Machine Learning for Discriminating Quantum Measurement Trajectories and Improving Readout

Easwar Magesan, Jay M. Gambetta, A. D. Córcoles, and Jerry M. Chow
IBM T.J. Watson Research Center, Yorktown Heights, New York 10598, USA
(Received 18 November 2014; published 18 May 2015)

Current methods for classifying measurement trajectories in superconducting qubit systems produce fidelities systematically lower than those predicted by experimental parameters. Here, we place current classification methods within the framework of machine learning (ML) algorithms and improve on them by investigating more sophisticated ML approaches. We find that nonlinear algorithms and clustering methods produce significantly higher assignment fidelities that help close the gap to the fidelity possible under ideal noise conditions. Clustering methods group trajectories into natural subsets within the data, which allows for the diagnosis of systematic errors. We find large clusters in the data associated with T_1 processes and show these are the main source of discrepancy between our experimental and ideal fidelities. These error diagnosis techniques help provide a path forward to improve qubit measurements.

DOI: 10.1103/PhysRevLett.114.200501 PACS numbers: 03.67.Lx, 03.65.Ta, 85.25.-j

Maximizing the information one can extract from a physical system requires the ability to perform accurate measurements. Our goal in this Letter is to provide methods for diagnosing measurement errors and increasing fidelities by using various machine learning (ML) algorithms. An important application of these techniques is in quantum information processing, where highly accurate operations and measurements are required to perform fault-tolerant information processing in the presence of noise [1]. We apply our methods in a superconducting qubit measurement system, and we anticipate that the generality of these techniques can be useful in a broader class of systems. Superconducting quantum bits (qubits) are becoming increasingly promising for experimentally demonstrating quantum protocols due to their long coherence times [2–4], high-fidelity multiqubit gate operations [5], and the ability to perform single-shot measurements [6–9] in a circuit quantum electrodynamics (CQED) architecture [10]. In the dispersive measurement scheme of CQED, a superconducting anharmonic oscillator, such as a transmon [11], is coupled to a resonator, producing a state-dependent shift of the resonator frequency. This allows for qubit measurements by driving the resonator and recording the output trajectory [12] in phase (I-Q) space. In practice, significant sources of random noise and systematic effects, such as T_1 processes (spontaneous decay), make single-shot trajectories appear complex and difficult to distinguish. There has been significant progress in reducing error rates and measurement times [13]; however, managing, classifying, and extracting useful information from the trajectory data is extremely important for improving readout as these systems scale to larger networks.

The outline of the Letter is as follows. We begin by describing the experimental system used to create the measurement data as well as how we characterize the measurement performance by the assignment fidelity $\mathcal{F}_a$ [defined in Eq. (1)]. First, we analyze the data using the current method, which is a simplified version of linear discriminant analysis (LDA), and obtain $\mathcal{F}_a = 0.9586$. LDA is the simplest ML classification algorithm and finds the plane that optimally separates the I-Q data trajectories under idealized assumptions of symmetric and Gaussian noise at each point in time. Next, we use the quadratic extension of LDA called quadratic discriminant analysis (QDA), which allows for noise asymmetry, and find noticeable improvement in the assignment fidelity. Finally, we remove assumptions on the noise and approach the problem from a purely geometric viewpoint using support vector machines (SVMs). The nonlinear SVM provides the largest improvement giving $\mathcal{F}_a = 0.9821$ ($\sim$2.4% increase), which indicates nonlinear effects such as T_1 events are likely present. To verify this, and to understand details of the noise, we use ML clustering algorithms to find natural subclasses in the data. We find a large subclass corresponding to T_1 events which validates our hypothesis. Accounting for these events, we find assignment fidelities much closer to those that should be attainable in our system under the ideal noise conditions assumed for the optimality of LDA.

Let us briefly make a few points about using ML methods. First, the methods we present here can be useful in a much broader context. Any measurement scheme that produces patterns in a geometric space can potentially benefit from more advanced ML methods. Investigating the applicability to different systems will depend on the details of each situation. Second, these methods are applicable even if we are trying to improve higher fidelity measurements than those considered here. The key is that these methods can be tailored according to the types of noise present. Third, ML methods have also been applied to other problems in quantum information such as phase estimation [14] and asymptotic state estimation [15].

TABLE I. System parameters.

Transmon 0–1 frequency	$\omega/2\pi = 5.415$ GHz
Resonator frequency	$\omega_R/2\pi = 6.693$ GHz
Transmon-resonator coupling	$g/2\pi = 42.3$ MHz
Resonator linewidth	$\kappa/2\pi = 1.21$ MHz
T_2^{echo}	22 μs
T_1	29 μs

Our system is a single transmon qubit (Q4) coupled to a readout resonator in a lattice of four superconducting qubits [16] (full details of this experiment are in Ref. [16] and a diagram of the four-qubit setup is given in the Supplemental Material [17]). The main parameters of the single-qubit system are provided in Table I. The measurement framework is the dispersive limit of CQED where $-(g/\Delta) \ll 1$ ($\Delta = \omega - \omega_R$). A full discussion of this framework is given in [10,17]. The general idea is that there is a qubit state-dependent dispersive shift of the readout resonator frequency, so driving near the bare resonator frequency corresponds to a quantum nondemolition measurement of the qubit state. A circuit diagram of the single-qubit apparatus is shown in Fig. 1. The control pulse is shown in black (top) and the readout pulse in red (bottom). The readout pulse is split into two signals, one of which is sent to the device while the other is mixed with the post-amplification signal. A single drive line enters the device and amplification of the output signal is provided by the (low-noise) Josephson parametric amplifier and HEMT amplifier. The I-Q field quadratures are measured and recorded, producing single-shot time-dependent data. An example of this data for a $|0\rangle$ state preparation is given in Fig. 2. I-Q plots for single-shot $|0\rangle$ and $|1\rangle$ preparations, as well as the means over these preparations, are given in Fig. 3. We see that, once the measurement drive is turned on, photons populate the cavity leading to a Stark shift of the qubits and an increase in the mean separation between the trajectories. Each single-shot trajectory can be complicated due to noise; however, there are enough shots to ensure smooth and well-separated means.

The assignment fidelity

$$\mathcal{F}_a = 1 - (\mathbb{P}[0|1] + \mathbb{P}[1|0])/2, \tag{1}$$

is a standard metric for characterizing how well a measurement assigns outcomes. Here, $\mathbb{P}[0|1]$ ($\mathbb{P}[1|0]$) is the probability of obtaining outcome "0" ("1") given the system was prepared in $|1\rangle$ ($|0\rangle$) and so $\mathcal{F}_a \in [0,1]$. Our data consist of 51 200 single-shot trajectories (shots), half initially prepared in $|0\rangle$ and the other half in $|1\rangle$ (denote these classes by C_0 and C_1). The first half of the trajectories is used as a training set for classification on the second half. The total measurement time T is 2.6 μs and $[0,T]$ is discretized into 163 time-points, so trajectories are represented by vectors $x \in \mathbb{R}^{326}$ (let $M = 326$) where the first (last) 163 entries correspond to the real (imaginary) parts of

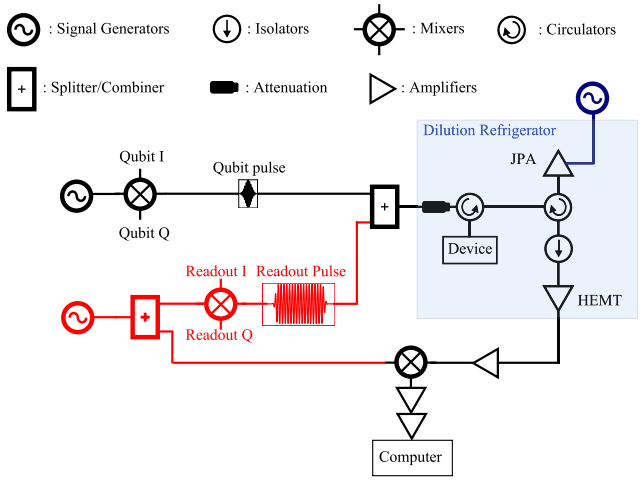

FIG. 1 (color online). Circuit diagram of the single-qubit setup. See text for details.

the trajectory. Hence, each $x(j)$ can be viewed as a real-valued random variable. The mean trajectories and covariance matrices for each class are denoted μ_0, μ_1 (see Fig. 3) and Σ_0, Σ_1, respectively.

The current method of classifying trajectories [22] is based on LDA [24] and assumes the noise is highly idealized; the noise at each time is assumed to be (1) Gaussian distributed, (2) uncorrelated with noise at other times, and (3) symmetric between C_0, C_1. Alternatively, this can be phrased as Σ_0 and Σ_1 are Gaussian, diagonal, and equal. Under these assumptions, x is associated to

$$f_{\text{LDA}}(x) = x^T[\Sigma_0^{-1}(\mu_0 - \mu_1)], \tag{2}$$

which is then assigned as 0 or 1 according to an appropriate threshold. Using this method in our experiment gives $\mathcal{F}_a = 0.9586$.

In reality, the noise is far from these ideal conditions and one of the main goals of this Letter is to deal with these

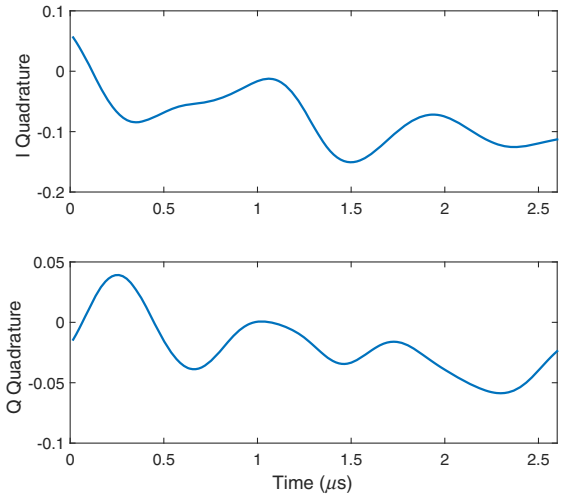

FIG. 2 (color online). Measured output of I and Q quadratures for state $|0\rangle$ single-shot trajectory.

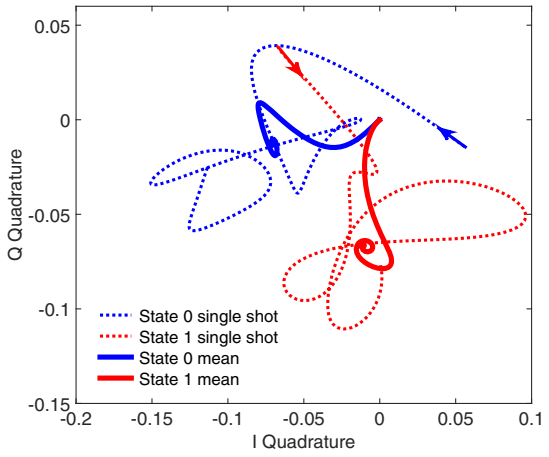

FIG. 3 (color online). Mean trajectories and single shots for $|0\rangle$ (blue) and $|1\rangle$ (red) preparations. The $|0\rangle$ ($|1\rangle$) single-shot trajectory [blue (red) dotted line] has arrows pointing up (down) and to the left (right). The mean trajectories of $|0\rangle$ (blue solid line) and $|1\rangle$ (red solid line) have steady states of $\sim(-0.07, -0.02)$ and $(-0.01, -0.07)$.

more realistic scenarios. Before doing this, we can ask what fidelity we would expect from the parameters of our system if the noise did satisfy these ideal conditions. As shown in [17], this ideal fidelity, denoted $\mathcal{F}_{\mathrm{id}}$, is

$$\mathcal{F}_{\mathrm{id}} = 0.9999 \pm 0.0001. \quad (3)$$

Hence, there is a large discrepancy between $\mathcal{F}_{\mathrm{id}}$ and $\mathcal{F}_a$ that can be due to effects such as state-preparation errors and non-Gaussian or nonlinear noise. This discrepancy motivates us to investigate better methods of classifying trajectories.

Let us relax the unrealistic assumption of noise symmetry between C_0 and C_1. In this case, the optimal method is QDA [23], and each trajectory x is mapped to

$$f_{\mathrm{QDA}}(x) = -\frac{1}{2}x^T[\Sigma_0^{-1} - \Sigma_1^{-1}]x + x^T[\Sigma_0^{-1}\mu_0 - \Sigma_1^{-1}\mu_1], \quad (4)$$

and then assigned "0" or "1" according to an appropriate threshold. We computed $\mathcal{F}_a$ using the "fitcdiscr" function in MATLAB for four different methods: LDAD, LDA, QDAD, and QDA ("d" represents diagonal covariance matrix and LDAD is the method of [22]). The results are in the second column of Table II. Not surprisingly, we find QDAD improves upon LDAD, and allowing nondiagonal covariance matrices produces higher $\mathcal{F}_a$. The values in Table II are the sample means from 100 repetitions. The sample variances σ^2 are $\sim 1 \times 10^{-8}$ indicating stable or reproducible results.

A value of $\mathcal{F}_a$ for QDA was not attainable due to singular covariance matrices, which is a result of overfitting the data (having more variables than required from the correlation time in the trajectories). To remedy this, we perform dimensionality reduction using principal component analysis (PCA) [25] and find 99.9% of the variance in the data can be accounted for in a subspace of dimension ~20 (noise correlation time ~85–180 ns). Results with a PCA preprocessing step (using "princomp" in MATLAB) are in the third column of Table II. Not surprisingly QDA provides the highest $\mathcal{F}_a$ out of all cases considered.

TABLE II. Assignment fidelities for discriminant analysis methods.

Method	All time points	PCA
LDAD	0.9586	0.9557
LDA	0.9701	0.9586
QDAD	0.9627	0.9648
QDA	…	0.9712

These classification methods assume Gaussian noise and better methods are needed to deal with realistic noise. We approach this in two ways. The first is via the SVM [26,27], which requires no assumptions on the noise and can be extended to nonlinear discriminating surfaces. The second is to utilize "clustering" methods in ML to naturally group the data from which we perform multiclass classification.

The linear SVM is a quadratic program based on maximizing the minimum distance of a data point to a hyperplane separating the data. The nonlinear SVM is derived by defining a kernel that maps the data to a higher-dimensional space. The linear SVM in the higher-dimensional space allows for nonlinear discrimination in $\mathbb{R}^M$. Because of its generality and simplicity, we chose a radial basis function kernel.

We implemented the SVM using the MATLAB "fitcsvm" function and classification was repeated 100 times. The mean values with the optimal soft-margin parameter are contained in Table III (see [17] for details). Sample variances σ^2 of $\mathcal{F}_a$ are approximately 1.9×10^{-8}. The nonlinear SVM produces the highest assignment fidelity out of all methods considered, indicating nonlinear effects are present. We hypothesize the main factor producing the nonlinearity is T_1 events.

Our second method for implementing a nonlinear classifier combines classification and clustering algorithms. Clustering naturally groups the data into subsets and is "unsupervised" since it requires no training data. We implement k-means clustering [17,28] and anticipate similar results can be obtained with other standard methods such as heirarchical clustering. We used the MATLAB "kmeans" function to find $k = 3$ clusters in each of C_0

TABLE III. Assignment fidelities for SVM methods.

Method	All time-points	PCA
Linear SVM	0.9753	0.9571
Nonlinear SVM	0.9821	0.9739

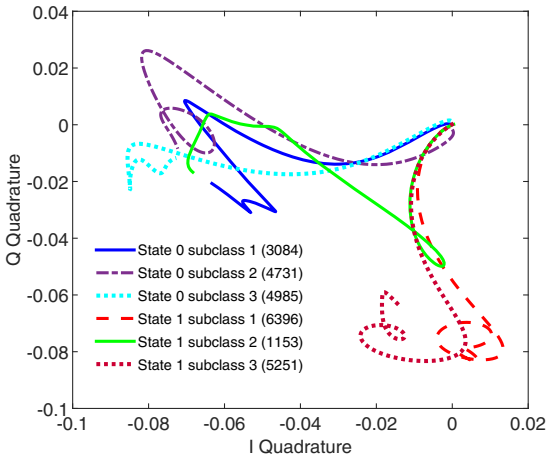

FIG. 4 (color online). Subclasses found from k-means algorithm. C_0 and C_1 have three subclasses, the trajectory representing each subclass is the mean over all subclass trajectories. Subclasses of C_0 (purple dashed-dotted line, blue solid line, light blue dotted line) initially move up and left. Subclasses of C_1 (red dashed line, green solid line, dark red dotted line) initially move down and right. The T_1 subclass (green solid line) of C_1 initially moves down and right but abruptly changes its path up and left. Legend numbers are subclass sizes.

and C_1. We chose $k = 3$ to take into account both variance and systematic effects. The mean trajectories and size of the six subclasses are given in Fig. 4. C_0 is split relatively evenly into the subclasses $S_{0,1}$, $S_{0,2}$, $S_{0,3}$ (blue colored trajectories) that mainly capture variance. We do not see a C_0 subclass corresponding to ground state heating; however, we implemented k means for $k = 7$ and found a heating subclass of size ~230 (see Fig. 2 in [17]).

C_1 has strikingly different properties as subclass $S_{1,2}$ (green solid line) is comprised of T_1 processes. $S_{1,1}$ and $S_{1,3}$ (red trajectories) are similar in size and mainly capture variance. The key point is that we have found explicit shot indices for T_1 events. We verified that $S_{1,2}$ is comprised of

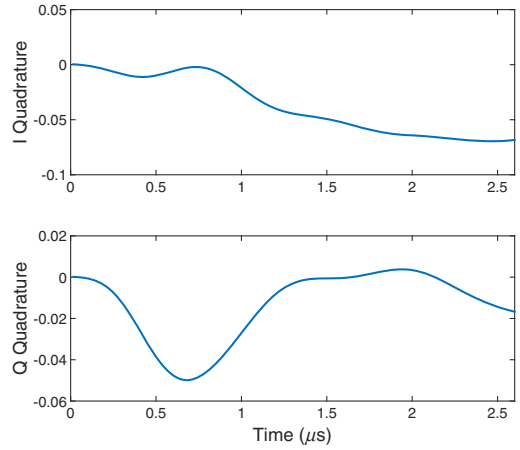

FIG. 5 (color online). I and Q quadratures for T_1 subclass mean trajectory.

TABLE IV. Assignment fidelities from multi-class classification.

Method	All time-points	PCA
Multi-LDA	0.9768	0.9689
Multi-SVM	0.9784	0.9717
TotalBoost	0.9527	0.9413
RUSBoost	0.9788	0.9723

T_1 trajectories by performing k means with $k = 4$ (see Fig. 3 in [17]). From Fig. 4, ~9% of the $|1\rangle$ preparations result in a T_1 event, which is consistent with the percentage calculated from system parameters [17], $1-e^{-2.6/29} \sim 8.6\%$. The I and Q quadratures of the T_1 subclass are contained in Fig. 5 where the T_1 jump is seen most clearly in Q.

To perform classification, we lift the T_1 subclass $S_{1,2}$ to a class C_2 of its own, redefine $C_1 = S_{1,1} \cup S_{1,3}$, keep C_0 as before, and perform multiclass classification on C_0, C_1, and C_2. We implemented four multiclass algorithms in MATLAB: multiclass LDA, multiclass SVM, "TotalBoost," and "RUSBoost." The latter two are examples of boosting algorithms [29] and RUSBoost [30] is particularly useful since it is tailored to the case of one class (here C_2) being significantly smaller than the rest. The results are in Table IV. We again see an increase in $\mathcal{F}_a$ over the discriminant analysis methods of Table II. Not surprisingly, RUSBoost provides the most significant increase. We repeated the k-means algorithm 50 times with random initializations and found it to be stable (σ^2 of $\mathcal{F}_a \sim 3 \times 10^{-6}$).

Out of all methods considered, nonlinear SVMs produce the greatest increase in $\mathcal{F}_a$ (0.9586 to 0.9821). All methods are relatively stable with reproducible assignment fidelities (each method was repeated ~100 times; sample means of $\mathcal{F}_a$ are the table values and sample variances are $\sim 1 \times 10^{-8}$).

While we have improved $\mathcal{F}_a$ to 0.9821, we are still far from $\mathcal{F}_{\text{id}} = 0.9999$. We hypothesize much of the remaining discrepancy comes from T_1 events. To investigate this, we propose the simple diagnostic test of replacing each T_1 event from the k-means algorithm with a random trajectory from $S_{1,2} \cup S_{1,3}$. This provides a measure of $\mathcal{F}_a$ when T_1 is negligible. The means of 100 samples for each classification method are in Table V (variances are $\sim 1 \times 10^{-8}$). Nonlinear SVM produces the highest value of $\mathcal{F}_a$, however, for all methods $\mathcal{F}_a > 0.99$, which is more consistent

TABLE V. Assignment fidelities with replacement of T_1 events.

Method	All time points	PCA
LDAD	0.9920	0.9909
LDA	0.9921	0.9928
QDAD	0.9918	0.9908
QDA	…	0.9927
Linear SVM	0.9936	0.9943
Nonlinear SVM	0.9945	0.9949

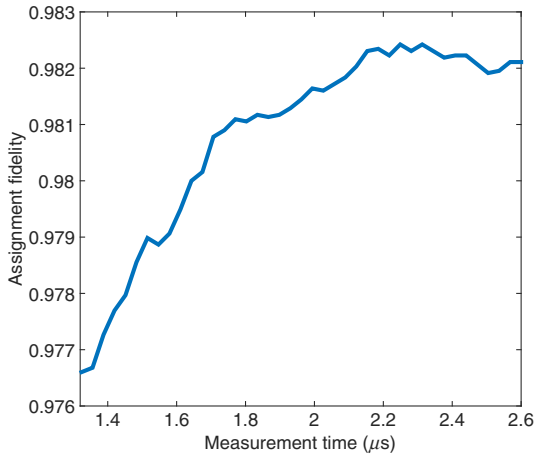

FIG. 6 (color online). Varying measurement time.

with $\mathcal{F}_{\mathrm{id}} = 0.9999$. This confirms T_1 events are the significant reason for not reaching $\mathcal{F}_{\mathrm{id}}$.

One attempt to reduce the significance of T_1 is to reduce T; however, this implies the trajectories will spend less time near their steady states, and noise variance will dominate. To observe this, we truncated the trajectories to different T and calculated $\mathcal{F}_a$ using the nonlinear SVM. From Fig. 6, $T = 2.6$ μs appears close to optimal. A much shorter measurement time of $\sim$1.2 μs (not shown in Fig. 6) is needed to achieve an assignment fidelity of 0.9586, which, we recall, is the value obtained using the simplest ML method of [22]. This is a strong message that better classifiers can allow for shorter measurement times. Longer measurement times than the current 2.6 μs decrease $\mathcal{F}_a$ due to an increase in T_1 events.

To conclude, we have utilized ML to understand and improve the readout in a superconducting system. More sophisticated classification algorithms can potentially allow for shorter measurement times and increase assignment fidelities. Nonlinear SVMs provided the largest increase in assignment fidelity, 0.9586 to 0.9821 ($\sim$2.4%). Clustering helped diagnose the prevalence of systematic effects by finding clusters in the data corresponding to single-shot identification of heating and T_1 effects. We verified that T_1 events are a significant source of error as the assignment fidelity increases to 0.9945 when the T_1 cluster is replaced with typical trajectories. This is more consistent with the ideal fidelity, and the remaining discrepancy can be due to effects such as heating and state-preparation errors. Moving forward, we expect these methods will help provide insight for improving readout, especially when nonlinear and non-Gaussian effects are present.

We acknowledge support from ARO under Contract No. W911NF-14-1-0124 and IARPA under Contract No. W911NF-10-1-0324. We acknowledge helpful discussions with Oliver Dial, Stefan Filipp, Blake Johnson, Jim Rozen, Colm Ryan, Marcus Silva, and Matthias Steffen.

[1] P. Shor, in *Proceedings of the 37'th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE Press, Burlington, VT, 1996).
[2] H. Paik *et al.*, Phys. Rev. Lett. **107**, 240501 (2011).
[3] J. B. Chang *et al.*, Appl. Phys. Lett. **103**, 012602 (2013).
[4] R. Barends *et al.*, Phys. Rev. Lett. **111**, 080502 (2013).
[5] R. Barends *et al.*, Nature (London) **508**, 500 (2014).
[6] F. Mallet, F. R. Ong, A. Palacios-Laloy, F. Nguyen, P. Bertet, D. Vion, and D. Esteve, Nat. Phys. **5**, 791 (2009).
[7] N. Bergeal, F. Schackert, M. Metcalfe, R. Vijay, V. E. Manucharyan, L. Frunzio, D. E. Prober, R. J. Schoelkopf, S. M. Girvin, and M. H. Devoret, Nature (London) **465**, 64 (2010).
[8] J. E. Johnson, C. Macklin, D. H. Slichter, R. Vijay, E. B. Weingarten, J. Clarke, and I. Siddiqi, Phys. Rev. Lett. **109**, 050506 (2012).
[9] D. Ristè, J. G. van Leeuwen, H.-S. Ku, K. W. Lehnert, and L. DiCarlo, Phys. Rev. Lett. **109**, 050507 (2012).
[10] A. Blais, R.-S. Huang, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. A **69**, 062320 (2004).
[11] 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, Phys. Rev. A **76**, 042319 (2007).
[12] J. Gambetta, A. Blais, M. Boissonneault, A. A. Houck, D. I. Schuster, and S. M. Girvin, Phys. Rev. A **77**, 012112 (2008).
[13] E. Jeffrey *et al.*, Phys. Rev. Lett. **112**, 190504 (2014).
[14] A. Hentschel and B. C. Sanders, Phys. Rev. Lett. **104**, 063603 (2010).
[15] M. Gu and W. Kotowski, New J. Phys. **12**, 123032 (2010).
[16] A. D. Córcoles *et al.*, arXiv:1410.6419.
[17] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.114.200501 for information on measurements in CQED, amplifier theory, and ML algorithms, which includes Refs. [18–21].
[18] M. Boissonneault, J. M. Gambetta, and A. Blais, Phys. Rev. A **79**, 013819 (2009).
[19] J. Gambetta, A. Blais, D. I. Schuster, A. Wallraff, L. Frunzio, J. Majer, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. A **74**, 042318 (2006).
[20] C. M. Caves, Phys. Rev. D **26**, 1817 (1982).
[21] D. Aloise, A. Deshpande, P. Hansen, and P. Popat, Mach. Learn. **75**, 245 (2009).
[22] C. A. Ryan, B. R. Johnson, J. M. Gambetta, J. M. Chow, M. P. da Silva, O. E. Dial, and T. A. Ohki, Phys. Rev. A **91**, 022118 (2015).
[23] T. M. Cover, IEEE Trans. Electron. Comput. **EC-14**, 326 (1965).
[24] R. A. Fisher, Annals of Eugenics **7**, 179 (1936).
[25] K. Pearson, Philos. Mag. **2**, 559 (1901).
[26] B. E. Boser, I. M. Guyon, and V. N. Vapnik, in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT '92* (ACM, New York, 1992), pp. 144–152.
[27] C. Cortes and V. Vapnik, Mach. Learn. **20**, 273 (1995).
[28] S. Lloyd, IEEE Trans. Inf. Theory **28**, 129 (1982).
[29] C. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2007).
[30] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans **40**, 185 (2010).