Monitoring fast superconducting qubit dynamics using a neural network

G. Koolstra,1,2 N. Stevenson,1 S. Barzilí,3,4 L. Burns,3,4 K. Siva,1 S. Greenfield,3,5 W. Livingston,1 A. Hashim,1,2 R. K. Naik,1,2 J. M. Kreikebaum,1,6 K. P. O’Brien,7 D. I. Santiago,1,2 J. Dressel,3,4 and I. Siddiqi1,2,6

1Quantum Nanoelectronics Laboratory, Department of Physics, University of California at Berkeley, Berkeley, CA 94720, USA
2Computational Research Division, Lawrence Berkeley National Lab, Berkeley, CA 94720, USA
3Institute for Quantum Studies, Chapman University, Orange, CA 92866, USA
4Schmid College of Science and Technology, Chapman University, Orange, CA 92866, USA
5Department of Physics and Astronomy, University of Southern California, Los Angeles CA 90089, USA
6Materials Sciences Division, Lawrence Berkeley National Lab, Berkeley, CA 94720, USA
7Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

(Dated: April 25, 2022)

Weak measurements of a superconducting qubit produce noisy voltage signals that are weakly correlated with the qubit state. To recover individual quantum trajectories from these noisy signals, traditional methods require slow qubit dynamics and substantial prior information in the form of calibration experiments. Monitoring rapid qubit dynamics, e.g. during quantum gates, requires more complicated methods with increased demand for prior information. Here, we experimentally demonstrate an alternative method for accurately tracking rapidly driven superconducting qubit trajectories that uses a Long Short-Term Memory (LSTM) artificial neural network with minimal prior information. Despite few training assumptions, the LSTM produces trajectories that include qubit-readout resonator correlations due to a finite detection bandwidth. In addition to revealing rotated measurement eigenstates and a reduced measurement rate in agreement with theory for a fixed drive, the trained LSTM also correctly reconstructs evolution for an unknown drive with rapid modulation. Our work enables new applications of weak measurements with faster or initially unknown qubit dynamics, such as the diagnosis of coherent errors in quantum gates.

I. INTRODUCTION

Weak measurements allow the observer to reconstruct the dynamics of a quantum system, and to track the evolution of a wavefunction before its collapse to an eigenstate. For superconducting quantum circuits in particular, reconstructing individual quantum trajectories has served as a tool to monitor quantum jumps, track diffusion statistics, generate entanglement via measurement, coherently control quantum evolution using feedback, and implement continuous quantum error correction. In weak measurements with superconducting qubits coupled to a readout resonator, the dynamics of the latter are typically much faster than the former. Therefore, the field exiting the resonator is a direct measure of the qubit dynamics, which can be readily computed using quantum filters built on generalizations of Bayes’ rule, Markovian stochastic master equations, or stochastic path integrals. However, such quantum filters are not well-suited to cases where the qubit dynamics are necessarily fast, such as during a rapid entangling gate or syndrome measurement in an error-correction scheme. Interpreting the output cavity field in this non-Markovian regime for the qubit is challenging.

A quantum filter obtained by training a Long Short-Term Memory (LSTM) neural network is an attractive alternative for reconstructing trajectories in more challenging regimes. An LSTM is a variant of recurrent neural network with persistent memory, which is well suited for time series data with long temporal correlations. They have recently found applications in quantum physics ranging from quantum state tomography to qubit noise spectroscopy. The first application of an LSTM quantum filter to weak measurements successfully tracked slower qubit dynamics with a fidelity comparable to a standard quantum filter. We build upon this initial success to show that LSTM-based filters are not constrained to slow qubit dynamics, but can learn rapid and non-trivial dynamics robustly without additional prior information (e.g. from calibration measurements) which would be required by traditional methods.

In this work, we demonstrate that an LSTM accurately reconstructs quantum state trajectories of a driven superconducting qubit coupled to a readout resonator, even for qubit dynamics faster than the relaxation timescale of the resonator. Our LSTM trains entirely on experimental observations, compensates for the delays and correlations originating from the limited detection bandwidth, and outperforms conventional reconstruction methods in the case of fast qubit dynamics. By extracting and separating the coherent dynamics and measurement backaction learned by the LSTM, we observe nontrivial drive-dependent corrections consistent with theory that includes the joint dynamics of the qubit and resonator (information not given to the LSTM). The dominant corrections include a rotation of the measured eigenstates and
II. EXPERIMENTAL SYSTEM

Our qubit-resonator system (Fig. 1 and Appendix A) consists of a superconducting transmon qubit (frequency $\omega_{ge}/2\pi = 5.473$ GHz) capacitively coupled to a superconducting coplanar waveguide resonator ($\omega_{res}/2\pi = 6.679$ GHz). The interaction Hamiltonian in this regime is dispersive

$$H_{int} = \hbar \chi a^\dagger a \sigma_z,$$  

where $\hbar$ is the reduced Planck’s constant, $\chi/2\pi = 0.47$ MHz is the qubit-dependent resonator frequency shift, $a^\dagger (a)$ is the creation (annihilation) operator for the readout resonator, and $\sigma_z$ is the qubit Pauli $z$-operator. We additionally drive the qubit (at its Stark-shifted transition frequency) to induce Rabi oscillations about the $x$-axis at a variable Rabi frequency $\Omega$.

To continuously monitor this system we apply a weak probe tone to the resonator at the frequency midpoint between qubit-shifted resonances, which populates the resonator with an approximate mean photon number of $\bar{n} \approx 0.3$. After interacting with the qubit, the resonator field escapes the resonator at a rate $\kappa/2 \approx 2\pi \times 0.8$ MHz. We amplify both in-phase ($I$) and quadrature ($Q$) parts of this field using a traveling wave parametric amplifier (TWPA), perform a heterodyne measurement of its complex amplitude, and digitize the resulting pair of noisy quadrature voltages with a 1 ns sampling time. After a chosen monitoring duration $0 < T_m < 8$ µs of variable length (Fig. 1c) we turn off the Rabi drive and projectively measure the qubit along one of the cardinal axes of the Bloch sphere ($\sigma_{x,y,z}$). We use this final projective measurement as a training label and for verification via conditioned final state tomography.

Our phase-preserving detection technique [16, 39] gives rise to two distinct types of measurement backaction: nonunitary partial collapse from information that distinguishes the qubit states, and unitary phase drift from fluctuations in resonator photon number [10]. Our detection technique is robust to drifts in amplifier gain, spans several GHz, and in contrast to phase-sensitive detection of just the informational quadrature [32], allows for simultaneous weak measurement of multiple qubits [10] in future experiments.

Before passing a monitoring record into a quantum filter, we digitally filter it to remove qubit-independent high-frequency noise, then coarse-grain it into longer time bins while ensuring rapidly driven qubit dynamics is not undersampled. We then pass the coarse-grained digital record $\{(I_k, Q_k)\}_{k=0}^n$ into both the LSTM filter and a more traditional Bayesian quantum filter, and compare the resulting qubit trajectories expressed in the Bloch coordinates $\{(x_k, y_k, z_k)\}_{k=0}^n$.

We train the LSTM (Fig. 1b) to reconstruct trajectories by feeding it a subset of experimental records $\{(I_k, Q_k)\}_{k=0}^n$ reserved for training, which are of varying length $T_m$ and labeled by their final projective measurement basis and outcome. During training the LSTM adjusts its network parameters to minimize the cross-entropy loss, which is computed from predictions at $t = T_m$ and associated projective measurement outcomes (for details see Appendix D). We have found that the LSTM network converges better than a simple feedforward neural network, which we attribute to the LSTM’s ability to track long correlations in the cell state. After the LSTM converges we feed the trained model individual measurement records (Fig. 1c) from separate data sets not used for training. In the rest of this work, we analyze the trajectories $\{(x_k, y_k, z_k)\}_{k=0}^n$ predicted by the

![Figure 1](image-url)
FIG. 2. Breakdown of adiabatic trajectory reconstruction. (a) The histogram of weakly driven trajectories of length $T_m = 8.0$ μs (reconstructed by the LSTM with $dt = 40$ ns) shows rapid trajectory diffusion due to measurement backaction. The colorbar represents the trajectory probability density at each timestep. The LSTM produces trajectories (example in red) comparable to those from a steady-state Bayesian filter (yellow). (b) LSTM validation based on tomographic measurements immediately following the LSTM prediction, for the trajectories shown in (a) where $2Ω/κ = 0.2$. The dashed line with slope 1 indicates perfect validation. The inset shows the Bayesian filter validation for the same data set with slightly smaller RMS error ($ε_{z, BF} = 0.5$, $ε_{x, BF} = 0.5$, $ε_{y, BF} = 0.5$). This is partly because $ε_{z, BF}$ are computed using the entire dataset, whereas $ε_{x, y, z, LSTM}$ only have access to trajectories not used for training. (c) For fast qubit dynamics ($2Ω/κ = 2.0$) outside the adiabatic regime the predictions of the steady-state Bayesian filter and LSTM diverge (trajectory $dt = 20$ ns). (d) Validation errors averaged over the three qubit coordinates vs. Rabi frequency $2Ω/κ$, showing a breakdown of the steady-state Bayesian filter for $2Ω/κ > 1$ (dots, BF), while the LSTM validation error (squares) remains small. The performance of the Bayesian filter improves with additional numerical prior information of the expected evolution of the $z$-conditioned resonator fields (BF + numerics, and Appendix E) set by the measurement efficiency $η$. Here, $Γ_m$ is defined as the single quadrature measurement dephasing rate which is set by the amplitude of the weak measurement tone.

III. QUBIT TRAJECTORY RECONSTRUCTION WITH A NEURAL NETWORK

Qubit dynamics faster than the resonator relaxation rate κ/2 can prevent the conditioned coherent steady-states in the resonator from adiabatically following their associated qubit states [10]. To investigate the onset of this non-adiabatic behavior beyond the coherent-state approximation [8, 10], we experimentally vary the rate of the qubit dynamics by tuning Rabi frequency $Ω$ from slow dynamics ($2Ω/κ ≪ 1$) to fast dynamics that are well outside the adiabatic regime ($2Ω/κ > 1$), training the LSTM to predict trajectories for each Rabi frequency independently.

We first use the LSTM to reconstruct qubit dynamics with a weak Rabi drive ($2Ω/κ = 0.2$), where conventional steady-state methods can still accurately reconstruct trajectories. A histogram of reconstructed trajectories (Fig. 2a) shows oscillations due to the Rabi drive as well as diffusion due to measurement backaction, which slowly collapses the trajectories towards $±Z$. The competition between (i) the Rabi drive, (ii) trajectory collapse and diffusion at rate $ηΓ_m$, and (iii) trajectory dephasing at rate $2(1−η)^2Γ_m$ confines trajectories to a Bloch sphere with reduced radius (dashed lines in Fig. 2a). Here, $Γ_m$ is set by the measurement efficiency $η$. When comparing individual trajectories, the LSTM produces trajectories similar to a conventional Bayesian filter approach. This Bayesian filter method sequentially estimates the quantum state from a measurement record trained network from such test data.
using a well-known initial state and calibrated values for $\Omega$, the ensemble decay rate and qubit-conditioned resonator output fields. We quantify the error of both methods self-consistently by averaging projective measurement results of trajectories with similar predictions \[16, 17, 32\]. The averaged tomography results closely follow the LSTM predictions for all three Bloch coordinates (Fig. 2b) with RMS error ($\varepsilon_x^{\text{LSTM}}, \varepsilon_y^{\text{LSTM}}, \varepsilon_z^{\text{LSTM}}$) = (4.0, 4.9, 4.3) $\times 10^{-2}$ (definition in Eq. \[1\]). This error largely reflects our imperfect knowledge of the true quantum state from a finite number of projective measurements, see Appendix F. More training data is likely to improve both the estimate of the true quantum state and the LSTM accuracy, since the LSTM training loss does not saturate when training on increasing fractions of available training data. Nevertheless, even with a finite size training dataset considered here ($\sim 4 \times 10^5$ voltage records), the LSTM learns an accurate representation of the reduced qubit dynamics, with a lower bound on the accuracy of $1 - \varepsilon_{\text{avg}}^{\text{LSTM}}/2 = 0.978$. For large Rabi frequencies the LSTM trajectories remain faithful, even when the qubit dynamics exceeds the relaxation rate of the resonator $\kappa/2$ (Figs. 2c–d). In contrast, the Bayesian filter’s validation error increases sharply past $2\Omega/\kappa \approx 1$, because this method is sensitive to two errors related to the sampling time of voltage records, $dt$. The first error, due to stepwise application of fast coherent dynamics, grows with $\Omega dt$ [40], whereas the second error due to temporal correlations grows for $dt < 2/\kappa$ [41]. Consequently, for fast Rabi frequencies $2\Omega/\kappa > 1$ there is no choice of $dt$ for which the standard Bayesian filter produces accurate trajectories. Experimentally the breakdown of the Bayesian filter coincides with a notable delay between oscillations in the measured voltage records and the qubit coordinate $z(t)$ (see Fig. 12), which occurs because the resonator retains photons for a time $\kappa^{-1}$. In addition, the coherent oscillation amplitude in the voltage record reduces significantly for $2\Omega/\kappa > 1$, because instead of measuring the instantaneous qubit state, photons measure the time-averaged measurement operator [42]. This effect is similar to the suppression of current oscillations in a semiconducting point contact detector [43].

As a first order correction of the resonator memory effects, we adjust the Bayesian filter based on independent master equation simulations of the resonator field amplitudes conditioned on the qubit states (Fig. 2a), BF + numerics and further details in Appendix K. While this reduces the RMS error at moderate Rabi frequencies, the average error remains large for $2\Omega/\kappa > 1$. Only after an analytic treatment of the resonator memory, and including calibrated parameters such as $\chi$ and $\kappa$, does the Bayesian filter reach similar error as the LSTM for all Rabi frequencies (Fig. 2a, BF + analytics and further details in Appendix K). This indicates that with enough prior information the Bayesian filter with analytic correction reproduces the essential ingredients of the LSTM trajectories. However, the LSTM offers an accurate reconstruction method requiring no prior knowledge of coupling rates to the environment.

### IV. RESONATOR MEMORY CORRECTIONS TO QUBIT TRAJECTORIES

To gain further insight into why the LSTM outperforms the Bayesian filter for $2\Omega/\kappa > 1$, we perform statistical analysis on the LSTM trajectories. First, we bin the LSTM trajectories inside the Bloch sphere (Fig. 3k) and calculate the Bloch vector increments $\{d(y_k, z_k)\}_{k=0}^m$, where each pair can be seen as a random variable due to the stochastic nature of the measurement backaction (Fig. 3h, dots). Next, we compute the average drift $(d\bar{y}, d\bar{z})$ and diffusion from the sample mean and covariance matrix on $\{(d(y_k, z_k)\}_{k=0}^m$, respectively. The covariance matrix

$$ C(y, z) = \begin{pmatrix} \sigma_{d_y}^2 & \sigma_{d_y d_z} \\ \sigma_{d_y d_z} & \sigma_{d_z}^2 \end{pmatrix} $$

is convenient to quantify diffusion since its largest eigenvalue $\lambda_{\text{max}}$ represents the magnitude of the backaction, and the associated eigenvector $\xi$ indicates the direction of the backaction. The vector $\mathbf{v} = (v_y, v_z) = \lambda_{\text{max}}\xi$ thus quantifies diffusion magnitude and direction, and is visualized in Fig. 3b. By separating the drift and diffusion as described above, we can separately compare the average drift to Lindbladian dynamics, and the diffusion to the measurement backaction dynamics.

In the limit of a weak Rabi drive, the expected drift and diffusion can be directly calculated from the Stochastic master equation in Eqs. \[E1\]-\[E3\]. Therefore, in the plane of the Rabi drive ($x = 0$), we expect the average drift to take on the form

$$ d\bar{y} = -\Gamma y dt + \Omega z dt $$

$$ d\bar{z} = -\Omega y dt. $$

Similarly, the expected diffusion is

$$ \text{Var}(dy) = -2\eta \Gamma_m z y dt $$

$$ \text{Var}(dz) = 2\eta \Gamma_m (1 - z^2) dt. $$


Here, $\Gamma_d$ is the ensemble average dephasing, $\Gamma_m$ is the single quadrature measurement rate and $\eta$ is the total measurement efficiency. In the remainder of this section we will first study the effect of the increasing Rabi drive on the average drift, and then study deviations from the backaction model of Eqs. (8) and (9).

In Fig. 3, we show an example of the average drift for $2\Omega/\kappa = 0.6$. The trajectory drift (red arrows) shows good agreement with the expected drift from the Hamiltonian $H = \Omega/2 \sigma_x$ and dissipation pulling trajectories towards the origin at a rate $\Gamma_d/2\pi$ (gray arrows). From a fit of the average drift to Eqs. (8)-(9), using $\Omega$ and $\Gamma_d$ as free parameters, we find $\Gamma_d/2\pi = 0.188 \pm 0.007$ MHz. For increasing Rabi drive strengths, Eqs. (8)-(9) continue to fit the average drifts. Interestingly, we observe a monotonic decrease in $\Gamma_d/2\pi$ (Fig. 3d), which we attribute to decoupling of the qubit from the weak measurement when the Rabi splitting of the dressed qubit exceeds the cavity linewidth $\kappa$. This decoupling behavior is well captured by a master equation simulation of the qubit-resonator system (Fig. 3, solid line). This master equation simulation also correctly shows a small Rabi drive asymptote $\Gamma_d \rightarrow 2\Gamma_m$ due to measurement dephasing and another asymptote $\Gamma_d \rightarrow 3\Gamma_1/4$ due to natural relaxation, when qubit and resonator are completely decoupled.

Since the average drift follows the expected Lindbladian dynamics even for fast Rabi rates, we now shift our attention to the measurement backaction, which we visualize by plotting the scaled eigenvectors $\mathbf{v}$ (Fig. 3a) of $C(y,z)$. At low Rabi frequencies ($2\Omega/\kappa = 0$), we find both informational backaction, vanishing at the poles $\{\pm Z\}$ of the Bloch sphere, and phase backaction, consistent with a dispersive heterodyne measurement (see Appendix H). Furthermore, a fit of the extracted diffusion $(v_y,v_z)$ to Eqs. (8)-(9), along with $\Gamma_d/2\pi = 0.175$ MHz determined from Fig. 13a, gives the efficiency of the measurement chain $\eta = 2\Gamma_m/\Gamma_d = 0.185 \pm 0.002$. This value is in agreement with an independent calibration measurement ($\eta = 0.188 \pm 0.003$ (see Fig. 13).

For large Rabi drives, we observe two significant corrections to the expected measurement backaction of Eqs. (8)-(9): a tilt of the measurement poles in the plane of the Rabi drive towards $\{\pm Y\}$ and a reduced trajectory diffusion rate $2\eta \Gamma_m$ (Fig. 3k). These corrections complicate extracting the diffusion rate. Therefore, we first determine the rotation $\theta$ from the average angle between $\mathbf{v}$ and the z-axis from bins near the origin (Fig. 4, shaded square). Next, we apply a coordinate rotation $(y',z') \rightarrow (y \cos \theta - z \sin \theta, y \sin \theta + z \cos \theta)$ and measurement $\langle v_{y'}, v_{z'} \rangle$ to Eqs. (8)-(9). The fit results for $\theta$ and the measurement rate $2\eta \Gamma_m$ are shown in Fig. 13, squares. Intuitively, the tilt occurs because the Rabi drive drags the qubit state counterclockwise in the $yz$-plane while photons in the resonator measure the qubit $z$ coordinate for a characteristic time $\tau$. This simple picture results in a delay between qubit state and measurement record which is proportional to $\Omega \tau$, consistent with the linear increase in $\theta$ for small Rabi frequencies in Fig. 4.
FIG. 4. Resonator memory corrections to the measurement backaction from LSTM trajectories. (a) Trajectory diffusion map for $2\Omega/\kappa = 0.6$, obtained by computing the eigenvectors of the covariance matrix Eq. (2) for trajectories binned in the $yz$-plane. See also Figs. 3a,b. The tilt towards $|\pm Y\rangle$ in the plane of the Rabi drive is in contrast to the prediction from Eqs. (5)–(9). (b) The theory prediction, detailed in Appendix K, includes resonator memory effects and reproduces the tilt. (c) For small $\Omega/\kappa$ the measurement axis tilts linearly with the Rabi frequency $\theta = 2\Omega/\kappa$ (dashed line). The experimental tilt of the measurement eigenstates in the $yz$-plane (red squares) is accompanied by a decrease in measurement rate (blue squares). Solid lines are fits to Eq. (11) (red) and Eq. (12) (blue), respectively. The error bars for $\theta$ are estimated from the imprecision in determining the tilt angle. The error bars for the measurement rate obtained from fitting are smaller than the markers.

A major advantage of the LSTM is that trajectory reconstruction does not require prior knowledge of system and environment parameters, which may fluctuate or may be hard to calibrate. To highlight this advantage, we perform a new set of weak measurements where we vary the Rabi frequency sinusoidally with a period of 1.8 $\mu$s, thereby obscuring the prior information necessary for traditional trajectory reconstruction methods. The ability to reconstruct trajectories with a priori unknown time-dependencies is important for a wide range of experiments, such as continuous quantum error correction or weak measurements of a highly non-Markovian environment [48].

The resulting histogram of LSTM trajectories (Fig. 5h) shows alternating periods of trajectory bunching, where the trajectories are protected from measurement backaction, and diffusion, where the weak measurement clearly imparts stochastic kicks to the qubit. We unravel this non-trivial dynamics by dividing the LSTM trajectories in 0.2 $\mu$s-long time windows and analyze the average drift and diffusion as described in the previous section. During each 0.2 $\mu$s window, the trajectories only partly cover the $yz$-plane, but we can still extract the Rabi frequency from the average drift and measurement rate from the diffusion (Fig. 5g). Without prior knowledge of the time-dependence of $\Omega$, $\Gamma_d$ or $\Gamma_m$, we correctly recover the sinusoidal applied Rabi frequency (Fig. 5b, blue), and find that it anti-correlates strongly with the extracted measurement rate (Fig. 5b, red), consistent with the results in Fig. 4. A sinusoidal fit to the instantaneous measurement rate shows a delay of $\Delta t = 0.14 \pm 0.06$ $\mu$s with respect to the instantaneous Rabi frequency, consistent with the resonator relaxation time $2/\kappa$. The results of Fig. 5 demonstrate that the LSTM correctly reconstructs...
VI. CONCLUSIONS

We have shown that LSTM recurrent neural networks can outperform standard quantum filters if the qubit dynamics is faster than the resonator linewidth. In this previously unexplored regime of weak measurement, the LSTM correctly recovers Lindbladian dynamics and distorted measurement backaction without prior knowledge of the resonator memory. For the strongest Rabi drives the LSTM validation error remains low; however, in this case the resonator temporally averages most of the qubit signal, reducing the best trajectory estimate to an ensemble-averaged trajectory. Therefore, the LSTM is most useful in the intermediate regime $2\Omega/\kappa < 2$, where the measurement backaction is distorted, but the backaction strength is non-zero and trajectories are therefore non-trivial. We further showed that the LSTM is not constrained to constant Hamiltonians. Even when underlying parameters vary on a sub-microsecond timescale the LSTM accurately recovers quantum trajectories, which allows parameter estimation with high time resolution, and may enable trajectory reconstruction for qubits connected to strong non-Markovian environments where coupling rates can vary with time.

To further improve the LSTM accuracy for applications such as parameter estimation, it is possible to add physical constraints to the LSTM loss function. In addition, other neural network architectures such as tensor networks or models based on dilated causal convolutions may improve prediction accuracy, though such comparisons require more training data. Finally, a possible extension of our LSTM is state tracking in larger Hilbert spaces. This requires more projective measurement pulses (e.g. at least 4 for single qutrits or 9 for two qubits) compared with the single qubit state tracking, which only requires 3 tomography axes. Given the single qubit data acquisition time of approximately three hours, extending our LSTM tracking to single qutrits or two qubits seems feasible, though optimizing data transfer from the digitizer to computer memory may be desirable to keep data acquisition efficient.

The authors would like to thank K. Nowrouzi, B. Mitchell and A. Morvan for experimental assistance, and L. B. Nguyen and Y. Kim for valuable comments on the manuscript. We gratefully acknowledge insightful conversations with L. Martin, É. Genois, A. DiPaolo, J. Gross and M. Laskin. This work was supported by the U.S. Army Research Laboratory and the U.S. Army Research Office under contract/grant number W911NF-17-S-0008. L. B., S. G. and J. D. acknowledge additional support from NSF-BSF Grant Award No. 1915015.

APPENDIX A: EXPERIMENTAL SETUP

The experiments in this work were performed with a chip of 8 superconducting transmon qubits, cooled to 10 mK in a BlueFors XLD dilution refrigerator. Room-temperature and cryogenic electronics for qubit control and measurement are shown in Fig. Qubit control pulses are generated by upconversion of intermediate frequency (IF) pulses generated by a Keysight PXI arbitrary waveform generator (AWG) via in-phase quadrature (IQ) modulation of a continuous wave (CW) local oscillator (LO) tone, sourced by a Keysight MXG N5183B at 5.415 GHz. Both I and Q components are sourced at 1 GS/s
between 66 and 261 MHz. The phase and DC offsets between the I and Q waveforms are tuned to eliminate the opposite sideband and LO leakage due to mixer non-idealities, while band-pass filtering at room-temperature reduces noise from the AWG. Readout pulses are generated with the same AWG and are similarly upconverted with a 6.83 GHz LO tone from a separate Keysight MXG N5183B. Control pulses are passed through a DC block, and all pulses travelling to the sample are attenuated at each stage, with a further K&L low-pass filter at the base stage. The upconversion chain schematic for both the qubit control and readout lines is detailed in the inset labeled “ctrl” and “RO”.

Reflected measurement signals are redirected by a circulator to a measurement chain outfitted with superconducting coaxial cable, where they are amplified by a traveling wave parametric amplifier (TWPA) at 10 mK, a high electron mobility transistor (HEMT) at 4K, and a low noise room temperature amplifier, before being downconverted to IQ IF components with the same 6.83 GHz LO tone used for RO upconversion. The downconverted signals are then amplified and filtered to reduce high-frequency amplifier noise before being digitized at 1 GS/s by an Alazar analog-to-digital converter (ADC) and demodulated in software. The downconversion chain schematic is detailed in the inset labeled “Demod”.

The TWPA is pumped with a CW tone, sourced by a Hittite HMC M2100 at 7.42 GHz and 16.5 dBm before being attenuated and filtered by low-pass K&L and copper powder filters.
TABLE I. Calibrated resonator and qubit parameters for the device used in this work. Note that $\omega_{\text{res}}$ and $\omega_{\text{ge}}$ are the bare readout resonator and qubit frequencies, respectively and that $T_1$ and $T_{2,\text{Ramsey}}$ were measured without a weak measurement tone ($\varepsilon/2\pi = 0$). The error bars on $T_1$ and $T_{2,\text{Ramsey}}$ reflect the standard deviation of 100 repeated measurements during a 3-hour window.

| Parameter Description | Measured value |
|-----------------------|----------------|
| $\omega_{\text{res}}/2\pi$ | Bare cavity frequency | 6679 MHz |
| $\kappa/2\pi$ | Cavity linewidth | 1.56 MHz |
| $\omega_{\text{ge}}/2\pi$ | Transmon frequency (empty cavity) | 5473 MHz |
| $\alpha/2\pi$ | Transmon anharmonicity | -270 MHz |
| $T_1$ | Transmon relaxation time ($\Omega = 0$) | 61 $\pm$ 7 $\mu$s |
| $T_{2,\text{Ramsey}}$ | Ramsey dephasing time (empty cavity) | 70 $\pm$ 9 $\mu$s |
| $\chi/2\pi$ | Half of the dispersive cavity shift | 0.47 MHz |
| $\varepsilon/2\pi$ | Weak measurement strength | 0.43 $\pm$ 0.01 MHz |

APPENDIX B: DEVICE PARAMETERS AND STABILITY

Table I summarizes the relevant parameters of the qubit and readout resonator used in this work. We obtain the resonator frequency $\omega_{\text{res}}$ and cavity linewidth $\kappa/2\pi$ through spectroscopic measurements, and we measure the qubit frequency $\omega_{\text{ge}}$, the anharmonicity $\alpha/2\pi$ and $T_{2,\text{Ramsey}}$ through repeated Ramsey measurements on the g-e transition ($\omega_{\text{ge}}, T_{2,\text{Ramsey}}$) and e-f transition ($\alpha/2\pi$). Lastly, we fit the weak measurement strength $\varepsilon/2\pi$ by comparing the measured Rabi decay time as function of Rabi frequency to a master equation simulation using QuTiP [53]. The errorbar on $\varepsilon$ is due uncertainty in the fit.

Acquiring a training data set (i.e. one initial qubit state with three tomography axes) for a single Rabi frequency takes approximately three hours. As the neural network learns a representation of the qubit Hamiltonian, it is important that the parameters remain stable during the acquisition.

Since readout infidelity and unexpected decay or dephasing directly affect the training labels, we have characterized the stability of $T_1$, $T_{2,\text{Ramsey}}$ as well as the readout fidelity over a three-hour time span. For $T_1$ and $T_{2,\text{Ramsey}}$ we observe fluctuations on the order of 10-15% of their respective values, which is likely due to nearby two-level systems that couple weakly to the qubit. Since the fluctuations are modest, and average values of 61 and 70 $\mu$s remain much larger than the maximum trajectory length of 8.0 $\mu$s, we do not expect decay time fluctuations to affect the performance of the neural network.

We characterize readout infidelity by repeatedly preparing the qubit in the ground and excited state followed by a measurement of the qubit state. We only observe small fluctuations of the correctly classified state over a three-hour window and find average correctly classified probabilities $P(0|0) = 0.995 \pm 0.005$ and $P(1|1) = 0.984 \pm 0.004$. We observe that the average readout infidelity of $1 - \frac{1}{2}(P(0|0) + P(1|1)) = 1.0\%$ affects the accuracy of the trained neural network, but fluctuations in this quantity do not play a significant role.

The qubit studied in this work is part of a larger 8-qubit quantum processor, and is capacitively coupled to two neighboring qubits, each via a coupling bus with an effective coupling strength of approximately 2 MHz. To avoid any undesired qubit dynamics due to the neighboring qubits, we post-select the voltage records, conditioning on all qubits starting in the ground state before the Rabi sequence, and the neighboring qubits remaining in the ground state after the Rabi sequence. Typically the post-selection removes 5% of the voltage records.

APPENDIX C: PROCESSING OF MEASUREMENT RECORDS

Traditional processing of weak measurement records includes digital filtering and choosing a coarse grained $dt$ that minimizes temporal correlations in the measurement record, to ensure that the appropriate stochastic master equation can be applied faithfully to the measured data [16, 17]. As explained in the main text, for fast qubit dynamics it becomes impossible to coarse grain the record in a way that produces uncorrelated measurement records without undersampling the data. We strike a balance between undersampling and correlation by coarse graining the measurement records as listed in Table II. Note that all records are sampled faster than the uncorrelated condition $dt > 2/\kappa = 0.2 \mu$s, and the two fastest datasets are sampled more finely than all others ($dt = 20$ ns). Feeding the Bayesian quantum filter correlated measurement records is expected to contribute to the RMS error in Fig. 2d. Interestingly, the Bayesian filter remains surprisingly accurate up to $2\Omega/\kappa \approx 1$. 
Note that all voltage records within a data set have equal sequence length, since sequences with Rabi drive durations smaller than $T_m$ are zero-padded.

| $2\Omega/\kappa$ | Init qubit state | Data set size | Traj. for LSTM analysis | Longest $T_m$ (µs) | $T_m$ increment (ns) | $dt$ (ns) | Sequence length |
|-----------------|-----------------|--------------|------------------------|---------------------|---------------------|----------|--------------|
| 0.0             | $|$+$Y$         | $454 \times 10^4$ | 46,579                 | 8.0                 | 200                 | 40       | 200          |
| 0.0             | $|$-$Y$         | $454 \times 10^4$ | 46,540                 | 8.0                 | 200                 | 40       | 200          |
| 0.2             | $|$+$Z$         | $283 \times 10^4$ | 29,027                 | 8.0                 | 200                 | 40       | 200          |
| 0.6             | $|$+$Z$         | $283 \times 10^4$ | 29,029                 | 8.0                 | 200                 | 40       | 200          |
| 1.0             | $|$+$Z$         | $283 \times 10^4$ | 29,014                 | 8.0                 | 200                 | 40       | 200          |
| 1.4             | $|$+$Z$         | $283 \times 10^4$ | 29,062                 | 8.0                 | 200                 | 40       | 200          |
| 2.0             | $|$+$Z$         | $353 \times 10^4$ | 36,054                 | 5.0                 | 100                 | 20       | 250          |
| 3.2             | $|$+$Z$         | $354 \times 10^4$ | 36,118                 | 5.0                 | 100                 | 20       | 250          |

APPENDIX D:  TRAINING OF THE NEURAL NETWORK

1. Training data acquisition details

The pulse sequence for training data acquisition is given in Fig. 7a. For each experimental run, the qubit is heralded in the ground state $|0\rangle$ with a strong measurement, resulting in a large mean photon occupation number $\bar{n}$ in the cavity. The measurement tone amplitude is then reduced to the weak measurement setpoint ($\bar{n} \approx 0.3$), and a 5 µs idling time allows $\bar{n}$ to reach steady state. Next, we prepare the qubit in one of the cardinal points of the Bloch sphere with a 30 ns preparation pulse, which is detuned from $\omega_y$ to account for the Stark shift induced by the non-zero cavity $\bar{n}$. At $t = 0$ we introduce $\sigma_z$ dynamics by quickly ramping up the Rabi drive (ramp time 5 ns), while continuing to weakly measure the qubit along $\sigma_z$. This yields voltage records $(I_n, Q_n)$ of variable duration $T_m$, which ranges between 0 and 8 µs. Finally, a projective measurement along one of the principal axes of the Bloch sphere $\sigma_{x,y,z}$ is performed by first applying a single qubit gate to rotate the qubit and then a strong measurement ($\bar{n} \gg 1$). The projective measurement results are binary variables $P_i \in \{0,1\}$ and represent the training labels for the neural network.

2. Training details

In this work we use a long short-term memory recurrent neural network, a model that recurrently updates a layer of virtual neuron-like nodes as it processes a time-serialized input, and features a memory cell to retain past information of the time series. At time $t$, the layer is encoded in a vector $\tilde{h}_t$, and propagates to time $t+1$.

A typical data set with single initial qubit state, and a single Rabi frequency consists of $N \approx 4 \times 10^3$ repetitions of $M = 40$ different length voltage records, repeated for three tomography axes. We use 90% of the approximately $3MN = 4 \times 10^5$ voltage records for training and the remaining 10% for cross-validation (see Table II).

The training code uses the Tensorflow 2 library in Python and consists of several epochs where training samples are fed forward and backwards through the neural network in mini-batches of size $N_i = 512$. During the training weights and biases of the network are adjusted to minimize a cost function $L$, which contains the following three components (similar to Ref. [47]):

- A cross-entropy loss associated with the projective measurement result at the end of a voltage record

$$L_{CE} = -\frac{1}{N_i}\sum_{i=1}^{N_i} P_i \log(s_i) + (1 - P_i) \log(1 - s_i),$$

where $P_i$ and $s_i$ are the tomography result (0 or 1) and probability assigned by neural network ($0 \leq s_i \leq 1$) for voltage record $i$ in the mini-batch, respectively.

- A mean-squared-error for deviating from the known initial state at time $t = 0$, $L_{init}$

- A Rectified Linear Unit (ReLU) activation on the purity of the quantum state $L_{purity}$ at all times, which tries to enforce that the Bloch vector does not lie outside the Bloch sphere for all $t$.

We construct the loss function as a weighted sum of the components above, but note that the main contribution to the loss function comes from $L_{CE}$.

We use the ADAM optimizer [57] to obtain the LSTM parameters that minimize the cost function $L$, and adjust the step size of the stochastic gradient descent (i.e. learning rate) $\xi$ in a cyclical fashion (Fig. 7a) to prevent getting stuck in local minima [58]. We have experimented with adding dropout and regularization, techniques which aim
to improve generalization on unseen data [59], but have not observed improvements in \( \mathcal{L} \) computed on the validation dataset. Since dropout randomly resets a fraction of trainable parameters between epochs, it destroys temporal correlations between successive LSTM cells which may be hard to learn from the training data. In addition, we believe that the lack of improvement could be due to the noisy nature of the voltage records and to the stochastic nature of the projective measurements (training labels), which could prevent overfitting even in the absence of dropout and regularization.

During the training process we monitor the value of \( \mathcal{L} \) for both the training and cross-validation samples (for an example see Fig. 7). The neural network is converged when \( \mathcal{L} \) computed on the training dataset no longer decreases after an annealing cycle [33]. The cross-validation samples are then passed through the network once more for both the training and cross-validation samples (for an example see Fig. 7b). The neural network is converged when the training process we monitor the value of \( \mathcal{L} \) computed on the training dataset no longer decreases after an annealing cycle [33]. The cross-validation samples are then passed through the network once more for both the training and cross-validation samples (for an example see Fig. 7b). The neural network is converged when the training process.

We consider the Stochastic master equation (in Itô form) for heterodyne measurement of a driven qubit, where the Rabi drive is applied in the \( yz \) plane [61]:

\[
\begin{align*}
    dx &= -\Gamma_d x \, dt - \sqrt{2\eta}\Gamma_m z x \, dW_1 - \sqrt{2\eta}\Gamma_m y \, dW_2 \quad (E1) \\
    dy &= -\Gamma_d y \, dt + \Omega z \, dt - \sqrt{2\eta}\Gamma_m z y \, dW_1 + \sqrt{2\eta}\Gamma_m x \, dW_2 \quad (E2) \\
    dz &= -\Omega y \, dt + \sqrt{2\eta}\Gamma_m (1 - z^2) \, dW_3. \\
\end{align*}
\]

These equations include informational backaction (\( dW_1 \)) causing trajectories to collapse to the poles \( |\pm Z\rangle \) at a rate \( \eta \Gamma_m \), phase backaction (\( dW_2 \)) at an equal rate due to the quantum fluctuations in the cavity photon number \( \bar{n} \), and dephasing towards the origin of the Bloch sphere at a rate \( 2(1 - \eta)\Gamma_m \). Note that we define \( \Gamma_m \) as the decoherence rate due to measurement of a single quadrature, and since we measure both \( I \) and \( Q \) quadratures the total dephasing rate includes a factor of 2. Further note that for a unit efficiency measurement setup, individual trajectories diffuse quickly but remain pure. Therefore, the reduced radius is due to limited measurement efficiency \( \eta \).

To further quantify the reduced radius observed in the experimental trajectories, we calculate the mean change of the purity, defined as \( |r|^2 = x^2 + y^2 + z^2 \):

\[
\left\langle \frac{d|r|^2}{dt} \right\rangle = -2[(1 - \eta)\Gamma_m + \Gamma_m] (|r|^2 - z^2) + 2\eta\Gamma_m (1 - z^2) + 2\eta\Gamma_m |r|^2 z^2. \\
\left( E4 \right)
\]

We can simplify this expression in the fast Rabi regime (\( \Omega \gg \Gamma_m \)), where \( z^2 \approx |r|^2/2 \) evaluated over one Rabi cycle. This yields

\[
\left\langle \frac{d|r|^2}{dt} \right\rangle = -2\eta\Gamma_m - (2 + \eta)\Gamma_m |r|^2 + \eta\Gamma_m |r|^4. \\
\left( E5 \right)
\]

From Eq. \( \left( E5 \right) \) we calculate the steady state solution for the reduced radius

\[
|r| = \sqrt{\frac{2 + \eta}{2\eta}} - \sqrt{\left(\frac{2 + \eta}{2\eta}\right)^2 - 2}. \\
\left( E6 \right)
\]
APPENDIX F: ACCURACY OF THE PREDICTED TRAJECTORIES BY THE LSTM

In Fig. 2 of the main text, we demonstrate the accuracy of the LSTM trajectories for a slow Rabi drive $2\Omega/\kappa = 0.2$. The purpose of this appendix is to define the accuracy and to discuss error sources.

To quantify the LSTM accuracy, we bin all predictions immediately before a projective measurement associated with each trajectory and average the projective measurement results for trajectories with similar predictions. This technique is frequently used to experimentally verify an estimate of the qubit state with an independent measurement [15, 62]. We use the following definition of the weighted root mean square (RMS) error for the $x$ coordinate ($y$ and $z$ are similar):

$$
\varepsilon_x = \sqrt{\frac{\sum_{i=1}^{N_x} N_i (x_{\text{LSTM},i} - \langle x_{\text{proj},i} \rangle)^2 / \sum_{i=1}^{N_x} N_i}{N_x}},
$$

where $N_x$ is the number of bins, $x_{\text{LSTM},i} \in [-1, 1]$ is the predicted value by the LSTM for bin $i$, and $\langle x_{\text{proj},i} \rangle$ is the average of projective measurements for predictions within $\delta = 2/N_x$ of $x_{\text{LSTM},i}$. The definition of Eq. (F1) captures any inaccuracy caused by the LSTM (e.g. inefficient model, insufficient training data) but is also sensitive to the uncertainty in our best estimate of the ground truth $\langle x_{\text{proj},i} \rangle$, which scales as $1/\sqrt{N_i}$. This becomes a problem for rare predictions, when the number of trajectories $\tilde{N}_i$ is small. Therefore, by weighing each bin by the number of predictions in that bin $\tilde{N}_i$, $\varepsilon_x$ becomes less sensitive to limited statistics for rare predictions.

In Fig. 9 we show that the uncertainty in the projective measurement dominates any errors caused by the LSTM, since the absolute error $|\langle x_{\text{LSTM},i} \rangle - \langle x_{\text{proj},i} \rangle|$ closely follows the expected uncertainty in $\langle x_{\text{proj},i} \rangle$, estimated from the Bernoulli distribution:

$$
\sigma(x_{\text{proj},i}) = \sqrt{\frac{1 - \langle x_{\text{proj},i} \rangle^2}{N_i}}.
$$

We observe similar trends in the absolute error and the uncertainty estimated from Eq. (F2) for all Bloch coordinates (Fig. 9b-c), and error magnitudes are consistent. In other words, our ability to quantify the LSTM error is limited by our knowledge of the true quantum state. To further assess the LSTM performance, more trajectories are needed to reduce our uncertainty in the estimate of the true quantum state. Alternatively, it is possible to train the LSTM on simulated trajectories for which the true quantum state is known [47].

APPENDIX G: CAVITY CORRECTIONS TO THE BAYESIAN FILTER

Quantum trajectory theory describes how an observer’s estimate of a quantum state evolves as it is updated with a weak measurement record [8, 62]. The measurement record is translated to quantum state evolution by first applying the unitary evolution and then updating the state with the measurement record at each time step, as the backaction of the measurement on the state can be derived [25]. Thus, with knowledge of the initial state and the Hamiltonian driving unitary evolution, the density matrix $\rho_{ij}(t_i)$ can be repeatedly updated.

We follow the procedure outlined in Refs. [10, 63]. For each time $t_i$ in the voltage record, we first apply the Lindbladian dynamics

$$
\rho'(t_i) = \exp(i\lambda dt) \rho(t_i),
$$

where $\lambda$ is a $4 \times 4$ matrix and $\rho(t_i) = (\rho_{00}, \rho_{01}, \rho_{10}, \rho_{11})$. Including a Rabi drive detuning $\Delta$ and a Rabi drive mis-
alignment $\Omega_y$, $\lambda$ takes the form

$$
\lambda = \begin{pmatrix}
0 & -\Omega_z & -\Omega_x & -i\Gamma_1 \\
\Omega_z & 0 & -\Delta & 0 \\
-\Omega_x & \Delta & 0 & 0 \\
-\Omega_z & 0 & i\Gamma_1 & 0 \\
\end{pmatrix}
$$

where

$$
\alpha = \omega_{\text{ge}} - \omega_{\text{rabi}} \\
\Delta = \omega_{\text{ge}} - \omega_{\text{rabi}} \\
\Gamma_1 = \frac{1}{\tau_{\text{rabi}}}
$$

Note that $\lambda$ describes the same Lindbladian dynamics that leads to Eqs. (E1)-(E3) provided $\Omega_y = 0$, and $dt = t_{i+1} - t_i$ is the time increment of the voltage record. The parameters $\Omega_{x,y,z}$, $\Delta$ and $\Gamma_1$ are obtained from a fit to the projective measurement results. Next, we update the density matrix taking into account measurement record $(I(t_i), Q(t_i))$ and the partially updated density matrix $\rho'(t_i)$:

$$
\rho_{11}(t_{i+1}) = \frac{\rho'_{11}(t_i) / \rho_{00}(t_i) e^{\alpha_i}}{1 + \rho'_{11}(t_i) / \rho_{00}(t_i) e^{\alpha_i}}
$$

$$
\rho_{00}(t_{i+1}) = 1 - \rho_{11}(t_{i+1})
$$

$$
\rho_{10}(t_{i+1}) = \rho'_{10}(t_i) \sqrt{\frac{\rho_{11}(t_{i+1}) / \rho_{00}(t_{i+1})}{\rho_{11}(t_i) / \rho_{00}(t_i)}} e^{-i\beta_i} e^{-2(1-\eta)\Gamma_m dt}
$$

$$
\rho_{01}(t_{i+1}) = \rho'_{01}(t_i)
$$

where $\alpha_i = \hat{I}_i \Delta I / \sigma^2$, $\beta_i = \hat{Q}_i \Delta I / 2\sigma^2$, $\sigma^2$ is the variance of the noise in the voltage record and finally

$$
\hat{I}_i = I(t_i) - \frac{I_0 + I_1}{2}
$$

$$
\hat{Q}_i = Q(t_i) - Q_0
$$

$$
\Delta I = I_1 - I_0.
$$

$I_0$ and $I_1$ are the steady state coherent state amplitudes conditioned on the qubit in the ground state and excited state, respectively. These state update equations are exact in the absence of a Rabi drive, and if the qubit decay rates are small compared with $\kappa$. (10). Fig. 2 in the main text shows that these equations accurately reconstruct trajectories up to $2\Omega/\kappa \approx 1$ even though they do not contain the cavity.

While $\Delta I$ can be obtained with a simple calibration measurement when $\Omega = 0$, it is impossible to calibrate $\Delta I$ in the presence of a Rabi drive, since the measurement term ($\sigma_z$) and Rabi drive ($\sigma_x$) do not commute. One might assume that $\Delta I$ does not change with the applied Rabi frequency, but this leads to a relatively large RMS error even at moderate Rabi drives. We can improve the performance of the Bayesian filter reconstruction at moderate Rabi drive by adjusting $\Delta I$ based on prior information obtained from a QuTiP numerical simulation which includes qubit and coupling to a cavity. From these simulations (Fig. 10b) we find that the steady state coherent state amplitudes conditioned on the qubit state ($\alpha_{g,e}$) substantially decrease as the Rabi frequency increases. Therefore, to correct for cavity effects, we scale $\Delta I$ by the steady state value of $\alpha_{e}(\Omega) - \alpha_{g}(\Omega)/(\alpha_{e}(0) - \alpha_{g}(0))$ (Fig. 10b), which yields the triangles in Fig. 2 of the main text with substantially smaller RMS error.

While the scaling factor for $\Delta I$ partially corrects for the cavity memory effect, the RMS error remains large for $2\Omega/\kappa \gg 1$, since temporal correlations in the voltage record are large. In this case the simple steady-state Bayesian state update equations are no longer valid, and the most straightforward way to further reduce the validation error is by including the full cavity in the stochastic master equation. However, as stated in the main text, this requires precise knowledge of coupling rates to the cavity, which can be hard to calibrate and may depend on the Rabi frequency (similar to $\Delta I$). In addition, it is difficult to experimentally verify whether the cavity modes are tracked accurately, as this would require cavity state tomography.

In Fig. 2 we show a further improvement in valida-
FIG. 10. (a) QuTiP master equation simulation of a driven qubit dispersively coupled to the readout cavity, showing coherent state amplitudes vs. time, conditioned on the qubit state. We show two pairs of \( \alpha_g \) and \( \alpha_e \) for \( 2\Omega/\kappa = 0.4 \) and \( 2\Omega/\kappa = 2.0 \) obtained from the expectation value \( \alpha_g \approx \text{Tr}(\rho_{q+c}\langle g \rangle / P_g) \), where \( a \) is the cavity photon annihilation operator and \( \rho_{q+c} \) is the joint qubit-cavity density matrix. (b) We use the relative distinguishability, defined as \( \alpha_g(\Omega) - \alpha_e(\Omega)/\alpha_g(0) - \alpha_e(0) \), to scale \( \Delta I \) and partially correct for cavity effects in the Bayesian filter method.

APPENDIX H: VERIFICATION OF HETERODYNE MEASUREMENT BACKACTION

A phase preserving measurement technique gives rise to both informational and phase backaction. In Fig 11 we verify that the undriven trajectories reconstructed by the LSTM indeed show the signatures predicted by Eqs. (E1)-(E3).

We analyze the measurement backaction for dynamics in the \( x = 0, y = 0 \) and \( z = 0 \) planes separately and the plots show good agreement with the expected measurement backaction, since we observe phase backaction in the \( xy \) plane and informational backaction in the \( xz \) and \( yz \) planes. This suggests that the LSTM accurately predicts heterodyne measurement backaction from experimental observations, and that the parametrization of the measurement backaction correctly describes the experiment for \( 2\Omega/\kappa \ll 1 \).

APPENDIX I: COMPARISON BETWEEN VOLTAGE RECORDS AND RECONSTRUCTED TRAJECTORIES

To further understand the breakdown of the Bayesian filter reconstruction demonstrated in Fig 2b of the main text it is helpful to compare the averaged voltage records with the averaged qubit trajectories reconstructed by the LSTM. For fast qubit dynamics (Fig 12b) we observe a large phase difference between the oscillations in the voltage record and those in \( z(t) \), because the cavity memory delays photons escaping to the transmission line while the qubit rotates quickly. Conventional stochastic master equations or Bayesian filters that do not include the cavity memory assume photons measure the qubit state \( z(t) \) instantaneously and therefore, this phase difference signals the breakdown of those methods. Note that this phase difference already becomes noticeable at intermediate Rabi frequencies (Fig. 12b), consistent with the increase in RMS error at \( 2\Omega/\kappa = 0.6 \) shown in Fig 2b of the main text.

APPENDIX J: EFFICIENCY CALIBRATION

To verify whether the LSTM correctly captures the measurement backaction we compare the extracted measurement rate from the LSTM with an independent calibration measurement, which extracts the separation \( S = (\Delta V/\sigma)^2 \) of conditional qubit histograms (Fig 13b) after weak measurement of variable duration \( T_m \). The characteristic measurement time measures how quickly we gain information about the qubit state in the presence of experimental inefficiencies, and is defined by the time it
Theory

Diffusion extracted from LSTM trajectories

FIG. 11. Comparison between theoretical heterodyne measurement backaction (gray quiver maps) and reconstructed measurement backaction by the LSTM (red quiver maps) for an undriven qubit initially prepared in $|\pm Y\rangle$. The theory is obtained by setting $z, y, x = 0$ in the measurement backaction term of Eqs. (E1)-(E3) for backaction in the $xy$, $xz$ and $yz$ plane, respectively. The reconstructed quiver maps do not cover the full $xy$ and $xz$ plane due to the limited measurement efficiency $\eta$ but agree qualitatively with the theory.

FIG. 12. Comparison between averaged voltage records and averaged $z$ trajectories reconstructed by the LSTM for increasing Rabi frequencies (from left to right). Both the trajectories and voltage records are averaged $M = 5 \times 10^3$ times.

takes to reach a histogram separation of $\Delta V = 2\sigma$:

$$\tau_m = 4 \left( \frac{dS}{dT_m} \right)^{-1}. \quad (J1)$$

In the undriven case where the coherent states in the cavity have reached steady state, the single quadrature measurement rate $\Gamma_m$ is given by [61]

$$\Gamma_m = \frac{1}{2\eta\tau_m}, \quad (J2)$$

where $\eta$ is the total efficiency of the measurement chain. In addition, from the definition of the measurement rate we have $\eta = 2\Gamma_m/\Gamma_d$ [22], where the factor of two accounts for both quadratures and $\Gamma_d$ is the decay rate of the ensemble average (Fig 13c). The results of the calibration are shown in Fig. 13b, c and from Eq. (J1) we find $\tau_m = 4.85 \pm 0.08$ $\mu$s and $\Gamma_d/2\pi = 0.175$ MHz which gives $\eta = 0.188 \pm 0.003$. This value agrees with the value found from analyzing the stochastic map at $\Omega = 0$ ($\eta = 0.185 \pm 0.002$). Note that this efficiency calibration is not valid when the coherent states of the cavity are not in equilibrium, for example when the qubit is subject to a Rabi drive or when $\Gamma_d > \kappa$. In those cases the methods presented in Ref. [65] could offer a way to calibrate the efficiency, but this is beyond the scope of this work.

APPENDIX K: DERIVATION OF THE CORRECTIONS TO THE MEASUREMENT BACKACTION

1. Problem setup

A superconducting transmon is measured by coupling it to a strongly detuned readout resonator, which shifts and splits the resonator frequency by a contrast of $2\chi$ between the two qubit-state-conditioned resonances. The ratio $2\chi/\kappa$ of this contrast to the resonance half-width $\kappa/2$ characterizes the distinguishability between resonances, and thus the amount of qubit information per resonator photon that can be stored in correlations. More precisely, given an input drive $d(t)e^{-i\omega_d t}$ tuned to the midpoint between qubit resonances, photons in the resonator mode $\hat{a}$ encode this qubit information in their relative phase, which follows from the Heisenberg evolution of the resonator mode in the rotating frame of the drive, $\hat{a}'(t) = -(\kappa/2)[1 + i(2\chi/\kappa)\hat{z}(t)]\hat{a}(t) + \sqrt{\kappa}\hat{d}(t)$. 
2. General derivation

When the qubit $\hat{z}$ is stationary and the drive $\langle \hat{d} \rangle = -i\varepsilon/\sqrt{\kappa}$ fluctuates around a constant mean, the resulting steady state $\hat{\rho}^{ss.} = -i\sqrt{\bar{n}} \exp(-i\phi)$ has a Lorentzian mean photon number $\bar{n} = |2\varepsilon/\kappa|^2/(1 + (2\chi/\kappa)^2)$ and qubit-dependent phase $\phi = \arctan[(2\chi/\kappa)\bar{z}]$, with a maximum phase contrast of $\Delta\phi_{\text{max}} = 2\arctan(2\chi/\kappa)$. A homodyne measurement aligned with the quadrature of maximum separation thus has a maximum amplitude contrast $\Delta\bar{a}_\text{max} = 2\sqrt{\bar{n}} \sin(\Delta\phi_{\text{max}}/2) = \sqrt{\bar{n}}(4\chi/\kappa)/\sqrt{1 + (2\chi/\kappa)^2}$ that sets the rate $\gamma_m = (\eta\varepsilon/2)|\Delta\bar{a}_\text{max}|^2 = \eta(8\chi^2\bar{n}/\kappa)/(1 + (2\chi/\kappa)^2)$ at which maximally separated steady-states can be distinguished by the photon amplitudes escaping the resonator at rate $\kappa/2$ and being successfully collected with efficiency $\eta$. The uncollected photons and residual qubit-resonator entanglement further contribute to a total qubit ensemble-dephasing rate due to measurement $\Gamma_m \approx \gamma_m/\eta$.

In the presence of a qubit drive, the resonator response additionally filters the evolution of $\hat{z}(t)$ to produce an effectively adiabatic response to a time-ordered geometric series of its delay-averages. That is, with a similarly constant drive $\langle \hat{d} \rangle = -i\varepsilon/\sqrt{\kappa}$ and $t \gg 2/\kappa$ to let transients decay, the resonator evolution has the recurrence relation

$$\hat{a}(t) = -i\frac{2\varepsilon}{\kappa} - i\frac{2\chi}{\kappa} \int_0^t \hat{z}(t-\tau) \hat{a}(t-\tau) e^{-\kappa \tau^2/2} d\tau,$$

which implies

$$\hat{a}(t) = -i\frac{2\varepsilon}{\kappa} J \sum_{n=0}^\infty \left[ -i\frac{2\chi}{\kappa} \int_0^t \hat{z}(t-\tau) dP(\tau) \right]^n. \quad (K1)$$

The convolution kernel in the Green’s function is an exponential probability distribution $dP(\tau) = \kappa e^{-\kappa \tau^2/2} d\tau/2$ over delay-times $\tau$, normalized as $\int_0^\infty dP(\tau) = 1$ with mean and variance both equal to the time constant $\tau_c = 2/\kappa$.

When $\hat{z}$ varies slowly on the timescale $\tau_c$ it can be approximately pulled outside the integral of Eq. (K1) to yield the standard steady-state solution but with a time-dependent phase $\phi(t) = \arctan[(2\chi/\kappa)\hat{z}(t)]$ that adiabatically tracks the qubit evolution. The next-order approximation treats the evolution as approximately linear within the exponential envelope $\hat{z}(t-\tau) \approx \hat{z}(t) - \hat{z}'(t)\tau_c$, which additionally delays the response to the qubit by the mean delay time $\tau_c = 2/\kappa$ to produce the effective phase $\phi(t) = \arctan[(2\chi/\kappa)\hat{z}(t-\tau_c)]$. For more rapid evolution, part of the evolution is averaged, thus reducing the measurement contrast while rotating the measurement basis.

3. Applying the theory to constant Rabi drive

In the case of a constant Rabi drive, the delay-average in Eq. (K1) can be computed directly. Assuming dominant harmonic evolution $\hat{z} = \Omega \hat{y}$, $\hat{y} = -\Omega \hat{z}$, repeated integration-by-parts of the delay-averaged $\hat{z}(t)$ when $t \gg \tau_c$ yields a pair of geometric series defining an effective $\hat{z}_\text{eff}(t)$ characterized by an adiabaticity parameter $(2\Omega/\kappa)$,

$$\hat{z}_\text{eff}(t) = \int_0^t \hat{z}(t-\tau) dP(\tau)$$

$$= \frac{1}{1 + (2\Omega/\kappa)^2} \hat{z}(t) - \frac{(2\Omega/\kappa)}{1 + (2\Omega/\kappa)^2} \hat{y}(t)$$

$$= \sqrt{\eta_{\text{avg}}} \left[ \cos \theta \hat{z}(t) - \sin \theta \hat{y}(t) \right]. \quad (K3)$$

The averaging both attenuates the eigenvalue contrast of $\hat{z}$ by an efficiency factor $\eta_{\text{avg}}$ and rotates the observable coupled to the resonator by an angle $\theta$ in the $yz$-plane.
The tilt angle and efficiency are thus
\[ \theta = \arctan(\Omega \tau) = \arctan(2\Omega / \kappa) \]  
\[ \sqrt{\eta_{\text{avg}}} = \cos \theta = \frac{1}{\sqrt{1 + (2\Omega / \kappa)^2}}. \]

At longer times \( t \gg \tau \), the geometric series in Eq. (K1) then yields the standard steady state, but with a phase angle that depends upon the efficiency-delay-averaged observable \( \hat{z}_{\text{eff}}(t) \) that is rotated by \( \theta \) and with eigenvalues reduced by \( \hat{z}_{\text{eff}}^2 = \eta_{\text{avg}} \). This tilt can be understood equivalently as two simultaneous measurements along \( z \) and \( y \) with differing measurement rates \( \gamma \) and \( \gamma_y = \eta_{\text{avg}} \gamma_m \sin^2 \theta \), that compete to rotate the effective measurement poles.

In the main text we fit the tilt angle and measurement rate extracted from the LSTM to the expressions for \( \theta \) and \( \eta_y = \eta_{\text{avg}} \gamma_m \) from Eq. (K5). These results are shown in Fig. [3].

**APPENDIX L: DERIVATION OF MEASUREMENT BACKACTION MAPS WITH CAVITY EFFECTS**

To predict measurement backaction statistics shown in Fig. [4], it is convenient to analyze an unnormalized Bloch-vector description \( \hat{s} = (y, z, p) \) restricted to the \( yz \)-plane, where \( p \in (0, 1) \) is the total probability and the standard Bloch coordinates are obtained after renormalization, \( \hat{S} = \mathcal{N}(\hat{s}) = (y/p, z/p) \). Focusing on the informational part of the measurement, a collected signal scaled to a measured observable \( \hat{z} \) will have the form \( r \, dt = (\hat{z}) \, d\tau + \sqrt{\gamma} D W \) integrated over a small time bin \( dt \), where \( \tau = 1/2 \eta \Gamma_m \) is a characteristic timescale, \( \Gamma_m \) is the induced ensemble-dephasing rate, \( \eta \in [0, 1] \) is the efficiency, and \( dW \) with \( \langle dW \rangle = 0 \) and \( \langle dW^2 \rangle = dt \) is a Weiner-increment of zero-mean Gaussian noise.

Observing such a signal \( r \) induces a partial state collapse that has the form of a hyperbolic boost matrix \( [60, 67] \),

\[ \hat{s}^{(r, dt, \tau)} = \hat{M}_z(r, dt, \tau) \hat{s} \]  
\[ \hat{M}_z(r, dt, \tau) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cosh(r \, dt / \tau) & \sinh(r \, dt / \tau) \\ 0 & \sinh(r \, dt / \tau) & \cosh(r \, dt / \tau) \end{bmatrix}, \]

Similarly, the measurement of a rotated observable \( \hat{z}(\theta) \) that is tilted by an angle \( \theta \) in the \( yz \)-plane and further attenuated by an efficiency \( \eta_z \) is readily obtained using a rotation matrix,

\[ \hat{M}_{z(\theta)}(r, dt, \tau, \eta_z) = \hat{R}_x(\theta) \hat{M}_z(r, dt, \tau / \eta_z) \hat{R}_x^{-1}(\theta) \]

\[ \hat{R}_x(\theta) = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \]

Computing the normalized state increment to linear order in \( r \, dt \) yields the expected stochastic state increment,

\[ \langle \hat{S}(r, dt, \tau, \eta_z, \theta) \rangle = \mathcal{N}(\hat{M}_{z(\theta)} \hat{s} - \hat{s}) = \begin{bmatrix} dy \\ dz \end{bmatrix} \]

\[ = \eta_z \frac{r \, dt}{\tau} \begin{bmatrix} -yz \cos \theta + (1 - y^2) \sin \theta \\ -zy \sin \theta + (1 - z^2) \cos \theta \end{bmatrix} + O(r \, dt)^2. \]

Ensemble-averaging the signal \( r \) produces its mean \( \langle \hat{S} \rangle = z \cos \theta + y \sin \theta \), yielding the mean backaction,

\[ \langle \hat{dS} \rangle = \begin{bmatrix} \langle dy \rangle \\ \langle dz \rangle \end{bmatrix} \]

\[ \approx \eta_z \frac{r \, dt}{\tau} \begin{bmatrix} -yz \cos \theta + (1 - y^2) \sin \theta \\ -zy \sin \theta + (1 - z^2) \cos \theta \end{bmatrix}, \]

from collapse, suppressing the distinct ensemble-dephasing terms appearing at second-order, \( \langle (r \, dt)^2 \rangle = \tau^2 + O(d \tau)^2 \).

To compute the variance of this informational backaction, it is convenient to decompose the random variable \( r \) into a convex mixture, \( r \, dt = (+1 \, dt + \sqrt{\gamma} D W^+) p_+ + (-1 \, dt + \sqrt{\gamma} D W^-)(1 - p_+) \) of two independent zero-mean Gaussian random variables, \( D W^+ \) and \( D W^- \), conditioned on definite eigenvalues \( \pm \hat{z} \).

Due to the linear prefactor of \( r \) in Eq. (L5), the variances can be readily computed using the conditional variance rule: the variance of the mixture is the mixture of the conditional variances plus the variance of the mixture means, \( \text{Var}(Y) = \text{Var}(Y|X) \times + \langle (Y|X)^2 \rangle_X - \langle (Y|X) \rangle_X^2 \). Computing \( \text{Var}(r \, dt) \), keeping the lowest order in \( dt \), and taking the square root then yields the standard deviation,

\[ \text{Std}(dS) = \sqrt{\text{Var}(dS)} \approx \sqrt{2 \eta_z \frac{d \tau}{\tau} \begin{bmatrix} -yz \cos \theta + (1 - y^2) \sin \theta \\ -zy \sin \theta + (1 - z^2) \cos \theta \end{bmatrix}}. \]

Note that the directionality of the standard deviation has been preserved when taking the square root.

Using the tilt angle \( \theta = -\theta = -\arctan(2\Omega / \kappa) \) and the additional time-averaging efficiency factor \( \eta_\gamma = \eta_{\text{avg}} = \cos^2 \theta = 1/(1 + (2\Omega / \kappa)^2) \) outlined in the previous section produces the variance vector plots in Fig. [4] of the main text.

---

[1] C. W. Gardiner and M. J. Collett, Input and output in damped quantum systems: Quantum stochastic differential equations and the master equation, Physical Review A 31, 3761 (1985).

[2] A. Barchielli, Measurement theory and stochastic differential equations in quantum mechanics, Physical Review A 34, 1642 (1986).
[3] L. Diósi, Continuous quantum measurement and itô formalism, Physics Letters A 129, 419 (1988).
[4] V. P. Belavkin, Quantum continual measurements and a posteriori collapse on crp, Communications in Mathematical Physics 146, 611 (1992).
[5] H. J. Carmichael, Quantum trajectory theory for cascaded open systems, Phys. Rev. Lett. 70, 2273 (1993).
[6] H. M. Wiseman and G. J. Milburn, Quantum theory of field-quadrature measurements, Physical Review A 47, 642 (1993).
[7] A. N. Korotkov, Selective quantum evolution of a qubit state due to continuous measurement, Physical Review B 63, 115403 (2001).
[8] J. Gambetta, A. Blais, M. Boissonneault, A. A. Houck, D. I. Schuster, and S. M. Girvin, Quantum trajectory approach to circuit QED: Quantum jumps and the Zeno effect, Physical Review A - Atomic, Molecular, and Optical Physics 77, 1 (2008).
[9] A. N. Korotkov, Quantum Bayesian approach to circuit QED measurement, in Quantum Machines: Measurement and Control of Engineered Quantum Systems (Oxford University Press, Oxford, 2011) pp. 333–356.
[10] A. N. Korotkov, Quantum Bayesian approach to circuit QED measurement with moderate bandwidth, Physical Review A 94, 1 (2016).
[11] R. Vijay, D. H. Slichter, and I. Siddiqi, Observation of quantum jumps in a superconducting artificial atom, Physical review letters 106, 110502 (2011).
[12] M. Hatridge, S. Shankar, M. Mirrahimi, F. Schackert, K. Geerlings, T. Brecht, K. M. Sliwa, B. Alsd, L. Frunzio, S. M. Girvin, R. J. Schoelkopf, and M. H. Devoret, Quantum back-action of an individual variable-strength measurement, Science 339, 178 (2013).
[13] L. Sun, A. Petrenko, Z. Leghtas, B. Vlastakis, G. Kirchmair, K. M. Sliwa, A. Narla, M. Hatridge, S. Shankar, J. Blumoff, L. Frunzio, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, Tracking photon jumps with repeated quantum non-demolition parity measurements, Nature 511, 444 (2014).
[14] U. Vool, S. Shankar, S. O. Mundhada, N. Ofek, A. Narla, K. Sliwa, E. Zalys-Geller, Y. Liu, L. Frunzio, R. J. Schoelkopf, S. M. Girvin, and M. H. Devoret, Continuous Quantum Nondemolition Measurement of the Transverse Component of a Qubit, Physical review letters 117, 133601 (2016).
[15] K. W. Murch, S. J. Weber, C. Macklin, and I. Siddiqi, Observing single quantum trajectories of a superconducting quantum bit, Nature 502, 211 (2013).
[16] P. Campagne-Ibarcq, P. Six, L. Bretheau, A. Sarlette, M. Mirrahimi, P. Rouxen, and B. Huard, Observing quantum state diffusion by heterodyne detection of fluorescence, Phys. Rev. X 6, 011002 (2016).
[17] Q. Ficheux, S. Jezouin, Z. Leghtas, and B. Huard, Dynamics of a qubit while simultaneously monitoring its relaxation and dephasing, Nature Communications 9, 1926 (2018).
[18] S. Hacohen-Gourgy, L. S. Martin, E. Flurin, V. V. Ramasesh, K. B. Whaley, and I. Siddiqi, Quantum dynamics of simultaneously measured non-commuting observables, Nature 538, 491 (2016).
[19] N. Roeh, M. E. Schwartz, F. Motzoi, C. Macklin, R. Vijay, A. W. Eddins, A. N. Korotkov, K. B. Whaley, M. Sarovar, and I. Siddiqi, Observation of measurement-induced entanglement and quantum trajectories of remote superconducting qubits, Physical Review Letters 112, 1 (2014).
[20] A. Chantasri, M. E. Kimchi-Schwartz, N. Roch, I. Siddiqi, and A. N. Jordan, Quantum trajectories and their statistics for remotely entangled quantum bits, Phys. Rev. X 6, 041052 (2016).
[21] R. Vijay, C. Macklin, D. H. Slichter, S. J. Weber, K. W. Murch, R. Naik, A. N. Korotkov, and I. Siddiqi, Stabilizing Rabi oscillations in a superconducting qubit using quantum feedback, Nature 490, 77 (2012).
[22] G. De Lange, D. Riste, M. Tiggelman, C. Eichler, L. Tornberg, G. Johansson, A. Wallraff, S. Schouten, and L. DiCarlo, Reversing quantum trajectories with analog feedback, Physical Review Letters 112, 080501 (2014).
[23] Z. K. Minev, S. O. Mundhada, S. Shankar, P. Reinhold, R. Gutiérrez-Jaurégui, R. J. Schoelkopf, M. Mirrahimi, H. J. Carmichael, and M. H. Devoret, To catch and reverse a quantum jump mid-flight, Nature 570, 200 (2019).
[24] L. S. Martin, W. P. Livingston, S. Hacohen-Gourgy, H. M. Wiseman, and I. Siddiqi, Implementation of a canonical phase measurement with quantum feedback, Nature Physics 16, 1046 (2020).
[25] S. Hacohen-Gourgy and L. S. Martin, Continuous measurements for control of superconducting quantum circuits, Advances in Physics: X 5, 1813626 (2020).
[26] R. Mohseninia, J. Yang, I. Siddiqi, A. N. Jordan, and J. Dressel, Always-On Quantum Error Tracking with Continuous Parity Measurements, Quantum 4, 358 (2020).
[27] J. Atalaya, A. N. Korotkov, and K. B. Whaley, Error-correcting bacon-shor code with continuous measurement of noncommuting operators, Phys. Rev. A 102, 022415 (2020).
[28] Y.-H. Chen and T. A. Brun, Continuous quantum error detection and suppression with pairwise local interactions, Phys. Rev. Research 2, 043093 (2020).
[29] W. P. Livingston, M. S. Blok, E. Flurin, J. Dressel, A. N. Jordan, and I. Siddiqi, Experimental demonstration of continuous quantum error correction (2021), arXiv:2107.11398.
[30] K. Jacobs and D. A. Steck, A straightforward introduction to continuous quantum measurement, Contemporary Physics 47, 279 (2006).
[31] A. Chantasri, J. Dressel, and A. N. Jordan, Action principle for continuous quantum measurement, Physical Review A 88, 042110 (2013).
[32] E. Flurin, L. S. Martin, S. Hacohen-Gourgy, and I. Siddiqi, Using a recurrent neural network to reconstruct quantum dynamics of a superconducting qubit from physical observations, Phys. Rev. X 10, 011006 (2020).
[33] A. Graves, Supervised Sequence Labelling with Recurrent Neural Networks (Springer-Verlag Berlin Heidelberg, Berlin Heidelberg, 2012).
[34] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. 91, 045002 (2019).
[35] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, Recurrent neural network wave functions, Phys. Rev. Research 2, 023355 (2020).
[36] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, Reconstructing quantum states with generative models, Nature Machine Intelligence 1, 155 (2019).
E. Genois, J. A. Gross, A. Di Paolo, N. J. Steven- 
F. Yan, S. Gustavsson, J. Bylander, X. Jin, F. Yoshihara, 
D. Szombati, A. Gomez Frieiro, C. M¨ uller, T. Jones, 
H. M. Wiseman and G. J. Milburn, Open quantum sys-
S. Krastanov, K. Head-Marsden, S. Zhou, S. T. Flam-
A. Shabani, J. Roden, and K. B. Whaley, Efficient quantum filtering for quantum feedback control, Phys. Rev. A 91, 012118 (2015).
H. M. Wiseman and G. J. Milburn, Open quantum systems, in Quantum Measurement and Control (Cambridge University Press, 2009) p. 97–147.
D. Szombati, A. Gomez Freiéro, C. Müller, T. Jones, M. Jerger, and A. Fedorov, Quantum rilling: Protecting a qubit from measurement back action, Phys. Rev. Lett. 124, 070401 (2020).
T. M. Stace and S. D. Barrett, Continuous quantum measurement: Inelastic tunneling and lack of current oscillations, Phys. Rev. Lett. 92, 136802 (2004).
F. Yan, S. Gustavsson, J. Bylander, X. Jin, F. Yoshihara, D. G. Cory, Y. Nakamura, T. P. Orlando, and W. D. Oliver, Rotating-frame relaxation as a noise spectrum analyser of a superconducting qubit undergoing driven evolution, Nature Communications 4, 2337 (2013).
J. Gambetta, A. Blais, D. I. Schuster, A. Wallraff, L. Frunzio, J. Majer, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Qubit-photon interactions in a cavity: Measurement-induced dephasing and number splitting, Phys. Rev. A 74, 042318 (2006).
M. Naghiloo, A. N. Jordan, and K. W. Murch, Achieving optimal quantum acceleration of frequency estimation using adaptive coherent control, Phys. Rev. Lett. 119, 180801 (2017).
E. Genois, J. A. Gross, A. Di Paolo, N. J. Steven-son, G. Koolstra, A. Hashim, I. Siddiqi, and A. Blais, Quantum-tailored machine-learning characterization of a superconducting qubit, PRX Quantum 2, 040355 (2021).
A. Shabani, J. Roden, and K. B. Whaley, Continuous measurement of a non-Markovian open quantum system, Physical Review Letters 112, 23 (2014).
S. Krastanov, K. Head-Marsden, S. Zhou, S. T. Flamnia, L. Jiang, and P. Narang, Unboxing quantum black box models: Learning non-markovian dynamics (2020), arXiv:2009.03902.
L. Banchi, E. Grant, A. Rocchetto, and S. Severini, Modelling non-markovian quantum processes with recurrent neural networks, New Journal of Physics 20, 0 (2018).
T. Beucler, M. Priftiha¨r, S. Rasp, J. Ott, P. Baldi, and P. Gentile, Enforcing analytic constraints in neural networks emulating physical systems, Phys. Rev. Lett. 126, 098302 (2021).
A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, Wavenet: A generative model for raw audio (2016), arXiv:1609.03499.
J. Johansson, P. Nation, and F. Nori, QuTip 2: A python framework for the dynamics of open quantum systems, Computer Physics Communications 184, 1234 (2013).