Nonequilibrium quasiparticle excitations degrade the performance of a variety of superconducting circuits. Understanding the energy distribution of these quasiparticles will yield insight into their generation mechanisms, the limitations they impose on superconducting devices, and how to efficiently mitigate quasiparticle-induced qubit decoherence. To probe this energy distribution, we systematically correlate quasiparticle relaxation and excitation with charge-parity sensitive transmon qubit, and find that quasiparticle-induced excitation events are the dominant mechanism behind the residual excited-state population in our samples. By itself, the observed quasiparticle distribution would limit $T_1$ to $\approx 200 \mu s$, which indicates that quasiparticle loss in our devices is on equal footing with all other loss mechanisms. Furthermore, the measured rate of quasiparticle-induced excitation events is greater than that of relaxation events, which signifies that the quasiparticles are more energetic than would be predicted from a thermal distribution describing their apparent density.

In this Letter, we report signatures of hot nonequilibrium quasiparticles (QPs) ubiquitous in aluminum superconducting devices have been recognized in a wide variety of systems, including Josephson junction (JJ) based superconducting qubits [1–13], kinetic-inductance [14–16] and quantum-capacitance [17] detectors, devices for current metrology [18], Andreev qubits [19–21], and proposed Majorana qubits [22,23]. While recent efforts to reduce the density of QPs in superconducting qubits have shown some improvement in the relaxation times of devices limited by QP-induced loss [11,24–26], understanding the energy distribution of nonequilibrium QPs may shed light on their source and further help to mitigate their effects. Furthermore, it has been suggested that “hot” nonequilibrium QPs may be responsible for the residual excited-state population seen in superconducting qubits at low temperatures [8,27,28], though this has yet to be confirmed directly.

In this Letter, we report signatures of hot nonequilibrium QPs observed in the correlations between qubit transitions and QP-tunneling events. An offset-charge-sensitive transmon qubit was used to directly detect switches in the charge parity of the transmon islands associated with individual QPs tunneling across the JJ [9]. We correlated these charge-parity switches with transitions between the ground and first-excited states of the transmon, and found that QP tunneling accounts for $\approx 30\%$ of all qubit relaxation events and $\approx 90\%$ of excitation events. The measured ratio of the QP-induced excitation and relaxation rates is greater than 1, which is at odds with a thermal distribution accounting for their estimated density, defining what we refer to as a “hot” energy distribution of tunneling QPs. These results confirm previous suspicions that nonequilibrium QPs are responsible for the residual excited-state population in transmon qubits [8,27,28], and emphasize the need for further understanding of QP-induced loss.

Ideally, QPs in superconducting devices would be in thermal equilibrium with their thermal anchor ($T \approx 20 \text{ mK}$ for dilution refrigerators), and their spontaneous generation would be exponentially suppressed by the superconducting gap $\Delta$. However, there is an observed fraction of broken Cooper pairs $\chi_{QP}^0 \approx 10^{-8} – 10^{-6}$ [1,3,4,11,25,27,29–31] which is orders of magnitude greater than would be predicted in thermal equilibrium. In a transmon [32], QP tunneling across the JJ will always change the excess charge on the islands by $1e$, switching the charge parity of the junction electrodes between “even” and “odd” [2]. Tunneling QPs couple to the phase across the JJ [4,6], and consequently can induce qubit transitions (Fig. 1). If the QPs were in thermal equilibrium, the values of $\chi_{QP}^0$ quoted above would correspond to an effective QP temperature of 130–190 mK. Under this assumption, QP-induced relaxation of the qubit should vastly outweigh QP-induced excitation. As we will show, this is not observed in our devices, indicating that this effective temperature does not adequately describe the QP energy distribution.

To directly probe the interaction between nonequilibrium QPs and a transmon qubit, we slightly relax the transmon-defining condition that the Josephson coupling energy $E_J$ is much greater than the charging energy $E_C$ [7]. In this regime, the ground-to-excited-state-transition frequency $f_{01} = (E_1 - E_0)/\hbar$ has a measurable dependence on charge...
FIG. 1. QP-induced transitions in transmon qubits. (a) Density of states $\nu_\sigma$ versus the reduced energy $\epsilon/\Delta$ in the leads of a superconductor-insulator-superconductor (SIS) JJ, in the excitation representation. Gray arrows represent tunneling processes of QPs, shown as purple dots. Dashed, dotted, and solid lines correspond to relaxation, excitation, and interband transitions of the qubit, respectively, with associated inelastic QP scattering. (b) The two lowest energy levels of an offset-charge-sensitive transmon qubit (vertical axis not to scale) as a function of offset charge $n_g$, in units of $2\epsilon$. These levels are shifted depending on the charge parity (even or odd) of the qubit, and $E_0$ and $E_1$ are time-averaged energies of the ground and first-excited states, respectively, assuming ergodic fluctuations of $n_g$ and/or charge parity. Arrows correspond to those in (a).

parity, switching between $f_{01} \pm \delta f_{01}$ when a QP tunnels across the JJ [the qubit energies switch between the blue and red lines in Fig. 1(b)] [7,9]. The deviation $\delta f_{01}$ is a sinusoidal function of the dimensionless offset charge $n_g$, which undergoes temporal fluctuations due to reconfiguration of mobile charges in the environment. Because $\hbar \delta f_{01}(n_g) \ll k_B T$, QP-tunneling dynamics will not depend strongly on $n_g$. The authors of Ref. [9] took advantage of this frequency splitting to track $n_g$, map the charge parity onto the state of a transmon, and correlate qubit relaxation with parity switches [33]. Extending their experiment, we extract not only the QP-induced relaxation rate, but also the QP-induced excitation rate by detailed modeling of the correlations between charge-parity switches and qubit transitions.

We focus below on a single transmon qubit with average frequency $f_{01} = 4.400$ GHz and $E_j/E_C = 23$, corresponding to a maximum even-odd splitting $2\delta f_{01}(0) = 3.18$ MHz. The average measured relaxation time $T_1 = 95 \mu$s is on par with state-of-the-art transmons, and the equilibrium ground state population $P^0_{01} = 0.74$ corresponds to an effective qubit temperature of 160 mK. Data from a second sample with similar parameters are discussed in Supplemental Material [34]. Chips were mounted in an Al 3D rectangular readout cavity [38] and anchored to the mixing chamber of a cryogen-free dilution refrigerator at 20 mK.

The slow background fluctuations of $n_g$ were tracked by monitoring $\delta f_{01}(n_g)$ using the Ramsey sequence depicted in Fig. 2(a). The carrier frequency of the Gaussian $\pi/2$ pulses is chosen to be $f_{01}$, which is symmetrically detuned from the even and odd charge-parity states at all values of $n_g$. This ensures that the phase evolution of even- and odd-parity states on the equator of the Bloch sphere will interfere constructively, resulting in Ramsey fringes [Fig. 2(b)] characterized by a single oscillation frequency $\delta f_{01}(n_g)$ and a decay constant $T_2$ that is insensitive to fast charge-parity switches. Repeated Ramsey experiments [Fig. 2(c)] show that $n_g$ fluctuates on a timescale of minutes, which is long enough to perform experiments that rely on prior knowledge of $\delta f_{01}(n_g)$.

Using a similar pulse sequence [Fig. 3(a)], we map the charge parity of the transmon onto the qubit state [9]. Two $\pi/2$ pulses, now about orthogonal axes, are separated by a delay $\tau(n_g) = 1/4\delta f_{01}(n_g)$, which constitutes an effective $\pi$ pulse conditioned on charge parity ($\pi_{e,o}$). This charge-parity-mapping operation only discerns between transition frequencies greater than or less than $f_{01}$, and we refer to these as “even” and “odd” charge-parity states, respectively, despite the inability to measure absolute parity. The relative phase of the $\pi/2$ pulses controls whether the $\pi_{e,o}$ sequence is conditioned on even or odd charge parity. The charge parity $P = (2M_1 - 1)(2M_2 - 1)$ is calculated in postprocessing. To observe QP-tunneling events in real time, we repeated the charge-parity-mapping sequence every $\Delta t_{exp} = 10 \mu$s for ~600 ms [Fig. 3(b)]. The power spectral density $S_{pp}$ of these parity fluctuations was averaged over 20 independent charge-parity jump traces (Fig. 3). $S_{pp}$ was fit to the characteristic Lorentzian of a random telegraph signal, from which a parity-switching timescale $T_p = 77 \pm 1 \mu$s and mapping fidelity $\mathcal{F} = 0.91 \pm 0.01$ were obtained [34]. Each jump trace was acquired after confirming that $\delta f_{01}(n_g) > 1$ MHz by the monitoring of $n_g$ described above. This conditioning was
During the correspondingly shorter δ, corresponding to of charge-parity fluctuations, with a Lorentzian fit (orange curve) introduced to increase the fidelity F of the parity mapping, as δf_{01}(n_g) is less sensitive to fluctuations in n_g at near-maximum δf_{01}(n_g); also, the qubit is less likely to dephase during the correspondingly shorter τ(n_g).

The fact that T_P ≈ T_1 hints at the possibility that our transmon may be limited by QP-induced dissipation. Following Ref. [33], the total relaxation rate Γ_{10} can be decomposed into the sum of two contributions: the rate of relaxation accompanied by a charge-parity switch (Γ_{01}^{TP}), which we attribute solely to QP-induced loss, and the rate of relaxation from charge-parity-conserving mechanisms (Γ_{01}^{ee}), such as dielectric loss. As there is no preferred parity, these transition rates are symmetric under exchange of even and odd (Γ_{ij}^{pp} = Γ_{ij}^{pp'} and Γ_{ij}^{oe} = Γ_{ij}^{oe'}). Similarly to the total relaxation rate, the total excitation rate is given by Γ_{01} = Γ_{01}^{ee} + Γ_{01}^{TP}.

We resolve these distinct contributions by concatenating two parity-mapping sequences (outcomes p and p') separated by a variable delay τ [Fig. 4(a), inset]. This measurement determines both the charge-parity and qubit state before and after τ, which allows us to correlate qubit transitions with QP-tunneling events. From our data, we compute 〈p_j; p'_j|τ〉, the probability of measuring outcome m_3 = j after a delay τ given that m_3 = i, with or without a parity switch (p_j = ±1 or +1, respectively).

To model these quantities, we employ a master equation describing the flow of probability between different system states:

\[ \rho_{ij}^{\alpha} = -(\Gamma_{ii}^{oe} + \Gamma_{ii}^{tp} + \Gamma_{ii}^{eo})\rho_{ii}^{\alpha} + \Gamma_{ii}^{eo}\rho_{i}^{\alpha} + \Gamma_{ii}^{tp}\rho_{i}^{\alpha} + \Gamma_{ii}^{ee}\rho_{i}^{\alpha}, \]

where \( \rho_{ij}^{\alpha} \) is the probability of finding the system in qubit state \( i \) and charge parity \( \alpha \), and \( \overline{\alpha} \) is read as “not \( \alpha \).” We evolve the above model numerically with initial conditions determined by \( M_2 \) and \( P \), and fit all eight conditional probabilities 〈p_j; p'_j|τ〉, a subset of which are shown in Figs. 4(a) and 4(b).

In addition, we calculate the charge-parity autocorrelation function \( \langle PP'\rangle_{ij}(τ) \), again conditioned on \( m_2 = i \) and \( m_3 = j \), respectively [Fig. 4(d)], and fit to functions of the form [34]:

\[ \langle PP'\rangle_{ij}(τ) = ρ_{ij}^{\alpha}(0)(ρ_{ij}^{\alpha}(τ) - ρ_{ij}^{\alpha}(0))/ρ_{ij}^{\alpha}(τ) + ρ_{ij}^{\alpha}(0) . \]
The maximum correlation $\langle PP'\rangle_{ij}(0)$ is limited by the fidelity of the correlation measurement, and qualitatively, the deviation of $\langle PP'\rangle_{ij}(0)$ from this maximum amplitude is related to the ratio $\Gamma_{ij}^{\text{ee}} / \Gamma_{ij}$ [Fig. 4(d)]. Equations (1) and (2) do not account for any measurement infidelities, which can skew the observed correlations. These include parity- and qubit-state-dependent errors, such as spontaneous qubit transitions during the parity-mapping sequence, as well as global errors such as pulse infidelity due to uncertainty in $\delta f_{01}(n_p)$. We stress that proper modeling of these errors is necessary to accurately extract the conditional rates. Taking into account these considerations, we fit all eight permutations of $\rho(j, p'|i)(\tau)$ and the four $\langle PP'\rangle_{ij}(\tau)$ curves simultaneously to the master equation model (solid lines in Fig. 4). For more details on the model and fit, see Supplemental Material [34]. The slight disagreement at short $\tau$ may be due to measurement-induced qubit transitions that could be present even at low readout power [39,40].

From our model with measurement errors taken into account, we extract $1/\Gamma_{10}^{\text{ee}} = 110 \pm 1 \mu$s, $1/\Gamma_{10}^{\text{eo}} = 77 \pm 1 \mu$s, $1/\Gamma_{10}^{\text{eo}} = 447 \pm 7 \mu$s, $1/\Gamma_{10}^{\text{eo}} = 400 \pm 5 \mu$s, $1/\Gamma_{10}^{\text{eo}} = 182 \pm 1 \mu$s, and $1/\Gamma_{10}^{\text{eo}} = 6500 \pm 900 \mu$s. Quoted parameter standard deviations reflect the uncertainty in the data, calculated using standard statistical techniques [41]. As a check of consistency, we calculate $T_1 = (\Gamma_{10}^{\text{eo}} + \Gamma_{10}^{\text{ee}} + \Gamma_{01}^{\text{eo}} + \Gamma_{01}^{\text{ee}})^{-1}$, $\rho_{\text{eo}} = (\Gamma_{10}^{\text{eo}} + \Gamma_{10}^{\text{ee}})T_1$, and $T_p \approx 2/\Gamma_{01}^{\text{eo}} + \Gamma_{11}^{\text{eo}} + \Gamma_{10}^{\text{eo}} + \Gamma_{01}^{\text{ee}}$, and find that they agree with the independently measured values quoted above [42]. A second transmon was found to have similar rates [34].

These rates have implications for our understanding of nonequilibrium QPs in our transmon qubits. First, the limit on $T_1$ of this sample imposed by QPs is $(\Gamma_{10}^{\text{eo}} + \Gamma_{10}^{\text{ee}})^{-1} = 211 \pm 3 \mu$s, compared to a limit of $(\Gamma_{10}^{\text{eo}} + \Gamma_{10}^{\text{ee}})^{-1} = 177 \pm 2 \mu$s imposed by all other loss mechanisms. This puts QP-induced dissipation on par with the sum of all other dissipation channels, contributing significantly to qubit relaxation $\Gamma_{10}^{\text{eo}} / \Gamma_{10} = 0.29 \pm 0.01$. Second, the ratio $\Gamma_{01}^{\text{eo}} / \Gamma_{01}^{\text{ee}} = 0.94 \pm 0.02$ indicates that QP-induced excitation accounts for the vast majority of the residual transmon excited-state population [Fig. 4(a)], confirming previous suspicions [8,28]. Finally, $\Gamma_{01}^{\text{eo}} / \Gamma_{10}^{\text{ee}} = 1.12 \pm 0.02$, which is direct evidence of a highly energetic distribution of QPs. Naïvely applying Fermi-Dirac statistics and detailed balance yields $\Gamma_{01}^{\text{eo}} / \Gamma_{10}^{\text{ee}} = \exp(-hf_{01}/k_BT_{\text{eff}}^{\text{QP}})$, which predicts a negative effective QP temperature $T_{\text{eff}}^{\text{QP}} \approx -2 \text{ K}$ in our device. This is evidence that the QP energy distribution is not localized near the gap edge, but has a characteristic energy greater than $\Delta + hf_{01}$. Conversely, $\Gamma_{01}^{\text{eo}} / \Gamma_{10}^{\text{ee}} = 0.03 \pm 0.01$, indicating that the non-QP dissipative baths coupled to the transmon are relatively “cold” [Fig. 4(b)], with an effective temperature $\sim 60 \text{ mK}$. The observation that $\Gamma_{11}^{\text{ee}} > \Gamma_{01}^{\text{ee}}$ is not yet explained by theoretical predictions [33]. We note that some weak dependence of QP dynamics on $E_J/E_C$ is expected, and following Appendix A of Ref. [33], we find that the QP-induced transition rates vary by less than a factor of 2 in the range $23 < E_J/E_C < 100$, with lower $E_J/E_C$ corresponding to increased QP sensitivity. To first order in perturbation theory, the ratio $\Gamma_{01}^{\text{eo}} / \Gamma_{10}^{\text{eo}}$ will not depend on $E_J/E_C$.

We repeated the correlation measurement (Fig. 4) at various mixing-chamber temperatures $T$ (Fig. 5). We find that all parity-switching rates $\Gamma_{ij}^{\text{eo}}$ increase after $\sim 140 \text{ mK}$, at which point $T_1$, $T_p$, and $\Gamma_{01}^{\text{eo}} / \Gamma_{10}^{\text{ee}}$ all begin to decrease. Modeling the temperature dependence of these rates requires some ansatz about the QP energy distribution, which is typically assumed to be localized near the gap edge [4,6]. While this assumption appears not to be valid for QPs in our system, we use it to compare our results with other reports of QP density $x_{\text{QP}}$ in superconducting circuits. If we further assume that the populations of nonequilibrium QPs and equilibrium QPs [6] are independent, the total $x_{\text{QP}}$ is the sum

$$x_{\text{QP}} = x_{\text{QP}}^0 + \sqrt{2\pi k_BT/\Delta}e^{-\Delta/k_BT}. \quad (3)$$

Here $\Delta = 205 \mu \text{eV}$, consistent with dc measurements of similar films (Δ increases with reduction of Al thickness) [43]. The QP-induced relaxation rate $\Gamma_{10}^{\text{eo}}$ should scale linearly with $x_{\text{QP}}$ [4,6]. We see this approximate scaling in

![FIG. 5. Temperature dependence of qubit-state-conditioned parity-switching rates. (a) Above $\sim 140 \text{ mK}$, all rates begin to increase, and $\Gamma_{01}^{\text{eo}} / \Gamma_{10}^{\text{ee}} \leq 1$ suggests that thermally generated QPs begin to outnumber nonequilibrium QPs. (b) $1/\Gamma_{10}^{\text{ee}}$ normalized by its base-temperature value $1/\Gamma_{10}^{\text{eo}}$ as a function of temperature. The solid black line is a fit to the thermal dependence of $x_{\text{QP}}^0 / x_{\text{QP}}$, which gives $x_{\text{QP}}^0 \approx 1 \times 10^{-7}$. (c) $\Gamma_{01}^{\text{eo}} / \Gamma_{10}^{\text{ee}}$ compared to predictions from detailed balance, assuming QPs are thermalized with the cryostat. Gray dashed line indicates the value above which $T_{\text{eff}}^{\text{QP}} \lesssim 0$.](157701-4)
yields our data [Fig. 5(b)] with a slight decrease in $\Gamma_{10}^{\text{se}}$ with increasing temperature that is not predicted by our simple model, but has been previously observed [4]. This model yields $x_{\text{QP}}^0 \approx 1 \times 10^{-7}$, which agrees with other recent experiments [1,10,11,24,25].

Thus, we have shown that QPs are more energetic than a Fermi-Dirac distribution accounting for their apparent density $x_{\text{QP}}^0$ would suggest. Further quantitative analysis of the measured parity-switching rates, together with modeling of QP dynamics in our Al films, could reveal the energy range of QP-generating excitations. Proper filtering of rf lines, light-tight shielding [44,45], and well-thermized components are now standard ingredients for reducing the QP density which were included in our measurement setup [34]. One should note that the authors of Ref. [9] reported $T_p$ one order of magnitude greater than what we have presented, with one experimental difference being a Cu readout cavity instead of a superconducting Al cavity.

In conclusion, the correlations between charge-parity switches and qubit transitions in an offset-charge-sensitive transmon indicate that QP-induced loss can be responsible for a significant fraction of dissipation in state-of-the-art superconducting qubits. Additionally, we confirm that hot QPs with a highly excited energy distribution are responsible for the residual excited-state population at low temperature in our samples. The techniques described above, building upon Ref. [9], provide a tool to distinguish the influences of various experimental factors on QP generation and assess QP-reduction techniques, such as induced Abrikosov vortices [11,24,25,31] or galvanically connected QP traps [13,46–52].

We acknowledge insightful discussions with Gianluigi Catelani, Leo DiCarlo, Yvonne Gao, Leonid Glazman, Ioan Pop, Dan Prober, Rob Schoelkopf, and Uri Vool. Facilities use was supported by the Yale Institute for Nanoscience and Quantum Engineering (YINQE), the Yale SEAS cleanroom, and NSF MRSEC DMR 1119826. This research was supported by ARO under Grant No. W911NF-14-1-0011, by MURI-ONR under Grant No. N00014-16-1-2270, and NSF DMR Grant No. 1603243. G. d. L. acknowledges support from the European Union through the Marie Skłodowska-Curie Grant Agreement No. 656129. M. Houzet acknowledges support from the European Union’s FP7 programme through the Marie-Skłodowska-Curie Grant Agreement No. 600382.

*\text{kyle.serniak@yale.edu}

\text{michel.devoret@yale.edu}


[34] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.121.157701 for more information, which includes in Ref. [35–37].


[42] This approximate relation for $T_P$ is due to the fact that in the aforementioned charge-parity jump experiment (Fig. 3) the qubit is taken out of equilibrium by the pulse sequence. The extracted $T_P$ approximately averages the conditional parity-switching rates corresponding to the qubit states $|0\rangle$ and $|1\rangle$.


