LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-810931

Improving Qubit Readout with Hidden Markov Models

L. A. Martinez, Y. J. Rosen, J. L. DuBois

May 28, 2020

Physics Review A

Improving Qubit Readout with Hidden Markov Models*

Luis A. Martinez, Yaniv J. Rosen, and Jonathan L. DuBois
Lawrence Livermore National Laboratory, CA 94550, USA

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, the method also eliminates the qubit-dependent readout time optimization requirement in 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].

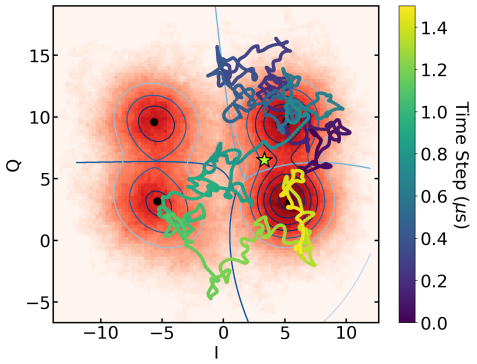


FIG. 1. Running average (colored line) of the heterodyned signal of a single shot in a two-qubit four-state system. The different colors correspond to time evolution. The star marker denotes the demodulated IQ value over the entire measurement time. Bayes classifier and contour lines representing probability distributions shown for reference.

Random noise and qubit decay processes reduce readout fidelity. In Fig. 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, Fig. 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

* (DOI: 10.1103/PhysRevA.102.062426)

example illustrates that, apart from optimizing hardware parameters [10], choosing an appropriate readout integration time plays an important role in mitigating qubit relaxation.

II. METHODS

The experimental quantum platform is a 3D cavity QED [17, 21] system utilizing a strong-projective dispersive measurement scheme [22]. 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 ω_r is applied to the readout resonator, filtered, and amplified. At this stage, as indicated in Fig. 2(a), a fidelity limit is imposed by the signal to noise ratio (SNR) and the CQED system parameters [10]. 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 ω_{LO}, to an intermediate frequency $\Omega_{IF} = \omega_r - \omega_{LO}$. The IF signal is then digitally decomposed in quadrature and demodulated by picking out the Fourier component at Ω_{IF}. The demodulation process integrates both the in-phase and quadrature components for a total integration time T_{int} (Fig. 2(b)). At this stage the readout fidelity may be optimized by appropriately tuning of the integration time. Insufficient state distinguishability may arise for relatively short integration times, whereas qubit decay obfuscates the readout signal for long times. The demodulation of each single shot results is a (I, Q) coordinate and the histogram of several shots forms the probability distribution in the IQ-plane as illustrated in Fig. 2(c).

Preparing a state classifier involves a training procedure in which a training dataset is used to learn the model parameters (Fig. 2(d)). For the multivariate Gaussian (MVG) model, the training solely consists of learning the mean array $\vec{\mu}$ and covariance matrix $\mathbf{\Sigma}$. The goodness-of-fit of the MVG model is severely affected for longer readout times because the IQ distributions for states exhibiting state transitions become skewed and, therefore, non-Gaussian [23]. Support vector machines (SVM) provide both supervised and unsupervised learning capabilities [18, 24]. 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, an optimal integration time is required for maximum readout fidelity.

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, $\mathcal{F}_{id}$, defined by the misclassification probability, is computed from the integration of the overlapped regions of the projected Gaussian probability distributions [14];

$$\mathcal{F}_{id} = \frac{1}{2}(1 + \mathrm{erf}(\sqrt{\frac{R}{8}})). \tag{1}$$

Here, R is a measure of the separation between the two distributions in question [14].

$$R = \frac{(\langle S_0 \rangle - \langle S_1 \rangle)^2}{\mathrm{var}(S)} \tag{2}$$

S denotes the measurement outcome after the integration of the signal, i.e. the demodulated value, and $\mathrm{var}(S)$ 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. However, for relatively long integration times ($T_{int} \gtrsim 5\% \times T_1$), the distributions are skewed by relaxation transitions and a Gaussian model is no longer adequate.

If the readout state probability distributions are not Gaussian, the method of calculating fidelity described above in Eq. (1) is not suitable. Instead, fidelity of classification systems may generally be assessed through various statistical figures of merit [25]. Here we use the assignment fidelity, ($\mathcal{F}_a$), to compare the MVG, SVM, and HMM classifiers. The assignment fidelity is adapted from a more general distortion measure which is equivalent to the infidelity. Introduced by Shanon's information theory [26], it gives a quantitative measure for performance of classification systems which generalizes to the confusion matrix formalism illustrated in Fig. 2(e). For the two state case the assignment fidelity is

$$\mathcal{F}_a = 1 - \frac{1}{2}(P(0|1) + P(1|0)), \tag{3}$$

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 fidelity of the gate used to prepare the state, 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. HMMs have been

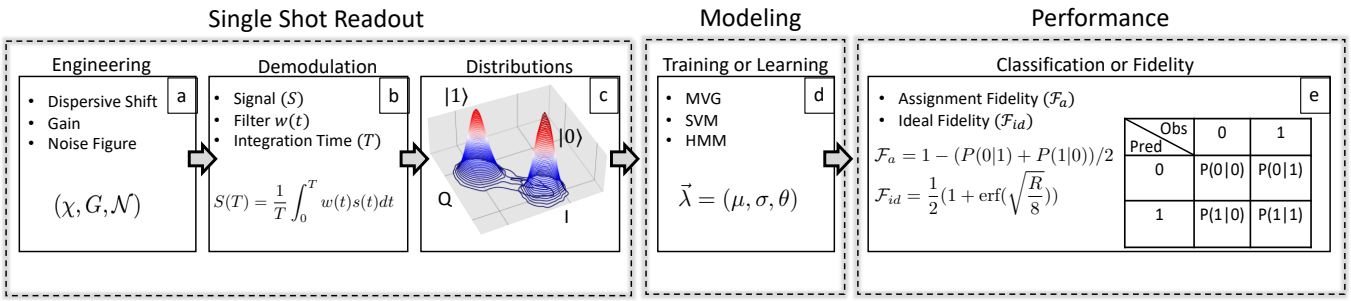


FIG. 2. Summary of Qubit Readout. (a) Factors determining readout fidelity begin with the quantum hardware; Hamiltonian parameters, and SNR. (b) Apart from hardware, fidelity may be improved at the demodulation stage. Conventional demodulation scheme requires tuning to an optimal integration time in which the distributions are approximately Gaussian (c). (d) Regression analysis can be used to train models for improved classification performance. (e) Classification errors are extracted from the misclassification probabilities.

used to measure the relaxation time of the nuclear spin state of a nitrogen-vacancy defect in diamond [27], and more recently in a proposed robust readout scheme for bosonic systems in the dispersive coupling regime [28]. For this work the relevant parameters (see appendix section 2) of a HMM are [16]:

$$\begin{aligned} O &= \text{(I,Q) pair observation sequence} \\ D &= \{d\}, \text{ set of qubit states} \\ N &= \text{number of qubit states in the model} \\ \pi &= \text{set of initial state distributions} \\ B_d &= \text{(I,Q) pair probability distribution} \\ &\quad \text{given state } d \in D \\ \mathbf{A} &= [a_{ij}], \text{state transition matrix} \end{aligned} \tag{4}$$

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 Δt, resulting in a single observation in the form of an IQ pair; $O_i = (I_i, Q_i)$. Therefore, each shot becomes a discretized sequence of observations of size $n = t/\Delta t$, where n is an integer. The emission probability distribution, $B_d \equiv P(O|d)$, is the probability that the observation pair (I, Q) was emitted from state d. For qubit readout, B_d corresponds to the 2-dimensional probability distributions in IQ-space that randomize the readout measurement based on the hidden state (e.g., Fig. 2(c)). The hidden states are identified with the qubit states; for two states $D = \{|0\rangle, |1\rangle\}$. The Markov assumption requires that the state d_i be only dependent on the preceding state d_{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 $\mathbf{A} = [a_{ij}]$ gives the transition probabilities between the qubit states. The off-diagonal elements of the form a_{ij} for $i > j$ are associated with qubit state relaxation, and elements in which $i < j$ can be attributed to qubit excitations, e.g. due to heating. The initial state distributions π represent the initial state probability of state d, i.e. the priors. For example, in the ideal case in which the excitation rate is zero ($a_{01} = 0$), the transition matrix ($\mathbf{A}$) for the two state case is given by

$$a_{ij} = \begin{bmatrix} 1 & 0 \\ 1 - e^{-\Delta t/T_{1,eff}} & e^{-\Delta t/T_{1,eff}} \end{bmatrix}, \tag{5}$$

where $i, j \in \{0, 1\}$, and $T_{1,eff}$ is the effective relaxation time which accounts for measurement induced dephasing [29]. In practice, the heating rate is not necessarily zero and can be extracted from the learned transition matrix element, $\Gamma_{01} = a_{01}/\Delta t$.

A HMM is defined by the set of parameters $\lambda = (\mathbf{A}, B, \pi, N)$. There are three well known solved problems with hidden Markov models, and they are briefly re-summarized in the following [16]. First, given a HMM $\lambda = (\mathbf{A}, B, \pi, N)$ and a sequence of observations O one can determine the probability of the sequence given the model λ, $P(O|\lambda)$. This probability can be computed in a straightforward fashion by summing the product of the observation sequence probability and the state sequence probability over all possible state sequences. However, this process is computationally intensive. In practice, either of the so-called forward or backward algorithms are used to compute $P(O|\lambda)$ [16]. Note that both the forward and backward algorithms enable the efficient applicability of HMMs. Second, given a model λ 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 d 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 [30]. This key feature enables the unsupervised training of HMMs.

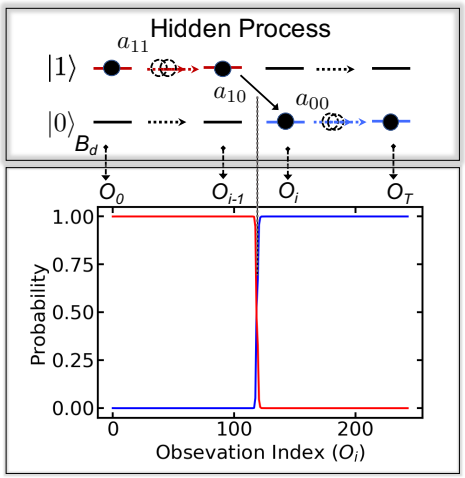


FIG. 3. State path prediction with hidden Markov models for readout of a single 20 μs shot. Each shot is discretized into a sequence of 80 ns observations (O_i), and each observation is an (I,Q) pair emitted from the hidden state having an emission probability distribution B_d. The forward-backward algorithm computes the probability of being in the ground (blue) or excited (red) state, $\{|0\rangle, |1\rangle\}$ respectively. The probability the qubit was in the excited state at the start of the measurement was 99.98%.

The application of the forward-backward algorithm via HMMs for single shot readout yields a predicted state path that is based on the probability of being in state d at each observation O_i of the sequence. Because the forward-backward algorithm finds the best-fit state sequence given an observation sequence and model, the result is the determination of the qubit state at each index with high probability. In particular, the determination of the qubit state d_0 at the start of the readout measurement, can be made with high probability even in cases where a relaxation transition occurs during the readout measurement. This is of key importance in qubit applications where the quantity of interest is the qubit state immediately at the start of the readout measurement. For example, Fig. 3 illustrates the predicted state-path of a 20 μs single shot using the forward-backward algorithm. The red line indicates the probability of being in the excited state $d_i = |1\rangle$ at each observation index, while the blue line represents the probability of being in the ground state $d_i = |0\rangle$ at each observation. It can been seen that for this single shot a relaxation transition is predicted near observation index $i = 125$, yet the probability the qubit was in the excited state at the start of the readout measurement ($d_0 = |1\rangle$) is 99.98%.

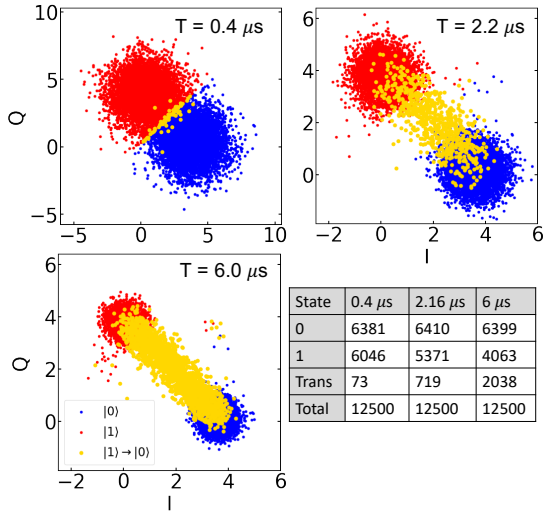
C. HMM Implementation

The HMM readout scheme was implemented in Python with the Pomegranate package [31]. Preparing a HMM in Pomegranate could be achieved by "baking" a model if the model parameters, $\lambda = (\mathbf{A}, B, \pi, N)$, 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 Δt. 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, corresponding to the intermediate frequency ($\Omega_{IF} = 25$ MHz) of the signal, we sought the shortest possible time to ensure that the probability distributions formed by the discretized IQ observation segments, B_d, remained Gaussian. However, while training hidden Markov models, we found that the first minimum at 40 ns (one intermediate-frequency period) led to large statistical variations in the learned parameters. The next autocorrelation minimum which led to consistent HMM parameters corresponded to a segment demodulation time of $\Delta t = 80$ ns (two intermediate-frequency periods).

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 25 ns π-pulse $R_x(\pi)$, followed by a $W = 20$ μs 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 250 ns before the detection of the readout pulse, and another 250 ns 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 500 ns. Note that due to this delay we expect the assignment fidelity, defined by Eq. (3), to be limited by the starting state population which is on the order of $\exp\big(-0.5/14.46\big) = 96.6\%$, where we used the effective relaxation time $T_{1,eff} = 14.46$ μs as extracted from the HMM scheme (discussed below). After these adjustment, each single shot consisted of a sequence $\{O_i\}$ of 243 IQ observations (further details of our 3D cQED system can be found in [21]).

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 $T_{1,eff}$ 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 [29, 32] (operating in the strong projective regime). Therefore, to determine the accuracy in estimating $T_{1,eff}$



State	0.4 μs	2.16 μs	6 μs
0	6381	6410	6399
1	6046	5371	4063
Trans	73	719	2038
Total	12500	12500	12500

FIG. 4. Single shot state discrimination with a hidden Markov model (HMM) initial state classifier for three different readout times (T). A HMM is used to classify each shot into one of three cases; (blue) shots predicted to have started and remain in the ground state during the readout, (red) shots predicted to have started and remain in the excited state, and (yellow) shots in which a relaxation transition occurred during the readout. The table lists the counts.

with unsupervised learning of HMMs, we generated 31 simulated datasets with the relaxation rates varying linearly from 1 μs to 16 μs. The learned relaxation rates were then calculated from the transition matrix element, $T_{1,eff} = (-80/\ln(a_{11}))$ ns, where a_{11} was extracted from unsupervised learning using the Baum Welch algorithm. The standard deviation of the differences between the actual and learned $T_{1,eff}$ values was 0.175 μs, and indicated that we could estimate the effective $T_{1,eff}$ to within 1.25% in this range (see appendix). For our experimental dataset the learned $T_{1,eff}$ value was (14.460±.175) μs. In comparison, measuring T_1 using the standard $R_x(\pi) \rightarrow$ *variable darktime* $\rightarrow$ *readout pulse* method resulted in $T_1 = 21 \pm 1$ μs. The difference of approximately 32% between the conventionally measured T_1 and the learned value was consistent with the qubit induced dephasing [29, 32] caused by the readout amplitude of ~ 20 photons [21], estimated via a AC-Stark shift calibration [33]. Similarly, the $|0\rangle \rightarrow |1\rangle$ excitation rate estimated from the a_{01} transition matrix element was $\Gamma_{01} = 9.6 \pm 0.2$Hz, which is reasonable given previous residual excited-state population measurements [34].

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, indicating good consistency in unsupervised training of hidden Markov models via Pomegranate (see appendix for details).

III. MAIN RESULTS

A. Hidden Markov Model State Classifier

The HMM classifier was implemented by using unsupervised learning on a training data subset consisting of 2000 shots prepared in the ground state and 2000 shots prepared in the excited state. For single shot classification the starting state probabilities of the HMM were then modified such that $\pi_{|0\rangle} = \pi_{|1\rangle} = 0.50$. The classification scheme was based on the state-path predicted by the forward-backward algorithm, and a "0" or "1" label 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 Fig. 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. Most importantly, with the HMM readout scheme transitions can be detected with high certainty while maintaining high fidelity in the determination of the starting state of the measurement. This is illustrated by labeling in yellow those shots in which a transition was predicted.

Next, the excited state assignment fidelity $\mathcal{F}_{a,1} = 1 - P(0|1)$ 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. Fig. 5(a) shows the classification readout fidelity, $\mathcal{F}_{a,1}$, for the excited state as a function of the readout time in units of $T_{1,eff}$. The assignment fidelities for the HMM, SVM, and MVG classifiers were 96.5%±0.4%, 95.9%±0.4%, and 96.1%±0.4%, 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 ~ 1 μs. 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 ns were trimmed from the start of each shot and state preparation errors were estimated to be less than 1% [21], 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 Fig. 5(a).

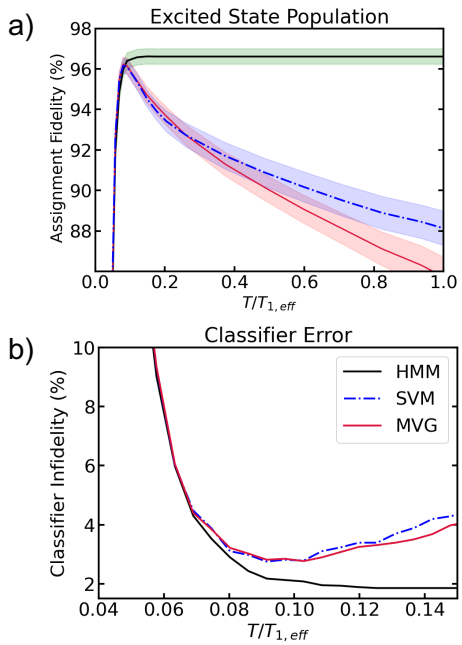


FIG. 5. a) Classifier fidelity in units of the effective relaxation time $T_{1,eff}$ for the excited state. The HMM scheme achieves a maximum assignment fidelity of 96.5%±0.4%, while the SVM and MVG methods achieve a fidelity of 95.9% ± 0.4%, and 96.1% ± 0.4%, respectively. Shaded regions indicate errors. The HMM scheme is robust against qubit relaxation time, eliminating the need for readout time optimization. b) Total classification error determined with a simulated dataset in which the starting state is prepared with 100% preparation fidelity demonstrates that 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 having 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 $T_{1,eff} = 14.46$ μs. Independent, identically distributed random samples of IQ values were then drawn from the learned multi-variate 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, $1-\mathcal{F}_a = (P(0|1)+P(1|0))/2$, is shown in Fig. 5(b). 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

Gaussian distributions from the classified shots can be obtained by using state path prediction with the HMM scheme, a feature that is not possible with SVM or MVG based classifiers. In turn, this feature makes it is possible to compute the fidelity as defined 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. We may extract the post-filtered 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 index. If the HMM scheme provides accurate state discrimination, we expect the sampled distributions of many shots to result in Gaussian distributions of equal variance for a given integration time. Note that in the case when $T_{int} >> T_1$ this no longer holds.

The resulting IQ plot for $T_{int} = 1.2$ μs is shown in Fig. 6(a). To compute the fidelity of the HMM filtered data, the HMM-filtered IQ distributions were projected onto the axis connecting the two centroids [14]. 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.

Table I. Maximum Fidelity from Gaussian Fits

T_{int}	$0.72\mu s$	$1.2\mu s$	$2.16\mu s$
Ideal	$(99.14 \pm .04)\%$	$(99.92 \pm .02)\%$	$(99.9987 \pm .0005)\%$
HMM	$(99.12 \pm .04)\%$	$(99.91 \pm .06)\%$	$(99.998 \pm .003)\%$

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

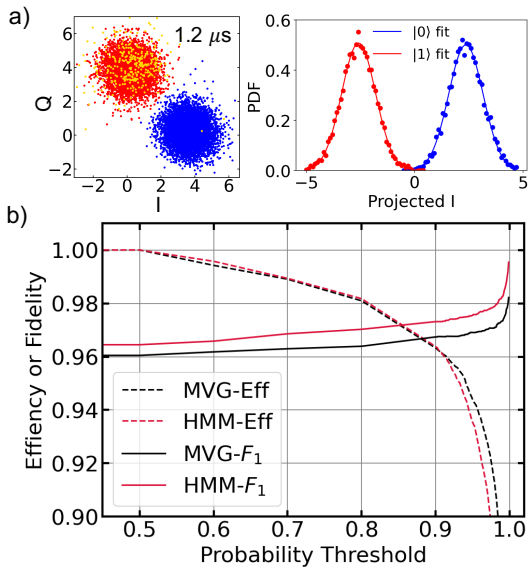


FIG. 6. a) IQ scatter plot and equal-variance Gaussian fits of ground (blue) and excited state (red) shots filtered with the HMM state discriminator for a 1.2 μs demodulation time. Shots predicted to start in the excited state which transitioned to the ground state during the readout are shown in yellow points on the left panel. Colored-matched markers on the right panel represent the projected distributions, solid lines represent the corresponding fits. Fitting errors are smaller than the markers. b) Fidelity may be improved by rejecting low probability shots via a threshold parameter, but only at the expense of readout efficiency.

accepted shots versus attempted shots. Fig. 6(b) shows the efficiency and the excited state assignment fidelity for both, the HMM and MVG schemes. In this case we selected the optimal readout time of approximately $5\% \times T_{1_{eff}}$ for the MVG classifier and a arbitrary time of $10\% \times T_{1,eff}$ for the HMM scheme. 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 qudit 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)

Appendix

1. Bootstrap results

The statistical variations on unsupervised learning of the HMM parameters were obtained by using a bootstrap technique. For the bootstrap technique we randomly generated 100 bootstrap data subsets from the full experimental dataset. The sampling for each data subset was done with replacement, and each subset consisted of 2000 records in the ground state and 2000 records in the excited state. Each bootstrapped data subset was then used to train a hidden markov model (HMM) via the Baum Welch algorithm and statistics were generated on the variations on the learned parameters. The variations for the learned means of the IQ distributions are shown in Fig. 7 along with standard deviations. Similarly, the variations for the transition matrix elements are shown in Fig. 8.

2. HMM with 2D Gaussian distributions

This section is intended to illustrate the notation used in hidden Markov models (HMM) with an example. Suppose we have observed the following sequence of observations, $O_{0:T} = \{O_0, O_1, O_2, \ldots, O_T\}$, and that there is a corresponding sequence of hidden states, $d_{0:T} = \{d_0, d_1, d_2, \ldots, d_T\}$, from which these observations were emitted, probabilistically. Here, d represents the hidden state explicitly, e.g. $d = |0\rangle$, or $d = |1\rangle$, and the index simply maintains the order in the sequence. In a HMM

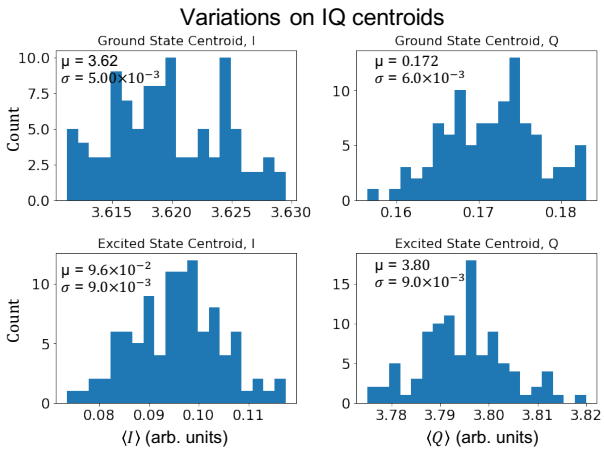


FIG. 7. Statistical variations extracted by the bootstrap method on the learned values of the means of the Gaussian IQ distributions for both the ground and excited state.

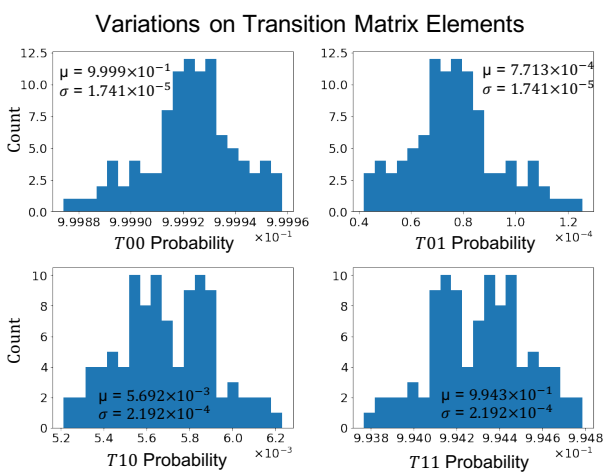


FIG. 8. Statistical variations extracted by the bootstrap technique on the learn values of the transition matrix ($\mathbf{A}$).

we wish to calculate the probability distributions over the many possible hidden-state configurations a sequence O of length T can be observed. This is achieved by computing the joint probability of a hidden-state sequence $d_{0:T}$ and its corresponding observation sequence $O_{0:T}$,

$$\begin{aligned} &P(d_{0:T}, O_{0:T}) = \\ &P(d_0)P(O_0|d_0)P(d_1|d_0)P(d_2|d_1)P(O_1|d_1)P(O_2|d_2)\cdots \\ &= P(d_0)P(O_0|d_0)\prod_{i=1}^{i=T} P(d_i|d_{i-1})P(O_i|d_i). \end{aligned} \tag{6}$$

In this notation $P(d_0) \equiv \pi_{d_0}$ represents the probability of starting in state d_0. The conditional probability $P(d_i|d_{i-1})$ represents the probability of transitioning to state d_i given we were in state d_{i-1}, thus, it defines the transition matrix $\mathbf{A}$. The conditional probability $P(O_i|d_i)$, known as the emission distribution in the formalism of HMMs, is the probability of observation O_i given state d_i. For this example we assume the emission distributions are represented by 2-dimensional equal-variance (σ^2) Gaussians, therefore, the emission distributions (B_d) are given by

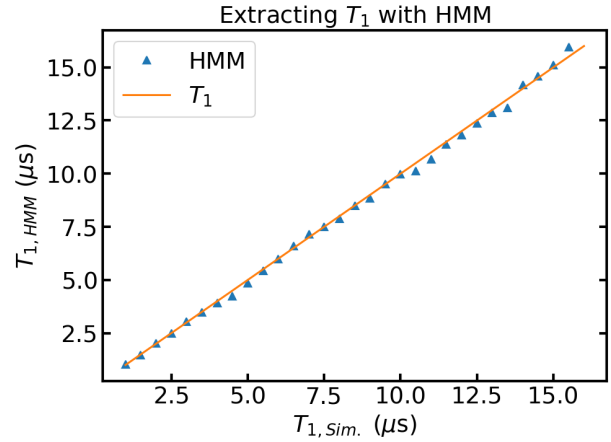


FIG. 9. Simulated datasets varying T_1 from 1 to 16 microseconds were used to trained several HMMs via unsupervised learning using the Baum Welch algorithm. The y-axis represents the T_1 values extracted from HMM and they are plotted against the actual simulated values.

$$\begin{aligned} &B_d(I,Q) \equiv P(O_i|d_i) \\ &= \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}[(I-\bar{I}_{d_i})^2 + (Q-\bar{Q}_{d_i})^2]\right), \end{aligned} \tag{7}$$

where $\bar{I}_{d_i}, \bar{Q}_{d_i}$ represent the center of the Gaussian distribution for state d_i, and the observation O_i is defined by the (I, Q) coordinate that results from a single shot measurement. For a strong-projective single-shot qubit readout measurement the IQ coordinate is the result of the demodulated heterodyne signal, where I represents the in-phase component and Q the out-of-phase component. In general the emission distributions need not be Gaussian, and can in principle be modeled by various probability distributions.

3. Verifying T_1 accuracy with HMM

To extract the accuracy in estimating the effective T_1 value with HMM, we generated several simulated datasets with T_1 times varying linearly from 1μs to 16μs. Unsupervised learning using the Baum Welch algorithm was used to learn the parameters, and the T_1 was estimated from the transition matrix element $a_{11} = exp(-\Delta t/T_{1,HMM})$,

$$T_{1,HMM} = -\frac{80}{\ln(a_{11})}\ ns, \tag{8}$$

where we set $\Delta t = 80$ ns, which corresponds to 80 ns $\times$ 2 GHz $=$ 160 samples for our 2 GHz ADC. The results

are shown in Fig. 9. The standard deviation of the differences (normalized) was 1.25%. The mean deviation between the actual and learned values was 0.0576 ns with a standard deviation of 0.175 ns. This indicates that we can estimate T_1 to within 1.25%. Fitting a linear function $y = mx$ gives $m = 0.994 \pm .003$, which is a difference of $m_{exp} - m_{theo} = 1 - 0.994 = .006$, showing good stability over the entire sampled range.

4. Confirming $T_{1,eff}$ via starting state probabilities

Due to measurement induced dephasing from the strong readout pulse, the T_1 extracted during a readout measurement is expected to be reduced from the typical measured T_1 value, i.e. not during the readout pulse. As an alternative to using the transition matrix to calculate the effective T_1, i.e. during the readout measurement, we confirmed the value by using unsupervised learning with HMMs to extract the starting state distributions $\pi = (\pi_{|0\rangle}, \pi_{|1\rangle})$, i.e. the priors. This was accomplished by varying the demodulation start time (i.e. starting point of integration) for the experimental dataset while extracting the starting state distributions with HMM. This is time consuming, but uses the HMM's optimization to calculate the best fit for the model parameters $(\mathbf{A}, B, \pi, N)$ and learn the starting state distributions. Fig. 10shows the results along with the fitted exponential.

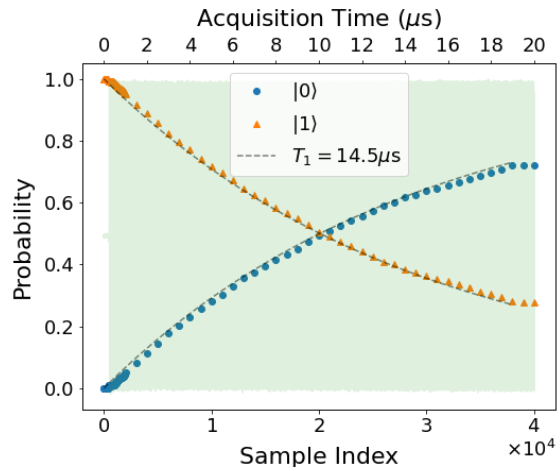


FIG. 10. The effective T_1 time extracted from the unsupervised learning of the starting state distributions by sweeping the start time of the integration during the demodulation.

[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. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. 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. Liang, 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. D. OConnell, H. Wang, A. N. Cleland, and J. M. Martinis, Physical review letters **100**, 247001 (2008).

[7] M. A. Rol, C. C. Bultink, T. E. OBrien, S. R. de Jong, L. S. Theis, X. Fu, F. Luthi, R. F. L. Vermeulen, J. C. de Sterke, A. Bruno, D. Deurloo, R. N. Schouten, F. K. Wilhelm, and L. DiCarlo, 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. Kurpiers, S. Gasparinetti, P. Magnard, A. Potočnik, Y. Salathé, M. Pechal, M. Mondal, M. Oppliger, 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, É. Dumur, F. Foroughi, J. Puertas, S. Leger, K. Bharadwaj, J. Delaforce, *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, ieee assp 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, Quantum Science and Technology **5**, 03LT01 (2020).

[21] X. Wu, S. L. Tomarken, N. A. Petersson, L. A. Martinez, Y. J. Rosen, and J. L. DuBois, Phys. Rev. Lett. **125**, 170502 (2020).

[22] A. Wallraff, D. I. Schuster, A. Blais, L. Frunzio, J. Majer, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Physical review letters **95**, 060501 (2005).
[23] J. Gambetta, W. A. Braff, A. Wallraff, S. M. Girvin, and R. J. 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] A. Dréau, P. Spinicelli, J. R. Maze, J.-F. Roch, and V. Jacques, Physical review letters **110**, 060502 (2013).
[28] C. T. Hann, S. S. Elder, C. S. Wang, K. Chou, R. J. Schoelkopf, and L. Jiang, Physical Review A **98**, 022305 (2018).
[29] M. Boissonneault, J. M. Gambetta, and A. Blais, Physical Review A **79**, 013819 (2009).
[30] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, The annals of mathematical statistics **41**, 164 (1970).
[31] J. Schreiber, The Journal of Machine Learning Research **18**, 5992 (2017).
[32] T. Picot, A. Lupacşcu, S. Saito, C. J. P. M. Harmans, and J. E. Mooij, Physical Review B **78**, 132508 (2008).
[33] D. I. Schuster, A. Wallraff, A. Blais, L. Frunzio, R.-S. Huang, J. Majer, S. M. Girvin, and R. J. Schoelkopf, Physical Review Letters **94**, 123602 (2005).
[34] X. Y. Jin, A. Kamal, A. P. Sears, T. Gudmundsen, D. Hover, J. Miloshi, R. Slattery, F. Yan, J. Yoder, T. P. Orlando, S. Gustavsson, and W. D. Oliver, Physical Review Letters **114**, 240501 (2015).