Rapid High-Fidelity Single-Shot Dispersive Readout of Superconducting Qubits

T. Walter, P. Kurpiers, S. Gasparinetti, P. Magnard, A. Potočnik, Y. Salathé, M. Pechal,
M. Mondal, M. Oppliger, C. Eichler, and A. Wallraff

Department of Physics, ETH Zurich, CH-8093 Zurich, Switzerland

The speed of quantum gates and measurements is a decisive factor for the overall fidelity of quantum protocols when performed on physical qubits with a finite coherence time. Reducing the time required to distinguish qubit states with high fidelity is, therefore, a critical goal in quantum-information science. The state-of-the-art readout of superconducting qubits is based on the dispersive interaction with a readout resonator. Here, we bring this technique to its current limit and demonstrate how the careful design of system parameters leads to fast and high-fidelity measurements without affecting qubit coherence. We achieve this result by increasing the dispersive-interaction strength, by choosing an optimal linewidth of the readout resonator, by employing a Purcell filter, and by utilizing phase-sensitive parametric amplification. In our experiment, we measure 98.25% readout fidelity in only 48 ns, when minimizing readout time, and 99.2% in 88 ns, when maximizing the fidelity, limited predominantly by the qubit lifetime of 7.6 μs. The presented scheme is also expected to be suitable for integration into a multiplexed readout architecture.

I. INTRODUCTION

High-fidelity single-shot qubit readout is quintessential for real-time quantum feedback schemes used, for example, in error correction [1,2], teleportation [3,4], and state initialization [5,6]. It is also a key element in fundamental tests of quantum mechanics, such as loophole-free Bell tests [7–9]. Experimental progress in these areas may ultimately lead to fault-tolerant quantum computation, for which one of the most promising platforms is built on superconducting circuits and qubits. The standard technique to probe these qubits relies on the state-dependent dispersive frequency shift imposed by the qubit on a coupled resonator [10,11]. While averaging was required early on to determine the qubit state [11], advances in quantum-limited amplification [12–15] allowed for the observation of quantum jumps and the discrimination of qubit states in single-shot measurements [16,17]. In an attempt to further improve readout fidelities and measure multiple qubits simultaneously, Purcell filters [18–20] and broadband parametric amplifiers were developed [21–23]. With these advances, state discrimination is now predominantly limited by qubit decay during the time of measurement [22,24,25].

In order to overcome this limitation, one can reduce the measurement time, which we achieve in this work by increasing the dispersive interaction strength and by choosing an appropriate resonator linewidth. In addition, we use a Purcell filter to protect the qubit from radiative decay, and we efficiently detect the microwave fields with a quantum-limited phase-sensitive amplifier. In our readout experiments, we discriminate between the qubit ground and excited states with a fidelity in excess of 98% in less than 50 ns. Furthermore, we show that the measurement time

can be reduced by using a shaped "two-step" readout pulse that populates the resonator faster with microwave photons than a simple square pulse does [19,26].

II. DISCUSSION OF THE EXPERIMENT

Our sample device, depicted in Fig. 1(a), consists of a transmon qubit (purple) coupled capacitively with rate $g/2\pi = 208$ MHz to a readout resonator (green). The qubit, which is fabricated from shadow-evaporated aluminum, has a transition frequency of $\omega_q/2\pi = 6.316$ GHz, an anharmonicity of $\alpha/2\pi = -340$ MHz and a lifetime $T_1 = 7.6$ μs. The qubit state is controlled by applying microwave pulses through a capacitively coupled drive line (teal). The readout resonator, together with the remaining on-chip elements, is fabricated from a niobium thin film sputtered on a sapphire substrate using photolithography and reactive-ion etching. The readout resonator has a center frequency of $\omega_r/2\pi = 4.754$ GHz. The detuning $\Delta = \omega_q - \omega_r$ is much larger than g, which results in a dispersive coupling between qubit and resonator with rate $\chi/2\pi = -7.9$ MHz. The resonator is coupled through a Purcell-filter cavity [19,27] to an external measurement line [the orange structure in Fig. 1(a)]. The effective resonator linewidth $\kappa_{\mathrm{eff}} = 4Q_p J^2/(\omega_p + 4\delta_p^2 Q_p^2/\omega_p)$ is controlled by the coupling strength $J/2\pi = 25$ MHz between the readout resonator and the Purcell filter and the detuning $\delta_p/2\pi = 2$ MHz between the two. With the quality factor of the Purcell-filter cavity $Q_p = 74$, we find $\kappa_{\mathrm{eff}}/2\pi = 37.5$ MHz.

The sample is probed in a dilution refrigerator at a temperature of 10 mK using a standard microwave frequency measurement setup [Fig. 5(a)]. A Josephson parametric dimer (JPD) amplifier [28] operated in the phase-sensitive
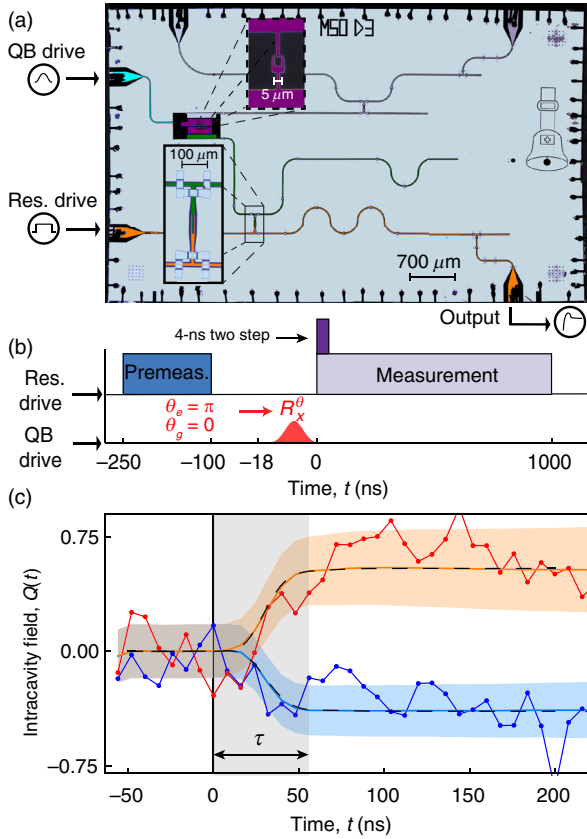
FIG. 1. (a) False-color micrograph of the experimental sample (see the text for details). (b) Pulse scheme including the preselection procedure, the gated (light purple) and the two-step measurement pulse (dark and light purple); see the text for details. (c) A characteristic ground (blue) and excited (red) single-shot trajectory plotted together with the mean of all ground and excited trajectories (light-blue and light-red solid lines, respectively) with their respective standard deviations (blue and orange shaded regions). The black dashed line shows the theoretically expected dynamics. The gray region depicts the typical integration time τ of the experiment. QB denotes qubit.

mode is used to optimize the measurement efficiency to achieve $\eta = 0.66$, with a 3-dB bandwidth of 27 MHz and a gain of 26 dB [13,15]; see Appendix C for details.

To read out the qubit, we apply a coherent tone at $t = 0$ to the input port of the Purcell filter [orange in Fig. 1(a)] at the center frequency $\omega_d = \omega_r$ between the two state-dependent resonator frequencies. The power is chosen to maximize the readout fidelity and corresponds to $n_{\text{drive}} = 2.5 \pm 0.25$ photons in the readout resonator, which we have calibrated based on an ac-Stark shift measurement; see Appendix A. At $t = -250$ ns, we apply a 150-ns-long measurement pulse to the device; see the pulse sequence in Fig. 1(b). We use the last 50 ns of this pulse to determine the state of the qubit and later analyze only those traces for which the qubit is found in the ground state. We refer to this procedure as preselection (Appendix B). Before preparing the qubit in either the ground or excited state by applying a 18-ns R_x^π

derivative-removal-by-adiabatic-gate (DRAG) pulse [29,30], which ends at $t = 0$, we wait for about 100 ns $\approx$ $25/\kappa_{\text{eff}}$ for the resonator to decay back to its vacuum state after preselection. In addition to the typical square ("gated") measurement pulse generated by gating the signal source, we also utilize a two-step pulse, which has an additional 4-ns high-power segment that drives the resonator into its steady state more rapidly [19,26]; see the pulse sequence in Fig. 1(b).

A field-programmable gate array (FPGA) with an analog-to-digital converter (ADC) samples the output signal in 8-ns time bins. The amplified quadrature of the JPD is chosen to maximize the contrast between the mean ground- and excited-state responses in a single quadrature $Q(t)$. As shown in Fig. 1(c), the ground- and excited-state responses can be clearly distinguished in a single shot of a measurement. With the qubit prepared in the excited state, due to qubit relaxation, the averaged response $\langle Q_e(t) \rangle$ (the solid orange line) slightly decays toward the ground-state response $\langle Q_g(t) \rangle$, resulting in the corresponding standard deviation $\langle \Delta Q_e(t) \rangle$ (the shaded orange region) to grow in time. These trends limit the distinguishability between the ground- and excited-state trajectory and highlight the need for fast readout.

The theoretical dynamics of the Purcell filter [27] (the dashed black lines), including the analog and digital filtering of the measurement line (see Sec. III and Appendix D), shows very good agreement with the averaged trajectories and allows us to calibrate the starting time of the measurement. We choose $t = 0$ to be the time at which the square readout pulse arrives at the input of the Purcell filter.

In order to optimize the distinguishability between the ground and excited states, given a certain integration time τ, we evaluate the integrated readout quadrature value $q_\tau = \sqrt{\kappa_p} \int_0^\tau dt Q(t) W(t)$, where the weighting function $W(t) \propto |\langle Q_e(t) - Q_g(t) \rangle|$ is proportional to the average difference between the ground- and excited-state responses and is normalized to $\int_0^\tau dt W^2(t) = 1$.

We perform 60 000 repetitions of the experiment, alternating the qubit preparation between the ground and excited states, and populate histograms of q_τ, as shown in Fig. 2. The ground- and excited-state histograms are simultaneously fit to the sum of two Gaussians, as described in Appendix B. From these fits, we extract a decision boundary, or threshold q_τ^{th}, at the intersection point of the two fitted state distributions. The measurement is then characterized by its fidelity, defined as [31]

$$F = 1 - P(e|g) - P(g|e), \tag{1}$$

where $P(x|y)$ is the probability that the qubit prepared in state y is measured to be in state x. The fidelity is extracted directly from the experimental data as 1 minus the sum of the fraction of ground-state preparation events with
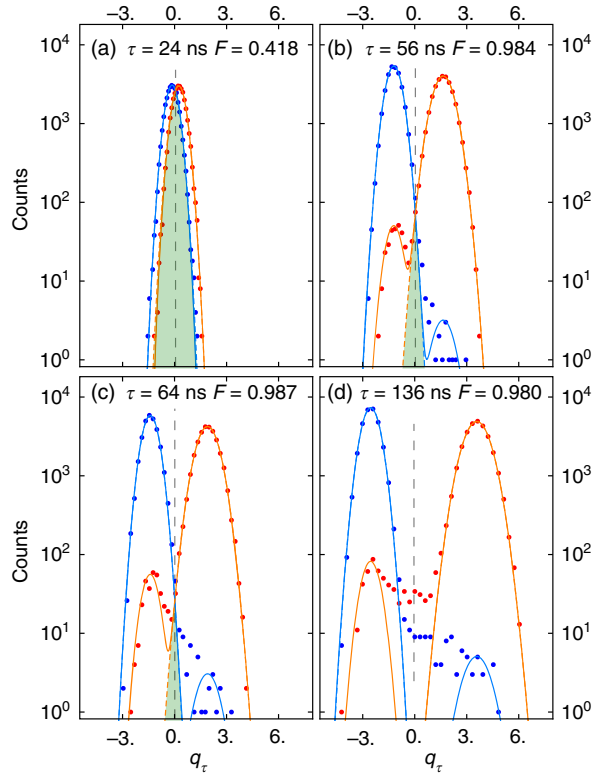
FIG. 2. (a)–(d) Histograms of the integrated quadrature amplitude q_τ at selected integration times between 24 and 136 ns after preparing the qubit in the ground (blue) or excited (red) state. The single shots are performed with a gated measurement pulse and optimal drive power. The solid lines are fits to a double-Gaussian model whose individual components are indicated by dashed lines. The green areas depict the overlap error. The dashed gray line indicates the qubit-state threshold.

$q_\tau \geq q_\tau^{\text{th}}$, and the fraction of excited-state preparation events with $q_\tau < q_\tau^{\text{th}}$, denoted as ϵ_g and ϵ_e, respectively.

At short measurement times where $\tau \lesssim 40$ ns, as shown in Fig. 2(a), the mean state trajectories have yet to reach their maximum separation, leaving a large overlap of the two distributions (the green area). However, at longer integration times, the separation grows sufficiently compared to the standard deviations, allowing us to clearly resolve the two states. With continued integration, the uncertainties further reduce, reaching a maximal distinguishability at $\tau = 64$ ns; see Fig. 2(c). For even longer integration times, transitions between the two states during the measurement, due to spontaneous emission, thermal excitation or transitions induced by the readout tone, limit the fidelity. These transitions lead to the discrepancies seen in Fig. 2(d) between the model and data and are discussed at the end of Appendix B in more detail.

We further analyze the data by separating the overlap error from transition errors. The overlap error ϵ_o between the two state distributions is defined as the normalized green area indicated in Fig. 2 and is a measure for how

distinguishable the two distributions are [19,31]. It can be further decomposed into contributions from the ground and excited states, $\epsilon_o = \epsilon_{o,g} + \epsilon_{o,e}$. The remaining transition errors $\tilde{\epsilon}_x = \epsilon_x - \epsilon_{o,x}$, where x is the state of the qubit, are due to failed preparation events and transitions during the measurement. The transition errors are found to be larger for a prepared excited state and increase with τ proportional to $1 - e^{-\tau/T_1}$ due to spontaneous emission. The overlap errors $\epsilon_{o,x}$ decrease with integration time; see Fig. 3.

For $\tau = 56$ ns, we obtain a fidelity of $98.42 \pm 0.07\%$ and an overlap error of $\epsilon_o = 0.39 \pm 0.01\%$, equal to that reported in Ref. [19] but achieved in roughly a third of the integration time. We also have an average assignment
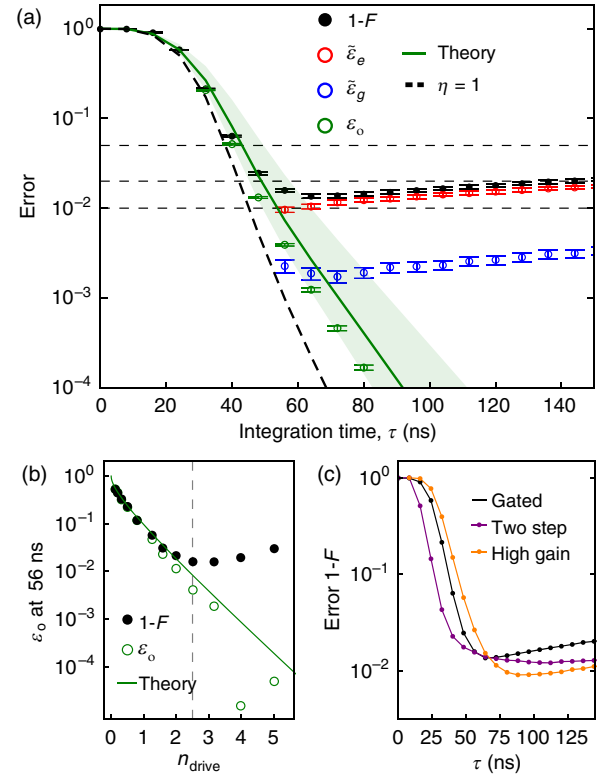

FIG. 3. (a) Errors ϵ of the gated single-shot experiment at optimal drive strength $n_{\text{drive}} = 2.5$ as a function of integration time τ. The solid black circles represent the infidelity extracted from the data directly. The open circles are excited-state (red), ground-state (blue), and overlap errors (green) extracted from the fits. The solid green line is the numerical solution of the full model for the overlap error [Eq. (D1)] with the independently measured system parameters, including the finite amplifier bandwidth, while the black dashed curve assumes $\eta = 1$. (b) Results of the overlap error ϵ_0 at 56 ns extracted from Gaussian fits (the green circles) with the gated single-shot experiments versus measurement drive power n_{drive}. The green solid line is the theoretical prediction using the full overlap model [Eq. (D1)], while the black dots are the measured infidelity $1 - F$. (c) Experimental results of infidelity versus integration time for the gated (black), two-step (purple), and high-amplifier-gain (orange) measurements.

fidelity of 99.2%, higher than that reported in Ref. [32] while using less than a fourth of the integration time. We also achieve this higher fidelity with a qubit lifetime of 70% and 30% of those in the two references [19,32], respectively. This result further emphasizes the effectiveness of fast, optimized readout. At this integration time, the normalized ground-state error is $\tilde{e}_g = 0.23 \pm 0.04\%$ and the normalized excited-state error is $\tilde{e}_e = 0.96 \pm 0.06\%$. The limited qubit lifetime $T_1 = 7.6\ \mu$s accounts for 76% of the excited-state error. We attribute the remaining part to measurement-induced mixing [33]. This hypothesis is supported by its equivalence to the ground-state error and the low thermal occupation of the qubit (see Appendix B). This mixing error is a consequence of driving the system harder in order to reduce the overlap faster and reach a higher fidelity in a shorter time. However, some errors may be introduced by imperfect state preparation and this estimate for the measurement-induced mixing is, therefore, an upper bound.

We also perform the gated measurements at different drive strengths and find that the overlap error decreases monotonically with the measurement power [Fig. 3(b)]. This behavior is in good agreement with theory [31] (Appendix D). However, the overall error of the readout protocol reaches a minimum at $n_{\text{drive}} = 2.5$ before rising again due to measurement-induced mixing.

To further improve the readout speed at the optimal power, we repeat the experiment with the two-step pulse shape introduced earlier [19,26], which more rapidly drives the resonator to its steady state. The readout fidelity during the rise time of the signal is consistently 8 ns faster than in the gated measurements, as shown in Fig. 3(c), and reaches a nearly identical fidelity of $F = 98.25 \pm 0.05\%$ in only 48 ns.

We also study the dependence of readout fidelity on the measurement efficiency η by increasing the gain of the JPD to 35 dB. The increased gain causes a reduction of the amplifier bandwidth to 10 MHz but increases the efficiency to $\eta = 0.75$ (see Appendix C). In this configuration, we measure the highest fidelity at 99.2%; however, the integration time of 88 ns required to achieve this fidelity is longer due to the reduced bandwidth [see the orange line in Fig. 3(c)].

III. MODEL USED FOR OPTIMIZING THE READOUT

A. Dynamics of readout circuit

For phase-sensitive amplification, the information acquired about the qubit state is proportional to the difference between the ground- and excited-state responses in one quadrature $S(t) = \sqrt{\kappa_p}|Q_e(t) - Q_g(t)|$, where κ_p is the linewidth of the Purcell filter [27]. This relation is approximately equal to $S(t) \approx \sqrt{\kappa_p}|\beta_e(t) - \beta_g(t)|$, where $\beta_{g,e}$ are the complex intracavity field amplitudes of the Purcell filter when the qubit is in the ground or excited state. In

the quasi-steady-state (QSS) limit given by $J \ll \kappa_p$, the full system dynamics can be reduced to the dynamics of the readout resonator, such that $S(t) \approx \sqrt{\kappa_{\text{eff}}}|\alpha_e(t) - \alpha_g(t)|$, where $\alpha_{e,g}$ are the complex intracavity field amplitudes of the readout resonator. By considering $\omega_d = \omega_r = \omega_p$, for a gated measurement pulse, we find an analytical solution for

$$S(t) \approx \sqrt{\frac{n_{\text{drive}}}{n_{\text{crit}}}\left|\frac{\alpha}{(1+\alpha/\Delta)}\right|\frac{1}{|\chi/\kappa_{\text{eff}}|[1+(2\chi/\kappa_{\text{eff}})^2]}}$$
$$\times \left(2\left|\frac{\chi}{\kappa_{\text{eff}}}\right| + e^{-\pi\kappa_{\text{eff}}t}\sin(2\pi\chi t) - 2e^{-\pi\kappa_{\text{eff}}t}\cos(2\pi\chi t)\right),$$
$$(2)$$

where χ is approximately given by

$$\chi = \frac{g^2}{\Delta}\frac{\alpha}{(\Delta+\alpha)} = \frac{\alpha}{4n_{\text{crit}}}\frac{1}{(1+\frac{\alpha}{\Delta})} \qquad (3)$$

for a transmon qubit when the population of the resonator is much less than the critical photon number, $n_{\text{crit}} = \Delta^2/4g^2$ [10].

The readout fidelity after a measurement time τ is directly related to the integrated information rate $s(\tau) = (1/\sqrt{\tau})\int S(t)dt$ [see Eq. (D1)]. In the following, we therefore study $s(\tau)$ and optimize it with respect to the accessible system parameters.

We first consider the influence of the effective linewidth κ_{eff} at otherwise constant parameters α, $n_{\text{drive}}/n_{\text{crit}}$, Δ, and χ. For integration times well into steady state, we find $s(\tau)$ to be maximized for $|\chi/\kappa_{\text{eff}}| = 0.5$, independent of all other parameters [31]; see Fig. 4(a). However, at integration times $\chi\tau \lesssim 4.5$, one could achieve a higher fidelity by increasing κ_{eff} with respect to χ. This optimization increases $S(t)$ at short times while reducing its steady-state value. We note that this initial speedup could also be achieved by using a two-step pulse while maintaining $\chi/\kappa = 0.5$, independent of the integration time.

When χ—and therefore κ_{eff}—becomes comparable to κ_p, one must solve the complete system dynamics beyond the QSS solution, including the dynamics of the Purcell filter [27]. We plot the full solution with the experimentally realized system parameters as the blue dashed line in Fig. 4(a), which shows a slower convergence to the steady-state ratio than the QSS solution.

For the optimal ratio $\chi/\kappa_{\text{eff}} = 0.5$, we also see from Fig. 4(a) that the time required to reach steady state is inversely proportional to χ. This relationship emphasizes that, in order to achieve fast readout, a large χ is favorable. Therefore, if one can maintain the optimal ratio, then a larger χ will more rapidly accumulate a signal. By successfully realizing these concepts in the experiment, we are able to achieve a higher single-shot fidelity at shorter integration times than previously reported, where it is less
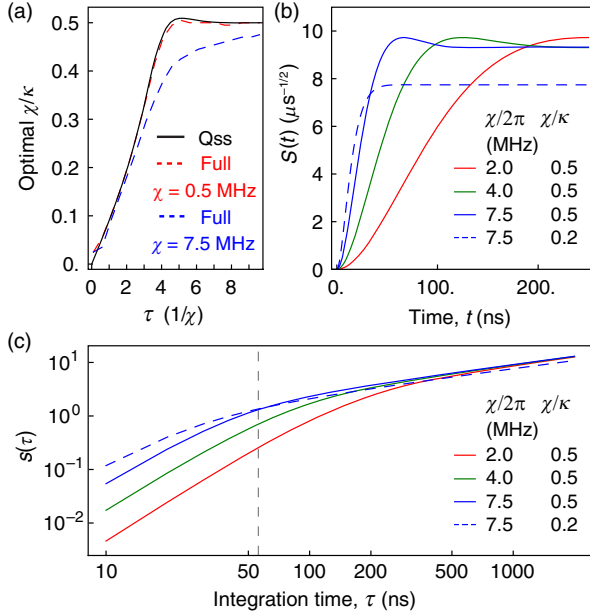
FIG. 4.   (a) Ratio χ/κ that maximizes $s(\tau)$ as a function of integration time τ. The solid black line represents the quasi-steady-state solution, while the dashed lines are the numerical solutions with our circuit parameters with the full model [27], which includes the dynamics of the Purcell filter. (b) Measurement signal $S(t)$ as a function of time t for the indicated values of χ. The solid lines indicate a ratio $\chi/\kappa = 0.5$, while the dashed line is for our experimentally used ratio of 0.2, chosen for being optimal for short integration times below 50 ns. (c) Integrated signal $s(\tau)$, as a function of integration time τ, for the indicated values of χ and χ/κ. The gray dashed line represents the typical integration time used in our experiment. For both $S(t)$ and $s(\tau)$, we use $\alpha/2\pi = -340$ MHz, $\Delta/2\pi = 1562$ MHz, and $n_{\text{drive}} = n_{\text{crit}}/5$.

affected by qubit relaxation time T_1. We choose the ratio $|\chi/\kappa_{\text{eff}}| = 0.2$ in our experiment to optimize for an integration time of 50 ns or less using the gated measurement pulse. This optimization choice is illustrated in Figs. 4(b) and 4(c), where the integrated signal is larger for $|\chi/\kappa_{\text{eff}}| = 0.2$ than for $|\chi/\kappa_{\text{eff}}| = 0.5$ for $\tau < 50$ ns.

B. Constraints

The optimal set of parameters are typically subject to experimental constraints of the system. With the use of Purcell filters [18–20], $|\chi/\kappa_{\text{eff}}| \approx 1/2$ can be maintained for larger values of χ, without reaching a Purcell limited qubit lifetime [34]. In this case, the limitation for readout speed derives directly from the upper bound of χ.

For larger χ, a large anharmonicity is advantageous; see Eq. (3). However, with the use of a transmon qubit, a large α reduces the qubit's coherence time through charge dispersion [35]. Charge dispersion scales exponentially with the ratio of the Josephson energy to the charging energy, E_J/E_C [34] and, therefore, with the qubit frequency and anharmonicity. For our sample, we choose an anharmonicity which results in an estimated upper bound for the

dephasing time of 150 μs with a qubit near 6.3 GHz. Our sample, however, does not realize this upper bound due to other dephasing mechanisms and fabrication limitations.

The second constraint on χ derives from its inverse relationship to the critical photon number $n_{\text{crit}} \propto 1/\chi$ [Eq. (3)]. The dispersive approximation remains valid only in the limit $n_{\text{crit}} \gg 1$, leading to an upper bound on χ. Furthermore, n_{crit} could limit the number of photons that can be used for readout without driving unwanted qubit transitions. The probability for such a transition in the dispersive regime is approximately proportional to the parameter [10,36]

$$\lambda = \sin^2\left[\frac{1}{2}\tan^{-1}\left(\sqrt{\frac{n_{\text{drive}} + 1}{n_{\text{crit}}}}\right)\right], \qquad (4)$$

implying that a larger $n_{\text{drive}}/n_{\text{crit}}$ could lead to additional transition errors. This constraint on the number of photons used for measurement is not strict. For a dispersive measurement, the optimal drive strength is determined by balancing the rising measurement-induced mixing errors and the declining overlap error (see Ref. [19,37] and Fig. 3). It is worth noting that alternative readout schemes have been investigated both theoretically and experimentally which achieve high-fidelity quantum-nondemolition measurements in a regime where $n_{\text{drive}} \gg n_{\text{crit}}$ (see Refs. [32,38]). As a suitable trade-off between these limitations, we choose $n_{\text{crit}} \approx 13$ and find $n_{\text{drive}} \approx n_{\text{crit}}/5$, resulting in $\lambda \sim 1.8\%$. In our experiment, this choice leads to a measurement-induced error $< 0.23\%$ at $\tau = 56$ ns.

Additionally, the mixing rates of the qubit states during measurement may also depend on higher-order transmon states [37,38], as well as the power spectrum of the flux noise $\Gamma_{\uparrow\downarrow} \propto (n_{\text{drive}}/n_{\text{crit}})N_f(\pm\Delta)$, where N_f is the flux-noise power at detuning $\pm\Delta$ [39]. Slichter *et al.* [40] measured a $N_f \propto 1/f$ flux-noise dependency and, therefore we design a large $g/2\pi \approx 210$ MHz to keep the detuning $\Delta/2\pi \approx 1.6$ GHz sufficiently large for our chosen n_{crit}. We further utilize a positive detuning Δ to exploit the asymmetry in χ with respect to Δ, which allows for a large dispersive shift, $|\chi/2\pi| > 7.5$ MHz, and remain within these constraints.

IV. CONCLUSIONS

In this work, we demonstrate an increase in speed and fidelity of qubit readout by optimizing the circuit parameters. Furthermore, we identify constraints on χ imposed by the nature of the transmon qubit and its dispersive interaction. The limits to the qubit lifetime and dephasing time imposed by our choice of circuit parameters are expected to be 600 and 150 μs, respectively, while other mechanisms not considered may impose shorter times. We therefore believe our design concepts are extensible to a multiplexed readout architecture [41].

While improvements in the qubit T_1 and in the amplifier performance are, at this point, well understood for further improving the readout, the role of measurement-induced mixing as a limiting factor remains an aspect worthy of further investigation.

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

APPENDIX A: SETUP DETAILS AND SAMPLE CHARACTERIZATION

In this section, we describe the measurement setup and discuss the basic sample characterization measurements and the ac-Stark shift calibration of the readout power.

A measurement signal applied to the sample exits the output port of the bandpass Purcell filter and passes through an isolator, a circulator, and a directional coupler before entering the JPD; see Fig. 5(a). The signal then passes through additional amplification stages before it is down-converted to 250 MHz and, finally, digitized at the ADC connected to the FPGA. The ADC samples the signal at a 1-GHz rate. Within the FPGA, the signal is digitally down-converted to dc and integrated for 8 ns in order to remove the down-converted low-frequency noise and local-oscillator (LO) leakage from the analog down-conversion process. The results of the FPGA calculations are then stored for the analysis discussed in the main text.

The implementation of a bandpass Purcell filter improves the lifetime of our qubit beyond the limit imposed by the readout-resonator bandwidth through the Purcell effect [42], which is shown in Fig. 5(b). The measured
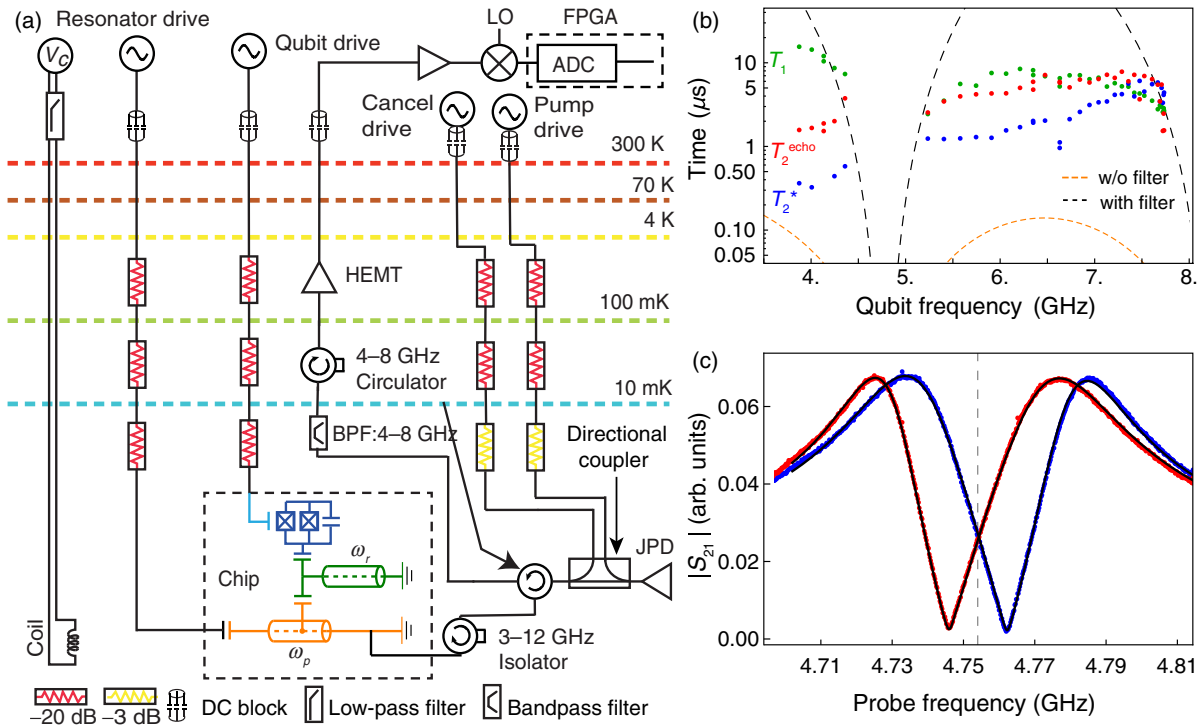

FIG. 5. (a) Schematic of the experimental setup. (b) Measured lifetime T_1, Ramsey decay time $T_2^\star$, and spin-echo time T_2^{echo} as a function of the qubit frequency. The black and orange dashed lines are the theoretical Purcell lifetimes with and without the Purcell filter. (c) Measured transmission of readout line with the qubit in its ground (blue) and excited (red) state. The solid black lines are fits to the input-output model of the circuit used to extract the relevant parameters of the system.

lifetime (the green dots) is between 30 and 100 times longer than the T_1 times predicted without the Purcell filter (the orange dashed line). However, it is often shorter than the limit imposed by the Purcell filter (the black dashed line), which we attribute to limitations in our fabrication techniques. The dephasing time of the qubit at its maximum resonant frequency is T_1 limited, which is itself Purcell limited and therefore, at this point, $T_2^* = 5~\mu s \approx 2T_1$.

We characterize the sample by performing resonator spectroscopy. We plot the transmission through the Purcell filter as a function of probe frequency for the ground (blue) and excited (red) qubit states in Fig. 5(c). From these transmission spectra, the values of ω_p, ω_r, J, χ, and $Q_p = \omega_p/\kappa_p$ are extracted from the fit parameters of the input-output model of the circuit given by

$$|S_{21}|_{\pm} \propto \frac{\kappa_p}{\frac{(\gamma+\kappa_p)}{2} + i(\omega_p - \omega) + \frac{2J^2}{\gamma + 2i(\omega_r \pm \chi - \omega)}}, \quad \text{(A1)}$$

where γ represents the small internal loss rates of the Purcell filter and the readout resonator, which are assumed to be equal. Furthermore, the qubit frequency and anharmonicity are determined by standard qubit spectroscopy [43].

We calibrate the photon number in the resonator n_{drive} by measuring the qubit frequency as a function of measurement drive power P_{rf}. Here, n_{drive} is assumed to be proportional to the power P_{out} transmitted through the device and measured at the FPGA; see Fig. 6(a). The proportionality factor between the output power and the intracavity photon number is extracted using the known χ. Because of its self-Kerr nonlinearity, the output power is proportional to the input power only at low readout powers [Fig. 6(b)].

APPENDIX B: PRESELECTION AND HISTOGRAM ANALYSIS

Preselection rejects instances in which the qubit is detected in the excited state from further analysis. Not doing so would lead to systematic errors in the experiment. Excited-state detection could be due to thermal excitation, or residual excitation of the qubit from previous measurement runs. In our experiment, we integrate the last 50 ns of the premeasurement pulse, resulting in a mean integrated measurement value q_p for determining the state of the qubit. The histogram of these integrated values, Fig. 6(c), is fit to a Gaussian distribution (the light-blue line), and the state discrimination threshold (the dashed vertical line) is set to 99% of the fitted cumulative distribution function. This threshold guarantees that the majority of thermally excited states are rejected.

We find that the fraction of rejected events increases with resonator drive strength n_{drive}; see Fig. 6(d). This increase is due to measurement-induced mixing, which is expected to vanish with decreasing drive strength. For $n_{\text{drive}} \to 0$, we
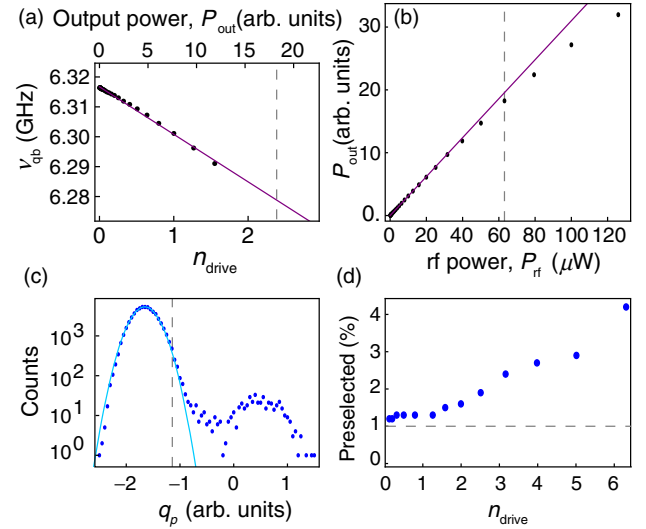

FIG. 6. (a) Qubit frequency ν_{qb} (the black dots) as a function of measured power P_{out} with a linear fit (the purple line). The calibrated photon number is shown on the bottom axis and the gray dashed line indicates the optimal drive power used in the experiment. (b) Measured output power as a function of input power P_{rf}. The purple line is a linear fit to the low-power data, while the gray dashed line indicates the optimal drive power. (c) Mean value q_p of the last 50 ns of the preselection pulse (the blue points) for $n_{\text{drive}} = 2.5$ with a Gaussian fit (the blue line). The gray dashed line is the threshold used to remove thermally excited states from further analysis set at the 99% cumulative distribution function of the Gaussian fit. (d) Percentage of removed data after preselection analysis as a function of drive power. The gray dashed line indicates the minimum percent of traces which could be removed due to our choice of the preselection-state discrimination threshold.

find an excess of $0.3 \pm 0.1\%$ of events rejected beyond the set threshold of 1%, which we attribute to the residual thermal population of the qubit; see Fig. 6(d). This low thermal population is a byproduct of the Purcell filter, which, due to its limited bandwidth and detuning from the qubit, further reduces the effective thermal photon flux in the readout resonator at the qubit frequency. This is an additional benefit of using a Purcell filter to improve the fidelity of single-shot readout that was not discussed in Ref. [19].

For the analysis of the measurement errors discussed in the text, the model used to fit the single-shot histograms is given by

$$\begin{bmatrix} C_g(q_\tau) \\ C_e(q_\tau) \end{bmatrix} = \begin{bmatrix} A_{gg} & A_{eg} \\ A_{ge} & A_{ee} \end{bmatrix} \begin{bmatrix} \text{PDF}[\mathcal{N}(\mu_g, \sigma_g), q_\tau] \\ \text{PDF}[\mathcal{N}(\mu_e, \sigma_e), q_\tau] \end{bmatrix}$$

where $\text{PDF}[\mathcal{N}(\mu_x, \sigma_x)]$ is a normal probability distribution function with mean and standard deviations, and μ_x, σ_x, and C_x are the expected counts of traces prepared in qubit state $x \in \{g, e\}$ as a function of the filtered quadrature value q_τ.

A_{if} represents the amplitudes with the qubit prepared in the desired state $i \in \{g, e\}$ and detected in the final state $f \in \{g, e\}$. This double-Gaussian model does not account for transitions that occur during the measurement, but instead only for ones that occur between the state preparation and the beginning of the measurement. We also study a model which does account for transitions during the measurement, and it produces an overlap error equal to our double-Gaussian model, within the error bounds.

APPENDIX C: CALIBRATION OF MEASUREMENT EFFICIENCY

We perform power-spectral-density measurements of the noise added by the amplifiers in the output line to calibrate the measurement efficiency $\eta = \eta_{\phi\mathrm{amp}}\eta_{\mathrm{loss}}$ of our detection chain characterized predominantly by the parametric amplifier's noise performance, $\eta_{\phi\mathrm{amp}}$, and the loss before this amplifier, η_{loss}. We achieve parametric gain by pumping the amplifier [28] with a coherent tone at the readout frequency. The maximum gain is adjusted by varying the pump power and the flux bias through the superconducting quantum-interference device (SQUID) arrays. The pump tone is interferometrically canceled by adjusting the phase and the amplitude of a cancellation tone coupled via the directional coupler to the output of the JPD. This interference minimizes pump leakage back toward the sample and avoids saturation of the following amplifiers along the output line. The frequency-dependent gain peaks at $G_0 = 19.7$ dB [see Figure 7(a)] at the desired signal frequency of 4.754 GHz used in the experiment. By sweeping the signal power, we determine the 1-dB compression point of the amplifier [Fig. 7(c)], which is about 10 dB higher than the power used for measuring the qubit state in our experiment.

Next, we characterize the amplifier's measurement efficiency when operating in the phase-sensitive mode $\eta_{\phi\mathrm{amp}}$, i.e., amplifying and detecting only one quadrature of the signal field. In order to create a single quadrature signal distinguishable from the pump, we apply two coherent signals of equal power and opposite detuning from the pump tone. We then adjust the phase of the two tones so that the resulting signal is oriented along the amplified quadrature, achieving a phase-sensitive gain, $4G_0$. For this
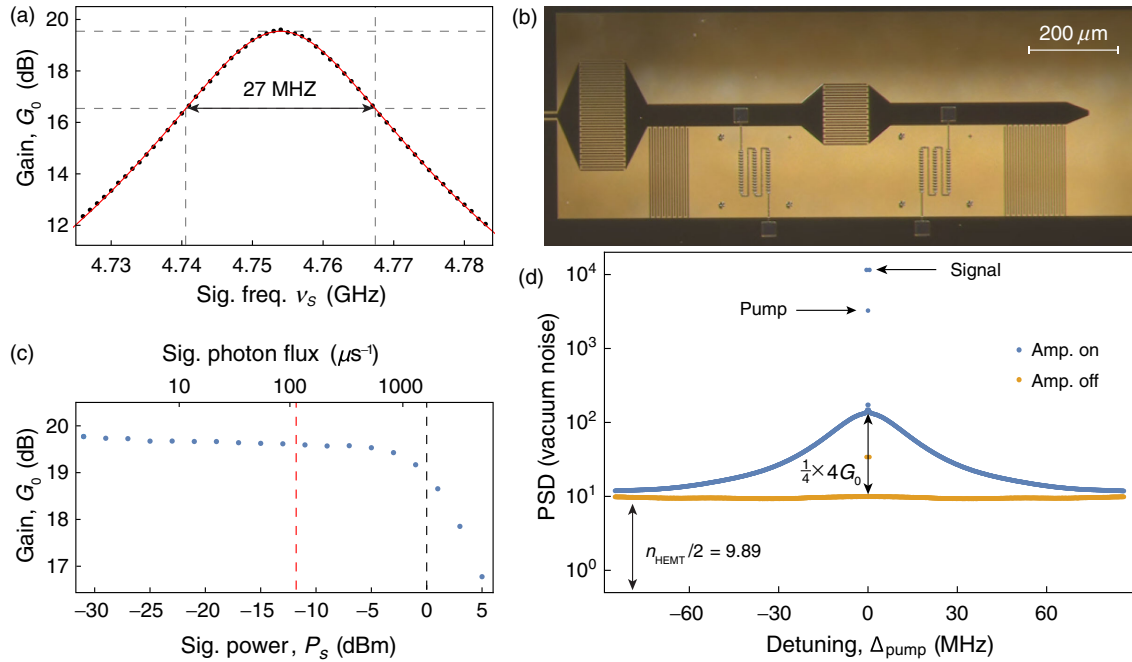

FIG. 7. (a) Measurement of the gain G_0 of the amplifier versus the signal frequency ν_s. The red line is a Lorentzian fit to extract the peak gain of approximately 19.7 dB and a 3-dB bandwidth of 27 MHz. (b) Micrograph of the Josephson parametric dimer used in the experiment made from etched niobium (black) on sapphire (yellow). The circuit consists of two capacitively coupled lumped-element nonlinear resonators, each with a large finger capacitor and an array of 45 SQUIDs (shadow-evaporated aluminum, gray in the picture) acting as a nonlinear inductor element. The left resonator is capacitively coupled to the output line of the measurement setup. (c) Measurement of the gain G_0 versus signal power P_s to extract the 1-dB compression point of the amplifier at the experimentally used bias point. The red dashed line is the optimal readout power used in the experiment, while the gray dashed line depicts the 1-dB compression power. (d) Power spectral density (PSD) of the amplified quadrature versus signal frequency detuning from the pump tone with the parametric amplifier on (blue) and off (yellow) used to calibrate the measurement efficiency of the output line. The y axis is scaled such that the distance between the peak of the parametric amplifier-on spectrum (blue) and the amplifier-off noise spectrum (yellow) is equal to the vacuum noise times the gain, i.e., $1/4 \times 4G_0$.

measurement, the ADC samples the output signal for 8.192 μs with 1-ns resolution, resulting in a frequency resolution of approximately 112 kHz. We therefore choose to detune the signals 448 kHz above and below the pump, which is an integer multiple of this resolution. Finally, we measure the noise performance of the amplified quadrature by measuring its power spectral density (PSD). The down-conversion local-oscillator phase is also adjusted to maximize the gain along the real quadrature of the complex signal entering the ADC, as is done for the main experiment. We use the FPGA to take the real component of the complex signal and compute the autocorrelation function and average the results over 10^6 times. The Fourier transform of these data then reveals the noise spectral density of this quadrature, as seen in Fig. 7(d). When the JPD is on, the overall noise level increases compared to when it is off following the frequency-dependent gain of the amplifier. The two coherent signal peaks can be easily distinguished from the noise floor in both the amplifier-on and amplifier-off spectra, as well as the pump tone for the amplifier-on case.

With the assumption that the noise of the parametric amplifier is dominated by amplified vacuum noise, the spectral-density measurements are scaled so that the noise increase when turning on the JPD is equal to $4G_0 \times 1/4$. In this expression, $1/4$ represents the noise at the input of the parametric amplifier in the amplified quadrature and $4G_0$ is the gain of this quadrature. The resulting noise offset, $n_{\text{HEMT}}/2$, is 9.89 times larger than the vacuum level. This offset is predominately due to the noise added by the following HEMT amplifier at the 4-K temperature stage and is divided by 2 to account for the single quadrature measurement. Following Frii's formula for noise performance of cascaded amplifier chains [44], this added noise is divided by the gain of the parametric amplifier, resulting in $\eta_{\phi\text{amp}} = [1 + n_{\text{HEMT}}/(2G_0)]^{-1}$. We estimate that, with a gain of $4G_0 = 26$ dB, the noise performance is about $\eta_{\phi\text{amp}} = 0.92$. If we adjust the pump settings to increase the gain to $4G_0 = 35$ dB, at the cost of bandwidth, we find the amplifier can achieve $\eta_{\phi\text{amp}} = 0.99$.

Finally, we determine the transmission efficiency η_{loss} between the output of the sample and the input of the parametric amplifier by comparing the coherent signal power in the scaled PSD measurement to the expected power calibrated with the ac-Stark shift measurement, i.e., $P_{\text{out}} = (\chi/J)^2 \kappa_p n_{\text{drive}}$. The $(\chi/J)^2$ prefactor is a result of driving the circuit through the input of the Purcell filter and not the readout resonator directly [27]. The ratio between the scaled-PSD-measurement signal power, and P_{out} is equal to the transmission efficiency and found to be $\eta_{\text{loss}} \approx 0.75$, corresponding to approximately 1.3-dB attenuation. This estimated attenuation is well within the specified insertion loss of the elements between the sample and the amplifier. The measured total efficiency of the line is determined to be $\eta = \eta_{\text{loss}}\eta_{\phi\text{amp}} = 0.66$ for 26 dB gain and $\eta = 0.75$ for 35 dB gain.

APPENDIX D: OVERLAP ERROR MODEL

We model the readout as a measurement of a random variable from the qubit-state Gaussian distributions, which have opposite means, $\pm S(t)/2$ (see Sec. III), and equal variance, $\sigma^2(t) = 1/(4\eta)$, at time t. With no state transitions, the optimal threshold is at the intersection of the two distributions and, including both analog and digital filtering, we find

$$\epsilon_o(\tau) = \text{erfc}\left(\sqrt{\frac{1}{8} \frac{\int_0^\tau S(t) f(\tau - t) dt}{\sqrt{\int_0^\tau \sigma^2(t) f(\tau - t) dt}}}\right), \quad \text{(D1)}$$

where $f(\tau - t)$ is the response function of the relevant filter elements for the measurement of s_τ. In our experiment, $f(\tau - t)$ consists of the narrow-bandwidth parametric amplifier, the 8-ns boxcar integration of the FPGA and the mode-matched digital weighting function $W(t)$. We use this model, along with the numerical solutions of $S(t)$ using the full system equations (with no quasi-steady-state assumption [27]), to determine the theory curves (the green areas) in Figs. 3(a) and 3(b).

[1] David P. DiVincenzo, Fault-tolerant architectures for superconducting qubits, Phys. Scr. **T137**, 014020 (2009).

[2] R. Barends *et al.*, Superconducting quantum circuits at the surface code threshold for fault tolerance, Nature (London) **508**, 500 (2014).

[3] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters, Teleporting an Unknown Quantum State via Dual Classical and Einstein-Podolsky-Rosen Channels, Phys. Rev. Lett. **70**, 1895 (1993).

[4] L. Steffen, Y. Salathe, M. Oppliger, P. Kurpiers, M. Baur, C. Lang, C. Eichler, G. Puebla-Hellmann, A. Fedorov, and A. Wallraff, Deterministic quantum teleportation with feedforward in a solid state system, Nature (London) **500**, 319 (2013).

[5] J. E. Johnson, C. Macklin, D. H. Slichter, R. Vijay, E. B. Weingarten, John Clarke, and I. Siddiqi, Heralded State Preparation in a Superconducting Qubit, Phys. Rev. Lett. **109**, 050506 (2012).

[6] D. Ristè, J. G. van Leeuwen, H.-S. Ku, K. W. Lehnert, and L. DiCarlo, Initialization by Measurement of a Superconducting Quantum Bit Circuit, Phys. Rev. Lett. **109**, 050507 (2012).

[7] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt, Proposed Experiment to Test Local Hidden-Variable Theories, Phys. Rev. Lett. **23**, 880 (1969).

[8] Markus Ansmann, H. Wang, Radoslaw C. Bialczak, Max Hofheinz, Erik Lucero, M. Neeley, A. D. O'Connell, D. Sank, M. Weides, J. Wenner, A. N. Cleland, and John M. Martinis, Violation of Bell's inequality in Josephson phase qubits, Nature (London) **461**, 504 (2009).

[9] B. Hensen, H. Bernien, A. E. Dreau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellan, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson, Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres, Nature (London) **526**, 682 (2015).

[10] A. Blais, R.-S. Huang, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, Cavity quantum electrodynamics for superconducting electrical circuits: An architecture for quantum computation, Phys. Rev. A **69**, 062320 (2004).

[11] A. Wallraff, D. I. Schuster, A. Blais, L. Frunzio, J. Majer, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Approaching Unit Visibility for Control of a Superconducting Qubit with Dispersive Readout, Phys. Rev. Lett. **95**, 060501 (2005).

[12] C. M. Caves, Quantum limits on noise in linear amplifiers, Phys. Rev. D **26**, 1817 (1982).

[13] B. Yurke, M. L. Roukes, R. Movshovich, and A. N. Pargellis, A low-noise series-array Josephson junction parametric amplifier, Appl. Phys. Lett. **69**, 3078 (1996).

[14] M. A. Castellanos-Beltran, K. D. Irwin, G. C. Hilton, L. R. Vale, and K. W. Lehnert, Amplification and squeezing of quantum noise with a tunable Josephson metamaterial, Nat. Phys. **4**, 929 (2008).

[15] A. A. Clerk, M. H. Devoret, S. M. Girvin, Florian Marquardt, and R. J. Schoelkopf, Introduction to quantum noise, measurement, and amplification, Rev. Mod. Phys. **82**, 1155 (2010).

[16] François Mallet, Florian R. Ong, Agustin Palacios-Laloy, François Nguyen, Patrice Bertet, Denis Vion, and Daniel Esteve, Single-shot qubit readout in circuit quantum electrodynamics, Nat. Phys. **5**, 791 (2009).

[17] R. Vijay, D. H. Slichter, and I. Siddiqi, Observation of Quantum Jumps in a Superconducting Artificial Atom, Phys. Rev. Lett. **106**, 110502 (2011).

[18] M. D. Reed, B. R. Johnson, A. A. Houck, L. DiCarlo, J. M. Chow, D. I. Schuster, L. Frunzio, and R. J. Schoelkopf, Fast reset and suppressing spontaneous emission of a superconducting qubit, Appl. Phys. Lett. **96**, 203110 (2010).

[19] Evan Jeffrey, Daniel Sank, J. Y. Mutus, T. C. White, J. Kelly, R. Barends, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. Megrant, P. J. J. O'Malley, C. Neill, P. Roushan, A. Vainsencher, J. Wenner, A. N. Cleland, and John M. Martinis, Fast Accurate State Measurement with Superconducting Qubits, Phys. Rev. Lett. **112**, 190504 (2014).

[20] Nicholas T. Bronn, Yanbing Liu, Jared B. Hertzberg, Antonio D. Córcoles, Andrew A. Houck, Jay M. Gambetta, and Jerry M. Chow, Broadband filters for abatement of spontaneous emission in circuit quantum electrodynamics, Appl. Phys. Lett. **107**, 172601 (2015).

[21] J. Y. Mutus, T. C. White, R. Barends, Yu Chen, Z. Chen, B. Chiaro, A. Dunsworth, E. Jeffrey, J. Kelly, A. Megrant, C. Neill, P. J. J. O'Malley, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, K. M. Sundqvist, A. N. Cleland, and John M. Martinis, Strong environmental coupling in a Josephson parametric amplifier, Appl. Phys. Lett. **104**, 263513 (2014).

[22] C. Macklin, K. O'Brien, D. Hover, M. E. Schwartz, V. Bolkhovsky, X. Zhang, W. D. Oliver, and I. Siddiqi, A near-quantum-limited Josephson traveling-wave parametric amplifier, Science **350**, 307 (2015).

[23] Tanay Roy, Suman Kundu, Madhavi Chand, A. M. Vadiraj, A. Ranadive, N. Nehra, Meghan P. Patankar, J. Aumentado, A. A. Clerk, and R. Vijay, Broadband parametric amplification with impedance engineering: Beyond the gain-bandwidth product, Appl. Phys. Lett. **107**, 262601 (2015).

[24] Philip Krantz, Andreas Bengtsson, Michael Simoen, Simon Gustavsson, Vitaly Shumeiko, W. D. Oliver, C. M. Wilson, Per Delsing, and Jonas Bylander, Single-shot read-out of a superconducting qubit using a Josephson parametric oscillator, Nat. Commun. **7**, 11417 (2016).

[25] Yanbing Liu, Srikanth J. Srinivasan, D. Hover, Shaojiang Zhu, R. McDermott, and A. A. Houck, High fidelity readout of a transmon qubit using a superconducting low-inductance undulatory galvanometer microwave amplifier, New J. Phys. **16**, 113008 (2014).

[26] D. T. McClure, Hanhee Paik, L. S. Bishop, M. Steffen, Jerry M. Chow, and Jay M. Gambetta, Rapid Driven Reset of a Qubit Readout Resonator, Phys. Rev. Applied **5**, 011001 (2016).

[27] Eyob A. Sete, John M. Martinis, and Alexander N. Korotkov, Quantum theory of a bandpass Purcell filter for qubit readout, Phys. Rev. A **92**, 012325 (2015).

[28] C. Eichler, Y. Salathe, J. Mlynek, S. Schmidt, and A. Wallraff, Quantum-Limited Amplification and Entanglement in Coupled Nonlinear Resonators, Phys. Rev. Lett. **113**, 110502 (2014).

[29] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, Simple Pulses for Elimination of Leakage in Weakly Nonlinear Qubits, Phys. Rev. Lett. **103**, 110501 (2009).

[30] J. M. Gambetta, F. Motzoi, S. T. Merkel, and F. K. Wilhelm, Analytic control methods for high-fidelity unitary operations in a weakly nonlinear oscillator, Phys. Rev. A **83**, 012308 (2011).

[31] J. Gambetta, W. A. Braff, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, Protocols for optimal readout of qubits using a continuous quantum nondemolition measurement, Phys. Rev. A **76**, 012325 (2007).

[32] C. C. Bultink, M. A. Rol, T. E. O'Brien, X. Fu, B. C. S. Dikken, C. Dickel, R. F. L. Vermeulen, J. C. de Sterke, A. Bruno, R. N. Schouten, and L. DiCarlo, Active Resonator Reset in the Nonlinear Dispersive Regime of Circuit QED, Phys. Rev. Applied **6**, 034008 (2016).

[33] 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).

[34] Jens Koch, Terri M. Yu, Jay Gambetta, A. A. Houck, D. I. Schuster, J. Majer, Alexandre Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Charge-insensitive qubit design derived from the Cooper pair box, Phys. Rev. A **76**, 042319 (2007).

[35] J. A. Schreier, A. A. Houck, Jens Koch, D. I. Schuster, B. R. Johnson, J. M. Chow, J. M. Gambetta, J. Majer, L. Frunzio, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Suppressing charge noise decoherence in superconducting charge qubits, Phys. Rev. B **77**, 180502 (2008).

[36] Mostafa Khezri, Eric Mlinar, Justin Dressel, and Alexander N. Korotkov, Measuring a transmon qubit in circuit QED: Dressed squeezed states, Phys. Rev. A **94,** 012347 (2016).

[37] Daniel Sank, Zijun Chen, Mostafa Khezri, J. Kelly, R. Barends, B. Campbell, Y. Chen, B. Chiaro, A. Dunsworth, A. Fowler *et al.*, Measurement-Induced State Transitions in a Superconducting Qubit: Beyond the Rotating Wave Approximation, Phys. Rev. Lett. **117,** 190503 (2016).

[38] Maxime Boissonneault, J. M. Gambetta, and Alexandre Blais, Improved Superconducting Qubit Readout by Qubit-Induced Nonlinearities, Phys. Rev. Lett. **105,** 100504 (2010).

[39] Maxime Boissonneault, J. M. Gambetta, and Alexandre Blais, Dispersive regime of circuit QED: Photon-dependent qubit dephasing and relaxation rates, Phys. Rev. A **79,** 013819 (2009).

[40] D. H. Slichter, R. Vijay, S. J. Weber, S. Boutin, M. Boissonneault, J. M. Gambetta, A. Blais, and I. Siddiqi, Measurement-Induced Qubit State Mixing in Circuit QED from Up-converted Dephasing Noise, Phys. Rev. Lett. **109,** 153601 (2012).

[41] Y. Chen, D. Sank, P. O'Malley, T. White, R. Barends, B. Chiaro, J. Kelly, E. Lucero, M. Mariantoni, A. Megrant, C. Neill, A. Vainsencher, J. Wenner, Y. Yin, A. N. Cleland, and J. M. Martinis, Multiplexed dispersive readout of superconducting phase qubits, Appl. Phys. Lett. **101,** 182601 (2012).

[42] A. A. Houck, J. A. Schreier, B. R. Johnson, J. M. Chow, Jens Koch, J. M. Gambetta, D. I. Schuster, L. Frunzio, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Controlling the Spontaneous Emission of a Superconducting Transmon Qubit, Phys. Rev. Lett. **101,** 080502 (2008).

[43] D. I. Schuster, A. Wallraff, A. Blais, L. Frunzio, R.-S. Huang, J. Majer, S. M. Girvin, and R. J. Schoelkopf, ac Stark Shift and Dephasing of a Superconducting Qubit Strongly Coupled to a Cavity Field, Phys. Rev. Lett. **94,** 123602 (2005).

[44] D. M. Pozar, *Microwave Engineering* (John Wiley & Sons, Hoboken, NJ, 2012).