sami Lewis for helpful intuition-building lessons about design and operation of transmon qubits. Finally, we thank Edgard Bonilla for insightful discussions about statistical distributions and tools for constructing and analyzing the energy relaxation signals. This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. This work was supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Science Center and the U.S. Department of Energy, Office of Science, High-Energy Physics Program Office.

RL led the organization and development of the simulations and analysis chain used in this work, and drafted the manuscript. RL, DBa, LH, EFF, and RK motivated the overall study and provided guidance in the scoping and presentation of this work. IH performed G4CMP phonon transport simulations useful in refining insights made in Section II. DT, SD, and KA contributed to development of the QDR package. DT also provided insight into community standards for energy resolution and calibration that motivated additional shaping of paper narrative. DBa, LH, EFF, RK, IH, KA, DBo, GB, GC, AC, RG, KS, and SS provided feedback and thoughtful discussions throughout the development of this work.

---

[1] A. Arbey and F. Mahmoudi, Dark matter and the early universe: A review, *Progress in Particle and Nuclear*

*Physics* **119**, 103865 (2021).

[2] C. Bartram *et al.* (ADMX Collaboration), Search for in-

- visible axion dark matter in the 3.3 – 4.2  $\mu\text{eV}$  mass range, *Phys. Rev. Lett.* **127**, 261803 (2021).
- [3] J. Aalbers *et al.* (LUX-ZEPLIN Collaboration), First dark matter search results from the lux-zeplin (lz) experiment, *Phys. Rev. Lett.* **131**, 041002 (2023).
- [4] E. Aprile *et al.* (XENON Collaboration 7), Dark matter search results from a one ton-year exposure of xenon1t, *Phys. Rev. Lett.* **121**, 111302 (2018).
- [5] R. Essig *et al.*, Snowmass2021 Cosmic Frontier: The landscape of low-threshold dark matter direct detection in the next decade, arXiv e-prints , arXiv:2203.08297 (2022), arXiv:2203.08297 [hep-ph].
- [6] C. W. Fink *et al.*, Characterizing tes power noise for future single optical-phonon and infrared-photon detectors, *AIP Advances* **10**, 10.1063/5.0011130 (2020).
- [7] S. L. Watkins, *Athermal Phonon Sensors in Searches for Light Dark Matter*, Ph.D. thesis (2022).
- [8] 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**, 032010 (2021).
- [9] C. W. Fink *et al.*, Performance of a large area photon detector for rare event search applications, *Applied Physics Letters* **118**, 022601 (2021), <https://pubs.aip.org/aip/apl/article-pdf/doi/10.1063/5.0032372/13706989/022601.1.online.pdf>.
- [10] D. J. Temples *et al.*, Performance of a kinetic inductance phonon-mediated detector at the nexus cryogenic facility (2024), arXiv:2402.04473 [physics.ins-det].
- [11] A. V. Dixit *et al.*, Searching for dark matter with a superconducting qubit, *Phys. Rev. Lett.* **126**, 141302 (2021).
- [12] M. D. Shaw *et al.*, Quantum capacitance detector: A pair-breaking radiation detector based on the single cooper-pair box, *Phys. Rev. B* **79**, 144511 (2009).
- [13] P. Echternach *et al.*, Single photon detection of 1.5 thz radiation with the quantum capacitance detector, *Nat Astron* **2**, 90 (2018).
- [14] S. Danilin *et al.*, Quantum-enhanced magnetometry by phase estimation algorithms with a single artificial atom, *npj Quantum Inf* **4**, <https://doi.org/10.1038/s41534-018-0078-y> (2018).
- [15] C. Wilen *et al.*, Correlated charge noise and relaxation errors in superconducting qubits., *Nature* **594**, 369 (2021).
- [16] C. Wang *et al.*, Measurement and control of quasiparticle dynamics in a superconducting qubit., *Nat Commun* **5**, <https://doi.org/10.1038/ncomms6836> (2014).
- [17] V. Iaiia *et al.*, Phonon downconversion to suppress correlated errors in superconducting qubits., *Nat Commun.* **13**, <https://doi.org/10.1038/s41467-022-33997-0> (2022).
- [18] M. McEwen *et al.*, Resolving catastrophic error bursts from cosmic rays in large arrays of superconducting qubits., *Nat. Phys.* **18**, 107 (2022).
- [19] P. M. Harrington *et al.*, Synchronous detection of cosmic rays and correlated errors in superconducting qubit arrays (2024), arXiv:2402.03208 [quant-ph].
- [20] C. W. Fink, C. P. Salemi, B. A. Young, D. I. Schuster, and N. A. Kurinsky, The Superconducting Quasiparticle-Amplifying Transmon: A Qubit-Based Sensor for meV Scale Phonons and Single THz Photons, (2023), arXiv:2310.01345 [physics.ins-det].
- [21] K. Ramanathan, Quantum parity detectors for dark matter (2023), cPAD 2023.
- [22] M. Kelsey *et al.*, G4cmp: Condensed matter physics simulation using the geant4 toolkit, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **1055**, 168473 (2023).
- [23] R. Barends *et al.*, Coherent josephson qubit suitable for scalable quantum integrated circuits, *Phys. Rev. Lett.* **111**, 080502 (2013).
- [24] E. Yelton *et al.*, Modeling phonon-mediated quasiparticle poisoning in superconducting qubit arrays (2024), arXiv:2402.15471 [quant-ph].
- [25] S. M. Griffin *et al.*, Multichannel direct detection of light dark matter: Target comparison, *Phys. Rev. D* **101**, 055004 (2020).
- [26] R. Anthony-Petersen *et al.*, A Stress Induced Source of Phonon Bursts and Quasiparticle Poisoning, arXiv e-prints , arXiv:2208.02790 (2022), arXiv:2208.02790 [physics.ins-det].
- [27] T. F. Harrelson, I. Hajar, O. A. Ashour, and S. M. Griffin, Theoretical investigation of decoherence channels in athermal phonon sensors (2023), arXiv:2109.10988 [cond-mat.mtrl-sci].
- [28] A. G. Every, Formation of phonon-focusing caustics in crystals, *Phys. Rev. B* **34**, 2852 (1986).
- [29] S. B. Kaplan, Acoustic matching of superconducting films to substrates, *J Low Temp Phys* **37**, 343 (1979).
- [30] S. B. Kaplan *et al.*, Quasiparticle and phonon lifetimes in superconductors, *Phys. Rev. B* **14**, 4854 (1976).
- [31] J. Martinis, Saving superconducting quantum processors from decay and correlated errors generated by gamma and cosmic rays., *npj Quantum Inf* **7**, <https://doi.org/10.1038/s41534-021-00431-0> (2021).
- [32] A. G. Kozorezov *et al.*, Fano fluctuations in superconducting-nanowire single-photon detectors, *Phys. Rev. B* **96**, 054507 (2017).
- [33] J. N. Ullom *et al.*, Energy-dependent quasiparticle group velocity in a superconductor, *Phys. Rev. B* **58**, 8225 (1998).
- [34] Y. Dong *et al.*, Measurement of quasiparticle diffusion in a superconducting transmon qubit, *Applied Sciences* **12**, 10.3390/app12178461 (2022).
- [35] S. Y. Hsieh and J. L. Levine, Diffusion of quasiparticles in superconducting aluminum films, *Phys. Rev. Lett.* **20**, 1502 (1968).
- [36] P. J. de Visser *et al.*, Number fluctuations of sparse quasiparticles in a superconductor, *Phys. Rev. Lett.* **106**, 167004 (2011).
- [37] S. E. de Graaf *et al.*, Two-level systems in superconducting quantum devices due to trapped quasiparticles, *Science Advances* **6**, eabc5055 (2020), <https://www.science.org/doi/pdf/10.1126/sciadv.abc5055>.
- [38] S. Diamond *et al.*, Distinguishing parity-switching mechanisms in a superconducting qubit, *PRX Quantum* **3**, 040304 (2022).
- [39] G. Catelani *et al.*, Relaxation and frequency shifts induced by quasiparticles in superconducting qubits, *Phys. Rev. B* **84**, 064517 (2011).
- [40] C. Wang *et al.*, Towards practical quantum computers: transmon qubit with a lifetime approaching 0.5 milliseconds., *npj Quantum Inf* **8**, <https://doi.org/10.1038/s41534-021-00510-2> (2022).
- [41] W. Nelson *et al.*, The egs4 code system (1985).
- [42] T. Walter *et al.*, Rapid high-fidelity single-shot dispersive readout of superconducting qubits, *Phys. Rev. Appl.* **7**, 054020 (2017).
- [43] S. R. Golwala, *Exclusion Limits on the WIMP-Nucleon*

- Elastic Scattering Cross Section from the Cryogenic Dark Matter Search*, Ph.D. thesis, UC Berkeley (1993).
- [44] R. A. Moffatt *et al.*, Spatial imaging of charge transport in silicon at low temperature, *Applied Physics Letters* **114**, 032104 (2019), [https://pubs.aip.org/aip/apl/article-pdf/doi/10.1063/1.5049691/14521319/032104\\_1.online.pdf](https://pubs.aip.org/aip/apl/article-pdf/doi/10.1063/1.5049691/14521319/032104_1.online.pdf).
- [45] K. Stifter *et al.*, Paper in preparation, (2024).
- [46] A. Kulikov, R. Navarathna, and A. Fedorov, Measuring effective temperatures of qubits using correlations, *Phys. Rev. Lett.* **124**, 240501 (2020).
- [47] J. Wenner *et al.*, Excitation of superconducting qubits from hot nonequilibrium quasiparticles, *Phys. Rev. Lett.* **110**, 150502 (2013).
- [48] L. Chen *et al.*, Transmon qubit readout fidelity at the threshold for quantum error correction without a quantum-limited amplifier., *npj Quantum Inf* **9**, <https://doi.org/10.1038/s41534-023-00689-6> (2023).
- [49] M. Beauchamp, On numerical computation for the distribution of the convolution of n independent rectified gaussian variables, *Journal de la Société de statistique de Paris* **159**, 88 (2018).
- [50] Y. He *et al.*, Control system of superconducting quantum computers, *J Supercond Nov Magn* **35**, 11 (2022).
- [51] R. L. Workman and Others (Particle Data Group), Review of Particle Physics, *PTEP* **2022**, 083C01 (2022).
- [52] S. Dutta *et al.*, Superfluid density in conventional superconductors: from clean to strongly disordered, *Journal of Physics: Condensed Matter* **34**, 335601 (2022).
- [53] J. Zhang *et al.*, Characterization of the thermal properties of ir/pt bilayer transition edge sensors, *J Low Temp Phys* **210**, 194–208 (2023).
- [54] N. Zobrist *et al.*, Design and performance of hafnium optical and near-ir kinetic inductance detectors, *Applied Physics Letters* **115**, 213503 (2019), [https://pubs.aip.org/aip/apl/article-pdf/doi/10.1063/1.5127768/13286696/213503\\_1.online.pdf](https://pubs.aip.org/aip/apl/article-pdf/doi/10.1063/1.5127768/13286696/213503_1.online.pdf).
- [55] E. Vavagiakis *et al.*, Developing almn films for argonne tes fabrication, *J Low Temp Phys* **199**, 408 (2020).
- [56] V. Novotny *et al.*, Single superconducting energy gap in pure niobium, *J Low Temp Phys* **18**, 147 (1975).
- [57] J. M. Chow *et al.*, Detecting highly entangled states with a joint qubit readout, *Phys. Rev. A* **81**, 062325 (2010).
- [58] M. McEwen and others., Resisting high-energy impact events through gap engineering in superconducting qubit arrays (2024), arXiv:2402.15644 [quant-ph].