Single quasiparticle trapping in aluminum nanobridge Josephson junctions

E. M. Levenson-Falk, 1 F. Kos, 2 R. Vijay, 1, 3 L. Glazman, 2 and I. Siddiqi 1, * 

1 Quantum Nanoelectronics Laboratory, Department of Physics, University of California, Berkeley CA 94720 
2 Department of Physics, Yale University, New Haven CT 06520 
3 Department of Condensed Matter Physics and Materials Science, Tata Institute of Fundamental Research, Mumbai 400005, India 
(Dated: October 28, 2013)

We present microwave measurements of a high quality factor superconducting resonator incorporating two aluminum nanobridge Josephson junctions in a loop shunted by an on-chip capacitor. Trapped quasiparticles (QPs) shift the resonant frequency, allowing us to probe the trapped QP number and energy distribution and to quantify their lifetimes. We find that the trapped QP population obeys a Gibbs distribution above 75 mK, with non-Poissonian trapping statistics. Our results are in quantitative agreement with the Andreev bound state model of transport, and demonstrate a practical means to quantify on-chip QP populations and validate mitigation strategies in a cryogenic environment.

Superconducting Josephson junctions, with their non-dissipative nonlinearity, form the basis of sensitive magnetometers, 1, 2 ultra-low-noise analog amplifiers, 3–5 and quantum bits. 6 The presence of quasiparticle (QP) excitations degrades or even disrupts the operation of many of these circuits. If QPs tunnel across or become trapped in a junction, they cause dissipation or excess noise that can limit qubit coherence and reduce the sensitivity of amplifier and magnetometer devices. Recent measurements have confirmed that such nonequilibrium excitations are present in superconducting circuits, even at very low temperatures, 7–9 well below the energy gap. Moreover, QP tunneling dynamics have also been studied in superconducting qubits, 10 providing valuable information on QP lifetimes.

In this letter, we quantify the trapping behavior of QPs in a resonant LC circuit consisting of a capacitively shunted nanoscale superconducting quantum interference device (SQUID). In this architecture, the Josephson junctions are not tunnel junctions but rather submicron constrictions with ≈ 700 conduction channels, each of which can be viewed as having a pair of quantized energy levels with a magnetic-flux-dependent gap. At finite flux bias, the trapping probability is large and a single trapped QP effectively neutralizes one channel, resulting in a shift of the junction kinetic inductance and correspondingly the resonant frequency of the SQUID circuit. The shift due to a single trapped QP is readily resolved provided the quality factor (Q) of the resonator is sufficiently high. We use this effect to infer the average number of trapped QPs (ntrap) and their energy distribution function. Furthermore, using microwave pulses and time domain measurements we resonantly excite the QPs, infer the gap between the trap states and the quasiparticle continuum (ΔA), and measure the associated trapping time. Our measurements are in quantitative agreement with the Andreev level picture of conduction channels, and illustrate a powerful method to characterize nonequilibrium QP populations in superconducting devices—a valuable new tool for optimizing quantum circuits.

In the semiconductor representation of a Josephson junction, the supercurrent is carried by a set of conduction channels, with the ith channel having a transmittivity τi between 0 and 1. For each channel, there is a pair of Andreev bound states with energies

\[ E^{(i)}_{\pm} = \pm \frac{\Delta}{2} \left( 1 - \tau_i \sin^2 \frac{\delta}{2} \right), \]

where δ is the gauge-invariant superconducting phase difference across the junction and Δ is the superconducting gap. 11–13 Energy spectra for τ = 1 and 0.5 are shown in Fig. 1(a). At zero temperature, the lower bound state is occupied and the upper state is unoccupied. However, for τi > 0 and δ > 0, the upper bound state energy drops below +Δ, and so forms a subgap trap state for any QPs which may exist in the bulk superconductor. The upper bound state, when occupied by a QP, carries current in the opposite direction as the lower state, and so occupation of both states leads to a total critical current of zero for that conduction channel. This “poisoning” of a conduction channel causes a shift in both the critical current and the inductance of the junction, as the poisoned channel no longer participates in supercurrent transport. Previous work has probed the trapping of QPs in few-channel quantum point contact junctions by performing switching measurements of the critical current. 14–16 In contrast, our approach realizes a dispersive measurement of the Josephson inductance of many-channel nanobridge junctions.

Our device consists of two variable-thickness aluminum nanobridge Josephson junctions arranged symmetrically in a small SQUID loop. This geometry allows us to apply a phase bias δ = φ mod 2π to both junctions, where φ ≡ Φ/Φ0 is the normalized flux through the loop in terms of the magnetic flux quantum Φ0. Fabrication details are contained in the supplementary material. The nanobridges are 8 nm thick, 100 nm long, and 25 nm wide, connecting banks which are 80 nm thick and 750 nm wide, in a
\[ \begin{align*}
\text{FIG. 1. (Color online) (a) Illustration of the Andreev bound states for channels with } & \tau = 1 \text{ (solid lines) and } \tau = 0.5 \text{ (dashed lines). The Andreev gap } \Delta A \text{ is illustrated for the } \tau = 1 \text{ channel. (b) Simplified measurement schematic. The resonator is biased with a static flux } \Phi_{dc}; \text{ measurements are performed by passing a weak probe tone (blue) through a circulator and analyzing the reflected signal. A strong excitation tone (red) may be applied via a directional coupler in order to alter the distribution of QPs in the junctions. (c) False-color SEM image of device, showing the nanoSQUID (circled), meander inductor (orange), and IDC (blue). A superconducting fast flux line, not used in this experiment, is shown off to the left in grey. (d) Detailed false-color SEM image of the nanobridge junction, showing the bridge (orange) connecting the thick banks (blue and green, highlighting the two layers of evaporation).}
\end{align*} \]

\[ \begin{align*}
\text{FIG. 2. (Color online) (a) Resonance lineshapes at different flux biases, plotted as a function of frequency detuning from the 0-QP resonance. The single Lorentzian peak present at } 0 \text{ flux begins to develop multiple broad humps as flux bias increases. (b) Temperature dependence of the resonance lineshape, showing the decrease in QP trapping as temperature rises. All data are shown at } \phi = 0.464. \text{ At the lowest temperature, multiple humps are resolvable, indicating multiple QP trapping numbers. At higher temperatures, the 2-QP and then the 1-QP peak become suppressed, leading to a single Lorentzian resonance. At even higher temperatures (T = 250 mK), the resonance broadens, shrinks, and moves to lower frequency; we attribute this to loss originating from bulk QP transport in the resonator, as the QP density becomes quite high.}
\end{align*} \]

2 × 2 µm SQUID loop. The SQUID is placed in series with a linear inductance of 1.2 nH and this combination is shunted by an interdigitated capacitor (IDC) to form a weakly nonlinear parallel LC oscillator with resonant frequency \( \omega_0 = 2\pi \times 4.72 \text{ GHz} \) at zero magnetic flux. SEM images of the device are shown in Fig. 1(c) and (d). The oscillator is isolated from the external environment by small IDCs, causing it to be critically coupled with \( Q_{ext} \approx Q_{int} \approx 2Q = 5.3 \times 10^3 \), thus giving a linewidth of 180 kHz. A simplified schematic of our measurement setup is shown in Fig. 1(b). The oscillator is enclosed in multiple layers of thermal and magnetic shielding, and attached to the base stage of a cryogen-free dilution refrigerator with a base temperature of 10 mK. Static flux bias is applied via a superconducting coil which sits underneath the small copper box housing the device. We perform microwave reflectometry, injecting power via heavily attenuated lines through a cryogenic circulator and passing the reflected signal through a semiconductor HEMT amplifier at 3 K before multiple amplification stages at room temperature. The sample temperature may be accurately controlled using a resistive heater at the base stage.

Resonance curves as a function of flux, plotted as the real part of the reflected signal, are shown in Fig. 2(a). At low flux bias \( \phi < 0.2 \), the resonance has a Lorentzian lineshape. However, at larger flux biases the resonance peak shrinks and begins to develop a second broad hump at lower frequency, indicating the development of another resonance peak; at the highest flux biases the resonance shows a second, even broader hump (resembling a long tail) at even lower frequency. This behavior is indicative of QPs trapping in the junction, raising its inductance and thus lowering the resonant frequency of the device. By averaging many resonance traces together on a timescale that is long compared to the QP dynamics (see supplementary material), we integrate over all possible configurations of QPs in the junction, weighted by their probabilities. Thus, the resonance lineshape tells us the average configuration of QPs trapped in the junction. As flux bias grows, \( E_A \) drops, and so the likelihood of a QP trapping in the Andreev excited state grows. This means that the probability that
0 QPs are trapped shrinks, and so the associated 0-QP resonance peak shrinks. In a junction with many conduction channels with different values of \( \tau \) there are a range of values of \( E_A(\tau) \), leading to differing trapping probabilities. Since the different channels have different untrapped inductances, there is a distribution of resonant frequencies with 1 trapped QP \( \omega_0(\tau) \), resulting in a broad 1-QP resonance peak. A similar argument applies to the 2-QP peak, where the range of trapping configurations is even broader.

We next measure the resonance at finite flux bias as a function of temperature; data at \( \phi = 0.464 \) is shown in Fig. 2(b). At low temperatures, 2 broad humps are resolvable in the resonance curves, indicating trapping of 1 and 2 QPs. As temperature rises within the range 10-200 mK, first the 2-QP and then the 1-QP hump are suppressed, as \( \bar{n}_{\text{trap}} \) decreases. This is to be expected, as hotter QPs are less likely to occupy lower-energy states (i.e., trap states). At higher temperatures (\( T = 200 \) mK), \( \bar{n}_{\text{trap}} \) starts to rise again due to increasing QP density in the resonator. At even higher temperatures (\( T = 250 \) mK), the resonance peak becomes shorter and broader while shifting to lower frequency, even at 0 flux. We attribute this effect to increased loss in the resonator due to bulk QP transport, as the normalized quasiparticle density \( x_{qp} \) becomes quite large at 250 mK.

We next attempt to quantitatively describe the QP trapping behavior by fitting the resonance curves. We first extract the junction inductance \( L_J(\phi) \) by fitting the flux tuning curve of the resonance to find the participation ratio \( q(\phi) \) of the junction inductance in the total inductance (see supplementary material). This procedure gives a value \( L_J(0) \approx 35 \) pH for each junction. Using the Dorokhov distribution\(^{17}\) of the channel transmittivities \( \rho(\tau) \) (applicable to diffusive nanobridges), we find the effective number of channels \( N_e = 3h^2/(2\Delta e^2 L_J) = 683 \). We define the QP configuration \( \{n_i\} \), where \( n_i = 0, 1 \) is the number of QPs in the \( i \)th channel, neglecting the possibility of having two QPs in the same channel since the number of channels is large. The resonant frequency \( \omega_0 \) and thus the response function \( S(\omega) \) of the oscillator is a function of this configuration, \( S(\omega, \omega_0(\{n_i\})) \).

We measure the response on time scales which are long compared to the QP dynamics, and thus average over all configurations weighted by their probabilities:

\[
\overline{S}(\omega) = \sum_i p(\{n_i\}) S(\omega, \omega_0(\{n_i\})).
\]  

(2)

We consider only the resonances resulting from 0-2 trapped QPs, weighted by their probabilities \( P_{0,1,2} \), with \( \bar{n}_{\text{trap}} = P_1 + 2P_2 \). Thus, the resonator response is given by

\[
\overline{S}(\omega) = P_0 S(\omega, \omega_0^{(0)}) + P_1 \int_0^1 d\tau \rho(\tau) f(E_A(\tau)) S(\omega, \omega_0(\tau)) + P_2 \int_0^1 d\tau_1 \int_0^1 d\tau_2 \rho(\tau_1) \rho(\tau_2) f(E_A(\tau_1)) f(E_A(\tau_2)) S(\omega, \omega_0(\tau_1, \tau_2)).
\]  

(3)

Here \( \omega_0^{(0)} \), \( \omega_0(\tau) \), and \( \omega(\tau_1, \tau_2) \) are the resonant frequencies with no trapped QPs, a single QP trapped in Andreev level formed by channel of transmittivity \( \tau \), and two QPs trapped in two Andreev states (channels characterized by \( \tau_1 \) and \( \tau_2 \)), respectively. We assume that the trapped QPs obey a Gibbs distribution, \( f(E_A) \propto e^{-E_A/k_B T} \). We fit the resonance with the \( P_k \) as the only free parameters; sample fits at \( T = 75 \) mK for \( \phi = 0 \) and 0.464 are shown in Fig 3(a). The theory produces excellent agreement with the data for \( \bar{n}_{\text{trap}} \lesssim 1 \), suggesting that trapped QPs follow a thermal Gibbs distribution with temperature equal to the fridge temperature of 75 mK. We note that in regimes where \( \bar{n}_{\text{trap}} \sim 1 \), the 3-QP contribution to the resonance is non-negligible.
at the few percent level; this causes the fit $P_k$ to sum to less than 1. Fitting the 3-QP resonance is computationally intensive, and so we restrict our analysis to the first two QP peaks. We note that in general the $P_k$ are not Poisson-distributed (even accounting for the neglected 3-QP peak).\textsuperscript{18}

We repeat our fitting procedure at all temperatures below 250 mK. Values of $\pi_{\text{trap}}$ between 75-200 mK are plotted in Fig. 3(b), as well as extracted values of $x_{qp}$. The 75 mK value of $x_{qp} = 1 \times 10^{-6}$ is consistent with other measurements of aluminum superconducting circuits.\textsuperscript{8,19} This indicates that our method can reproduce the results of other techniques for measuring the mean density of QPs, while also shedding light on their distribution. Although our theory fits the data well in the range 75-200 mK, we find that below 75 mK we cannot fit the data with a Gibbs distribution at the fridge temperature. There are likely two causes for this discrepancy: at low temperatures $\pi_{\text{trap}}$ grows, and so the 3-QP peak becomes significant; also, below 75 mK the QPs may not be thermally distributed, as thermalization mechanisms such as inelastic electron-phonon scattering will be suppressed at low temperatures due to the falling phonon density.\textsuperscript{20}

The overall population of nonequilibrium QPs is likely due to remnant infrared radiation leaking through our sample shielding and impacting the device. Indeed, a systematic reduction in $x_{qp}$ was observed with the addition of several layers of radiation shielding over the course of these experiments; the reported data correspond to the lowest value of $x_{qp}$ observed. This highlights the utility of this simple circuit as a tool for characterizing excess radiation leakage.

Our QP trapping theory can also be applied in a different limit, where $\pi_{\text{trap}}$ is large enough such that $P_0 \approx 0$. In this limit, one may use the saddle-point approximation to integrate over a continuous quasiparticle occupation (rather than summing discrete occupation numbers).

We can then write the resonant response as a Lorentzian centered at $\omega_0(\pi_{\text{trap}}) = \omega_0^{(0)} + \sum_i p_i \frac{\partial \omega_0}{\partial n_i}$, convolved with a Gaussian of width $\delta \omega_0^2 = \sum_i p_i (\frac{\partial \omega_0}{\partial n_i})^2$, where $p_i$ is the probability of having a QP in $i$th channel. To reach this regime, we deliberately raise $x_{qp}$ and thus $\pi_{\text{trap}}$ by heating a radiator near the sample box. Infrared radiation leaks into the box and hits the resonator, creating QPs.\textsuperscript{9} Fig. 4 shows a resonance curve at 100 mK and $\phi = 0.432$, with a theory fit (dashed line) giving $\pi_{\text{trap}} = 7$. Also plotted for comparison is a resonance at zero flux, which is well-fit by a Lorentzian (indicating $\pi_{\text{trap}} = 0$). The best fit to the $\phi = 0.432$ data (solid line) is obtained by allowing the Gaussian width to vary independently from the Lorentzian centerpoint; extracting QP number from each gives $\pi_{\text{trap}} = 19$ and 7, respectively. The discrepancy is likely due to the limitations of the short-channel approximation accepted in Eq. (1);\textsuperscript{21,22} it may also be related to the non-Poissonian nature of the low-$\pi_{\text{trap}}$ data, since the Gaussian is just the high-$\pi_{\text{trap}}$ limit of Poisson statistics.

By illuminating the junction with a microwave tone of frequency $f_{esc} > \Delta_A/h \equiv (\Delta - E_A(\delta, \tau))/h$, it is possible to promote a QP out of its Andreev trap and back up into the continuum, thus restoring the relevant conduction channel to its original properties (see Fig. 1(a)). When such an excitation tone is on, the resonant response retains its Lorentzian lineshape at all flux biases, with only a small decrease in the 0-QP resonance peak at high flux bias. We attribute this small decrease to an imperfect excitation efficiency, as the QP trapping probability becomes larger at large phase biases, but the microwave excitation power (i.e. QP excitation rate) is limited, as an excitation tone at very high power will begin to bias the junctions into nonlinearity and shift the resonance. The fact that the resonance lineshape remains constant as a function of flux indicates that when no QPs are trapped other effects such as flux noise and phase slips in the junctions are negligible.

In order to probe the energy spectrum of trapped QPs, we sweep the frequency of the applied microwave excitation tone and measure the response of the resonator as a function of flux (i.e. phase bias). Spectroscopy data at 100 mK is shown in Fig. 5, plotted as the real part of the reflected probe signal at the 0-QP resonant frequency minus its value with the bias tone off; a positive shift indicates a greater probability of 0 QPs, and thus shows that trapped QPs have been cleared from the junction. At a given flux bias, a minimum excitation tone frequency is required to observe any response; this frequency grows with increasing flux bias, and has a value consistent with
\[ \Delta A(\phi, \tau) \text{ for } \tau \approx 0.8 - 1. \] At high excitation frequencies \( (f_{\text{exc}} > \Delta A(\phi, 1)) \), the response saturates. The width of this transition from no response to saturation is consistent with the Dorokhov distribution for channel transmission. Finally, we note that the saturated response grows with flux bias, as the 0-QP resonance with no bias becomes more suppressed and thus the effect of the excitation tone becomes more pronounced.

\[ \Delta f_{\text{A}} > f_{\text{exc}} \]

FIG. 5. (Color online) Spectroscopy data, showing the shift in the 0-QP resonance peak height versus flux \( \phi \) and excitation tone frequency \( f_{\text{exc}} \), referred to its height with the excitation tone off. The data show a threshold excitation frequency which grows with flux, consistent with \( \Delta A(\phi) \), and a height shift which grows with flux, indicating a growing \( \overline{n}_{\text{trap}} \).

By pulsing the excitation tone and monitoring the response of the 0-QP resonance, it is possible to probe the dynamics associated with QP trapping and excitation. Data at base temperature are shown in the supplementary material; during both trapping and excitation the resonance evolves exponentially in time towards steady state. The excitation time scales inversely with excitation tone power and is relatively constant as a function of flux. Typical values range from 1 to 5 \( \mu s \) for the excitation powers used. The retrapping time shortens as flux increases, ranging from 60 to 15 \( \mu s \). The scaling as a function of flux appears consistent with electron-phonon relaxation being the dominant mechanism for QP trapping, although the basic theory does not correctly predict the order of magnitude of the retrapping time (see supplementary material for more details).

In conclusion, we have measured ensemble QP trapping in a phase-biased nanobridge. By using a dispersive measurement of a narrow linewidth resonator, we are able to resolve single QPs trapped in junctions with \( \sim 700 \) channels, while keeping the junction in the superconducting state at fixed phase bias. The trapped QPs obey a thermal Gibbs distribution for temperatures above 75 mK, both in the limit of low and high average trapped QP number. Applying a microwave bias tone to the junction results in a frequency-dependent clearing of trapped QPs, with a retrapping time ranging from 15-60 \( \mu s \). Future work can further probe the mechanisms of QP trapping and thermalization and investigate any correlations between trapping/untrapping events in nearby junctions. By engineering the parameters of the resonant circuit, it should be possible to achieve continuous single-shot measurement of the QP configuration. Finally, we note that the number of trapped QPs is a sensitive probe of the QP density, thus allowing our device to be used to evaluate the quality of the radiation shielding used in a cryogenic setup.

The authors acknowledge N. Roch, Ç. Girit, G. Catelani, M. Devoret, A. Eddins, S. Hacohen-Gourgy, and R. Schoelkopf for useful discussions. Funding was provided in part by an NSF GRFP grant (EMLF) at UC Berkeley, by AFOSR under Grant FA9550-08-1-0104, by DOE contract DE-FG02-08ER46482 (theory of resonance line shapes), and by IARPA via the Army Research Office W911NF-09-1-0369 (analysis of experimental data) at Yale University.
[15] L. Bretheau, Ç. Girit, L. Tosi, M. Goffman, P. Joyez, H. Pothier, D. Esteve, and C. Urbina, Comptes Rendus Physique 13, 89 (2012).
[16] L. Bretheau, Ç. O. Girit, H. Pothier, D. Esteve, and C. Urbina, Nature 499, 312 (2013).
[17] O. Dorokhov, JETP Lett 36, 318 (1982).
[18] For $T = 100$ mK the measured values are $P_0 = 0.66$, $P_1 = 0.16$, $P_2 = 0.12$, while the Poisson distribution $P_k = \bar{n}_{trap}^k \exp(-\bar{n}_{trap})/k!$ gives $P_0 = 0.66$, $P_1 = 0.26$, $P_2 = 0.06$.
[19] 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, Phys. Rev. B 84, 024501 (2011).
[20] A. Zazunov, V. S. Shumeiko, G. Wendin, and E. N. Bratus’, Phys. Rev. B 71, 214505 (2005).
[21] R. Vijay, J. D. Sau, M. L. Cohen, and I. Siddiqi, Phys. Rev. Lett. 103, 087003 (2009).
[22] R. Vijay, E. M. Levenson-Falk, D. H. Slichter, and I. Siddiqi, Applied Physics Letters 96, 3 (2010).
Supplement to “Single quasiparticle trapping in aluminum nanobridge Josephson junctions”
FABRICATION DETAILS

The device is fabricated using conventional double-angle shadow-mask evaporation and a liftoff process. First, a two-layer PMMA/copolymer resist stack is spun onto a silicon substrate. The nanobridge resonator structure is then defined lithographically in an SEM running at 30 keV. The SQUID pattern is a 750 nm wide, $2 \times 2 \mu m$ square washer, with two 100 nm gaps spanned by the 25 nm wide nanobridges. After development in a solution of 1 part methyl isobutyl ketone to 3 parts isopropanol at $-15^\circ C$, the sample is put in an electron-beam evaporator with a base pressure of $\sim 10^{-8}$ Torr. The evaporation is performed in two stages. First, 8 nm of aluminum are evaporated at normal incidence, depositing metal on the substrate everywhere exposed by the SEM beam. Next, the sample is tilted in-situ to an angle of $36^\circ$. This causes the nanobridge region to be occluded, as there is no longer a line-of-sight from the substrate to the evaporation source through the upper resist layer due to the steep aspect ratio of the nanobridge resist window. We evaporate 80 nm of aluminum in this configuration, defining banks which are $8 + 80 \cos(36^\circ) = 73$ nm thick, connected by the 8 nm bridge. Finally, the resist is lifted off by soaking in acetone, leaving only the resonator with variable-thickness bridge defined.

EXTRACTING $q_0$ AND $L_J$

We model the nanobridge SQUID as two identical nanobridge junctions in parallel. The resonant frequency of the device is given by $\omega_0 = (C(L + L_J/2))^{-1/2}$, where $C$ is the total capacitance, $L_J$ is the inductance of a single nanobridge, and $L$ is the linear inductance of the circuit. Defining the junction participation ratio $q$ by

$$q(\phi) \equiv \frac{L_J(\phi)/2}{L + L_J(\phi)/2}$$

$$q_0 \equiv q(\phi = 0),$$

one can write

$$\omega_0(\phi) = \omega_0(0)[1 + q_0 \frac{L_J(\phi) - L_J(0)}{L_J(0)}]^{-1/2}.$$  \hspace{1cm} (3)

In order to fit $q_0$, we assume a KO-1 current-phase relation (CPR) for the nanobridges,\textsuperscript{31} thus giving

$$\omega_0(\phi) = \omega_0(0)[1 + q_0 \frac{\sin \frac{\phi}{2} \tanh^{-1} \sin \frac{\phi}{2}}{1 - \sin \frac{\phi}{2} \tanh^{-1} \sin \frac{\phi}{2}}]^{-1/2}.$$  \hspace{1cm} (4)
Here, we use $\delta = \pi \phi$. Fitting the flux tuning of the resonant frequency with this equation gives $q_0 = 0.015$. We performed EM simulations of the oscillator structure in Microwave Office using the AXIEM solver and found good agreement with experimentally measured quantities. These simulations gave $L = 1.2 \, \text{nH}$, and so we find $L_J(0) = 2Lq_0/(1 - q_0) = 36 \, \text{pH}$.

**THEORETICAL MODEL**

We consider a weak link with $N_e$ channels, i.e. $N_e$ Andreev states. Let $n_i = 0, 1$ denote the number of quasiparticles (QPs) trapped in the $i$-th channel. For a given configuration $\{n_i\}$ of QPs, we can find the resonant frequency of the circuit, $\omega_0(\{n_i\})$. We denote the frequency-dependent response function as $S(\omega, \omega_0(\{n_i\}))$, since it depends on both $\omega$ and resonant frequency, which is a function of QP configuration. In the experiment we measure the average value of the response function, $\bar{S}(\omega)$:

$$\bar{S}(\omega) = \sum_{\{n_i\}} p(\{n_i\}) S(\omega, \omega_0(\{n_i\})).$$

(5)

where $p(\{n_i\})$ is the probability to find the system in the configuration $\{n_i\}$. We assume it can be written as a product $p(\{n_i\}) = \prod_i p_i(n_i)$. The average number of QPs is then $\bar{n}_{\text{trap}} = \sum_i n_i$, where $p_i = p_i(n_i = 1)$. Assuming $p_i \ll 1$, the probability to have zero QPs is $e^{-\bar{n}_{\text{trap}}}$. The expression for $\bar{S}(\omega)$ can be written as a convolution:

$$\bar{S}(\omega) = \int d\Omega F(\Omega) S(\omega, \Omega),$$

(6)

where

$$F(\Omega) = \sum_{\{n_i\}} p(\{n_i\}) \delta(\Omega - \omega_0(\{n_i\})).$$

(7)

We distinguish two different cases – large average number of QPs ($\bar{n}_{\text{trap}} \gg 1$) and small average number of QPs ($\bar{n}_{\text{trap}} \lesssim 1$). In the case of large average number of QPs, we evaluate $F(\Omega)$ using the saddle point approximation as follows:

$$F(\Omega) = \int \frac{d\alpha}{2\pi} \sum_{\{n_i\}} p(\{n_i\}) e^{i\alpha(\Omega - \omega_0(\{n_i\}))} = \int \frac{d\alpha}{2\pi} \sum_{\{n_i\}} p(\{n_i\}) \exp[i\alpha(\Omega - \omega_0^{(0)} - \sum_i \frac{\partial \omega_0}{\partial n_i} n_i)].$$

(8)
Here we have expanded $\omega_0(\{n_i\})$ around $\omega_0^{(0)}$, corresponding to no QPs, since each QP only slightly shifts the resonant frequency. Writing the probability $p(\{n_i\})$ as a product of $p_i$’s we get:

$$F(\Omega) = \int \frac{d\alpha}{2\pi} \exp[i\alpha(\Omega - \omega_0^{(0)}) + \sum_i \ln(1 - p_i + p_i e^{-i\alpha \partial \omega_0 / \partial n_i})].$$ \hspace{1cm} (9)

The saddle point approximation of this integral then gives a Gaussian for $F(\Omega)$:

$$F(\Omega) \propto \exp \left[ -\frac{(\Omega - \omega_0^{(0)}) - \sum_i p_i \partial \omega_0 / \partial n_i}{2 \sum_i p_i (\partial \omega_0 / \partial n_i)^2} \right],$$ \hspace{1cm} (10)

yielding the Gaussian referred to in the main text with center frequency $\omega_0 = \omega_0^{(0)} + \sum_i p_i \partial \omega_0 / \partial n_i$ and width $\delta \omega_0^2 = \sum_i p_i (\partial \omega_0 / \partial n_i)^2$. For $S(\omega, \Omega)$ we assume a Lorentzian lineshape:

$$S(\omega, \Omega) = \frac{\Gamma / \pi}{(\omega - \Omega)^2 + \Gamma^2},$$ \hspace{1cm} (11)

where the width $\Gamma$ is easily measured from the resonance at zero flux. Thus, in the case of large number of QPs, the resonance is a convolved Gaussian-Lorentzian with its maximum shifted from the 0-QP peak. The 0-QP peak is suppressed by a factor $e^{-\bar{n}_{trap}}$ and thus is too small to be observed. In the case of a diffusive nanobridge we can find the exact expression for the inductance $L_J$ with a general distribution of QPs. The channel transmittivities $\tau$ in a diffusive nanobridge are given by the Dorokhov distribution:

$$\rho(\tau) = \frac{\pi \hbar G}{2e^2} \frac{1}{\tau \sqrt{1 - \tau}},$$ \hspace{1cm} (12)

where $G$ is the normal-state conductance of the nanobridge. The number of channels is not well defined for this distribution, but we can define the effective number of channels as

$$N_e = \left( \frac{\int_0^1 d\tau \rho(\tau) \tau}{\int_0^1 d\tau \rho(\tau) \tau^2} \right)^2 = 3\pi \hbar G / 2e^2.$$ \hspace{1cm} (13)

Assuming that the quasiparticles in the nanobridge are distributed according to a given distribution $f(E)$, the inductance is given by:

$$1 / L_J = \int_0^1 d\tau \rho(\tau) e^2 \Delta \tau \frac{\cos \delta + \tau \sin \frac{\delta}{2}}{\hbar^2} \frac{\tau^4}{(1 - \tau \sin^2 \frac{\delta}{2})^{3/2}} [1 - 2f(E_A(\delta, \tau))],$$ \hspace{1cm} (14)

where $E_A(\delta, \tau)$ is the Andreev energy of a channel with transmittivity $\tau$ (see Eq. (1) in the main text). It is clear from Eq. (14) that the presence of QPs changes the expectation value
of the inductance. Similarly, fluctuations of the number or distribution of trapped QPs cause fluctuations of $L_J$. Since the resonant frequency depends on the inductance through Eq. (3), QPs therefore cause a resonant frequency shift and broaden the resonance. Assuming $f(E)$ is small and given by the Gibbs distribution with some low temperature $k_B T \ll |\Delta - \Delta \cos \frac{\delta}{2}|$, we get the following expressions:

$$
\omega_0^{(0)} - \omega_0 = \omega_0^{(0)} g(\delta) \frac{\bar{n}_{\text{trap}}}{N_e}, \quad \delta \omega_0^2 = \frac{1}{4} (\omega_0^{(0)})^2 g^2(\delta) \frac{\bar{n}_{\text{trap}}}{N_e^2},
$$

(15)  

$$
g(\delta) \equiv \frac{3q(\delta)}{2} \frac{\left| \cos \frac{\delta}{2} \right|}{1 - \sin^2 \frac{\delta}{2} \tanh^{-1} \sin \frac{\delta}{2}}.
$$

(16)  

All quantities appearing in the above expressions are measured except $\bar{n}_{\text{trap}}$. Therefore, we may fit the resonance to the convolved Gaussian-Lorentzian described above with $\bar{n}_{\text{trap}}$ as the only fitting parameter. However, we find out that such a procedure does not fit the data well. Thus, we take the frequency shift and the width as independent fitting parameters, and extract $\bar{n}_{\text{trap}}$ from the best fit value of each of the parameters using Eq. (15). As explained in the main text, we find that the value of $\bar{n}_{\text{trap}}$ extracted from the width of the resonance is about a factor of 3 greater than the value $\bar{n}_{\text{trap}}$ extracted from the shift of the resonance.

In the case of small number of QPs ($\bar{n}_{\text{trap}} \lesssim 1$) the main resonance peak comes from the 0-QP resonance, which is suppressed by a factor $e^{-\bar{n}_{\text{trap}}} \lesssim 1$. Additional contributions to the response function come from the 1-QP and 2-QP resonances, while the multi-QP contributions are further suppressed. We can thus keep only terms corresponding to 0, 1, and 2 QPs in the expression for $F(\Omega)$, Eq. (7):

$$
F(\Omega) = e^{-\bar{n}_{\text{trap}}} \sum_i p_i \delta(\Omega - \omega_0^{(0)}) + \sum_i p_i p_j \delta(\Omega - \omega_0^{(i,j)}),
$$

(17)  

where $\omega_0^{(0)}$, $\omega_0^{(i)}$, and $\omega_0^{(i,j)}$ are the resonant frequencies at zero, one, and two trapped QPs, respectively; $i$ and $j$ label channels.

We write the resonant frequency with a QP trapped in a channel with transmittivity $\tau^{(i)}$ and energy $E_A^{(i)}$ as

$$
\omega_0^{(i)} \equiv \omega_0(\tau^{(i)}) = \omega_0^{(0)} - \frac{q}{2} \omega_0^{(0)} \cdot L_j^{(0)} \Delta \frac{1}{L_j^{(i)}}, \quad \Delta \frac{1}{L_j^{(i)}} = \frac{2e^2 \tau_i \Delta \left( \cos \delta + \tau_i \sin^4 \frac{\delta}{2} \right)}{\hbar^2 (E_A^{(i)}) / \Delta^3}.
$$

(18)  

Here $\tau_i$ and $\Delta \frac{1}{L_j^{(i)}}$ are the transmittivity of the $i$th channel and its contribution to the inductance, respectively.
Using Eq. (18) in expression (17) and transforming the sums into integrals by averaging over the distribution of transmission coefficients, we find

\[ F(\Omega) = e^{-\bar{n}_{\text{trap}}}[\delta(\Omega - \omega_0(0)) + \int_0^1 d\tau \rho(\tau)p(E_A(\tau))\delta(\Omega - \omega(\tau))] \]

\[ + \int_0^1 d\tau_1 \int_0^1 d\tau_2 \rho(\tau_1)\rho(\tau_2)p(E_A(\tau_1))p(E_A(\tau_2))\delta(\Omega - \omega(\tau_1, \tau_2))] \]. (19)

We assume a Gibbs distribution for the QPs:

\[ p(E_A(\tau)) \equiv \bar{n}_{\text{trap}}f(E_A(\tau)) = \bar{n}_{\text{trap}} Ne^{-E_A(\tau)/k_B T}, \] (20)

where the normalization factor \( N \) is determined from the condition \( \int d\tau \rho(\tau)p(E_A(\tau)) = \bar{n}_{\text{trap}} \). Finally, convolving the expression for \( F(\Omega) \) with the Lorentzian lineshape given in Eq. (11), we find the response function (see Eq. (2) in the main text):

\[ \mathcal{S}(\omega) = P_0 S(\omega, \omega_0(0)) + P_1 \int_0^1 d\tau \rho(\tau)f(E_A(\tau))S(\omega, \omega_0(\tau)) \]

\[ + P_2 \int_0^1 d\tau_1 \int_0^1 d\tau_2 \rho(\tau_1)\rho(\tau_2)f(E_A(\tau_1))f(E_A(\tau_2))S(\omega, \omega_0(\tau_1, \tau_2)) \], (21)

where \( P_k \) is the probability of having exactly \( k \) trapped QPs. Our data typically do not agree with the Poisson-distributed values of \( P_k = \bar{n}_{\text{trap}}^k \exp(-\bar{n}_{\text{trap}})/k! \) that are expected for rare, independent trapping events. Because of this, we allow the \( P_k \) to vary independently, improving the fits. We may also use temperature as an additional fitting parameter, but even allowing \( T \) to vary freely gives \( P_k \) which are not Poisson-distributed.

**RETRAPPING TIME MEASUREMENTS**

We probe the resonator at its 0-QP resonant frequency with a constant tone. We mix the reflected tone down to dc with an IQ demodulation setup and measure the two quadratures with a fast digitizer circuit. We then pulse the QP excitation tone at a frequency which is significantly above \( \Delta_A \). This removes some percentage of the QPs out of the junction trap states. We test the pulse’s efficacy by increasing its power, and determine that at the power at which the response saturates there are no trapped QPs. Fig. S1(a) shows a sample
FIG. 1. (Color online) (a) Representative measurement of QP excitation and retrapping at $\phi = 0.476$. The shaded region indicates the time during which the excitation pulse is on. We fit the exponential rise and fall to extract excitation and retrapping time constants, respectively. (b) QP excitation (top) and retrapping (bottom) time constants as a function of flux. The excitation time is roughly constant as a function of flux for a given power, while the retrapping time decreases with $\phi$, in qualitative agreement with the theoretical prediction of electron-phonon relaxation (solid blue line).

The pulse response signal was too small to accurately measure.

We attempt to fit the retrapping time by assuming electron-phonon relaxation as the
dominant mechanism. The crude model we use disregards the influence of disorder on the electron-phonon interaction, and the effect of bridge geometry. The electron-phonon interaction Hamiltonian then takes the form:

\[ H_{e-ph} = \frac{1}{\sqrt{N}} \sum_{k,q,\sigma} \alpha q^{1/2} (a_{k+q,\sigma}^\dagger b_q + a_{k,\sigma}^\dagger a_{k+q,\sigma} b_q^\dagger), \]  

where \( a \) and \( b \) are the annihilation operators and \( k \) and \( q \) are the momenta for electrons and phonons, respectively. We want to relate the retrapping time in our experiment, \( \tau_T \), to the QP recombination time \( \tau_R \) in a bulk superconductor that was measured in earlier experiments. The recombination rate of a QP with momentum \( p \) in the superconductor is given by Fermi’s Golden Rule as:

\[ w_R(p) = \frac{2\pi}{N} \sum_{p_2} \sum_q \alpha^2 q |u_p v_{p_2} + u_{p_2} v_p|^2 (N_q + 1) f_{p_2} \delta_{p+p_2,q} \delta(E_p + E_{p_2} - \omega_q), \]  

where \( u_p \) and \( v_p \) are Bogoliubov amplitudes of a QP with momentum \( p \) and energy \( E_p \), and \( N_q, f_p \) are distribution functions of phonons and QPs, respectively. In the case where the QP energy is close to the gap, \( E_p \approx \Delta \), and the QPs are Gibbs-distributed at a low temperature \( k_B T \ll \Delta \), the integrals in the above expression for the recombination rate can be evaluated to give:

\[ w_R = 4\pi x_{qp} b \Delta^3, \]  

where \( b = 2V \alpha^2 / \pi^2 N s^3 v_F \), with \( s \) and \( v_F \) representing the speed of sound and Fermi velocity, respectively. The recombination time is defined as \( \tau_R = 1/w_R \).

Next we consider the single-channel point contact with one Andreev state \( \alpha_{A\sigma}^\dagger |0\rangle \), which can be expressed in terms of the electron operators as:

\[ \alpha_{A\sigma} = \sum_k \left( u_{Ak}^L a_{k\sigma}^L + v_{Ak}^L a_{k-\sigma}^L + u_{Ak}^R a_{k\sigma}^R + v_{Ak}^R a_{k-\sigma}^R \right). \]  

Indices \( L \) and \( R \) denote left and right lead, respectively. The retrapping rate \( w_T \) can be written as \( w_T = \sum_\mu w_{\mu \rightarrow A} f(E_\mu) \), where the sum goes over all single-QP states and \( f(E_\mu) \) is the initial distribution function of QPs. The transition rate from a given QP state \( \mu \) to the Andreev level is given by Fermi’s Golden Rule as:

\[ w_{\mu \rightarrow A} = 2\pi \sum_q |\langle 0 | \alpha_{A\sigma}^\dagger b_q H_{e-ph} \alpha_{\mu\sigma}^\dagger |0\rangle|^2 \delta(E_\mu - E_A - \omega_q). \]
The retrapping rate can then be expressed in the form:

\[ w_T = -\frac{i}{N} \sum_q q^2 f(E_A + \omega_q) \sum_{i,j,k_1,k_4} [u_{Ak_1}^{i*} u_{Ak_4}^{j} G_{ij}(k_1 - q, k_4 - q; E_A + \omega_q) \]
\[ + v_{Ak_1}^{i*} v_{Ak_4}^{j} G_{ij}(k_1 + q, k_4 + q; -E_A - \omega_q) + v_{Ak_1}^{i*} u_{Ak_4}^{j} F_{ij}^+(k_1 + q, k_4 - q; E_A + \omega_q) \]
\[ + u_{Ak_1}^{i*} v_{Ak_4}^{j} F_{ij}(k_1 - q, k_4 + q; E_A + \omega_q)] \]  

(27)

Here \( G_{ij} = G_{ij}^R - G_{ij}^A \) and \( F_{ij}^\pm = F_{ij}^{\pm R} - F_{ij}^{\pm A} \) are electron-electron Green’s functions in the point contact. The indices \( i, j \) in the above sum are going over the values \( L, R \). Describing the point contact with transmittivity \( \tau \) using the tunneling Hamiltonian,\(^{84}\) it is possible to find the exact expressions for the Green’s functions and the Bogoliubov amplitudes of the Andreev states. However, the integrals in Eq. (27) are difficult to evaluate for arbitrary \( \tau \), so we only consider the limit of a shallow Andreev level, in which case the amplitudes are given by:

\[ |u_{Ak}^{L,R}|^2 = \frac{(\Delta^2 - E^2_A)^{3/2}}{2\pi \nu_0(E^2_k - \Delta^2)^2}, \]  

(28)

\[ v_{Ak}^{L} = e^{i\delta/2} u_{Ak}^{L} = e^{-i\delta/2} u_{Ak}^{R*} = v_{Ak}^{R*}, \]  

(29)

where \( \nu_0 \) is the normal-state density of states at Fermi level. For the crude estimate of the retrapping time, we use the bulk superconductor Green’s functions. Taking as the initial distribution of QPs a Gibbs distribution at a low temperature \( k_B T \ll \Delta - E_A \), we can evaluate Eq. (27):

\[ w_T = \pi x_{qp} b\Delta(\Delta - E_A)^2. \]  

(30)

Defining the retrapping time as \( \tau_T = 1/w_T \) and comparing with Eq. (24), we get:

\[ \tau_T = \left( \frac{2\Delta}{\Delta - E_A} \right)^2 \tau_R. \]  

(31)

A fit to the relaxation time measurements using this functional form, with \( E_A(\phi) = \Delta|\cos \frac{\phi \pi}{2}| \), is shown in Fig. S1(b). The functional form derived for \( \tau_T(\phi) \) agrees reasonably with the measurements. However, the value \( \tau_R = 0.3 \mu s \) which we extract with the help of Eq. (31) is much lower than the value \( \tau_R = 100 \mu s \) reported in earlier experiments at 250mK.\(^2\) We note, however, that our derivation of Eq. (31) was very crude; we assumed the clean superconductor limit, did not take into account the geometry of the junction, and considered only the limit of a shallow Andreev level. A more complete theoretical treatment may be able to resolve the discrepancy.
[S1] I. O. Kulik and A. N. Omelyanchuk, JETP Lett. 21, 96 (1975).

[S2] C. M. Wilson and D. E. Prober, Phys. Rev. B 69, 094524 (2004).

[S3] S. B. Kaplan, C. C. Chi, D. N. Langenberg, J. J. Chang, S. Jafarey, and D. J. Scalapino, Phys. Rev. B 14, 4854 (1976).

[S4] F. Kos, S. E. Nigg, and L. I. Glazman, Phys. Rev. B 87, 174521 (2013).