

# Supplemental Material for “Observation of classical-quantum crossover of $1/f$ flux noise and its paramagnetic temperature dependence”

C. M. Quintana,<sup>1</sup> Yu Chen,<sup>2</sup> D. Sank,<sup>2</sup> A. G. Petukhov,<sup>3</sup> T. C. White,<sup>2</sup> Dvir Kafri,<sup>4</sup> B. Chiaro,<sup>1</sup> A. Megrant,<sup>2</sup> R. Barends,<sup>2</sup> B. Campbell,<sup>1</sup> Z. Chen,<sup>1</sup> A. Dunsworth,<sup>1</sup> A. G. Fowler,<sup>2</sup> R. Graff,<sup>2</sup> E. Jeffrey,<sup>2</sup> J. Kelly,<sup>2</sup> E. Lucero,<sup>2</sup> J. Y. Mutus,<sup>2</sup> M. Neeley,<sup>2</sup> C. Neill,<sup>1</sup> P. J. J. O’Malley,<sup>1</sup> P. Roushan,<sup>2</sup> A. Shabani,<sup>4</sup> V. N. Smelyanskiy,<sup>4</sup> A. Vainsencher,<sup>2</sup> J. Wenner,<sup>1</sup> H. Neven,<sup>4</sup> and John M. Martinis<sup>1,2</sup>

<sup>1</sup>*Department of Physics, University of California, Santa Barbara, California 93106, USA*

<sup>2</sup>*Google Inc., Santa Barbara, California 93117, USA*

<sup>3</sup>*NASA Ames Research Center, Moffett Field, California 94035, USA*

<sup>4</sup>*Google Inc., Venice, CA 90291, USA*

(Dated: December 6, 2016)

## QUASISTATIC FLUX NOISE MEASUREMENT

We refer to noise slower than the experimental repetition rate as quasistatic. To measure the quasistatic flux noise in  $\Phi_t^x$  (i.e., noise in  $\epsilon$ ), we use a pulse sequence similar to that used in Ref. [1] combined with the signal processing techniques used in Refs. [2, 3]. This allows us to obtain data both well below and well above 1 Hz, the latter being achieved by directly processing a binary sequence of single-shot measurements.

The measurement works as follows. We treat each experimental repetition as if it had a static flux offset, and repeatedly measure a function  $f(\Phi_t^x(t))$  that is sensitive to fluctuations in  $\Phi_t^x$  but not to fluctuations in  $\Phi_{\text{ba}}^x$ . We do this by performing the pulse sequence illustrated in Fig. 1(a). We initialize the qubit in its ground state (by energy relaxation) in the single-well regime at zero tilt. We then symmetrically raise the barrier so that in the absence of noise in  $\Phi_t^x$  there would be probability 0.5 to end up in the left or right well, completely uncorrelated with any previous or future measurement. Deviations from  $P = 0.5$  correspond to deviations away from zero tilt in  $\Phi_t^x$ . We calibrate this experiment by measuring  $P_{|R\rangle}$  (the probability of ending up in the right well) as a function of applied tilt bias  $\Phi_t^x$ , as shown in Fig. 1(b). Since we accurately know the applied flux, this curve gives a direct calibration  $P_{|R\rangle}(\Phi_t^x)$  between physical flux and probability, which (as detailed below) can be used to convert between probability noise and flux noise as long as the flux excursions are small enough that they remain on the linear part of the curve (close to  $P_{|R\rangle} = 0.5$ ). The exact functional form of this curve depends on the ramp rate due to factors including non-adiabatic transition and thermalization, but it is possible to choose a ramp rate such that the flux fluctuations remain within the linear part of the curve. We verify this by looking at the raw measurement statistics of the noise measurement when nominally parked at  $P_{|R\rangle} = 0.5$ . For example, Fig. 1(c) gives a histogram of probabilities obtained by averaging every 100 consecutive single-shot samples, along with a Gaussian fit. We also checked experimentally that the

quasistatic sensitivity to the barrier bias,  $dP_{|R\rangle}/d\Phi_{\text{ba}}^x$ , was less than  $\frac{1}{100} \cdot dP_{|R\rangle}/d\Phi_t^x$ , which is certainly negligible upon taking the square when comparing the relative contributions of incoherent flux noise.

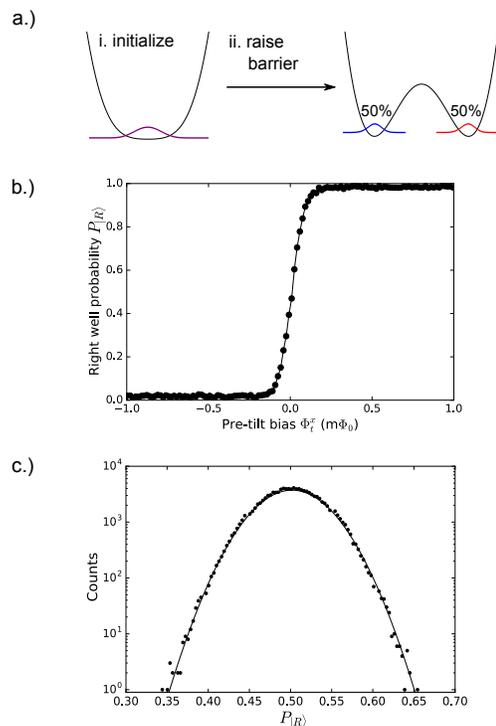


Figure 1. (a) Illustration of pulse sequence for the quasistatic flux noise measurement. (b) Calibration curve giving probability to end up in the right well as a function of external tilt flux bias. (c) Semi-log histogram of probabilities (obtained by averaging every 100 consecutive single-shot measurements) showing a Gaussian distribution with a standard deviation small enough to be within the linear part of the flux-probability curve.

For flux noise data below  $\sim 0.1$  Hz, we measure a time series of  $P_{|R\rangle}$  (each value being the average of a few hundred consecutive stats) over a total period of  $\sim 24$  hours and use the same signal processing techniques as in Ref. [2] to obtain  $S_{\Phi_t^x}^+(f)$ . For the data above  $\sim 0.1$  Hz, we instead obtain  $S_{\Phi_t^x}^+(f)$  by processing time series of  $N \approx 1$

million single-shot measurements taken with a regularly spaced sampling interval  $\delta t$  (ranging from 10 – 100  $\mu\text{s}$ ) and use a variant of the signal processing techniques used in Ref. [3]. We refer to this method as a “1-bit detector measurement” because it involves keeping all single-shot measurement results without explicitly computing probabilities. Given an underlying sequence of flux fluctuations  $\{\delta\Phi_n\}$  in the qubit, we obtain a finite probabilistic binary sequence  $\{x_n\}$  of length  $N$ , with  $x_n \in \{-1, +1\}$  corresponding to each single-shot measurement, where  $-1/+1$  are assigned to the outcomes  $|L\rangle/|R\rangle$  respectively. The probability  $P_{x_n}$  of obtaining  $-1$  or  $+1$  is related to the underlying sequence  $\{\delta\Phi_n\}$  according to

$$P_{x_n}(x) = \delta(x-1) \left[ \frac{1}{2} + \frac{dP}{d\Phi_t^x} \delta\Phi_n \right] + \delta(x+1) \left[ \frac{1}{2} - \frac{dP}{d\Phi_t^x} \delta\Phi_n \right], \quad (1)$$

where for short we now use  $P$  to denote  $P_{|R\rangle}$  and we have assumed a linear probability-flux curve. Defining the DFT coefficients as

$$\tilde{x}_k \equiv \sum_{n=0}^{N-1} x_n e^{-i2\pi nk/N}, \quad (2)$$

then the expected value for the periodogram (PSD estimate) of  $\{x_n\}$  can be related to the underlying power spectral density of  $\{\delta\Phi_n\}$  according to

$$\begin{aligned} \langle |\tilde{x}_k|^2 \rangle &= \sum_{n,m=0}^{N-1} \langle x_n x_m \rangle e^{-i2\pi(n-m)k/N} \\ &= \sum_{n,m=0}^{N-1} \left[ 4 \left( \frac{dP}{d\Phi_t^x} \right)^2 \delta\Phi_n \delta\Phi_m + \delta_{n,m} \right] e^{-i2\pi(n-m)\frac{k}{N}} \\ &= 4 \left( \frac{dP}{d\Phi_t^x} \right)^2 \left| \sum_{n=0}^{N-1} \delta\Phi_n e^{-i2\pi nk/N} \right|^2 + N \\ &= 4 \left( \frac{dP}{d\Phi_t^x} \right)^2 \left| \widetilde{\delta\Phi}_k \right|^2 + N, \end{aligned} \quad (3)$$

where in the second line we have used the relation for the correlation

$$\langle x_n x_m \rangle = \begin{cases} 4 \left( \frac{dP}{d\Phi_t^x} \right)^2 \delta\Phi_n \delta\Phi_m & n \neq m \\ 1 & n = m \end{cases}$$

computed from (1). We can then convert to a physical single-sided flux noise PSD after assigning a sampling interval time  $\delta t$  to the sequence according to

$$\begin{aligned} S_{\Phi}^+(f) &= \frac{2T}{N^2} \left| \widetilde{\delta\Phi}_{k=fT} \right|^2 \\ &= \frac{2\delta t}{N} \frac{1}{4 \left( \frac{dP}{d\Phi_t^x} \right)^2} \langle |\tilde{x}_{k=fT}|^2 \rangle - 2 \frac{1}{4 \left( \frac{dP}{d\Phi_t^x} \right)^2} \delta t. \end{aligned} \quad (4)$$

where  $T = N\delta t$  is the total acquisition time for the dataset of  $N$  samples. The normalization convention for  $S_{\Phi}^+$  is chosen so that the total power in the signal is obtained by integrating over positive frequencies only. It is normalized such that the deterministic signal  $\Phi(t) = A \sin(2\pi ft)$ , with  $f$  in the baseband of the sampled DFT, has total power  $A^2/2$  as physically expected. The ability to correctly extract the underlying  $S_{\Phi}^+(f)$  with these formulas was verified by numerically feeding in an artificially generated noise source into a numerical simulation of our measurement, which was in turn fed into the data analysis software.

Equation (4) tells us that the PSD of our measurement sequence  $S_x^+(f)$  will be a combination of the underlying  $S_{\Phi}^+(f)$  plus a white noise floor. This noise floor makes sense because if there were no flux noise at all, we would expect shot noise from a perfectly uncorrelated probability of 0.5 for each measurement result. This white noise floor can be substantial: for typical parameters of  $\delta t = 10 - 100 \mu\text{s}$  and  $dP/d\Phi_t^x \approx 2000/\Phi_0$ , it has an equivalent flux noise amplitude of order  $1 \mu\Phi_0/\sqrt{\text{Hz}}$ . Since this is the typical strength expected for the intrinsic  $1/f$  flux noise of the device, this means the white noise floor will dominate the signal above  $\sim 1 \text{ Hz}$ .

Fortunately, there is a way to process the data that allows one to drastically reduce the shot noise floor and infer  $S_{\Phi}^+(f)$  from the measurement of  $S_x^+(f)$  without noticeably distorting the underlying flux noise signal [3]. The idea is inspired by the technique of using two separate detectors to sample a signal and computing the cross spectrum to throw away the detectors’ contribution to the noise. By breaking  $\{x_n\}$  into two interleaved series and taking the cross-spectral density (CSD) of the two sub-series, in the limit of infinite statistics this would completely eliminate the white noise floor, as shown below. The intuition behind this is that the interleaving removes the zero-delay autocorrelation term because the point  $\delta\Phi_n$  never “sees itself” in the sum. However, the actual datasets are finite, so as discussed below this cancellation will not be perfect. Regarding the underlying  $1/f$  signal itself, at least at low enough frequencies, the two interleaved flux noise signals should be approximately equal because the noise is highly correlated, and so the spectrum should not become distorted. This will be shown numerically below.

Mathematically, we define the two subsequences

$$x'_n = x_{2n}, \quad x''_n = x_{2n+1} \quad (5)$$

$n = 0, 1, 2, \dots, M-1$ , where  $M = N/2$ . We define a CSD for these interleaved sequences multiplied by a particular

frequency-dependent phase factor:

$$\langle \tilde{x}'_k (\tilde{x}''_k)^* \rangle e^{i2\pi k/N} \quad (6)$$

$$\begin{aligned} &= \sum_{n,m=0}^{M-1} \langle x'_n x''_m \rangle e^{i2\pi(n-m)k/M} e^{i2\pi k/2M} \\ &= 4 \left( \frac{dP}{d\Phi_t^x} \right)^2 \sum_{n,m=0}^{M-1} \delta\Phi_{2n} \delta\Phi_{2m+1} e^{-i2\pi(n-m-\frac{1}{2})k/M}, \end{aligned} \quad (7)$$

where the extra phase factor in line (6) corrects for the phase shift arising from the time domain offset between the two interleaved sequences. This phase offset ensures that the interleaved CSD of a deterministic sinusoidal signal with integer wave index  $q$  is itself real.

Equation (7) says that if we perform the interleave and take the CSD in the absence of underlying flux noise, we would observe absolutely no noise (since  $\delta\Phi_n$  is identically zero). Of course, the only reason the noise cancellation is exact is due to the use of the expectation value operator,  $\langle \cdot \rangle$ , which means an average over an ensemble of infinitely many realizations of  $\{x_n\}$  generated by a given  $\{\delta\Phi_n\}$ . In reality, we deal with finite sequences ( $N \approx 10^6$ ), and instead of (6), we can only compute the CSD coefficients  $\tilde{x}'_k (\tilde{x}''_k)^* e^{i2\pi k/N}$  for a finite number of finite sequences. The simplest way to understand the effect of interleaving on the white noise floor in the actual experiment, then, is to think about a random walk of phasors in the complex plane. First, let us consider how we can reduce the noise floor within a single dataset, and then we will consider averaging over multiple datasets. When  $\{\delta\Phi_n\}$  is identically zero (no flux noise), each CSD coefficient of the 1-bit detector measurement will itself a Gaussian distributed complex random variable, with uniformly distributed phase and some distribution of magnitude that is peaked close to  $\delta t / (dP/d\Phi_t^x)^2$ . Thus, the white noise floor will not actually be reduced unless we perform some sort of averaging, either across frequency bins or across datasets. Since it is informative to plot flux noise on a log-log scale, it is natural to use a logarithmic averaging scheme where the number of neighboring frequency bins whose CSD coefficients are averaged together is proportional to frequency. In other words, the number of bins per decade of frequency used in the averaging is constant, as is done for the quasistatic flux noise data shown in the main paper. Since the number of points  $N$  averaged in a frequency bin is chosen to scale as  $f$ , and sum of  $N$  random phasors scales in magnitude as  $\sqrt{N}$ , then upon taking the average over each frequency bin we would expect the white noise floor power to decrease as  $1/\sqrt{N}$ , meaning as  $1/\sqrt{f}$ . Thus, the act of interleaving and taking a coherent logarithmic frequency average within a given dataset amounts to a  $1/\sqrt{f}$  filter for the white noise floor. This can be seen from the slope of the observed filtered white noise floor in the numerical simulation of Fig. 2.

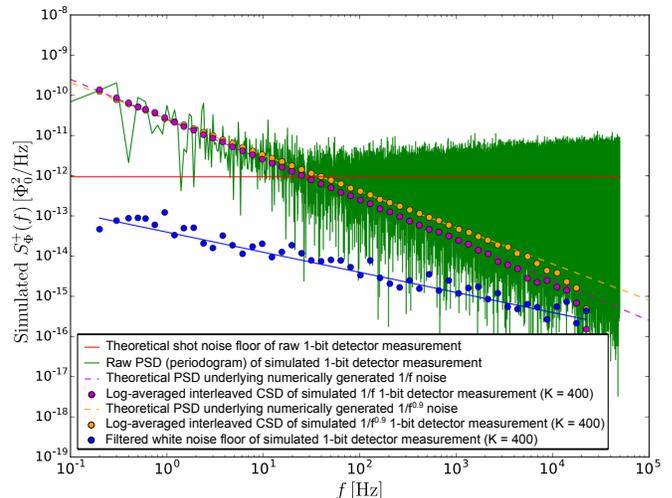


Figure 2. Numerical simulation of the 1-bit detector flux noise measurement for artificially generated  $1/f$  and  $1/f^{0.9}$  noise processes. The red line shows the expected white noise level for the raw 1-bit detector measurement without interleaving. Green shows the raw PSD of the simulated 1-bit detector measurement for a single realization of the numerically generated  $1/f$  noise. The dashed magenta line shows the theoretical PSD used to generate the artificial  $1/f$  noise. Magenta circles show the result of the simulated 1-bit detector measurement using the same processing that was used on the experimental data (interleaved + log-averaged over frequency, averaged over  $K = 400$  realizations of the numerically generated  $1/f$  noise). Orange: same but for numerically generated  $1/f^{0.9}$  noise. The blue circles show the white noise floor of the simulated measurement after interleaving and averaging, with the blue line its theoretically expected level.

However, we find that the filtered suppression of the white noise floor from averaging over frequency bins within a single dataset alone is not enough to bring the white noise floor below the qubit flux noise signal at high frequencies. Because of this, we further average the CSD coherently over  $K$  datasets, with  $K$  a few hundred, before taking the real part. This gives a factor of  $\sqrt{K} \approx 10$  further reduction in the filtered white noise floor power without distorting the underlying (presumably) correlated flux noise signal, according to the numerical simulation of the measurement shown in Fig. 2. This simulation uses artificially generated  $1/f^\alpha$  noise signals with  $\alpha = 1$  and  $\alpha = 0.9$  and magnitude at 1 Hz equal to  $5 \mu\Phi_0/\sqrt{\text{Hz}}$ , the value extracted from experiment. We note that the discretely generated artificial  $1/f$  noise extends an order of magnitude higher in frequency than the sampling frequency of the simulated 1-bit detector measurement. This was done purposefully to make sure there is no influence from aliasing. We find that the interleaving technique greatly reduces the effects of aliasing that would otherwise be present from the substantial amount of  $1/f$  noise that is likely present above the Nyquist frequency of the measurement. From the simulation, we see that the 1-bit detector measurement

after interleaving and coherently averaging over datasets would faithfully reproduce the  $1/f$  power spectrum without any significant distortions except for the highest half-decade of frequency. The filtered white noise is well below the simulated  $1/f$  noise (below the highest half-decade of frequency), which means it is even further below the experimentally measured flux noise (because of the significant ‘‘bump’’ observed starting at 10 – 100 Hz in the experimental data)<sup>1</sup>. As an extra consistency check, we note that the extracted flux noise data in the experiment was not materially changed (apart from the highest factor of 2 in frequency) whether we coherently averaged over 250 or 500 datasets.

### CONSISTENT DEFINITION OF $S_{\Phi}^+(f)$ AT LOW AND HIGH FREQUENCIES

Because the low and high frequency flux noise are measured by very different methods, we must be careful to have consistent definitions of  $S_{\Phi}^+(f)$  at low and high frequencies. In other words, in Fig. 2(e) of the main text, we must be sure we are plotting the same physical quantity at low and high frequencies, without any discrepant factors of 2 or  $2\pi$ . Such a discrepancy could for example affect the best-fit value of  $\alpha$  in an interpolating power law between the two frequency ranges.

To infer  $S_{\Phi}^+(f)$  at low frequencies, we measure the discrete time sequence  $\Phi_n$ , where  $\Phi_n$  is a classical real number, over  $N$  discrete time steps indexed by  $n \in \{0, 1, \dots, N-1\}$  and separated by the physical sampling interval  $\delta t$  (meaning total data acquisition time  $T = (N-1)\delta t \approx N\delta t$ ). We then estimate the single-sided PSD by computing

$$S_{\Phi}^+(f) = \frac{2T}{N^2} \langle |\tilde{\Phi}_{k=fT}|^2 \rangle, \quad (8)$$

where the DFT coefficients  $\Phi_k$  are defined by

$$\tilde{\Phi}_k \equiv \sum_{n=0}^{N-1} \Phi_n e^{-i2\pi nk/N}. \quad (9)$$

and  $\langle \cdot \rangle$  denotes an ensemble average since  $\tilde{\Phi}_k$  is itself a random variable.

At high frequencies, we instead infer the flux noise spectral density through Fermi’s golden rule [4] using the fact that  $1/T_1 = \Gamma_{\uparrow} + \Gamma_{\downarrow}$ , which implies the relation

$$\begin{aligned} S_{\Phi}^+(f) &\equiv S_{\Phi}(f) + S_{\Phi}(-f) \\ &= \frac{\hbar^2}{T_1} \frac{1}{|\langle 0 | \frac{d\hat{H}}{d\Phi_{\tau}} | 1 \rangle|^2} = \frac{\hbar^2 L^2}{T_1} \frac{1}{|\langle 0 | \hat{\Phi} | 1 \rangle|^2}, \end{aligned} \quad (10)$$

<sup>1</sup> We note that the ‘‘bump’’ in noise around 100 Hz in the experimental data is unchanged when the pulse tube compressor of the dry dilution refrigerator is turned off.

where [4]

$$S_{\Phi}(\omega) \equiv \int_{-\infty}^{\infty} d\tau e^{i\omega\tau} \langle \Phi(\tau)\Phi(0) \rangle. \quad (11)$$

Here,  $\Phi$  could be an operator, but for simplicity we can assume it’s a real number, since it is sufficient to check whether or not the two definitions (8) and (10) for  $S_{\Phi}^+$  coincide for a classical incoherent flux noise source acting on the qubit. Showing this will turn out to be equivalent to deriving the Wiener-Khinchin theorem for a stationary stochastic process.

First, we can write (9) in the continuum limit  $N \rightarrow \infty$ ,  $\delta t \rightarrow 0$  with  $T$  held constant, so that  $\Phi(t = n\delta t) = \Phi_n$  and  $\tilde{\Phi}_{k=fT} \rightarrow \frac{1}{\delta t} \int_0^T \Phi(t) e^{-i2\pi ft} dt$ . Since  $\Phi(t)$  is a random process, so is  $\tilde{\Phi}(f)$ , so we keep the  $\langle \cdot \rangle$  before taking the limit  $T \rightarrow \infty$ . We can obtain its expectation value by taking the limit  $T \rightarrow \infty$  after plugging the continuum limit expression into (8),

$$\begin{aligned} S_{\Phi}^+(\omega) &= \lim_{T \rightarrow \infty} \frac{2}{T} \left\langle \left| \int_0^T \Phi(t) e^{-i\omega t} dt \right|^2 \right\rangle \\ &= \lim_{T \rightarrow \infty} \frac{2}{T} \int_0^T \int_0^T dt dt' e^{i\omega t} e^{-i\omega t'} \langle \Phi(t)\Phi(t') \rangle \\ &= \lim_{T \rightarrow \infty} \frac{2}{T} \int_0^T \int_0^T dt dt' e^{-i\omega(t-t')} \langle \Phi(t-t')\Phi(0) \rangle, \end{aligned} \quad (12)$$

where in the last step we have assumed that  $\Phi(t)$  is a stationary process. To continue, we note that the integrand [call it  $f(t, t')$ ] inside the double integral is a function of  $\tau \equiv t' - t$  alone, meaning that  $f(\tau) = f(t - t')$  is constant along lines of constant  $\tau$  defined by the equation  $t' = t + \tau$  within the  $t-t'$  plane. We can therefore convert the double integral into a one-dimensional integral in  $f(\tau)$ , by integrating the diagonal ‘‘slices’’ formed by such lines across the two-dimensional domain of integration. The domain of integration is the square  $[0, T] \times [0, T]$  in the  $t-t'$  plane, which is covered by diagonal strips parameterized by  $\tau$  ranging from  $-T$  to  $T$ . The area of each diagonal strip corresponding to  $\tau$  with infinitesimal width  $d\tau$  is  $\sqrt{2}(T - |\tau|) \frac{d\tau}{\sqrt{2}} = (T - |\tau|)d\tau$ , so we can convert  $\int_0^T \int_0^T dt dt' f(t - t')$  to  $\int_{-T}^T d\tau f(\tau)$ , meaning that (12) becomes

$$\begin{aligned} S_{\Phi}^+(\omega) &= \lim_{T \rightarrow \infty} \frac{2}{T} \int_{-T}^T d\tau e^{i\omega\tau} \langle \Phi(\tau)\Phi(0) \rangle (T - |\tau|) \\ &= \lim_{T \rightarrow \infty} 2 \int_{-T}^T d\tau e^{i\omega\tau} \langle \Phi(-\tau)\Phi(0) \rangle \left( 1 - \frac{|\tau|}{T} \right) \\ &= 2 \int_{-\infty}^{\infty} d\tau e^{i\omega\tau} \langle \Phi(\tau)\Phi(0) \rangle, \end{aligned} \quad (13)$$

where in the last line we have used the property that the autocorrelation function  $\langle x(t+\tau)x(t) \rangle$  is an even function of  $\tau$ . Comparing this to (11) and (10) shows that the two definitions of  $S_{\Phi}^+$  are indeed equivalent.

### FLUX NOISE AT HIGH AND LOW FREQUENCIES CHANGES SIMILARLY BETWEEN SAMPLES

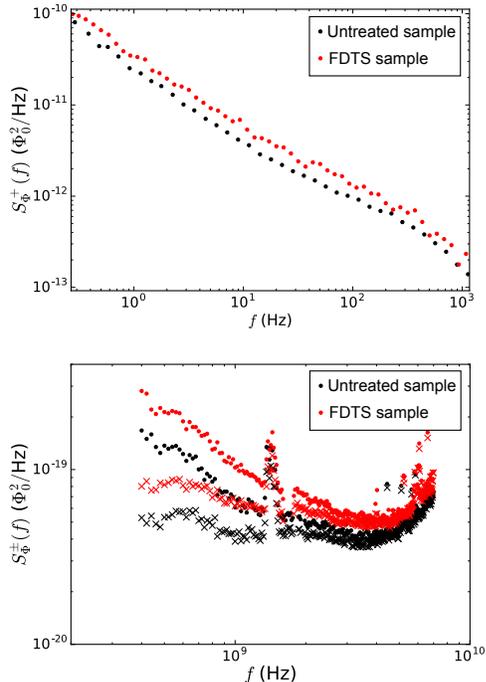


Figure 3. Comparison of flux noise at low and high frequencies between two different samples at base temperature. The samples were nominally identical apart from differing fabrication post-treatments.

Fig. 3 shows low and high frequency flux noise data for nominally the same qubit on two different chips, with the second chip seeing extra post-processing in the form of a downstream oxygen ash clean and the application of a (nominal) monolayer of perfluorodecyltrichlorosilane (FDTS) via molecular vapor deposition. The second sample also sat covered in photoresist for 6 months longer than the first sample. The flux noise power just below 1 GHz changes by a factor of 1.6, while the flux noise power around 1 Hz changes by a factor of approximately 1.5. Consistent with this we also observe that the Ramsey decay time  $T_{\phi_2}$  away from zero tilt was  $\sim 30\%$  lower on the FDTS sample when measured at a point with the same sensitivity of  $f_{10}$  to tilt flux (this degradation was reproducible at several bias points).

### SURFACE HYDROGEN AS POSSIBLE ORIGIN OF DISSIPATION PEAK AT 1.4 GHz

The mysterious “mode” at 1.4 GHz, where there is a peak in dissipation, is present at the same frequency and similar strength in all qubits with different chip and sample box sizes, different qubit geometries, and differing

types of filtering and attenuation on the bias lines, including the addition of attenuation on the output of the readout line. There is reason to believe that this feature is something physically intrinsic to the materials of the qubit, because it persists even when significantly changing the filtering and attenuation on all the coaxial lines going to the qubit box, and is furthermore independent of chip size and box size, and so is unlikely to be due to coupling to an environmental electromagnetic mode.

Recently, using on-chip ESR techniques, de Graaf *et al.* found evidence for coupling of superconducting resonators to the ground state hyperfine transition of surface hydrogen, whose presence is claimed to be due to physical adsorption as a by-product of water dissociation [5, 6]. It so happens that the ground state hyperfine transition frequency of hydrogen is 1.42 GHz, the frequency of the feature observed in this work. Below, we provide a rough calculation to determine what surface density of weakly bound hydrogen atoms would be needed to explain the peak in dissipation observed in the fluxmon.

We suppose that there is a distribution of hydrogen atoms weakly bound to the surface of the fluxmon’s coplanar waveguide (CPW) segment, so that the ground state hyperfine levels are roughly those of an isolated H atom. First, we write down the interaction Hamiltonian between the fluxmon and a single such atom (electron plus proton) at a lateral position  $z$  ( $-W/2 \leq z \leq W/2$ ) on the surface of the CPW center trace (Fig. 4),

$$\begin{aligned} \hat{H}_{\text{int}} &= -\boldsymbol{\mu}_e \cdot \mathbf{B}(z) - \boldsymbol{\mu}_p \cdot \mathbf{B}(z) \\ &= -\mu_B \hat{B}_z(z) \sigma_e^z - \frac{g_p e \hbar}{2m_p} \hat{B}_z(z) \sigma_p^z, \end{aligned} \quad (14)$$

where we have assumed the magnetic field generated by the qubit at the atom is in the  $z$  direction (see Fig. 4).

The magnetic field generated by the qubit on the surface of the metal at position  $z$  is related to the qubit current operator according to

$$\hat{B}_z(z) = \frac{\mu_0 \hat{I}}{\pi W} \frac{1}{\sqrt{(1 - \frac{2z}{W})(1 + \frac{2z}{W})}} \quad (15)$$

This equation is valid as long as  $z$  is further than  $\sim \lambda^2/2t \approx 50$  nm from the edge of the CPW strip [7]. Note

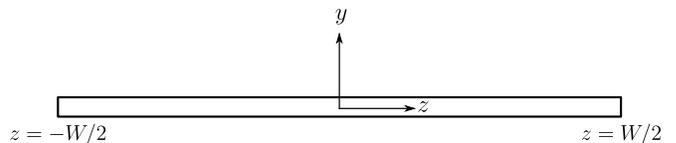


Figure 4. Cross section of center trace of fluxmon CPW (treated as a thin film) and definition of coordinates. Fluxmon current flows into the page ( $x$  direction), generating a magnetic field in the  $z$  direction.  $W \approx 10 \mu\text{m}$  is the width of the CPW center trace.

that it is normalized so that effectively half the current flows on the bottom and half flows on the top side of the CPW. Since the CPW center trace dominates the surface magnetic participation, we have ignored the CPW ground planes.

We claim that the qubit's magnetic field (which is uniform over the hydrogen atom) will couple to the transition between the singlet hyperfine ground state  $|S\rangle$  and the  $m = 0$  triplet state  $|T_0\rangle$  1.4 GHz above. We will want to do a Fermi's golden rule calculation to calculate the induced  $|e\rangle \rightarrow |g\rangle$  transition rate in the qubit. This requires computing the matrix element between initial and final states,  $|i\rangle = |e\rangle|S\rangle$  and  $|f\rangle = |g\rangle|T_0\rangle$ , where  $|S\rangle$  and  $|T_0\rangle$  are the singlet and  $m = 0$  triplet states

$$\begin{aligned} |S\rangle &= \frac{1}{\sqrt{2}}(|\uparrow_e\rangle|\downarrow_p\rangle - |\downarrow_e\rangle|\uparrow_p\rangle), \\ |T_0\rangle &= \frac{1}{\sqrt{2}}(|\uparrow_e\rangle|\downarrow_p\rangle + |\downarrow_e\rangle|\uparrow_p\rangle) \end{aligned} \quad (16)$$

We compute for a hydrogen atom at position  $z$

$$\langle f|\hat{H}_{\text{int}}|i\rangle(z) = \langle g|\hat{I}|e\rangle\mu_B \frac{\mu_0}{\pi W} \frac{1}{\sqrt{(1 - \frac{2z}{W})(1 + \frac{2z}{W})}}, \quad (17)$$

where we have dropped a second term due to the proton because it is much smaller (by a factor of  $\sim \frac{m_e}{m_p}$ ). We recognize  $\langle g|\hat{I}|e\rangle = \frac{1}{L}\langle g|\hat{\Phi}|e\rangle$  as the qubit persistent current  $I_p$ , which at 1.4 GHz is approximately  $0.3 \mu\text{A}$ .

Now we are ready to use Fermi's golden rule to compute the decay rate induced in the qubit by a large collection of hydrogen atoms. We need to know the density of states of hyperfine splittings. Since the peak in qubit dissipation at 1.4 GHz has a linewidth of 150 MHz, significantly wider than the dephasing-limited qubit linewidth of 20 - 30 MHz, we assume the hyperfine splittings are spread out over  $\sim 100$  MHz, presumably due to an inhomogeneous weak interaction with the  $\text{AlO}_x$  surface. A naive interpretation Ref. [8] suggests that this spread might be possible if the hydrogen was weakly bound to O atoms with a mean distance of  $\sim 4.5$  Bohr radii with an inhomogeneous spread of  $\sim 0.2$  Bohr radii. Further theoretical study of the surface physics is needed to determine whether or not this spread is physically realistic.

In any case, this gives us the density of states  $\rho(E_{10})$  to use in Fermi's golden rule,

$$\Gamma_{|e\rangle \rightarrow |g\rangle} = \frac{2\pi}{\hbar} \int_{-W/2+50\text{nm}}^{W/2-50\text{nm}} dz |\langle f|\hat{H}_{\text{int}}|i\rangle(z)|^2 \rho(E_{10}), \quad (18)$$

where for simplicity we use a length cutoff of  $50 \text{ nm} \approx \lambda$  from the CPW edge in order to have a convergent integral. Using the true current distribution that correctly treats the edge is not expected to significantly change the answer.

Using the fact that the fluxmon has a length of 2 nm (i.e., the dimension going into the page in Fig. 4), we back out that to get the experimentally observed excess dissipation  $\Gamma_{|e\rangle \rightarrow |g\rangle} \approx 1/(200 \text{ ns})$ , we must have an areal density of  $\sim 1 \times 10^{16}/\text{m}^2$ , or about 1 H atom per 10 nm, on the  $\text{AlO}_x$  surface.

### CHECKING FOR DISTORTION OF EXTRACTED $S_{\Phi_t^x}^{\pm}(f)$ FROM NONLINEAR CROSSTALK

In the main text, to deduce the high frequency flux noise  $S_{\Phi_t^x}^{\pm}(f)$  in the main qubit loop we measure  $T_1$  at zero tilt, because at degeneracy only flux noise in  $\varepsilon(\Phi_t^x)$  and not noise in  $\Delta(\Phi_{\text{ba}}^x)$  would induce transitions between qubit energy eigenstates. However, this is no longer strictly true if the two junctions in the DC SQUID are not perfectly symmetric, which is to be expected due to fabrication imperfections. It can be shown [9] that if the junction asymmetry is  $d \equiv \frac{E_{J1} - E_{J2}}{E_{J1} + E_{J2}}$ , any flux threading the DC SQUID loop will lead to an offset in tilt flux according to the nonlinear relation

$$\Delta\Phi_t^x = \frac{\Phi_0}{2\pi} \tan^{-1}(d \tan[\pi\Phi_{\text{ba}}^x/\Phi_0]), \quad (19)$$

implying that noise in  $\Phi_{\text{ba}}^x$  leads to noise in  $\Phi_t^x$  according to the differential transfer function

$$\frac{d\Phi_t^x}{d\Phi_{\text{ba}}^x} = \frac{d}{2} \frac{1}{d^2 \sin^2(\pi\Phi_{\text{ba}}^x/\Phi_0) + \cos^2(\pi\Phi_{\text{ba}}^x/\Phi_0)}. \quad (20)$$

By measuring  $\Delta\Phi_t^x$  as a function of  $\Phi_{\text{ba}}^x$  and subtracting out any contribution from linear geometric crosstalk between barrier and tilt bias lines, we can extract this intrinsic nonlinear transfer function experimentally. For the data in the main text, we obtain  $(\frac{d\Phi_t^x}{d\Phi_{\text{ba}}^x})^2 < .005$  over the range of  $\Delta(\Phi_{\text{ba}}^x)$  measured, corresponding to a junction asymmetry of  $\sim 1\%$ . This suggests that dissipation from incoherent flux noise is likely dominated by noise in  $\Phi_t^x$  and not  $\Phi_{\text{ba}}^x$ , since the  $1/f$  noise in  $\Phi_{\text{ba}}^x$  should be less than or comparable to the noise in  $\Phi_t^x$  (using conventional Ramsey experiments [2] at zero tilt we obtain the noise in the DC SQUID flux  $\Phi_{\text{ba}}^x$  has magnitude  $1 - 2\mu\Phi_0/\sqrt{\text{Hz}}$  at 1 Hz, compared to  $\sim 5\mu\Phi_0/\sqrt{\text{Hz}}$  for  $\Phi_t^x$ ). However, this does not exclude the possibility of the relative strength of noise in  $\Phi_t^x$  and  $\Phi_{\text{ba}}^x$  changing greatly between 1 Hz and GHz, but given that we see similar  $T_1$ 's over several samples with different junction asymmetries  $d$ , this seems not to be the case. Note that this analysis does not rule out noise from surface spins fluctuating on the wiring of the DC SQUID, because this wire is shared by both barrier and tilt loops; rather, it only implies that noise from such fluctuators would affect the  $T_1$  data via induced noise in  $\Phi_t^x$  and not due to induced noise in  $\Phi_{\text{ba}}^x$ .

## CHECKING FOR DISSIPATION FROM NON-EQUILIBRIUM QUASIPARTICLES

The frequency dependence of the fluxmon  $T_1$  at zero tilt below  $\sim 3$  GHz is also consistent with quasiparticle dissipation if one were to assume a large enough population of non-equilibrium quasiparticles in the system. We give theoretical and experimental arguments as to why this is unlikely to be a dominant effect in our system, including a test of the effects of magnetic vortices.

Quasiparticles with energy near the superconducting gap  $\Delta$  can absorb energy from the qubit when they tunnel across one of the Josephson junctions, while “hot” quasiparticles with energies more than  $\hbar\omega_{10}$  from the gap can excite the qubit. If the energies of all quasiparticles influencing the qubit are sufficiently less than  $2\Delta$ , then for an arbitrary quasiparticle occupation distribution  $f(E)$ , the decay and excitation rates induced on the qubit are given by [10]

$$\Gamma_{i \rightarrow f} = \sum_{j=1,2} \left| \langle f | \sin \frac{\hat{\phi}_j}{2} | i \rangle \right|^2 S_{\text{qp}}^j(\omega_{if}), \quad (21)$$

where  $j$  indexes the two Josephson junctions of the fluxmon and

$$S_{\text{qp}}^j(\omega) = \frac{32EJ_j}{\pi\hbar} \int_0^\infty dx \rho((1+x)\Delta)\rho((1+x)\Delta + \hbar\omega) \times f[(1+x)\Delta](1 - f[(1+x)\Delta + \hbar\omega]) \quad (22)$$

is the double-sided quasiparticle current spectral density, with  $\rho(E) = E/\sqrt{E^2 - \Delta^2} \approx 1/\sqrt{2(E - \Delta)/\Delta}$  the normalized quasiparticle density of states. The formula (22) works for  $\omega < 0$  by simply swapping the arguments of  $f$  and replacing  $\omega$  with  $-\omega$ . In a non-tunable gap flux qubit (i.e., a single-junction fluxmon), at zero tilt the junction would be biased at  $\pi$ , meaning that quasiparticle dissipation ( $\propto |\langle 0 | \sin \frac{\hat{\phi}}{2} | 1 \rangle|^2$ ) would vanish (physically, this is due to destructive interference between electron-like and hole-like tunneling [11]). However, for the gap-tunable fluxmon, even though the effective dynamical phase  $\hat{\phi} = (\hat{\phi}_1 + \hat{\phi}_2)/2$  can be biased at  $\pi$  at zero tilt, the phase of the individual junctions are not. Instead, flux quantization dictates that  $\hat{\phi}_1 = \hat{\phi} - \pi\Phi_{\text{ba}}/\Phi_0$ ,  $\hat{\phi}_2 = \hat{\phi} + \pi\Phi_{\text{ba}}/\Phi_0$ , meaning that the matrix element can be non-zero and quasiparticle dissipation can occur even at zero tilt.

While a thermal distribution of quasiparticles would be much too small to explain the observed dissipation, this does not rule out the possibility of non-equilibrium quasiparticles. A reasonable model for computing the distribution of non-equilibrium quasiparticles is outlined in Refs. [12–14]. Here, quasiparticles are assumed to be injected at some energy or range of energies well above the gap and, via electron-phonon scattering and recombination, relax to some steady-state distribution that is

essentially independent of the injection energy as long as the injection energy is high enough. Using the steady state equations outlined in Ref. [13], we can calculate the expected distribution of nonequilibrium quasiparticles as a function of injection rate and phonon temperature  $T_{\text{ph}}$ , and then using (21) and (22) numerically calculate the resulting up and down transition rates to obtain the quasiparticle-induced  $T_1$  and stray population. We choose an injection rate that leads to a quasiparticle density that best matches the measured  $T_1$  data while choosing the phonon temperature to be equal to the fridge temperature. We assume the same injection rate for all four fridge temperatures used in the main paper. The injection energies were between  $2.1\Delta$  and  $2.2\Delta$ , though changing this energy range does not materially affect the result. The results are summarized in Fig. 8, where we plot the computed  $f(E)$ ,  $T_1$  and stray population/effective temperature induced on the qubit on top of the  $T_1$  and stray population data used in Fig. 3 of the main paper. The best fit yields a quasiparticle density  $n_{\text{qp}} \equiv 2D(E_F) \int_{\Delta}^{\infty} \rho(E)f(E) dE$  of  $3.5/\mu\text{m}^3$ , or a fractional quasiparticle density  $x_{\text{qp}} = n_{\text{qp}}/n_{\text{cp}} = n_{\text{qp}}/(D(E_F)\Delta) = 1.3 \times 10^{-6}$ , where  $D(E_F)$  is the density of electron states at the Fermi energy.

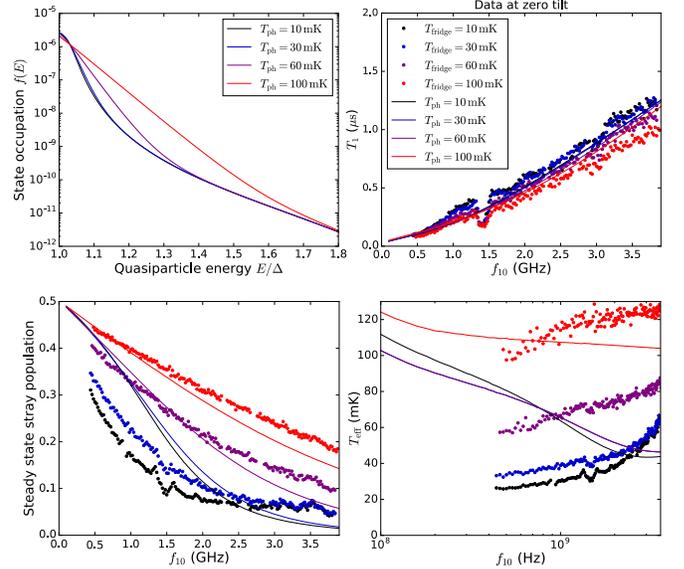


Figure 5. Data vs. fit to nonequilibrium quasiparticle theory. Although a quasiparticle density can be chosen large enough to roughly match  $T_1$  vs. frequency, a simultaneous fit to stray population is not possible within this model.

We make several observations about this calculation. While the fits do match the  $T_1$  data reasonably well below  $\sim 4$  GHz, the stray population is quite off and does not have nearly a large enough dependence on temperature at low frequencies. We also see that the quasiparticle-induced qubit effective temperature does not approach a constant value at low frequencies like the data appears to do, and like the stray populations, the predicted effective

temperatures do not match the data well at all, especially for the lowest three temperatures. Furthermore, there are several experimentally-based estimates substantially below  $x_{\text{qp}} = 1.3 \times 10^{-6}$  in the superconducting qubit literature, for example  $1 \times 10^{-8}$ ,  $4 \times 10^{-8}$ , and  $3 \times 10^{-7}$  in Refs. [15–17], and  $2.2 \times 10^{-7}$  in Ref. [18] (where the nonequilibrium quasiparticle density explained both  $T_1$  and stray population), so it is not unreasonable that  $x_{\text{qp}}$  is similarly low in our system, especially given the stringent level of light-tight filtering used in our setup [19]. In addition, we note that if the non-equilibrium quasiparticle density was indeed  $x_{\text{qp}} = 1.3 \times 10^{-6}$ , the  $T_1$  limit imposed on our standard Xmon qubits would be  $7 \mu\text{s}$  at 3 GHz, but Xmons consistently achieve  $T_1$  a factor of 10 higher than this at 3 GHz, and furthermore show a defect structure and opposite overall scaling in frequency from that predicted from quasiparticles. We have no reason to believe the quasiparticle density would be any higher in the fluxmon device given that the device is fabricated with identical materials and chip mount, and the fridge wiring is essentially identical to that used for Xmon experiments. If the quasiparticles were being introduced by the long (few  $\mu\text{s}$ ) readout pulses used for the fluxmon, we would expect  $T_1$  to depend on the repetition time between experiments. We do not see any change in  $T_1$  as we vary the repetition time from  $50 \mu\text{s}$  to  $1000 \mu\text{s}$ .

One more experimental check we can perform is to see if  $T_1$  improves after inducing magnetic vortices into the Al film of the sample. Abrikosov vortices, which have quasinormal cores, form when a thin-film superconductor is cooled through  $T_c$  in the presence of a magnetic field. It is well-known that such vortices trap quasiparticles, and have even been shown to significantly decrease quasiparticle-induced dissipation in superconducting qubits [17]. In particular, for a thin-film Al transmon qubit vortices were shown to significantly decrease quasiparticle-induced dissipation with a modest applied field of  $\sim 10 \text{ mG}$  [17]. To check if there is any improvement in  $T_1$  to be gained from the presence of vortices, we added a magnetic coil to the setup directly outside the qubit box, which we used to apply a several different magnetic fields between  $-30$  and  $30 \text{ mG}$  to the qubit chip as it cooled through its superconducting transition. The result of these magnetic field cools are shown in Fig. 7. We see only degradation of qubit  $T_1$  with applied magnetic field. This data suggests that quasiparticles are not playing a dominant role in qubit dissipation in the fluxmon device.

We also note that, as depicted in Fig. 6 for the zero field data, the raw  $T_1$  traces of 1-state population versus time fit very well to an exponential decay as opposed to showing a measurable non-exponential behavior that could be due to fluctuation in quasiparticle number [20].

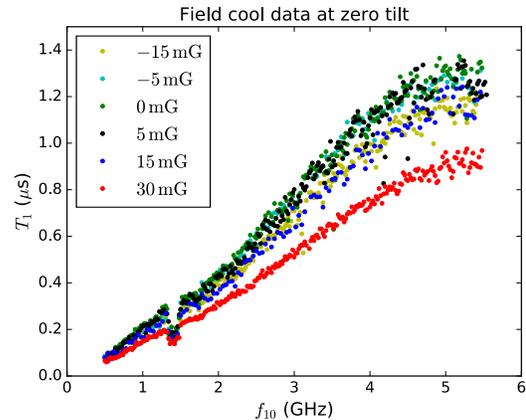


Figure 6. Field cool data. Inducing vortices only degrades  $T_1$ .

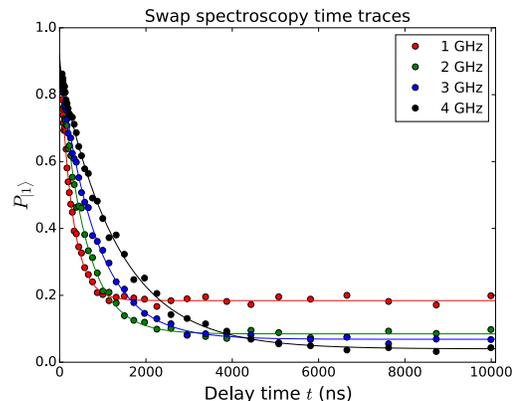


Figure 7. Swap spectroscopy time traces at zero applied field show an exponential time dependence. State preparation infidelity is due to imperfect  $\pi$  pulse given the short  $T_2$  time, and relaxation during the detuning pulse.

### LOW AND HIGH FREQUENCY NOISE CUTOFFS FOR MACROSCOPIC RESONANT TUNNELING RATES

In the main text, it was argued that the low-frequency cutoff  $f_l$  in the noise integral for the MRT tunneling linewidth  $W$  should be the inverse of the total experimental data acquisition time, whereas the  $f_l$  for the integral for the reorganization energy  $\varepsilon_p$  should be the tunneling rate near maximum tunneling. The latter physically makes sense as the reorganization energy should not depend on the time at which the experiment was performed. In other words, only dissipation at frequencies higher than the slowest timescale of the tunneling can affect this energy. However, different instances of the tunneling experiment may have different quasistatic flux offsets  $\delta\varepsilon = 2I_p\delta\Phi_{\xi}^x$  as discussed earlier in the context of the low frequency flux noise measurement, which can give additional broadening of the MRT tunneling peak. For simplicity, let us take the Gaussian approx-

imation to the lineshape described in the main text,  $\Gamma(\varepsilon) = \sqrt{\frac{\pi}{8}} \frac{\Delta^2}{\hbar W} \exp\left[-\frac{(\varepsilon - \varepsilon_p)^2}{2W^2}\right]$ , so that what is actually measured after averaging is the quantity

$$\bar{\Gamma}(\varepsilon) = \int_{-\infty}^{\infty} d(\delta\varepsilon) \Gamma(\varepsilon + \delta\varepsilon) p(\delta\varepsilon), \quad (23)$$

where  $p(\delta\varepsilon) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\delta\varepsilon)^2}{2\sigma^2}}$  describes the Gaussian distribution of quasistatic flux fluctuations. Performing the integration yields

$$\bar{\Gamma}(\varepsilon) = \frac{\pi}{8} \frac{\Delta^2}{\sqrt{W_0^2 + \sigma^2}} e^{-\frac{(\varepsilon - \varepsilon_p)^2}{2(W_0^2 + \sigma^2)}}, \quad (24)$$

which shows that the original  $W_0$  (obtained from integrating down to  $f_l$  equal to the maximum tunneling rate) is broadened by the r.m.s. quasistatic flux fluctuations via addition in quadrature. Since  $W^2 = 4I_p^2 \int_{f_l}^{f_h} df S_{\Phi}^+(f)$ , this amounts to extending  $f_l$  down to the inverse total experimental averaging time.

For the high frequency cutoff to the tunneling rate integral, we use the oscillation frequency of the inverted potential barrier. Previous theoretical and experimental work in macroscopic quantum tunneling has shown this to be the physical high-frequency cutoff [21–24]. Nevertheless, we find that the tunneling rate near resonance according to full integral in Eq. (3) of the main text is not materially affected by this high frequency cutoff.

### TEMPERATURE INDEPENDENCE OF THE CLASSICAL LOW FREQUENCY FLUX NOISE

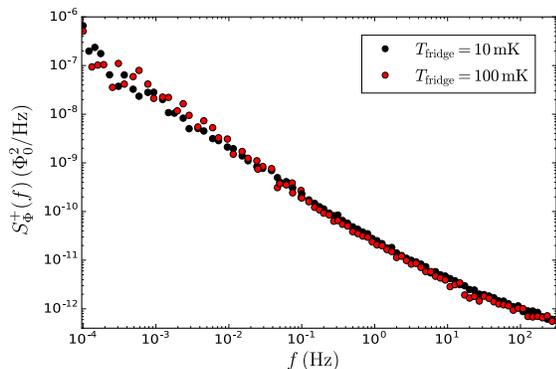


Figure 8. Quasistatic flux noise measurement is not noticeably changed between fridge temperatures 10 mK and 100 mK.

We do not see any systematic temperature dependence of the quasistatic noise at low frequencies or of the Ramsey decay time  $T_{\phi 2}$ , consistent with a picture of a paramagnetic environment with temperature-independent relaxation times (see the following sections).

### FLUCTUATION-DISSIPATION RELATION BETWEEN $S_{\Phi}^{\pm}(\omega)$ AND MAGNETIC SUSCEPTIBILITY

Here we give a simple, generic argument for how the two-sided spectrum  $S_{\Phi}^{\pm}(\omega)$  is related to the dynamic susceptibility  $\chi(\omega) = \chi'(\omega) + i\chi''(\omega)$  of the magnetic environment. In particular, we outline how  $S_{\Phi}^-(\omega)$  may be related to the absorptive part of the environment's linear response, and that its  $1/T$  dependence is consistent with a paramagnetic environmental susceptibility.

The most common assumption in models for  $1/f$  noise is that the noise comes from a collection of dynamical fluctuators, each characterized by an exponential decay with a single relaxation time. The  $1/f$  scaling then arises from an exponentially broad distribution of these relaxation times (see discussion in next section). The fluctuators could be represented by weakly interacting spin clusters of various sizes on the qubit surface, where the relaxation time is exponentially dependent on cluster size. An alternative model invokes a spin diffusion mechanism where collective diffusion modes play the role of individual fluctuators, and the broad distribution of relaxation times arises from the highly non-uniform distribution of the magnetic field around the superconducting metal of the flux qubit loop.

Whatever the microscopic source of these fluctuators may be, each fluctuator will have an effective magnetic moment operator  $M_n$  that will couple flux into the qubit according to  $\Delta\Phi = g_n M_n$ . The two-sided flux noise spectrum in the qubit can then be written as

$$\begin{aligned} S_{\Phi}(\omega) &= \int_{-\infty}^{\infty} d\tau e^{i\omega\tau} \langle \hat{\Phi}(\tau) \hat{\Phi}(0) \rangle \\ &= \int_{-\infty}^{\infty} d\tau e^{i\omega\tau} \left\langle \sum_{n,m} g_n M_n(\tau) g_m M_m(0) \right\rangle \\ &\approx \sum_n |g_n|^2 S_{M_n}(\omega), \end{aligned} \quad (25)$$

where  $S_{M_n}$  is the full two-sided spectral density of magnetization noise, and in the last line we have assumed negligible correlations between fluctuators (spin clusters/diffusion modes). We now use the fluctuation-dissipation theorem, which states that the equilibrium fluctuations of each moment are related to the dissipative part of its response to a non-equilibrium perturbation according to

$$S_{M_n}(\omega, T) = \hbar(1 + \coth[\hbar\omega/2k_B T]) \chi_n''(\omega, T), \quad (26)$$

where  $\chi(\omega, T) = \chi'(\omega, T) + i\chi''(\omega, T)$  is the frequency-domain linear response function (dynamical susceptibility) to a magnetic field. Inserting this relation above

yields

$$S_{\Phi}(\omega, T) = \sum_n \hbar |g_n|^2 \left[ 1 + \coth \left( \frac{\hbar\omega}{2k_B T} \right) \right] \chi_n''(\omega, T). \quad (27)$$

Since  $\chi''(\omega, T)$  is an odd function of  $\omega$ , looking at the antisymmetric part shows that the experimental  $1/T$  dependence of  $S_{\Phi}^-(f)$  below the classical-quantum crossover is consistent with a dynamical susceptibility  $\chi_n''$  that scales as  $1/T$  for all fluctuators. This is consistent with previous measurements the  $1/T$  dependence of the static susceptibility in SQUIDS [25], assuming a temperature-independent distribution of relaxation times.

### $1/f^\alpha$ SCALING NEAR THE CROSSOVER AND HIGH-FREQUENCY CUTOFF

Here we discuss the  $1/f^\alpha$  form of the noise and its modifications due to a possible high frequency cutoff for the relaxation times of the magnetic fluctuators. We argue that while our data clearly shows that such a cutoff must be of order  $k_B T/h$  or higher, a model where the cutoff is a few times  $k_B T$  fits the temperature dependence data slightly better than one with a much higher cutoff.

To obtain a  $1/f^\alpha$  scaling, we assume a single relaxation time  $\tau_n$  for each fluctuator. The dynamic susceptibility of a single such fluctuator is given by a standard Drude formula  $\chi_n(\omega, T) = \frac{\chi_n(0, T)}{1 + i\omega\tau_n}$ , meaning  $\chi_n''(\omega, T) = \chi_n(0, T) \frac{\omega\tau_n}{1 + \omega^2\tau_n^2}$ . For paramagnetic spins,  $\chi_n(0, T) \propto 1/T$ . In the limit of many fluctuators with different relaxation times, we can convert (27) into an integral over  $\tau$  with an effective weight for each  $\tau$ :

$$S_{\Phi}(\omega) = \hbar \left[ 1 + \coth \left( \frac{\hbar\omega}{2k_B T} \right) \right] \int_{\tau_{\min}}^{\tau_{\max}} \rho(\tau) \frac{1}{T} \frac{\omega\tau}{1 + \omega^2\tau^2}, \quad (28)$$

where  $\tau_{\min}$  and  $\tau_{\max}$  are lower and upper cutoffs for the relaxation times, and we have included a uniform  $1/T$  factor in the integrand under the assumption that all the fluctuators are paramagnetic.<sup>2</sup> In the classical limit,<sup>3</sup>

$$S_{\Phi}^+(\omega \ll k_B T/\hbar) \propto k_B \int_{\tau_{\min}}^{\tau_{\max}} d\tau \rho(\tau) \frac{\tau}{1 + \omega^2\tau^2}, \quad (29)$$

$$S_{\Phi}^-(\omega \ll k_B T/\hbar) \propto \frac{\hbar}{T} \int_{\tau_{\min}}^{\tau_{\max}} d\tau \rho(\tau) \frac{\omega\tau}{1 + \omega^2\tau^2}. \quad (30)$$

<sup>2</sup> We have neglected a possible dependence of the static susceptibility  $\chi(0, \tau, T)$  on  $\tau$  because it should be very weak. For instance, if we consider superparamagnetic clusters with different values for the total spin  $S$  that fluctuate by tunneling through anisotropy barriers  $U \propto S^2$ , then for a given  $S$ ,  $\chi(0) \propto \frac{1}{3}S(S+1)$  while  $\log \tau \propto U \propto S^2$ . Thus  $\chi(0, \tau, T)$  depends on  $\tau$  only logarithmically and can be safely replaced by its average value.

<sup>3</sup> We note that in the fully classical limit  $\hbar \rightarrow 0$ ,  $S_{\Phi}^-$  will vanish while  $\chi''(\omega)$  (the dissipation) does not need to vanish.

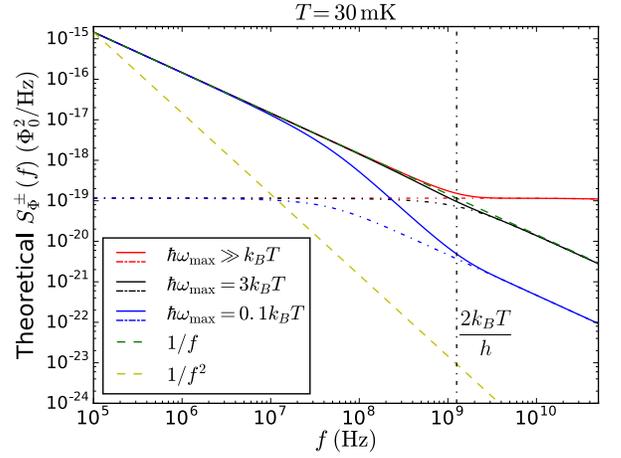


Figure 9.  $1/f$  noise according to the finite high frequency cutoff model (31) for different values of  $\omega_{\max}$ , showing the three qualitative ‘flavors’ the noise scaling near the classical-quantum crossover. Dash-dotted lines are  $S_{\Phi}^-$  while solid lines are  $S_{\Phi}^+$ . The ‘smoothest’ transition through the cutoff where  $S_{\Phi}^+(f)$  remains close to  $1/f$  for all frequencies is achieved for  $\omega_{\max} \approx 3k_B T/h$ .

As before, from this we can see that our experimental data below the classical-quantum crossover is consistent with an environment of magnetic fluctuators with a paramagnetic static susceptibility  $\chi(0)$ , under the assumption that  $\rho(\tau)$  is independent of temperature.

If we postulate  $\rho(\tau) \propto \frac{1}{\tau}$  between  $\tau_1$  and  $\tau_2$ , then performing the integration leads to  $S_{\Phi}^+(\omega) \propto 1/\omega$  assuming  $\frac{1}{\tau_2} \ll \omega \ll \frac{1}{\tau_1}$  ( $\omega_{\min} \ll \omega \ll \omega_{\max}$ ), which is the usual picture of  $1/f$  noise in the classical limit. However, more precisely performing the full integration (28) without assuming anything about  $\omega$  relative to  $T$  or  $\tau_{\min}/\max$  yields for the full spectrum

$$S_{\Phi}(\omega) \propto \frac{\hbar}{T} \left[ 1 + \coth \left( \frac{\hbar\omega}{2k_B T} \right) \right] \tan^{-1}(\omega\tau) \Big|_{\tau_{\min}}^{\tau_{\max}} \quad (31)$$

As long as  $\tau_{\min}/\max$  don’t depend exponentially on  $T$ , the temperature dependence of  $S_{\Phi}^-$  will again be given by that of the static susceptibility. Let us assume that  $\omega_{\min} \ll k_B T/h$  (justified by the presence of  $1/f$  flux noise well below 1 GHz) and look at the shape of the classical-quantum crossover for different  $\omega_{\max}$ . Fig. 9 shows that there are three qualitative types of scaling behavior of  $S_{\Phi}^{\pm}(f)$  around the crossover. If  $\omega_{\max} \ll k_B T/h$ , then  $S_{\Phi}^+(f)$  would approach the crossover from below as  $1/f^2$ , inconsistent with the data. On the other hand, if  $\omega_{\max} \gg k_B T/h$ , then  $S_{\Phi}^+(f)$  will turn into white noise just above the crossover. There is also an intermediate regime  $\omega_{\max} \approx 3k_B T/h$  where  $S_{\Phi}^+(f)$  is very close to  $1/f$  for all frequencies except for a slight deviation at the crossover point.

In Fig. 10, we compare the phenomenological thermo-

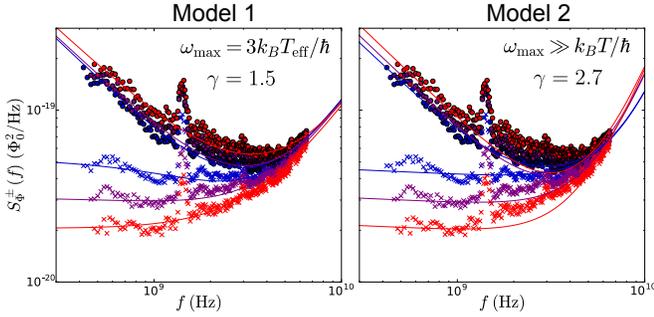


Figure 10. Comparison of two models with the constraint that the ohmic exponent  $\gamma$  be the same at all temperatures. The model with a finite, temperature-dependent  $\omega_c \approx 3k_B T_{\text{eff}}/\hbar$  fits the temperature dependence data somewhat better than the phenomenological model with  $\alpha = 1.05$  used in the main text at base temperature. Here,  $T_{\text{eff}}$  is the temperature deduced from the data below the crossover. However, there is no significant difference between the fits just looking at 30 mK alone.

dynamic power law model used in the main text,

$$S_{\Phi}^{\text{phen.}}(\omega) = A \frac{\omega}{|\omega|^{\alpha}} [1 + \coth(\hbar\omega/[2k_B T_A])] \quad (32)$$

$$+ B\omega|\omega|^{\gamma-1} [1 + \coth(\hbar\omega/[2k_B T_B])], \quad (33)$$

to the finite high-frequency cutoff model (31). The former implicitly assumes that  $\omega_{\text{max}} \gg k_B T/\hbar$ , so that any deviation of  $S_{\Phi}^+$  from a perfect  $1/f$  scaling near the classical-quantum crossover would be due to temperature alone. Although this seems to fit our data at base temperature, we find that incorporating a finite  $\omega_{\text{max}}$  fits the higher temperature data slightly better under the conditions of the fit (constant  $\alpha$  and  $\gamma$  for all temperatures). Namely, we use the “intermediate” type of crossover,  $\omega_{\text{max}} = 3k_B T_{\text{eff}}$ , where  $T_{\text{eff}}$  is the limiting effective temperature deduced from the stray population data below the crossover. A high frequency cutoff that scales linearly with temperature might have a natural physical meaning. For instance, if the inverse relaxation times  $\tau_n^{-1}$  are determined by spin-phonon interactions, both  $\tau_n^{-1}$  and  $\omega_{\text{max}}$  will be proportional to the number of phonons with energy close to the typical Zeeman splittings of the fluctuators, which scales linearly with  $T$  (since the Zeeman splittings should be  $\ll k_B T$  even for clusters).

### IMPLICATIONS OF HIGH FREQUENCY CUTOFF FOR SPIN DIFFUSION

We mention one more mechanism that would give a frequency-dependent  $\alpha$  due to a finite  $\omega_{\text{max}}$ , but with a different functional form. Namely, we consider spin diffusion, which was proposed by Faoro and Ioffe [26] to explain  $1/f$  noise in SQUIDs and further explored in the context of D-Wave flux qubits by Lanting *et al.* [27]. We conclude that given the  $1/f$  scaling of  $S_{\Phi}^+(f)$  near 1 GHz, spin diffusion is unlikely to be the source of the

$1/f$  noise near the classical-quantum crossover unless i.) the spin diffusion constant is several orders of magnitude higher than estimates in the literature [26, 27] or ii.) the spin density is substantially inhomogeneous, leading to a shallower power law at high frequencies, in which case a separate physical mechanism needs to be invoked for the  $1/f$  power law at lower frequencies.

Within the spin diffusion model, the total spin is conserved and spin excitations will diffuse around the surface of the superconducting qubit metal, generating flux noise by coupling to a non-uniform distribution of the magnetic field. In Ref. [26], the mechanism of diffusion was proposed to be an RKKY interaction mediated through the superconductor. The diffusion equation for the coarse-grained magnetization is

$$\partial M_{\alpha}(\mathbf{r}, t)/\partial t = \mathcal{D}\nabla^2 M_{\alpha}(\mathbf{r}, t), \quad (34)$$

where  $\mathcal{D}$  is the diffusion coefficient and  $\alpha = x, y, z$ . Eq. (34) can be solved using the Laplace transform  $M_{\alpha}(\mathbf{r}) \propto e^{-\Gamma_n t} \varphi_n(\mathbf{r})$ , which leads to the eigenvalue problem

$$\nabla^2 \varphi_n(\mathbf{r}) = -\Gamma_n \varphi_n(\mathbf{r}), \quad (35)$$

with periodic boundary conditions on the surface of the qubit metal. Solving for the eigenmodes  $\varphi_n(\mathbf{r})$  allows one to express the dynamic magnetic susceptibility from the Green’s function for Eq. (34):

$$\chi''_{\alpha\beta}(\mathbf{r}, \mathbf{r}', \omega) = \delta_{\alpha\beta} \chi(0, T) \sum_n \varphi_n(\mathbf{r}) \frac{\omega \Gamma_n}{\omega^2 + \Gamma_n^2} \varphi_n(\mathbf{r}'). \quad (36)$$

The flux noise then becomes a sum of Lorentzians corresponding to each diffusion mode

$$S_{\Phi}(\omega) = \hbar\omega \chi(0, T) \left[ 1 + \coth\left(\frac{\hbar\omega}{2k_B T}\right) \right] \sum_n b_n^2 \frac{\Gamma_n}{\omega^2 + \Gamma_n^2}, \quad (37)$$

where  $b_n^2 = \sum_{\alpha} b_{\alpha n}^2$ ,  $b_{\alpha n} = \int d\mathbf{r} b_{\alpha}(\mathbf{r}) \varphi_n(\mathbf{r})$ , are coupling factors describing how the local moment couples to the qubit’s magnetic field and by reciprocity the qubit loop itself. Therefore the spin diffusion model naturally leads to a broad set of relaxation times  $\tau_n = 1/\Gamma_n$  with a temperature-independent distribution function  $\rho(\tau_n) = b_n^2$  given by the form factors. The temperature dependence of  $S_{\Phi}$  will then arise solely from that of the static magnetic susceptibility  $\chi(0, T)$ .

In the limit of a flat wire (thin film), one would expect a  $1/f^{\alpha=1}$  power law for  $\omega_{\text{min}} \ll \omega \ll \omega_{\text{max}}, k_B T/\hbar$ . But allowing for a finite aspect ratio or for inhomogeneity of the spin density on the surface, one can have an exponent  $\alpha \neq 1$ . In particular, if the spins are concentrated near edges we can have  $\alpha < 1$  [28]. An analytic approximation for the noise summation is [28]

$$S_{\Phi}(\omega) = A \left[ 1 + \coth\left(\frac{\hbar\omega}{2k_B T}\right) \right] \frac{\hbar\omega}{k_B T} \int_0^{\infty} dx \frac{x^{3-2\alpha} e^{-\frac{x}{\sqrt{\omega_{\text{max}}}}}}{\omega^2 + x^4}. \quad (38)$$

One would expect the high frequency cutoff  $\omega_{\max}$  for  $\Gamma_n$  to be given by  $\omega_{\max} \approx \mathcal{D}/\ell^2$ , where  $\ell$  is the smallest dimension associated with the qubit geometry, which for our device should be the metal thickness of  $\sim 100$  nm. The highest estimates in the literature for  $\mathcal{D}$  are  $10^8 - 10^9$  nm<sup>2</sup>/s [26], which would imply a physically expected cutoff of  $\omega_{\max}/(2\pi) \sim 100$  kHz.<sup>4</sup> Therefore, for spin diffusion to be relevant to the classical-quantum crossover, we would need either a much larger  $\mathcal{D}$  and/or  $\alpha < 1$ . If  $\alpha < 1$ , this would mean the quasistatic flux noise (which shows  $\alpha = 1$  below 10 Hz) is not spin diffusion noise, but it may be possible to have different physical mechanisms in different frequency ranges with the noise power still scaling similarly between samples at high and low frequencies.

- 
- [1] T. Lanting, A. J. Berkley, B. Bumble, P. Bunyk, A. Fung, J. Johansson, A. Kaul, A. Kleinsasser, E. Ladizinsky, F. Maibaum, R. Harris, M. W. Johnson, E. Tolkacheva, and M. H. S. Amin, *Phys. Rev. B* **79**, 060509 (2009).
- [2] D. Sank, R. Barends, R. C. Bialczak, Y. Chen, J. Kelly, M. Lenander, E. Lucero, M. Mariantoni, A. Megrant, M. Neeley, P. J. J. O'Malley, A. Vainsencher, H. Wang, J. Wenner, T. C. White, T. Yamamoto, Y. Yin, A. N. Cleland, and J. M. Martinis, *Phys. Rev. Lett.* **109**, 067001 (2012).
- [3] F. Yan, J. Bylander, S. Gustavsson, F. Yoshihara, K. Harrabi, D. G. Cory, T. P. Orlando, Y. Nakamura, J.-S. Tsai, and W. D. Oliver, *Phys. Rev. B* **85**, 174521 (2012).
- [4] R. J. Schoelkopf, A. A. Clerk, S. M. Girvin, K. W. Lehnert, and M. H. Devoret, *Proc. SPIE* **5115**, 356 (2003).
- [5] S. E. de Graaf, A. A. Adamyan, T. Lindström, D. Erts, S. E. Kubatkin, A. Y. Tzalenchuk, and A. V. Danilov, *arXiv preprint arXiv:1609.04562* (2016).
- [6] Y.-H. Lu and H.-T. Chen, *Phys. Chem. Chem. Phys.* **17**, 6834 (2015).
- [7] T. Van Duzer and C. W. Turner, *Principles of superconductive devices and circuits* (1981).
- [8] I. Tupitsyn and S. Kotochigova, *J. Res. Natl. Inst. Stand. Technol.* **103**, 205 (1998).
- [9] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, *Phys. Rev. A* **76**, 042319 (2007).
- [10] G. Catelani, R. J. Schoelkopf, M. H. Devoret, and L. I. Glazman, *Phys. Rev. B* **84**, 064517 (2011).
- [11] I. M. Pop, K. Geerlings, G. Catelani, R. J. Schoelkopf, L. I. Glazman, and M. H. Devoret, *Nature* **508**, 369 (2014).
- [12] J. M. Martinis, M. Ansmann, and J. Aumentado, *Phys. Rev. Lett.* **103**, 097002 (2009).
- [13] 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).
- [14] J. Wenner, Y. Yin, E. Lucero, R. Barends, Y. Chen, B. Chiaro, J. Kelly, M. Lenander, M. Mariantoni, A. Megrant, C. Neill, P. J. J. O'Malley, D. Sank, A. Vainsencher, H. Wang, T. C. White, A. N. Cleland, and J. M. Martinis, *Phys. Rev. Lett.* **110**, 150502 (2013).
- [15] D. Ristè, C. Bultink, M. Tiggelman, R. Schouten, K. Lehnert, and L. DiCarlo, *Nat. Commun.* **4**, 1913 (2013).
- [16] U. Vool, I. M. Pop, K. Sliwa, B. Abdo, 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, *Phys. Rev. Lett.* **113**, 247001 (2014).
- [17] C. Wang, Y. Y. Gao, I. M. Pop, U. Vool, C. Axline, T. Brecht, R. W. Heeres, L. Frunzio, M. H. Devoret, G. Catelani, *et al.*, *Nature communications* **5** (2014).
- [18] X. Y. Jin, A. Kamal, A. P. Sears, T. Gudmundsen, D. Hover, J. Miloshi, R. Slattery, F. Yan, J. Yoder, T. P. Orlando, S. Gustavsson, and W. D. Oliver, *Phys. Rev. Lett.* **114**, 240501 (2015).
- [19] R. Barends, J. Wenner, M. Lenander, Y. Chen, R. C. Bialczak, J. Kelly, E. Lucero, P. O'Malley, M. Mariantoni, D. Sank, H. Wang, T. C. White, Y. Yin, J. Zhao, A. N. Cleland, J. M. Martinis, and J. J. A. Baselmans, *Appl. Phys. Lett.* **99**, 113507 (2011).
- [20] F. Yan, S. Gustavsson, A. Kamal, J. Birenbaum, A. P. Sears, D. Hover, T. J. Gudmundsen, D. Rosenberg, G. Samach, S. Weber, J. L. Yoder, T. P. Orlando, J. Clarke, A. J. Kerman, and W. D. Oliver, *arXiv preprint arXiv:1508.06299* (2016).
- [21] M. Büttiker and R. Landauer, *Phys. Rev. Lett.* **49**, 1739 (1982).
- [22] A. J. Leggett, *Phys. Rev. B* **30**, 1208 (1984).
- [23] J. M. Martinis, M. H. Devoret, D. Esteve, and C. Urbina, *Physica B: Condensed Matter* **152**, 159 (1988).
- [24] D. Esteve, J. M. Martinis, C. Urbina, E. Turlot, M. H. Devoret, H. Grabert, and S. Linkwitz, *Physica Scripta* **1989**, 121 (1989).
- [25] S. Sendelbach, D. Hover, A. Kittel, M. Mück, J. M. Martinis, and R. McDermott, *Phys. Rev. Lett.* **100**, 227006 (2008).
- [26] L. Faoro and L. B. Ioffe, *Phys. Rev. Lett.* **100**, 227005 (2008).
- [27] T. Lanting, M. H. Amin, A. J. Berkley, C. Rich, S.-F. Chen, S. LaForest, and R. de Sousa, *Phys. Rev. B* **89**, 014503 (2014).
- [28] A. G. Petukhov, V. N. Smelyanskiy and J. M. Martinis (in preparation).

---

<sup>4</sup> Conversely, a low-frequency cutoff of  $\sim 10$  Hz would be expected, which would mean the noise below 1 Hz would need to have a different physical mechanism, unless the diffusion constant is much

lower than expected. The “bump” in flux noise observed at intermediate frequencies in the main paper could potentially be due to spin diffusion, assuming it is not suppressed by spin relaxation.