

Efficient Qubit Calibration by Binary-Search Hamiltonian Tracking

Fabrizio Berritta,^{1,2} Jacob Benestad,³ Lukas Pahl,^{1,4} Melvin Mathews,^{1,5} Jan A. Krzywdka,⁶
Réouven Assouly,¹ Youngkyu Sung,^{1,4} David K. Kim,⁷ Bethany M. Niedzielski,⁷
Kyle Serniak,^{1,7} Mollie E. Schwartz,⁷ Jonilyn L. Yoder,⁷ Anasua Chatterjee,^{2,8}
Jeffrey A. Grover,¹ Jeroen Danon,³ William D. Oliver,^{1,4,9} and Ferdinand Kuemmeth^{2,10,11,*}

¹*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

²*Center for Quantum Devices, Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark*

³*Center for Quantum Spintronics, Department of Physics, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway*

⁴*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

⁵*Department of Information Technology and Electrical Engineering, ETH Zürich, 8093 Zürich, Switzerland*

⁶*(aQa^L) Applied Quantum Algorithms — Lorentz Institute for Theoretical Physics & Leiden Institute of Advanced Computer Science, Universiteit Leiden, The Netherlands*

⁷*Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA 02421, USA*

⁸*QuTech and Kavli Institute of Nanoscience, Delft University of Technology, Delft, The Netherlands*

⁹*Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

¹⁰*Institute of Experimental and Applied Physics, University of Regensburg, 93040 Regensburg, Germany*

¹¹*QDevil, Quantum Machines, 2750 Ballerup, Denmark*

(Dated: August 26, 2025)

We present and experimentally implement a real-time protocol for calibrating the frequency of a resonantly driven qubit, achieving exponential scaling in calibration precision with the number of measurements, up to the limit imposed by decoherence. The real-time processing capabilities of a classical controller dynamically generate adaptive probing sequences for qubit-frequency estimation. Each probing evolution time and drive frequency are calculated to divide the prior probability distribution into two branches, following a locally optimal strategy that mimics a conventional binary search. The scheme does not require repeated measurements at the same setting, as it accounts for state preparation and measurement errors. Its use of a parametrized probability distribution favors numerical accuracy and computational speed. We show the algorithm’s efficacy by stabilizing a flux-tunable transmon qubit, leading to improved coherence and gate fidelity. As benchmarked by gate set tomography, the FPGA-powered control electronics partially mitigates non-Markovian noise, which is detrimental to quantum error correction. The mitigation is achieved by dynamically updating and feeding forward the qubit frequency. Our protocol highlights the importance of feedback in improving the calibration and stability of qubits subject to drift and can be readily applied to other qubit platforms.

I. INTRODUCTION

Quantum processing units (QPUs) with tens of qubits are becoming increasingly common [1]. Efforts to reduce error rates have resulted in the first demonstrations of quantum error correction [2, 3], though these advances come with costly calibration overhead as the number of qubits increases. Temporal instability in QPUs arises from numerous stochastic noise channels [4], and the lowest-performing outlier qubits often limit overall performance [5]. Tackling time-dependent fluctuations and outlier qubits in large QPUs requires scalable and efficient calibration methods to ensure stable, fault-tolerant operation while minimizing calibration downtime. We introduce here a protocol that features exponential scaling of the calibration precision versus the number of measurements until it is limited by decoherence. This scaling is

the result of a locally optimal strategy that maximizes the expected precision for the next measurement.

Online Hamiltonian learning [6] offers a promising strategy for addressing drifts in stochastic qubit parameters through real-time estimation [6–13], facilitated by modern field-programmable gate array (FPGA) hardware advancements. Although various estimation methods [14–17] have been proposed to enhance calibration efficiency, no experimental implementation has yet achieved real-time adaptive estimation in a resonantly driven qubit with a parametrized probability distribution. This work fills the gap by the experimental demonstration of a real-time Bayesian estimation approach applied to a superconducting qubit.

To demonstrate the adaptive and efficient calibration protocol, we employ a flux-tunable transmon qubit [18, 19]. Transmons are largely insensitive to charge noise by design. The tunable qubit frequency is controlled by a magnetic field and it is susceptible to random flux variations, known as flux noise. Typically, flux-tunable transmons are thus operated at bias points (sweet spots) that

* ferdinand.kuemmeth@ur.de

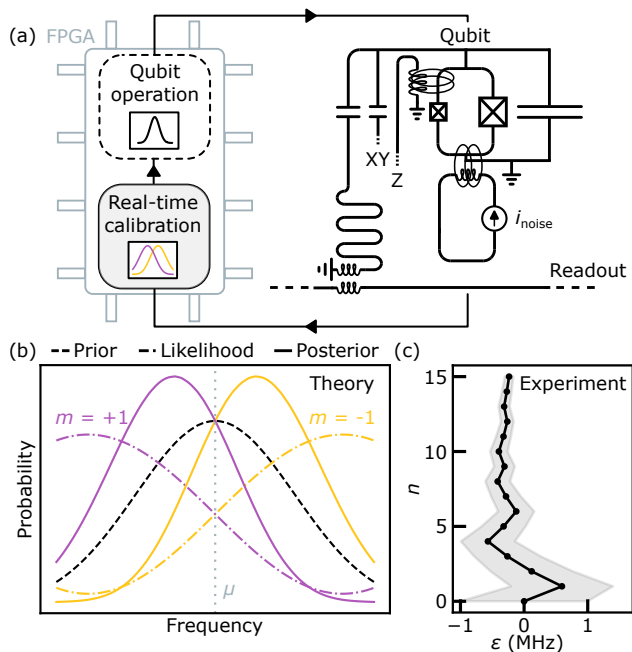


FIG. 1. **Qubit implementation and frequency binary search.** (a) Experiment schedule, alternating between periods of quantum information processing (dashed box), and periods for efficiently calibrating the qubit frequency (gray box). The transmon (Qubit) is controlled by microwave pulses (XY) that are based on previous qubit measurements (Readout). (b) Evolution of the probability distribution $\mathcal{P}(\epsilon)$ during the frequency binary search algorithm. For each Ramsey probing cycle, the probing time and frequency detuning are chosen such that the two possible likelihood functions (dash-dotted lines) multiplied with the prior distribution (dashed) yield posterior distributions (solid) that are shifted left or right, while the standard deviation is reduced (see main text). (c) The controller updates the probability distribution after each measurement. The dots show the resulting expectation value (ϵ) and the shaded area marks the 68% credible interval.

are first-order insensitive to small flux changes. However, operating them away from these bias points may be necessary to reduce frequency crowding in large qubit arrays [20, 21] or minimize coupling to two-level fluctuators. To maintain high-fidelity operation despite the increased sensitivity to flux noise, one may use real-time frequency estimation and stabilization [22].

The so-called frequentist approach is the most common method for estimating fluctuations in Hamiltonian parameters of superconducting qubits. Frequentist estimation methods infer an observable from the observed frequencies of measurement outcomes in repeated experiments and they do not yield an exponential scaling with the number of measurements. For instance, in Ref. [22] the qubit frequency fluctuation is computed in real time by repeated Ramsey interferometry cycles. In the probing Ramsey cycle, the Bloch vector is initialized on the equator of the Bloch sphere, evolves under its Hamiltonian, and is projected back onto the \hat{z} -axis after a fixed

(*non-adaptive*) evolution time. The measurement outcome is averaged over many probing cycles. The fraction of times the system is measured in the excited state, for instance, is linearly mapped to the qubit frequency fluctuation, as further detailed in the next section. The drawbacks of such frequentist methods are that (i) they are not optimally efficient, (ii) there is a trade-off between frequency sensitivity and probing range, and (iii) the qubit must already be approximately calibrated for the stabilization to work. In contrast, a Bayesian estimation approach assigns probabilities to hypotheses and naturally incorporates real-time optimization techniques [6, 8, 9, 12, 13]. At any given time, the current knowledge of the parameters can be used to *adaptively* select the optimal experimental settings for the next probing cycle. While such techniques have been widely adopted in nitrogen-vacancy centers and semiconductor spin qubits over the past decade, they have not been similarly applied to superconducting qubits.

In this work, we employ an adaptive Bayesian estimation scheme to calibrate a transmon qubit, which circumvents the limitations of the frequentist approach mentioned above. Our protocol involves programming a commercial controller with an integrated FPGA to generate dynamic microwave pulses, enabling real-time qubit frequency calibration through a binary search algorithm inspired by Ref. [23]. We use the real-time capabilities of the controller (on the qubit coherence timescale) to select optimal parameters within an adaptive estimation cycle. Specifically, the controller dynamically adjusts the drive frequency and the evolution time of each probing cycle based on previous measurement results. We demonstrate the algorithm's efficacy through improved effective qubit coherence T_2^* and fidelity.

While the majority of theoretical research concentrates on Markovian noise [24], non-Markovian noise sources such as $1/f$ flux noise, which exhibit memory effects, are common in solid-state qubit platforms [13]. They introduce significant overhead for error mitigation [25, 26], their dynamical decoupling [27, 28] is not universally effective and may not align with specific experimental goals. In order to realize fault-tolerant quantum computing with error-corrected solid-state qubits, non-Markovian noise likely needs to be reduced or eliminated [29]. We show how our real-time frequency tracking protocol reduces such noise as validated by gate set tomography [30], whereas Ref. [22] relied solely on randomized benchmarking [31]. While randomized benchmarking blends the non-Markovian noise with other error types by introducing random gate sequences, gate set tomography assumes Markovian noise and can therefore detect non-Markovianity in the system through model violations. Our scheme enables intermittent and efficient calibration of qubit frequencies, making it ideal for stable quantum circuit execution in the presence of drift.

II. METHOD

A. Device

We use a 2×2 superconducting qubit array operated at the mixing-chamber stage (below 30 mK) of a dilution refrigerator. The array design is based on Ref. [32], and we calibrate one of its flux-tunable transmon qubits. A commercial controller (Quantum Machines OPX+ and Octave) applies high-frequency waveforms to the lines for qubit control and single-shot readout. A Yokogawa GS200 provides the DC flux bias through a bias-tee at room temperature (see the Supplemental Material [33] for details on the experimental setup). The transmon comprises a DC SQUID with asymmetric junctions, and on-chip Z and XY control lines [see Fig. 1(a)]. The qubit is dispersively coupled to a coplanar waveguide resonator for state readout [18, 19].

The real-time calibration we develop here can be embedded in an operation loop as sketched in Fig. 1(a). The calibration protocol is used to estimate the qubit frequency, after which the resulting estimate is used to adjust the frequency of the drive accordingly, leading to a decreased sensitivity to flux noise during qubit operation.

Our goal is to compensate for qubit frequency fluctuations particularly when it is operated away from the sweet spot. Thus we tune the transmon where the effect of flux noise is the strongest, at $\Phi_{\text{ext}} = \Phi_0/4$, where $\Phi_0 = h/(2e)$ is the superconducting magnetic flux quantum and Φ_{ext} is the external magnetic flux applied via a mutual coupling to the Z line. Our qubit nominally has a transition frequency $f_q \approx 3.78$ GHz and in the rotating frame its Hamiltonian is

$$\frac{\mathcal{H}(t)}{h} = -\frac{\Delta f - \varepsilon(t)}{2}\sigma_z, \quad (1)$$

where σ_z is the Pauli z -matrix, $\Delta f = f_d - f_q$ is the frequency detuning between f_q and the chosen rotating frame frequency f_d , and $\varepsilon(t)$ represents the time-dependent shift of the qubit frequency due to flux noise.

B. Frequency binary search by Bayesian estimation

The frequency binary search (FBS) algorithm involves a Ramsey probing cycle where the qubit is first prepared in a superposition state $|\psi\rangle = (|0\rangle - i|1\rangle)/\sqrt{2}$ using a $X_{\pi/2}$ rotation around the \hat{x} -axis of the Bloch sphere. The state preparation is followed by a period of free evolution for a duration τ , during which the qubit state acquires a phase $\phi = 2\pi \int_0^\tau \delta_q(t') dt'$, where $\delta_q(t) = \Delta f - \varepsilon(t)$. A second $X_{\pi/2}$ pulse is then applied, and the state of the qubit is measured using dispersive readout. By the quasistatic noise approximation, we assume $\varepsilon(t)$ to be constant on the scale of a few probing cycles, resulting in $\phi = 2\pi[\Delta f - \varepsilon(t)]\tau$. In the following, we drop the time dependence of $\varepsilon(t)$ for ease of notation.

We assume that the probability of measuring an outcome $m \in \{-1, 1\}$ corresponding to the states $|0\rangle$ and $|1\rangle$ is given by the likelihood function

$$P(m|\varepsilon, \Delta f, \tau) = \frac{1}{2} + \frac{m}{2} \{\alpha + \beta e^{-\tau/T} \cos[2\pi(\Delta f - \varepsilon)\tau]\}, \quad (2)$$

where the parameters $\alpha = -0.02$ and $\beta = 0.6$ are coefficients that together capture state preparation and readout errors, while T describes a coherence time that limits the duration of the evolution time τ (see the Supplemental Material [33] for further details on how they impact the algorithm performance). We set $T = 10 \mu\text{s}$ based on the Hahn echo coherence time $T_{2E} \approx 12 \mu\text{s}$. We choose a value slightly lower than the estimated T_{2E} to be on the safe side in case the intrinsic decoherence rate fluctuates on longer time scales. For reference, $T_2 \approx 42 \mu\text{s}$ at $\Phi_{\text{ext}} = 0$. The qubit fluctuation ε is what we want the controller to estimate.

We emphasize that in this experiment the controller *adaptively* computes in real time both the Ramsey evolution time τ and the drive frequency f_d (and thus Δf) for each probing cycle. This differs from previously employed approaches where (i) the evolution time τ is fixed [22] and (ii) only either the evolution time τ or the effective rotating frame frequency is chosen adaptively [8, 12, 34, 35].

Adaptive control of the phase theoretically allows for one bit of information to be gained per experiment in an error-free qubit. The estimation begins with the longest sensing time, which corresponds to the smallest frequency sensitivity [36], and proceeds with shorter sensing times that yield larger frequency resolutions. To improve robustness against experimental errors, the strategy can be modified by repeating measurements with reduced information gain per measurement [37, 38], which has been experimentally demonstrated using nitrogen-vacancy centers [34, 39, 40]. However, this phase estimation approach presents some limitations: (i) an extensive parameter search is required to determine the optimal number of repetitions for each sensing time [34], (ii) a Bayesian approach would yield a multi-modal probability distribution that requires many more than two parameters to be stored in memory, and (iii) the estimate becomes only useful after the very last cycle.

We also highlight that using a fixed evolution time τ in non-adaptive probing cycles [22] limits the frequency range to $(-\frac{1}{2\tau}, +\frac{1}{2\tau})$, which results in a trade-off between range and frequency sensitivity. This trade-off does not apply to the binary search algorithm introduced here.

In the quasistatic approximation, we want to find the optimal sequence of τ 's and f_d 's to estimate ε in as few measurements as possible. Using Bayes' rule we write

$$\mathcal{P}_{n+1}(\varepsilon) \propto \mathcal{P}_n(\varepsilon) P(m_{n+1}|\varepsilon, \Delta f_{n+1}, \tau_{n+1}), \quad (3)$$

where $\mathcal{P}_n(\varepsilon)$ denotes the probability distribution for ε after the n^{th} probing cycle, which thus depends on all previously used Δf_i and τ_i and all measurement outcomes m_i (i.e., $i \leq n$).

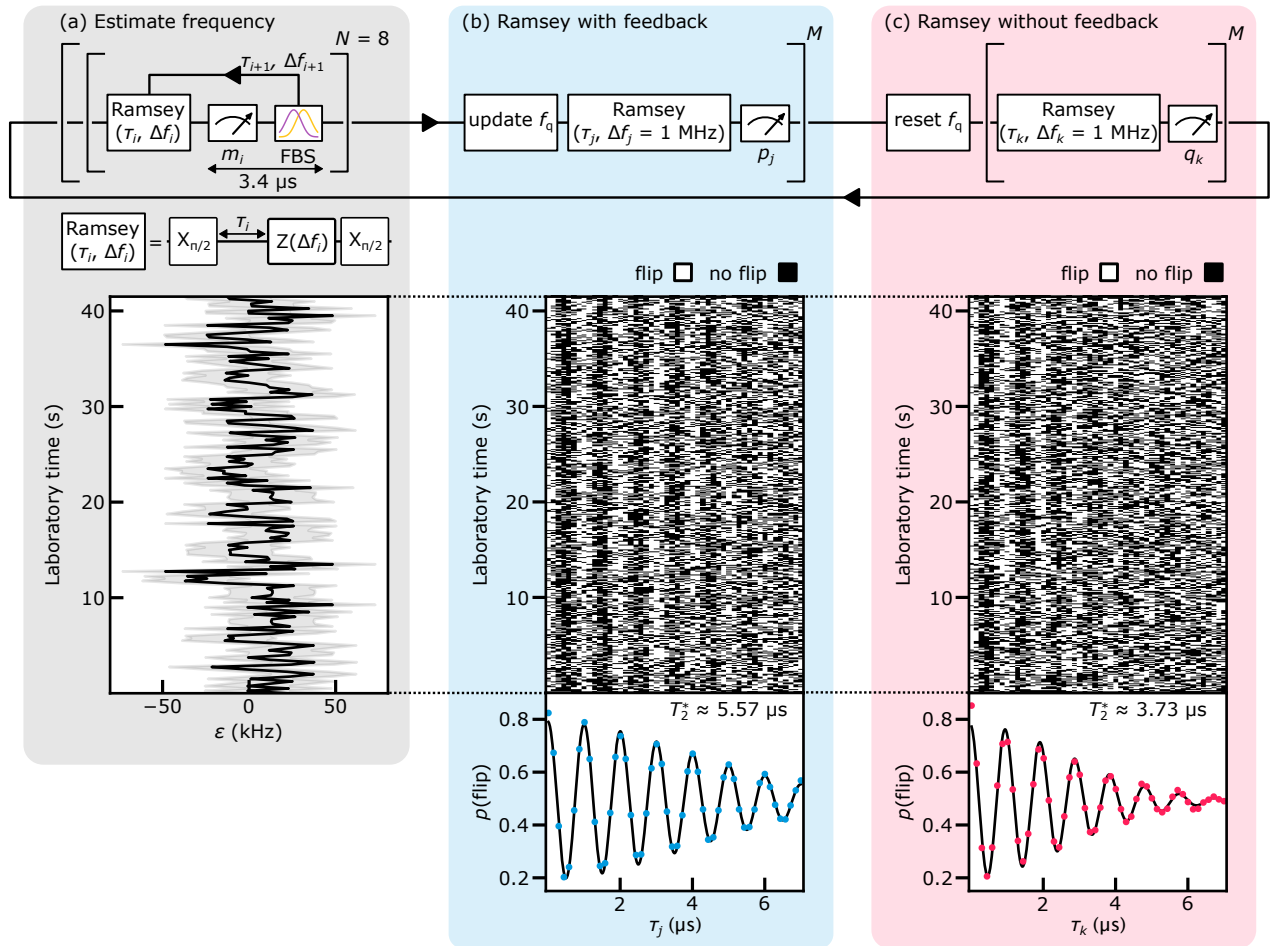


FIG. 2. **Frequency binary search validation by suppressed dephasing of a qubit in a feedback-controlled rotating frame.** (upper diagram) One loop (solid arrows) represents one repetition of the protocol. **(a)** For each estimate of the frequency shift ε , the controller uses $N = 8$ Ramsey probing cycles with adaptive probe times τ_i and detuned frequencies Δf_i , computed in real time. Bottom panel: $\langle \varepsilon \rangle$ (solid line) and 68% credible interval (shaded area) of the final probability distribution $\mathcal{P}_8(\varepsilon)$ of all estimates performed during the ≈ 41.5 s of the experiment. **(b)** After each estimation, we task the controller to perform a Ramsey cycle with evolution time τ_j , while setting $\Delta f_j = 1$ MHz by adjusting f_d using the latest estimate $\langle \varepsilon \rangle$. These interleaved FBS estimations and Ramsey cycles are repeated $M = 50$ times with evenly spaced τ_j . Single-shot measurement outcomes p_j are plotted in the middle panel as white/black pixels, and the fraction of flipped outcomes in each column is shown as a blue dot in the lower panel. **(c)** Subsequently the controller resets all frequencies to the offline-calibrated values, i.e., assumes $\langle \varepsilon \rangle = 0$, and performs again $M = 50$ evenly spaced Ramsey cycles of set $\Delta f_k = 1$ MHz. Single-shot measurement outcomes q_k are plotted as white/black pixels in the middle panel, and the fraction of flipped outcomes in each column is shown as a red dot in the lower panel. Comparing (b) to (c), the coherence time improves by $\approx 49\%$ with feedback.

In the following we assume $\mathcal{P}_n(\varepsilon)$ to be a Gaussian, which well approximates the actual distribution (see the Supplemental Material [33] for more details). The advantage of using the Gaussian approximation is that it allows the distribution to be conveniently described using just two parameters: the mean μ and the standard deviation σ .

For each individual probing cycle, to minimize the expected posterior variance, in a greedy approach the optimal experiment $(\tau, \Delta f)$ is the one that divides the prior distribution into equal (or approximately equal) left and right portions [see Fig. 1(b)] so that a half period of Eq. (2) is comparable to the width of the prior distribution,

performing something similar to a binary search partitioning [41]. The approach of using Gaussians and partitioning the posterior distributions in a similar fashion was introduced by Ref. [23]. However, in that case, the lack of phase control in the likelihood prevents the algorithm from determining the sign of ε , making it ineffective for priors where $\mu/\sigma \rightarrow 0$ [42]. While fitting to a bimodal Gaussian has been suggested as a solution [43], this approach still does not resolve the ambiguity of the sign. In this work, we address both aspects by dynamically updating f_d (and thus Δf) in real time on the controller. This allows the controller to consistently use the optimal likelihood function and perform the most ef-

ficient estimation using only a Gaussian distribution.

Two steps are needed to implement the binary search in the controller. The first step is to determine optimal experiment parameters τ and Δf based on the prior distribution. Then, one must approximate the resulting posterior $\mathcal{P}_{n+1}(\varepsilon)$ to a Gaussian. This can be done using equations implemented directly on the controller in real time.

The prior distribution, dashed black line in Fig. 1(b), in the Gaussian approximation is given by

$$\mathcal{P}_n(\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma_n^2} \exp\left\{-\left(\frac{\varepsilon - \mu_n}{\sqrt{2}\sigma_n}\right)^2\right\}, \quad (4)$$

where μ_n is our guess for what ε is with uncertainty σ_n . In order to separate the prior at the center by the likelihood functions, the detuned frequency must satisfy

$$\Delta f_{n+1} = \frac{1+2l}{4\tau_{n+1}} + \mu_n, \quad (5)$$

where l can be any integer number; we set it to be zero. While this is formally the optimal solution for $\alpha = 0$, it is a reasonable approximation for the optimum at other realistic values of α . Given this constraint, the optimal evolution time τ that minimizes the posterior variance is given by:

$$\tau_{n+1} = \frac{\sqrt{16\pi^2\sigma_n^2 + 1/T^2} - 1/T}{8\pi^2\sigma_n^2}. \quad (6)$$

After measuring $m_{n+1} = \pm 1$, the posterior distribution is obtained by using Eq. (3), after which it is fit to a Gaussian using the method of moments, i.e., by calculating its mean and variance [23]

$$\mu_{n+1} = \mu_n - \frac{2\pi m_{n+1} \beta \sigma_n^2 \tau_{n+1} e^{-\tau_{n+1}/T - 2\pi^2 \sigma_n^2 \tau_{n+1}^2}}{1 + m_{n+1} \alpha}, \quad (7a)$$

$$\sigma_{n+1}^2 = \sigma_n^2 - \frac{4\pi^2 \beta^2 \sigma_n^4 \tau_{n+1}^2 e^{-2\tau_{n+1}/T - 4\pi^2 \sigma_n^2 \tau_{n+1}^2}}{(1 + m_{n+1} \alpha)^2}, \quad (7b)$$

and using these values to construct the new prior to be a Gaussian distribution with mean and variance μ_{n+1} and σ_{n+1}^2 [see Fig. 1(b)]. This scheme is repeated N times to obtain a sufficiently narrow distribution with exponential scaling with the number of measurements given by Eq. (7b) [44].

In Fig. 1(c) we illustrate the resulting evolution of $\mathcal{P}_n(\varepsilon)$ as a function of the measurement number n for one representative estimation sequence, with an initial prior with $\sigma_0 = 1$ MHz and $\mu_0 = 0$, and $N = 15$. The black dots show the expectation value $\langle \varepsilon \rangle$, and the shaded area indicates its 68% credible interval, narrowing here down to 67 kHz after 15 measurements in less than 100 μ s. With similar parameters, this efficient frequency estimation could replace currently used periodic calibration routines by narrowing the qubit frequency within

a user-defined error budget. As the FBS relies only on single-qubit operations, it can be extended to multi-qubit systems without introducing additional overhead.

In the following, we optimize the number of single-shot measurements N we use based on the specific experiment we are performing. We test the performance of the FBS by performing Ramsey repetitions to verify that it finds the correct qubit frequency, and we use randomized benchmarking and gate set tomography to assess the impact of the FBS on the qubit fidelity and drift. We emphasize again that the goal of these experiments is not to increase qubit coherence time or fidelity as much as possible, but simply to confirm that the information gained from the FBS is most likely correct.

III. RESULTS

A. Qubit coherence time T_2^*

We thus validate the FBS by rapid estimation of the shift of the qubit frequency ε and demonstrating an extended coherence of the flux-tunable transmon qubit.

The fluctuating parameter ε is estimated from the probing sequence shown in the top part of Fig. 2(a): For each probe cycle, the Bloch vector is positioned on the equator using an $X_{\pi/2}$ pulse. After precessing for a time τ_i , a virtual $Z(\Delta f_i)$ gate [45] is performed, introducing a phase offset $\phi = 2\pi \Delta f_i \tau_i$. Here, virtual means that the phase offset is added to the subsequent $X_{\pi/2}$ pulse, which projects the Bloch vector back onto the \hat{z} -axis. The virtual Z gate controls the phase $\Delta f \tau$ of the likelihood function [Eq. (2)] based on Eq. (5). This is followed by a measurement, after which the qubit state s_i , ground ($s_i = 0$) or excited ($s_i = 1$), is assigned by thresholding the demodulated dispersive readout signal on the controller.

The Bayesian probability distribution of ε is updated after comparing the measurement outcome s_i to the previous one s_{i-1} , as we do not initialize the qubit to the ground state at the beginning of each cycle to reduce the probing cycle period (thus yielding higher feedback bandwidth). It follows that in Eq. (2) $m_i = 2|s_i - s_{i-1}| - 1$. For instance, if in the previous measurement the qubit was in the ground state $s_{i-1} = 0$ and now it is in the excited state ($s_i = 1$) then $m_i = +1$. The duration of the probe cycle is short compared to the measured $T_1 \approx 80 \mu$ s at quarter flux: the average Ramsey evolution time is $\approx 4.36 \mu$ s, the time used to read out the qubit is 1.44 μ s, and the time used to subsequently cool down the resonator again (to deplete it from any residual photons after the readout pulse) is $\approx 2 \mu$ s.

The controller is programmed to start from a prior with $\mu_0 = 0$ and $\sigma_0 = 30$ kHz based on previously measured qubit frequency fluctuations. After each measurement, the controller updates the probability distribution, resulting in a narrowing of the width of the distribution [cf. Fig. 1(c)]. In the middle panel of Fig. 2(a) we plot

the experimentally found final posterior probability distributions $\mathcal{P}_N(\varepsilon)$ from all estimations done in the full experiment, where we consistently use $N = 8$ measurements per estimation sequence. For plotting clarity we down-sample to ≈ 14 ms.

To confirm that these narrowed distribution functions indeed represent more accurate and correct knowledge about ε , we interleave the estimations with a series of Ramsey repetitions, half of them making use of the outcome of the estimations and half of them not, as shown schematically in the upper part of Fig. 2. After each estimation sequence of ε , the controller updates the qubit-frequency parameter f_q (in software) to compensate for the estimated shift $\langle\varepsilon\rangle$, so that the total expected detuning becomes $\Delta f_j - \langle\varepsilon\rangle = 1$ MHz. Then it performs one Ramsey cycle with evolution time τ_j , after which it again estimates ε . The intentional detuning of 1 MHz applied via the virtual Z gate improves the visibility of the Ramsey fringes, making them easier to fit. This cycle is repeated $M = 50$ times, while τ_j is increased linearly from 0 to $7\ \mu\text{s}$. Each row in the middle panel of Fig. 2(b) shows the result of one such set of 50 Ramsey cycles, where we plot all single-shot measurement outcomes p_j as white/black pixels. After each set of $M = 50$ Ramsey cycles with feedback, the controller resets f_q (in software) to the offline-calibrated value, tuning to $\Delta f_k = 1$ MHz. It then performs the same series of 50 Ramsey cycles as before, linearly stepping the evolution time τ_k from 0 to $7\ \mu\text{s}$. The rows in the middle panel of Fig. 2(c) show the single-shot measurement outcomes of this part of the protocol. After completing the Ramsey cycles without feedback, the whole protocol is repeated, as indicated in the top of Fig. 2, starting with a new estimation of ε . In the new estimation sequence, the controller starts from a prior with μ_0 equal to the estimated ε of the previous sequence.

To highlight the effect of the feedback, we plot the averages of all Ramsey repetitions in the lower panels of Fig. 2(b,c). In both cases, we fit the decaying signal using Gaussian envelopes (solid line), yielding $T_2^* = (3.73 \pm 0.11)\ \mu\text{s}$ without feedback and $T_2^* = (5.57 \pm 0.09)\ \mu\text{s}$ with feedback, corresponding to a $\approx 49\%$ improvement when including feedback. This increased coherence of the qubit demonstrates that the narrowed distribution function obtained from the estimation procedure indeed reflects more accurate knowledge about the fluctuating parameter ε . Notably, our approach requires fewer single-shot measurement outcomes ($N = 8$) compared to $N = 20$ of Ref. [22]. The T_2^* coherence time achieved with feedback is ultimately limited by the Hahn echo value due to the lower bandwidth of the feedback loop (approximately 16 kHz for $N = 8$), compared to the much higher bandwidth of the dynamically decoupled Ramsey ($1/\tau \gtrsim 100$ kHz). This explains why the Hahn echo is less sensitive to quasistatic fluctuations than the feedback-stabilized Ramsey experiment.

We perform another experiment for 6 hours, from which we extract the noise power spectral density and

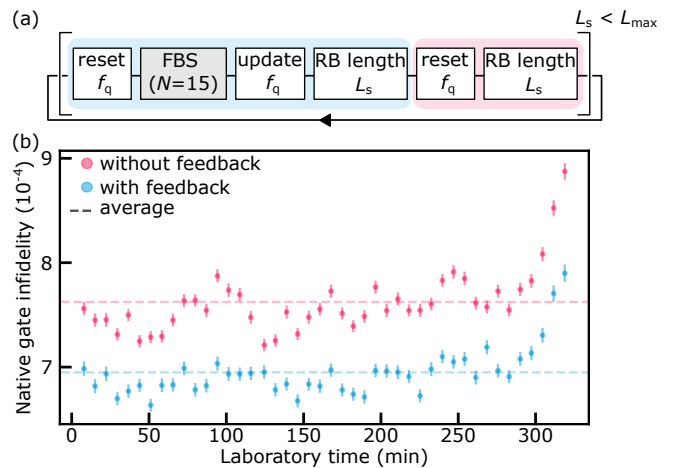


FIG. 3. **Frequency binary search validation by randomized benchmarking.** (a) A randomized benchmarking repetition of length L_s combined with the FBS feedback (with $N = 15$ probing cycles). (b) Randomized benchmarking is repeated 44 times with the feedback (blue dots) and without (red dots). The native gate errors are extracted from the fit to the data, which consists of 30 random Clifford sequences with and without the feedback. The error bars show the 68% confidence interval for the fitted gate error. Dashed lines are the averaged gate infidelities.

show that the FBS keeps track of the qubit frequency also over longer time scales (see the Supplemental Material [33]). Overall, the results presented in this section show how the FBS can efficiently find and stabilize the qubit frequency.

B. Single-qubit gate fidelity

In this section we further validate the FBS calibration by showing improvement of the single-qubit gate fidelity by randomized benchmarking (RB) [31].

The pulse sequence is shown in Fig. 3(a): the controller resets the qubit-frequency parameter f_q (in software), and it performs the FBS with $N = 15$ probing sequences, starting with $\mu_0 = 0$ and $\sigma_0 = 200$ kHz [46]. At the end of the estimation, f_q is updated (in software) and an RB sequence of depth L_s is performed. An interleaved RB measurement without feedback follows by resetting the qubit frequency to the offline-calibrated value. The maximum circuit depth (L_{\max}) is 2,300, and the repetition is averaged 10,000 times. We implement DRAG (derivative reduction by adiabatic gate) for our 20 ns-long pulses to suppress leakage errors [18, 19] in the RB experiment. As with the previous experiment, the qubit is not initialized in the ground state at the beginning of each cycle; the controller keeps track of whether the state is different or not compared to the previous measurement. Every 1,000 averages, the threshold that classifies the demodulated dispersive readout signal is updated online in the con-

troller by taking the average of 10,000 single-shot measurements after performing an $X_{\pi/2}$ pulse to the qubit.

The controller performs the RB experiment for 6 hours, yielding a native gate infidelity of $(7.6 \pm 0.3) \times 10^{-4}$ without feedback and $(6.9 \pm 0.2) \times 10^{-4}$ with feedback [dashed lines in Fig. 3(b)]. The single-qubit gate infidelities are higher than the decoherence limit [47] approximated by $t_{\text{gate}}(1/T_1 + 1/T_\phi)/3 \approx 6 \times 10^{-4}$, given $t_{\text{gate}} \approx 20$ ns, $T_1 \approx 80$ μ s and the exponential part of the pure dephasing time $T_\phi \approx 13$ μ s. The feedback protocol always performs better than without feedback, and with less spread around the mean value as a result of the stabilization. Some of the drifts of the infidelity remain correlated, which we tentatively attribute to other factors (e.g., changes in T_1). The improved fidelities of single-qubit gates by feedback by interleaved RB are presented in the Supplemental Material [33]. We attribute the smaller relative improvement in the single-qubit gate qubit fidelity, compared to coherence in the previous section, to the larger final value of $\sigma_{N=15} \approx 90$ kHz, as opposed to $\sigma_{N=8} \approx 24$ kHz, which resulted from different initial prior distributions.

C. Reduction of non-Markovian noise

So far, we have demonstrated an efficient qubit calibration protocol by instructing a controller to generate adaptive probing sequences in real time. The calibration has been validated by improved coherence and fidelity. Next, we investigate if our estimation protocol also reduces the non-Markovian noise in the system, this time estimated by gate set tomography (GST) [30]. While GST has been employed to validate the mitigation of drifts in semiconductor spin qubits [13] by real-time estimation, we present its application for validating drift mitigation in a superconducting qubit by real-time feedback.

GST is a calibration-free method for benchmarking and characterizing operations in a quantum processor and we implement it by using the pyGSTi [48] software package. GST relies on running circuits designed to amplify certain types of errors. Each circuit, as shown in the right panel of Fig. 4(a), consists of the gate to be characterized, $\{G_1, G_x, G_y, G_x G_y, G_x G_x G_y\}$ (the germs), sandwiched between two fiducial circuits from the set $\{I, X_{\pi/2}, Y_{\pi/2}, X_{\pi/2} X_{\pi/2}, X_{\pi/2} X_{\pi/2} X_{\pi/2}, Y_{\pi/2} Y_{\pi/2} Y_{\pi/2}\}$. These are applied to the qubit after initialization, to generate the state ρ_i , and before the measurement in the computational basis, to measure the operator E_j . The variable L_s is the maximum depth used to construct the base circuit from the germ [30]. By running the circuit multiple times, we extract the measurement probabilities $P_{ji} = \text{Tr}\{E_j G[\rho_i]\}$, which are then fitted to the error model, providing a maximum likelihood estimator of the transfer matrix for all of the gates. The quality of the fit can be used to detect violations of the Markovian model of the noise [30, 49].

We incorporate the feedback loop inside the GST protocol using the scheme shown in the left panel of Fig. 4(a), implemented on the controller. We find suitable parameters for testing the FBS to be $N = 6$ per estimation sequence with feedback, with initial $\mu_0 = 0$ and $\sigma_0 = 30$ kHz. Each repetition is executed 100 times for each of the 616 sequences constructed for GST. In the new estimation sequence, the controller starts from a prior with μ_0 equal to the estimated ε of the previous repetition.

The measured discrepancy between the Markovian model and the data ($2\Delta \log L_s$ as defined in Ref. [30]) is shown in Fig. 4(b) using the color of the squares: The grayscale is used to reflect the total discrepancy and a red color signals a model violation, indicating non-Markovian noise at 95% confidence interval. The rows and columns of the inner matrices represent the fiducial operations used for state preparation and measurement respectively. For the data with feedback, represented in the right panel of Fig. 4(b), a clear decrease in the number of red squares in comparison to the feedback-free result of the left panel is visible.

To further quantify the reduction of non-Markovian noise, the controller repeats the GST experiment 30 times, spanning a laboratory time of 40 minutes. In Fig. 4(c), we plot, as a function of the maximum length, the total violation of the model. While there is no clear improvement for maximum lengths of 1 and 2, it is evident for longer sequences. As shown in the Supplementary Material [33], feedback allows to significantly decrease the amplitude of non-Markovian noise, identified as frequency fluctuations below the inverse of feedback timescale. It complements the previous analysis of gate set tomography data, which is consistent with the expectation that longer circuits are more sensitive to non-Markovian noise [30]. We attribute the residual discrepancy between the model and the data, especially where the G_x germ is used, to larger $|\varepsilon|$ fluctuations not captured by the chosen initial prior distribution $\mathcal{P}_0(\varepsilon)$. Further improvements would require a more quantitative analysis of the trade-off between the initially chosen uncertainty, the estimation bandwidth, and the white noise floor of the estimation method (see the Supplemental Material [33]).

Regarding gate performance, the GST protocol yields an $X_{\pi/2}$ -pulse infidelity of $(2.6 \pm 0.3) \times 10^{-2}$ and for the $Y_{\pi/2}$ -pulse $(1.6 \pm 0.3) \times 10^{-2}$, both with and without feedback. These values are two orders of magnitude larger than those reported by RB, which further suggests the dominant role of coherent noise, to which RB is less sensitive [50]. In summary, the decrease in Markovian model violation for longer GST sequences indicates that the FBS feedback protocol mitigates a significant portion of non-Markovian noise.

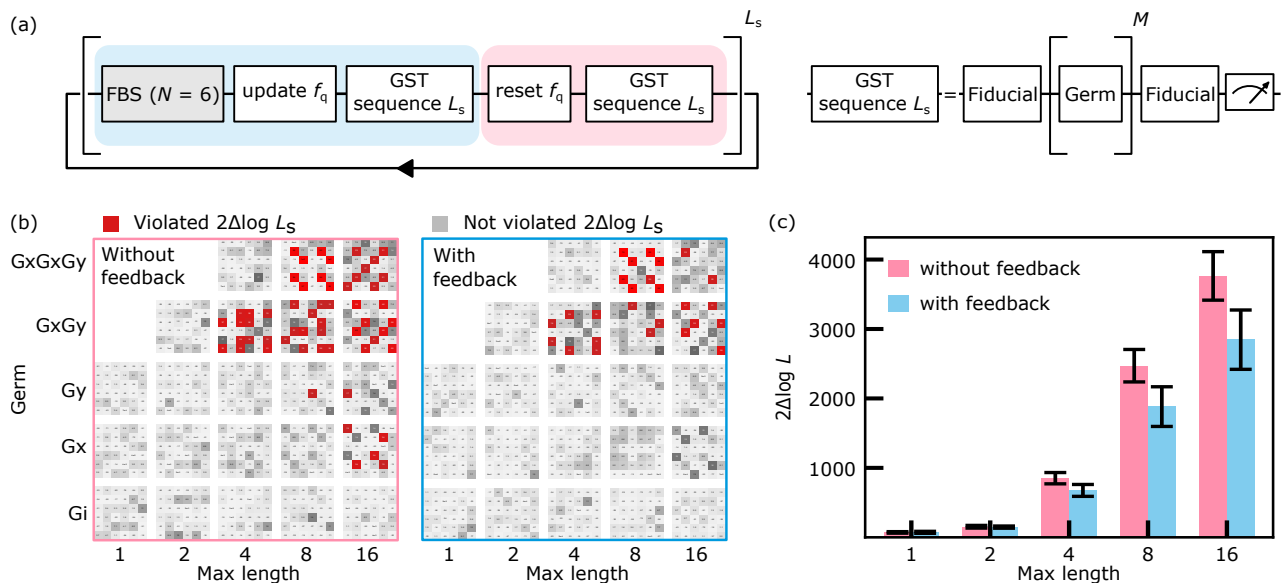


FIG. 4. **Gate set tomography and model violation.** (a) A gate set tomography (GST) sequence combined with the FBS feedback (with $N = 6$ probing cycles). (b) Model violation plot without (left) and with (right) FBS feedback. The red marks reveal detection of model violation at a confidence level of more than 95% and the gray boxes indicate statistical fluctuations. (c) The FBS feedback decreases the total amount of the log-likelihood ratios $2\Delta\log L$ at maximum lengths of 4, 8, and 16, extracted from 30 GST runs. Error bars show the 68% credible interval.

IV. OUTLOOK

Our work presents the experimental demonstration of a greedy adaptive Bayesian estimation protocol for the frequency of a resonantly driven qubit, improving its performance using only $N = 6$ single-shot measurements leveraging the exponential scaling of the algorithm with the number of measurements. This small number is to be compared with the tens [22] or hundreds of single-shot measurements of previous works [6, 13, 51] in resonantly driven qubits. Our approach reduces qubit-frequency drift by real-time Bayesian estimation and feedback, implemented via a low-latency FPGA-based qubit control system.

The binary search estimation algorithm allows control pulses to compensate for qubit frequency fluctuations caused by flux noise, improving coherence and fidelity without sacrificing frequency sensitivity or range. We validate the protocol using gate set tomography, which further corroborates our claim that our adaptive feedback loop reduces the effects of non-Markovian noise. This may facilitate quantum error correction methods, which generally assume Markovian noise.

Our protocol assumes that a Gaussian distribution can approximate the frequency fluctuations and that these remain constant over the time scale of a few measurements. In this work, the controller updates the drive frequency in real time, but the scheme can be adapted to adjust the qubit frequency by modifying the flux bias instead. The estimation bandwidth is limited by the relatively slow measurement time of a few microseconds. From this per-

spective, our work represents a worst-case scenario, underlying the efficacy of our experimental technique.

Traditional real-time approaches for continuous parameter estimations have relied on simple linear or trigonometric functions or precomputed lookup tables [6, 8, 9, 12, 13, 51]. In contrast, our FPGA implementation performs real-time Bayesian inference and processes widely used Gaussian distributions, resulting in a numerical complexity higher than previous works. At the same time, the parametrization of Gaussian distributions allows the FPGA to perform fewer and faster computations compared to traditional approaches using histograms and particle filtering.

We anticipate further improvements possibly by the implementation of a physics-informed prior [12] that accounts for $1/f$ noise. Also, the online calibration of single-qubit gates is a possible direction with available hardware, for single-qubit corrections in two-qubit gates by actual (not virtual) Z gates.

Adaptive Bayesian techniques could be implemented for frequency tracking in the presence of two-level fluctuators [52, 53]. Specific to our flux-tunable transmon qubit, while the feedback bandwidth is limited by readout and resonator cooldown time, higher bandwidth can be achieved by adding a Purcell filter [18, 19] which protects the qubit from relaxing into its environment. Our protocol may favor using symmetric junctions to increase the frequency range of the qubit, without worrying about increased sensitivity to flux noise as shown in Ref. [22].

Although qubit calibration by Bayesian inference can be relatively easily integrated with existing classical con-

trol hardware, it requires prior knowledge of the system’s parametrization. We expect that quantum model learning agents [6], which are more challenging to implement efficiently in real time, will gain wider adoption. We anticipate further research that merges theoretical and hardware advances, ultimately eliminating the need for a case-to-case modeling of experimental parameters.

Beyond superconducting qubits, our scheme offers new insights into real-time calibration of any qubits manipulated by resonant pulses. The protocol is not restricted to flux noise but is compatible with any source of low-frequency noise. We envision future applications in mitigating quasistatic electrical and nuclear noise in other solid-state qubit platforms.

This work advances quantum control by implementing an adaptive Bayesian technique to calibrate the qubits frequencies in real time. Our algorithm is a locally optimal solution by minimizing the expected estimator variance under a Gaussian distribution approximation, making it appealing for real-time calibration in large QPUs.

V. AUTHOR CONTRIBUTIONS

FB, JB, and JAK conceptualized the experiment. FB led the measurements and data analysis, and wrote the manuscript with input from all authors. FB, LP, MM, RA, AC, JAG, WDO, and FK performed the experiment with theoretical contributions from JB, JAK, and JD. YS designed the device, which was fabricated by DKK and BMN under supervision of KS, MES, and JLY. JAG, JD,

WDO, and FK supervised the project.

VI. ACKNOWLEDGMENTS

We gratefully acknowledge Patrick Harrington and Max Hays for fruitful discussions. This work received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreements 101017733 (QuantERA II), 951852 (QLSI), EU-REKA Eurostars 3 (ECHIDNA), European Research Council (ERC grant 856526), the Novo Nordisk Foundation under Challenge Programme NNF20OC0060019 (SolidQ), the Inge Lehmann Programme of the Independent Research Fund Denmark, the US Army Research Office (ARO) under Award No. W911NF-24-2-0043, the QuSpin Mobility Grant, the INTFELLES-Project No. 333990, which is funded by the Research Council of Norway (RCN), the Dutch National Growth Fund (NGF) as part of the Quantum Delta NL programme, and from the Danish Agency for Higher Education and Science (DAHES, grant 2076-00014B). It was also funded in part by the US Army Research Office (ARO) Multidisciplinary University Research Initiative (MURI) W911NF-18-1-0218, in part by the US Army Research Office (ARO) under Award No. W911NF-23-1-0045, and in part under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the US Air Force or the US Government.

-
- [1] T. Ichikawa, H. Hakoshima, K. Inui, K. Ito, R. Matsuda, K. Mitarai, K. Miyamoto, W. Mizukami, K. Mizuta, T. Mori, Y. Nakano, A. Nakayama, K. N. Okada, T. Sugimoto, S. Takahira, N. Takemori, S. Tsukano, H. Ueda, R. Watanabe, Y. Yoshida, and K. Fujii, Current numbers of qubits and their uses, *Nature Reviews Physics* **6**, 345 (2024).
 - [2] E. Campbell, A series of fast-paced advances in quantum error correction, *Nature Reviews Physics* **6**, 160 (2024).
 - [3] R. Acharya, L. Aghababaie-Beni, I. Aleiner, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, N. Astrakhantsev, J. Atalaya, *et al.*, Quantum error correction below the surface code threshold [10.48550/arXiv.2408.13687](https://arxiv.org/abs/10.48550/arXiv.2408.13687) (2024).
 - [4] H. Ball, W. D. Oliver, and M. J. Biercuk, The role of master clock stability in quantum information processing, *npj Quantum Information* **2**, 1 (2016).
 - [5] M. Mohseni, A. Scherer, K. G. Johnson, O. Wertheim, M. Otten, N. A. Aadit, K. M. Bresniker, K. Y. Cam-sari, B. Chapman, S. Chatterjee, G. A. Dagnew, A. Esposito, F. Fahim, M. Fiorentino, A. Khalid, X. Kong, B. Kulchytsky, R. Li, P. A. Lott, I. L. Markov, R. F. McDermott, G. Pedretti, A. Gajjar, A. Silva, J. Sorebo, P. Spentzouris, Z. Steiner, B. Torosov, D. Venturelli, R. J. Visser, Z. Webb, X. Zhan, Y. Cohen, P. Ronagh, A. Ho, R. G. Beausoleil, and J. M. Martinis, How to build a quantum supercomputer: Scaling challenges and opportunities [10.48550/arXiv.2411.10406](https://arxiv.org/abs/10.48550/arXiv.2411.10406) (2024).
 - [6] V. Gebhart, R. Santagati, A. A. Gentile, E. M. Gauger, D. Craig, N. Ares, L. Banchi, F. Marquardt, L. Pezzè, and C. Bonato, Learning quantum systems, *Nature Reviews Physics* **5**, 141 (2023).
 - [7] K. Reuer, J. Landgraf, T. Fösel, J. O’Sullivan, L. Beltrán, A. Akin, G. J. Norris, A. Remm, M. Kerschbaum, J.-C. Besse, *et al.*, Realizing a deep reinforcement learning agent for real-time quantum feedback, *Nature Communications* **14**, 7138 (2023).
 - [8] M. J. Arshad, C. Bekker, B. Haylock, K. Skrzypczak, D. White, B. Griffiths, J. Gore, G. W. Morley, P. Salter, J. Smith, I. Zohar, A. Finkler, Y. Altmann, E. M. Gauger, and C. Bonato, Real-time adaptive estimation of decoherence timescales for a single qubit, *Physical Review Applied* **21**, 024026 (2024).
 - [9] F. Berritta, T. Rasmussen, J. A. Krzywda, J. van der Heijden, F. Fedele, S. Fallahi, G. C. Gardner, M. J. Manfra, E. van Nieuwenburg, J. Danon, A. Chatterjee, and F. Kuemmeth, Real-time two-axis control of a spin qubit, *Nature Communications* **15**, 1676 (2024).
 - [10] N. Dumoulin Stuyck, A. E. Seedhouse, S. Serrano, T. Tanttu, W. Gilbert, J. Y. Huang, F. Hudson, K. M.

- Itoh, A. Laucht, W. H. Lim, C. H. Yang, A. Saraiva, and A. S. Dzurak, Silicon spin qubit noise characterization using real-time feedback protocols and wavelet analysis, *Applied Physics Letters* **124**, 114003 (2024).
- [11] N. R. Vora, Y. Xu, A. Hashim, N. Fruitwala, H. N. Nguyen, H. Liao, J. Balewski, A. Rajagopala, K. Nowrouzi, Q. Ji, *et al.*, ML-powered FPGA-based real-time quantum state discrimination enabling mid-circuit measurements [10.48550/arXiv.2406.18807](https://arxiv.org/abs/10.48550/arXiv.2406.18807) (2024).
- [12] F. Berritta, J. A. Krzywda, J. Benestad, J. van der Heijden, F. Fedele, S. Fallahi, G. C. Gardner, M. J. Manfra, E. van Nieuwenburg, J. Danon, A. Chatterjee, and F. Kuemmeth, Physics-informed tracking of qubit fluctuations, *Physical Review Applied* **22**, 014033 (2024).
- [13] J. Park, H. Jang, H. Sohn, J. Yun, Y. Song, B. Kang, L. E. Stehouwer, D. D. Esposti, G. Scappucci, and D. Kim, Passive and active suppression of transduced noise in silicon spin qubits, *Nature Communications* **16**, 78 (2025).
- [14] S. Kimmel, G. H. Low, and T. J. Yoder, Robust calibration of a universal single-qubit gate set via robust phase estimation, *Physical Review A* **92**, 062315 (2015).
- [15] R. D. McMichael and S. M. Blakley, Simplified algorithms for adaptive experiment design in parameter estimation, *Physical review applied* **18**, 054001 (2022).
- [16] T. Hurant, K. Sun, Z. Jia, J. Kim, and K. R. Brown, Few-shot, robust calibration of single qubit gates using Bayesian robust phase estimation [10.48550/arXiv.2407.18339](https://arxiv.org/abs/10.48550/arXiv.2407.18339) (2024).
- [17] B. de Neeve, A. V. Lebedev, V. Negnevitsky, and J. P. Home, Time-adaptive phase estimation [10.48550/arXiv.2405.08930](https://arxiv.org/abs/10.48550/arXiv.2405.08930) (2024).
- [18] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A quantum engineer's guide to superconducting qubits, *Applied Physics Reviews* **6**, 021318 (2019).
- [19] A. Blais, A. L. Grimsmo, S. M. Girvin, and A. Wallraff, Circuit quantum electrodynamics, *Reviews of Modern Physics* **93**, 025005 (2021).
- [20] J. B. Hertzberg, E. J. Zhang, S. Rosenblatt, E. Magesan, J. A. Smolin, J.-B. Yau, V. P. Adiga, M. Sandberg, M. Brink, J. M. Chow, *et al.*, Laser-annealing Josephson junctions for yielding scaled-up superconducting quantum processors, *npj Quantum Information* **7**, 129 (2021).
- [21] C. Berke, E. Varvelis, S. Trebst, A. Altland, and D. P. DiVincenzo, Transmon platform for quantum computing challenged by chaotic fluctuations, *Nature communications* **13**, 2495 (2022).
- [22] A. Vepsäläinen, R. Winik, A. H. Karamlou, J. Braumüller, A. D. Paolo, Y. Sung, B. Kannan, M. Kjaergaard, D. K. Kim, A. J. Melville, B. M. Niedzielski, J. L. Yoder, S. Gustavsson, and W. D. Oliver, Improving qubit coherence using closed-loop feedback, *Nature Communications* **13**, 1932 (2022).
- [23] C. Ferrie, C. E. Granade, and D. G. Cory, How to best sample a periodic probability distribution, or on the accuracy of Hamiltonian finding strategies, *Quantum Information Processing* **12**, 611 (2013).
- [24] The commonly used term *non-Markovian noise* refers to the non-unitary evolution of a quantum system that cannot be modeled by the Markovian master equation. Such non-Markovian decoherence arises from the environment's finite memory, which introduces correlations between consecutive measurement outcomes.
- [25] H. Hakoshima, Y. Matsuzaki, and S. Endo, Relationship between costs for quantum error mitigation and non-Markovian measures, *Physical Review A* **103**, 012611 (2021).
- [26] J. F. Kam, S. Gicev, K. Modi, A. Southwell, and M. Usman, Detrimental non-Markovian errors for surface code memory [10.48550/arXiv.2410.23779](https://arxiv.org/abs/10.48550/arXiv.2410.23779) (2024).
- [27] L. Viola, E. Knill, and S. Lloyd, Dynamical decoupling of open quantum systems, *Physical Review Letters* **82**, 2417 (1999).
- [28] P. Szańkowski, G. Ramon, J. Krzywda, D. Kwiatkowski, and L. Cywiński, Environmental noise spectroscopy with qubits subjected to dynamical decoupling, *Journal of Physics: Condensed Matter* **29**, 333001 (2017).
- [29] D. Pataki, Á. Márton, J. K. Asbóth, and A. Pályi, Coherent errors in stabilizer codes caused by quasistatic phase damping, *Physical Review A* **110**, 012417 (2024).
- [30] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, *Quantum* **5**, 557 (2021).
- [31] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized benchmarking of quantum gates, *Physical Review A* **77**, 012307 (2008).
- [32] Y. Sung, L. Ding, J. Braumüller, A. Vepsäläinen, B. Kannan, M. Kjaergaard, A. Greene, G. O. Samach, C. McNally, D. Kim, A. Melville, B. M. Niedzielski, M. E. Schwartz, J. L. Yoder, T. P. Orlando, S. Gustavsson, and W. D. Oliver, Realization of high-fidelity CZ and ZZ-free iSWAP gates with a tunable coupler, *Physical Review X* **11**, 021058 (2021).
- [33] See Supplemental Material, which includes Refs. [54–57].
- [34] C. Bonato, M. S. Blok, H. T. Dinani, D. W. Berry, M. L. Markham, D. J. Twitchen, and R. Hanson, Optimized quantum sensing with a single electron spin using real-time adaptive measurements, *Nature Nanotechnology* **11**, 247 (2015).
- [35] T. Joas, S. Schmitt, R. Santagati, A. A. Gentile, C. Bonato, A. Laing, L. P. McGuinness, and F. Jelezko, Online adaptive quantum characterization of a nuclear spin, *npj Quantum Information* **7**, 56 (2021).
- [36] A. M. Childs, J. Preskill, and J. Renes, Quantum information and precision measurement, *Journal of modern optics* **47**, 155 (2000).
- [37] R. S. Said, D. W. Berry, and J. Twamley, Nanoscale magnetometry using a single-spin system in diamond, *Physical Review B* **83**, 125410 (2011).
- [38] P. Cappellaro, Spin-bath narrowing with adaptive parameter estimation, *Physical Review A* **85**, 030301 (2012).
- [39] N. M. Nusran, M. U. Momeen, and M. V. G. Dutt, High-dynamic-range magnetometry with a single electronic spin in diamond, *Nature Nanotechnology* **7**, 109 (2012).
- [40] G. Waldherr, J. Beck, P. Neumann, R. S. Said, M. Nitsche, M. L. Markham, D. J. Twitchen, J. Twamley, F. Jelezko, and J. Wrachtrup, High-dynamic-range magnetometry with a single nuclear spin in diamond, *Nature Nanotechnology* **7**, 105 (2012).
- [41] We clarify that this is not a true binary search in the sense of gaining exactly one bit of information per measurement. The search does, however, follow a binary search tree where the two options at each node provide the

most information within the approximations we use. A true *quantum* binary search could be implemented for a decoherence-free qubit with ideal initialization and read-out [36]. The focus here is on estimating noise in a physical qubit.

- [42] In the limit $\mu/\sigma \rightarrow 0$, for $m = +1$, the likelihood function with zero phase (a squared cosine) has a single maximum at $\mu = 0$ within the 95% credible interval of the prior distribution, and the corresponding posterior distribution can be approximated by a Gaussian. In contrast, for $m = -1$ in the same limit, the zero-phase likelihood function (a squared sine) exhibits two global maxima within the 95% credible interval and a minimum at the maximum prior value, resulting in a bimodal posterior distribution that cannot be adequately approximated by a single Gaussian.
- [43] J. Benestad, J. A. Krzywda, E. van Nieuwenburg, and J. Danon, Efficient adaptive Bayesian estimation of a slowly fluctuating Overhauser field gradient, *SciPost Physics* **17**, 014 (2024).
- [44] Note that we do not claim exponential scaling with sensing time, which is fundamentally bounded by the Heisenberg limit $\sigma^2(\varepsilon) \sim 1/\tau^2$. Instead, our protocol achieves exponential scaling with the number of single-shot measurements N .
- [45] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, Efficient Z gates for quantum computing, *Physical Review A* **96**, 022330 (2017).
- [46] In the RB experiment, μ_0 is reset to zero before each estimation sequence, which explains the increased σ_0 and N compared to the Ramsey experiment of the previous section, where we do not reset μ_0 before each estimation sequence. We believe that further improvements in the single-qubit gate fidelity could be achieved by using the same settings as in the Ramsey experiment. Still, this work focuses on demonstrating the enhancement provided by FBS, rather than achieving the highest possible fidelity for this particular setup.
- [47] P. J. J. O'Malley, J. Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, I.-C. Hoi, E. Jeffrey, A. Megrant, J. Mutus, C. Neill, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Korotkov, A. N. Cleland, and J. M. Martinis, Qubit metrology of ultralow phase noise using randomized benchmarking, *Physical Review Applied* **3**, 044009 (2015).
- [48] E. Nielsen, K. Rudinger, T. Proctor, A. Russo, K. Young, and R. Blume-Kohout, Probing quantum processor performance with pyGSTi, *Quantum Science and Technology* **5**, 044002 (2020).
- [49] A. Hashim, L. B. Nguyen, N. Goss, B. Marinelli, R. K. Naik, T. Chistolini, J. Hines, J. Marceaux, Y. Kim, P. Gokhale, *et al.*, A practical introduction to benchmarking and characterization of quantum computers [10.48550/arXiv.2408.12064](https://arxiv.org/abs/10.48550/arXiv.2408.12064) (2024).
- [50] Y. R. Sanders, J. J. Wallman, and B. C. Sanders, Bounding quantum gate error rate based on reported average fidelity, *New Journal of Physics* **18**, 012002 (2015).
- [51] T. Nakajima, A. Noiri, K. Kawasaki, J. Yoneda, P. Stano, S. Amaha, T. Otsuka, K. Takeda, M. R. Delbecq, G. Allison, *et al.*, Coherence of a driven electron spin qubit actively decoupled from quasistatic noise, *Physical Review X* **10**, 011060 (2020).
- [52] B.-J. Liu, Y.-Y. Wang, T. Sheffer, and C. Wang, Observation of discrete charge states of a coherent two-level system in a superconducting qubit [10.48550/arXiv.2401.12183](https://arxiv.org/abs/10.48550/arXiv.2401.12183) (2024).
- [53] F. Ye, A. Ellaboudy, and J. M. Nichol, Stabilizing an individual charge fluctuator in a Si/SiGe quantum dot [10.48550/arXiv.2407.05439](https://arxiv.org/abs/10.48550/arXiv.2407.05439) (2024).
- [54] B. Paquelet Wuetz, D. Degli Esposti, A.-M. J. Zwerver, S. V. Amitonov, M. Botifoll, J. Arbiol, A. Sammak, L. M. Vandersypen, M. Russ, and G. Scappucci, Reducing charge noise in quantum dots by using thin silicon quantum wells, *Nature communications* **14**, 1385 (2023).
- [55] J. S. Rojas-Arias, Y. Kojima, K. Takeda, P. Stano, T. Nakajima, J. Yoneda, A. Noiri, T. Kobayashi, D. Loss, and S. Tarucha, The origins of noise in the Zeeman splitting of spin qubits in natural-silicon devices [10.48550/arXiv.2408.13707](https://arxiv.org/abs/10.48550/arXiv.2408.13707).
- [56] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, Efficient measurement of quantum gate error by interleaved randomized benchmarking, *Physical Review Letters* **109**, 080505 (2012).
- [57] P. J. Huber, *Robust statistics*, Wiley series in probability and mathematical statistics (J. Wiley & sons, New York, 1981).

Supplemental Material for “Efficient Qubit Calibration by Binary-Search Hamiltonian Tracking”

Fabrizio Berritta

*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA and
Center for Quantum Devices, Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark*

Jacob Benestad and Jeroen Danon

*Center for Quantum Spintronics, Department of Physics, Norwegian
University of Science and Technology, NO-7491 Trondheim, Norway*

Lukas Pahl and Youngkyu Sung

*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA and
Department of Electrical Engineering and Computer Science,
Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Melvin Mathews

*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA and
Department of Information Technology and Electrical Engineering, ETH Zürich, 8093 Zürich, Switzerland*

Jan A. Krzywda

*$\langle aQa^L \rangle$ Applied Quantum Algorithms — Lorentz Institute for Theoretical Physics &
Leiden Institute of Advanced Computer Science, Universiteit Leiden, The Netherlands*

Réouven Assouly and Jeffrey A. Grover

Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

David K. Kim, Bethany M. Niedzielski, Mollie E. Schwartz, and Jonilyn L. Yoder
Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA 02421, USA

Kyle Serniak

*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA and
Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA 02421, USA*

Anasua Chatterjee

*Center for Quantum Devices, Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark and
QuTech and Kavli Institute of Nanoscience, Delft University of Technology, Delft, The Netherlands*

William D. Oliver

*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
Department of Electrical Engineering and Computer Science,
Massachusetts Institute of Technology, Cambridge, MA 02139, USA and
Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Ferdinand Kuemmeth*

*Center for Quantum Devices, Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark
Institute of Experimental and Applied Physics, University of Regensburg, 93040 Regensburg, Germany and
QDevil, Quantum Machines, 2750 Ballerup, Denmark*

(Dated: August 26, 2025)

CONTENTS

Experimental setup	2
The frequency binary search algorithm	4

Qubit stabilization over six hours	5
Power spectral density with and without feedback	5
Effect of the feedback on the non-Markovian dephasing	7
Interleaved randomized benchmarking	9
On the impact of state and preparation measurement errors and dephasing	10
Effect of not including SPAM errors in the likelihood model	11
Validity of the Gaussian Approximation	12
References	13

EXPERIMENTAL SETUP

The measurements are performed in a Bluefors XLD-600 dilution refrigerator with a base temperature of 20 mK and the setup is sketched in Fig. S1. The Quantum Machines OPX+ and Octave are used for the XY control of the qubit and readout signal, and both microwave pulses are generated by single-sideband modulation with suppressed carrier. For the qubit coherence experiment in the main text, each drive pulse is 28-ns long, with a cosine rise and fall envelope of 5 ns each. For the randomized benchmarking and gate set tomography experiments, the pulses are 20-ns long and calibrated by DRAG.

The OPX+ includes real-time classical processing with fast analog feedback programmed in QUA software. The OPX+ and Octave are frequency-locked by a Quantum Machines OPT (not shown). The readout resonator linewidth $\kappa_r/(2\pi) \approx 0.35$ MHz, and the microwave readout tone, approximately 7.08 GHz, is filtered and attenuated at room temperature by passive components. The readout tone is attenuated in the cryostat to remove excess thermal photons from higher-temperature stages, and low-pass filtered at the mixing chamber. The device is shielded magnetically with a superconducting can. The transmitted signal from the feedline goes through a circulator placed after the sample to remove noise of the Josephson traveling wave parametric amplifier (JTWPA) and dump it in a $50\ \Omega$ terminator. We preamplify the readout signal at base temperature using a Holzworth HS9000 Synthesizer.

To prevent noise from higher-temperature stages from reaching the JTWPA and the sample, two microwave isolators are placed after the JTWPA. A high-electron mobility transistor amplifier thermally anchored at the 3 K stage further amplifies the readout signal. At room temperature, the readout line is again amplified. A Yokogawa GS200 provides the DC bias flux through a bias-tee terminated by $50\ \Omega$ from the AC side, as we perform virtual Z gates in this work.

Device parameters		
Transition frequency	$\omega_q/(2\pi)$ (GHz)	3.78
Anharmonicity	$\eta/(2\pi)$ (MHz)	-248
Relaxation time	T_1 (μ s)	80
Ramsey decay time	T_2^{*a} (μ s)	3.3
Spin-echo decay time	T_{2E} (μ s)	12
Ramsey decay time ($\Phi_{\text{ext}} = 0$)	T_2^{*b} (μ s)	42
Readout resonator frequency	$\omega_r/(2\pi)$ (GHz)	7.08
Readout resonator linewidth	$\kappa_r/(2\pi)$ (MHz)	0.35

TABLE S1. Device parameters at $\Phi_{\text{ext}} = \Phi_0/4$ if not specified otherwise.

^a Measured over six hours.

^b Measured over one minute.

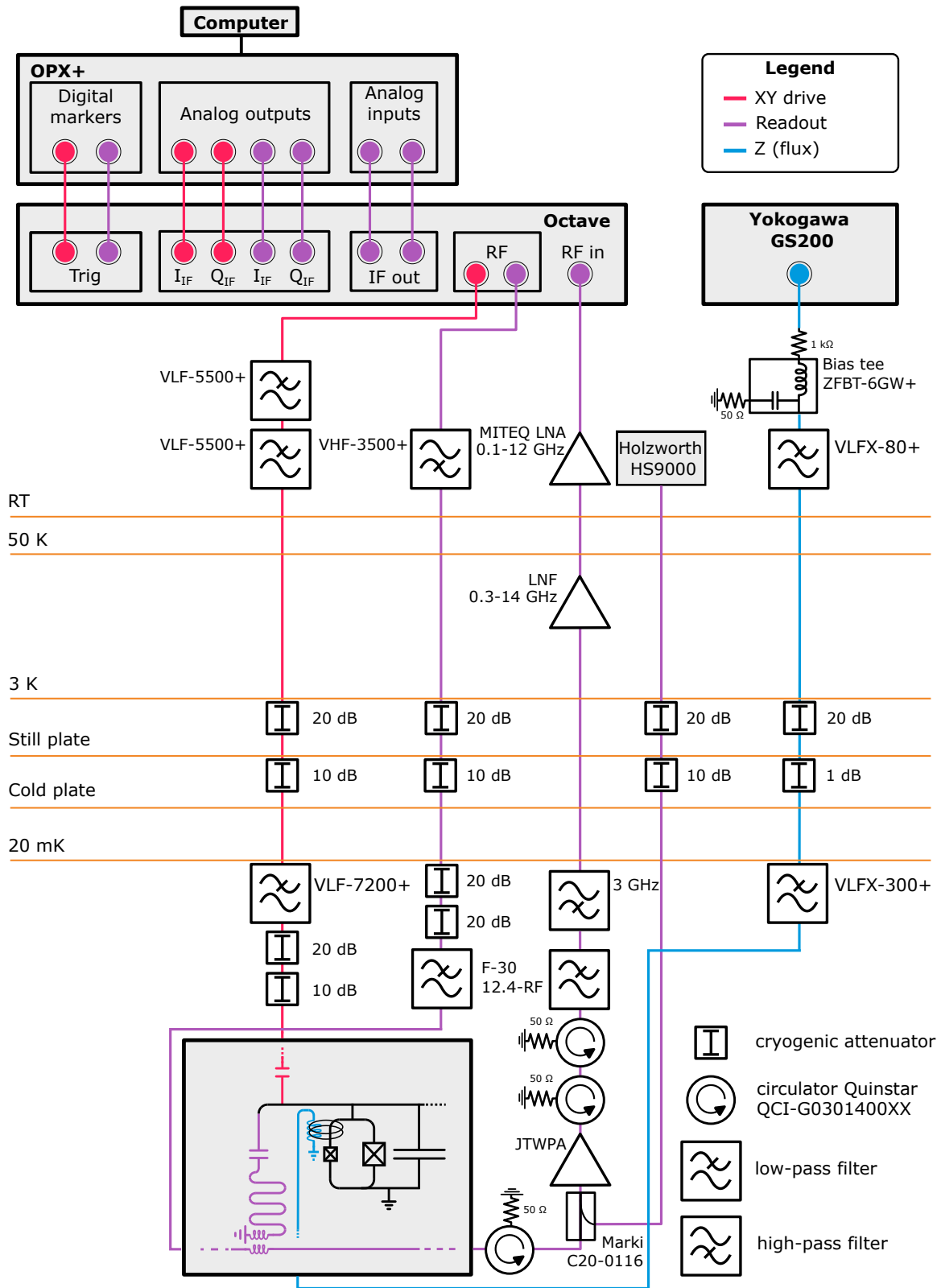


FIG. S1. **Experimental setup BF2.** The cryostat is a Bluefors XLD-600 dilution refrigerator with a base temperature lower than 30 mK. A Quantum Machines OPX+ and Octave are used for the XY drive pulses and readout. The setup includes a Yokogawa GS200 for DC flux biasing.

THE FREQUENCY BINARY SEARCH ALGORITHM

Here we summarize the FBS algorithm described in the main text, together with a concise pseudocode format as seen in Algorithm 1. The inputs to the algorithm are μ_0 and σ_0 to parameterize the initial Gaussian prior (i.e., roughly the search area for the frequency ε to be estimated), in addition to the number of update steps N and SPAM error and dephasing coefficients α , β and T . Each update step starts by calculating the qubit precession time τ to balance that the oscillation period of the likelihood is closest possible to the width (specifically $2\pi\sigma$) of the prior distribution while also not reaching so long times that more information is lost through dephasing than necessary (formally, one calculates the τ that minimizes the expected posterior variance). Based on τ , the frequency detuning Δf is then calculated so that the inflection point of the likelihood function is located in the middle of the prior. This combination of τ and Δf divides the prior in a left and right branch based on the outcome of a Ramsey measurement $m \in \{-1, 1\}$. The parameters μ and σ for the Gaussian used to approximate the true posterior distribution are then calculated based on a method of moments fit where these parameters are taken to be the mean and standard deviation of the true posterior. This update sequence is repeated, now with μ and σ parameterizing the new prior, until it has been performed a total of N times.

Algorithm 1 Frequency binary search algorithm

```

1: procedure FREQUENCYBINARYSEARCH( $\mu_0, \sigma_0, N, \alpha, \beta, T$ )
2:    $\mu \leftarrow \mu_0$ 
3:    $\sigma \leftarrow \sigma_0$ 
4:    $n \leftarrow 0$ 
5:   for  $n < N$  do
6:      $\tau \leftarrow \frac{\sqrt{16\pi^2\sigma^2+1/T^2}-1/T}{8\pi^2\sigma^2}$ 
7:      $\Delta f \leftarrow \frac{1}{4\tau} - \mu$ 
8:      $m \leftarrow \text{RAMSEY}(\tau, \Delta f)$   $\triangleright m \in \{-1, 1\}$  is the experiment outcome
9:      $\mu \leftarrow \mu - \frac{2\pi m \beta \sigma^2 \tau e^{-\tau/T-2\pi^2\sigma^2\tau^2}}{1+m\alpha}$ 
10:     $\sigma \leftarrow \sqrt{\sigma^2 - \frac{4\pi^2\beta^2\sigma^4\tau^2 e^{-2\tau/T-4\pi^2\sigma^2\tau^2}}{(1+m\alpha)^2}}$ 
11:     $n \leftarrow n + 1$ 
12:  end for
13:  return  $\mu, \sigma$ 
14: end procedure

```

QUBIT STABILIZATION OVER SIX HOURS

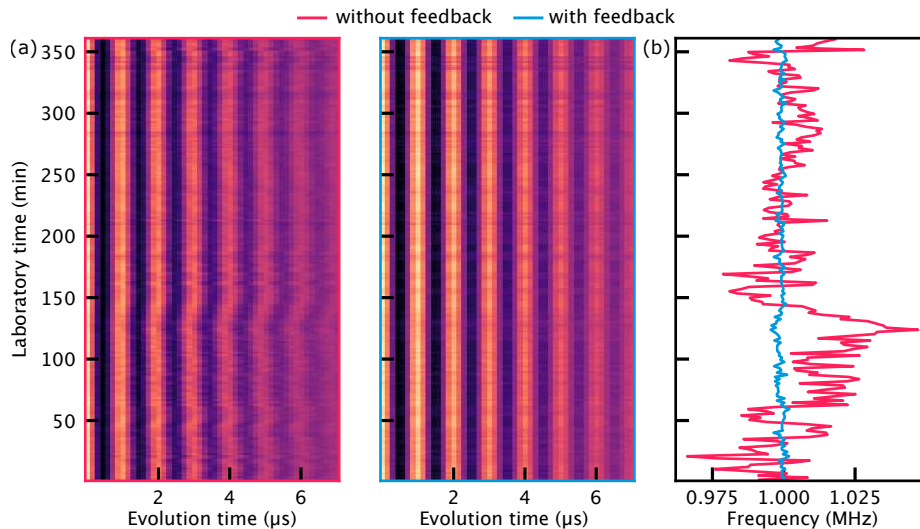


FIG. S2. **Qubit frequency stabilization over 6 hours.** (a) The Ramsey experiment is repeated for 6 hours without (left) and with (right) feedback by the frequency binary search. With the feedback, the qubit frequency is locked during the whole duration of the experiment. (b) Time dependence of the measured frequency detuning extracted from Ramsey measurements in (a). The red trace is taken with a fixed drive frequency f_d , and the blue trace is taken with a feedback-controlled f_d .

We perform the protocol explained in Fig. 2 of the main text for hours as shown for the interleaved measurements in Fig. S2(a), except this time the controller uses $N = 6$ single-shot measurements for the FBS estimation instead of $N = 8$. In the left (right) panel the feedback is turned off (on) for over 6 hours. Figure S2(b) shows the values of Δf obtained from Ramsey measurements with feedback off (red curve) and on (blue curve). The feedback significantly reduces the fluctuation of Δf . By averaging across laboratory time panels (a) and (b), T_2^* improves from $(3.32 \pm 0.10) \mu\text{s}$ without feedback to $(4.81 \pm 0.13) \mu\text{s}$ with feedback, which results in an average 44% improvement, similar to Fig. 2 of the main text.

Power spectral density with and without feedback

We use the 6-hour-long Ramsey measurements to extract the power spectral density (PSD) of the qubit frequency fluctuations with and without the feedback. We combine four different methods to extract the PSD across seven orders of magnitude (the result is shown in Fig. S3):

- At low frequencies (red points), we use the frequency time-series from Fig. S2 and compute the PSD using Welch's method. In this approach, each data point is separated by an approximately constant delay of 104.37 s, which consists of a measurement period (30.19 s) and data transfer (74.18 s).
- At intermediate frequencies (orange points), we repeat this procedure inside each of 207 runs, using an averaging window of 1,000 repetitions, which effectively corresponds to 3.02 s window duration. Empirically, we found that smaller averaging windows lead to artificial noise due to errors in fitting the frequency to the averaged oscillating signal.
- At high frequencies (green points), we adapt the method of Ref. [1], and relate the spectral density of the frequency fluctuations to the correlation function of the initial state population:

$$P_0(t; \tau) = \frac{1}{2} + \frac{1}{2} \left(\alpha + \beta e^{-\tau/T} \cos(2\pi[\epsilon_0 + \delta\epsilon(t)]\tau) \right) \approx A_0(\tau) + B_1(\tau)\delta\epsilon(t)\tau + B_2(\tau)\delta\epsilon(t)^2\tau^2,$$

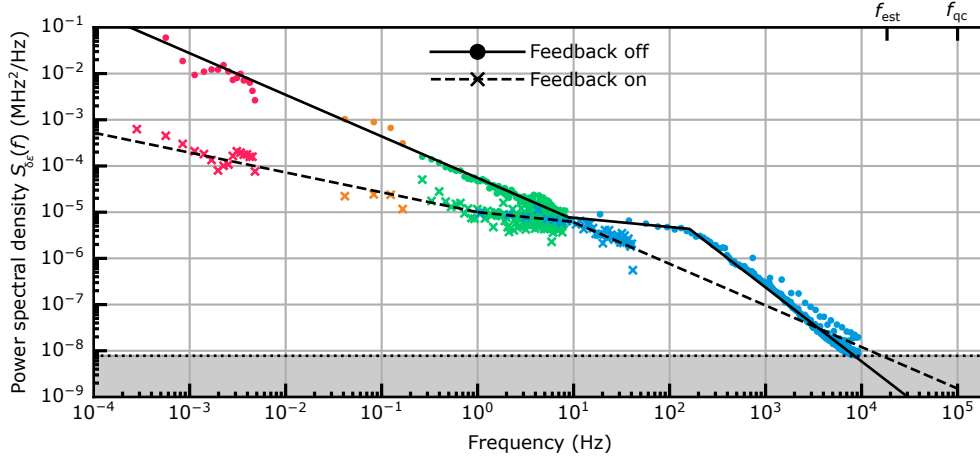


FIG. S3. **Power Spectral Density of frequency fluctuations with and without feedback.** We use four different methods to process 6 hours of Ramsey data and estimate the PSD across eight (five) orders of magnitude for data without (with) feedback. We perform a piecewise fit to a power-law spectrum, i.e., $S_f(\delta\epsilon) = A(f_1/f)^\alpha$, yielding three different regions for the data without feedback (drawn with solid black lines). For the data with feedback we use two regions (dashed black lines). The horizontal dotted line, denotes the white-noise floor corresponding to the experimentally measured $T_2 = 12 \mu\text{s}$. We plot a one-sided PSD. The extracted parameters are given in Tab. S2.

where

$$\begin{aligned} A_0(\tau) &= \frac{1}{2} + \frac{\alpha}{2} + \frac{\beta e^{-\tau/T}}{2} \cos(2\pi\epsilon_0\tau), \\ B_1(\tau) &= -\pi\beta e^{-\tau/T} \sin(2\pi\epsilon_0\tau), \\ B_2(\tau) &= -\pi^2\beta e^{-\tau/T} \cos(2\pi\epsilon_0\tau), \end{aligned}$$

with t denoting laboratory time, in contrast to τ , which stands for a probing time. In this model, $\delta\epsilon(t)$ can be seen as the coarse-grained fluctuations in the qubit frequency, which, for considered probing times of $\tau < 7 \mu\text{s}$ and expected noise amplitudes of $\sigma_{\delta\epsilon} \sim 0.01 \text{ MHz}$, results in relatively small random phases $\delta\phi = 2\pi\delta\epsilon(t)\tau \ll 1$, which validates the expansion. The correlation function of P_0 is now related to the correlation function of the frequency fluctuations $R_{\delta\epsilon}(t) = \langle \delta\epsilon(t)\delta\epsilon(0) \rangle$ as

$$C_P(t; \tau) = \langle P_0(t)P_0(0) \rangle - \langle P_0(0) \rangle^2 \approx B_1^2 R_{\delta\epsilon}(t)\tau^2 + B_2^2 R_{\delta\epsilon}(t)^2\tau^4.$$

The above relation holds for any τ ; however, it is convenient to use a τ that (i) makes the $B_2 \propto \cos(2\pi\epsilon_0\tau)$ term small, (ii) is large enough to provide a good signal-to-noise ratio, but (iii) is small enough not to be affected by relaxation. By optimizing the signal-to-noise ratio, we pick $\tau_0 = \{3.721, 3.864, 4.149, 4.292\} \mu\text{s}$ probing times, which confirms the observation that the optimal $\tau_0 \approx T_2^*$ [1]. At this point, we use the Wiener-Khinchin theorem to find the spectral density as the Fourier transform of the extracted correlation function, i.e.,

$$S(f) = \int dt e^{-i2\pi ft} R_\epsilon(t) \approx \int dt \frac{\hat{C}_P(t; \tau_0)}{B_1^2 \tau_0^2} e^{-i2\pi ft}, \quad (\text{S.1})$$

where $\hat{C}_P(t; \tau)$ is the estimated correlation function of the population, averaged over the 207 runs and waiting times τ_0 . In a single repetition, the correlation function can be empirically computed as

$$\hat{C}_P(k\Delta t; \tau_0) = \frac{1}{N} \sum_{n=1}^N x_{\tau_0}[n]x_{\tau_0}[n+k] - \frac{1}{N^2} \left(\sum_{n=1}^N x_{\tau_0}[n] \right)^2 \quad (\text{S.2})$$

where $x_{\tau_0}[n] = \pm 1$ is the n -th measurement of the population at time τ_0 and $\Delta t \approx 3.5 \text{ ms}$ is the time between two measurements.

- Finally, at the highest frequencies (blue points), we rely on the time series of frequency fluctuations extracted from Bayesian estimation. For the data without feedback, we directly use the frequencies estimated in the FBS

scheme, allowing us to extract the PSD up to the inverse of the estimation cycle $f_{\max} \approx 10^4$ Hz. For the noise with feedback, we perform offline Bayesian estimation using 4 repetitions, i.e., 200 probing times, which in total takes $\Delta t = 4 \times 3.5 \text{ ms} = 14 \text{ ms}$ and sets $f_{\max} \approx 60$ Hz. For the estimation, we use a Gaussian prior centered around the previous estimate with an initial $\sigma_0 = 2 \text{ kHz}$. Larger σ_0 resulted in an artificially flat spectrum due to large uncertainty in estimation, while smaller σ_0 introduced artificial low-pass filtering. For both the feedback and no-feedback data, we converted the frequency fluctuations time trace to a PSD using Welch's method.

Together, the extracted values of the PSDs are plotted in Fig. S3 as a function of frequency. We use dots for no-feedback data and crosses for data with feedback, and use different colors for the four methods described above. Both PSDs are piecewise fitted with a power law $S(f) = A(f_1/f)^\alpha$, where A and α are the fitting parameters and $f_1 = 1 \text{ Hz}$. The extracted values of A and α are shown in Tab. S2.

To cross-check different methods of estimating power spectral density, we compare in Fig. S4 time traces of frequency fluctuations obtained in the low- and mid-frequency methods against Bayesian estimation data with a moving average, obtaining relatively good agreement.

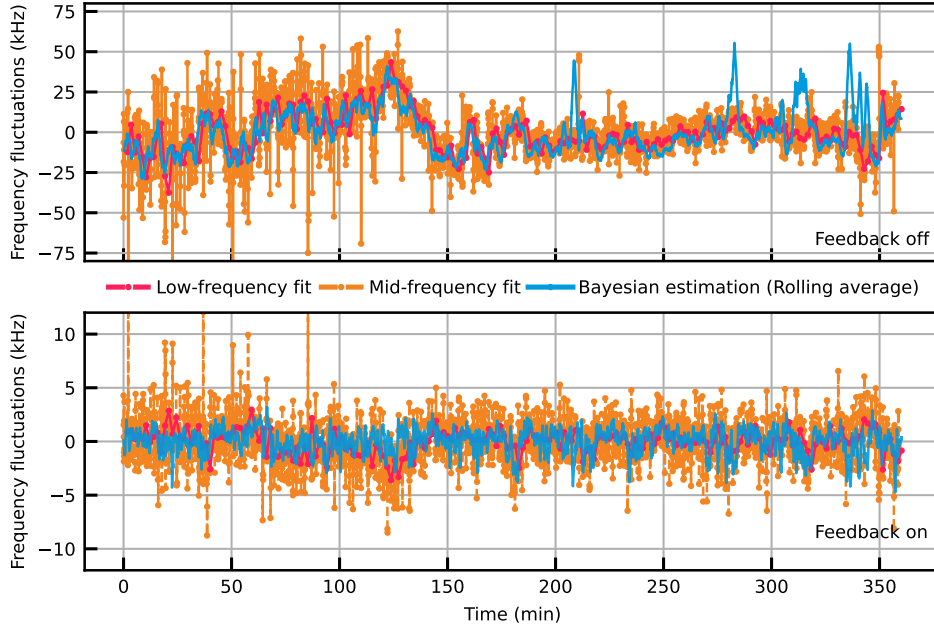


FIG. S4. **Frequency fluctuations extracted from 6-hours Ramsey.** We cross-validate methods by plotting the frequency fluctuations as a function of laboratory time. For the Bayesian estimation data we apply a rolling average with $N = 10,000$ for the data without feedback and $N = 1,000$ for the data with feedback. Furthermore we downsample the averaged Bayesian signal by a factor of $M = 100$.

Effect of the feedback on the non-Markovian dephasing

Based on the extracted power spectral density, we can estimate the quasistatic dephasing time, using the relation [2]

$$T_2^* = \frac{1}{\sqrt{2\pi}\sigma_f}, \quad \sigma_f^2 = \int_{f_0}^{f_{qc}} df S(f), \quad (\text{S.3})$$

where $S(f)$ is the one-sided spectrum, i.e., $S(f) = S_2(f) + S_2(-f)$ in terms of the two-sided spectrum $S_2(f)$. The lower bound f_0 is set by the duration of the experiment used to determine T_2^* . For the upper bound f_{qc} we assumed that the PSD is dominated by frequencies below the inverse of the qubit cycle time $\approx 10 \mu\text{s}$, yielding $f_{qc} = 10^5 \text{ Hz}$ as upper cut-off frequency.

For each part of the PSD that fits a power law $S(f) = A(f_1/f)^\alpha$ between frequencies f_{\min} and f_{\max} , we can thus

estimate its contribution to the noise amplitude as

$$\int_{f_{\min}}^{f_{\max}} df S(f) = \begin{cases} \frac{Af_1}{1-\alpha} (f_{\min}/f_1)^{1-\alpha} (\kappa^{1-\alpha} - 1) & \text{for } \alpha \neq 1, \\ Af_1 \ln(\kappa) & \text{for } \alpha = 1, \end{cases} \quad (\text{S.4})$$

where $\kappa = f_{\max}/f_{\min}$. The total σ_f^2 in Eq. (S.3) can then be expressed as a sum of the contributions from different regions of the PSD using the piecewise fits shown in Fig. S3. We present the results for each piecewise region of the PSD in Tab. S2 below.

TABLE S2. Integrated PSD (σ_f^2) and various T_2 times with feedback status and fit parameters. The σ_f^2 values are calculated using $\sigma^2 = \int_{f_{\min}}^{f_{\max}} A(f_1/f)^\alpha df$. The T_2^* sums up the total contribution $1/(\sqrt{2\pi}\sqrt{\sum \sigma_i^2})$ from the different regions. $T_E = 12 \mu\text{s}$ is the measured Hahn echo time, used for estimating \hat{T}_2 in both groups, according to relation $1/\hat{T}_2 = 1/T_2^* + 1/T_E$.

Feedback	f_{\min} (Hz)	f_{\max} (Hz)	A (MHz ² /Hz)	α	σ_f^2 (MHz ²)	T_2^* (μs)	T_E (μs)	\hat{T}_2 (μs)	T_2 (μs)
off	1.0×10^{-4}	8.9×10^0	0.55×10^{-4}	0.90	4.664×10^{-4}	4.55	12	3.30	3.4
	8.9×10^0	1.6×10^2	0.12×10^{-4}	0.20	8.364×10^{-4}				
	1.6×10^2	1.0×10^5	1.48×10^{-2}	1.60	1.147×10^{-3}				
on	1.0×10^{-4}	1.0×10^0	0.10×10^{-4}	0.43	1.745×10^{-5}	7.13	12	4.47	4.8
	1.0×10^0	1.0×10^1	0.10×10^{-4}	0.21	6.539×10^{-5}				
	1.0×10^1	1.0×10^5	0.48×10^{-4}	0.90	9.134×10^{-4}				

We start from the low-frequency region $f < 10$ Hz, in which feedback provided more than a factor of two reduction in noise root mean square, i.e., from $\sigma_{\text{off}} = 21.6$ kHz to $\sigma_{\text{on}} = 9.1$ kHz. However, for the data with the feedback, finite exponent $\alpha_{\text{on}} > 0$ suggests that filtering properties are worse for low-frequency noise. One explanation is that a narrow prior in the FBS protocol leads to underestimating large deviations from the average frequency, which are statistically associated with low-frequency noise.

The extracted PSD suggests that decoherence is more likely to be dominated by higher frequencies $f > 10$ Hz. In this regime the improvement is weaker, of the order of 40%, reflected in a reduction from $\sigma_{\text{off}} \approx 45$ kHz to $\sigma_{\text{on}} \approx 31$ kHz. We highlight however that in this regime the extracted PSD relies on sparse data, which includes extrapolation of an initial trend extracted from Bayesian data (feedback on). Also for feedback-off data, the kink in the spectral density might represent a change in the dominating physical mechanism responsible for the fluctuations, but could also reflect too narrow prior distribution resulting in an apparent low-pass filter.

In total, the estimated quasistatic dephasing times of $T_{2,\text{off}}^* = 4.55 \mu\text{s}$ and $T_{2,\text{on}} = 7.13 \mu\text{s}$ are slightly longer than experimentally measured $T_{2,\text{off}} = 3.4 \mu\text{s}$ and $T_{2,\text{on}} = 4.8 \mu\text{s}$. As a consequence, the predicted feedback-related improvement of 56% slightly exceeds the reported 44%. Most likely, this is due to dephasing caused by high-frequency noise, with a characteristic frequency above $f_{\text{qc}} = 10^5$ Hz. Such an explanation is consistent with the larger overestimation of the data with feedback, in case of which a relatively larger portion of dephasing can be attributed to unfiltered high-frequency noise.

To further verify this hypothesis, we use a heuristic formula for estimated total dephasing time \hat{T}_2^* , i.e.

$$\frac{1}{\hat{T}_2} = \frac{1}{T_2^*} + \frac{1}{T_E} \quad (\text{S.5})$$

where the spin-echo time T_E is approximately equal to the dephasing time due to high-frequency noise. Substituting the experimentally measured $T_E = 12 \mu\text{s}$, the estimated \hat{T}_2 moves closer to experimental values, i.e., $\hat{T}_{2,\text{off}} \approx 3.30 \mu\text{s}$ and $\hat{T}_{2,\text{on}} \approx 4.47 \mu\text{s}$, both within 10% error. Similarly, the predicted, feedback-related improvement of $\approx 35\%$ is consistent with the experiment. Its slightly smaller value is likely related to larger underestimation of dephasing time for the data with the feedback, caused by extrapolation of the PSD in the high frequency regime.

INTERLEAVED RANDOMIZED BENCHMARKING

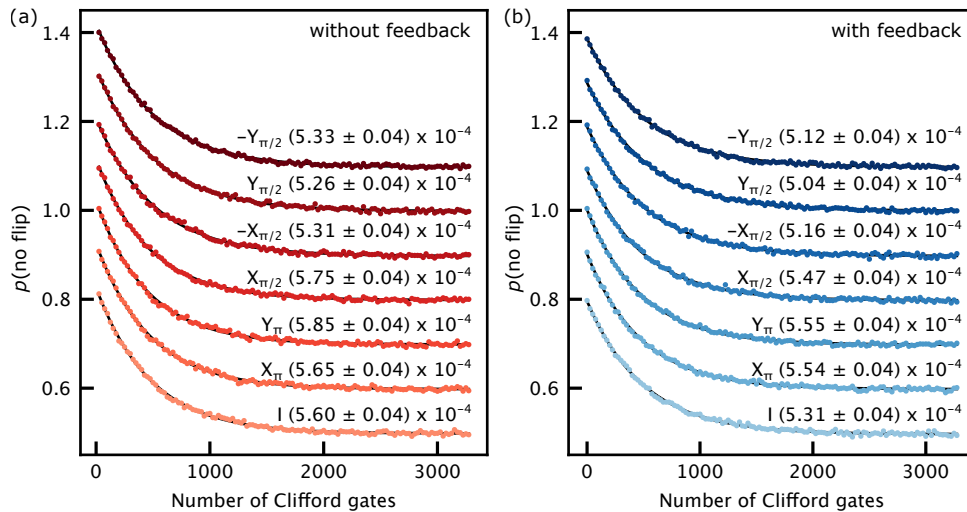


FIG. S5. **Interleaved randomized benchmarking with and without frequency binary search.** (a) Sequence infidelities for interleaved randomized benchmarking without frequency estimation by FBS. Traces are offset by 0.1 for clarity. Interleaved sequences are annotated with corresponding single-qubit gates and extracted natural gate infidelities. (b) Interleaved with (a) using frequency estimation by FBS. For all gates, the infidelity is slightly lower with feedback by a few 10^{-5} .

We perform interleaved randomized benchmarking (RB) [3] with and without feedback. The protocol consists of interleaving the gate \mathcal{C} of interest with random gates from the Clifford sequence. A final gate is performed to make the total sequence equal to the identity operation. In Fig. S5 we plot the fraction of not-flipped state $p(\text{no flip})$ as a function of the number of Clifford gates in a random sequence, terminated by a Clifford gate that would in principle bring the qubit back to the initial state. For each gate, 30 random sequences are generated and they are averaged 1,000 times. We find a $X_{\pi/2}$ -pulse gate infidelity improvement from $(5.75 \pm 0.04) \times 10^{-4}$ to $(5.47 \pm 0.04) \times 10^{-4}$, and for the $Y_{\pi/2}$ -pulse from $(5.26 \pm 0.04) \times 10^{-4}$ to $(5.04 \pm 0.04) \times 10^{-4}$. The interleaved RB infidelities are in the same order of magnitude as the standard RB shown in Fig. 3 of the main text. For each gate, feedback shows a similar improvement.

ON THE IMPACT OF STATE AND PREPARATION MEASUREMENT ERRORS AND DEPHASING

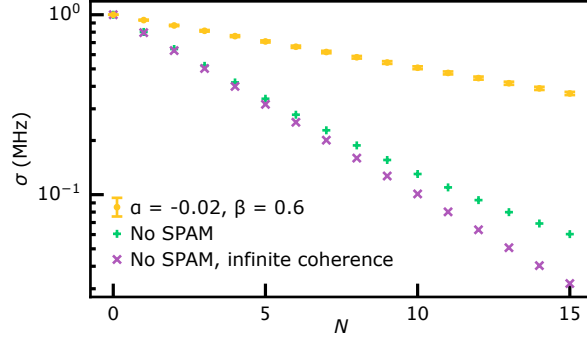


FIG. S6. **Impact of SPAM errors and dephasing.** As the algorithm becomes aware of errors it becomes more conservative. As a reference we plot the posterior standard deviation σ without SPAM errors (ie. $\alpha = 0$ and $\beta = 0$) both for $T \rightarrow +\infty$ (purple) and $T = 10 \mu\text{s}$ (green). For the case when $\alpha \neq 0$ we plot the average σ over several simulations (yellow) with error bars (extending two standard deviations of σ to each side), as it now becomes measurement dependent.

In Fig. S6 we plot the scaling of the FBS uncertainty σ as a function of the number of measurements N for the case (i) no state and preparation measurement errors (SPAM), no decoherence, (ii) no SPAM, with decoherence $T = 10 \mu\text{s}$, and (iii) including SPAM and decoherence. By including SPAM of the type $\beta \neq 0$ and dephasing T in the statistical model for how measurements are generated, the algorithm becomes more conservative in how much the posterior distribution variance is reduced as the visibility is decreased and the algorithm now has less trust in measurement outcomes. Having SPAM errors of the type $\alpha \neq 0$ reflects a bias towards one of the states being generated more than the other, and by making the algorithm aware of this bias it attempts to compensate by increasing the weight of the minority outcomes while reducing the weight of the majority outcomes.

The linlog plot in Fig. S6 of σ versus N demonstrates the scaling of the uncertainty in the algorithm. An exponential scaling of the posterior variance persists for as long as $\tau \ll T$, as the expected posterior variance in this case can be approximated as

$$\mathbb{E}[\sigma^2] = \sigma^2 - \frac{4\pi^2\beta^2\tau^2\sigma^4}{1-\alpha^2} e^{-2\tau(1/T+2\pi^2\sigma^2\tau)} \approx \sigma^2 - \frac{4\pi^2\beta^2\tau^2\sigma^4}{1-\alpha^2} e^{-4\pi^2\sigma^2\tau^2}, \quad (\text{S.6})$$

where the optimal time τ (presented in the main text) is now approximately

$$\tau \approx \frac{1}{2\pi\sigma}, \quad (\text{S.7})$$

so that the variance scales as a constant factor

$$\mathbb{E}[\sigma^2] \approx \left(1 - \frac{\beta^2 e^{-1}}{1-\alpha^2}\right) \sigma^2 = \xi \sigma^2. \quad (\text{S.8})$$

As a function of n , this then gives the scaling

$$\sigma_n^2 \sim e^{n \ln \xi}, \quad (\text{S.9})$$

where $\ln \xi < 0$. A larger α and β will have the effect of giving a slower exponential scaling by bringing ξ closer to one, while T serves as a sort of ‘‘cutoff’’ parameter that suppresses the exponential scaling when eventually τ becomes comparably long to T (illustrated by the green curve in Fig. S6 flattening out compared to the purple curve starting from $N \approx 6$).

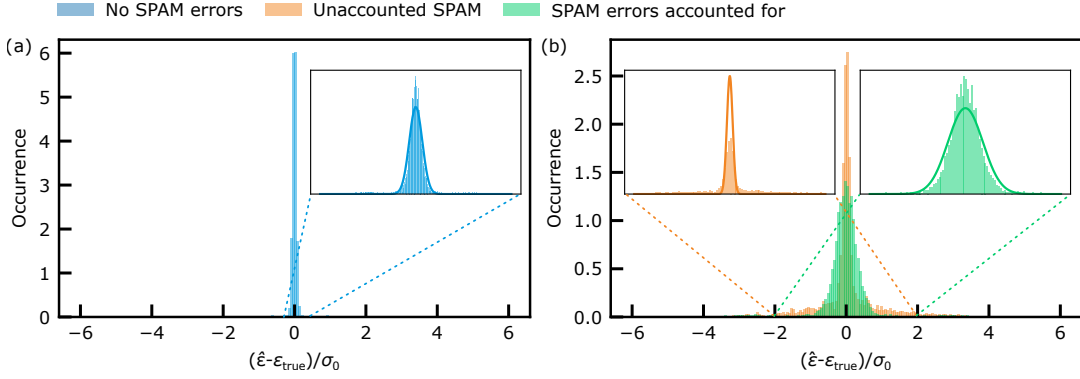


FIG. S7. **Effect of not including SPAM errors in the likelihood model.** (a) Simulated distribution of estimation errors after $N = 15$ steps without SPAM errors. (b) Simulated estimation errors after $N = 15$ steps with SPAM errors modeled using $\alpha = -0.02$ and $\beta = 0.6$, where the orange histogram shows the distribution of errors if the Bayesian update is performed assuming no SPAM errors (i.e $\alpha = 0$ and $\beta = 1$) while the green histogram shows the distribution of errors when the correct values $\alpha = -0.02$ and $\beta = 0.6$ are used in the Bayesian update model. Insets show zoomed-in histograms together and the Gaussian distributions of errors that the algorithm expects to see given all the final posterior variances (solid lines). All histograms consist of 5,000 simulated experiments and are normalized.

EFFECT OF NOT INCLUDING SPAM ERRORS IN THE LIKELIHOOD MODEL

The Bayesian scheme presented in this work only focuses on estimation of a single parameter in order to obtain a simple parametric solution to the update equations, assuming reasonably good estimates for the other parameters of the model are known beforehand, by fitting to the data of simpler preliminary experiments. In the context of the FBS, this means providing estimates of the dephasing parameter T and the SPAM error coefficients α and β . Here we investigate through simulations how neglecting SPAM errors in the model for how the measurement data is generated impacts estimation errors in the case of a model where such SPAM errors are present. Fig. S7(a) shows the simulated distribution of estimation errors from the true frequency if perfect initialization and readout could be achieved, i.e., the data generating function is

$$P(m|\varepsilon, \Delta f, \tau) = \frac{1}{2} \left(1 + m e^{-\tau/T} \cos(2\pi(\Delta f - \varepsilon)\tau) \right), \quad (\text{S.10})$$

while Fig. S7(b) shows simulated estimation error distributions when measurements are generated according to

$$P(m|\varepsilon, \Delta f, \tau) = \frac{1}{2} \left(1 + m[\alpha + \beta e^{-\tau/T} \cos(2\pi(\Delta f - \varepsilon)\tau)] \right), \quad (\text{S.11})$$

with $\alpha = -0.02$ and $\beta = 0.6$, where the orange histogram is the result when mistakenly assuming the model Eq. (S.10) and the green histogram is the result when updating according to the correct model in Eq. (S.11). We see that updating according to the wrong model parameters α and β results in tails of large estimation errors. A large mismatch in the true value of α and its estimated value would have the effect of biasing the frequency estimation error towards positive or negative values, but here α is small enough to not to have much impact. Accounting for the SPAM errors by choosing the correct values of α and β has the effect of reducing the occurrence of large outlier errors and any estimation error bias. The errors are still larger than in the case of no SPAM errors, which is due to the update equations becoming more conservative when reducing the posterior variance. This is a result of the information gathered from each measurement becoming less certain, and so more measurements are necessary to obtain as low estimation errors. The fact that most of the estimation errors can still be reduced by more measurements is an important contrast to the case of the simpler model that does not account for the SPAM errors present. Even though this simpler model has more occurrences of very low errors, the many outliers in the tails exceed the amount of large errors the algorithm would expect if it was correct, with these outliers presumably being estimates that most often cannot be improved much by further gathering of measurements. These differences in what sort of estimation errors the algorithm expects versus what errors are actually realized are shown in the insets of Fig. S7. The simulations here consist of 5,000 attempts to get estimates $\hat{\varepsilon}$ for different true values $\varepsilon_{\text{true}}$ drawn from the initial prior, which here is the standard normal distribution, i.e. $\varepsilon_{\text{true}} \sim \mathcal{N}(0, 1)$.

VALIDITY OF THE GAUSSIAN APPROXIMATION

We approximate the true posterior, starting from a Gaussian prior with an oscillating likelihood, as Gaussian. However, this assumption may not always hold, especially for certain parameter choices where the oscillatory likelihood can create posteriors with multiple peaks. For instance, specific values of τ can lead to multimodal posteriors. Here we show that an optimal choice of τ results in fairly Gaussian posteriors, as shown in Fig. S8(a). To quantify how different the true posterior is from the Gaussian we use to approximate it we can calculate the Kullback–Leibler (KL) divergence

$$\mathcal{D}_{\text{KL}}(p||q) = \int_{-\infty}^{\infty} p(\varepsilon) \log_2 \left(\frac{p(\varepsilon)}{q(\varepsilon)} \right) d\varepsilon, \quad (\text{S.12})$$

with p here denoting the true distribution and q the Gaussian approximation. In Fig. S8(b) we plot the KL divergence (green) for a range of values τ around the optimal value, together with the resulting posterior standard deviations (black), for the case of $m = -1$ (dashed) and $m = 1$ (solid). At the value τ_{opt} that minimizes the expected posterior standard deviation the KL divergence is still relatively small, indicating that the true posterior still looks much like the approximated Gaussian.

We select Δf to align the likelihood’s inflection point with the prior’s mean, and then choose τ to minimize the posterior variance. This choice makes the likelihood’s period comparable with the prior’s standard deviation, smoothing out the oscillations in the prior’s tails.

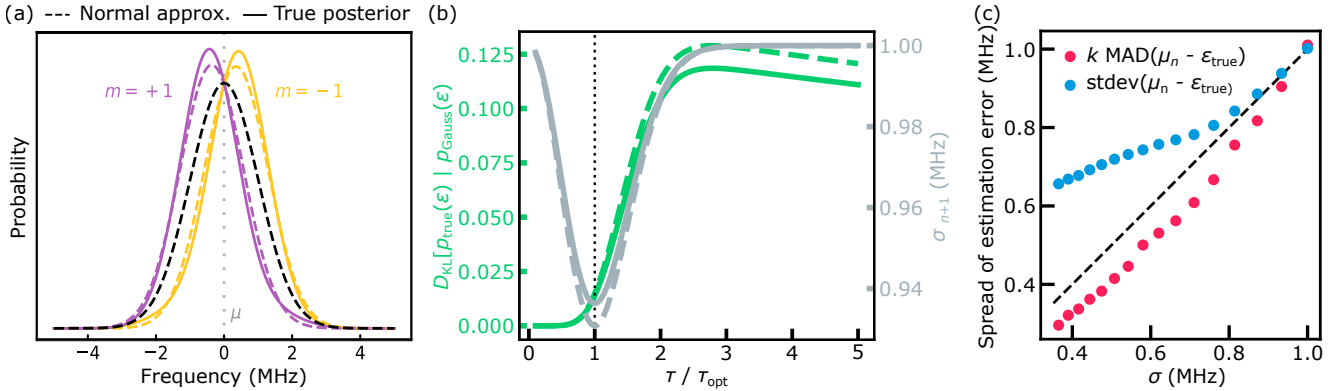


FIG. S8. **Validity of the Gaussian approximation.** (a) Approximated Gaussian distributions (dashed lines) versus true posteriors (solid lines) from a Gaussian prior of variance 1 MHz^2 (black dashed lines), with a dephasing time $T = 10 \mu\text{s}$ and SPAM errors $\alpha = -0.02$ and $\beta = 0.6$. While the oscillatory likelihood adds some skewness and broader tails, the posteriors are close to Gaussian, with deviations mainly seen in small shifts and kurtosis mismatches. (b) KL divergence between the true posterior and its Gaussian method of moments fit and the posterior standard deviation σ for the measurement outcomes $m = -1$ (dashed) and $m = +1$ (solid) at different choices of τ . (c) Posterior standard deviation as an estimator of error. The posterior distributions in the Bayesian framework provide estimators for the error in the estimate ε , here given by σ . Due to a few outliers of large errors, σ turns out to be a poor estimate for the standard deviation of errors, ie. it does not follow the diagonal black line. Instead, σ is a good estimator of the median absolute deviation (assuming the measurement generating model is correct), which is related to σ by a factor k (for normally distributed data $k = 1.4826$).

To see how well the always Gaussian approximation works across an estimation run of $N = 15$ single-shot measurements we plot the posterior standard deviation versus the actual spread of errors, quantified in two different ways, as seen in Fig. S8(c). The 16 dots for each of the two quantities show the error for $N = 0$ to $N = 15$ measurements (positioned right to left). The data here is generated by performing 5,000 simulations, where we have assumed that the likelihood function Eq. (S.11) with $\alpha = -0.02$, $\beta = 0.6$ and $T = 10 \mu\text{s}$ is indeed the correct measurement generating function. It turns out that σ seems to be a poor estimator for the actual standard deviation of the errors (blue dots). This is presumably because the standard deviation is very sensitive to outliers, and these outliers (although still quite rare) occur more often than expected due to approximating the true posteriors by Gaussians which have slightly slimmer tails. However, as seen in Fig. S7 the error distribution still overlaps very well with a zero-centered Gaussian with the average σ as standard deviation. The posterior standard deviation σ instead turns out to be quite a good estimator for the median absolute deviation (MAD), which is less sensitive to outliers. One can relate the MAD to a standard deviation σ by a factor k , so $\sigma = k \cdot \text{MAD}$. To be consistent with a normal distribution one sets

$k = 1.4826$ [4]. Thus the Gaussian approximation works well as long as it is acceptable to have some more outliers in terms of estimation error compared to what would be generated if the distribution truly was Gaussian (although the number of outliers is still small compared to expected outcomes).

* ferdinand.kuemmeth@ur.de

- [1] J. S. Rojas-Arias, Y. Kojima, K. Takeda, P. Stano, T. Nakajima, J. Yoneda, A. Noiri, T. Kobayashi, D. Loss, and S. Tarucha, The origins of noise in the Zeeman splitting of spin qubits in natural-silicon devices [10.48550/arXiv.2408.13707](https://arxiv.org/abs/10.48550/arXiv.2408.13707).
- [2] B. Paquelet Wuetz, D. Degli Esposti, A.-M. J. Zwerver, S. V. Amitonov, M. Botifoll, J. Arbiol, A. Sammak, L. M. Vandersypen, M. Russ, and G. Scappucci, Reducing charge noise in quantum dots by using thin silicon quantum wells, *Nature communications* **14**, 1385 (2023).
- [3] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, Efficient measurement of quantum gate error by interleaved randomized benchmarking, *Physical Review Letters* **109**, 080505 (2012).
- [4] P. J. Huber, *Robust statistics*, Wiley series in probability and mathematical statistics (J. Wiley & sons, New York, 1981).