Continuous Real-Time Detection of Quasiparticle Trapping in Aluminum Nanobridge Josephson Junctions

J. T. Farmer,1, 2 A. Zarassi,1, 2 D. M. Hartsell,1, 2 E. Vlachos,1, 2 H. Zhang,3, 2 and E. M. Levenson-Falk1, 2

1) Department of Physics & Astronomy, Dornsife College of Letters, Arts, & Sciences, University of Southern California, Los Angeles, CA 90089, USA
2) Center for Quantum Information Science & Technology, University of Southern California, Los Angeles, CA 90089, USA
3) Department of Electrical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089

(Dated: 7 September 2021)

Nonequilibrium quasiparticles are ubiquitous in superconducting electronics. These quasiparticles can trap in the internal Andreev bound states of a phase-biased Josephson junction, providing a mechanism for studying their presence and behavior. We characterize a quasiparticle trapping detector device based on a two-junction aluminum nanobridge superconducting quantum interference device incorporated into a transmission-line resonator. When flux-biased, distinct resonant frequencies develop depending on the trapped quasiparticle number. We demonstrate continuous detection of up to 3 trapped quasiparticles, with detection of a trapped quasiparticle with signal-to-noise ratio of 27 in 5 µs. We describe initial measurements of quasiparticle behavior and discuss the possible optimization and application of such detector devices.

Superconducting qubits and other low-temperature superconducting electronics have ubiquitous populations of quasiparticles (QPs) far above their thermal equilibrium prevalence. These QPs can cause loss, spurious excitations, and spectral noise when tunneling across qubit junctions. Even when QP populations are extraordinarily low, rare bursts of QPs can induce correlated errors that are difficult to address with error correction algorithms. QPs may be generated by stray infrared photons, cosmic rays and other high-energy radiation sources, or materials defects. Many experiments have probed QP behavior via their tunneling across Josephson junctions in charge-sensitive transmons, giving valuable insight into their effects on qubits. However, these measurements are discrete and cannot distinguish between 0 and 2 tunneling events. Trapping measurements provide a tool for continuous, nonsaturating monitoring of QP behavior. Furthermore, QP traps have been proposed as a tool to mitigate QPs’ effects on qubits, as QPs may diffuse great distances after being generated. The trapping process itself is thus worthy of study, in addition to providing insight into bulk QP behavior.

In this letter we characterize a device optimized for continuous, non-saturating measurements of QP trapping in Andreev states. Using microwave reflectometry we are able to continuously detect 0, 1, and 2 more trapped quasiparticles in 5 ms with a signal-to-noise ratio (SNR) of 27. By altering the detector bias we are able to distinguish 3, 2, and 1 or fewer QPs. We discuss straightforward improvements that can further improve SNR and allow detection of many more trapped QPs at a single bias point. Our device provides a prototype for detectors optimized for continuous measurements of QP behavior and properties and for studies of the dynamics of Andreev states coupled to resonant cavities.

In the semiconductor picture of the Josephson effect, supercurrent is carried by electrons/holes traveling in 1 dimensional conduction channels. At the junction boundaries, the electron (hole) reflects as a hole (electron). This Andreev reflection causes a ±e charge transfer, where e is the elementary charge, and thus transmits a Cooper pair across the junction. Each channel forms a pair of Andreev bound states with energies

\[ E_{A\pm} = \pm \Delta \sqrt{1 - \sin^2 \frac{\delta}{2}} \]  

where \( \Delta \) is the superconducting gap, \( \tau \) is the transmissivity of the channel, and \( \delta \) is the phase bias across the junction. In the semiconductor picture the Fermi energy is 0, so at temperatures \( T \ll \Delta/k_B \), normally the upper Andreev state is unoccupied and the lower state is occupied, carrying the supercurrent across the junction. This is the channel’s \( |g \rangle \) state with total energy \(-E_A\). QPs in the bulk only exist at energies greater than \( \Delta \) (or as unoccupied states below \(-\Delta\)), so it is energetically favorable for a QP to drop into the unoccupied upper Andreev state, bringing the channel to the \( |o \rangle \) state with 0 energy. This “poisons” the channel, eliminating it from carrying supercurrent and increasing the Josephson inductance. The lower-level QP may also be promoted to the upper Andreev level, bringing the channel to the \( |e \rangle \) and producing twice the inductance shift of the \( |o \rangle \) state; or this QP may be cleared from the junction completely, creating another degenerate 0-energy \( |o \rangle \) state with 0 supercurrent. Researchers have demonstrated continuous monitoring of QPs trapping in point contact and semiconductor nanowire junctions; however, these junctions have few channels and so quickly saturate as QP detectors.

Three dimensional aluminum nanobridge Josephson
junctons achieve good phase confinement and nonlinearity in an all-superconducting design. Importantly, nanobridges comprise many conduction channels (\(\sim 100 - 1000\)) with \(\tau \sim 1\), approximately following the Dorokhov distribution \(\rho(\tau) = \frac{\pi G}{2e^2} \frac{1}{\tau \sqrt{1 - \tau}}\), where \(G\) is the junction’s normal-state conductance. When 2 identical nanobridges are placed in a loop, forming a superconducting quantum interference device (SQUID), the junctions’ phase bias \(\delta\) is simply set by the flux bias \(\Phi\) as \(\delta = \pi \phi\), where \(\phi \equiv \Phi / \Phi_0\) and \(\Phi_0\) is the flux quantum. In an Al nanobridge SQUID near half-flux \((\delta \approx \pi/2\)) as \(E_{A+}/h \approx 29\) GHz and has linewidth \(\kappa = 2\tau \times 250\) kHz, largely set by the coupling to the microwave feedline. This device was imaged and found to have junctions that appear nearly identical visually; past studies indicate they may thus be treated as symmetric. Trapping in either junction produces a similar resonant frequency shift, and so for QP detector applications they may be thought of as a single junction with twice the number of channels. We note that the resonator energy per photon is much smaller than the trap depth at high flux bias, so absorption of resonator photons should not appreciably affect the quasiparticle states.

Our measurement setup is shown schematically in Fig. 1(c). A signal generator provides a drive tone at \(\omega_0(\phi) \approx \omega_0(\phi) - \kappa/2\). A power splitter sends half of this power into the dilution refrigerator where it is attenuated then (in some cooldowns) filtered by K&L 12 GHz and Eccosorb low pass filters. The drive tone is circulated to reflect off our device, which is flux tunable via a CPW resonator in which the center trace is terminated by a DC coil in the packaging, and amplified by a travelling wave parametric amplifier (TWPA) which is pumped at 8.078 GHz. The reflected signal is amplified before IQ demodulation with reference to the original signal. The in-phase (I) and quadrature (Q) components are 15 MHz low pass filtered before digitization at 300 MSA/s. Raw data is down-sampled to 10 MSA/s before saving.

We first characterize our device with ensemble-averaged vector network analyzer (VNA) measurements of the resonance. Fig. 1(a) shows resonance measurements at flux biases of \(\phi = 0\) and \(\phi = 0.49\), taken on a cooldown in which the K&L and Eccosorb filters were removed. The \(\phi = 0.49\) trace in orange shows two shallow peaks at \(\sim 0.5\) and 1 MHz below the main resonance. These are the resonance peaks with 1 and 2 trapped QPs, respectively, showing the resonance shifting due to the change in nanobridge inductance. These ensemble measurements average over all possible QP trapping configurations, and so a resonance peak amplitude corresponds to the probability of that configuration. We then move on to time-domain IQ measurements as described above. Panels (b) and (c) are log-scale histograms showing 30 seconds of continuous IQ data for \(\phi = 0\) (in blue) and \(\phi = 0.47\) (in orange), respectively. The data shown has been integrated by convolving with a Gaussian window of effective integration time \(\sqrt{2\pi\sigma} = 3\) \(\mu s\). In the finite-flux data of panel (c) we can immediately see 3 distinct modes with excellent separation in the log scale plots. The darkest peak (lower left), with the most counts by far, is due to the response with 0 QPs in Andreev traps. The next darkest (upper center) is from having 1 QP trapped while the lightest mode (middle right) is from 2 QPs in Andreev traps and/or excitation of a single channel into the \(|c\rangle\) state. Trapping of more than 2 QPs moves the resonance multiple linewidths and thus saturates the change in response to additional trapping, so these counts lie on top of the 2-QP distribution. We later discuss methods for avoiding this saturation. By probing
at a frequency close to the 2-QP resonant frequency, we are able to observe 0, 1, 2, and 3 QP modes, confirming that we do indeed see multiple QPs trapping and not simply the \( |e\rangle \) state of a single channel; data is shown in the supplementary material\(^7\). We have verified that these modes are indeed due to Andreev trapping of QPs by measuring the weights of each mode in the presence of a “clearing tone” at 17 GHz\(^7\). A trapped QP may absorb a photon from this tone and so be promoted back into the bulk continuum; the frequency was chosen to be greater than the trap depth and because it happens to couple efficiently into the device. We find that the tone causes the 1- and 2-QP mode counts decrease while the 0-QP mode counts increase. We have also observed that both the separations and the weights of the modes increase as a function of flux, which agrees qualitatively with a QP trapping picture\(^9\).

![Image](image1.png)

FIG. 2. [color online] (a) Ensemble measurement of resonator response at 0 flux (blue) and at \( \phi = 0.49 \) (orange). When \( \phi = 0.49 \), distinct peaks are visible roughly 0.5 and 1 MHz below resonance, corresponding to 1 and 2 trapped QPs, respectively. (b) Histogram of continuous IQ data taken at 0 flux for 30 s. Data has 10 MHz sample rate and has been convolved with a Gaussian window with effective integration time of 3 \( \mu s \). (c) Data taken at \( \phi = 0.47 \) with the same procedure as panel b. The darkest mode is due to the resonance with 0 trapped QPs. The second darkest, located near \((I,Q) = (12 \text{ mV}, 15 \text{ mV})\), corresponds to 1 trapped QP and the last mode corresponds to 2 or more trapped QPs (as this mode corresponds to the resonance moving far from the drive frequency).

We now turn to extracting the QP trap occupation as a function of time from the continuous IQ data. To optimize our analysis procedures we choose data that stresses the detector’s capabilities. We use data from a cooldown with the K&L and Eccosorb filters, which reduces QP generation by infrared radiation and thus shows a much lower QP number than that shown in Fig. 2, and use a low resonator drive power chosen to ensure the drive does not affect the QP configuration\(^22\). We increase the integration time to 5 \( \mu s \) to compensate for this loss of SNR. We first fit histograms using a Gaussian mixture expectation-maximization algorithm implemented by the Python module available from scikit-learn\(^23\). This module takes in a subset of data, assigns each point to one of the specified number of modes, then tweaks assignments and mode parameters to maximize the total likelihood for all data and all modes. The result is a set of 3 Gaussian modes describing the data, shown by their 1-\( \sigma \) (solid) and 2-\( \sigma \) (dashed) contours overlaying the histogram in Fig. 3(a). We then assign each time-series point to the mode with the highest posterior probability after Bayesian updating with a 3 sample rolling memory window to update the prior\(^24\). Unfortunately, the Gaussian mixture fit is unreliable in terms of quality of fit and reproducibility of Gaussian mode parameters, which can vary significantly even when refitting the same data with the same initial guess. This unreliability is apparent from the fit distributions shown, which do not faithfully represent the means of the 1- and 2-QP modes.

To improve the fits we need to “initialize” the trapped QP configuration, thereby isolating each Gaussian mode for independent fitting. This is challenging, as we have no direct control of the trapped QP number. Fortunately, the Gaussian mixture procedure assigns most data points to the correct occupation. We use this initial assignment to fit the mean lifetime of each mode \( \langle \tau_i \rangle \) and then identify periods in the time series when the extracted QP occupation is stationary for at least \( 4\langle \tau_i \rangle \). By stitching these “quiet periods” together, we build up large distributions of data points that are pre-assigned to modes. Each distribution is then independently fit to a Gaussian to fix its mean and covariance. Finally, the full dataset is fit to a mixture of 3 Gaussians with these same means and covariances, with the weight of each mode as the only free parameters. Fig. 3(b) shows the result, with 1-\( \sigma \) (solid) and 2-\( \sigma \) (dashed) contours overlaying the data histogram. It is immediately evident that the “quiet periods” method produces a better quality of fit.

We extract QP occupation from time series as before with updated Gaussian parameters. Fig. 3(c) shows two 0.5 ms sections of data with a 5 \( \mu s \) Gaussian convolution. These sections were chosen to demonstrate switching events and are far more “active” than typical data, which mostly stays in the 0 QP mode. The background shading represents the extracted QP trap occupation (light blue for 0 QPs, dark blue for 1, and orange for 2 or more). Transitions between the 3 configurations are clearly visible and appear to be faithfully captured by the assignment algorithm. The detector’s SNR at our operating parameters (~ 4 photon drive power, 5 \( \mu s \) integration), defined as the separation of mode centers in the IQ plane squared divided by product of their standard deviations along the line between them, is 27 for 0 − 1 distinguishability, 32 for 1 − 2, and 30 for 0 − 2. The 0 − 1 SNR gives a detector noise floor of \( 6.1 \times 10^{-4} \).
QPs /\sqrt{\text{Hz}}\), i.e., we can detect 0.00061 of the signal from a QP trapping with SNR = 1 after 0.5 s of integration, assuming a stationary occupation.

We now briefly describe the QP behavior measured with our device. We see a mean trap occupation of \(n_{\text{QP}} = 0.0185\), state occupation probabilities of \(P_0 = 0.983\), \(P_1 = 0.0155\), \(P_2 = 0.00148\), and state lifetimes of \(\tau_0 = 728 \, \mu s\), \(\tau_1 = 12.7 \, \mu s\), \(\tau_2 = 4.73 \, \mu s\). These lifetimes are corrected for the detector bandwidth assuming Poisson switching processes\(^{29,36}\). The distributions of lifetimes do appear Poissonian in the long-time limit, but better SNR (discussed below) may resolve fast events which may show non-Poisson behavior. We also note that we see transitions between all three of the 0, 1, and 2 QP modes. We attribute the \(0 \to 2\) transitions to either correlated trapping of 2 QPs in less than a detector bandwidth or to direct \(|g\rangle \to |e\rangle\) excitation of a single channel; spectroscopic measurements of the trapped QPs should be able to distinguish between these processes\(^{29}\). Future work will probe switching rates as a function of bias and environmental parameters (e.g., flux and temperature), analyze correlations between switching events, and develop more sophisticated state-assignment algorithms that do not assume independent (Poisson) switching.

We note that our device is not fully optimized for high sensitivity. While the resonant frequency was kept low in order to be less than the trap depth, raising the resonance slightly to \(\sim 6 \, \text{GHz}\) by shortening the waveguide would increase the participation ratio of the nanobridge inductance to total inductance while still remaining far below trap-clearing frequencies. Similarly, reducing the resonator’s characteristic impedance from 50 \(\Omega\) to an easily-achievable \(\sim 30 \, \Omega\) would further reduce the linear inductance, increasing sensitivity. Additionally, the TWPA used as a first-stage amplifier had a moderate \(\approx 15 \, \text{dB}\) gain (due to being operated near the edge of its bandwidth) and adds noise above the quantum limit. Adding a standard parametric amplifier with a near-quantum-limited noise temperature and 20 dB of gain as a preamplifier will further improve SNR. We may also trade off some of this sensitivity and increase the resonator bandwidth, thus allowing for detection of more than 2 QPs without saturating the response. We also note the possibility of detecting higher numbers of trapped QPs by probing the device with multiple probe tones simultaneously. By performing heterodyne measurement of probe tones centered on, e.g., the 1-, 3-, and 5-QP resonant frequencies, and correlating the measured outcomes, we should be able to detect up to 6 trapped QPs with a similar device.

In conclusion, we have developed a device for the ultra-low-noise continuous detection of up to 2 quasiparticles trapping in Andreev bound states. Our device is capable of detecting a trapped QP with SNR of 27 in 5 \(\mu s\), giving it a noise floor of \(6.1 \times 10^{-4} \, \text{QPs} / \sqrt{\text{Hz}}\). Straightforward extensions are possible to higher sensitivity and QP saturation number. Our device can be used for QP studies including statistical analysis of trapping and untrapping rates and trap occupation, spectroscopic measurements of trapped QP energy distributions, effects of environmental variables such as temperature, and testing of QP mitigation techniques.

The authors would like to acknowledge J. Aumentado and L. Glazman for useful discussions. We acknowledge MIT Lincoln Laboratory and IARPA for providing the TWPA used in this work. This work was funded by the AFOSR YIP under grant FA9550-19-1-0060 and by the NSF DMR under grant DMR-1900135. All data presented is available upon request to the corresponding author.

\(^{1}\)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)

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)

G. Catelani, J. Koch, L. Frunzio, R. J. Schoelkopf, M. H. Devoret, and L. I. Glazman, “Quasiparticle relaxation of superconducting qubits in the presence of flux,” Phys. Rev. Lett. 106, 077002 (2011)

D. Rainis and D. Loss, “Majorana qubit decoherence by quasiparticle poisoning,” Phys. Rev. B. 85, 174533 (2012)

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. Glarman, and R. J. Schoelkopf, “Measurement and control of quasiparticle dynamics in a superconducting qubit,” Nat. Commun. 5, 10.1038/ncomms6836

U. Vool, I. M. Pop, K. Stiwa, B. Abdol, C. Wang, T. Brecht, Y. Y. Gao, S. Shankar, M. Hatridge, G. Catelani, M. Mirrahimi, L. Frunzio, R. J. Schoelkopf, L. I. Glazman, and M. H. Devoret, “Non-poisonous quantum jumps of a fluxonium qubit due to quasiparticle excitations,” Phys. Rev. Lett. 113, 247001 (2014)

L. I. Glazman and G. Catelani, “Bogoliubov quasiparticle diffuseness in superconducting qubits,” SciPost Phys. Lect. Notes 3, 031 (2021)

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, 1573–1577 (2010)

K. Serniak, M. Hays, G. De Lange, S. Diamond, S. Shankar, L. D. Burkett, L. Frunzio, M. Houset, and M. H. Devoret, “Hot nonequilibrium quasiparticles in transmon qubits,” Phys. Rev. Lett. 121, 157701 (2018)

A. Somoroff, Q. Ficheux, R. A. Mencia, H. Xiong, R. V. Kuzmin, and V. E. Manucharyan, “Millisecond coherence in a superconducting qubit,” (2021) arXiv:2103.08578

A. P. Vepsäläinen, A. H. Karamlou, J. L. Orrell, A. S. Dogra, B. Loer, F. Vasmelos, D. K. Kim, J. A. Melville, B. M. Niedziewski, J. L. Yoder, S. Gustavsson, J. A. Formaggio, B. A. VanDevender, and W. D. Oliver, “Impact of ionizing radiation on superconducting qubit coherence,” Phys. Rev. B. 551–556 (2020)

M. Hays, G. de Lange, K. Serniak, D. J. van Woerkom, D. Bouman, P. Kroogstrup, J. Nygard, A. Gersdol, and M. H. Devoret, “Direct microwave measurement of andreev-bound-state dynamics in a semiconductor-nanowire josephson junction,” Phys. Rev. Lett. 121, 047001 (2018)

A. P. Vepsäläinen, A. H. Karamlou, J. L. Orrell, A. S. Dogra, B. Loer, F. Vasmelos, D. K. Kim, J. A. Melville, B. M. Niedziewski, 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, 551–556 (2020)

M. Hays, G. de Lange, K. Serniak, D. J. van Woerkom, D. Bouman, P. Kroogstrup, J. Nygard, A. Gersdol, and M. H. Devoret, “Continuous monitoring of a trapped superconducting spin,” Nat. Phys. 16, 1103–1107 (2020)

R.-P. Riwar, A. Hosseinikhani, L. D. Burkhart, 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)

J. M. Martinis, “Saving superconducting quantum processors from decay and correlated errors generated by gamma and cosmic rays:” npj Quantum Information 7, 1–9 (2021)

T. Patel, I. V. Pecheneshchik, B. A. VanDevender, M. G. Vavilov, and R. McDermott, “Phonon-mediated quasiparticle poisoning of superconducting microwave resonators,” Phys. Rev. B 96, 220501 (2017)

A. F. Andreev, “Thermal conductivity of the intermediate state in superconductors,” JETP 19, 1228 (1964).

See supplementary material.

R. Vijay, J. D. Sau, M. L. Cohen, and I. Siddiqi, “Optimizing anharmonicity in nanoscale weak-link Josephson junction oscillators,” Phys. Rev. Lett. 103, 087003 (2009)

R. Vijay, E. M. Levenson-Falk, D. H. Sliechter, and I. Siddiqi, “Approaching ideal weak link behavior with three dimensional aluminum nanobridges,” Applied Physics Letters 96, 3 (2010)

E. M. Levenson-Falk, R. Vijay, and I. Siddiqi, “Nonlinear microwave response of aluminum weak-link Josephson oscillators,” Applied Physics Letters 98, 3 (2011)

O. Dorokhov, “Transmission coefficient and the localization length of an electron in N bound disordered chains,” JETP Lett. 36, 259 (1982).

C. Macklin, K. O’Brien, D. Hover, M. Schwartz, V. Bolkovskiy, X. Zhang, W. Oliver, and I. Siddiqi, “A near-quantum-
limited josephson traveling-wave parametric amplifier,” Science 350, 307–310 (2015).

35 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” Journal of Machine Learning Research 12, 2825–2830 (2011).

36 O. Naaman and J. Aumentado, “Poisson transition rates from time-domain measurements with a finite bandwidth,” Phys. Rev. Lett. 96, 100201 (2006).

37 T. Kulik and A. Omel’Yanchuk, “Contribution to the microscopic theory of the josephson effect in superconducting bridges,” Soviet Journal of Experimental and Theoretical Physics Letters 21, 96 (1975).
I. SUPPLEMENT TO “CONTINUOUS REAL-TIME DETECTION OF QUASIPARTICLE TRAPPING IN ALUMINUM NANOBRIDGE JOSEPHSON JUNCTIONS”

A. Fabrication details

We start with a high-resistivity intrinsic silicon wafer with 5/50 nm Ti/Au markers bordering a 1 cm square chip. AZ1512 photoresist is spun on at 4000 RPM for 60 seconds, followed by a pre-exposure bake at 100°C for 60 seconds. We use shadow-mask photolithography to pattern the coplanar waveguide resonator and the chip is developed in AZ400K at room temperature for 45 seconds. Aluminum is deposited via e-beam evaporation to a thickness of 100 nm. Lift-off is performed by soaking in acetone for 30 minutes followed by a 30 second sonication.

To fabricate the junctions, we next spin the chip with MMA EL6 at 3200 RPM for 60 seconds followed by a 5 minute bake at 170°C. This results in a layer of MMA around 100 nm thick. We then spin a layer of PMMA A2 at 3600 RPM followed by a 5 minute bake at 180°C to achieve a 60 nm layer of PMMA on top. The SQUID is patterned with e-beam lithography using a beam current of ~18 pA and a 2 nm step size, with some intentionally low dose elements on one side of the pattern that define undercuts – regions where the MMA develops away but a lip of PMMA remains. The junctions are defined as 100 nm long lines bridging the two sides of the SQUID. The chip is developed in a -15°C chilled beaker of MIBK:IPA 1:3 for 60 seconds, followed immediately by drying under a strong jet of N₂. The timing is crucial as any delay between development and drying will warm the developer and result in overdeveloped features. Aluminum is deposited via e-beam evaporation to 8 nm at direct incidence, then the substrate stage is tilted to 35 degrees and another 62 nm thickness is accumulated on chip. This angled evaporation is necessary to define the 3 dimensional nanobridge as the 35 degree incidence blocks the line of sight to the substrate in the junction line while allowing it in the SQUID leads. Lift-off is performed by a 2 hour soak in acetone warmed to 45°C, followed by a 10 second sonication.

The last step is to make superconducting contact between the resonator and the SQUID. AZ1512 is spun on as before and photolithography defines a pair of rectangles for each device. This is developed in AZ400K as before. Each pair of rectangles opens a window that covers a portion of the SQUID contact pad and the resonator. The chip is then loaded into the e-beam evaporator and an end-Hall ion source running Argon is used to remove surface oxides from the SQUID and resonator portions exposed in these windows. Our recipe is 80 V at 3.5 A discharge (resulting in ~50 eV neutral Argon bombardment) at 35 degree incidence for 270 seconds. We find that this is enough to remove the native oxide layer while not etching through the thinnest (70 nm) layer of Aluminum. Without breaking vacuum, we deposit 250 nm of Al to ensure good contact between resonator and SQUID. Lift-off is performed by a 30 minute soak in acetone followed by 10 second sonication.

B. Fitting flux tuning and participation ratio

To characterize the junction properties we fit the device resonant frequency, ω₀, as a function of applied flux, φ, and extract the Josephson inductance participation ratio at zero flux, q₀. To fit this, we need to derive the flux tuning equation. We assume we have two symmetric junctions which each approximately obey the KO-1 current-phase relation (CPR) of an ideal short diffusive metallic weak link37,

\[ I(\delta) = \frac{\Delta N_e}{4\varphi_0} \cos \frac{\delta}{2} \tanh^{-1}(\sin \frac{\delta}{2}), \]  

where Δ is the superconducting gap, Ne is the effective number of channels, \( \varphi_0 \) is the reduced flux quantum, and \( \delta = \pi \phi \) is the phase bias across the junction. We note that this approximation of symmetric junctions has been shown to be valid for devices with nanobridges which appear visually similar to each other, as our device’s nanobridges do31,32. This gives a SQUID inductance,

\[ L_S(\phi) = \frac{L_{j,0}}{2} \left[ 1 - \sin \frac{\pi \phi}{2} \tanh^{-1}\left(\sin \frac{\pi \phi}{2}\right) \right], \]  

where the Josephson inductance per bridge at zero flux is

\[ L_{j,0} = \frac{8\varphi_0^2}{\Delta N_e}. \]  

With the definition of the Josephson inductance participation ratio as

\[ q(\phi) = \frac{L_S(\phi)}{L + L_S(\phi)} \]
and $q_0 \equiv q(0)$, we can rewrite the resonant frequency function as

$$\omega(\phi) = \omega_0 \left[ 1 + q_0 \frac{\sin \frac{\pi \phi}{2} \tanh^{-1}(\sin \frac{\pi \phi}{2})}{1 - \sin \frac{\pi \phi}{2} \tanh^{-1}(\sin \frac{\pi \phi}{2})} \right]^{-1/2}. \quad (S5)$$

![Graph](image)

FIG. S1. The extracted resonant frequency as a function of flux. The solid curve is the fit to the equation in the supplemental text. We extract a Josephson inductance participation ratio at zero flux of 0.01341. We see hysteretic branches in the flux tuning, as expected for a SQUID with junctions which have non-sinusoidal CPRs.

We measure $S_{11}$ as a function of applied flux to extract the resonant frequency, then fit it to the above equation. The results for the device in the main text are presented in Fig. S1. We find our Josephson participation ratio at zero flux is $q_0 \approx 0.01341$. We obtain the resonator’s linear inductance of $L = 1.83 \text{ nH}$ via finite element EM simulations. From this and the participation ratio we extract the SQUID inductance at zero flux as $L_S(0) \approx 24 \text{ pH}$, corresponding to an effective channel count $N_e \approx 660$ per junction.

We note that our 100 nm long nanobridges have CPRs which are not quite equal to the KO-1 formula. However, we have found these junctions’ CPRs are very well approximated by an ideal KO-1 junction in series with a small linear inductance. This approximation leads to an insignificant modification of the participation ratio, as this small extra linear inductance simply adds to the large resonator linear inductance, and so we neglect it. We also note that the flux tuning is hysteretic in a small range around $\phi = 0.5$. This is expected for any SQUID incorporating junctions with a non-sinusoidal CPR, and is well fit by our model. We conduct most measurements at a flux of $\phi = 0.47$ to ensure deep traps while remaining in the favored branch.

C. Power dependence

The approximately 4.3 GHz resonant frequency of our device means no single resonator photon has enough energy to excite a QP out of a deep trap state. However, multi-photon processes can occur, as there is a finite probability of multiple resonator photons for any strength of coherent drive. To find a measurement tone power that does not affect QP trapping, we take ensemble-averaged VNA measurements of the resonance as a function of probe power. Results at $\phi = 0.47$ are shown in Fig. S2. We find that at measurement powers below -141 dBm the resonator response is power-independent. As we increase power above this level, we find that the weight of the response near
the 1-QP resonant frequency begins to decrease and the weight near the 0-QP resonant frequency begins to increase. We attribute these changes to clearing of QPs from the trap states. To optimize SNR while ensuring the detector is not affecting the QP configuration, we choose a measurement power just below -141 dBm. We note that at higher measurement powers the resonator response becomes nonlinear, leading to increased sensitivity per QP at the expense of bandwidth and saturation at a lower QP number. However, in our device this nonlinear regime happens at powers which are already affecting the QP configuration, and so we neglect it in our analysis.

![Resonator Response](image)

**FIG. S2.** Ensemble-averaged resonator response at $\phi = 0.47$ as a function of measurement power. At powers below -141 dBm the response is roughly power-independent. As we increase the measurement power beyond -139 dBm, the weight of the 1-QP mode begins to decrease and the weight of the 0-QP mode increases.

D. Quiet periods Gaussian fit and occupation extraction

We first fit the IQ voltage-vs-time data using the scikit-learn Gaussian Mixture module described in the main text to find the conditional probability of a measurement outcome given a mode $P(\vec{x}|M_i)$ and the prior probability of each mode $P(M_i)$. We then run this through a Bayesian state identification algorithm (described below) to extract a rough occupation. The next task is to separate out subsets of data in which the mode (i.e. the trapped QP number) is stationary for a significant portion of time. To do this, we first fit the lifetimes of each occupation by building a histogram of the times spent in each mode. We take the portion of this histogram above the mean and fit it to an exponential to extract the apparent lifetime of the mode $\tau^*_i$. For each mode, we separate out only those subsets of data in which the mode is constant for longer than $4\tau^*_i$. We trim the first and last 5% of the lifetime from each subset before stitching them together, which should exclude any ringdown effects. This provides 3 datasets, one for each mode, where we can confidently assign the data to a certain QP occupation.

We separately histogram each data set and fit each to a 2D Gaussian distribution. Using the means and covariances of each fit as fixed values, we perform another fit of a histogram of the original full dataset to a sum of all 3 Gaussians. In this fit, the relative weights of the 3 modes are the only free parameters. This final step is of critical importance as the weights the modes ($P(M_i)$) are a significant contributor to the calculation of posterior probabilities used in extracting the occupation. Since we expect the resonator to remain in the same mode for many consecutive points in the time series, we apply a Bayesian inference scheme with a memory window of 3 points. i.e., the probability that the $k_{th}$ data point $\vec{x}_k$ belongs to the $i_{th}$ mode is

$$P(M_i|\vec{x}_k, \vec{x}_{k-1}, \vec{x}_{k-2}) = \frac{P(\vec{x}_k|M_i)P(M_i|\vec{x}_{k-1}, \vec{x}_{k-2})}{\sum_{j=0}^{2} P(\vec{x}_k|M_j)P(M_j|\vec{x}_{k-1}, \vec{x}_{k-2})}$$
This gives an apparent state lifetime $\tau$ extracted the apparent lifetimes outside the ground state following the method developed by Aumentado and Naaman.

Finally, the quasiparticle occupation at the $k_{th}$ point in the time series is taken as the mode $M_i$ with maximum $P(M_i|x_k, x_{k-1}, x_{k-2}) \forall i \in \{0, 1, 2\}$.

E. Clearing tone

As described in the main text, we expect the trap depth to be $\Delta - E_A(\phi = 0.47) \approx 10.7$ GHz. To test that our observed behavior is due to quasiparticles trapping in Andreev states, we attempt to clear these trapped quasiparticles by injecting a tone with sufficient energy to excite the trapped quasiparticles into the continuum of available states. As described in the main text, a conduction channel can take on 3 states, shown schematically in Fig. S5(a). The $A$ has a QP in the lower Andreev state and an empty upper state. The 2 degenerate modes of similar devices.

We show data in Fig. S4. The distributions of lifetimes for each of the 0-, 1-, and 2-QP modes show an exponential dependence at long times, but show a distinct peak at low times. We attribute this peak to the finite detection SNR produces occasional “blips” from the setup, couples energy efficiently into the resonator. Ensemble measurements of the reflection coefficient show that, when driven strongly, the resonance curve recovers the single Lorentzian shape that it has at 0 flux. We also see that the second and third modes apparent in the continuous IQ measurements have greatly reduced amplitudes.

F. Quasiparticle transition rates

Using our extracted QP state as a function of time, we are able to determine the lifetimes of the transition rates. We show data in Fig. S3. The distributions of lifetimes for each of the 0-, 1-, and 2-QP modes show an exponential dependence at long times, but show a distinct peak at low times. We attribute this peak to the finite detection SNR produces occasional “blips” from the setup, couples energy efficiently into the resonator. Ensemble measurements of the reflection coefficient show that, when driven strongly, the resonance curve recovers the single Lorentzian shape that it has at 0 flux. We also see that the second and third modes apparent in the continuous IQ measurements have greatly reduced amplitudes.

This frequency was chosen as it is higher than the trap depth and, due to details of our device and measurement setup, couples energy efficiently into the resonator. Ensemble measurements of the reflection coefficient show that, when driven strongly, the resonance curve recovers the single Lorentzian shape that it has at 0 flux. We also see that the second and third modes apparent in the continuous IQ measurements have greatly reduced amplitudes, indicating that QPs spend far less time in trap states as the 17 GHz to ne clears them out. Combined with the qualitative flux dependence of trapping—trapping modes move farther apart and become stronger as the flux bias becomes deeper—we take this as definitive evidence of QP trapping, similar to that shown in ensemble measurements of similar devices.

We note that the frequency of the clearing tone can be swept to spectroscopically measure the exact energies of trapped QPs. We plan such spectroscopic measurements in future studies.

G. Trapping and Excitation Mechanisms

As discussed in the main text, a conduction channel can take on 3 states, shown schematically in Fig. S5(a). The ground state $|g$ with energy $-E_A$ has a QP in the lower Andreev state and an empty upper state. The 2 degenerate
FIG. S3. Microwave response of the resonance at flux bias 0.45 as a function of clearing tone power. The legend in (a) shows the approximate power of the clearing tone at the plane of the device. The red curve is at low power and shows a significant bump on the low frequency side due to averaging over many configurations of QP occupation. At high clearing tone power (blue curve) we see the resonator response narrows, becomes more symmetric, and appears to be more over-coupled. We stress that the ratio of internal loss to external coupling is not actually changing; rather, the response is taller because we are no longer averaging over other configurations of QP occupation. (b) Histogram of time series data with the clearing tone at the same power as the red trace in (a). (c) Histogram of time series data with the clearing tone at the same power as the blue trace in (a). Note that any trapped quasiparticles are quickly excited back above the gap, greatly reducing the weights of the 1- and 2-QP modes and increasing the weight of the 0-QP mode.

⟩_o states with 0 energy either have the lower QP excited out of its state or have an extra QP trapped in the higher state. The excited state⟩_e with energy +E_A has the lower QP excited into the upper state. We note that both⟩_o states cause the same resonant frequency shift in our device, while a channel being excited to the⟩_e state has the same effect as two channels excited to their⟩_o states. As the gaps between Andreev states and between the lower Andreev state and the continuum are much larger than the gap between the upper state and the continuum, we anticipate that trapping of QPs in the upper states will be much more frequent than excitation from the lower state. However, we do see transitions between all 3 modes in our data. The 0-2 mode transitions are either the result of 2 QPs trapping in different channels (2 channels moving to their⟩_o states) in less than a detector bandwidth, or a single channel being excited to its⟩_e state. Clearing-tone measurements should be able to distinguish these processes: a clearing tone will take a single channel in⟩_e to⟩_o, causing a frequency shift equivalent to clearing 1 QP. If 2 QPs are trapped, bringing 2 channels to⟩_o, then a clearing tone will clear both and bring both channels back to⟩_g, producing twice the frequency shift. We note that the presence of a 3-QP mode, discussed below, indicates that at least 2 and likely 3 channels (and QPs) are involved. We expect higher QP numbers to be involved in an environment with a higher background QP population, e. g. one in which there is much more thermal photon radiation.
FIG. S4. Histograms of mode lifetimes for the 0- (a), 1- (b), and 2-QP (c) modes. All lifetimes are exponentially distributed at long times, but show a distinct peak at low times which we attribute to the finite detection bandwidth $\tau_{\text{det}}$. We extract the apparent lifetimes $\tau_i^*$ by fitting the behavior above a cutoff time, defined as the mean measured lifetime. These lifetimes are then adjusted to correct for the finite detector bandwidth as described in the text, giving $\tau_0 = 728 \mu$s, $\tau_1 = 12.7 \mu$s, $\tau_2 = 4.73 \mu$s.

H. Higher Number Detection

In order to confirm that we really see multiple QPs trapping and not just a single channel excited to the $\langle e \rangle$ state, we tune our device to measure the 3-QP mode. Time series data in the main text is shown with the probe tone on or near the 1-QP resonant frequency. As the frequency shift per trapped QP is larger than the resonance linewidth, trapping 2 QPs fully saturates the response of the detector and the modes representing higher trap numbers will lie on top of the 2-QP mode in the IQ plane. However, we can rectify this issue straightforwardly by decreasing the frequency of the probe tone so that it lies closer to the resonances of these higher QP modes. This allows, for instance, distinguishing 2 vs 3 QPs at the expense of our ability to distinguish 0 vs 1. We show this procedure in Fig. S5. We step the probe frequency from halfway between the 0- and 1-QP resonant frequencies (panel b), to the 1-QP resonance (panel c), to halfway between 1- and 2-QP (panel d), to the 2-QP resonance (panel e). We initially see 3 modes, which we attribute to 0, 1, and 2 trapped QPs, circled in red, blue, and green respectively. As we move the probe frequency lower, a fourth mode (circled in purple) appears, breaking away from the previously-saturated 2-QP mode. We note that this data was taken with a higher measurement power which may affect the QP configuration and which caused the resonance to be slightly nonlinear. We used this higher power because the 3-QP mode was extremely rare in this cooldown (due to the low background QP density) and so was difficult see easily in a histogram if the SNR was not very high. This will not be a concern in situations where the mean trapped QP number is higher, or if the SNR improvements discussed in the main text are made.
FIG. S5. (a) Schematic description of the ground state $|g\rangle$, 2 degenerate first excited states $|o\rangle$, and second excited state $|e\rangle$ of a channel. Both $|o\rangle$ states produce the same resonant frequency shift, while the $|e\rangle$ state produces the same shift as 2 channels entering their $|o\rangle$ states. (b-e) Response of the device to different probe tone frequencies. The probe tone is stepped to progressively lower frequencies, starting with the midpoint of the 0- (red) and 1-QP (blue) mode resonances in (b) and ending with the 2-QP mode (green) frequency in (e). As the probe moves close enough, a 3-QP mode (purple) becomes apparent.