We demonstrate the application of pattern recognition algorithms via hidden Markov models (HMM) for qubit readout. This scheme provides a state-path trajectory approach capable of detecting qubit state transitions and makes for a robust classification scheme with higher starting state assignment fidelity than when compared to a multivariate Gaussian (MVG) or a support vector machine (SVM) scheme. Therefore, also eliminating the qubit-dependant readout time optimization requirement with current schemes. Using a HMM state discriminator we estimate fidelities reaching the ideal limit. Unsupervised learning gives access to transition matrix, priors, and IQ distributions, providing a toolbox for studying qubit state dynamics during strong projective readout.

I. INTRODUCTION

Quantum processors employing superconducting qubits are now reaching new milestones in their simulation [1–4] and computational capabilities [5]. There are numerous technical challenges in the implementation of a fault tolerant quantum processor, but at the core is the ability to generate high fidelity gates [6, 7], perform quantum error correction [8, 9], and the ability to make high-fidelity qubit readout measurements [10]. In particular, high-fidelity single shot qubit readout enables faster quantum protocols while simultaneously allowing for reduced errors in their characterization. Apart from improving $T_1$ times of superconducting qubits [11, 12], optimizing hardware design and configuration [10], and invoking new qubit-cavity coupling schemes [13], readout fidelity may be improved by applying classification schemes utilizing machine learning algorithms [14, 15].

Here we demonstrate the application of pattern recognition algorithms, via hidden Markov models [16], to the heterodyned readout signal of a superconducting qubit [17]. The Markov structure allows for a state-path trajectory approach by discretizing each shot into a sequence of uncorrelated segments. The result is a robust starting state classification scheme with higher fidelity than when compared with multivariate Gaussian (MVG) and support vector machines (SVM) classifiers [18]. The advantage arises from the ability to detect transitions with high probability and, thus, circumvent measurement obfuscation caused by qubit state relaxation. In addition, the application of hidden Markov models for qubit readout can naturally be extended to multi-level qudit systems. Unsupervised learning with hidden Markov models provide the capability of extracting distribution parameters, transition matrices, and starting state probabilities (priors), therefore, providing a valuable toolbox for qubit readout and measurement error correction [19, 20].

This paper is organized as follows. First, an example illustrating the evolution of the readout signal in the IQ plane is presented. Followed by a description of the experimental system used to generate the experimental data. We continue with a brief description of the MVG and SVM classifiers and define the fidelity metrics before detailing the implementation of the hidden Markov model (HMM) classifier. Next, we extract the statistical variations associated with training HMMs and calculate classification errors. Finally, we calculate the readout fidelity of a HMM state classifier and compare it with the ideal fidelity metric defined in reference [14].

Random noise and qubit decay ($T_1$) processes reduce readout fidelity. In figure 1 the trajectory of a single shot measurement for two coupled qubits in a 3D CQED system [21] is tracked in the IQ plane. For reference, figure 1 also includes a Bayes classifier trained on several single shots for each qubit state. The contour lines represent the probability distributions learned with a general mixture model. From the running average of the heterodyned signal (colored line) we see the signal starts near the prepared $|0,1\rangle$ state (purple), wanders around the IQ plane, and finally decays to the ground state $|0,0\rangle$ (yellow). Integration over the total readout time, denoted by the star marker, illustrates that this shot would had been classified to state $|0,1\rangle$ with low probability. This example illustrates that, apart from optimizing hardware parameters [10], choosing an appropriate readout integration time plays an important role in mitigating qubit relaxation.
FIG. 2. Summary of Qubit Readout. (A) Factors determining readout fidelity begin with the quantum hardware; Hamiltonian parameters, and SNR. (B) Apart from hardware, fidelity may be improved at the demodulation stage. Conventional demodulation scheme requires tuning to an optimal integration time in which the distributions are approximately Gaussian (C). (D) Regression analysis can be used to train models for improved classification performance. (F) Classification errors are extracted from the misclassification probabilities.

II. METHODS

The experimental quantum platform is a 3D cavity QED system utilizing a strong-projective dispersive measurement scheme. In this platform the qubit state information is encoded in the amplitude and phase of the readout signal. For a single shot measurement, a readout pulse of width $W$ and radio frequency $\omega_r$ is applied to the readout resonator, filtered, and amplified. At this stage, as indicated in figure 2A, a fidelity limit is imposed by the signal to noise ratio (SNR) and the CQED system parameters. For the purpose of this work, these parameters are to be associated with hardware and, therefore, assumed to be fixed after some initial optimization. Next, the amplified signal is mixed down, with a RF-mixer and a local oscillator tone at frequency $\omega_{LO}$, to an intermediate frequency $\Omega_{IF} = \omega_r - \omega_{LO}$. The IF signal is then digitally decomposed in quadrature and de-modulated by picking out the Fourier component at $\Omega_{IF}$.

The demodulation process integrates both the in-phase and quadrature components for a total integration time $T_{int}$ (figure 2B). As noted in figure 2A at this stage the integration time plays an important role in readout fidelity, as qubit decay processes obfuscate the readout signal for relatively long measurement times. For this reason, an optimal integration time is determined. The demodulated result of each single shot is a coordinate $(I,Q)$, and the histogram of several shots forms the probability distribution represented in the IQ-plane as illustrated in figure 2C.

Preparing a state classifier involves a training procedure in which a training dataset is used to learn the model parameters (figure 2D). For the multivariate Gaussian (MVG) model the training solely consist of learning the mean $\mu$ and covariance matrix $\Sigma$. In this case the distributions are allowed to be elongated towards one axis, and the goodness-of-fit of the MVG model is severely affected for longer readout times as the distributions become skewed and non-Gaussian. Support vector machines (SVM) provide both supervised and unsupervised learning capabilities. Because SVMs are geometric models and can, therefore, partially circumvent random noise processes, they provide excellent classification results by finding an optimal hyperplane which provides maximum separation between clusters. However, SVMs also succumb to $T_1$ effects, and so, for maximum performance an optimal readout time must be used as well.

A. Fidelity Metrics

Before describing HMMs we define the fidelity metrics used herein, and remind the reader we are operating in the strong projective measurement limit. In the absence of qubit state decay and assuming Gaussian noise, the IQ distributions for each readout state are Gaussian. The ideal fidelity $F_{id}$, defined by the misclassification probability, is computed from the integration of the overlapped regions of the projected Gaussian probability distributions;

$$F_{id} = \frac{1}{2} \left(1 + \text{erf} \left( \sqrt{\frac{R}{8}} \right) \right).$$

Here, $R$ is a measure of the separation between the two distributions in question given by

$$R = \frac{\left(\langle S_0 \rangle - \langle S_1 \rangle \right)^2}{\text{var}(S)},$$

where $S$ denotes the measurement outcome after the integration of the signal, i.e. the demodulated value, and $\text{var}(S)$ is the variance of $S$. In practice, the IQ probability distributions are also well modeled by Gaussian distributions so long we operate in the limit where the integration time is much less than the relaxation time ($T_1$). However, for relatively long integration times ($T_{int} \gtrsim 5\% T_1$), the distributions are skewed by relaxation transitions and a Gaussian model is no longer adequate.

When using classification schemes, the probability distributions of each readout state may not be Gaussian and...
the method of calculating fidelity described above is not adequate. Instead, in classification systems fidelity may be assessed through various statistical figures of merit \[ F_{\lambda} \]. Here we use the assignment fidelity, \( F_{\lambda} \), to compare the MVG, SVM, and HMM classifiers. The assignment fidelity is adapted from a more general distortion measure (infidelity) introduced by Shannon’s information theory \[ 26 \], and gives a quantitative measure for performance of classification systems which generalizes to the confusion matrix formalism illustrated in figure 2F. For the two state case the assignment fidelity is \[ F_{\lambda} = 1 - \frac{1}{2}(P(0|1) + P(1|0)), \] (3)

where \( P(i|j) \) is the probability the label \( i \) is assigned when state \( j \) is prepared. Note, that this definition makes \( P(i|j) \) dependent on the preparation (gate) fidelity, and the starting state population at the start of the readout measurement.

B. Hidden Markov Models for Qubit Readout

Hidden Markov models are a special case of Bayesian networks in which underlying (hidden) stochastic Markov processes yield observations which themselves are associated with a probability distribution. The relevant parameters of a HMM are \[ 10 \]:

\[
\begin{align*}
T &= \text{length of observation sequence} \\
O &= (O_0, O_1, ... O_{T-1}) = \text{IQ pair observation sequence} \\
B &= \text{IQ probability distributions} \\
N &= \text{number of qubit states in the model} \\
Q &= \{q_0, q_1, ..., q_{N-1}\} = \text{Markov process qubit states} \\
A &= \text{state transition matrix} \\
\pi &= \text{initial state distributions}
\end{align*}
\]

Preparing a readout measurement as a Markov chain requires partitioning a single shot of total time \( t \) into several uncorrelated segments. Note that the total readout time \( t \) does not necessarily correspond to the readout pulse width \( W \), and typically \( t < W \). Each segment is demodulated for a short time interval \( \Delta t \), resulting in a single observation in the form of a IQ pair; \( O_i = (I_i, Q_i) \). Therefore, each shot becomes a discretized sequence of observations of size \( T = t/\Delta t \), where \( T \) is an integer. The emission distributions \( B = \{B_i\} \) correspond to the 2-dimensional probability distributions that randomize the measurement based on the hidden state (e.g., figure 2F). The hidden states are identified with the qubit states; for two states \( Q = \{0, 1\} \). The Markov assumption requires that the state \( q_i \) be only dependent on the preceding state \( q_{i-1} \). This is satisfied since in this regime, the probability of transitioning from the excited state in one observation segment to the ground state in the next is fixed by \( P_e(\Delta t) = e^{-\Delta t/T_1} \).

The transition matrix \( A = [a_{ij}] \) gives the transition probabilities between the qubit states. State transitions of the form \( q_i \rightarrow q_j \) for \( i > j \) are associated with qubit state relaxation \( (T_1) \), and those transitions in which \( i < j \) can be attributed to excitations (“heating”). The initial state distributions \( \pi = \{\pi_i\} \) represent the initial state probability of state \( q_i \) (i.e. the priors). For example, in the ideal case in which the excitation rate is zero \( (a_{01} = 0) \), the transition matrix \( (A) \) for the two state case is given by

\[
a_{ij} = \begin{bmatrix}
1 & 0 \\
1 - e^{-\Delta t/T_1,eff} & e^{-\Delta t/T_1,eff}
\end{bmatrix},
\] (5)

where \( i, j \in \{0, 1\} \), and \( T_1,eff \) is the effective relaxation time which accounts for measurement induced dephasing \[ 27 \]. In practice, the heating rate is not necessarily zero and can be extracted from the learned transition matrix.

A HMM is defined by the set of parameters \( \lambda = (A, B, \pi, N) \). There are three well known solved problems with hidden Markov models, and they are briefly re-summarized next \[ 16 \]. First, given a HMM \( \lambda = (A, B, \pi, N) \) and a sequence of observations \( O \) one can determine the probability of the sequence given the model \( \lambda \), \( P(O|\lambda) \). This probability can be computed in a straightforward fashion by summing, over all possible state sequences, the product of the probability of the observation sequence for a specific state sequence times the probability of that specific state sequence. However, this process is computationally intensive. In practice, either of the so-called forward or backward algorithms are used to compute \( P(O|\lambda) \) (see ref. \[ 16 \]). Note that both the forward and backward algorithms enable the
efficient applicability of HMMs. Second, given a model \( \lambda \) and observation sequence \( O \), an optimal hidden state sequence can be computed. This feature of HMMs allows for the prediction of the optimal state sequence by calculating the probability of being in state \( q_i \) at observation \( O_i \). The predicted hidden state sequence is then composed by selecting the most probable state at each observation. And third, given an observation sequence \( O \) and the number of hidden states \( N \), the model parameters can be computed by solving the maximum likelihood problem using the Baum-Welch algorithm [28]. This feature enables unsupervised training of HMMs.

Figure 3 summarizes the HMM scheme for a two-qubit state model, and illustrates the predicted state-path of a single shot using the forward-backward algorithm. The red line indicates the probability of being in the excited state, while the blue line represents the probability of being in the ground state. It can been seen that for this shot a relaxation transition is predicted near observation index \( i = 125 \). Nevertheless, the starting state \( (O_0) \) is identified with high probability.

C. HMM Implementation

The HMM readout scheme was implemented in python with the Pomegranate package [29]. Preparing a HMM in Pomegranate could be achieved by “baking” a model if the model parameters, \( \lambda = (A, B, \pi, N) \), were known. Alternatively, unsupervised training using the Baum Welch algorithm learned the model parameters from a dataset, but the number of states \( (N) \) was required a-priori. Since the segments must be uncorrelated, preparing the heterodyned readout signal required careful selection of the segment demodulation time \( (\Delta t) \). Hence, to find a suitable segment demodulation time we calculated the autocorrelation of the heterodyned readout signal and determined the point of minimum correlation. Although there were several points of minimum correlation, we sought the shortest possible time to ensure the IQ distributions remained Gaussian. However, while training hidden Markov models we found that the first minimum led to large variations in the learned parameters. The next autocorrelation minimum which led to consistent HMM parameters corresponded to a segment demodulation time of \( \Delta t = 80ns \) (160 sampled points for our 2 GHz ADC), or stated differently, two periods of the demodulation signal \( (\Omega_{TF} = 25 \text{ MHz}) \).

The complete experimental dataset consisted of 25,250 single shots prepared in the ground state, and 25,250 shots prepared in the excited state. Each excited state measurement shot was taken with a pulse sequence consisting of a 25ns \( \pi \)-pulse \( (R_{\pi}(\pi)) \), followed by a \( W = 20\mu s \) rectangular readout pulse. The ground state shots were obtained with the same readout pulse but without any qubit excitation preceding it. The readout signal had a delay time of approximately 250ns before the detection of the readout pulse, and another 250ns was trimmed from the readout signal to ensure the readout cavity was in steady state. Therefore, the total delay between the qubit pulse and observations was 500ns. Note that due to this delay we expect the assignment fidelity, defined by equation [5] to be limited by the starting state population which is on the order of \( e^{xp(-0.5/14.46)} = 96.6\% \), where we used the effective relaxation time \( T_{1,\text{eff}} = 14.46\mu s \) as extracted from the HMM scheme (discussed below).

D. Unsupervised Learning with HMMs

Before proceeding to the main results we discuss the validation procedures used to confirm the reliability and consistency of hidden Markov models via Pomegranate. First, since qubit readout with HMMs gives access to the transition rates during the measurement, it is possible to extract the effective relaxation rate \( (T_{1,\text{eff}}) \) under the influence of the readout signal. However, due to measurement induced decoherence the relaxation rate is not necessarily equal to the conventionally measured \( T_{1} \) [27, 30] (operating in the strong projective regime). Therefore, in order to determine the accuracy in estimating \( T_{1,\text{eff}} \) with unsupervised learning of HMMs, we generated 31 simulated datasets with the relaxation rates varying linearly from 1\( \mu s \) to 16\( \mu s \). The learned relaxation rates were then calculated from the transition matrix element, \( T_{1,\text{eff}} = (-80/\ln(a_{11}))\mu s \), where \( a_{11} \) was extracted from unsupervised learning using the
Baum Welch algorithm. The standard deviation of the differences between the actual and learned $T_{1,\text{eff}}$ values was 0.175µs, and indicated that we could estimate the effective $T_{1,\text{eff}}$ to within 1.25% in this range. For our experimental dataset the learned $T_{1,\text{eff}}$ value was (14.460 ± 0.175)µs. In comparison, measuring $T_1$ using the standard $R_\chi(\pi) \rightarrow \text{variable darktime} \rightarrow \text{readout pulse}$ method resulted in $T_1 = 21 ± 1$µs. The difference of approximately 32% between the conventionally measured $T_1$ and the learned value was consistent with the qubit induced dephasing [27, 30] caused by the readout amplitude of ∼ 20 photons [21], estimated via a AC-Stark shift calibration [31].

Next, bootstrap sampling techniques were used to extract the statistical variations in unsupervised training of HMMs via Pomegranate. The bootstrap technique consisted of generating 100 randomized subsets for each state from the experimental dataset. Each randomized bootstrapped subset consisted of a total of 2,000 ground state and 2,000 excited state single shot measurements. HMMs were then trained from each bootstrap subset using the Baum Welch algorithm. The standard deviation from the bootstrap technique on the learned transition matrix parameters was under 0.03%, and under 1% for the means of the IQ distributions. Thus, indicating good consistency in unsupervised training of hidden Markov models via Pomegranate.

### III. MAIN RESULTS

#### A. Hidden Markov Model State Classifier

For the implementation of the HMM classifier, unsupervised learning with a training data subset of 2000 shots prepared in the ground state and 2000 shots prepared in the excited state was used. For a classifier which is practical for single shot application the starting state probabilities were then modified such that $\pi_0 = \pi_1 = 0.50$. The classification scheme was based on the state-path predicted by the forward-backward algorithm, and a “0” or “1” was assigned based on the state that had the maximum starting state probability. The remaining 46,500 shots were used as a test dataset. In figure 4, 6,250 excited state shots and 6,250 ground state shots were classified with the HMM readout scheme for three different readout times. Shots that were predicted to start and remain in the excited state during the readout measurement were labeled in red, and those predicted to have started and remain in the ground state were labeled in blue. With the HMM readout scheme, transitions can be detected with high certainty while maintaining high fidelity in the determination of the starting state. This is illustrated by labeling in yellow those shots in which a transition was predicted.

Next, the excited state assignment fidelity $F_{a,1} = 1 - P(0|1)$ defined by equation (3) is compared for the HMM readout classifier against a multivariate Gaussian (MVG) classifier, and a support vector machines (SVM) classifier. Here the full dataset of 46,500 shots was classified, and errors for the HMM classifier were extracted from the bootstrap samples. Figure 5a shows the classification readout fidelity ($F_{a,1}$) for the excited state as a function of the demodulation time in units of $T_{1,\text{eff}}$. The assignment fidelities for the HMM, SVM, and MVG classifiers were 96.48%, 95.87%, and 96.05%, respectively. A striking difference between the datasets is that the HMM method is impervious to qubit state transitions. Since the HMM scheme can determine the starting state of shots that underwent a state transition with high probability, the readout fidelity remains fixed as a function of the readout time beyond ∼ 1µs. Note that for the HMM scheme the dominant source of misclassification was observed from shots that had a transitions within the first few observation segments. Since approximately 500 nanoseconds were trimmed from the start of each shot and state preparation errors were estimated to be less than 1%, the excited state assignment fidelity is expected to be limited by the starting state population which is on the order of $\exp(-0.5/14.46) = 96.6%$. This is consistent with the results presented in figure 5a.

![Figure 5a](image-url)

**FIG. 5.** a) Excited state classifier fidelity as a function of the readout time, in units of the effective relaxation time $T_{1,\text{eff}}$. The HMM scheme achieves a maximum assignment fidelity of 96.48%, while the SVM and MVG methods achieve a 95.87%, and 96.05% fidelity, respectively. The HMM scheme is robust against qubit relaxation time, eliminating the need for readout time optimization. b) The total classification error as determined with a simulated dataset in which the starting state is prepared with 100% preparation fidelity. The HMM scheme has a lower total classification error.

The single shot classification errors for the starting state were then extracted from a simulated dataset hav-
ing 100% preparation fidelity. The simulated dataset for the excited state was created by first generating state sequences (i.e. a sequence of ones and zeroes) having an exponential probability distribution with $T_{1,eff} = 14.46\mu s$. Independent, identically distributed (i.i.d) random samples of IQ values were then drawn from the learned multivariate Gaussian distributions according to the randomly generated state sequences. The ground state shots were simulated, with no transitions, by random sampling from a multi-variate Gaussian distribution. A plot of the total classifier error extracted from the simulated dataset, $1 - \mathcal{F}_n = \left( P(0|1) + P(1|0) \right)/2$, is shown in figure 5b. This measure quantifies the errors in the classification of the single shot measurements for both the excited and ground states. It can be seen that overall the HMM classifier has a misclassification error under 2% with a plateau of 1.86%, whereas the MVG (SVM) method reaches a minimum error of 2.75% (2.77%) before increasing as a function of the readout time.

B. Ideal Fidelity and Single Shot Efficiency

As already mentioned, the assignment fidelity shown in figure 3a is limited by the starting state population. However, because the HMM readout scheme relies on short discretize segments (making the IQ distributions Gaussian), it is possible to compute the fidelity from the integration of the overlap regions. This method of computing fidelity from the overlap differs from the assignment fidelity in that the former will yield the maximum fidelity achievable, whereas the latter will be limited by the starting state population. Since the HMM readout scheme can predict state-path sequences, we may extract the Gaussian IQ probability distributions as follows. Using state-path prediction with the forward-backward algorithm, each shot is demodulated until a transition is detected. When a transition is detected, the shot is split at the transition and demodulated in two sections. One section corresponds to when the qubit was predicted in the excited state for that shot, and the other corresponds to when the qubit was predicted in the ground state. In shots with no predicted transitions the signal is demodulated for the complete integration time ($T_{int}$). Demodulating in this fashion eliminates averaging over transitions, and relies on the ability of the HMM scheme to correctly predict the qubit state at each observation. If HMM provides good state discrimination we expect a Gaussian distribution for each state. Although each shot may have a different standard deviation due to the variable integration time, the sampled distributions for many shots over a given integration time ($T_{int}$) should result in distributions of equal variance so long the HMM method has accurate state discrimination.

The resulting IQ plot for $T_{int} = 1.2\mu s$ is shown in figure 6a. To compute the fidelity of the HMM filtered data, the HMM-filtered IQ distributions were projected onto the axis connecting the two centroids [12]. Each of the projected distributions were then fitted simultaneously with equal-variance single Gaussians. In contrast, the ideal fidelities were extracted by simultaneously fitting equal-variance double Gaussians to each projected distribution of the unfiltered IQ data [14]. Fidelities for both methods were then computed using equation (1). Table I shows the results of the computed fidelities for various integration times ($T_{int}$), and shows that the HMM scheme reaches the ideal fidelity limit.

![Image](image_url)

**FIG. 6. a)** IQ scatter plot and equal-variance Gaussian fits of ground (blue) and excited state (red) shots filtered with the HMM state discriminator. Yellow points indicate shots predicted to start in the excited state which transitioned to the ground state during the readout. b) Fidelity may be improved by rejecting low probability shots via a threshold parameter, but only at the expense of readout efficiency.

| $T_{int}$ | 0.72\mu s | 1.2\mu s | 2.16\mu s |
|----------|-----------|---------|-----------|
| Ideal    | (99.14 $\pm$ 0.04)% | (99.92 $\pm$ 0.02)% | (99.9987 $\pm$ 0.0005)% |
| HMM      | (99.12 $\pm$ 0.04)% | (99.91 $\pm$ 0.06)% | (99.998 $\pm$ 0.0003)% |

The above method serves to illustrate the flexibility of using the HMM scheme by having access to state-path prediction for each shot. An alternative method to arrive at the same conclusion is by allowing low-probability shot rejection with the HMM classifier, which increases the readout fidelity at the expense of reducing the readout efficiency. The efficiency is quantified by the ratio of accepted shots versus attempted shots. Figure 6b shows the efficiency and the excited state assignment fidelity for both, the HMM and MVG schemes. In this case the optimal readout time of approximately 5\%$T_{1,eff}$ for the MVG classifier and a arbitrary time of 10\%$T_{1,eff}$ for the HMM scheme was selected. It can be seen the HMM...
scheme achieved a higher assignment fidelity than the MVG scheme, while maintaining a comparable efficiency. We omit the SVM method since its classification scheme is based on a geometric approach and, thus, does not enable low probability shot rejection.

IV. CONCLUSION

We demonstrated that hidden Markov models allow for a robust state-path readout scheme via transition detection, thus, allowing for starting state determination with high fidelity. The HMM scheme also demonstrated consistent classification performance even for readout times comparable with qubit $T_1$ times, where in contrast, current state of the art schemes are hindered by qubit state decay. Thus, using the HMM readout scheme eliminates the need for optimizing the integration time, a process which must be tailored to each superconducting qubit due to individual variations in their $T_1$ times. Meanwhile, the state-path trajectory HMM scheme is compatible with real time control systems, e.g. quantum orchestration platforms, which can lead to measurement speed up. Furthermore, unsupervised learning with the Baum Welch algorithm provides a tool for learning about transitions rates between quantum states and distribution parameters, thus, giving easy access to information not accessible with current state-of-the-art qubit readout schemes. Indeed, owing to the Markovian nature of qubit relaxation, hidden Markov models are a natural platform for qubit readout, and which, can handle multi-level qubit systems as well.

We acknowledge the helpful conversations with Kater Murch, Xian Wu, Jacob Schreiber, and Jason Bernstein. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. The authors gratefully acknowledge support from the Department of Energy Office of Advanced Scientific Computing Research, Quantum Testbed Pathfinder Program under Award 2017-LLNL-SCW163, the National Nuclear Security Administration Advanced Simulation and Computing Beyond Moore’s Law program (LLNL-ABS-795437) and Lab Directed Research and Development award LDRD19-ERD-013. (LLNL-JRNL-810931)

[1] R. Barends, L. Lamata, J. Kelly, L. García-Álvarez, A. Fowler, A. Megrant, E. Jeffrey, T. White, D. Sank, J. Mutus, et al., Nature communications 6, 1 (2015).
[2] G. Wendin, Reports on Progress in Physics 80, 106001 (2017).
[3] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. Kimchi-Schwartz, J. McClean, J. Carter, W. De Jong, and I. Siddiqi, Physical Review X 8, 011021 (2018).
[4] Z. Yan, Y.-R. Zhang, M. Gong, Y. Wu, Y. Zheng, S. Li, C. Wang, F. Lian, J. Lin, Y. Xu, et al., Science 364, 753 (2019).
[5] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, et al., Nature 574, 505 (2019).
[6] E. Lucero, M. Hofheinz, M. Ansmann, R. C. Bialczak, N. Katz, M. Neeley, A. O’Connell, H. Wang, A. Cleland, and J. M. Martinis, Physical review letters 100, 247001 (2008).
[7] M. Rol, C. Bultink, T. O’Brien, S. De Jong, L. Theis, X. Fu, F. Luthi, R. Vermeulen, J. de Sterke, A. Bruno, et al., Physical Review Applied 7, 041001 (2017).
[8] M. D. Reed, L. DiCarlo, S. E. Nigg, L. Sun, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Nature 482, 382 (2012).
[9] A. D. Córcoles, E. Magesan, S. J. Srinivasan, A. W. Cross, M. Steffen, J. M. Gambetta, and J. M. Chow, Nature communications 6, 1 (2015).
[10] T. Walter, P. Kurnpiers, S. Gasparinetti, P. Magnard, A. Potočnik, Y. Salathé, M. Pechal, M. Mondal, M. Opppliger, C. Eichler, et al., Physical Review Applied 7, 054020 (2017).
[11] A. Nersisyan, S. Poletto, N. Alidoust, R. Manenti, R. Renzas, C.-V. Bui, K. Vu, T. Whyland, Y. Mohan, E. A. Sete, et al., arXiv preprint arXiv:1901.08042 (2019).
[12] A. P. Place, L. V. Rodgers, P. Mundada, B. M. Smitham, M. Fitzpatrick, Z. Leng, A. Premkumar, J. Bryon, S. Sussman, G. Cheng, et al., arXiv preprint arXiv:2003.00024 (2020).
[13] R. Dassonneville, T. Ramos, V. Milchakov, L. Planat, E. Dumur, F. Foroughi, J. Puertas, S. Leger, K. Bharadwaj, J. Delafosse, et al., Physical Review X 10, 011045 (2020).
[14] E. Magesan, J. M. Gambetta, A. D. Córcoles, and J. M. Chow, Physical review letters 114, 200501 (2015).
[15] A. Seif, K. A. Landsman, N. M. Linke, C. Figgatt, C. Monroe, and M. Hafezi, Journal of Physics B: Atomic, Molecular and Optical Physics 51, 174006 (2018).
[16] L. Rabiner and B. Juang, icesa magazine 3, 4 (1986).
[17] A. Blais, R.-S. Huang, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, Physical Review A 69, 062320 (2004).
[18] C. Cortes and V. Vapnik, Machine learning 20, 273 (1995).
[19] M. Sun and M. R. Geller, arXiv preprint arXiv:1810.10523 (2018).
[20] M. R. Geller, arXiv preprint arXiv:2002.01471 (2020).
[21] X. Wu, S. L. Tomarken, N. A. Petersson, L. A. Martinez, J. Y. Rosen, and J. L. DuBois, (2020), arXiv:2005.13165 [quant-ph].
[22] A. Wallraff, D. Schuster, A. Blais, L. Frunzio, J. Majer, M. Devoret, S. Girvin, and R. Schoelkopf, Physical review letters 95, 060501 (2005).
[23] J. Gambetta, W. Braff, A. Wallraff, S. Girvin, and R. Schoelkopf, Physical Review A 76, 012325 (2007).
[24] A. Ben-Hur, D. Horn, H. T. Siegelmann, and V. Vapnik, Journal of machine learning research 2, 125 (2001).
[25] M. Reagor, C. B. Osborn, N. Tezak, A. Staley, G. Prawiroatmodjo, M. Scheer, N. Alidoust, E. A. Sete,
N. Didier, M. P. da Silva, et al., Science advances 4, eaao3603 (2018).

[26] C. E. Shannon, IRE Nat. Conv. Rec 4, 1 (1959).

[27] M. Boissonneault, J. M. Gambetta, and A. Blais, Physical Review A 79, 013819 (2009).

[28] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, The annals of mathematical statistics 41, 164 (1970).

[29] J. Schreiber, The Journal of Machine Learning Research 18, 5992 (2017).

[30] T. Picot, A. Lupaşcu, S. Saito, C. Harmans, and J. Mooij, Physical Review B 78, 132508 (2008).

[31] D. Schuster, A. Wallraff, A. Blais, L. Frunzio, R.-S. Huang, J. Majer, S. Girvin, Schoelkopf, and RJ, Physical Review Letters 94, 123602 (2005).