ARTICLE OPEN

Check for updates

Transmon qubit readout fidelity at the threshold for quantum error correction without a quantum-limited amplifier

Liangyu Chen [1]✉, Hang-Xi Li [1], Yong Lu[2], Christopher W. Warren [1], Christian J. Križan [1], Sandoko Kosen [1], Marcus Rommel[1], Shahnawaz Ahmed[1], Amr Osman [1], Janka Biznárová[1], Anita Fadavi Roudsari [1], Benjamin Lienhard [3,4], Marco Caputo[5], Kestutis Grigoras[5], Leif Grönberg[5], Joonas Govenius[5], Anton Frisk Kockum[1], Per Delsing [1], Jonas Bylander [1]✉ and Giovanna Tancredi [1]✉

High-fidelity and rapid readout of a qubit state is key to quantum computing and communication, and it is a prerequisite for quantum error correction. We present a readout scheme for superconducting qubits that combines two microwave techniques: applying a shelving technique to the qubit that reduces the contribution of decay error during readout, and a two-tone excitation of the readout resonator to distinguish among qubit populations in higher energy levels. Using a machine-learning algorithm to post-process the two-tone measurement results further improves the qubit-state assignment fidelity. We perform single-shot frequency-multiplexed qubit readout, with a 140 ns readout time, and demonstrate 99.5% assignment fidelity for two-state readout and 96.9% for three-state readout–without using a quantum-limited amplifier.

INTRODUCTION

With the recent demonstrations of quantum error correction[1–4], superconducting circuits are one of the leading platforms toward the realization of a fault-tolerant quantum computer[5–7]. However, despite the remarkable progress, achieving fast and high-fidelity single-shot readout of the qubits' states remains a challenge. As a comparison, while the two-qubit-gate fidelities are approaching the 0.1% error threshold[8–11], readout errors are typically at the 1% level for two-state readout[12–15]. Similarly, the implementation of high-fidelity single- and two-qubit gates takes between 10 and 100 ns[10,11,16], while a readout measurement can take from hundreds of nanoseconds to a few microseconds[12–16]. Notably, readout performances with fidelity >99% in less than 100 ns readout time have been achieved in refs. [12,15] Further improvement in the readout of superconducting qubits is therefore crucial to reliably cross the threshold of efficient error correction, which is estimated to be less than 0.5% for the break-even point[7]. Moreover, having a fast and high-fidelity measurement scheme can boost the repetition rate for both quantum-computing and quantum-communication applications[17–21] and is essential for achieving fast reset protocols[15,22].

In superconducting circuits, the state of a qubit is generally read out by detecting the dispersive frequency shift of a resonator coupled to the qubit[23]. The predominant source of error is the relaxation of the qubit from the excited ($|1\rangle$) to the ground ($|0\rangle$) state during the readout. On short time scales, this error grows almost linearly with the ratio between the readout time τ_r and the qubit relaxation time T_1[12], and can be mitigated by reducing τ_r. Various high-power readout schemes have been exploited to decrease the measurement time[24,25]. Furthermore, Purcell filters[26–28] and quantum-limited or near-quantum-limited amplifiers[29–32] have

been implemented and, with the combination of both, a readout fidelity exceeding 99% within 100 ns has been demonstrated [12,13].

Here, we report the implementation of an improved readout scheme that boosts the state-assignment fidelity and reduces the contribution of decay error during readout. Our readout strategy exploits the higher energy levels of the qubit[14,26,33–36] and introduces a two-tone probing of the resonator to enhance the readout fidelity of multiple states. We demonstrate single-shot readout fidelity up to 99.5% (96.9%) for two-state (three-state) discrimination within 140 ns without using a quantum-limited amplifier. The techniques we present here offer significant protection against decay during readout, are straightforward to implement and can be readily integrated into state-of-the-art quantum-computing devices.

The circuit schematic is shown in Fig. 1. The device, whose fabrication is described in ref. [37] consists of three fixed-frequency transmon qubits[38] with transition frequencies $\omega_{q_i}/2\pi$ at 5.36, 5.40, and 5.46 GHz for $i = 1, 2, 3$, respectively. Note that there is no direct coupling element between the qubits. Each qubit is coupled with a strength g_i to a readout resonator of frequency $\omega_{r_i}/2\pi = 6.45$, 6.61, and 6.74 GHz. The three resonators are coupled with a strength J_i to a common Purcell filter that is embedded in the readout feedline[27]. The Purcell-to-resonator coupling rates $J_i/2\pi$ are designed to be at 60 MHz, while the qubit-resonator coupling rates $g_i/2\pi$ are much larger, about 250 MHz. The Purcell filter is centered at $\omega_f/2\pi = 6.726$ GHz with a linewidth of $\kappa_f/2\pi = 820.9$ MHz. The theoretical analysis and design details are discussed in Section I of the Supplementary Note. The device is cooled down to 10 mK and a microwave setup is used to measure the signal transmitted through the feedline. The complete experimental setup and device parameters resulting from basic characterization are reported in Section II of the Supplementary Note.

[1]Department of Microtechnology and Nanoscience, Chalmers University of Technology, 41296 Gothenburg, Sweden. [2]3rd Institute of Physics, University of Stuttgart, 70569 Stuttgart, Germany. [3]Department of Chemistry, Princeton University, Princeton, NJ 08544, USA. [4]Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA. [5]VTT Technical Research Centre of Finland, FI-02044 VTT, Finland. ✉email: liangyuc@chalmers.se; jonas.bylander@chalmers.se; tancredi@chalmers.se
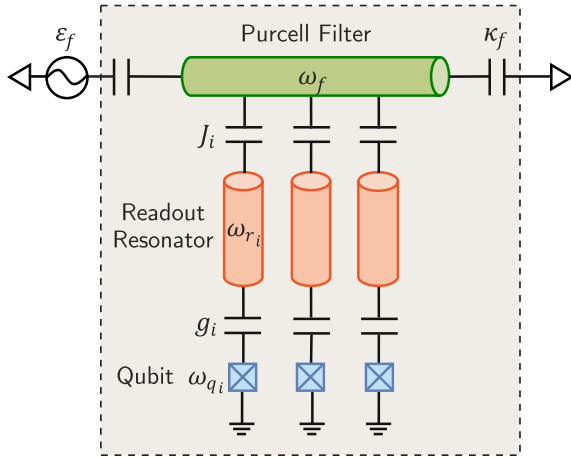
Fig. 1 Circuit schematic of the common-mode Purcell filter. The Purcell filter is a $\lambda/2$ coplanar waveguide resonator centered at ω_f, defined by a capacitor on each side. The filter is embedded in the readout feedline, and driven by the field ε_f. The output capacitance, represented by the Purcell-filter linewidth κ_f, is around an order of magnitude larger than the input capacitance such that the signal is guided towards the output port to measure transmission. Multiple resonators of resonant frequency ω_{r_i} couple to the Purcell filter with strength J_i within the filter bandwidth. The individual resonators are capacitively coupled with strength g_i to the qubits with transition frequency ω_{q_i}.

RESULTS

Exploiting higher energy levels

Typically, the qubit-excited-state decay is a major error source during readout, whose minimization requires performing the measurements in the shortest possible time[12,13]. To further mitigate this error, we first implement a shelving scheme that exploits the higher energy levels[14,26,33–36,39,40]. The pulse scheme is shown in Fig. 2a; a π_{12} and a π_{23} pulse are applied consecutively prior to the readout pulse so that the qubit population originally in the $|1\rangle$ state is transferred to the $|3\rangle$ state before readout. Thus, the qubit population that was in $|1\rangle$ will take a longer time to decay to $|0\rangle$ as the main relaxation channel is through cascading single-photon emission down the energy ladder, illustrated in Fig. 2b.

To quantify the possible improvement, we measure the population of the ground state $p_0(t)$ as a function of the delay time t when the qubit is prepared in $|0\rangle$, $|1\rangle$, $|2\rangle$, and $|3\rangle$ ($t = 0$ means that there is no delay between the last shelving pulse and the readout pulse). State preparation and measurement (SPAM) errors are mitigated by applying the inverse of the assignment matrix to the measurement results[41]. Then, the most probable physical state is acquired with a maximum likelihood estimator[42]. The data for Qubit 2 is shown in Fig. 2c. We find that when the qubit is prepared in $|1\rangle$, the $|1\rangle$-state population decays during the readout by $\epsilon = 1 - e^{-\tau_r/T_{01}} = 2.24\,\%$, with relaxation time $T_{01} = 6.18\,\mu s$, giving a significant contribution to the readout error. The duration of our optimized readout pulse[43–45] is $\tau_r = 140\,ns$, which is minimized by optimizing the readout-pulse amplitude without introducing significant readout-induced mixing that contributes to the readout errors[46,47].

We calculate the population $p_i(t)$ in the $|i\rangle$ state with the following rate equations:

$$\dot{p}_0(t) = p_1(t)/T_{01} + p_2(t)/T_{02} + p_3(t)/T_{03},$$
$$\dot{p}_1(t) = -p_1(t)/T_{01} + p_2(t)/T_{12} + p_3(t)/T_{13},$$
$$\dot{p}_2(t) = -p_2(t)/(T_{12} + T_{02}) + p_3(t)/T_{23}, \tag{1}$$
$$\dot{p}_3(t) = -p_3(t)/(T_{23} + T_{13} + T_{03}),$$
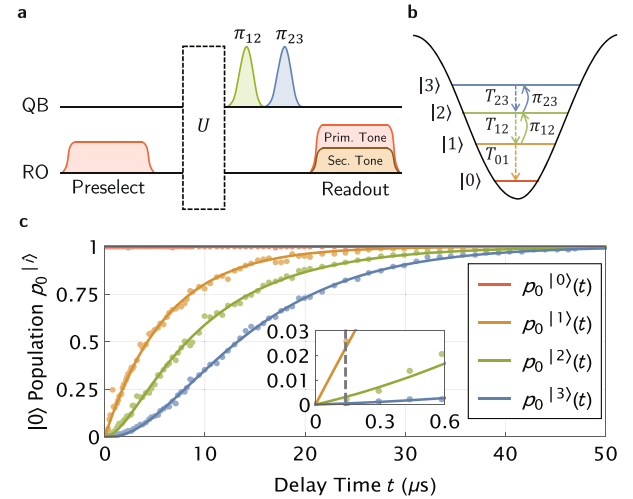
Fig. 2 Schematics and implementation of the shelving and two-tone readout technique. a A π_{12} and a consecutive π_{23} pulse are inserted between any experimental sequence U and the final readout pulse comprising two readout tones. This scheme can be implemented in any multi-level quantum processor platform without modifying the hardware design. **b** Qubit population can be transferred to the desired energy level with consecutive π_{ij} pulses. The population in state $|j\rangle$ decays to $|i\rangle$ with a rate $1/T_{ij}$. **c** The ground state $|0\rangle$ population p_0 is plotted as a function of the delay time t after the transmon is initially prepared in $|0\rangle$, $|1\rangle$, $|2\rangle$, or $|3\rangle$. The delay time t is counted from the end of the last qubit shelving pulse in the sequence. Points represent experimental data for Qubit 2 while continuous lines show fits of the data according to the solutions of the expanded rate equations including all non-sequential rates[34], which are presented in Section III of the Supplementary Note. The inset shows the population at short time scales with the dashed line marking the duration $\tau_r = 140\,ns$ of the readout pulse.

where T_{ij} is the relaxation time from the $|j\rangle$ to the $|i\rangle$ state as illustrated in the level diagram of Fig. 2b. In principle, the anharmonicity of the transmon is sufficient such that non-sequential decay through multi-level channels is exponentially suppressed. For example, the contribution of direct decay from $|2\rangle$ to $|0\rangle$ is found to be two orders of magnitude smaller than that from $|2\rangle$ to $|1\rangle$[34]. For simplicity, we neglect the non-sequential decay terms and solve for the evolution of any $|i\rangle$-state population as a function of time t when the qubit is initialized in the $|j\rangle$ state, denoted as $p_i^{|j\rangle}(t)$. Specifically, we first solve for $p_0^{|2\rangle}(t)$ and assume $p_2^{|2\rangle}(0) = 1$ in the absence of any error and neglect the effect of higher-energy levels by using the initial conditions $p_0^{|2\rangle} = p_1^{|2\rangle} = p_3^{|2\rangle} = 0$. We find:

$$p_0^{|2\rangle}(t) = 1 - \frac{T_{01}e^{-t/T_{01}}}{T_{01} - T_{12}} + \frac{T_{12}e^{-t/T_{12}}}{T_{01} - T_{12}}, \tag{2}$$

where $T_{12} \approx \frac{1}{2}T_{01}$ for typical transmon parameters[34,48]. The second and third terms in Eq. (2) are two decaying functions with opposite signs, hence the net result is no longer purely exponential. We also solve the rate equations of the system in Eq. (1) for $p_0^{|3\rangle}$, when the qubit is prepared in $|3\rangle$, with the initial conditions $p_0^{|3\rangle} = p_1^{|3\rangle} = p_2^{|3\rangle} = 0$, and find the following:

$$p_0^{|3\rangle}(t) = 1 - T_{01}{}^2/[(T_{01} - T_{12})(T_{01} - T_{23})] \cdot e^{-t/T_{01}}$$
$$+ T_{12}{}^2/[(T_{01} - T_{12})(T_{12} - T_{23})] \cdot e^{-t/T_{12}} \tag{3}$$
$$- T_{23}{}^2/[(T_{01} - T_{23})(T_{12} - T_{23})] \cdot e^{-t/T_{23}}.$$

This equation contains a combination of exponential decays with different signs as well. Using the solutions that include the

non-sequential decay terms, which are presented as Eq. (S.10–S.13) in Section III of the Supplementary Note, we fit the data in Fig. 2c. We find excellent agreement between the data and the full model. Experimentally, we find that there is a significant deviation at short time scales if the non-sequential decay T_{02}, T_{03}, and T_{13} are not taken into account[34], which is illustrated in Supplementary Fig. 2.

For Qubit 2, we find that the readout error from the $|1\rangle$ state is reduced from 2.24 % to 0.057 % after the application of a π_{12} pulse, and to 8.7×10^{-4} % after a π_{23} pulse. This is calculated by taking the difference between $p_0^{|2\rangle}(t = \tau_r)$ and $p_0^{|2\rangle}(0)$ in Fig. 2a. The remaining error is equivalent to the decay error of a qubit with $T_{01} = 85.8$ μs using the standard readout method, achieving an order-of-magnitude improvement. For a longer-lived qubit, the percentage of decay errors that can be suppressed with shelving continues to grow closer to unity[36]. However, other error contributions will likely become more prominent at this level.

Two-state readout with a primary tone

Having optimized the state-preparation pulses (see Methods section), we fine-tune the readout frequency to maximally separate the 2D in-phase and quadrature (IQ) histograms corresponding to the $|0\rangle$ and $|1\rangle$ states, as shown in Fig. 3a. Higher energy levels are indistinguishable from $|1\rangle$ in this
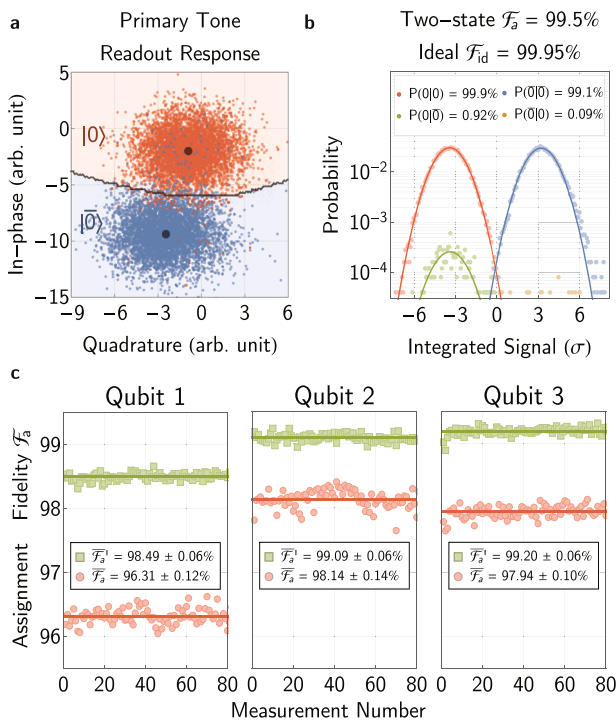
Fig. 3 Single-shot readout results of discriminating between $|0\rangle$ and $|\overline{0}\rangle$. a Integrated readout signal in the IQ plane with Qubit 2 prepared in $|0\rangle$ and $|1\rangle$. With the consecutive π_{12} and π_{23} pulses implemented prior to a 140 ns readout, we distinguish between $|0\rangle$ and $|\overline{0}\rangle$ (NOT $|0\rangle$). **b** The IQ-plane signals in **a** are projected onto an optimal axis, and the resulting histogram is fitted with a Gaussian distribution. The horizontal axis is normalized by the standard deviation σ. The calculated assignment fidelity $\mathcal{F}_a$ and ideal fidelity $\mathcal{F}_{id}$ are shown above the plot. The conditional probabilities $P(i|j)$ represent the probabilities of measuring state $|i\rangle$ given that the qubit is prepared in state $|j\rangle$. **c** Simultaneously measured single-shot readout assignment fidelities for the three qubits with ($\overline{\mathcal{F}'}$) and without ($\overline{\mathcal{F}_a}$) the application of the π_{12} and π_{23} pulses. The error statistics are calculated from the standard deviation of the measured set.

configuration, and we can only distinguish between $|0\rangle$ and $|\overline{0}\rangle$ (NOT $|0\rangle$). To calibrate the readout, we prepare the qubit in either $|0\rangle$ or $|1\rangle$ and add the π_{12} and π_{23} pulses to transfer the $|1\rangle$-state population to the $|3\rangle$ state before the readout, as illustrated in Fig. 2a. First, we start with a single readout tone in the pulse which is referred to as the primary tone. To discard the results for which the initial state is not $|0\rangle$, we include a preselection pulse, i.e., an additional readout measurement before the sequence starts[12,49]. Through this preselection procedure, thermal and residual populations in the qubits are filtered from the outcomes.

We perform simultaneous readout of all three qubits and calculate the single-qubit readout assignment fidelity $\mathcal{F}_a = 1 - [P(0|\overline{0}) + P(\overline{0}|0)]/2$, where $P(i|j)$ is the classification probability, i.e., the probability that the $|i\rangle$ state is assigned given that the $|j\rangle$ state is initially prepared. This measure of readout fidelity takes all the error contributions into account, including imperfect state preparation before the readout sequence. The assignment fidelity for 80 repetitions, each containing 50,000 shots, is shown in Fig. 3c with ($\mathcal{F}'$) and without ($\overline{\mathcal{F}_a}$) implementation of the shelving technique. The data demonstrate a reduction in the overall error rate by 57 % on average for the three qubits with the introduced readout scheme. We also compute the ideal readout fidelity $\mathcal{F}_{id}$ by integrating the overlapping area of the Gaussian probability distributions after projecting the IQ data onto an optimal axis[50]:

$$\mathcal{F}_{id} = \frac{1}{2}\left[1 + \mathrm{erf}\left(\sqrt{\frac{\mathrm{SNR}^2}{8}}\right)\right], \quad (4)$$

where the signal-to-noise ratio is

$$\mathrm{SNR} = \frac{|\langle S_0 \rangle - \langle S_{\overline{0}} \rangle|}{\sigma_{S_0}}, \quad (5)$$

with S_i being the set of measurement outcomes and σ_{S_0} being the standard deviation of the data set. For Qubit 2, the best assignment fidelity is 99.5 % while the ideal fidelity exceeds 99.95%; see Fig. 3b for detailed histograms. In particular, the readout errors associated with the $|0\rangle$ and the $|\overline{0}\rangle$ state are $\epsilon_{|0\rangle} = P(\overline{0}|0) = 0.09$ % and $\epsilon_{|\overline{0}\rangle} = P(0|\overline{0}) = 0.92$ %, respectively. The error from qubit decay during readout contributes to 0.03% of the $|\overline{0}\rangle$-state error $\epsilon_{|\overline{0}\rangle}$. This suggests that the fidelity is predominantly limited by other sources of error. The most dominant source of error is due to imperfect state preparation and application of the shelving pulses[12]. This contribution is at least 58% of the $\epsilon_{|\overline{0}\rangle}$ error and is calculated based on the coherence limit of the qubit and serves as a lower bound. The measurement-induced mixing is the next dominant source of error. We estimate this contribution to be at most 10% of $\epsilon_{|\overline{0}\rangle}$, since this affects both $\epsilon_{|0\rangle}$ and $\epsilon_{|\overline{0}\rangle}$ equally[46,47].

Three-state readout with two tones

Choosing the optimal readout frequency to attain the best two-state assignment fidelity leaves other higher-energy states indistinguishable from each other. However, the information of the $|2\rangle$-state population is crucial to detect leakage errors during gate calibrations and algorithms[2]. To access this information, we introduce a secondary readout pulse with a readout frequency that maximizes the separation between $|1\rangle$ and $|2\rangle$. This pulse is multiplexed with the primary pulse to perform the readout measurements simultaneously.

We also use the π_{12} and π_{23} pulses to implement the shelving scheme. As the initial $|1\rangle$-level population is transferred to the $|3\rangle$ state and the $|2\rangle$-population is transferred to the $|1\rangle$ state, an error in $|1\rangle$-state assignment will occur if a cascade of decays happens from $|3\rangle$ to $|1\rangle$. The contribution of decay error during readout is
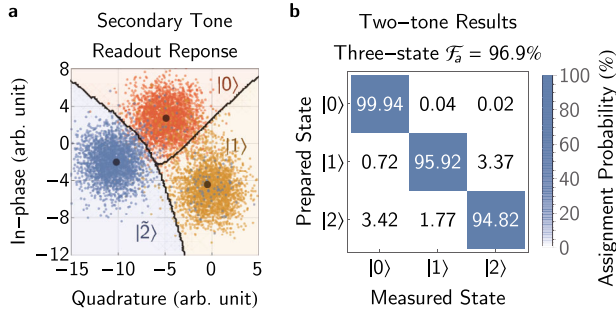
Fig. 4 Single-shot readout results for discriminating between the |0⟩, |1⟩, and |2⟩ state. a Integrated single-shot readout signal of the secondary tone for Qubit 2. The |0⟩ and |1⟩ states are distinguishable from the rest, while |2⟩ and |3⟩ have significant overlap, and are therefore being combined into a single classification: $|\tilde{2}⟩$. The frequency of the primary tone is identical to that shown in Fig. 3, which maximizes the distinction between |0⟩ and $|\bar{0}⟩$. The dots indicate the centers of the Gaussian distributions. **b** Three-state assignment matrix with the two-tone readout for Qubit 2, reconstructed using a neural network. Note that the most significant error contributions in the two-tone readout are the misclassification between |2⟩ and |0⟩ as well as that between |1⟩ and |2⟩.

Table 1. Truth table of the selection rule for the two-tone readout.

| Primary Result	Secondary Result	Before Readout	Initial State
|0⟩	|0⟩	|0⟩	|0⟩
$|\bar{0}⟩$	|1⟩	|3⟩	|1⟩
$|\bar{0}⟩$	$|\tilde{2}⟩$	|1⟩	|2⟩
|0⟩	$|\bar{0}⟩$	Overlap Error	
$|\bar{0}⟩$	|0⟩	Overlap Error	

The first two columns are the discrimination results from the two readout tones, respectively. There is a unique initial state if the results agree with each other. Otherwise, the shots where the two readout results disagree will be counted towards overlap error and discarded.

reduced, leading to an improvement in assignment fidelity. To quantify the improvement, we need to solve Eq. (1) for the evolution of population $p_1^{|3⟩}(t)$. The analytical solution is

$$
\begin{aligned}
p_1^{|3⟩}(t) &= T_{01}^2/[(T_{01} - T_{12})(T_{01} - T_{23})] \cdot e^{-t/T_{01}} \\
&\quad - T_{01}T_{12}/[(T_{01} - T_{23})(T_{12} - T_{23})] \cdot e^{-t/T_{12}} \\
&\quad + T_{01}T_{23}/[(T_{01} - T_{12})(T_{12} - T_{23})] \cdot e^{-t/T_{23}}.
\end{aligned}
\tag{6}
$$

With the qubit being in the |3⟩ state ($p_3^{|3⟩}(0) = 1$), the effective population accumulation in |1⟩ after a readout time of $\tau_r = 140$ ns is 0.035 %. This is two orders of magnitude smaller than the error from a direct decay of the |2⟩-state population with rate $1/T_{12}$, corresponding to 2.65%. Therefore, the contribution to the three-state readout error should be reduced by a similar factor.

The secondary tone is tuned up in the presence of the primary pulse. An initial estimate for the secondary readout frequency is where the |1⟩- and |2⟩-state responses are maximally separated in the IQ-plane (see Methods section). We then fine-tune this frequency such that |0⟩ and |3⟩ are distinguishable from each other as well. After optimization, the two readout pulses are typically a few MHz apart and are multiplexed in a single waveform for the readout. The transmitted signal is then processed with standard multiplexed readout techniques[13]. We obtain two complex voltages after signal integration, each containing a pair of in-phase and quadrature values. Overlap errors are then filtered by comparing the results in post-processing. The response of the secondary tone when Qubit 2 is
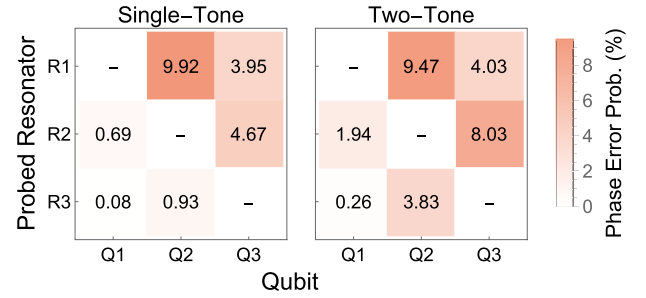
Fig. 5 Measurement-induced dephasing matrix for single-tone and two-tone readout. Each element represents the qubit dephasing while a pulse is targeting one of the readout resonators (R1, R2, or R3). Note that for the two-tone case, resonator 1 was not driven by an additional tone, so it served as a benchmark measurement.

prepared in |0⟩, |1⟩, |2⟩, and |3⟩, with 50,000 repetitions per state, is shown in Fig. 4a. Since |2⟩ and |3⟩ are indistinguishable, we combine the results together and relabel them as the $|\tilde{2}⟩$-state. We then formulate two methods to combine the results from the primary and secondary readout pulses to reconstruct the population initially prepared in |0⟩, |1⟩, and |2⟩. The first method is a truth table (see Table 1) that takes the individual measurement results from the two tones as a pair of input values. There exists a unique initial state that is compatible with both results. For example, if the primary result is $|\bar{0}⟩$ and the secondary result is |1⟩, then the initially prepared state must be |2⟩. In case these two results cannot reach a common decision due to overlap error, the measurement is discarded.

The second method to discriminate the qubit state utilizes a feedforward neural network (FNN) that was specifically developed for multiplexed readout[51] and treats the two data sets as a single system (see Methods section). The input to the neural network combines the in-phase (I[n]) and quadrature (Q[n]) data from the nth integrated signal of the two tones into a four-element vector $\{I_1[n], Q_1[n], I_2[n], Q_2[n]\}$. After being trained with a calibration data set, the neural network is able to classify two-tone results and give the initial qubit state as the output.

The resulting assignment matrix, shown in Fig. 4b, demonstrates an assignment fidelity of 96.9 % for the three-state readout. This result shows a significant improvement over the average 94.2 % assignment fidelity that we find using only a single readout pulse at an optimal readout frequency to distinguish between |0⟩, |1⟩, and |2⟩, with the overall error rate reduced by 47 %. The amount of suppression is achieved with the neural network that consistently outperforms the truth table by 13 % in an overall error rate reduction.

A feature of the resulting assignment matrix is that the population originally in |2⟩ has a higher probability to be misidentified as |0⟩ than as |1⟩, which is due to the shelving technique. Since the initial |2⟩-state population is transferred to |1⟩ before measurement, decaying to the ground state is more likely than the excitation back to higher energy levels.

We also investigate the effect of increased photon population in the resonators due to the secondary tone. A large photon number leads to significant measurement-induced mixing and readout crosstalk that contribute to the overall readout error[13,52]. We optimize the readout amplitude such that readout errors due to measurement-induced mixing remain small with the addition of the secondary tone. To investigate readout crosstalk, we determine the measurement-induced dephasing with and without the secondary tone[13]. We find that the probabilities of a phase error in untargeted qubits are a factor of three larger on average due to the increase in photon number, as shown in Fig. 5. Mitigation strategies may be required if this contribution becomes significant for error-correction algorithms.

In the design of future devices, the qubits could be grouped into physically separated readout lines depending on their designation,
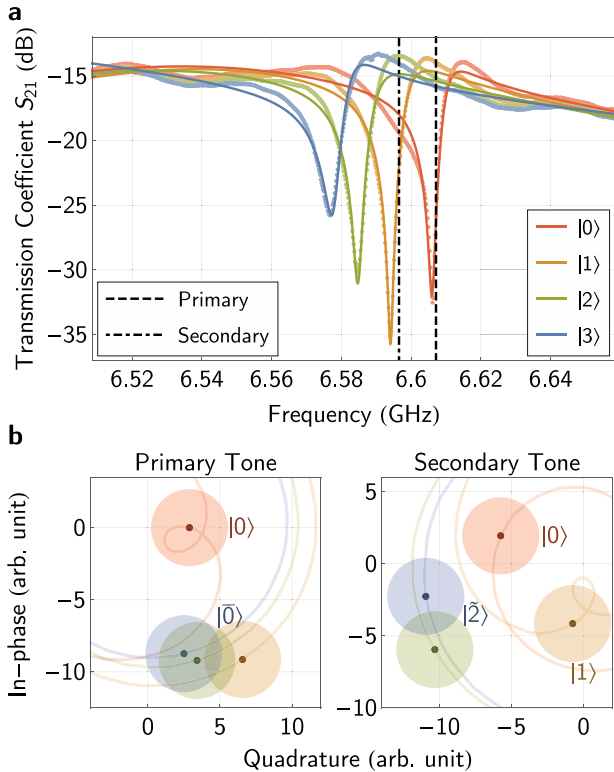
Fig. 6 Transmission coefficient S_{21} as a function of the driving frequency ω_d centered around Resonator 2. a Qubit-state-dependent transmission S_{21} of resonator R2 when Qubit 2 is prepared in $|0\rangle$, $|1\rangle$, $|2\rangle$, and $|3\rangle$, respectively. The colored dots represent the measured data while the solid lines show the fitted function (see Section I of the Supplementary Notes). The vertical lines represent the optimal readout frequencies for the primary (dashed line) and secondary tone (dash-dotted line). **b** Estimated readout response of the primary and secondary tones at their respective optimal frequencies. The solid lines represent the in-phase and quadrature data shown in **a**. The disks of respective color represent the estimated Gaussian envelope of the signal taking into account the added thermal noise.

e.g., ancilla or data qubits, and whether their measurements occur simultaneously. Moreover, the induced crosstalk could be further mitigated with other techniques such as machine-learning algorithms for discrimination and readout pulse shaping[51,53,54].

DISCUSSION

In conclusion, we have demonstrated that exploiting the higher energy levels of the qubit together with the implementation of a secondary readout tone lead to an improved readout-assignment fidelity of 99.5 % (96.9 %) for two-state (three-state) discrimination within 140 ns of readout time, reducing the overall error rate by 57 % (47 %) compared to our baseline. This result could be further improved with the use of a quantum-limited parametric amplifier.

The proposed pulse scheme is straightforward to implement in any measurement sequence to improve readout fidelity. Our scheme is also fully compatible with quantum error-correction algorithms that involve rapid, high-fidelity measurements of ancilla qubits. In the case of the surface code, single-qubit and measurement operations ideally do not disturb the encoded logical state stored on data qubits. The proposed techniques will only be performed on the ancilla qubits. Additionally, we envisage that to be fully utilized in the surface-code algorithm, the readout scheme could be accompanied by a reset protocol that targets higher-excited states after ancilla measurement[15].

To further develop these readout techniques, more complex methods in the construction of the primary and secondary readout tones could be explored. Sophisticated deep neural-network methods could also be employed to aid state classification of the two-tone readout results[51]. The possibility to generalize these techniques to further boost fidelity for multiplexed readout is a promising prospect for the future investigation of quantum computing with superconducting qubits.

METHODS

Pulse calibration

We optimize the parameters of the π_{12} and π_{23} pulses similar to the standard method developed for the π_{01} pulse. The pulses have a cosine envelope with lengths set to be 50 ns as a starting point. We first prepare the qubit in $|1\rangle$ and conduct Rabi and Ramsey-like experiments between the higher energy levels to optimize the amplitude and frequency of the drive pulse. For shorter pulse lengths a proper derivative removal by adiabatic gate (DRAG)[55,56] needs to be calibrated for the π_{12} and π_{23} pulses as well.

With the state-preparation pulses calibrated, we acquire the responses of the readout resonator when the qubit is prepared in $|0\rangle$, $|1\rangle$, $|2\rangle$, and $|3\rangle$, respectively, illustrated in Fig. 6a. We plot the in-phase and quadrature parts of the spectroscopy result, as shown in Fig. 6b. We calculate the distance between each point of a pair of trajectories, which represents the separation of the two respective state responses as a function of readout frequency. We find the readout frequencies that maximize $|0\rangle$-$|1\rangle$ separation for the primary tone, and $|1\rangle$-$|2\rangle$ separation for the secondary tone. As the readout tones are multiplexed into a single pulse, the frequency and phase of the secondary tone are fine-tuned to minimize the effect on the measurement performance of the primary tone. The readout power is adjusted such that the critical photon number is not surpassed, as shown in Section IV of the Supplementary Note.

Feedforward neural network

We utilize a feedforward neural network (FNN) with two hidden layers to discriminate the qubit state using the combined two-tone results. The choice of using an FNN over other methods such as support-vector machine (SVM) or non-linear support-vector machine (NSVM) is justified by the fact that an FNN could achieve greater performance when discriminating more than two states as well as better scalability[51]. The FNN is also capable of transfer learning, where retraining of the network during future re-calibration of the system is significantly more efficient[57]. On the other hand, an SVM or NSVM will need to be retrained from scratch every time. The advantage of using FNN is significant enough to warrant a wider application[51,53,58].

The network is implemented with Wolfram Mathematica. The input layer contains four nodes corresponding to the in-phase and quadrature components of the two-tone results, $\{I_1[n], Q_1[n], I_2[n], Q_2[n]\}$, of each individual single-shot measurement n. The first hidden layer contains 16 nodes, while the second layer has eight nodes. Each node comprised of the hidden layers is filtered by a scaled exponential linear unit (SELU), which acts as the nonlinear activation function. Finally, the output layer contains three nodes, representing the probability of the qubit being in state $|0\rangle$, $|1\rangle$, and $|2\rangle$, respectively. We specify an epoch of 100 and a learning rate of 0.0005 with a batch size of 64 as a starting point. The network is then trained with 8000 samples and validated with 2000 samples.

CODE AVAILABILITY

All relevant code supporting the document is available upon request. Please refer to Liangyu Chen at liangyuc@chalmers.se.

REFERENCES

1. Chen, Z. et al. Exponential suppression of bit or phase errors with cyclic error correction. *Nature* **595**, 383–387 (2021).
2. Krinner, S. et al. Realizing repeated quantum error correction in a distance-three surface code. *Nature* **605**, 669–674 (2022).
3. Zhao, Y. et al. Realization of an error-correcting surface code with superconducting qubits. *Phys. Rev. Lett.* **129**, 030501 (2022).
4. Google Quantum AI. Suppressing quantum errors by scaling a surface code logical qubit. *Nature* **614**, 676–681 (2023).
5. Lo, H.-K., Spiller, T. & Popescu, S., Introduction to Quantum Computation and Information (World Scientific, 1998). https://doi.org/10.1142/3724.
6. Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: Towards practical large-scale quantum computation. *Phys. Rev. A* **86**, 032324 (2012).
7. Martinis, J. M. Qubit metrology for building a fault-tolerant quantum computer. *npj Quantum Inform.* **1**, 15005 (2015).
8. Foxen, B. et al. Demonstrating a continuous set of two-qubit gates for near-term quantum algorithms. *Phys. Rev. Lett.* **125**, 120504 (2020).
9. Xu, Y. et al. High-fidelity, high-scalability two-qubit gate scheme for superconducting qubits. *Phys. Rev. Lett.* **125**, 240503 (2020).
10. Negîrneac, V. et al. High-fidelity controlled-Z gate with maximal intermediate leakage operating at the speed limit in a superconducting quantum processor. *Phys. Rev. Lett.* **126**, 220502 (2021).
11. Sung, Y. et al. Realization of high-fidelity CZ and ZZ-free iSWAP gates with a tunable coupler. *Phys. Rev. X* **11**, 021058 (2021).
12. Walter, T. et al. Rapid high-fidelity single-shot dispersive readout of superconducting qubits. *Phys. Rev. Appl.* **7**, 054020 (2017).
13. Heinsoo, J. et al. Rapid High-fidelity Multiplexed Readout of Superconducting Qubits. *Phys. Rev. Appl.* **10**, 034040 (2018).
14. Elder, S. S. et al. High-fidelity measurement of qubits encoded in multilevel superconducting circuits. *Phys. Rev. X* **10**, 011001 (2020).
15. Sunada, Y. et al. Fast readout and reset of a superconducting qubit coupled to a resonator with an intrinsic Purcell filter. *Phys. Rev. Appl.* **17**, 044016 (2022).
16. Blais, A., Grimsmo, A. L., Girvin, S. M. & Wallraff, A. Circuit quantum electrodynamics. *Rev. Mod. Phys.* **93**, 025005 (2021).
17. Campagne-Ibarcq, P. et al. Deterministic remote entanglement of superconducting circuits through microwave two-photon transitions. *Phys. Rev. Lett.* **120**, 200501 (2018).
18. Kurpiers, P. et al. Deterministic quantum state transfer and remote entanglement using microwave photons. *Nature* **558**, 264–267 (2018).
19. Kurpiers, P. et al. Quantum communication with time-bin encoded microwave photons. *Phys. Rev. Appl.* **12**, 044067 (2019).
20. Ilves, J. et al. On-demand generation and characterization of a microwave time-bin qubit. *npj Quantum Inform.* **6**, 34 (2020).
21. Magnard, P. et al. Microwave quantum link between superconducting circuits housed in spatially separated cryogenic systems. *Phys. Rev. Lett.* **125**, 260502 (2020).
22. Dassonneville, R. et al. Fast high-fidelity quantum nondemolition qubit readout via a nonperturbative Cross-Kerr coupling. *Phys. Rev. X* **10**, 011045 (2020).
23. Blais, A., Huang, R.-S., Wallraff, A., Girvin, S. M. & Schoelkopf, R. J. Cavity quantum electrodynamics for superconducting electrical circuits: an architecture for quantum computation. *Phys. Rev. A* **69**, 062320 (2004).
24. Reed, M. D. et al. High-fidelity readout in circuit quantum electrodynamics using the Jaynes-Cummings nonlinearity. *Phys. Rev. Lett.* **105**, 173601 (2010).
25. Gusenkova, D. et al. Quantum nondemolition dispersive readout of a superconducting artificial atom using large photon numbers. *Phys. Rev. Appl.* **15**, 064030 (2021).
26. Mallet, F. et al. Single-shot qubit readout in circuit quantum electrodynamics. *Nat. Phys.* **5**, 791–795 (2009).
27. Jeffrey, E. et al. Fast accurate state measurement with superconducting qubits. *Phys. Rev. Lett.* **112**, 190504 (2014).
28. Sete, E. A., Martinis, J. M. & Korotkov, A. N. Quantum theory of a bandpass Purcell filter for qubit readout. *Phys. Rev. A* **92**, 012325 (2015).
29. Yurke, B. et al. Observation of parametric amplification and deamplification in a josephson parametric amplifier. *Phys. Rev. A* **39**, 2519 (1989).
30. Mutus, J. Y. et al. Strong environmental coupling in a Josephson parametric amplifier. *Appl. Phys. Lett.* **104**, 263513 (2014).
31. Macklin, C. et al. A near quantum-limited Josephson traveling-wave parametric amplifier. *Science* **350**, 307–310 (2015).
32. Renger, M. et al. Beyond the standard quantum limit for parametric amplification of broadband signals. *npj Quantum Inform.* **7**, 160 (2021).
33. Martinis, J. M., Nam, S., Aumentado, J. & Urbina, C. Rabi oscillations in a large Josephson-junction qubit. *Phys. Rev. Lett.* **89**, 117901 (2002).
34. Peterer, M. J. et al. Coherence and decay of higher energy levels of a superconducting transmon qubit. *Phys. Rev. Lett.* **114**, 010501 (2015).
35. Jurcevic, P. et al. Demonstration of quantum volume 64 on a superconducting quantum computing system. *Quantum Sci. Technol.* **6**, 025020 (2021).
36. Wang, C., Chen, M.-C., Lu, C.-Y. & Pan, J.-W. Optimal readout of superconducting qubits exploiting high-level states. *Fundam. Res.* **1**, 16 (2021).
37. Kosen, S. et al. Building blocks of a flip-chip integrated superconducting quantum processor. *Quantum Sci. Technol.* **7**, 035018 (2022).
38. Koch, J. et al. Charge-insensitive qubit design derived from the cooper pair box. *Phys. Rev. A* **76**, 042319 (2007).
39. D'Anjou, B. & Coish, W. A. Enhancing qubit readout through dissipative sub-Poissonian dynamics. *Phys. Rev. A* **96**, 052321 (2017).
40. Cottet, N., Xiong, H., Nguyen, L. B., Lin, Y.-H. & Manucharyan, V. E. Electron shelving of a superconducting artificial atom. *Nat. Commun.* **12**, 6383 (2021).
41. Jattana, M. S., Jin, F., De Raedt, H. & Michielsen, K. General error mitigation for quantum circuits. *Quantum Inform. Process.* **19**, 414 (2020).
42. Baumgratz, T., Nüßeler, A., Cramer, M. & Plenio, M. B. A scalable maximum likelihood method for quantum state tomography. *New J. Phys.* **15**, 125004 (2013).
43. Schuster, D. I. et al. ac stark shift and dephasing of a superconducting qubit strongly coupled to a cavity field. *Phys. Rev. Lett.* **94**, 123602 (2005).
44. Gambetta, J. et al. Qubit-photon interactions in a cavity: measurement-induced dephasing and number splitting. *Phys. Rev. A* **74**, 042318 (2006).
45. Schuster, D. I. et al. Resolving photon number states in a superconducting circuit. *Nature* **445**, 515–518 (2007).
46. Sank, D. et al. Measurement-induced state transitions in a superconducting qubit: Beyond the rotating wave approximation. *Phys. Rev. Lett.* **117**, 190503 (2016).
47. Slichter, D. H. et al. Measurement-induced qubit state mixing in circuit qed from up-converted dephasing noise. *Phys. Rev. Lett.* **109**, 153601 (2012).
48. Wang, H. et al. Measurement of the decay of fock states in a superconducting quantum circuit. *Phys. Rev. Lett.* **101**, 240401 (2008).
49. Ristè, D., van Leeuwen, J. G., Ku, H.-S., Lehnert, K. W. & DiCarlo, L. Initialization by measurement of a superconducting quantum bit circuit. *Phys. Rev. Lett.* **109**, 050507 (2012).
50. Magesan, E., Gambetta, J. M., Córcoles, A. D. & Chow, J. M. Machine learning for discriminating quantum measurement trajectories and improving readout. *Phys. Rev. Lett.* **114**, 200501 (2015).
51. Lienhard, B. et al. Deep-neural-network discrimination of multiplexed superconducting-qubit states. *Phys. Rev. Appl.* **17**, 014024 (2022).
52. Bultink, C. C. et al. General method for extracting the quantum efficiency of dispersive qubit readout in circuit qed. *Appl. Phys. Lett.* **112**, 092601 (2018).
53. Duan, P. et al. Mitigating crosstalk-induced qubit readout error with shallow-neural-network discrimination. *Phys. Rev. Appl.* **16**, 024063 (2021).
54. Lienhard, B. Machine Learning Assisted Superconducting Qubit Readout, Ph.D. thesis, MIT (2021). https://hdl.handle.net/1721.1/140024.
55. Motzoi, F., Gambetta, J. M., Rebentrost, P. & Wilhelm, F. K. Simple pulses for elimination of leakage in weakly nonlinear qubits. *Phys. Rev. Lett.* **103**, 110501 (2009).
56. Gambetta, J. M., Motzoi, F., Merkel, S. T. & Wilhelm, F. K. Analytic control methods for high-fidelity unitary operations in a weakly nonlinear oscillator. *Phys. Rev. A* **83**, 012308 (2011).
57. Bengio, Y. in *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, Proceedings of Machine Learning Research, Vol. 27, edited by Guyon, I., Dror, G., Lemaire, V., Taylor, G. & Silver, D. (PMLR, Bellevue, Washington, USA, 2012) pp. 17–36. https://proceedings.mlr.press/v27/bengio12a.html.
58. Navarathna, R. et al. Neural networks for on-the-fly single-shot state classification. *Appl. Phys. Lett.* **119**, 114003 (2021).

AUTHOR CONTRIBUTIONS

FUNDING

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.