

Distinguishing Parity-Switching Mechanisms in a Superconducting Qubit

S. Diamond¹,* V. Fatemi¹,† M. Hays¹,‡ H. Nho¹, P. D. Kurilovich, T. Connolly, V. R. Joshi, K. Serniak¹,§ L. Frunzio¹, L. I. Glazman, and M. H. Devoret¹¶

Departments of Physics and Applied Physics, Yale University, New Haven, Connecticut 06520, USA



(Received 25 April 2022; accepted 1 September 2022; published 10 October 2022)

Single-charge tunneling is a decoherence mechanism affecting superconducting qubits, yet the origin of excess quasiparticle excitations (QPs) responsible for this tunneling in superconducting devices is not fully understood. We measure the flux dependence of charge-parity (or, simply, “parity”) switching in an offset-charge-sensitive transmon qubit to identify the contributions of photon-assisted parity switching and QP generation to the overall parity-switching rate. The parity-switching rate exhibits a qubit-state-dependent peak in the flux dependence, indicating a cold distribution of excess QPs that are predominantly trapped in the low-gap film of the device. Moreover, we find that the photon-assisted process contributes significantly to both parity switching and the generation of excess QPs by fitting to a model that self-consistently incorporates photon-assisted parity switching as well as interfilm QP dynamics.

DOI: [10.1103/PRXQuantum.3.040304](https://doi.org/10.1103/PRXQuantum.3.040304)

I. INTRODUCTION

A growing sector of electromagnetic radiation sensing and quantum information science relies on superconducting circuits due to their dissipationless nature. However, nonequilibrium quasiparticle excitations (QPs) notoriously present in the superconductors can cause dissipation and hinder the performance of superconducting devices. Nonequilibrium QPs can limit the sensitivity of kinetic inductance detectors [1], “poison” charge-sensitive devices such as single-Cooper-pair transistors [2–4], and cause decoherence in superconducting qubits [5–20]. Hybrid superconductor-semiconductor architectures are likewise susceptible to QP poisoning, which would limit Andreev qubits [21–23] and proposed Majorana-based qubits [24–26].

QP-induced decoherence in superconducting qubits is typically ascribed to tunneling of excess nonequilibrium

QPs across a Josephson junction (JJ), as shown schematically by the purple arrows in Fig. 1(a). In this mechanism, when a QP tunnels across the JJ, it couples to the phase across the junction and can thereby cause transitions of the qubit state. The rate of QP tunneling occurs in proportion to the QP density in the superconductors, by which previous experiments have inferred QP densities (normalized to the Cooper-pair density) in the range $x_{\text{QP}} = n_{\text{QP}}/n_{\text{CP}} \sim 10^{-9} - 10^{-5}$ [12,15,18–20,27–33]. Depending on levels of other sources of decoherence, QP densities in this range may limit qubit performance.

These QP densities are many orders of magnitude higher than expected for devices at thermal equilibrium with the approximately 30-mK base temperature of a dilution refrigerator, in which QP excitations are exponentially suppressed by the superconducting gap Δ_{Al} . This contrast points to the nonequilibrium nature of the QPs in the superconductors. Furthermore, attempting to explain measurements of QP-induced excitation and relaxation of the qubit by QP tunneling requires an assumption of their presence at high energies [19]. This is inconsistent with predictions that QPs relax to a distribution near the edge of the superconducting gap [11,34,35]. Despite the many observations of excess QPs across various types of qubits, a two-pronged question remains unanswered: how are nonequilibrium QPs generated and why do they appear to have a nonthermal distribution?

A possible mechanism to help answer these questions has been proposed by Houzet *et al.* [36]. There, it has been pointed out that a photon with sufficient energy to break a Cooper pair ($\geq 2\Delta_{\text{Al}} \sim 100$ GHz) may be efficiently absorbed at the JJ, generating a pair of QPs. Just as QP tunneling results in a single charge being transferred across

*spencer.diamond@yale.edu

†Current address: Department of Applied Physics, Cornell University, Ithaca, New York 14853, USA

‡Current address: Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

§Current address: MIT Lincoln Laboratory, Lexington, Massachusetts 02421, USA

¶michel.devoret@yale.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

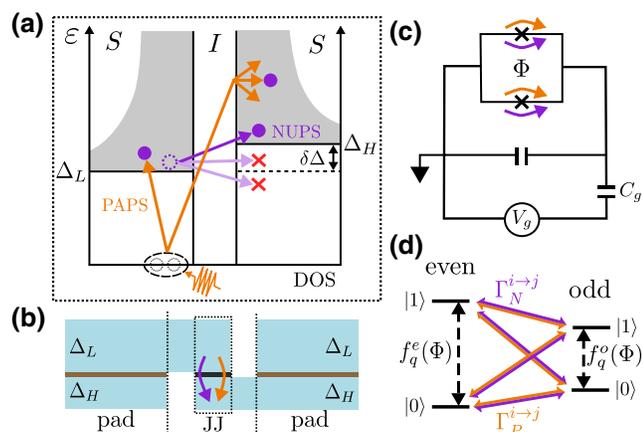


FIG. 1. (a) The superconducting density of states of the JJ in the excitation picture, with a difference in the superconducting gaps $\delta\Delta$ of the aluminum films. Two parity-switching mechanisms are illustrated. The first conserves the QP number while the second generates two QPs. Number-conserving parity switching (NUPS, purple): a preexisting QP may tunnel while exciting, relaxing, or without exchanging energy with the qubit. Some cases of QP-qubit interactions may be suppressed by lack of available final states (red cross). Photon-assisted parity switching (PAPS, orange): a photon with energy greater than $\Delta_L + \Delta_H$ may be absorbed at the JJ, breaking a Cooper pair to generate two QPs, with one tunneling across the JJ. (b) A cross-section sketch of the device, formed by one thinner higher-gap film and one thicker lower-gap film. The arrows represent NUPS (purple) and PAPS (orange) tunneling across the JJ. Like the JJ, the pads of the device consist of a bilayer of both films but have an approximately 10^6 times larger contact area. This large contact area ensures that QPs in the pads may rapidly redistribute between the two films as discussed in Sec. III. (c) The circuit diagram of a flux-tunable offset-charge-sensitive transmon. Fluctuations in the charge environment may induce offset charge $n_g = C_g V_g / 2e$. Both NUPS and PAPS across either JJ result in a switch of the parity, which appears as a sudden jump in n_g by $1/2$. (d) The level diagram showing the qubit ground ($|0\rangle$) and excited ($|1\rangle$) states separated into even- and odd-parity manifolds. Both NUPS and PAPS cause a switch between even and odd parity and may relax or excite the qubit. For $E_J/E_C \lesssim 30$, $|f_q^e - f_q^o| (n_g = 0) \gtrsim 500$ kHz.

the JJ, this process likewise switches the parity of the number of electrons that have tunneled across the JJ (referred to simply as the “parity” for the remainder of the work) and may cause decoherence of the qubit state [Fig. 1(a), orange arrows]. We label this process photon-assisted parity switching (PAPS) and it may contribute to QP-induced decoherence in two ways simultaneously: directly, by inducing qubit transitions during parity switches, and indirectly, as a generation mechanism of excess x_{QP} . The conventional tunneling of excess QPs we refer to as the number-conserving parity switching (NUPS) mechanism because, unlike PAPS, it conserves the number of QPs

[13,14]. We distinguish PAPS from QP generation by ionizing radiation [33,37,38], which likewise generates QPs but is assumed to do so primarily in the large pads of the device as opposed to at the JJ and thus does not directly cause a parity switch.

Additionally, PAPS causes qubit transitions imitating a nonthermal QP distribution when the typical photon energy is well above $2\Delta_{\text{Al}}$ [36]. For this reason, even in cases where the energy-relaxation time of the qubit is limited by other decoherence sources, PAPS may be the primary cause of anomalous excitation of the qubit. The amount of high-frequency radiation that reaches the qubit can be reduced with targeted filtering and shielding, which has been shown to improve qubit performance [29,32,39]. Even with these measures in place, recent experiments have shown that stray high-frequency photons are indeed absorbed resonantly by spurious antenna modes, inducing parity switching [40–42]. Given that high-frequency photon absorption at the junction can be responsible for QP generation and the appearance of a nonthermal QP distribution, it is imperative to experimentally distinguish the contributions of PAPS to parity switching and QP generation in superconducting qubits. However, it is difficult to discern the mechanism from a single measurement of the parity-switching rate or the QP-limited energy-relaxation time.

Here, we measure the parity-switching rate Γ in a flux (Φ)-tunable transmon sensitive to offset charge. As we describe below, the dependence of the parity-switching rate on the applied flux can distinguish PAPS from NUPS. In the flux dependence of the parity-switching rate $\Gamma(\Phi)$, we observe a peak that can be explained by a difference of superconducting gaps between the two aluminum films of the device matching the qubit transition energy. This gap difference enhances the contrast in flux dependence between PAPS and NUPS and also helps to demonstrate the thermalization of QPs in the device. Using a new measurement protocol to extract the parity-switching rates conditioned on the initial state of the qubit, we find evidence that excess QPs relax to a low energy distribution and are primarily trapped in the low-gap film. We develop a model that quantitatively fits the measured $\Gamma(\Phi)$ with a self-consistent combination of PAPS and NUPS and derive two new insights. First, PAPS is responsible for a significant fraction of parity switching. Second, PAPS generates excess QPs at a rate on par with the sum of all other mechanisms, which we observe by measuring $\Gamma(\Phi)$ in the presence of a controllable photon source operated at several powers. Due to these effects, estimates of x_{QP} obtained from measurements of Γ or QP-limited qubit relaxation that do not take into account PAPS or gap difference may be inaccurate. These results advance our understanding of QP dynamics in superconducting qubits and will inform approaches to mitigation of single-charge-tunneling decoherence.

II. EXPERIMENTAL DEVICE

A. Flux-tunable offset-charge-sensitive transmon

In this section, we describe the experimental device in which we distinguish the parity-switching mechanisms. The parity-switching rates are measured directly with an offset-charge-sensitive transmon [15,19,32]. A parity switching event, in which a single charge tunnels across the JJ, changes the parity $p \in \{e, o\}$ of the electron number in each of its two electrodes. Such an event appears as a sudden jump by $1/2$ in the reduced offset charge $n_g = C_g V_g / 2e$ (which is measured in units of $2e$). The plasmon eigenstates $i \in \{0, 1\}$ of the qubit can be separated into even- and odd-parity manifolds [Fig. 1(d)]. In the offset-charge-sensitive transmon regime ($E_J/E_C \lesssim 30$), the effective jump in n_g that occurs when NUPS or PAPS induce a switch in the parity results in an observable jump in the qubit transition frequency between two values $f_q \pm \delta f_q$, where $\delta f_q(n_g = 0) \gtrsim 500$ kHz. The parity dependence of the qubit frequency has been used previously to measure the parity-switching rates $\Gamma^{i \rightarrow j}$ correlated with qubit transitions from i to j and these rates have been inconsistent with parity switching solely due to a thermal distribution of the resident QPs [19].

Because the parity-switching rate depends on the density of states available for tunneling [Fig. 1(a)], changes to the qubit frequency can affect PAPS and NUPS differently. Thus, we replace the JJ with a dc superconducting quantum interference device (SQUID) in order to tune *in situ* the average frequency of the qubit $f_q(\Phi)$ with the applied flux [Fig. 1(c)]. For both NUPS and PAPS, the flux-dependent rates of parity switching causing a qubit transition from i to j ($\Gamma^{i \rightarrow j}(\Phi)$) depend on the single-charge-tunneling qubit matrix elements as well as factors accounting for the occupation and availability of QP states in the JJ films [17,35,36][Appendix A]. While applying flux changes the single-charge-tunneling matrix elements identically for NUPS and PAPS, the resulting change to the qubit frequency $f_q(\Phi)$ affects PAPS and NUPS differently due to the unique constraints on the energies of the QPs involved in each process. As a result, the rates of parity switching by NUPS ($\Gamma_N^{i \rightarrow j}(\Phi)$) and by PAPS ($\Gamma_P^{i \rightarrow j}(\Phi)$) can have starkly different dependencies on the flux.

The offset-charge-sensitive SQUID transmon in this experiment is fabricated with $E_{J1}/h = 2.465$ GHz and $E_{J2}/h = 8.045$ GHz. The large asymmetry of the JJs enables Γ to be measured at all values of the flux by mapping the parity onto the state of the qubit [15,19]. With $E_C/h = 0.352$ GHz, δf_q varies in the range 0.7–14.5 MHz as the mean even-odd qubit frequency f_q is tuned from $f_q = 5.0594$ GHz at $\Phi/\Phi_0 = 0$ to $f_q = 3.5624$ GHz at $\Phi/\Phi_0 = 0.5$ [Appendix B].

B. Gap differences in aluminum films

A key factor influencing the available density of states for parity switching is the difference in the superconducting gaps of the two superconductors on either side of the JJ tunneling barrier. The superconducting gap of thin-film aluminum increases with decreasing film thickness [43–45]. Previous works on Cooper-pair transistors [3,45] and Cooper-pair box qubits [27,28] have taken advantage of this effect to trap QPs in thicker aluminum films and reduce QP poisoning of the island. However, the effect of the gap difference on parity switching in transmon qubits has not been previously reported.

The offset-charge-sensitive transmon measured here consists of two aluminum films of 20 nm and 30 nm, respectively [Figs. 1(a) and 1(b), [46]]. For these thicknesses, the gaps are expected to differ by $\delta\Delta := \Delta_H - \Delta_L \sim 20 \mu\text{eV} \sim 5 \text{ GHz} \times h$ [43,45]. This gap difference has two important consequences for parity switching in transmons. First, a gap difference changes the proportionality of Γ_N to x_{QP} , as certain QP-qubit interactions are suppressed or enhanced depending on the value of $\delta\Delta$ relative to the qubit transition energy hf_q [Fig. 1(a), purple arrows]. Second, the low-gap film can act as a QP trap and we show evidence that the excess QPs indeed relax to a cold distribution in the low-gap film [Sec. III B]. While not apparent in a single measurement of the parity-switching rate, these effects are revealed by our flux-dependence measurements, as we describe in the following sections.

III. EVIDENCE FOR GAP DIFFERENCE AND COLD QP ENERGY DISTRIBUTION

A. Flux-dependent parity-switching rates

Here, we show a signature of the gap difference $\delta\Delta$ in the flux-dependent $\Gamma(\Phi)$. As we explain below, this effect is specific to the NUPS mechanism and aids in the differentiation of PAPS from NUPS. The parity-switching rate is measured as a function of the flux through the SQUID loop of the qubit using techniques developed and demonstrated in Refs. [15,19] [for additional detail, see Appendix B]. Due to the fast repetition rate and the symmetrized parity-conditional π pulses, the qubit spends approximately equal time in the ground and excited states during the experiment. As a result, the measured parity-switching rate Γ weighs the parity-switching rates when the qubit is in $|0\rangle$ and $|1\rangle$ approximately evenly: $\Gamma \approx 0.5(\Gamma^{0 \rightarrow 0} + \Gamma^{0 \rightarrow 1}) + 0.5(\Gamma^{1 \rightarrow 1} + \Gamma^{1 \rightarrow 0})$.

The flux dependence of Γ displays a peak near $\Phi/\Phi_0 \approx 0.325$ ($f_q \approx 4.12$ GHz) on top of a broader increase from $\Phi/\Phi_0 = 0$ to 0.5 [Fig. 2(a), upper]. This broader increase is predicted to result from the flux dependence of the single-charge-tunneling matrix elements. Here, we focus on the intermediate peak, which may be understood as an enhancement of NUPS when the qubit energy matches the

gap difference $hf_q \approx \delta\Delta$. Under this condition, the rates of NUPS-induced relaxation ($\Gamma_N^{1 \rightarrow 0}$) and excitation ($\Gamma_N^{0 \rightarrow 1}$) are enhanced due to the divergences of the superconducting density of states on both sides of the JJ, since a QP at the gap edge on one side can tunnel to the gap edge on the other side by exchanging energy with the qubit. The implied gap difference $\delta\Delta \approx hf_q \approx 20 \mu\text{eV}$ is consistent with reported gap measurements for films of these nominal thicknesses [43,45,47]. This peak indicates that a considerable fraction of parity switching is due to NUPS, since PAPS is not enhanced in the same way by $\delta\Delta = hf_q$, as we now explain. PAPS depends on the sum $\Delta_H + \Delta_L$, rather than the gaps individually, because the sum of the energies of two QPs generated by PAPS is determined by the absorbed photon energy hf_p [Fig. 1(a), orange arrows]. This is in contrast to NUPS, for which the final energy of the tunneling QP is constrained to be the approximately the same as the initial energy and may only differ by the qubit energy [Fig. 1(a), purple arrows]. The PAPS rate therefore does not exhibit a peak when $hf_q = \delta\Delta$ and shows only a smooth flux dependence due to the matrix elements [Appendix A].

B. Probing the QP distribution

The peak in Γ also provides insight into the energetic and spatial distributions of QPs in the device. The average energy of QPs in each film appears to be close to its respective gap; otherwise, there would be not be a significant peak when $\delta\Delta = hf_q$. This is consistent with predictions that QPs generated at energies above Δ_{Al} relax rapidly by emitting phonons toward a steady-state distribution with average energy close to the gap [11,34,35].

If the QPs are indeed efficiently relaxing to the low energies, it would also be natural to expect that they tend to reside in the low-gap film (i.e., they are trapped there [41,48,49]). The previously described measurement has not been directly sensitive to this, so we develop a new experiment to probe the QP densities in each film. Note that $\Gamma_N^{1 \rightarrow 0}$ is enhanced by the presence of QPs in the low-gap film and requires the qubit to be in the excited state. In contrast, $\Gamma_N^{0 \rightarrow 1}$ requires QPs in the high-gap film of the JJ and the qubit initially in the ground state. Thus, by measuring Γ with different initial qubit states, we can learn about the distribution of QPs in the two films.

In our new protocol, the decay of the parity auto-correlation function $\langle p(0)p(\tau) \rangle$ is measured with the sequence shown in Fig. 2(b), which controls the time the qubit spends in the ground and excited states with active feedback. During the delay between parity measurements, the qubit undergoes repeated blocks of a qubit-state measurement and a preparation into $|\psi_\theta\rangle = \cos(\theta/2)|0\rangle + \sin(\theta/2)|1\rangle$. In this way, the qubit is projected to the ground (excited) state with probability $\cos^2\theta/2$ ($\sin^2\theta/2$) by the ensuing measurement. With

this procedure, $\langle p(0)p(\tau) \rangle \propto e^{-2\Gamma\tau}$ [Fig. 2(c)], with $\Gamma = \cos^2\theta/2(\Gamma^{0 \rightarrow 1} + \Gamma^{0 \rightarrow 0}) + \sin^2\theta/2(\Gamma^{1 \rightarrow 0} + \Gamma^{1 \rightarrow 1})$. Ideally, polarization angle $\theta = 0$ would keep the qubit in $|0\rangle$ and result in measurement of $\Gamma^0 := \Gamma^{0 \rightarrow 0} + \Gamma^{0 \rightarrow 1}$; $\theta = \pi$ would keep the qubit in $|1\rangle$ and result in measurement of $\Gamma^1 := \Gamma^{1 \rightarrow 1} + \Gamma^{1 \rightarrow 0}$. In practice, $T_1(\Phi) \approx 20\text{--}70 \mu\text{s}$ limits the experimentally attainable polarizations. Instead, the qubit measurement record during the delay is used to estimate the fraction of the delay the qubit spends in each state [Appendix C]. Then, plotting Γ as a function of the average qubit-state measurement $\langle m_q \rangle$ during the feedback delay, we use a linear fit to extrapolate to $\langle m_q \rangle = 0, 1$ and infer the parity-switching rate conditioned on the qubit state $|0\rangle$ (Γ^0) and $|1\rangle$ (Γ^1) [Fig. 2(d)].

We repeat this measurement as a function of flux and find that only Γ^1 exhibits a clear peak at $\Phi/\Phi_0 \approx 0.325$ [Fig. 2(a), lower]. This suggests that the increased parity-switching rate when $hf_q \approx \delta\Delta$ is due to QPs tunneling from the low-gap edge to the high-gap edge by relaxing the qubit [Fig. 2(e), red]. From this, we deduce that generated QPs relax to a relatively cold steady-state distribution with an average energy from the low-gap edge $\langle \varepsilon - \Delta_L \rangle \ll \delta\Delta$. The thicker lower-gap aluminum film acts as a built-in QP trap [Fig. 2(f)].

Therefore, the gap difference helps to reduce parity switching in two distinct ways. In the low-to-high-gap direction, tunneling is reduced because of the lack of available states in the high-gap film. In the high-to-low-gap direction, tunneling is reduced because the effective x_{QP} in the high-gap film of the JJ is reduced due to trapping in the low-gap film on top of it. For a QP distribution thermalized at a phonon temperature $T_{ph} \ll \delta\Delta$, the parity-switching rate would be exponentially suppressed while the qubit is in the ground state if $hf_q < \delta\Delta$. However, in Fig. 2(a, lower), it is clear that Γ^0 is always of the same order as Γ^1 , suggesting that PAPS is also contributing significantly to parity switching in this device.

IV. PAPS CONTRIBUTION TO PARITY SWITCHING

A. Self-consistent QP dynamics model

In order to elucidate the contributions of both NUPS and PAPS, we develop a model that takes into account the gap difference as well as parity switching and QP generation by PAPS. With this model, we can take advantage of the different flux dependence of NUPS and PAPS to distinguish between the respective contributions from the fit to $\Gamma(\Phi)$. The total parity-switching rate is the sum of both parity-switching mechanisms:

$$\Gamma(\Phi) = \Gamma_P(\Phi, \bar{n}, f_p) + \Gamma_N(\Phi, T_{ph}, x_0(\Phi), x_3(\Phi), \delta\Delta). \quad (1)$$

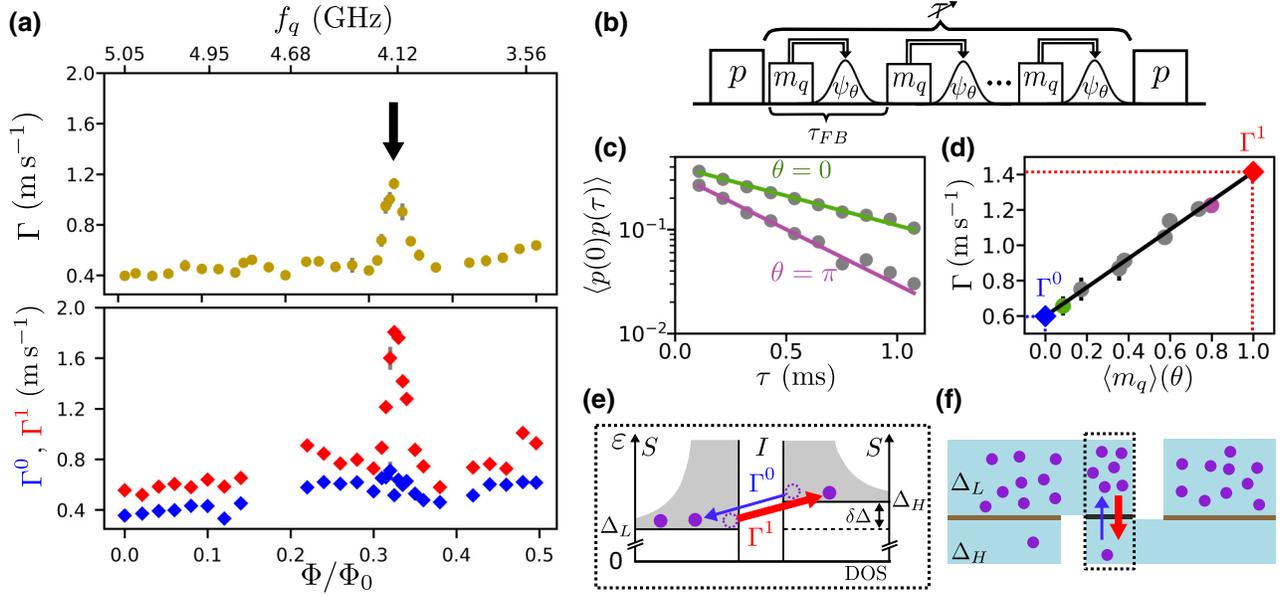


FIG. 2. (a) Upper: the flux dependence of the parity-switching rate $\Gamma \approx 0.5(\Gamma^0 + \Gamma^1)$, measured by fitting power spectral densities of jump traces of parity ($\Gamma^0 := \Gamma^{0 \rightarrow 0} + \Gamma^{0 \rightarrow 1}$, $\Gamma^1 := \Gamma^{1 \rightarrow 1} + \Gamma^{1 \rightarrow 0}$). The parity is measured by a Ramsey-like sequence that maps the parity onto the state of the qubit [15,19]. Lower: the flux dependence of the parity-switching rate when the qubit starts in the ground (blue) or excited (red) state, measured by the protocol described in (b). (b) The pulse sequence for measurement of Γ^0 , Γ^1 . Two measurements of the parity (p) are separated by a variable delay of length τ . For the duration of the delay, the qubit state is repeatedly measured (m_q , $4 \mu\text{s}$), followed by preparation of a superposition with polarization angle θ ($\psi(\theta)$), with feedback block repetition time $\tau_{FB} = 5.376 \mu\text{s}$. (c) The decay of the autocorrelation of the parity ($\langle p(0)p(\tau) \rangle$) for polarization angles $\theta = 0$ (green) and $\theta = \pi$ (pink) as a function of the delay τ between parity measurements ($\Phi/\Phi_0 = 0.335$). The parity autocorrelation function $\langle p(0)p(\tau) \rangle$ decays at the rate 2Γ , which depends on the θ -dependent time the qubit spends in $|0\rangle$ or $|1\rangle$. (d) The measured parity-switching rate Γ ($\Phi/\Phi_0 = 0.335$) (dots) as a function of the average measurement of the qubit state during τ ($\langle m_q \rangle$). The qubit-state-conditional parity-switching rates Γ^0 and Γ^1 (blue and red diamonds) are determined by extrapolation of a linear fit. (e) A schematic depicting NUPS in which the superconducting gap of the films differs by $\delta\Delta \approx hf_q$. At this flux, $\Gamma_N^{0 \rightarrow 1}$ and $\Gamma_N^{1 \rightarrow 0}$ are strongly enhanced by the divergent densities of states on both sides of the JJ. Γ^0 is lower than Γ^1 due to fewer QPs available for tunneling [Fig. 2(f)]. (f) A cross-section sketch of the device as in Fig. 1(b). Away from the JJ, QPs may tunnel between the two films over a very large contact area. QPs are shown predominantly in the low-gap film, illustrating the QP trapping effect, to which we attribute the rise in Γ^1 but not Γ^0 observed in the lower panel of (a).

Here, \bar{n} is the occupation of a high-frequency mode at f_P that couples to the JJ and induces PAPS at a per-photon rate calculated following Ref. [36]. In this model, the relative increase of Γ_P with Φ/Φ_0 varying from 0 to 0.5 is determined by f_P , while \bar{n} independently scales the magnitude of Γ_P . These are the two fitting parameters for the PAPS contribution to Γ . There is an ambiguity in the physical interpretation of these quantities, as it is possible for combinations of modes with varying occupations to give the same effective Γ_P [Appendix A]. The case for a narrow band of modes with dominant coupling to the JJ has been made previously [40,41] and is consistent with our data (discussed below).

The NUPS rate depends on the QP densities in the JJ films 0 and 3 (x_0 and x_3 ; “QP” is dropped for the x_{QP} of specific films for notational simplicity). In an isolated superconducting film, the steady state x_{QP} can be determined by balancing QP generation (g) with trapping (s)

and recombination (r):

$$\dot{x}_{QP} = g - sx_{QP} - rx_{QP}^2 = 0. \quad (2)$$

For our model, we extend this concept to all four films of the device (i.e., the high- and low-gap films on each side of the JJ), considering these dynamics in each film as well as tunneling between films. We separate the generation of QPs into two types. Generation by PAPS, $g_P(\Phi) = \Gamma_P(\Phi)/N_{CP}$, is included self-consistently with the flux-dependent $\Gamma_P(\Phi)$ contribution to $\Gamma(\Phi)$. Other pair breaking that does not directly result in a parity switch is accounted for by g_{other} , which is assumed to occur equally in each pad. Trapping at rate s may arise from vortices or gap inhomogeneities and is included as a fitting parameter also assumed to be the same for all four films. The recombination rate $r = 1/(120 \text{ ns})$ based on the literature [30] is also included in each film.

As described earlier [Sec. III B], the qubit-state dependence of the peak in $\Gamma(\Phi)$ indicates that QPs thermalize and become trapped in the low-gap films of each pad (films 0 and 2). This is supported by predictions that QPs relax by emitting phonons rapidly relative to the other dynamics in the system [34,49]. Under this assumption, we express the QP distributions as Fermi distributions thermalized at T_{ph} , $f(\epsilon, T_{\text{ph}}, \mu_{L(R)}) = 1/(e^{(\epsilon - \mu_{L(R)})/k_B T_{\text{ph}}} + 1)$, with nonzero chemical potential $\mu_{L(R)}$ accounting for excess QPs [27,50] on the left (right) side of the JJ. We take $T_{\text{ph}} \approx 50$ mK for the temperature of the device, based on the reading of a thermometer mounted near the sample cavity [Appendix D]. In the presence of the gap difference, this thermalization will result in the x_{QP} of the high-gap films being reduced from that of the low-gap films by a factor $e^{-\delta\Delta/k_B T_{\text{ph}}} \approx 0.008$.

Because the QPs are assumed to reside predominantly in the low-gap films, we can ignore the effects of trapping and recombination in the high-gap films, which have a negligible effect on the overall x_{QP} on each side. We can thus approximate the dynamics of x_{QP} in the low-gap films (x_0 and x_2):

$$\dot{x}_0 = g_P + g_{\text{other}} - s x_0 - r x_0^2 - \gamma_{03} x_0 + \gamma_{30} x_2 e^{-\delta\Delta/k_B T_{\text{ph}}}, \quad (3)$$

$$\dot{x}_2 = g_P + g_{\text{other}} - s x_2 - r x_2^2 + \gamma_{03} x_0 - \gamma_{30} x_2 e^{-\delta\Delta/k_B T_{\text{ph}}}. \quad (4)$$

Here, we include QP tunneling between films 0 and 3 by the rates γ_{03} and γ_{30} , which are the per-QP tunneling rates in each direction. These rates take into account T_{ph} and $\delta\Delta$ and are also flux dependent. If the qubit state was in thermal equilibrium with the QPs, x_0 and x_2 would be the same, because tunneling in each direction across the JJ would be balanced. However, as discussed in Sec. III A, the qubit is frequently π pulsed throughout the parity-mapping sequence, resulting in a nonthermal qubit population during measurement of Γ . The extra time the qubit spends in the excited state results in excess tunneling from film 0 to film 3 as compared to the reverse process, since Γ_N^{10} favors tunneling from low-gap to high-gap film due to the densities of states of the superconducting films. According to our model, the measurement therefore ‘‘pumps’’ QPs from film 0 to film 3, which produces a steady state x_{QP} in films 2 and 3 that is larger than in same-gap films 0 and 1 [51]. This effect is predicted to be particularly strong when $hf_q \approx \delta\Delta$.

Setting the left-hand side of Eqs. (3) and (4) equal to 0 results in coupled equations that can be solved to determine the steady state x_0 and x_2 . These values are flux dependent due to the flux dependence of $g_P(\Phi)$ and per-QP tunneling rates $\gamma_{03}(\Phi)$, $\gamma_{30}(\Phi)$. Finally, $\Gamma_N(\Phi)$ can be calculated given the x_{QP} of the JJ films ($x_0(\Phi)$ and

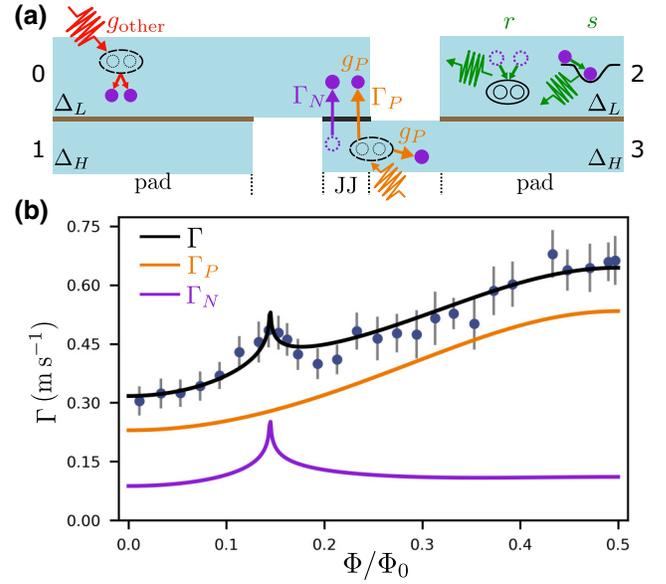


FIG. 3. (a) An illustration of the QP processes included in our model. The films are numbered 0–3 for reference in the text. The generation of QPs may occur by pair breaking in the pads (red) at rate g_{other} or by PAPS at the JJ (orange) at rate g_P . QPs tunnel across the JJ at rates Γ_P (orange) and Γ_N (purple). QPs are eliminated by recombination at rate r or are removed from the tunneling population by trapping at rate s in each film (green). (b) The parity-switching rate $\Gamma(\Phi)$ from $\Phi/\Phi_0 = 0$ to $\Phi/\Phi_0 = 0.5$. The fit to the self-consistent model (Sec. IV A) is indicated in black, with the Γ_N (purple) and Γ_P (orange) contributions shown explicitly. The peak near $\Phi/\Phi_0 \approx 0.145$ is explained by the $\delta\Delta \approx hf_q$ condition illustrated in Fig. 2(e).

$x_3(\Phi) = x_2(\Phi)e^{-\delta\Delta/k_B T_{\text{ph}}}$) and added to $\Gamma_P(\Phi)$ to determine $\Gamma(\Phi)$ [Eq. (2); for additional detail on the model, see Appendix E].

B. Distinguishing parity-switching mechanisms

In Fig. 3(b), we show $\Gamma(\Phi)$ for the same device measured during an earlier cool down, which shows a similar peak as seen in Fig. 2 but at a lower flux, corresponding to a value of $\delta\Delta$ that is $2.9 \mu\text{eV}$ higher. The cause of this shift is not known but it may be related to mechanisms causing JJ aging [52]. Fitting this data with the self-consistent model (black) yields $f_P = 112 \pm 2$ GHz, $\bar{n} = (1.9 \pm 0.2) \times 10^{-3}$, $\delta\Delta/h = 4860 \pm 5$ MHz. The values of g_{other} and s cannot be independently extracted from this fit, so we simply set $g_{\text{other}} = 0$ (below, this condition is relaxed, as we discuss). The QP densities in the low- and high-gap films that form the JJ (at $\Phi/\Phi_0 = 0$) are $x_0 \approx 6.2 \times 10^{-9}$ and $x_3 \approx 0.1 \times 10^{-9}$, respectively, corresponding to, on average, 65 QPs in the low-gap film and 0.7 QPs in the high-gap film. We decompose the fit Γ into its Γ_P (orange) and Γ_N (purple) components and see that the peak at $\Phi/\Phi_0 \approx 0.145$ is due to the effect of $\delta\Delta = hf_q$ on Γ_N , while Γ_P is insensitive to $\delta\Delta$ and increases

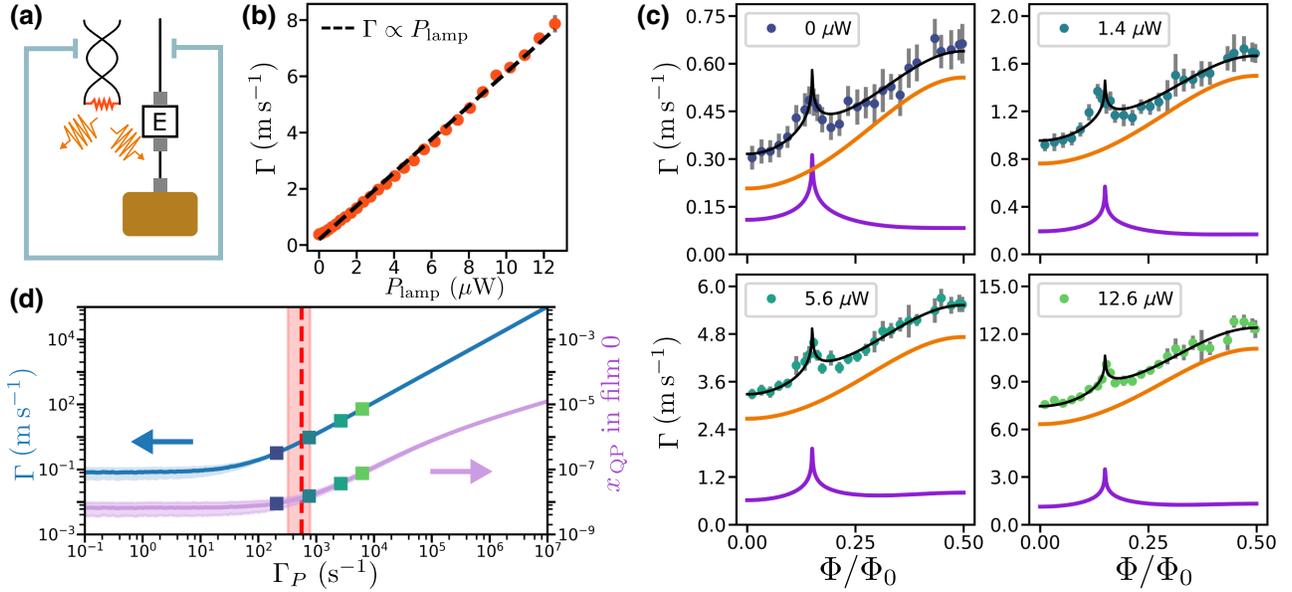


FIG. 4. (a) The manganin “lamp” resistor hangs in the aluminum can shielding the copper cavity containing the qubit. The coax lines include a high-frequency-absorbing Eccosorb CR-110 filter (“E”) in line just before the readout cavity, inside the can. The passing of current through the lamp causes its temperature to rise, resulting in additional radiation that can leak into the cavity and induce PAPS. Photons may enter the electromagnetic environment of the qubit through SMA connectors or packaging seams. (b) Γ as a function of the power dissipated by the manganin “lamp” resistor, measured at $\Phi/\Phi_0 = 0$. Γ increases approximately linearly with the power dissipated by the lamp (black dashed; for discussion, see Appendix F). (c) $\Gamma(\Phi)$ measured at $P_{\text{lamp}} = \{0, 1.4, 5.6, 12.6\} \mu\text{W}$. We perform a simultaneous fit to all four data sets allowing unique f_p and \bar{n} and shared s , g_{other} , and $\delta\Delta$ (solid black). The contributions of PAPS (orange) and NUPS (purple) indicate that as the lamp power increases, both types of parity switching increase. (d) The simulated $\Gamma(0)$ (blue, left axis) and x_0 (lavender, right axis) as a function of Γ_P , calculated using fit values from Fig. 4(c). Experimental values (squares) are marked for reference and the PAPS level at which $g_p = g_{\text{other}}$ is marked by the red dashed line. The reduction of Γ_P beyond the current background level ($P_{\text{lamp}} = 0$) would reduce Γ and x_{QP} but the improvement is limited due to g_{other} , which is comparable to g_p .

monotonically due to the matrix elements. Importantly, we find that $0.53 \leq \Gamma_P(\Phi)/\Gamma(\Phi) \leq 0.83$, indicating that both mechanisms contribute significantly to parity switching.

From the fit to Fig. 3(b), we extract a QP-induced excitation rate $\Gamma^{0 \rightarrow 1}(0) = 134 \text{ s}^{-1}$ and a QP-induced relaxation rate $\Gamma^{1 \rightarrow 0}(0) = 247 \text{ s}^{-1}$. The ratio $\Gamma^{0 \rightarrow 1}/\Gamma^{1 \rightarrow 0} \approx 0.53$ implies that QPs cause much more qubit excitation than would be expected by detailed balance at 50 mK, which would predict $\Gamma^{0 \rightarrow 1}/\Gamma^{1 \rightarrow 0} \approx 8 \times 10^{-3}$. A previous observation of this type of anomalous excitation has been interpreted as evidence of a nonthermal QP distribution of unclear origin [19]. Here, this apparently nonthermal ratio is interpreted as resulting from PAPS.

V. QP GENERATION BY PAPS

A. Varying photon incidence

Lastly, we introduce a variable photon source in order to deduce the contribution of PAPS to QP generation. While the fit in Fig. 3(b) determines the contribution of Γ_P to Γ , it does not determine whether the QPs responsible for the observed Γ_N are generated solely by Γ_P or

if other generation mechanisms also contribute. The generation contributions g_{other} cannot be determined from the data shown in Fig. 3 because Γ_N depends on x_0 and x_3 and the trapping rate s can compensate varying levels of g_{other} to yield the same QP densities. To obtain an estimate of g_{other} to compare to g_p , we vary the photon incidence and observe the response in x_{QP} .

To do this, we add a resistor that acts as a controllable source of additional photons for PAPS [Fig. 4(a)]. A 2 cm length of manganin wire (approximately 1.4Ω) is suspended inside the aluminum shield containing the copper cavity, though notably not in line with the coaxial cable for the microwave reflection measurement. When current is passed through this “lamp” resistor, we observe that $\Gamma(0)$ increases approximately linearly with the power dissipated by the lamp [Fig. 4(b), Appendix F]. We do not attribute this to an increase in temperature resulting in additional thermal QPs, because the temperature measured by a ruthenium oxide (RuO_2) thermometer installed at the bracket holding the copper three-dimensional (3D) readout cavity increased by only a few millikelvins. Instead, we attribute the increase of Γ to an increase in PAPS (including the added generation g_p), suggesting that the SMA

connections to the cavity and Eccosorb filter inside the aluminum shield may allow photons to leak into the input line [53].

The parity-switching rate $\Gamma(\Phi)$ is measured at $P_{\text{lamp}} = 0 \mu\text{W}$, $1.4 \mu\text{W}$, $5.6 \mu\text{W}$, and $12.6 \mu\text{W}$ [Fig. 4(c)] corresponding to the sequence of 0, 1, 2, and 3 mA of current through the $1.4\text{-}\Omega$ manganin filament. Since the overall value of Γ increases by 25 times from $P_{\text{lamp}} = 0 \mu\text{W}$ to $12.6 \mu\text{W}$ and the intermediate peak at $\Phi/\Phi_0 \approx 0.145$ is still visible, it is clear that Γ_N must be increasing along with Γ_P . This indicates additional x_{QP} generated by PAPS and constrains s and g_{other} [Appendix G]. We fit these curves simultaneously with the model described above, allowing unique \bar{n} and f_P for each P_{lamp} and assuming common s , g_{other} , and $\delta\Delta$ (fit values shown in Table I). The f_P and \bar{n} values in the table for nonzero lamp power correspond to the additional PAPS added by the lamp on top of the $P_{\text{lamp}} = 0 \mu\text{W}$ background.

From this fit, we observe that with the lamp off ($P_{\text{lamp}} = 0 \mu\text{W}$), the ratio of generation contributions $g_P/g_{\text{other}} = 0.37$, i.e., the rate at which PAPS generates QPs, is comparable to generation by other sources. The extracted trapping rate $s = 11 \text{ s}^{-1}$ is of the same order as the trapping rate estimated in Ref. [30]. Surprisingly, we find that the effective frequency of the additional PAPS-generating photons from the lamp is $f_P \approx 125 \text{ GHz}$ for each power. If the coupling of high-frequency photons to the qubit was broadband, we would expect this frequency to increase as the power of the resistor increases, due to the rising temperature of the emitting black body. Instead, the results may indicate that higher lamp power causes increased occupation of modes that are well matched to antenna modes of the qubit, which have relatively high absorption efficiency [40–42]. A detailed understanding of the spectrum and coupling of PAPS-inducing radiation in 3D transmons is left for future work.

B. Implications for mitigating parity-switching decoherence

Having established the QP dynamics, we may extrapolate Γ under different levels of Γ_P -inducing radiation. In Fig. 4(d), we simulate sweeping Γ_P for the device measured in this work. The blue curve shows the calculated $\Gamma(\Phi = 0)$ with $g_{\text{other}} = 8 \times 10^{-8} x_{\text{QP}}/s$, $s = 11 \text{ s}^{-1}$,

TABLE I. The model parameters corresponding to the fit shown in Fig. 4(c).

P_{lamp} (μW)	f_P (GHz)	\bar{n} $\times 10^{-3}$	s (s^{-1})	g_{other} (x_{QP}/s) $\times 10^{-8}$	$\delta\Delta$ (MHz)
0	109 ± 3	2.1 ± 0.3	11 ± 2	8 ± 3	4844 ± 3
1.4	125 ± 3	2.9 ± 0.1			
5.6	125 ± 2	12.8 ± 0.4			
12.6	124 ± 1	32.6 ± 0.8			

and $\delta\Delta = 4844 \text{ MHz}$ from the fit in Fig. 4(c), while the lavender curve depicts the corresponding x_0 as a function of Γ_P . The experimental values derived from the fits in Fig. 4(c) are marked and the Γ_P at which $g_P = g_{\text{other}}$ is marked by the red dashed line. In the $g_P \gg g_{\text{other}}$ regime, decreasing Γ_P efficiently lowers Γ . High-frequency absorbing filters in the rf lines have indeed been shown to significantly lower Γ [32]. The efficient increase of Γ with P_{lamp} emphasizes the previously observed importance of light-tight shielding surrounding the device [29, 54] to reduce the flux of such photons seen by SMA connectors below the last in-line filter.

The lower bound due to the QPs generated by non-PAPS mechanisms is $\Gamma \approx 81 \text{ s}^{-1}$, suggesting that removing Γ_P entirely would help reduce Γ only by a factor of 4 in this device, with x_0 reaching a plateau due to g_{other} . To further decrease the parity-switching rate, QPs generated in the pads of the device would need to be addressed. One such possible source is ionizing radiation, which has been shown to cause correlated errors in qubits across a single substrate attributed to bursts of QPs being generated by the impact and then tunneling [37]. In this device, we observe sudden occurrences of rapid parity switching directly [Appendix H], which supports the interpretation of QPs being the mechanism for these errors. The frequency and amount of energy deposited by these bursts as well as the time scale for decay of the generated QP density will determine the extent to which these impacts contribute to g_{other} , which is a subject for future investigation.

An alternative approach is to reduce the harmful impact of QP generation by preventing QPs from tunneling after they have been generated, via QP traps. These may be implemented by additional normal-metal or lower-gap superconductor traps [48,49] but a simpler solution may be to increase the gap difference between the aluminum films. For a fixed number of QPs in the device, all NUPS rates decrease exponentially with the difference between the gaps as discussed above.

The fit to our model yields a limit on the energy-relaxation time due to QPs of $T_1^{\text{QP}} = (\Gamma^{0 \rightarrow 1} + \Gamma^{1 \rightarrow 0})^{-1} = 2.2 \text{ ms}$ for this device, which is about one order of magnitude longer than T_1 of current state-of-the-art transmon qubits [39,54,55]. While parity-switching decoherence does not currently limit transmons, the eventual reduction of dielectric loss [56,57] will motivate further mitigation of parity switching. Recently, parity-switching rates well below the lower limit imposed by non-PAPS sources in this experiment have been measured [39,41,54,58], with $\Gamma < 1 \text{ s}^{-1}$ being measured in Refs. [41,58]. To reduce Γ , these experiments each use some combination of improved shielding or engineering of the coupling to high-frequency photons to reduce PAPS [41,42,54], a larger gap difference between the pads and the JJ to prevent QPs from tunneling [41], and normal metal reservoirs on the back side of the substrate to reduce QP generation by phonons [58].

VI. CONCLUSIONS AND OUTLOOK

We measure the flux dependence of parity switching to distinguish the contributions of photon-assisted parity switching (PAPS) and number-conserving parity switching (NUPS). In the flux dependence, we observe a peak that stems from NUPS in the presence of a difference between the superconducting gaps of the aluminum films of our device. The dependence of this peak on the flux dependence on the qubit state indicates that QPs relax into a low energy distribution in the low-gap aluminum film. We fit the flux dependence of parity switching with a model that takes into account QP dynamics between the films of the qubit and self-consistent generation of QPs by PAPS. From this fit, we conclude that parity switching in this device is consistent with comparable contributions of PAPS and NUPS. We also find that PAPS generate QPs at a rate similar to other processes that do not directly change the parity. This work shows that parity switching in transmon qubits cannot be understood in terms of solely NUPS defined by a single QP density. The roles of PAPS and the gap difference must be considered to accurately determine the QP densities from measurements of the parity-switching rate Γ or the QP-limited energy-relaxation time.

We note that from these results, we cannot rule out the alternative explanation that subgap states in the JJ cause enhanced tunneling at certain energies. However, the agreement of the observed $\delta\Delta$ with reported measurements of thin-film aluminum and the lack of an observed peak in a device with an intentionally large gap difference support the gap-difference interpretation (Appendix I). Future investigations probing a suite of devices with varying gap differences will further clarify the underlying physics. Additionally, qubits with different metallization of the pads and geometric footprints can be utilized to modify the spatial distribution of QPs and to elucidate the spectrum of PAPS-inducing radiation, respectively [40–42]. The framework introduced here will assist investigations of QP generation and impact in these devices going forward.

ACKNOWLEDGMENTS

We acknowledge helpful discussions with N. Frattini, A. Koottandavida, G. Catelani, M. Houzet, and R. J. Schoelkopf. Facilities use was supported by The Yale Institute for Nanoscience and Quantum Engineering (YINQE) and the Yale School of Engineering & Applied Science (SEAS) cleanroom. We also acknowledge the Yale Quantum Institute. The research was sponsored by the Army Research Office (ARO) and was accomplished under Grant No. W911-18-1-0212. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the ARO or the U.S. Government. The U.S. Government is authorized to reproduce

and distribute reprints for Government purposes notwithstanding any copyright notation herein. M.H.D. and L.F. are founders and L.F. is a shareholder of Quantum Circuits Inc.

S.D., K.S., M.H., and V.F. designed the device and experimental setup. S.D. fabricated the qubit with assistance from L.F. Author V.R.J. fabricated the parametric amplifier. S.D. performed the measurements with feedback from V.F., M.H., P.D.K., and T.C. Authors S.D., V.F., M.H., H.N., P.D.K., T.C., L.I.G., and M.H.D. developed the model and analyzed the data. S.D., V.F., and M.H.D. wrote the manuscript with feedback from all authors.

APPENDIX A: CALCULATING Γ_N AND Γ_P

1. Single-charge-tunneling qubit transition rates

In this appendix, we derive Γ_N and Γ_P following Refs. [14,36] and additionally incorporate the flux dependence of Γ for the SQUID device. The Hamiltonian for single-charge tunneling across the JJ and coupling to the phase degree of freedom $\hat{\phi}$ of the qubit is

$$\hat{H}_{\text{QP},\hat{\phi}} = t \sum_{r,l,s} e^{i\hat{\phi}/2} \hat{c}_{r,s}^\dagger \hat{c}_{l,s} + \text{H.c.}, \quad (\text{A1})$$

where t is the tunneling amplitude, $\hat{c}_{l,s}$ is the electron annihilation operator for the reservoir on the left side of the junction, and $s = \uparrow, \downarrow$ denotes the spin of the electron. We apply the Bogoliubov transformation, which diagonalizes the BCS Hamiltonian of the superconductors in the leads:

$$\hat{c}_{l\uparrow} = u_l \hat{\gamma}_{l\downarrow} + v_l \hat{\gamma}_{l\uparrow}, \quad \hat{c}_{l\downarrow}^\dagger = -v_l \hat{\gamma}_{l\downarrow} + u_l \hat{\gamma}_{l\uparrow}. \quad (\text{A2})$$

The operator $\hat{\gamma}_{l\downarrow}$ ($\hat{\gamma}_{l\downarrow}^\dagger$) is the operator for annihilation (creation) of a QP excitation on the left with spin down and u_l and v_l are the conventionally defined BCS coherence factors, which depend on the QP energy ϵ_l [50]. Applying this transformation to Eq. (A1), we see that two mechanisms of single-charge tunneling may occur:

$$\begin{aligned} \hat{H}_{\text{QP},\hat{\phi}} = t \sum_{r,l,s} & \left\{ \left[(u_r u_l - v_r v_l) \cos \frac{\hat{\phi}}{2} \right. \right. \\ & \left. \left. + i (u_r u_l + v_r v_l) \sin \frac{\hat{\phi}}{2} \right] \hat{\gamma}_{rs}^\dagger \hat{\gamma}_{ls} \right. \\ & \left. + \left[(u_r u_l + v_r v_l) \cos \frac{\hat{\phi}}{2} + i (u_r u_l - v_r v_l) \sin \frac{\hat{\phi}}{2} \right] \hat{\gamma}_{rs}^\dagger \hat{\gamma}_{ls}^\dagger \right\} \\ & + \text{H.c.} \end{aligned} \quad (\text{A3})$$

The first term, with $\hat{\gamma}_{rs}^\dagger \hat{\gamma}_{ls}$, accounts for NUPS as it empties a QP state on one side and fills a state on the other. The second term, with $\hat{\gamma}_{rs}^\dagger \hat{\gamma}_{ls}^\dagger$, generates QPs on both sides. In order to conserve energy, this process may only happen with the absorption of a photon of energy greater than

$\Delta_L + \Delta_H$ (the conjugate process annihilates QPs on both sides and emits a photon). Such radiation will couple to the qubit by imposing a time-dependent phase across the junction. Assuming that the phase increments φ_P induced by the electric field of the incident photon with frequency ω_P are small, the single-charge-tunneling operators are transformed by linear expansion of the trigonometric functions in the field-induced phase increments:

$$\begin{aligned}\cos \frac{\hat{\varphi}}{2} &\rightarrow \cos \frac{\hat{\varphi}}{2} - \varphi_P \sin \omega_P t \sin \frac{\hat{\varphi}}{2}, \\ \sin \frac{\hat{\varphi}}{2} &\rightarrow \sin \frac{\hat{\varphi}}{2} + \varphi_P \sin \omega_P t \cos \frac{\hat{\varphi}}{2}.\end{aligned}\quad (\text{A4})$$

We now apply Fermi's Golden Rule to calculate the rates of parity switching accompanied by a qubit transition from plasmonic eigenstate i to j . Application of Fermi's Golden Rule to the first term of Eq. (A3) gives the rate of NUPS accompanied by such a transition ($\Gamma_N^{i \rightarrow j}$). The second term of Eq. (A3) gives the rate of PAPS ($\Gamma_P^{i \rightarrow j}$), assuming the presence of high-frequency photons and expanding the single-charge-tunneling operators as above [Eq. (A4)]. We also use the Ambagaokar-Baratoff relation to express t in terms of the Josephson energy of the junction E_J and substitute the energy-dependent definitions of u and v , to find

$$\Gamma_N^{i \rightarrow j} = \frac{16E_J}{\pi \hbar} \left[\left| \langle i | \cos \frac{\hat{\varphi}}{2} | j \rangle \right|^2 S_{-N} + \left| \langle i | \sin \frac{\hat{\varphi}}{2} | j \rangle \right|^2 S_{+N} \right], \quad (\text{A5})$$

$$\Gamma_P^{i \rightarrow j} = \frac{\bar{n} g^2 \omega_r}{\pi \omega_q \omega_P} \left[\left| \langle i | \cos \frac{\hat{\varphi}}{2} | j \rangle \right|^2 S_{-P} + \left| \langle i | \sin \frac{\hat{\varphi}}{2} | j \rangle \right|^2 S_{+P} \right]. \quad (\text{A6})$$

Here, $\Gamma_P^{i \rightarrow j}$ is the total PAPS rate induced by photons with frequency ω_P . To calculate it, we multiply the transition rate induced by a single photon at ω_P by the average photon number in the mode \bar{n} . The coupling factor that determines the per-photon rate depends on the electric field amplitude and effective dipole length of the qubit. Following Ref. [36], we express this factor in terms of system parameters: the geometric coupling rate $g/2\pi = 331$ MHz, the readout resonator frequency $\omega_r/2\pi = 9.126$ GHz, and the qubit frequency at $\Phi/\Phi_0 = 0$ $\omega_q/2\pi = 5.0594$ GHz. Here, we only account for PAPS transitions between the ground and first excited states of the qubit. Photon-assisted transitions to higher states may also occur and add to Γ_P but inclusion of these transitions would not qualitatively alter these results.

The so-called QP structure factors $S_{\pm} = S_{\pm}^r + S_{\pm}^l$ include the BCS coherence factors, the QP distribution

functions, and the superconducting density of states ν . Tunneling from left to right and right to left are summed to obtain the total rate. The structure factors can be expressed as follows:

$$\begin{aligned}S_{\pm N}^{lr}(\omega_{ij}) &= \frac{1}{\Delta} \int_0^{\infty} d\varepsilon \frac{1}{2} \left(1 \pm \frac{\Delta_l \Delta_r}{\varepsilon(\varepsilon - \hbar\omega_{ij})} \right) \\ &\times f(\varepsilon, T_{\text{ph}}, \mu_l) \nu(\varepsilon, \Delta_l) \\ &\times [1 - f(\varepsilon - \hbar\omega_{ij}, T_{\text{ph}}, \mu_r)] \nu(\varepsilon - \hbar\omega_{ij}, \Delta_r),\end{aligned}\quad (\text{A7})$$

$$\begin{aligned}S_{\pm P}^{lr}(\omega_{ij}) &= \frac{1}{\Delta} \int_0^{\infty} d\varepsilon \frac{1}{2} \left(1 \pm \frac{\Delta_l \Delta_r}{\varepsilon(\hbar\omega_P - \varepsilon - \hbar\omega_{ij})} \right) \\ &\times [1 - f(\varepsilon, T_{\text{ph}}, \mu_l)] \nu(\varepsilon, \Delta_l) \\ &\times [1 - f(\hbar\omega_P - \varepsilon - \hbar\omega_{ij}, T_{\text{ph}}, \mu_r)] \nu \\ &\times (\hbar\omega_P - \varepsilon - \hbar\omega_{ij}, \Delta_r).\end{aligned}\quad (\text{A8})$$

Nonzero values of μ in $f(\varepsilon, T_{\text{ph}}, \mu) = 1/e^{(\varepsilon - \mu)/k_B T_{\text{ph}}} + 1$ describe distributions of QPs that are thermalized at T_{ph} but have excess number [27]. The total unassisted and photon-assisted parity-switching rates for a single junction device Γ_N and Γ_P are then calculated as the sums of the individual rates with qubit transition from i to j weighted by the qubit-state probabilities ρ_0, ρ_1 : $\Gamma_m = \rho_0(\Gamma_m^{0 \rightarrow 0} + \Gamma_m^{0 \rightarrow 1}) + \rho_1(\Gamma_m^{1 \rightarrow 1} + \Gamma_m^{1 \rightarrow 0})$, with $m \in \{N, P\}$.

In the SQUID transmon, single-charge tunneling across either JJ results in a parity switch. Therefore, to calculate the parity-switching rates in the SQUID transmon as a function of the flux Φ through the loop, we sum the respective rates across the JJs with E_{J1} (Γ_{m1}) and with E_{J2} (Γ_{m2}):

$$\Gamma_m(\Phi) = \Gamma_{m1}(\Phi) + \Gamma_{m2}(\Phi), \quad (\text{A9})$$

with $m \in \{N, P\}$. The SQUID-transmon Hamiltonian can be expressed in terms of both Josephson energies $\hat{H} = 4E_C(\hat{n} - n_g)^2 - E_{J1} \cos(\hat{\varphi} - \varphi_{\text{ext}}) - E_{J2} \cos \hat{\varphi}$, with the externally tunable flux $\varphi_{\text{ext}} = 2\pi \Phi/\Phi_0$. Equations (A5) and (A6) are modified to calculate the parity-switching rate for the individual junctions: $E_J \rightarrow E_{J1}$, $\hat{\varphi} \rightarrow \hat{\varphi} - \varphi_{\text{ext}}$ for Γ_{m1} ; $E_J \rightarrow E_{J2}$ for Γ_{m2} . The proportionality factor for photon absorption in Eq. (A6) also includes a factor of $E_{J1(2)}/E_{J1} + E_{J2}$, for $\Gamma_{m1(2)}$.

The flux dependence of Γ results from the tuning of the qubit frequency with the flux [see ω_{ij} in Eqs. (A7) and (A8)] as well as the flux-dependent matrix elements. In Fig. 5, we plot these matrix elements as a function of the flux. The green (purple) lines correspond to the matrix elements for single-charge tunneling across the JJ with $E_{J1} \approx 2.5$ GHz ($E_{J2} \approx 8$ GHz). The overall increase in $\Gamma_P(\Phi)$ from $\Phi/\Phi_0 = 0$ to $\Phi/\Phi_0 = 0.5$ results primarily from the increase of the $|\langle 0 | \sin(\hat{\varphi} - \varphi_{\text{ext}}/2) | 0 \rangle|^2$ and

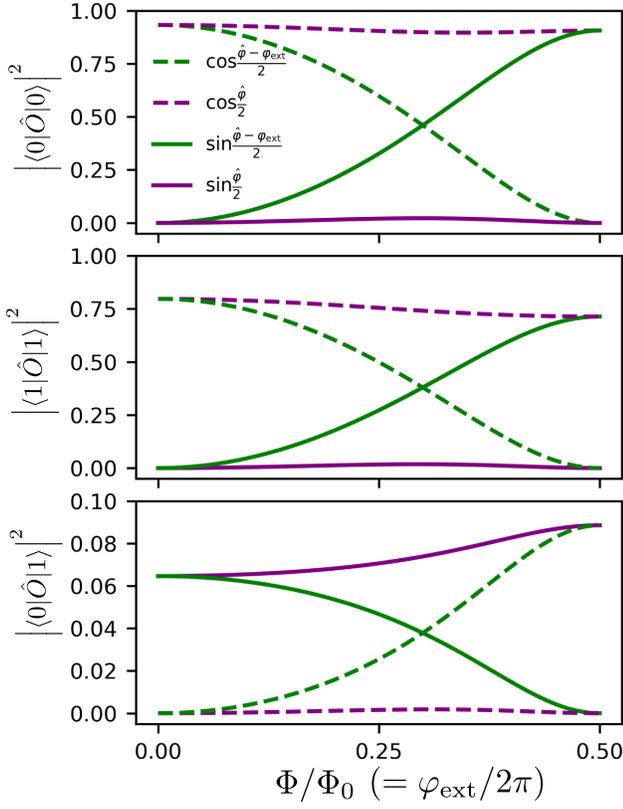


FIG. 5. The qubit matrix elements for single-charge-tunneling as a function of the flux through the SQUID loop, with $\Phi/\Phi_0 = \varphi_{\text{ext}}/2\pi$. The upper, middle, and lower panels correspond to matrix elements between plasmonic eigenstates $0 \leftrightarrow 0$, $1 \leftrightarrow 1$, and $0 \leftrightarrow 1$. The green (purple) lines correspond to the matrix elements for single-charge tunneling across the JJ with $E_{J1} \approx 2.5$ GHz ($E_{J2} \approx 8$ GHz).

$|\langle 1 | \sin(\hat{\varphi} - \varphi_{\text{ext}}/2) | 1 \rangle|^2$ matrix elements for single-charge tunneling across the lower- E_J JJ. Note that the wave functions $|i\rangle$ implicitly depend on φ_{ext} , ensuring that the choice of assignment for φ_{ext} does not affect the calculated rates.

2. Modeling Γ_P assuming single photon frequency f_P

For the purposes of this work, the true spectrum of the radiation inducing PAPS is not required. The total induced $\Gamma_P(\Phi)$ for arbitrary spectra of PAPS-inducing photons above $\Delta_L + \Delta_H$ can be approximated by an effective occupation \bar{n} of a single mode at frequency f_P : $\Gamma_P(\Phi) = \Gamma_P(\Phi, \bar{n}, f_P)$. Three examples of this are shown in Fig. 6. The relative flux dependence of the photon-assisted parity-switching rate, $\Gamma_P(\Phi)/\Gamma_P(0)$, depends on the absorbed photon frequency f_P . Lower f_P values induce a stronger relative increase with Φ , while f_P causes a weaker relative increase, as shown for 110 GHz (black dashed) and 300 GHz (black dot-dash) in Fig. 6. This is due to smaller photon energies generating QPs closer to the gap, where interference between electronlike and holelike tunneling of

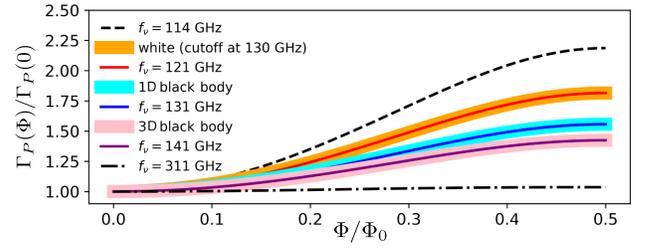


FIG. 6. The calculated relative flux dependence of Γ induced by various absorbed photon spectral densities. The thin lines correspond to Γ_P calculated for individual photon frequencies as described in Appendix A, with lower photon energies resulting in a larger increase in $\Gamma_P(\Phi)/\Gamma_P(0)$. The thick lines correspond to Γ_P calculated for the following spectral densities: a white spectrum with a cutoff at $2.5\Delta \approx 130$ GHz (orange), a one-dimensional black body at 1 K (cyan), and a 3D black body at 1 K (pink). Each of these spectral densities results in $\Gamma_P(\Phi)/\Gamma_P(0)$ that can be matched by PAPS induced by individual photon frequencies (thin colored lines).

QPs is stronger (see Eq. (A8): lower f_P results in larger S_{+P}/S_{-P} , which emphasizes the flux dependence of the matrix elements).

In Fig. 6, we calculate $\Gamma_P(\Phi)$ assuming the spectral densities of a white spectrum with a cutoff frequency of 130 GHz (orange), a one-dimensional black body at 1 K from 110 to 300 GHz (cyan), and a 3D black body at 1 K from 110 and 300 GHz (pink). We see that the relative flux dependence $\Gamma_P(\Phi)/\Gamma_P(0)$ for each spectral density can be fitted instead by PAPS induced by photons at a single frequency (solid curves). The absolute value of Γ_P will depend on the unknown attenuation of the radiation as it couples into the cavity, along with the possibly photon-frequency-dependent coupling rate to the qubit. Accordingly, for simplicity in our model, $\Gamma_P(\Phi)/\Gamma_P(0)$ is described by a single effective frequency f_P and an average photon number \bar{n} that determines the magnitude of the Γ_P contribution. Further experiments with different qubit geometries could be performed to elucidate the spectrum and coupling of radiation inducing PAPS, as done for two-dimensional (2D) qubits in Refs. [41,42].

APPENDIX B: MEASUREMENT OF Γ

1. Measurement protocol

In order to obtain a single measurement of Γ , we take the power spectral density (PSD) of a jump trace of the parity. The jump trace is measured using the Ramsey sequence first demonstrated in Ref. [15], which maps the parity onto the qubit state. The sign of the second $\pi/2$ pulse in the measurement sequence is alternated between parity measurements such that the sequences enact π pulses that alternate between being conditioned on the even or odd parity. While a measurement of Γ could be determined from a jump trace that is approximately 2 s long

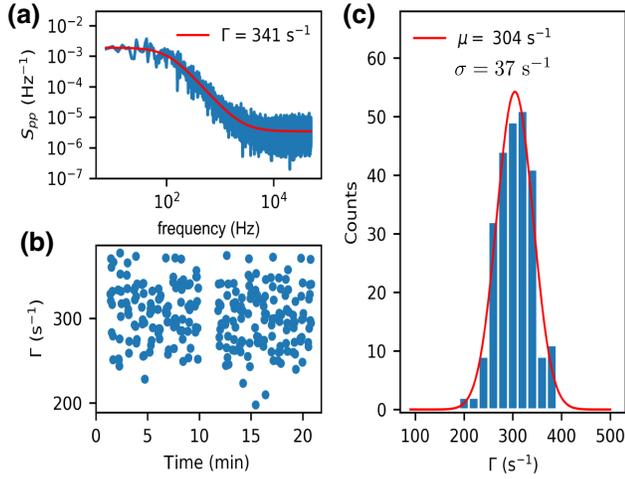


FIG. 7. (a) An example of the PSD of a 400-ms jump trace of parity. The PSD is fitted by a Lorentzian with characteristic rate Γ , the parity-switching rate, modified by a frequency-independent term due to infidelity of the parity-mapping sequence. (b) The time series of Γ measurements [e.g., Fig. 7(a)], showing the fluctuations of the measured value as a function of time. The gap in the data near 11 min corresponds to a time when Γ is not measurable due to the value of δf not meeting the criteria described in the text. (c) The histogram of the 250 measurements in (b). The histogram is fitted well by a Gaussian distribution. The mean and uncertainty μ and σ give the value and error bar of the individual Γ measurements as plotted in Figs. 2(a, upper), 3(b), 4(b), and 4(c).

with good signal-to-noise ratio (e.g., Fig. 7(a), in which $\Gamma = 341 \pm 2 \text{ s}^{-1}$), we observe that these measured values fluctuate in time [Fig. 7(b)]. In order to average over these fluctuations, we measure Γ over approximately 20 min.

For the data shown in the main text, we measure 25 jump traces, each comprised of 2×10^6 measurements of the parity repeated every $10 \mu\text{s}$. Each of these 20 s jump traces is chopped into 25 segments that are 400 ms long. We compute the PSD of each segment and average five PSDs together to obtain one value of Γ [Fig. 7(a)]. Finally, we fit the distribution of all Γ measurements with a Gaussian to determine the mean value of Γ [Fig. 7(c)], as well as an estimate of the fluctuations from the σ of the distribution. The width of this distribution does not become narrower when more measurements are included, suggesting that the width is due to fluctuation of Γ rather than measurement uncertainty.

In between each jump trace, the difference between the even- and odd-parity qubit frequencies is checked by a Ramsey experiment. If the value of δf meets two criteria, the next jump trace is measured. First, the value of δf has to exceed a threshold value such that the delay during the parity-mapping sequence $\tau = 1/4\delta f$ remains well below $T_2 \sim 2 \mu\text{s}$. Measuring at n_g with δf below this threshold would result in reduced fidelity of the measurement. Second, for fluxes at which the effective $E_J/E_C \lesssim 20$, the

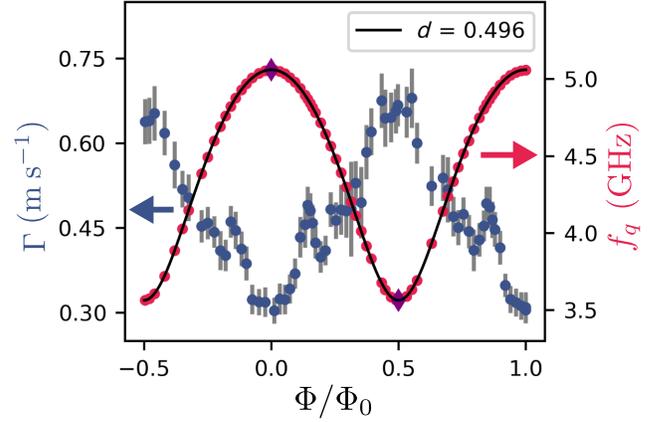


FIG. 8. The charge-parity-switching rate Γ (blue, left axis) and the mean even-odd-parity qubit frequency f_q (pink, right axis) as a function of the flux from $\Phi/\Phi_0 = -0.5$ to 1.0 . The qubit frequencies at $\Phi/\Phi_0 = 0$ and $\Phi/\Phi_0 = 0.5$ are taken to be the maximal and minimal qubit frequencies, respectively, and are measured at the beginning of the cool down: $f_{q,\text{min}} = 3.5624 \text{ GHz}$, $f_{q,\text{max}} = 5.0594 \text{ GHz}$ (purple diamonds). The black line shows a fit to just these two data points $f_q(0) = 5.0594 \text{ GHz}$ and $f_q(0.5) = 3.5624 \text{ GHz}$ with the expected flux dependence of the frequency $f_q(\Phi) = f_q(0)(\cos^2 \pi \Phi/\Phi_0 + d^2 \sin^2 \pi \Phi/\Phi_0)^{1/4}$, where $d = (E_{J1} - E_{J2})/(E_{J1} + E_{J2})$. The rest of the measurements are assigned flux values based on the qubit frequency according to this model.

dispersive shifts of the ground and excited states depend on n_g [32]. In order to measure the parity by mapping onto the qubit state, the ground and excited states need to have dispersive shifts such that the phase of the reflected measurement signal differs between $|0\rangle$ and $|1\rangle$. Therefore, the value of δf found by the Ramsey measurement needs to correspond to a value where the ground and excited states are separable. This range of usable δf values is determined by inspection at each flux prior to measurement of Γ . It is also verified that Γ does not itself depend on δf . If the value of δf does not meet these two criteria, the dc voltage on the readout pin of the 3D cavity is changed to induce a change in n_g and δf is measured again.

2. Measured Γ over wider flux range

In Fig. 8, we show the measured parity-switching rate Γ as a function of the flux from $\Phi/\Phi_0 = -0.5$ to 1.0 for the cool down corresponding to the data in Figs. 3 and 4. The intermediate peaks corresponding to $hf_q = \delta\Delta$ are also observed at $\Phi/\Phi_0 = -0.145$ and $\Phi/\Phi_0 = 0.855$, where this condition is also met.

APPENDIX C: MEASUREMENT OF Γ^0, Γ^1

In order to measure Γ^0 and Γ^1 , the parity-switching rates conditioned on the qubit being in $|0\rangle$ and $|1\rangle$, we would ideally measure Γ with the qubit spending nearly all of the

measurement time in $|0\rangle$ and $|1\rangle$, respectively. Measurement of Γ by fitting the PSD of a jump trace of parity does not allow this, because the individual parity measurements in the jump trace necessarily π pulse the qubit conditioned on the parity. Instead, we measure Γ by the decay of the parity autocorrelation function $\langle p(0)p(\tau) \rangle$. In this protocol, we separate two measurements of the parity by a variable delay τ . During this delay, we repeatedly measure the qubit state and use active feedback to control the mixture of Γ^0 and Γ^1 being measured. The qubit energy-relaxation time as a function of the flux varies from 20 to 70 μs . Since this is only somewhat longer than the time for qubit measurement (4 μs), the qubit state often changes between repeated measurements. Therefore, due to these jumps, the measurement of Γ while feeding back to $|0\rangle$ or $|1\rangle$ still results in a mixture of Γ^0 and Γ^1 . While the exact timing of such jumps is unknown, we estimate the fraction of the delay the qubit spends in $|0\rangle$ or $|1\rangle$ based on the qubit measurement record during τ . We repeat this measurement for different polarization angles to obtain values of Γ with different mixtures of Γ^0 and Γ^1 and we plot these values of Γ versus $\langle m_q \rangle$, the average qubit measurement during τ . Then, since $\Gamma \approx \langle m_q \rangle (\Gamma^1 - \Gamma^0) + \Gamma^0$, we extrapolate a linear fit of $\Gamma(\langle m_q \rangle)$ to obtain Γ^0 , Γ^1 .

To summarize, the measurement protocol at each flux point consists of the following:

- (1) Perform a Ramsey experiment to determine δf for the parity measurements p .
- (2) Measure p , delay τ (during which, repeatedly, the qubit is measured and prepared into $\psi(\theta)$ every 5.376 μs), and measure p . This is done 4 times, using all four combinations of the parity measurements (π_e, π_e) , (π_e, π_o) , (π_o, π_e) , and (π_o, π_o) , where π_e and π_o are π pulses conditioned on the even and odd parity, respectively.
- (3) Repeat step (2) 150 times.
- (4) Repeat steps (2) and (3) for each τ .
- (5) Repeat steps (2)–(4) for each θ .
- (6) Repeat steps (1)–(5) 50 times.

Steps (2)–(5) take approximately 30 s, during which n_g is typically stable. If it is found that n_g changes between successive Ramsey measurements, the data between them is not used since the parity measurements are unreliable. The full protocol takes approximately 1 h per flux point.

We now describe how θ is chosen and implemented in the pulse sequence. The extremal cases are simplest. In order to spend maximal time in $|0\rangle$, we prepare $\psi(\theta = 0)$. The feedback protocol is to π pulse the qubit if $m_q = 1$; otherwise, do not pulse. In order to spend maximal time in $|1\rangle$ ($\psi(\pi)$), the protocol is to π pulse the qubit if $m_q = 0$; otherwise, do not pulse. For angles in between, due to the nature of the feedback implementation with the field-programmable gate array, only one dynamical angle

can be used for rotation. As a result, the protocol requires an additional π pulse on either $|0\rangle$ or $|1\rangle$. For example, to prepare $\psi(\pi/4)$, the protocol could be: if $m_q = 0$, rotate by $\pi/4$; if $m_q = 1$, do a π pulse and then rotate by $\pi/4$. Alternatively, the extra π pulse could be enacted instead on $|0\rangle$ with $3\pi/4$ rotations to also prepare $\psi(\pi/4)$.

For the measurement in Fig. 2(d), eight polarization angles are used: $\theta = 0, \pi/6, \pi/3$, and $\pi/2$, with the extra π pulse on $|1\rangle$ [Fig. 9, triangles] and $\theta = \pi/2, 2\pi/3, 5\pi/6$, and π , with the extra π pulse on $|0\rangle$ [Fig. 9, squares]. Figure 9(a) shows the average qubit-state measurement $\langle m_q \rangle$ for all eight instances of θ as a function of the expected $\langle m_q \rangle = \sin^2(\theta/2)$. We observe that the measured $\langle m_q \rangle$ corresponding to angles in which the extra π pulse is enacted on $|0\rangle$ and those measured with the extra π pulse on $|1\rangle$ deviate from the expected value. We attribute this to two effects. First, energy relaxation tends to shift $\langle m_q \rangle$ toward the thermal value. Second, large detuning δf (up to approximately 14 MHz) between the pulse carrier frequency and the even- and odd-parity qubit frequencies affects the value of $\langle m_q \rangle$ when preparation of $\psi(\theta)$ is performed with the extra π

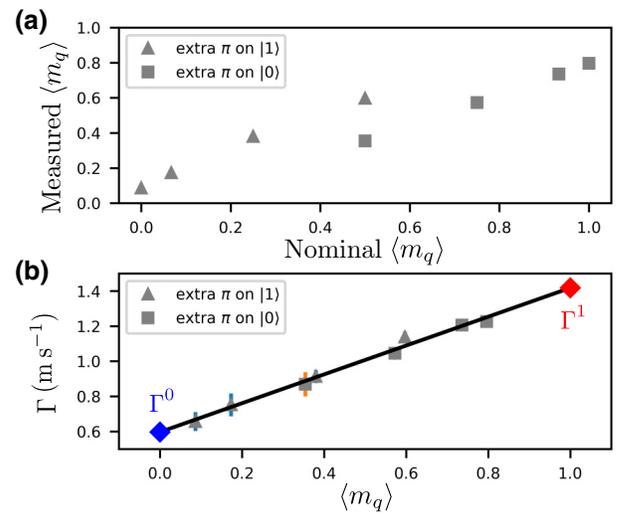


FIG. 9. (a) The average qubit-state measurement during the delay between parity measurements as a function of the nominal polarization angle θ ($\Phi/\Phi_0 = 0.335$). The observed $\langle m_q \rangle$ deviates from the nominal value $\langle m_q \rangle = \sin^2(\theta/2)$ due to energy relaxation as well as the large detuning between the carrier frequency of the qubit pulse and the even- and odd-parity qubit frequencies. (b) The charge-parity-switching rate Γ as a function of the average qubit-state measurement during feedback between parity measurements [Fig. 2(b)]. For each of the eight measurements of Γ (gray), a different state $\psi(\theta)$ is prepared after each qubit measurement m_q . The triangles represent measurements for which, during feedback, the qubit is π pulsed if it is in $|1\rangle$, followed by rotation to $\psi(\theta)$. The squares represent measurements for which, during feedback, the qubit is π pulsed if it is in $|0\rangle$, followed by rotation to $\psi(\theta)$. Both sets of measurements are fitted to a single line in order to determine Γ^0 (blue diamond) and Γ^1 (red diamond).

pulse. For example, $\delta f \approx 5$ MHz for the data depicted in Fig. 9 and thus the qubit state evolves significantly during the time between the π pulse and the nominal rotation to θ . Even with no intentionally added delay between these pulses, there is effectively time between the rotations due to the Gaussian shape of the pulses. Depending on the specific δf and the effective evolution time, the true prepared state ψ may differ significantly from the nominally expected value $\psi(\theta)$.

However, calculation of the functional dependence of $\langle m_q \rangle$ on δf and the pulse lengths is not necessary for this work. The measurements of Γ^0 and Γ^1 rely on measurement of Γ with different mixtures of $|0\rangle$ and $|1\rangle$ and not specific values of $\langle m_q \rangle$. Figure 9(b) shows that the parity-switching rate Γ can be fitted by a single line for feedback protocols with the extra π pulse on $|0\rangle$ or $|1\rangle$, since ultimately the value of Γ depends on the time spent in $|0\rangle$ and $|1\rangle$, which is proportional to $\langle m_q \rangle$. It is also checked that $\psi(0)$ and $\psi(2\pi)$ give the same values of Γ^0 and Γ^1 within measurement uncertainty, showing that the drive power itself does not influence the measured value of Γ .

APPENDIX D: MEASUREMENT SETUP AND DEVICE IMAGES

1. Device images

The offset-charge-sensitive SQUID transmon is comprised of electron-beam-evaporated aluminum on a sapphire substrate. The fabrication process is the same as that described in detail in the Supplemental Material of Refs. [19,32]. Images of a nominally identical device fabricated on the same wafer as the experimental device are shown in Fig. 10.

2. Measurement setup

The qubit is mounted in a 3D Cu microwave readout cavity and measured in reflection ($\kappa = 3.5$ MHz) as diagrammed in Fig. 11, with a superconducting nonlinear asymmetric inductive element (SNAIL) parametric amplifier providing initial amplification to achieve single-shot qubit-state readout. An Eccosorb filter is included inside the aluminum and magnetic shields, which has been demonstrated to reduce the parity-switching rate [32]. A 2-cm section of manganin wire (approximately 1.4Ω) is suspended in the aluminum can by superconducting leads and acts as an adjustable source of PAPS-inducing photons. A RuO₂ thermometer is mounted on the bracket holding the Cu readout cavity in order to monitor the temperature of the bracket as power is dissipated by the manganin wire. A dc voltage bias is added to the rf input line with a pair of bias tees, such that a dc voltage can be applied to the readout pin in the cavity in order to bias the offset charge. All lines enter the shields via a narrow

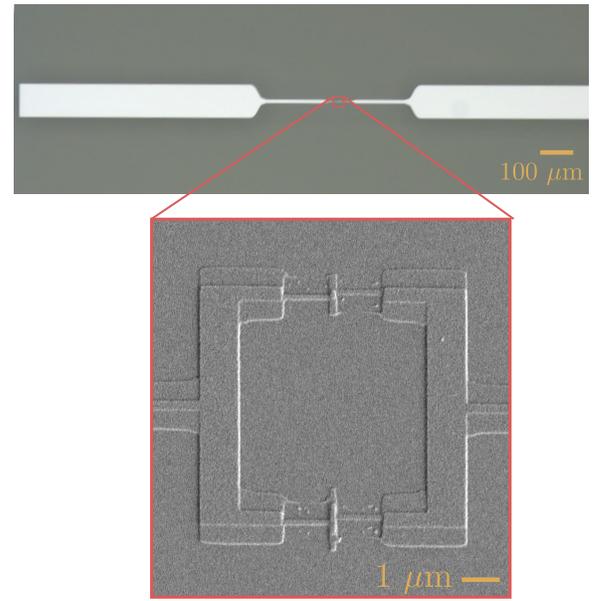


FIG. 10. Upper: an optical image of a qubit fabricated on the same wafer as the experimental device, which is nominally identical. The end of the right pad is not in view but is symmetric with the pad on the left. Lower: a scanning-electron-microscope image of the SQUID loop of the optically imaged device, in which the difference of the areas of the two Josephson junctions is apparent. The image contrast is enhanced using the GIMP image-manipulation software. The qubit is fabricated using the bridge-free technique in a liftoff process as described in Refs. [19,32].

slot opening and copper tape is used to make the slot as light-tight as possible.

APPENDIX E: FOUR-FILM MODEL FOR QP DYNAMICS

Here, we provide additional detail to Sec. IV A on the calculation of x_0 and x_3 , the QP densities of the JJ films that determine Γ_N . We consider the QP dynamics in all four films of the device in order to determine these values, taking into account generation of QPs by PAPS and tunneling across the JJ self-consistently with Γ_P and Γ_N , respectively. We first consider x_0 and x_1 , the QP densities of the films on the left side of the JJ in Fig. 3(a). These QP densities may change by the mechanisms of Eq. (2) (generation, trapping, and recombination) but also by tunneling between films of the device. QPs may tunnel between the low- and high-gap films 0 and 1 (t_{01}, t_{10}) as well as across the JJ between films 0 and 3 ($\gamma_{03}x_0, \gamma_{30}x_3$). The interfilm transport can be described by the coupled equations

$$\dot{x}_0 = g_0 - sx_0 - rx_0^2 - (t_{01} - t_{10}) - \gamma_{03}x_0 + \gamma_{30}x_3, \quad (\text{E1})$$

$$\dot{x}_1 = g_1 - sx_1 - rx_1^2 + (t_{01} - t_{10}). \quad (\text{E2})$$

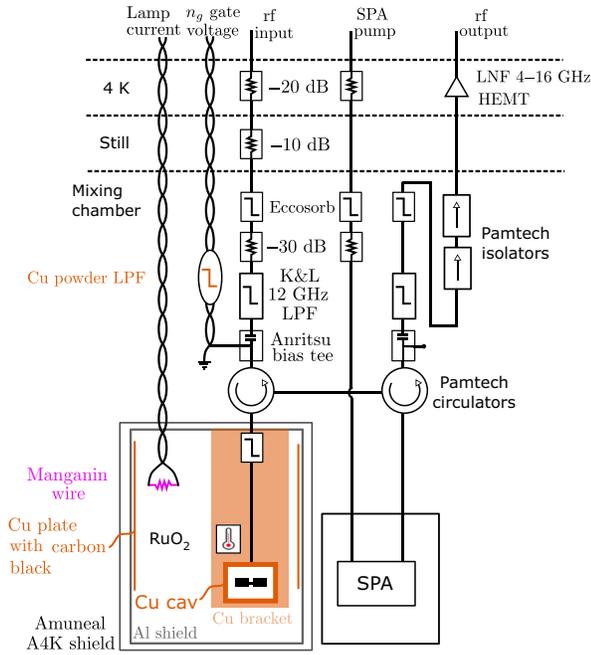


FIG. 11. The cryogenic measurement setup.

Below, we describe our approach to these interfilm tunneling terms.

While films 0 and 1 are separated by an AlO_x layer, they share a large contact area in the pads, so QPs tunnel rapidly between the films. In Sec. III B, we show evidence that QPs thermalize to distributions primarily located in the low-gap films on each side of the JJ. We therefore assume Fermi distributions for the QPs, $f(\varepsilon, T_{\text{ph}}, \mu_L) = 1/(e^{(\varepsilon - \mu_L)/k_B T_{\text{ph}}} + 1)$, with films 0 and 1 sharing the same nonzero chemical potential μ_L to describe an excess QP number independent of the temperature T_{ph} [27,50]. The QP densities in each film x_0 (x_1) are related to μ_L through the definition of $x_{\text{QP}} = 2 \int_0^\infty d\varepsilon v(\varepsilon, \Delta_{0(1)}) f(\varepsilon, T_{\text{ph}}, \mu_L)$. Due to the difference between the gaps, this thermalization results in $x_1 \approx x_0 e^{-\eta}$, where $\eta = \delta\Delta/k_B T_{\text{ph}}$. In our model, we therefore replace the tunneling rates between films 0 and 1 (t_{01} and t_{10}) with this assumption of rapid thermalization. We add Eqs. (E1) and (E2) and substitute $x_1 = x_0 e^{-\eta}$. The same approach is applied to films 2 and 3, which likewise share a large contact area in the opposite pad, reducing the number of independent QP densities in the system from four to two. This results in two coupled equations for x_0 and x_2 :

$$\dot{x}_0 = \frac{1}{1 + e^{-\eta}} [(g_0 + g_1) - (1 + e^{-\eta})sx_0 - (1 + e^{-2\eta})rx_0^2 - (\gamma_{03}x_0 - \gamma_{30}x_2 e^{-\eta})], \quad (\text{E3})$$

$$\dot{x}_2 = \frac{1}{1 + e^{-\eta}} [(g_2 + g_3) - (1 + e^{-\eta})sx_2 - (1 + e^{-2\eta})rx_2^2 + (\gamma_{03}x_0 - \gamma_{30}x_2 e^{-\eta})]. \quad (\text{E4})$$

As described in Sec. IV A, if the qubit was in thermal equilibrium with the QPs, the films on opposite sides of the JJ would share the same chemical potential. However, the qubit state is frequently pulsed during measurement of Γ and as a result, the tunnelings from films 0 to 3 and films 3 to 0 are not equal. The QP densities on the opposite side of the JJ (x_0 and x_2) differ in the steady state during the measurement due to this imbalance.

The rates γ_{03} and γ_{30} are the per-QP tunneling rates in the 0 to 3 and 3 to 0 directions, respectively. These may be calculated using only the S^{lr} or S^{rl} terms of $S_{\pm N}$ [Eq. (A7)] and assuming nonthermal qubit-state weights $\rho_0 = \rho_1 = 0.5$ caused by the measurement sequence. Because $\Gamma_{N,03}$ is approximately proportional to x_0 and independent of x_3 for small x_3 such as those existing in this experiment, the per-QP tunneling rate $\gamma_{03} := \Gamma_{N,03}/(x_0 N_{\text{CP},0})$ can be calculated. Similarly, $\gamma_{30} := \Gamma_{N,30}/(x_3 N_{\text{CP},2})$. The number of Cooper pairs in each film $N_{\text{CP},k} = 2D(\varepsilon_F)\Delta_k V_k$, with film number $k \in \{0, 1, 2, 3\}$, depends on the superconducting gap and volume. The single spin density of states at the Fermi energy $D(\varepsilon_F) = 0.72 \times 10^{29} \mu\text{m}^{-3} \text{J}^{-1}$ [59] and the volumes are approximately $100 \times 700 \times 0.03 \mu\text{m}^3$ for the low-gap films and $100 \times 700 \times 0.02 \mu\text{m}^3$ for the high-gap films.

These per-QP tunneling rates depend on T_{ph} and $\delta\Delta$ and have significant dependence on the flux. The imbalance $\Gamma_{N,03}/\Gamma_{N,30} > 1$ is largest when $hf_q = \delta\Delta$, since QPs at the low-gap edge in film 0 can absorb an excitation from the qubit to tunnel to the high-gap edge in film 3. Based on our fit, this results in 55% more QPs in films 2 and 3 as compared to films 0 and 1 at the flux for which $hf_q = \delta\Delta$.

We separate QP generation into generation by PAPS (g_P) and generation by other mechanisms (g_{other}), which break Cooper pairs but do not simultaneously result in a parity switch. While PAPS generates QPs only in JJ films 0 and 3, mechanisms contributing to g_{other} could generate QPs in all the films. Therefore, we have $g_0 + g_1 = (g_P + g_{\text{other},0}) + g_{\text{other},1} = g_P + g_{\text{other}}$, where we define $g_{\text{other}} := g_{\text{other},0} + g_{\text{other},1}$ as the total generation on one side of the JJ. Likewise, $g_2 + g_3 = g_{\text{other},2} + (g_P + g_{\text{other},3}) = g_P + g_{\text{other}}$, where we make the reasonable assumption that non-PAPS pair breaking occurs identically on opposite sides of the JJ, since the two pads have the same geometries. Generation by PAPS $g_P := \Gamma_P/N_{\text{CP},0}$ is calculated self-consistently with Γ_P and g_{other} is a flux-independent fitting parameter in our model.

In Sec. IV A, the additional approximation is made that since $e^{-\eta} \approx 0.008$ for $T_{\text{ph}} \approx 50$ mK and $\delta\Delta/k_B \approx 233$ mK, the high-gap films can essentially be removed from the determination of x_0 and x_2 . Equations (3) and (4) are obtained by substituting the above expressions for generation into Eqs. (E3) and (E4) and making the approximation $e^{-\eta} = 0$.

APPENDIX F: EFFECT OF LAMP

1. Dependence of Γ on P_{lamp}

We observe empirically that Γ increases approximately linearly with P_{lamp} , the power dissipated by the manganin resistor ‘‘lamp.’’ The emission spectrum of the lamp depends on the temperature of the lamp, which is determined by a balance of the dissipated heat and the thermal conductivity to the nearest cold heat sink (likely the mixing chamber). We expect dissipation to be uniform in the manganin filament of the lamp and independent of temperature for the range of our experiment. On the other hand, the thermal conductivity of the leads which determines the temperature of the lamp filament is itself temperature dependent and may vary along the length of the leads.

As described in Sec. VA, the lamp is required to compare QP generation that occurs with parity switching due to PAPS (g_P) and QP generation that does not occur with parity switching (g_{other}). For the purpose of constraining g_{other} , measurement at any increased Γ without increased device temperature is sufficient, since a second set of $\Gamma_P(\Phi)$ and $\Gamma_N(\Phi)$ constrains the model and fixes the strength of the trapping rate.

Nonetheless, we attempt to understand the linear dependence of Γ on the power dissipated by the lamp P_{lamp} with a simple model that takes into account the temperature dependence of the thermal conductivity of the (Nb-Ti)-Cu leads. We assume that the heat flow in the wire is constant in the steady state but the thermal conductivity of the Cu in the leads depends on the position x from the mixing chamber due to the temperature gradient [Fig. 12(a)]:

$$\kappa(x) \frac{\partial T}{\partial x} = P_{\text{lamp}}. \quad (\text{F1})$$

The temperature of the lamp T_{lamp} is the temperature at $T(x=l)$, where l is the length from the mixing chamber where the (Nb-Ti)-Cu leads are thermalized to the lamp. The thermal conductance per unit length $\kappa(x)$ depends on temperature according to the Weidemann-Franz law, $\kappa(x)\partial x = c_\kappa T(x)$. The proportionality factor c_κ is related to the electrical resistivity ρ and the Lorenz number L , $c_\kappa = L/\rho$ and we treat it as a fitting parameter here. Substituting for κ and integrating, we find that $T_{\text{lamp}} \propto P_{\text{lamp}}^{1/2}$:

$$T_{\text{lamp}} = \sqrt{P_{\text{lamp}} \frac{2l}{c_\kappa} + T_{\text{MC}}^2}. \quad (\text{F2})$$

The mixing-chamber temperature is $T_{\text{MC}} \approx 30$ mK.

Next, we propose that Γ_P is proportional to the integrated power P_{int} radiated from the lamp that enters the cavity and causes PAPS. Treating the lamp as a 3D black body, we find that integrating the radiated power from 100 to 300 GHz yields a dependence on the temperature

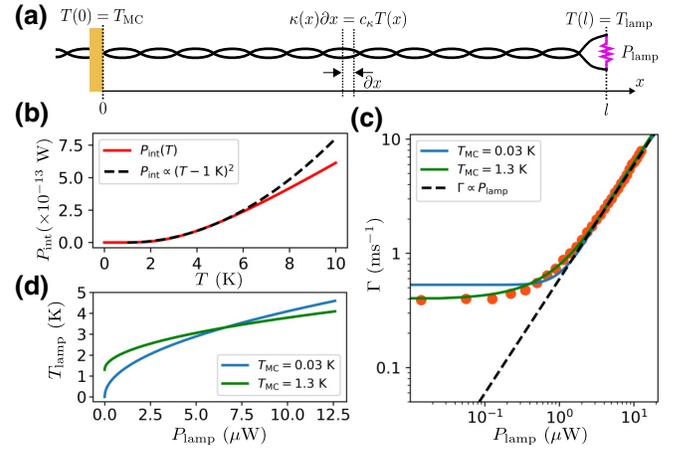


FIG. 12. (a) A diagram of the model for the temperature of the lamp T_{lamp} as a function of the power dissipated by the lamp, P_{lamp} . The thermal conductivity per unit length of the leads is assumed to be linearly proportional to the temperature, which varies as a function of position x along the leads. The temperature at the mixing chamber, $T(0) = T_{\text{MC}}$, is approximately 30 mK. (b) The power radiated by a 3D black body at temperature T integrated from 100 to 300 GHz. We observe that from 1 K to 5 K, the dependence is approximately quadratic in temperature with an offset of 1 K. At higher temperatures, $P_{\text{int}} \propto T$. (c) The measured Γ as a function of the power dissipated by the lamp (orange). The data are the same as in Fig. 4(b) but here shown in log-log scale. We fit them with the model in which $\Gamma \propto P_{\text{int}}(T_{\text{lamp}})$, where T_{lamp} depends on P_{lamp} according to Eq. (F2). In blue, we fit assuming that the lamp is at the mixing-chamber temperature when no power is dissipated ($T_{\text{MC}} = 0.03$ K). The model captures the linear behavior between approximately 1 and 10 μW but not the low-power dependence of Γ . In green, in an attempt to fit $\Gamma(P_{\text{lamp}} < 1 \mu\text{W})$, we allow T_{lamp} at $P_{\text{lamp}} = 0 \mu\text{W}$ to vary as a fitting parameter, which yields $T_{\text{MC}} = 1.3$ K. This is an unreasonably high temperature, indicating that this model does not accurately capture the low-power behavior. (d) The temperature of the lamp T_{lamp} as a function of P_{lamp} , calculated using the fitting parameters of (b). In blue, $T_{\text{lamp}}(0)$ is assumed to be the mixing-chamber temperature, while in green, $T_{\text{lamp}}(0) = 1.3$ K. In both cases, $T_{\text{lamp}} \propto P_{\text{lamp}}^{1/2}$.

of the lamp (T_{lamp}) that is approximately quadratic for temperatures of approximately 1–5 K [Fig. 12(b)]. The upper bound of this frequency range of photons that induce PAPS is determined by the constraint that the size of the qubit must be smaller than the wavelength of the photon. Higher-frequency photons with smaller wavelengths are increasingly likely to be absorbed in the pads of the device due to the larger absorption area.

In Fig. 12(c), the measured Γ is shown as a function of P_{lamp} . We fit the data with a simple model in which the lamp radiates the spectral density of a 3D black body at T_{lamp} and Γ is proportional to P_{int} , the integrated power between 100 and 300 GHz radiated by the lamp: $\Gamma = AP_{\text{int}}(T_{\text{lamp}}) + B$. In this model, A is an unknown

proportionality constant between Γ and P_{int} that includes attenuation and coupling of radiation to the qubit, which is assumed to be frequency independent for simplicity. This scale factor also includes the additional Γ resulting from increased Γ_N due to the QPs generated by the increased Γ_P (when $g_P \gg g_{\text{other}}$, Γ_N is approximately proportional to Γ_P). The lamp-power-independent offset B accounts for background PAPS from other sources and NUPS from QPs existing at $P_{\text{lamp}} = 0 \mu\text{W}$. Combining $T_{\text{lamp}} \propto P_{\text{lamp}}^{1/2}$ with $P_{\text{int}} \propto T_{\text{lamp}}^2$, we expect a linear dependence of Γ_P on P_{lamp} for a range corresponding to $T_{\text{lamp}} \approx 1\text{--}5 \text{ K}$. In the range of $P_{\text{lamp}} \approx 1\text{--}12.6 \mu\text{W}$, the data show this linear dependence.

The blue line shows a fit with $T_{\text{MC}} = 0.03 \text{ K}$, which models the linear part of the data reasonably well but does not capture the low-power behavior. The fitting parameter c_κ corresponds to an electrical resistivity of $\rho = 2.9 \times 10^{-9} \Omega\text{m}$. Shown on log-log scale, it is clear that $\Gamma(P_{\text{lamp}})$ is not linear for $P_{\text{lamp}} \lesssim 1 \mu\text{W}$. The fit to this range can be improved by allowing T_{MC} to vary (green) but yields $T_{\text{MC}} = 1.3 \text{ K}$, which is 40 times hotter than the mixing chamber. Thus, while this model does capture the observed linear behavior for $P_{\text{lamp}} \approx 1\text{--}12.6 \mu\text{W}$, the low- P_{lamp} behavior is not described by this simple model. This is reasonable given that the model assumes $T_{\text{lamp}} \gg T_{\text{MC}}$ and $g_P \gg g_{\text{other}}$.

2. Effect of lamp on temperature

To confirm that the manganin lamp is increasing Γ by PAPS rather than by increasing the temperature of the device and generating additional thermal QPs, we mount a RuO_2 thermometer on the bracket holding the copper cavity. The temperature measured by this thermometer is plotted for each measurement of Γ in Fig. 13(a). The difference between the temperature measured by this thermometer and the mixing-chamber temperature $T_{\text{MC}} \approx 30 \text{ mK}$ is attributed to the thermal resistance from the end of the bracket to the mezzanine to the mixing chamber. We observe that while the temperature increases by several millikelvins as the power dissipated by the lamp increases, this increase is far too small to generate a significant number of thermal QPs.

This is verified by performing a separate sweep in which we increase the temperature by means of a heater mounted on the mixing-chamber plate [Fig. 13(b), red]. We can see that when the bracket temperature measured by the RuO_2 thermometer increases due to the remote heater by similarly small amounts, there is minimal change in Γ . Only when the bracket RuO_2 thermometer reaches approximately 150 mK does the increase in Γ match what is produced by the $12.6 \mu\text{W}$ dissipated by the lamp. Figure 13(b) shows that the increase in Γ due to power dissipated by the lamp cannot be explained by an increased temperature of the sample.

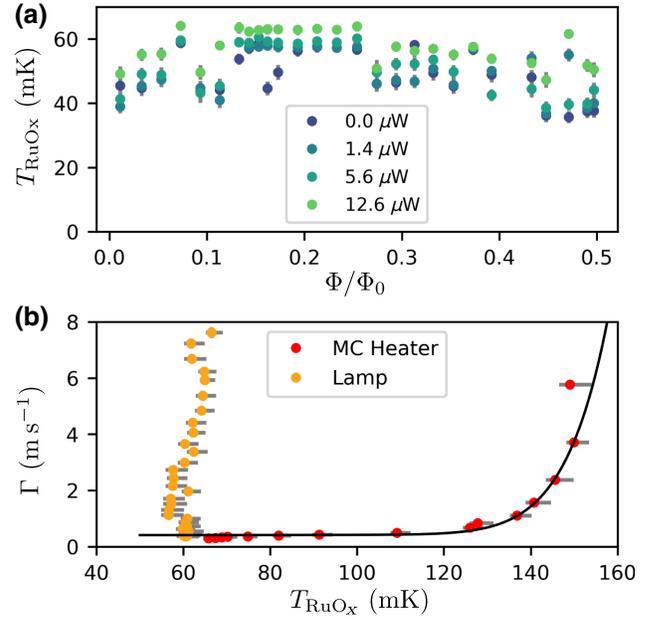


FIG. 13. (a) The temperature measured by the RuO_2 thermometer for the data of Fig. 4(c). The colors correspond to different lamp powers. The thermometer is mounted on the copper bracket supporting the 3D readout cavity in which the qubit is embedded. The temperature remains well below the temperature at which thermal QPs are generated for all values of Φ/Φ_0 and P_{lamp} . (b) The measured parity-switching rate Γ as a function of the temperature measured by the RuO_2 thermometer near the readout cavity. Two sweeps are shown. In yellow, the lamp power is swept [Fig. 13(b)] and here we observe that the bracket temperature does not change significantly as Γ increases by this mechanism. In red, the mixing-chamber temperature is swept by increasing the power dissipated by heaters on the mixing-chamber plate. The temperature measured by the RuO_2 thermometer increases as the mixing-chamber temperature rises but Γ does not increase dramatically until approximately 125 mK. This shows that the slight temperature increases shown in (a) cannot be responsible for the large increase in Γ . The black line shows a fit to Γ versus T , from which we may extract the average gap $\bar{\Delta}/h \approx 51.8 \text{ GHz}$.

This dependence of Γ on the mixing-chamber temperature is used to estimate the average gap of the two superconducting films $\bar{\Delta} = (\Delta_L + \Delta_H)/2$, since thermal activation of QPs will be sensitive to this value. The black solid line shows a fit to a model in which there is temperature-independent PAPS and temperature-dependent NUPS due to the QP energy distribution changing and thermal QPs being generated at higher temperature. The only fitting parameters used in this model are Γ_P and $\bar{\Delta}$. The fitted $\bar{\Delta}$ is not sensitive to the assumed values for trapping and recombination rates, which are fixed parameters in the model. In this case, we set $g_{\text{other}} = 0$, since the data cannot distinguish between a temperature-independent PAPS rate versus a temperature-independent background of excess

QPs. Repeating the fit with g_{other} as a fitting parameter and $\Gamma_P = 0$ yields the same Δ . In each case, we find $\Delta/h = 51.8$ GHz, which is consistent with reported superconducting-gap measurements of thin-film aluminum [43,45].

APPENDIX G: FIT SENSITIVITY TO TRAPPING RATE s

Figure 14 demonstrates the sensitivity of the fit shown in Fig. 4(c) to the trapping rate s . In Fig. 14(a), we plot the pseudo- R^2 goodness-of-fit metric for fits to the full data set fixing s (blue, left axis). We use the definition $R^2 = \sum_i [1 - (S_{\text{res},i}/S_{\text{tot},i})]/4$, $i \in \{0, 1, 2, 3\}$, where $S_{\text{res},i}$ and $S_{\text{tot},i}$ are the sum-of-squares difference between the data and the model ($S_{\text{res},i}$) and between the data and mean ($S_{\text{tot},i}$) measured at the i th lamp power.

We observe a maximum of R^2 near $s = 11$ s $^{-1}$ corresponding to the best fit. For $s \lesssim 6$ s $^{-1}$, generation by PAPS is sufficient to generate the approximate x_{QP} in the device, such that the contributions of g_{other} found by the fits in this range are negligible. For $s \gtrsim 6$ s $^{-1}$, the best fit g_{other} increases linearly with s in order to keep the x_{QP} at the value that best fits the data for $P_{\text{lamp}} = 0$ μW .

Figure 14(b) illustrates how values of s differing from the best-fit value affect the model. We observe that for a lower value of $s = 2$ s $^{-1}$, the additional QPs generated by the enhanced Γ_P with $P_{\text{lamp}} = 12.6$ μW cause a level of Γ_N that gives a peak at $\Phi/\Phi_0 \approx 0.145$ that is too large compared to the data. For a higher value of $s = 100$ s $^{-1}$, the strong trapping rate suppresses the additional QPs generated by the enhanced Γ_P with $P_{\text{lamp}} = 12.6$ μW , causing a level of Γ_N that gives a peak at $\Phi/\Phi_0 \approx 0.145$, which is too small compared to the data.

APPENDIX H: DIRECT OBSERVATION OF QUASIPARTICLE BURSTS

One possible source of pair-breaking energy for QP generation (i.e., a contribution to g_{other}) is ionizing radiation. The energy cascade from the ionization of atoms in the superconducting films or substrate, described in Refs. [33,37,38,60], ultimately results in “bursts” of QPs. Experimentally, evidence for these bursts has thus far come in the form of sudden drops in T_1 in superconducting qubits or drops in the quality factor of superconducting resonators. In this device, we find evidence of rapid parity switching, which further substantiates the hypothesis that these sudden drops in T_1 are due to QPs. By tuning the flux of our SQUID offset-charge-sensitive transmon such that $E_J/E_C \lesssim 20$, we are able to measure the parity directly via the parity-dependent dispersive shift as described in Ref. [32]. With a quantum-limited SNAIL parametric amplifier [61], we are able to discern the parity with a single 4- μs measurement of the readout cavity.

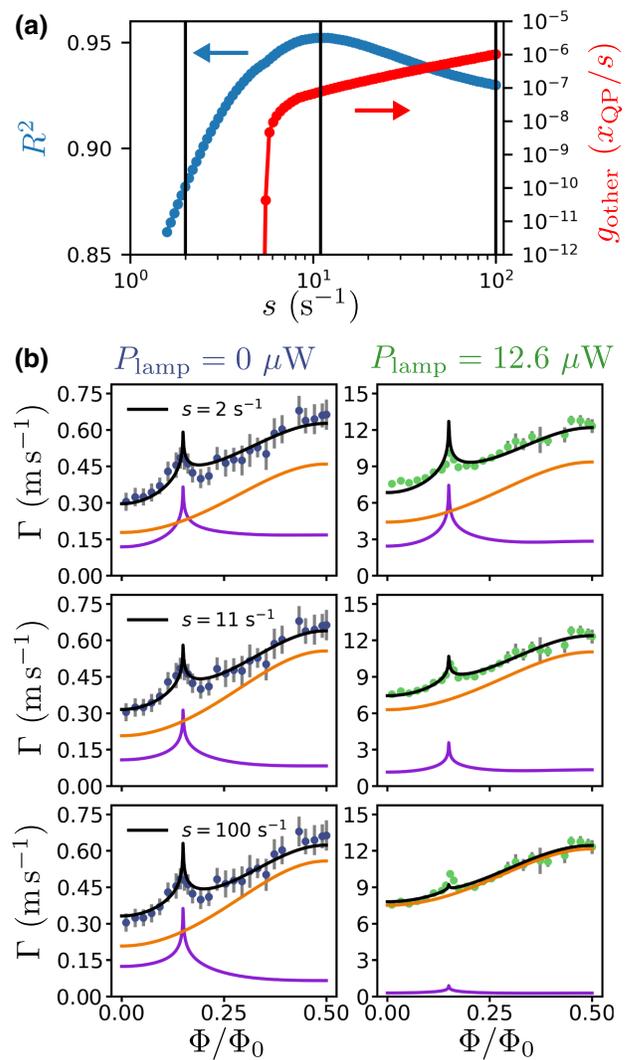


FIG. 14. (a) The pseudo- R^2 goodness-of-fit metric for fits to the full data set as a function of the trapping rate s (blue, left axis). The value of g_{other} obtained in each fit is also plotted for reference (red, right axis). The vertical black lines indicate values of s for which the fits are shown in (b). (b) Fits to the $P_{\text{lamp}} = 0$ μW (left) and $P_{\text{lamp}} = 12.6$ μW (right) data for the values of s marked by vertical lines in (a). The fits corresponding to lower values of R^2 (i.e., $s = 2$ s $^{-1}$ and $s = 100$ s $^{-1}$) appear to describe the intermediate peak at $\Phi/\Phi_0 = 0.145$ less well for the $P_{\text{lamp}} = 12.6$ μW as compared to the best fit ($s = 11$ s $^{-1}$).

In addition to generating high-energy phonons that break Cooper pairs and generate QPs, ionizing-radiation impacts can also cause n_g jumps due to the redistribution of charge in the substrate. With this direct detection technique, which tracks parity switching and charge jumps simultaneously, we are able to quantify how many of the observed burst events are correlated with an n_g jump. In Fig. 15, we show two examples of bursts: one in which the n_g is unchanged [Figs. 15(a)–(d)] and one in which n_g changes visibly before and after the burst [Figs. 15(e)–(h)].

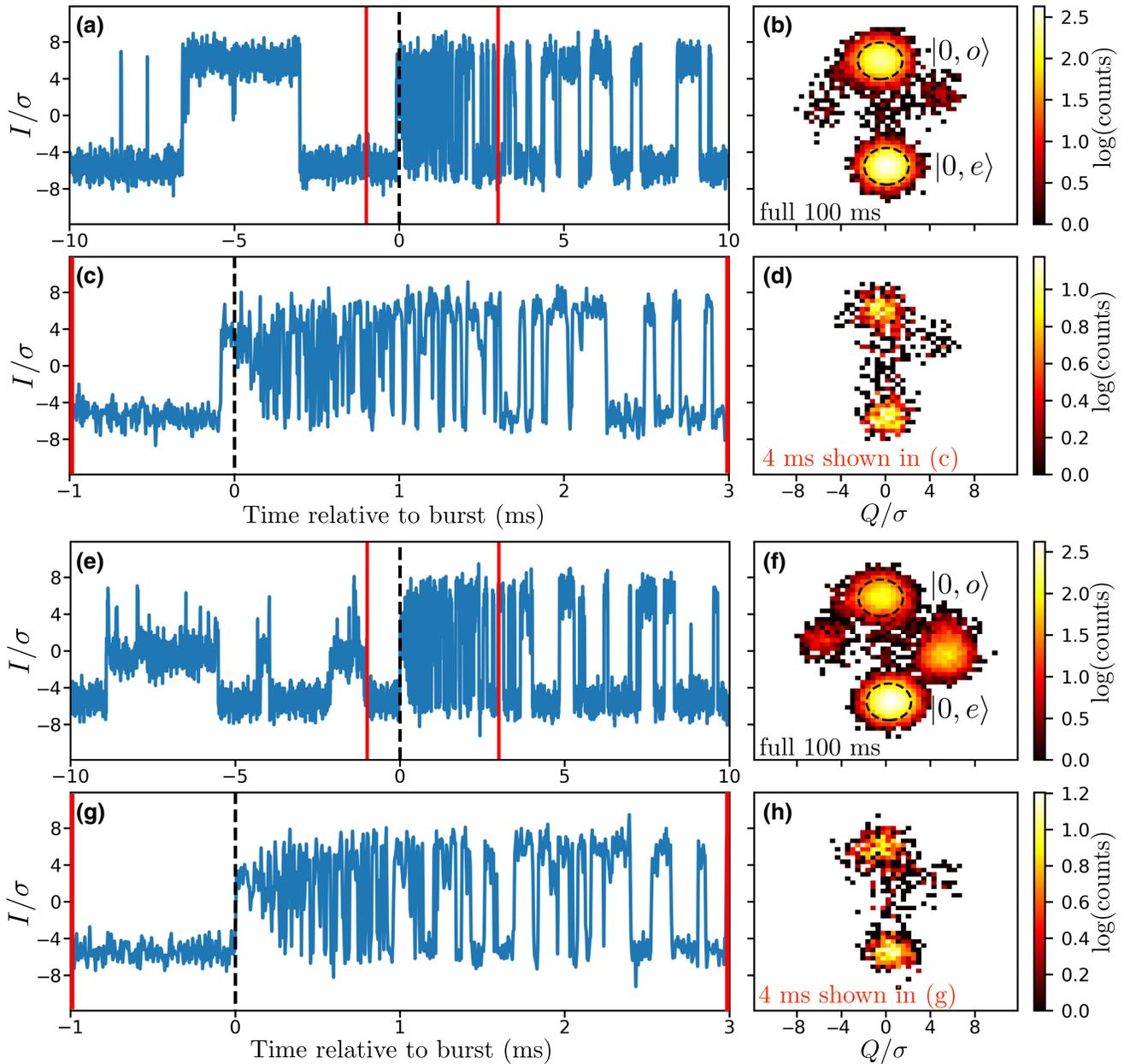


FIG. 15. (a) The time series of the I quadrature of qubit-state measurements repeated every $5 \mu\text{s}$, in units of the σ of the Gaussian distributions of the ground-state measurement histograms. Rapid switching between parity states begins at $t = 0$, which we attribute to a “burst” of QPs generated by ionizing radiation. (b) The 2D histogram of qubit-state measurements in the complex plane for the full 100-s time series from which the data in (a) are drawn, shown for state identification. Most of the counts form two Gaussian distributions with positive Q , which we attribute to the even- and odd-parity ground states, respectively. The counts at negative Q correspond to excited states of the qubit. The data are plotted in log scale for visibility of excited states and the dashed lines indicate 2σ . (c) An enlargement of (a), showing 4 ms around the onset of the burst. The value jumps to a value between $|0, e\rangle$ and $|0, o\rangle$ due to switching that is fast compared to the $1/(5 \mu\text{s})$ measurement repetition rate. This is followed by rapid switching between the parity states, which decays in frequency over time. (d) The same as (b), but for only the data in the window between the red lines shown in (c). The data are primarily in the ground states, showing that the data near $I \approx 0$ at $t = 0$ are due to rapid switching rather than a jump to an excited state. Panels (e) and (f) are analogous to (a)–(d), but show a burst correlated with a jump in n_g due to the redistribution of charge in the substrate resulting from the ionizing impact. At $t < 0$, the $|0, o\rangle$ state is located at $I/\sigma \approx 0$, $Q/\sigma \approx 6$ and parity switches occur between the even- and odd-parity ground states in this configuration. When the burst occurs at $t = 0$, the $|0, o\rangle$ state jumps to the same configuration as in (a)–(d), indicating a jump in n_g correlated with a burst.

In Fig. 15(a), we show a 20-ms-long jump trace of qubit measurements, with the I quadrature of the complex-valued signal plotted. Figure 15(b) shows the 2D histogram in the complex plane of the full 100-ms time series from which the data in Fig. 15(a) are drawn, which consists of 2×10^5 qubit measurements. The readout-pulse length and integration time are $4 \mu\text{s}$, with a $1\text{-}\mu\text{s}$ delay between measurements. We can see the measurements primarily form two Gaussian distributions near $Q = 0$, which we interpret as the even- and odd-parity ground states (even and odd are assigned arbitrarily). The remaining population corresponds to higher-energy states. At $t < 0$, we can see from the value of I that the qubit initially has even parity and then switches to odd parity at $t \approx -7$ ms and back to even at $t = -3$ ms. The background parity-switching rate at this flux value is $\Gamma \approx 600 \text{ s}^{-1} \approx 1/1.7 \text{ ms}$ during this cool down.

Figure 15(c) is an enlargement of a 4-ms section around a burst event, which we label as occurring at $t = 0$. We observe that at $t = 0$, the value of I jumps to a value between the values corresponding to $|0, e\rangle$ and $|0, o\rangle$. This is likely due to the parity switching much faster than the $5 \mu\text{s}$ repetition time of our measurement. We confirm that this is not a result of qubit excitation by checking the 2D histogram of points during this 4 ms segment, in which it is clear that smearing between $|0, e\rangle$ and $|0, o\rangle$ is much more prevalent than qubit excitation [Fig. 15(d)]. Following this jump, the qubit state switches between the $|0, e\rangle$ and $|0, o\rangle$ states much faster than the 1.7-ms lifetime characteristic of the majority of the jump trace. The parity-switching rate appears to decay significantly by $t = 10$ ms, although it is not clear that it has returned to the consistent background rate.

Figures 15(e)–15(h) are analogous to Figs. 15(a)–15(d), except that at $t = 0$, there is a sudden change in the dispersive shifts of the qubit states, signaling a jump in n_g in addition to the onset of rapid parity switching. In this data set, we search for bursts when the configuration of n_g -dependent dispersive shifts provides maximal distinguishability of the even- and odd-parity states [the configuration of Figs. 15(a)–15(d)]. In the event shown in Figs. 15(e)–15(h), the dispersive shift of the $|0, o\rangle$ state changes at $t = 0$, changing from $I/\sigma \approx 0$ to the optimal configuration in which $I/\sigma \approx 6$. Impacts that cause n_g jumps out of this optimal configuration into others are not detected; however, we note that they occur at the same rate as jumps into the optimal configuration.

We observe approximately 209 events in approximately 5.6 h of data. Of the 209 impacts, we find that 60 are correlated with an n_g jump. Additional experiments are required to interpret this correlation fraction and how it relates to factors such as the deposited energy and location of impacts. Adding in an equal number of undetected jumps out of the optimal configuration, we estimate an event rate of approximately $1/75 \text{ s}$. This is consistent with

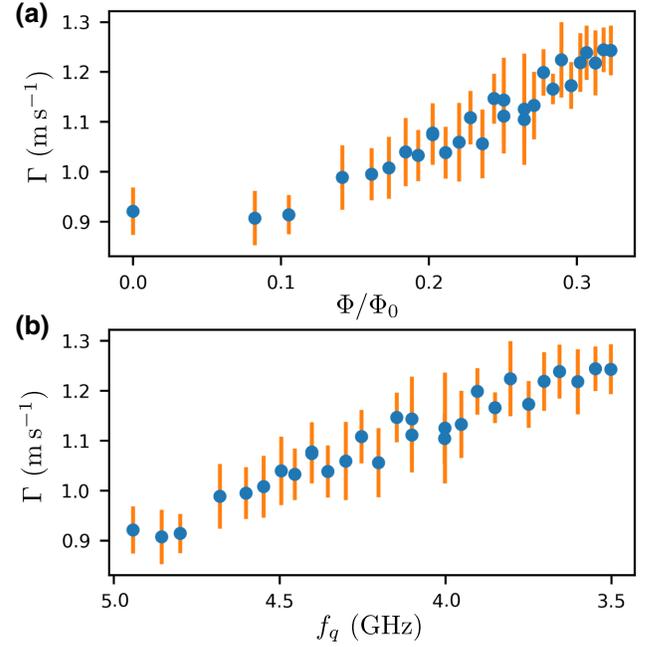


FIG. 16. (a) The parity-switching rate as a function of the flux for a device with 10 nm/100 nm thick electrodes. (b) The parity-switching rate as a function of the qubit frequency. We do not observe a sharp peak associated with the gap difference matching the qubit energy in this frequency range.

impact rates reported in Refs. [37,60,62] after taking into account the area of the substrate (our sapphire substrate is $3 \text{ mm} \times 15 \text{ mm}$). The extent to which bursts contribute to the background excess QP density and parity-switching rate depends on the energy deposited by the ionizing radiation, the frequency of the events, and the time scale on which the density decays. Further work is necessary to estimate the energy deposited by these events and to understand the decay of the QP density after a burst, which may involve several time scales [38,60].

APPENDIX I: FLUX DEPENDENCE IN DEVICE WITH LARGER GAP DIFFERENCE

As a consistency check, we measure the parity-switching rate in a device designed to have a large gap difference. The bottom and top electrodes are 10 nm and 100 nm thick, respectively, which we predict yields $\delta\Delta/h \approx 10 - 15 \text{ GHz}$ based on reported gap measurements [43,45]. Therefore, we should not observe a peak associated with the gap difference in the measured range of $5 \text{ GHz} \gtrsim f_q \gtrsim 3.5 \text{ GHz}$. In Fig. 16, the parity-switching rate as a function of flux $\Gamma(\Phi)$ is shown for this device. We observe no peak associated with the gap difference in this range. In this device, the JJs are more symmetric than those of the device discussed in the main text, such that the flux corresponding to $f_q \approx 3.5 \text{ GHz}$ is $\Phi/\Phi_0 \approx 0.33$.

Due to the suppression of NUPS by the gap difference as discussed in Sec. III, we predict the parity-switching rate to be reduced in this device relative to the 20 nm/30 nm thickness device presented in the main text. However, here we observe that $\Gamma(\Phi/\Phi_0 = 0) \approx 900 \text{ s}^{-1}$, a factor of approximately 3 higher than that measured for the 20 nm/30 nm device. While further work is required to determine the cause of this higher parity-switching rate, we believe that it is due to an increased level of PAPS. We find that the measured parity-switching rate for a single device fluctuates from cool down to cool down, which we attribute to variation in the light-tightness of the sample cavity and shields for each particular cool down. In order to isolate the effect of the gap difference, devices with different gap differences must be measured in the same package during a single cool down.

-
- [1] P. K. Day, H. G. LeDuc, B. A. Mazin, A. Vayonakis, and J. Zmuidzinas, A broadband superconducting detector suitable for use in large arrays, *Nature* **425**, 817 (2003).
- [2] M. T. Tuominen, J. M. Hergenrother, T. S. Tighe, and M. Tinkham, Experimental Evidence for Parity-Based $2e$ Periodicity in a Superconducting Single-Electron Tunneling Transistor, *Phys. Rev. Lett.* **69**, 1997 (1992).
- [3] J. Aumentado, M. W. Keller, J. M. Martinis, and M. H. Devoret, Nonequilibrium Quasiparticles and $2e$ Periodicity in Single-Cooper-Pair Transistors, *Phys. Rev. Lett.* **92**, 066802 (2004).
- [4] A. J. Ferguson, N. A. Court, F. E. Hudson, and R. G. Clark, Microsecond Resolution of Quasiparticle Tunneling in the Single-Cooper-Pair Transistor, *Phys. Rev. Lett.* **97**, 106603 (2006).
- [5] V. Bouchiat, D. Vion, P. Joyez, D. Esteve, and M. H. Devoret, Quantum coherence with a single Cooper pair, *Phys. Scr.* **T76**, 165 (1998).
- [6] Y. Nakamura, Y. A. Pashkin, and J. S. Tsai, Coherent control of macroscopic quantum states in a single-Cooper-pair box, *Nature* **398**, 786 (1999).
- [7] J. Männik and J. E. Lukens, Effect of Measurement on the Periodicity of the Coulomb Staircase of a Superconducting Box, *Phys. Rev. Lett.* **92**, 057004 (2004).
- [8] A. Guillaume, J. F. Schneiderman, P. Delsing, H. M. Bozler, and P. M. Echternach, Free evolution of superposition states in a single Cooper pair box, *Phys. Rev. B* **69**, 132504 (2004).
- [9] R. Lutchyn, L. Glazman, and A. Larkin, Quasiparticle decay rate of Josephson charge qubit oscillations, *Phys. Rev. B* **72**, 014517 (2005).
- [10] R. M. Lutchyn, L. I. Glazman, and A. I. Larkin, Kinetics of the superconducting charge qubit in the presence of a quasiparticle, *Phys. Rev. B* **74**, 064515 (2006).
- [11] J. M. Martinis, M. Ansmann, and J. Aumentado, Energy Decay in Superconducting Josephson-Junction Qubits from Nonequilibrium Quasiparticle Excitations, *Phys. Rev. Lett.* **103**, 097002 (2009).
- [12] M. Lenander, H. Wang, R. C. Bialczak, E. Lucero, M. Mariantoni, M. Neeley, A. D. O’Connell, D. Sank, M. Weides, J. Wenner, T. Yamamoto, Y. Yin, J. Zhao, A. N. Cleland, and J. M. Martinis, Measurement of energy decay in superconducting qubits from nonequilibrium quasiparticles, *Phys. Rev. B* **84**, 024501 (2011).
- [13] G. Catelani, R. J. Schoelkopf, M. H. Devoret, and L. I. Glazman, Relaxation and frequency shifts induced by quasiparticles in superconducting qubits, *Phys. Rev. B* **84**, 064517 (2011).
- [14] G. Catelani, S. E. Nigg, S. M. Girvin, R. J. Schoelkopf, and L. I. Glazman, Decoherence of superconducting qubits caused by quasiparticle tunneling, *Phys. Rev. B* **86**, 184514 (2012).
- [15] D. Ristè, C. C. Bultink, M. J. Tiggelman, R. N. Schouten, K. W. Lehnert, and L. DiCarlo, Millisecond charge-parity fluctuations and induced decoherence in a superconducting transmon qubit, *Nat. Commun.* **4**, 1913 (2013).
- [16] J. Wenner, Y. Yin, E. Lucero, R. Barends, Y. Chen, B. Chiaro, J. Kelly, M. Lenander, M. Mariantoni, A. Megrant, C. Neill, P. J. J. O’Malley, D. Sank, A. Vainsencher, H. Wang, T. C. White, A. N. Cleland, and J. M. Martinis, Excitation of Superconducting Qubits from Hot Nonequilibrium Quasiparticles, *Phys. Rev. Lett.* **110**, 150502 (2013).
- [17] G. Catelani, Parity switching and decoherence by quasiparticles in single-junction transmons, *Phys. Rev. B* **89**, 094522 (2014).
- [18] I. M. Pop, K. Geerlings, G. Catelani, R. J. Schoelkopf, L. I. Glazman, and M. H. Devoret, Coherent suppression of electromagnetic dissipation due to superconducting quasiparticles, *Nature* **508**, 369 (2014).
- [19] K. Serniak, M. Hays, G. de Lange, S. Diamond, S. Shankar, L. Burkhardt, L. Frunzio, M. Houzet, and M. Devoret, Hot Nonequilibrium Quasiparticles in Transmon Qubits, *Phys. Rev. Lett.* **121**, 157701 (2018).
- [20] U. Vool, I. Pop, K. Sliwa, B. Abdo, C. Wang, T. Brecht, Y. Gao, S. Shankar, M. Hatridge, G. Catelani, M. Mirrahimi, L. Frunzio, R. Schoelkopf, L. Glazman, and M. Devoret, Non-Poissonian Quantum Jumps of a Fluxonium Qubit due to Quasiparticle Excitations, *Phys. Rev. Lett.* **113**, 247001 (2014).
- [21] A. Zazunov, V. S. Shumeiko, E. N. Bratus’, J. Lantz, and G. Wendin, Andreev Level Qubit, *Phys. Rev. Lett.* **90**, 087003 (2003).
- [22] M. Hays, G. de Lange, K. Serniak, D. van Woerkom, D. Bouman, P. Krogstrup, J. Nygård, A. Geresdi, and M. Devoret, Direct Microwave Measurement of Andreev-Bound-State Dynamics in a Semiconductor-Nanowire Josephson Junction, *Phys. Rev. Lett.* **121**, 047001 (2018).
- [23] M. Hays, V. Fatemi, D. Bouman, J. Cerrillo, S. Diamond, K. Serniak, T. Connolly, P. Krogstrup, J. Nygård, A. Levy Yeyati, A. Geresdi, and M. H. Devoret, Coherent manipulation of an Andreev spin qubit, *Science* **373**, 430 (2021).
- [24] R. M. Lutchyn, J. D. Sau, and S. Das Sarma, Majorana Fermions and a Topological Phase Transition in Semiconductor-Superconductor Heterostructures, *Phys. Rev. Lett.* **105**, 077001 (2010).
- [25] D. Rainis and D. Loss, Majorana qubit decoherence by quasiparticle poisoning, *Phys. Rev. B* **85**, 174533 (2012).
- [26] D. Aasen, M. Hell, R. V. Mishmash, A. Higginbotham, J. Danon, M. Leijnse, T. S. Jespersen, J. A. Folk, C. M. Marcus, K. Flensberg, and J. Alicea, Milestones Toward

- Majorana-Based Quantum Computing, *Phys. Rev. X* **6**, 031016 (2016).
- [27] B. S. Palmer, C. A. Sanchez, A. Naik, M. A. Manheimer, J. F. Schneiderman, P. M. Echternach, and F. C. Wellstood, Steady-state thermodynamics of nonequilibrium quasiparticles in a Cooper-pair box, *Phys. Rev. B* **76**, 054501 (2007).
- [28] M. D. Shaw, R. M. Lutchyn, P. Delsing, and P. M. Echternach, Kinetics of nonequilibrium quasiparticle tunneling in superconducting charge qubits, *Phys. Rev. B* **78**, 024503 (2008).
- [29] R. Barends, J. Wenner, M. Lenander, Y. Chen, R. C. Bialczak, J. Kelly, E. Lucero, P. O'Malley, M. Mariantoni, D. Sank, H. Wang, T. C. White, Y. Yin, J. Zhao, A. N. Cleland, J. M. Martinis, and J. J. A. Baselmans, Minimizing quasiparticle generation from stray infrared light in superconducting quantum circuits, *Appl. Phys. Lett.* **99**, 113507 (2011).
- [30] C. Wang, Y. Y. Gao, I. M. Pop, U. Vool, C. Axline, T. Brecht, R. W. Heeres, L. Frunzio, M. H. Devoret, G. Catelani, L. I. Glazman, and R. J. Schoelkopf, Measurement and control of quasiparticle dynamics in a superconducting qubit, *Nat. Commun.* **5**, 5836 (2014).
- [31] S. Gustavsson, F. Yan, G. Catelani, J. Bylander, A. Kamal, J. Birenbaum, D. Hover, D. Rosenberg, G. Samach, A. P. Sears, S. J. Weber, J. L. Yoder, J. Clarke, A. J. Kerman, F. Yoshihara, Y. Nakamura, T. P. Orlando, and W. D. Oliver, Suppressing relaxation in superconducting qubits by quasiparticle pumping, *Science* **354**, 1573 (2016).
- [32] K. Serniak, S. Diamond, M. Hays, V. Fatemi, S. Shankar, L. Frunzio, R. Schoelkopf, and M. Devoret, Direct Dispersive Monitoring of Charge Parity in Offset-Charge-Sensitive Transmons, *Phys. Rev. Appl.* **12**, 014052 (2019).
- [33] A. P. Vepsäläinen, A. H. Karamlou, J. L. Orrell, A. S. Dogra, B. Loer, F. Vasconcelos, D. K. Kim, A. J. Melville, B. M. Niedzielski, J. L. Yoder, S. Gustavsson, J. A. Formaggio, B. A. VanDevender, and W. D. Oliver, Impact of ionizing radiation on superconducting qubit coherence, *Nature* **584**, 551 (2020).
- [34] S. B. Kaplan, C. C. Chi, D. N. Langenberg, J. J. Chang, S. Jafarey, and D. J. Scalapino, Quasiparticle and phonon lifetimes in superconductors, *Phys. Rev. B* **14**, 4854 (1976).
- [35] K. Serniak, Nonequilibrium quasiparticles in superconducting qubits, Ph.D. thesis, Yale University (2019).
- [36] M. Houzet, K. Serniak, G. Catelani, M. Devoret, and L. Glazman, Photon-Assisted Charge-Parity Jumps in a Superconducting Qubit, *Phys. Rev. Lett.* **123**, 107704 (2019).
- [37] C. D. Wilen, S. Abdullah, N. A. Kurinsky, C. Stanford, L. Cardani, G. D'Imperio, C. Tomei, L. Faoro, L. B. Ioffe, C. H. Liu, A. Opremcak, B. G. Christensen, J. L. DuBois, and R. McDermott, Correlated charge noise and relaxation errors in superconducting qubits, *Nature* **594**, 369 (2021).
- [38] J. M. Martinis, Saving superconducting quantum processors from decay and correlated errors generated by gamma and cosmic rays, *npj Quantum Inf.* **7**, 1 (2021).
- [39] C. Kurter, C. E. Murray, R. T. Gordon, B. B. Wymore, M. Sandberg, R. M. Shelby, A. Eddins, V. P. Adiga, A. D. K. Finck, E. Rivera, A. A. Stabile, B. Trimm, B. Wacaser, K. Balakrishnan, A. Pyzyna, J. Sleight, M. Steffen, and K. Rodbell, Quasiparticle tunneling as a probe of Josephson junction quality and capacitor material in superconducting qubits (2021), [ArXiv:arXiv:2106.11488](https://arxiv.org/abs/2106.11488).
- [40] O. Rafferty, S. Patel, C. H. Liu, S. Abdullah, C. D. Wilen, D. C. Harrison, and R. McDermott, Spurious antenna modes of the transmon qubit (2021), [ArXiv:2103.06803](https://arxiv.org/abs/2103.06803).
- [41] X. Pan, H. Yuan, Y. Zhou, L. Zhang, J. Li, S. Liu, Z. H. Jiang, G. Catelani, L. Hu, and F. Yan, Engineering superconducting qubits to reduce quasiparticles and charge noise (2022), [ArXiv:2202.01435](https://arxiv.org/abs/2202.01435).
- [42] C.-H. Liu, D. C. Harrison, S. Patel, C. D. Wilen, O. Rafferty, A. Shearrow, A. Ballard, V. Iaia, J. Ku, B. L. T. Plourde, and R. McDermott, Quasiparticle poisoning of superconducting qubits from resonant absorption of pair-breaking photons (2022), [ArXiv:2203.06577](https://arxiv.org/abs/2203.06577).
- [43] P. Chubov, V. V. Eremenko, and Y. A. Pilipenko, Dependence of the critical temperature and energy gap on the thickness of superconducting aluminum films, *Sov. Phys. JETP* **28**, 389 (1969).
- [44] T. Yamamoto, Y. Nakamura, Y. A. Pashkin, O. Astafiev, and J. S. Tsai, Parity effect in superconducting aluminum single electron transistors with spatial gap profile controlled by film thickness, *Appl. Phys. Lett.* **88**, 212509 (2006).
- [45] N. A. Court, A. J. Ferguson, and R. G. Clark, Energy gap measurement of nanostructured aluminium thin films for single Cooper-pair devices, *Supercond. Sci. Technol.* **21**, 015013 (2007).
- [46] In bridge-free fabrication of Al/AIO_x/Al JJs, the second aluminum film is typically deposited more thickly in order to ensure that it climbs the initial film.
- [47] We note that neither the observed excited state populations nor energy-relaxation rates can be responsible for the observed flux dependence of Γ .
- [48] R.-P. Riwar, A. Hosseinkhani, L. D. Burkhardt, Y. Y. Gao, R. J. Schoelkopf, L. I. Glazman, and G. Catelani, Normal-metal quasiparticle traps for superconducting qubits, *Phys. Rev. B* **94**, 104516 (2016).
- [49] R.-P. Riwar and G. Catelani, Efficient quasiparticle traps with low dissipation through gap engineering, *Phys. Rev. B* **100**, 144514 (2019).
- [50] L. Glazman and G. Catelani, Bogoliubov quasiparticles in superconducting qubits, *SciPost Phys. Lect. Notes*, 31 (2021).
- [51] The imbalance of tunneling across the Josephson junction prompts the allowance of $\mu_L \neq \mu_R$.
- [52] I. M. Pop, T. Fournier, T. Crozes, F. Lecocq, I. Matei, B. Pannetier, O. Buisson, and W. Guichard, Fabrication of stable and reproducible submicron tunnel junctions, *J. Vac. Sci. Technol. B, Nanotechnol. Microelectron.: Mater., Process., Meas., Phenom.* **30**, 010607 (2012).
- [53] The same lamp had no effect when implemented outside of the aluminum can surrounding the device.
- [54] R. T. Gordon, C. E. Murray, C. Kurter, M. Sandberg, S. A. Hall, K. Balakrishnan, R. Shelby, B. Wacaser, A. A. Stabile, J. W. Sleight, M. Brink, M. B. Rothwell, K. P. Rodbell, O. Dial, and M. Steffen, Environmental radiation impact on lifetimes and quasiparticle tunneling rates of fixed-frequency transmon qubits, *Appl. Phys. Lett.* **120**, 074002 (2022). publisher: American Institute of Physics
- [55] A. P. M. Place, *et al.*, New material platform for superconducting transmon qubits with coherence times exceeding 0.3 milliseconds, *Nat. Commun.* **12**, 1779 (2021).

- [56] J. M. Martinis, K. B. Cooper, R. McDermott, M. Steffen, M. Ansmann, K. D. Osborn, K. Cicak, S. Oh, D. P. Pappas, R. W. Simmonds, and C. C. Yu, Decoherence in Josephson Qubits from Dielectric Loss, *Phys. Rev. Lett.* **95**, 210503 (2005).
- [57] I. Siddiqi, Engineering high-coherence superconducting qubits, *Nat. Rev. Mater.* **6**, 875 (2021).
- [58] V. Iaiia, J. Ku, A. Ballard, C. P. Larson, E. Yelton, C. H. Liu, S. Patel, R. McDermott, and B. L. T. Plourde, Phonon downconversion to suppress correlated errors in superconducting qubits (2022), [ArXiv:2203.06586](https://arxiv.org/abs/2203.06586).
- [59] N. A. Court, A. J. Ferguson, R. Lutchyn, and R. G. Clark, Quantitative study of quasiparticle traps using the single-Cooper-pair transistor, *Phys. Rev. B* **77**, 100501 (2008).
- [60] M. McEwen, *et al.*, Resolving catastrophic error bursts from cosmic rays in large arrays of superconducting qubits, *Nat. Phys.* **18**, 107 (2022).
- [61] N. E. Frattini, V. V. Sivak, A. Lingenfelter, S. Shankar, and M. H. Devoret, Optimizing the Nonlinearity and Dissipation of a SNAIL Parametric Amplifier for Dynamic Range, *Phys. Rev. Appl.* **10**, 054020 (2018).
- [62] L. Grünhaupt, N. Maleeva, S. T. Skacel, M. Calvo, F. Levy-Bertrand, A. V. Ustinov, H. Rotzinger, A. Monfardini, G. Catelani, and I. M. Pop, Loss Mechanisms and Quasiparticle Dynamics in Superconducting Microwave Resonators Made of Thin-Film Granular Aluminum, *Phys. Rev. Lett.* **121**, 117001 (2018).