Mitigating Crosstalk-Induced Qubit Readout Error with Shallow-Neural-Network Discrimination

Peng Duan,[1,2,†] Zi-Feng Chen,[1,2,†] Qi Zhou,[1,2] Wei-Cheng Kong,[3] Hai-Feng Zhang,[1,2] and Guo-Ping Guo[1,2,3,*]

[1] *CAS Key Laboratory of Quantum Information, University of Science and Technology of China, Hefei, Anhui 230026, China*

[2] *CAS Center for Excellence and Synergetic Innovation Center in Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei, Anhui 230026, China*

[3] *Origin Quantum Computing Company Limited, Hefei, Anhui 230026, China*

(Received 19 March 2021; revised 2 August 2021; accepted 13 August 2021; published 31 August 2021)

Measurement of qubits plays a key role in quantum computation. In superconducting multiqubit quantum processors, a multiplexed readout scheme is widely used. In such a scheme, measurement of a qubit state may be influenced by the state of neighboring qubits, due to various crosstalk effects, which will degrade the readout fidelity. To reduce the impact of crosstalk, we model the digital signal processing system used in measurements as a shallow neural network and train it to become a state discriminator. Applying our method to a six-qubit superconducting quantum chip, we see an overall improved readout performance compared with a contemporary qubit-state discriminator. The readout crosstalk is decreased by more than 80%. The training and optimization process of the neural network consumes only about 10 s.

DOI: 10.1103/PhysRevApplied.16.024063

I. INTRODUCTION

High-fidelity single-shot qubit readout of a multiqubit system is an essential feature in scalable quantum computation [1]. For superconducting qubits, a state-of-the-art readout is implemented by probing the qubit-state-dependent frequency shift of a readout resonator coupled to the qubit based on dispersive interaction [2–4]. Remarkable progress in research has been made in rapid high-fidelity single-shot and multiplexed readout [5,6] for circuit quantum electrodynamics [7] systems of a few qubits [8–11]. To further improve the readout fidelity while scaling up the system [12,13], one of the main challenges is how we can deal with crosstalk [10,14–17]. There may be crosstalk-induced readout error when measurement of a qubit state is influenced by the neighboring qubits [14]. Such readout crosstalk effect may arise from spurious coupling between neighboring qubits [15–17], as well as unwanted coupling between readout resonators or between the qubit and the resonators of neighboring qubits [15,16]. It may also arise from cross excitation of neighboring (in the frequency domain) resonators due to the large spectral width of the probe pulse [10,15].

Mitigation strategies for such crosstalk effects have been studied before [10,18,19], with focus mainly on the hardware level. Examples are using an independent Purcell filter for each resonator [10] and enlarging the space and frequency distance between neighboring resonators [18,19]. One may wonder whether it is possible to mitigate the crosstalk-induced readout error at the software level, e.g., data processing or probe-pulse shaping, when there is an inevitable crosstalk effect in the hardware. Convolving the probe-pulse shape with a Gaussian kernel may help to reduce the crosstalk effect, but this prediction has not been experimentally verified [10].

From the perspective of data processing, in frequency-multiplexed readout, the implementation of parallel mode-matched filters [20] by weighted integration for each readout frequency effectively improves readout fidelity [21–23]. Nevertheless, these methods focus on information extraction from the readout frequency of each qubit with little consideration of the partial information leaked to other qubits. That is, matched filters with presumption of a simple noise model fail to capture multiqubit readout crosstalk.

Recently, machine learning algorithms have been applied to many problems in quantum information such as quantum state discrimination [24,25] and quantum dynamics reconstruction [26]. Here in this work, we use a shallow neural network (SNN) to model the process

*gpguo@ustc.edu.cn

†These authors contributed equally to this work.

of conventional digital signal processing (DSP) and train it for state discrimination in simultaneously multiplexed readout. We demonstrate that with this SNN-based method, the readout performance can be improved, and the crosstalk-induced error is largely suppressed.

II. OVERVIEW OF THE CLASSIFIER ARCHITECTURE

We explain briefly the typical DSP flow applied in qubit readouts and the crosstalk problem [16] present in the DSP domain. We then introduce the method to build the architecture of our SNN discriminator based on the DSP chain.

A. Readout and crosstalk problem

In a typical DSP chain, a demodulation in a heterodyning scheme is performed after a down-conversion and digitization of the reflected or transmitted tone from the readout resonator [27]. Demodulation involves multiple stages including mixing with an intermediate frequency (IF), filtering, and integration with a kernel, which may be an optimal demodulation envelope [21]. This three-stage process yields the in-phase (I) and quadrature (Q) components with encoded information regarding the qubit state. Repeated many times, single-shot measurements for the target qubit form two two-dimensional Gaussian statistical distributions in the I-Q plane. These two I-Q clouds correspond to the two possible qubit states. Projecting the clouds onto the axis for which their relative separation in the complex plane is maximized produces a pair of one-dimensional Gaussian curves. Further selection

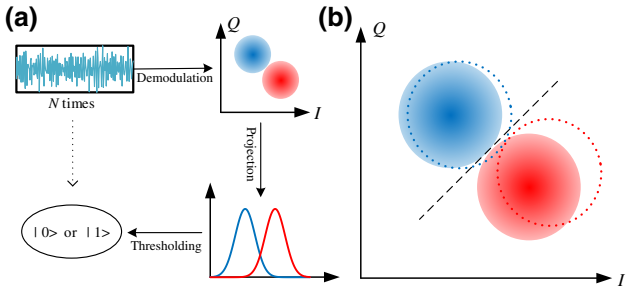

FIG. 1. (a) Diagram of the DSP flow in readout. The raw signal is mapped onto the I-Q plane by demodulation. Projection and thresholding then help to make a state decision. (b) Shifts in the I-Q clouds observed in a two-qubit system. The two solid circles are the I-Q clouds of Q1 with Q2 prepared in state $|0\rangle$ and the black dashed line is the optimum discriminant line that is perpendicular to the projection axis and crosses the threshold point. Dashed circles represent shifts in the I-Q clouds when preparing Q2 in state $|1\rangle$ and the previous decision boundary is no longer optimum.

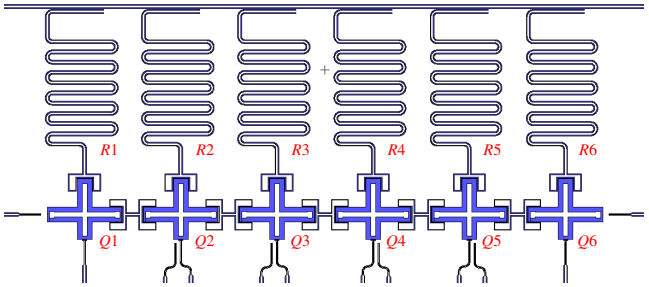

FIG. 2. Schematic layout of our superconducting quantum chip. Each qubit is capacitively coupled to its neighboring qubits and a readout resonator. All resonators share the same feedline. The intermediate frequencies related to the readout resonators range from approximately 500 MHz to approximately 600 MHz. The frequency spacing between individual resonator frequencies is close to the designed value of 20 MHz.

of an appropriate threshold acquires the optimized decision boundary for state discrimination. The DSP flow is depicted in Fig. 1(a).

However, in practice, the result after demodulation may be influenced by more than one qubit. For example, in our quantum chip featuring six transmon qubits [28,29], the readout resonator of Q4 (Fig. 2) may be affected by the state of Q3. As a result, if we prepare Q3 in the ground or excited state before measuring Q4 (Q1, Q2, Q5, and Q6 initialized in their ground states), the I-Q clouds have an obvious shift Fig. 1(b). We use $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$ to denote the four possible basis states of $|\mathrm{Q3,Q4}\rangle$. For the readout of Q4, distinguishing $|00\rangle$ from $|01\rangle$ and $|10\rangle$ from $|11\rangle$ would afford two discriminant lines (denoted as l_0 and l_1) on the I-Q plane. Nevertheless, l_0 and l_1 are not identical due to the shift of I-Q clouds. Cross assignment of state by the two lines would give a significant decline in the fidelity (see Table I). Here, the fidelity of the target qubit i is defined as [9]

$$F^{Qi} = 1 - [P(0_i|1_i) + P(1_i|0_i)]/2, \tag{1}$$

with $P(x_i|y_i)$ the assignment probability that the target qubit i prepared in state y is assigned as state x regardless of the state of nontarget qubits. In Table I, $F(i|l_j)$ $(i,j = 0,1)$ represents state assignment fidelity of Q4 by l_j when Q3 is in state $|i\rangle$.

TABLE I. Fidelities of Q4's readout under the influence of Q3. Here, $F(i|l_j)$ $(i,j = 0,1)$ represents state assignment fidelity of Q4 by l_j when Q3 is in state $|i\rangle$.

$F(i\|l_j)$	$F(0\|l_0)$	$F(1\|l_1)$	$F(0\|l_1)$	$F(1\|l_0)$
	0.9130	0.9150	0.9232	0.8470

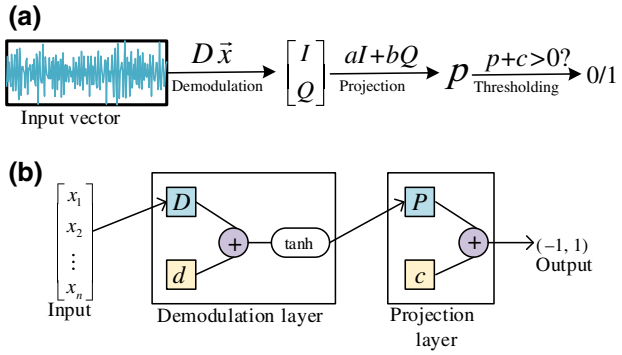

FIG. 3. (a) Matrix form of the DSP chain of the single-qubit readout. (b) Neural network for state assignment based on the DSP system in (a).

B. Architecture of SNN discriminator

The three-stage demodulation pipeline has been mathematically proven to be equivalent to a single kernel integration stage [30]. In other words, if we regard the signal sequence as an N-dimensional column vector $\vec{x}$, demodulation corresponds to left multiplying the signal by a 2-by-N matrix D, mapping the signal to the I-Q plane; specifically,

$$\begin{bmatrix} I \\ Q \end{bmatrix} = D\vec{x}. \tag{2}$$

Suppose the function of the discriminant line is $ax + by + c = 0$; projection may also be represented as matrix manipulation:

$$p = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} I \\ Q \end{bmatrix}. \tag{3}$$

Therefore, we may describe the data processing of the readout through linear algebra [Fig. 3(a)]. A slight modification of this architecture makes it into a two-layer neural network. The addition of biases d in the demodulation layer [Fig. 3(b)] could move the points in the I-Q plane; tanh is the activation function we use in the demodulation layer, whose nonlinearity may help to model the nonlinear effects in the readout crosstalk [31]. We refer to this SNN-based discriminator as SNND.

III. RESULTS

A. Experiment setup

In this section, we illustrate the optimization process for the SNND with experimental data from our device. As an example, we select Q1 as the target qubit.

1. Data preparation and preprocessing

In our experiment, we read out all the six qubits simultaneously with a square pulse through the common feedline, and the transmitted tone is sampled at 1.6 GS/s by an analog-to-digital converter. For our device, six qubits have 64 qubit-state permutations. We prepare 64 sets of 2000 records for the single-shot readout, each set initially prepared in one of 64 possible states and are divided into a randomized training and testing set (1/4 of them for training and 3/4 for testing). The following results are evaluated using the testing set. For the targeted qubit Q1, we label all the samples with 0/1 according to the prepared state of Q1. The optimal integration time for Q1 is 550 ns (470 ns for MFD; see Appendix B), which is a trade-off between the overlap error and the T_1 events [9]. The discrete time measurement record for a total of length N time points is denoted as $v(t_n)$. Before regarding each sample as an input vector for the SNND, we apply a two-step data preprocessing on the training set:

1. Bias. According to the standard demodulation result of the training set, we denote the center of all the I-Q points as (I_c, Q_c). Then we subtract a wave sequence from each sample, $v'(t_n) = v(t_n) - \sqrt{I_c^2 + Q_c^2}\cos(\omega_{\mathrm{IF}} t_n + \varphi)$ (for $n = 1, 2, \ldots, N$), where ω_{IF} are the IFs of readout resonator of Q1 and φ is the angle between vector (I_c, Q_c) and the I axis. Since the demodulation operation is linear, the two I-Q clouds of the biased sample are roughly symmetric about the origin.
2. Feature scaling. Also from the demodulation result, we can calculate the distance d_1 between the centers of the two I-Q clouds corresponding to the two states of Q1. We scale each sample after step 1 with a factor of $2/d_1$ to fix the separation distance of the two centers at 2.

2. Initialization

We initialize the SNND with the training set according to conventional DSP theory. As discussed in Ref. [30], the modified kernel for single-step demodulation is

$$k'_n = e^{-i\omega_{\mathrm{IF}} t_n} \sum_{l=1}^{L} k_l b_{l-n+1}, \tag{4}$$

where b_{l-n+1} are the coefficients of an M-tap finite-impulse response low-pass filter whose valid index runs from 1 to M and k_l is the N time points integration kernel [22]. Subsequently, for the demodulation layer, we initialize the two rows of D with the real part and imaginary part of k'_n, respectively. And we initialize the biases d of the demodulation layer to a zero vector. By inputting a small number of the training samples into the initialized demodulation layer, a rough location of the two I-Q clouds (modified by data preprocessing and tanh function) is acquired. We denote the angle between the line connecting their centers and the I axis as θ. Then we initialize the weights P of the projection layer to $[\cos\theta \;\; \sin\theta]$ and the bias c to zero.

3. Training

The loss function we selected is mean square error $E = (1/2)\sum_{s\in T}(o_s - i_s)^2$, where T represents the training set and o_s (i_s) is the output (input) state label of sample s. The training algorithm we use is gradient descent with momentum [32,33] and adaptive learning rate back-propagation [34] with initial learning rate of 0.01. In our experiment, 500 training events are enough for convergence in about 10 s, and SNND of each qubit is trained in parallel with the same training set.

4. Baseline

For the conventional DSP method, as discussed in Sec. II A, "matched filter" (MF) is the optimal filter in terms of signal-to-noise ratio using an integration kernel [k_l in Eq. (4)], which is proportional to the difference between the mean ground- and excited-state-readout signal [22] and normalized by its standard deviation. As for baseline, we train this MF-based discriminator for each qubit parallelly while nontarget qubits are initialized in their ground state. Note that this is widely applied in current single-qubit or multiple-qubit readout schemes and we refer to it as MFD.

B. Result analysis

In our experiment, the SNND (MFD) is applied for all the six qubits with the same testing set. The fidelities of the six parallelized single-qubit SNND (MFD) for our quantum processor are listed in Table II. Here the fidelity is defined by Eq. (1). The fidelities are improved for all six qubits when using SNND compared with MFD. The largest error reduction is 20.59% seen by Q4 and the average error reduction is about 8.5%.

The difference between assignment probability matrices [10] of SNND and MFD is shown in Fig. 4(a). From this matrix, we see a general fidelity improvement (positive diagonal values) and error reduction (negative off-diagonal values) of SNND over MFD for multiqubit state discrimination.

Moreover, for quantifying the crosstalk-induced readout error, we employ the cross-fidelity matrix introduced in Ref. [10]. The elements of this matrix are defined as

$$\mathbb{F}_{ij} = 1 - P\left(1_i \mid 0_j\right) - P\left(0_i \mid 1_j\right), \tag{5}$$

TABLE II. Fidelities of parallelized SNND and MFD. The last line presents the error reduction ϵ_r of SNND relative to MFD. The average error reduction is about 8.5%.

		F^{Q1}	F^{Q2}	F^{Q3}	F^{Q4}	F^{Q5}	F^{Q6}
MFD	%	96.35	93.02	95.63	90.77	82.54	85.15
SNND	%	**96.54**	**93.71**	**95.83**	**92.67**	**82.62**	**86.73**
ϵ_r	%	5.11	9.94	4.37	20.59	0.47	10.63

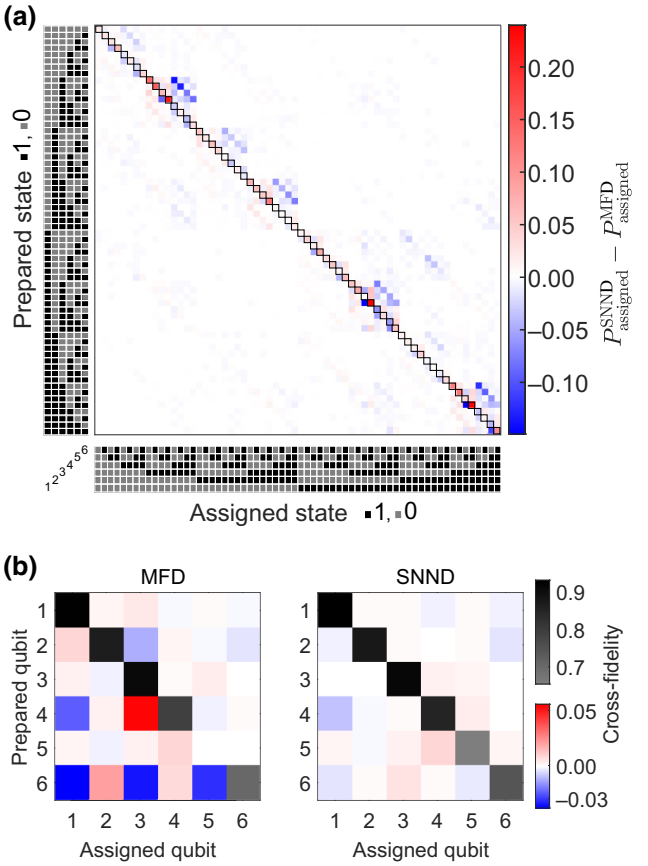

FIG. 4. (a) Difference between assignment probability matrices of trained SNND and MFD. $P(x|y)$ is the probability that the six qubits prepared in state y are assigned as state x. State x (or y) is denoted by a 6-bit binary string. Sixty-four possible prepared (assigned) states are in the order of the smallest to the largest binary string from left to right (top to bottom). The sum of the diagonal elements is 247.8% and the sum of the off-diagonal elements is −247.8%, indicating the state-assignment improvement of the trained SNNDs relative to the MFDs. (b) Cross-fidelity matrix of MFD (left) and SNND (right). The correlations between different qubits are largely reduced in the left-hand panel, indicating the crosstalk suppression of SNND.

where $P(x_i|y_j)$ is the probability that the qubit i is assigned as state x on the condition that qubit j is prepared in state y. Ideally, $P\left(1_i \mid 0_j\right) = P\left(0_i \mid 1_j\right) = 0.5$ for $i \neq j$ that the off-diagonal elements are 0s. The cross-fidelity matrices are shown in Fig. 4(b). The trained SNNDs acquire an overall crosstalk mitigation. Further, from the cross-fidelity matrix, we list the mean absolute cross fidelities $\left\langle\left|\mathbb{F}_{j=i\pm n}\right|\right\rangle$ for $n \in [1,5]$ of SNND and MFD in Table III. In general, readout crosstalk is drastically suppressed. Mean absolute value of $\mathbb{F}_{j\neq i}$ is decreased by 80.4%, from 0.0092 to 0.0018.

TABLE III. Mean absolute value of the readout correlations between Qi and Qj ($|i-j|=1,2,3,4,5$) extracted from the cross-fidelity matrix.

	$\langle\lvert\mathbb{F}_{j=i\pm1}\rvert\rangle$	$\langle\lvert\mathbb{F}_{j=i\pm2}\rvert\rangle$	$\langle\lvert\mathbb{F}_{j=i\pm3}\rvert\rangle$	$\langle\lvert\mathbb{F}_{j=i\pm4}\rvert\rangle$	$\langle\lvert\mathbb{F}_{j=i\pm5}\rvert\rangle$
MFD	0.0122	0.0041	0.0085	0.0076	0.0196
SNND	**0.0018**	**0.0016**	**0.0021**	**0.0022**	**0.0013**

IV. DISCUSSION AND OUTLOOK

In Sec. II A, we demonstrate the crosstalk-induced shift of I-Q clouds and the decline of readout fidelity. Here we continue to select Q4 as the target qubit with Q3 as spectator qubit. SNND achieves consistent optimal fidelities on any subset of the states of Q4 (data not shown). We infer that through learning SNND acquires the capability to gather leaked information to other qubits. We assess this hypothesis by monitoring the change in power spectrum of the weighted integration kernel of the demodulation layer [Fig. 5(a)]. In the power spectrum, the conventional weighted kernel has a single peak located at the IF of the readout resonator of Q4. However, for SNND, an additional small peak appears at the IF of the resonator of Q3 (see Appendix A for information of IFs of readout resonators). This implies that, to some extent, the trained SNND obtains access to mitigate crosstalk effect. The enhanced performance of the demodulation layer of SNND is evident [Figs. 5(b) and 5(c)]. The crosstalk-induced shift in the I-Q clouds is almost eliminated in the SNND case.

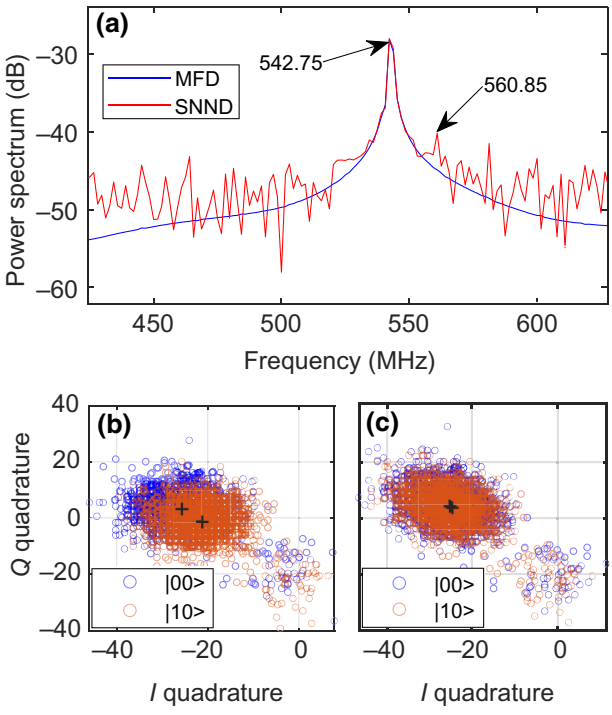

FIG. 5. (a) Power spectrum of weighted kernel D in the first Nyquist zone. For MFD (blue), the main frequency component is only related to the target qubit Q4; for SNND (red), an additional peak appears related to Q3. (b),(c) The I-Q clouds corresponding to the ground state of qubit Q4 involving an influence from Q3 demodulated by (b) the MF-based weighted kernel and (c) the trained weights D of SNND. Here $|\mathrm{Q3,Q4}\rangle$ denotes the state, and black crosses mark the centers of the clouds.

In conclusion, we demonstrate the enhanced readout performance and effective crosstalk suppression of SNND. In addition, SNND with a two-layer architecture with a DSP-based initialization provides a great advantage over training. And SNND is applicable to systems with larger numbers of qubits as long as the readout crosstalk effects are local. Here we also give a brief discussion about a further simplification of SNND. In fact, removing the activation function of the demodulation layer would not have much influence on the performance (see Appendix C); however, by doing this we could even integrate the two layers into one. That is, the output of SNND could be simplified by associative law of matrix multiplication:

$$\text{output} = P(D\vec{x}+d)+c = C\vec{x}+c'. \tag{6}$$

Here C is a row vector with the same dimension as the input vector. The architecture of Eq. (6) is actually a single multiply-accumulator with initial value c'. Consequently, the simplified SNND can be directly transplanted to a field-programmable gate array platform, especially when we have finished the training process on the upper computer. This would be extremely useful for rapid state assignment and real-time feedback control [30,35,36], as it can distinguish the state almost just at the moment the last data point is received.

ACKNOWLEDGMENTS

We thank OriginQ, Inc. for hardware support. This work is supported by the National Key Research and Development Program of China (Grant No. 2016YFA0301700), the National Natural Science Foundation of China (Grants

TABLE IV. Operation parameters of the six-qubit chip. Qubits are detuned from the sweet point ($\omega_{\mathrm{Qubit}}^{\mathrm{idle}}$) to a biased point ($\omega_{\mathrm{Qubit}}^{\mathrm{biased}}$) to limit the residual qubit-qubit interaction. Neighboring readout resonator frequencies differ by approximately 20 MHz. The qubit lifetimes T_1 and Ramsey coherence times $T_2^\star$ are measured at the biased point.

Number	$\omega_{\mathrm{Qubit}}^{\mathrm{idle}}/2\pi$ (GHz)	$\omega_{\mathrm{Qubit}}^{\mathrm{biased}}/2\pi$ (GHz)	$\omega_{\mathrm{Res}}/2\pi$ (GHz)	$\omega_{\mathrm{IF}}/2\pi$ (MHz)	T_1 (μs)	$T_2^\star$ (μs)
1	5.456	5.456	6.424	600.0	21	9
2	4.824	4.824	6.443	580.3	22	11
3	5.359	5.319	6.463	560.95	22	8
4	4.742	4.693	6.481	542.75	17	14
5	5.374	5.215	6.502	521.25	18	9
6	4.684	4.579	6.521	502.05	15	10

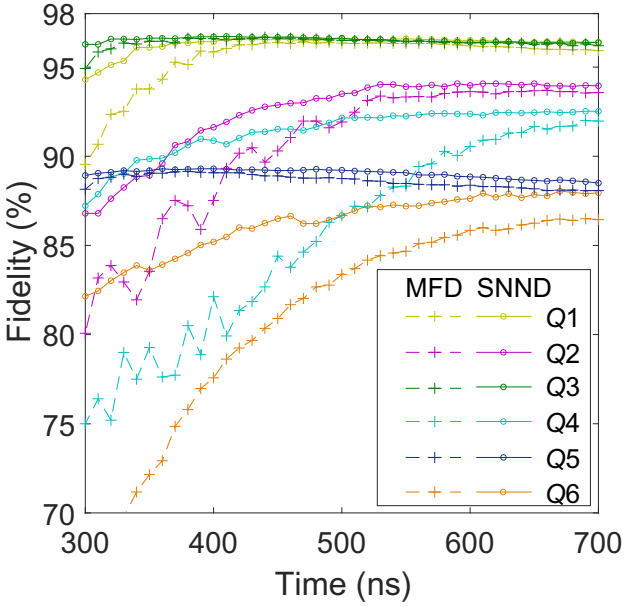

FIG. 6. Readout fidelity versus demodulation time.

No. 12034018 and No. 11625419), the Strategic Priority Research Program of the CAS (Grant No. XDB24030601), and the Anhui Initiative in Quantum Information Technologies (Grant No. AHY080000). This work was partially performed at the USTC Center for Micro and Nanoscale Research and Fabrication.

APPENDIX A: INFORMATION ABOUT THE SIX-QUBIT CHIP

The quantum chip we employ is KF C6-130, which has six Xmon-type qubits. Each qubit is capacitively coupled to its neighboring qubits, with coupling strength $J/2\pi \approx 10$ MHz, and to its quarter-wave readout resonator, with coupling strength $g/2\pi \approx 50$ MHz. All qubits can be individually controlled via an X-control line and Z-control line. All resonators are inductively coupled to a common feedline. The qubit and resonator operation parameters are listed in Table IV.

TABLE V. Optimal demodulation time and fidelity.

	MFD		SNND	
	Time (ns)	Fidelity (%)	Time (ns)	Fidelity (%)
Q1	470	96.42	550	96.60
Q2	660	93.70	610	94.07
Q3	440	96.67	410	96.72
Q4	690	92.00	700	92.52
Q5	380	89.15	400	89.30
Q6	690	86.50	670	88.05

APPENDIX B: OPTIMAL DEMODULATION TIME

In our experiment, we read out all the six qubits simultaneously with a square pulse of 1.5 μs through the common feedline. After an appropriate sampling delay, the transmitted tone is sampled at 1.6 GS/s by an analog-to-digital converter with total length of 700 ns (1120 time points). To find the optimal demodulation time with maximum readout fidelity, we train the SNND (MFD) for each qubit parallelly with different length of demodulation time as shown in Fig. 6. Note that during this calculation, we only use those data sets for which the nontarget qubits are in their ground state (crosstalk effects are not included). Nevertheless, we can still see the superiority of SNND over MFD, especially at shorter time where the I-Q clouds are not well separated. The results are summarized in Table V and the optimal demodulation times are used to obtain the results in the main text.

APPENDIX C: OTHER METHODS BASED ON MF AND LINEAR SNND

In the main text, as discussed in Sec. II A, the discriminant line of MFD is found according to the projection axis

TABLE VI. Readout correlations between Qi and Qj ($|i-j| = 1, 2, 3, 4, 5$) extracted from the cross-fidelity matrix and their mean value. The superscript "0" indicates that the discriminator is trained with training set where nontarget qubits are kept in their ground state, while the superscript "*" indicates that the training set is composed of all permutations of the six qubits. "Th" means that the discriminant line is found via projection and threshold method as in the main text, while "GM" means that the discriminant line is found with machine learning method based on Gaussian mixture model (here we use the function *fitcdiscr* in **MATLAB**). LSNND refers to the linear SNND.

	$n=1$	$n=2$	$\langle\lvert\mathbb{F}_{j=i\pm n}\rvert\rangle$ $n=3$	$n=4$	$n=5$	$\langle\lvert\mathbb{F}_{j\neq i}\rvert\rangle$
MFD^0-Th	0.0122	0.0041	0.0085	0.0076	0.0196	0.0092
MFD*-Th	0.0079	0.0035	0.0055	0.0078	0.0207	0.0071
MFD^0-GM	0.0098	0.0050	0.0044	0.009	0.0201	0.0080
MFD*-GM	0.0061	0.0028	0.0035	0.0092	0.0230	0.0062
SNND	**0.0018**	**0.0016**	**0.0021**	**0.0022**	**0.0013**	**0.0018**
LSNND	0.0020	**0.0016**	0.0029	0.0030	0.0023	0.0022

TABLE VII. Assigned fidelities of each qubit.

		F^{Q1}	F^{Q2}	F^{Q3}	F^{Q4}	F^{Q5}	F^{Q6}
MFD^0-Th	%	96.35	93.02	95.63	90.77	82.54	85.15
MFD^*-Th	%	96.36	93.36	95.64	92.10	82.48	85.96
MFD^0-GM	%	95.87	93.25	95.62	92.10	82.54	85.87
MFD^*-GM	%	96.37	93.48	95.63	92.52	82.54	86.03
SNND	%	**96.54**	93.71	**95.83**	**92.67**	**82.62**	**86.73**
LSNND	%	96.49	**93.73**	95.81	92.66	82.59	86.61

and threshold value. However, instead of projecting the I-Q clouds onto one axis, one can also find an optimum discriminant line with a machine learning method, such as K-Means or Gaussian mixture model in scikit-learn library. Furthermore, the training set can be composed of either all permutations of the qubit states or a specific subset, e.g., nontarget qubits are kept in their ground state during the training of MFD in the main text. Here we take all these variations into consideration, and compare their performance with SNND. We also calculate the performance of linear SNND where the nonlinear activation function is removed. The results are shown in Tables VI and VII. In all cases, SNND outperforms MF-based discriminators, especially in terms of readout crosstalk suppression. Moreover, removing the nonlinear activation function of SNND only has a limited influence and the linear SNND still outperforms MF-based discriminators.

[1] D. DiVincenzo, The physical implementation of quantum computation, Fortschritte der Physik **48**, 771 (2000).

[2] A. Wallraff, D. I. Schuster, A. Blais, L. Frunzio, R.-. S. Huang, J. Majer, S. Kumar, S. M. Girvin, and R. J. Schoelkopf, Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics, Nature **431**, 162 (2004).

[3] R. Bianchetti, S. Filipp, M. Baur, J. M. Fink, M. Göppl, P. J. Leek, L. Steffen, A. Blais, and A. Wallraff, Dynamics of dispersive single-qubit readout in circuit quantum electrodynamics, Phys. Rev. A **80**, 043840 (2009).

[4] J. Gambetta, A. Blais, M. Boissonneault, A. A. Houck, D. I. Schuster, and S. M. Girvin, Quantum trajectory approach to circuit qed: Quantum jumps and the zeno effect, Phys. Rev. A **77**, 012112 (2008).

[5] M. Jerger, S. Poletto, P. Macha, U. Hübner, A. Lukashenko, E. Il'ichev, and A. V. Ustinov, Readout of a qubit array via a single transmission line, EPL (Europhysics Letters) **96**, 40012 (2011).

[6] M. Jerger, S. Poletto, P. Macha, U. Hübner, E. Il'ichev, and A. V. Ustinov, Frequency division multiplexing readout and simultaneous manipulation of an array of flux qubits, Appl. Phys. Lett. **101**, 042604 (2012).

[7] 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).

[8] Y. Chen, D. Sank, P. O'Malley, T. White, and R. Barends *et al.*, Multiplexed dispersive readout of superconducting phase qubits, Appl. Phys. Lett. **101**, 182601 (2012).

[9] T. Walter, P. Kurpiers, S. Gasparinetti, P. Magnard, A. Potočnik, Y. Salathé, M. Pechal, M. Mondal, M. Oppliger, C. Eichler, and A. Wallraff, Rapid High-Fidelity Single-Shot Dispersive Readout of Superconducting Qubits, Phys. Rev. Appl. **7**, 054020 (2017).

[10] J. Heinsoo, C. K. Andersen, A. Remm, S. Krinner, T. Walter, Y. Salathé, S. Gasparinetti, J.-C. Besse, A. Potočnik, A. Wallraff, and C. Eichler, Rapid High-Fidelity Multiplexed Readout of Superconducting Qubits, Phys. Rev. Appl. **10**, 034040 (2018).

[11] L. Ranzani, M. Bal, K. C. Fong, G. Ribeill, X. Wu, J. Long, H.-S. Ku, R. P. Erickson, D. Pappas, and T. A. Ohki, Kinetic inductance traveling-wave amplifiers for multiplexed qubit readout, Appl. Phys. Lett. **113**, 242602 (2018).

[12] F. Arute, K. Arya, R. Babbush, D. Bacon, and J. C. Bardin *et al.*, Quantum supremacy using a programmable superconducting processor, Nature **574**, 505 (2019).

[13] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[14] M. Sarovar, T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, Detecting crosstalk errors in quantum information processors, Quantum **4**, 321 (2020).

[15] K. Rudinger, C. W. Hogle, R. K. Naik, A. Hashim, D. Lobser, D. I. Santiago, M. D. Grace, E. Nielsen, T. Proctor, S. Seritan, S. M. Clark, R. Blume-Kohout, I. Siddiqi, and K. C. Young, Experimental characterization of crosstalk errors with simultaneous gate set tomography, arXiv:2103.09890 (2021).

[16] D. Pitsun, A. Sultanov, I. Novikov, E. Mutsenik, B. Ivanov, A. Matanin, V. Polozov, E. Malevannaya, A. Ivanov, G. Fedorov, K. Delfanazari, I. Rodionov, and E. Il'ichev, Cross Coupling of a Solid-State Qubit to an Input Signal due to Multiplexed Dispersive Readout, Phys. Rev. Appl. **14**, 054059 (2020).

[17] P. Mundada, G. Zhang, T. Hazard, and A. Houck, Suppression of Qubit Crosstalk in a Tunable Coupling Superconducting Circuit, Phys. Rev. Appl. **12**, 054023 (2019).

[18] F. Hirayama, T. Irimatsugawa, H. Yamamori, S. Kohjiro, A. Sato, S. Nagasawa, D. Fukuda, H. Sasaki, M. Hidaka, Y. Sato, M. Ohno, and H. Takahashi, Interchannel crosstalk and nonlinearity of microwave squid multiplexers, IEEE Trans. Appl. Supercond. **27**, 1 (2017).

[19] O. Noroozian, P. K. Day, B. H. Eom, H. G. Leduc, and J. Zmuidzinas, Crosstalk reduction for superconducting microwave resonator arrays, IEEE Trans. Microw. Theory Tech. **60**, 1235 (2012).

[20] G. Turin, An introduction to matched filters, IRE Trans. Inf. Theory **6**, 311 (1960).

[21] 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).

[22] C. A. Ryan, B. R. Johnson, J. M. Gambetta, J. M. Chow, M. P. da Silva, O. E. Dial, and T. A. Ohki, Tomography via correlation of noisy measurement records, Phys. Rev. A **91**, 022118 (2015).

[23] C. C. Bultink, B. Tarasinski, N. Haandbæk, S. Poletto, N. Haider, D. J. Michalak, A. Bruno, and L. DiCarlo, General method for extracting the quantum efficiency of dispersive qubit readout in circuit qed, Appl. Phys. Lett. **112**, 092601 (2018).

[24] E. Magesan, J. M. Gambetta, A. D. Córcoles, and J. M. Chow, Machine Learning for Discriminating Quantum Measurement Trajectories and Improving Readout, Phys. Rev. Lett. **114**, 200501 (2015).

[25] M. Reagor, C. B. Osborn, N. Tezak, A. Staley, G. Prawiroatmodjo, M. Scheer, N. Alidoust, E. A. Sete, N. Didier, and M. P. da Silva *et al.*, Demonstration of universal parametric entangling gates on a multi-qubit lattice, Sci. Adv. 4, eaao3603 (2018).

[26] E. Flurin, L. S. Martin, S. Hacohen-Gourgy, and I. Siddiqi, Using a Recurrent Neural Network to Reconstruct Quantum Dynamics of a Superconducting Qubit from Physical Observations, Phys. Rev. X **10**, 011006 (2020).

[27] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A quantum engineer's guide to superconducting qubits, Appl. Phys. Rev. **6**, 021318 (2019).

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

[29] R. Barends, J. Kelly, A. Megrant, D. Sank, and E. Jeffrey *et al.*, Coherent Josephson Qubit Suitable for Scalable Quantum Integrated Circuits, Phys. Rev. Lett. **111**, 080502 (2013).

[30] C. A. Ryan, B. R. Johnson, D. Ristè, B. Donovan, and T. A. Ohki, Hardware for dynamic quantum computing, Rev. Sci. Instrum. **88**, 104703 (2017).

[31] G. Cybenko, Approximation by superpositions of a sigmoidal function, Math. Control Signals Syst. **2**, 303 (1989).

[32] S. Ruder, An overview of gradient descent optimization algorithms, arXiv:1609.04747 (2016).

[33] N. Qian, On the momentum term in gradient descent learning algorithms, Neural Netw. **12**, 145 (1999).

[34] T. P. Vogl, J. K. Mangis, A. K. Rigler, W. T. Zink, and D. L. Alkon, Accelerating the convergence of the back-propagation method, Biol. Cybern. **59**, 257 (1988).

[35] S. McHugh, B. A. Mazin, B. Serfass, S. Meeker, K. O'Brien, R. Duan, R. Raffanti, and D. Werthimer, A readout for large arrays of microwave kinetic inductance detectors, Rev. Sci. Instrum. **83**, 044702 (2012).

[36] Y. Salathé, P. Kurpiers, T. Karg, C. Lang, C. K. Andersen, A. Akin, S. Krinner, C. Eichler, and A. Wallraff, Low-Latency Digital Signal Processing for Feedback and Feedforward in Quantum Computing and Communication, Phys. Rev. Appl. **9**, 034011 (2018).