Faster and Higher Fidelity Qubit Readout with Machine Learning

Wei Tang
Princeton University
weit@princeton.edu

Jimmy Wu
Princeton University
jw60@cs.princeton.edu

Abstract

In this project, we apply classification techniques from machine learning to the task of quantum computing qubit state recognition. Current challenges to qubit readout include the low accuracy of traditional methods, as well as the problem of qubit decoherence, which limits the window of time in which we can perform readout before a qubit collapses. We address these challenges by developing machine learning models with both high accuracy and fast readout times. Compared to traditional methods, our methods increase the readout fidelity by 15% while simultaneously halving the amount of time required to do reliable qubit readout.

1 Introduction

Quantum computers are in an entirely different paradigm of computation from classical computers. Their unique properties have led to theoretical results that show potentially exponential speedup over classical computers in certain applications such as number factoring [1]. Instead of being in either a 0 or 1 binary state as in a classical bit, a qubit exists in a superposition of both states. Figure 1 shows a bloch sphere representation of a qubit. These unique properties are key to the promised potential of quantum computing.

With recent advances in the field, quantum computing has come near the inflection point where having sufficiently many qubits would provide more computational power than any classical computer in existence [2]. However, current physical implementations for qubits still face significant challenges. Specifically, qubit readout, which refers to the process of determining the state of a qubit, is critical for allowing us to make sense of the results from a quantum computation.

The quantum hardware research community has mainly targeted the qubit readout problem by developing more reliable hardware. This is usually achieved by fine-tuning the qubit control signals [3], building better hardware chips [4], or improving implementation methods [5]. Traditional methods for qubit state differentiation, such as the threshold cutting method and the matched filter method, rely on hand-crafted heuristics and give low accuracies [6]. Although advances in hardware technology is still very much crucial to building more functional quantum computers, there is also opportunity to achieve higher readout fidelity using only software methods, specifically machine learning methods.

Additionally, quantum computers currently face a serious challenge called qubit decoherence. Decoherence refers to the phenomenon in which a quantum system reverts back to classical, and is caused by interactions with the environment, which decay or eliminate the quantum behaviour of particles. Different qubit implementation technologies have different decoherence times. The decoherence time places an inherent limit on the window of time in which we can perform qubit readout. For this reason, a short readout time is crucial to ensure that a qubit is able to retain its the quantum information during the readout, lest it collapses from interaction with the environment before the readout is completed.

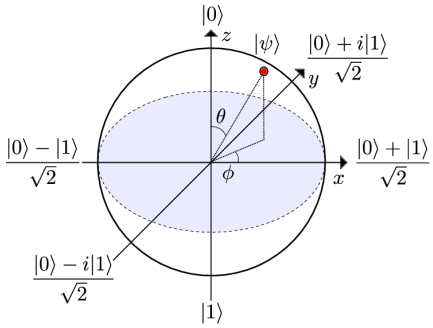

Figure 1: A bloch sphere representation of a qubit. Qubits can exist in superpositions of the 0 and 1 states.

In this project, we tackle the two aforementioned challenges by applying machine learning techniques to both increase the readout fidelity and reduce the time required for accurate readout.

2 Related Work

Prior methods that do not use machine learning include adaptive measurement techniques [7], and techniques that use the Jaynes-Cummings nonlinearity [8]. Other methods use thresholding or maximum likelihood [9], or adaptive maximum likelihood [7], but they require a long readout time.

Prior work has also attempted to use machine learning techniques for qubit assignment [10, 11, 12]. Due to the differences in experimental setup, our accuracy results are not directly comparable. However, prior work has only contrasted traditional method with machine learning methods for a single experimental setup, with no effort at examining the differences in performances across combinations of experimental setups. In addition, these works focus on improving the training and prediction speed of machine learning models [12], while our focus is on shortening the required readout time while preserving high readout fidelity.

3 Data

Our data comes from a 2-qubit superconducting experiment setup from Professor Andrew Houck's group at Princeton [13]. We collect single shot voltage readout traces from the superconducting quantum bits for various combinations of readout power and sampling frequency. The goal is to predict from a qubit's voltage trace whether it is in the ground state or the excited state.

The sampling frequency affects the quality of the voltage trace depending on how close it is to the signal's frequency, and the readout power affects the separation of the traces for the two states. A readout power that is too low results in too little separation to discern the two distinct states, while a power that is too high results in too much noise in the readout. We evaluate across all the different combinations of sampling frequency and readout power to find the optimal combination to use.

The ground truth of the dataset is obtained based on the assumption that we are able to prepare a qubit in the ground state and apply a quantum `NOT` gate, both with perfect fidelity. The quantum `NOT` gate simply rotates a ground state qubit to the excited state, and vice versa. The readout voltage traces of the ground states are obtained by directly measuring the prepared ground state qubits, while those of the excited states are obtained by measuring the qubit after application of a `NOT` gate.

The entire dataset is collected at 36 power levels and 31 frequencies, for a total of 1,116 subdatasets. Each subdataset contains 16,384 ground and excited traces each, and corresponds to a particular combination of power and frequency. Each individual trace is a measurement of the voltage along channel I and channel Q for 2,048 time steps, giving us a 2 by 2,048 dimensional time series data example. Overall, each of the 1,116 subdatasets have 32k (16k each for ground and excited) examples, each of dimension (2, 2048). In our classification methods, we flatten each (2, 2048) sample to have shape (1, 4096). The readout for each trace consists of 2,048 total timesteps, with timesteps

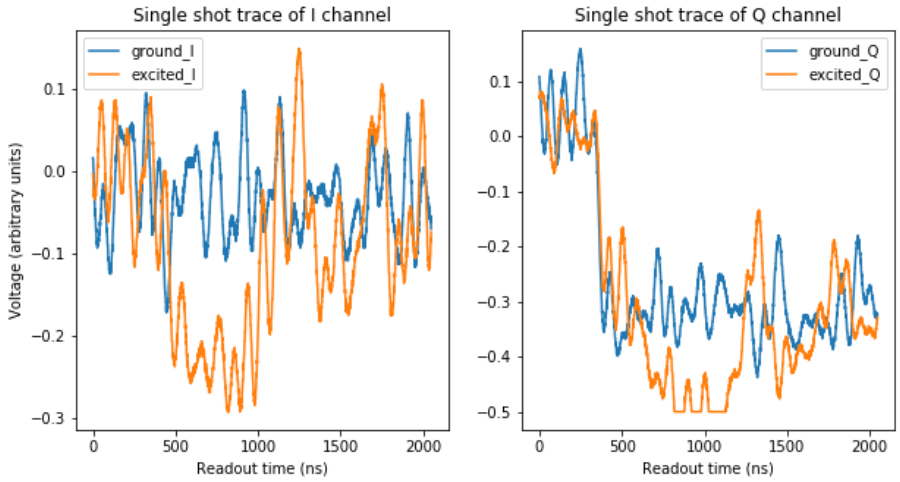

Figure 2: Example of a single shot trace in the ground and excited states for both I and Q channels. The ground I (excited I) and ground Q (excited Q) curves show the two channels of voltage readout for the same ground (excited) state qubit. The curves are arranged based on the I and Q channels for comparison.

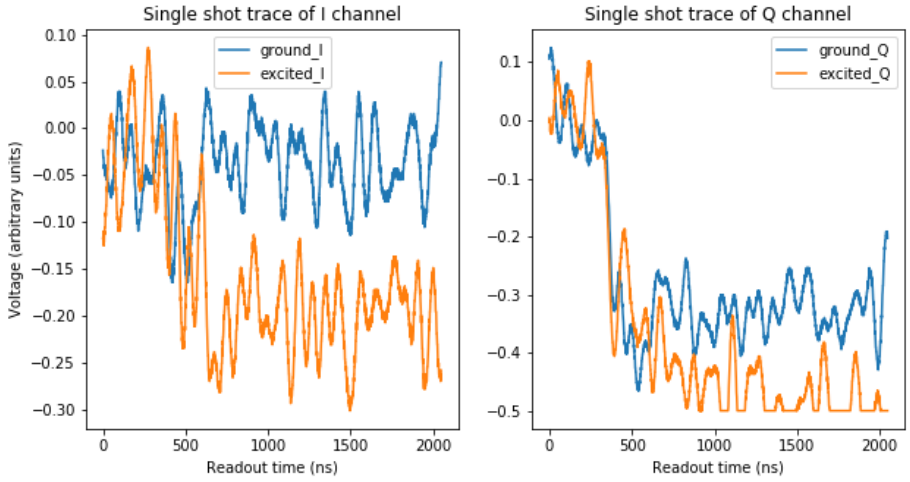

Figure 3: Example of a single shot trace in the ground and excited states for both I and Q channels. This example looks more separated than the one shown in Figure 2.

of length 1 ns. Each readout trace contains signals from two independent channels, I and Q. To simulate shorter readout times, we crop each trace to the appropriate readout time length.

Figure 2 shows an example of two single shot traces, one in the ground state and one in the excited state. Note that individual readout traces from the ground and excited states can significantly overlap with each other and look quite similar upon crude inspection. This is due to electrical noise, heat noise, and other systematic errors in the readout chain which are very difficult to model and quantify. For comparison, Figure 3 shows a different set of traces that exhibits better separation between the two states.

In order to better visualize the readouts, we can remove effect of system noise by simply averaging the 16,384 ground and excited state traces in a subdataset. Figure 4 shows an example of an averaged trace, in which we can clearly observe three distinct stages of the readout process. During the initialization stage, the initial 310 timesteps of both channels are almost identical for the ground and

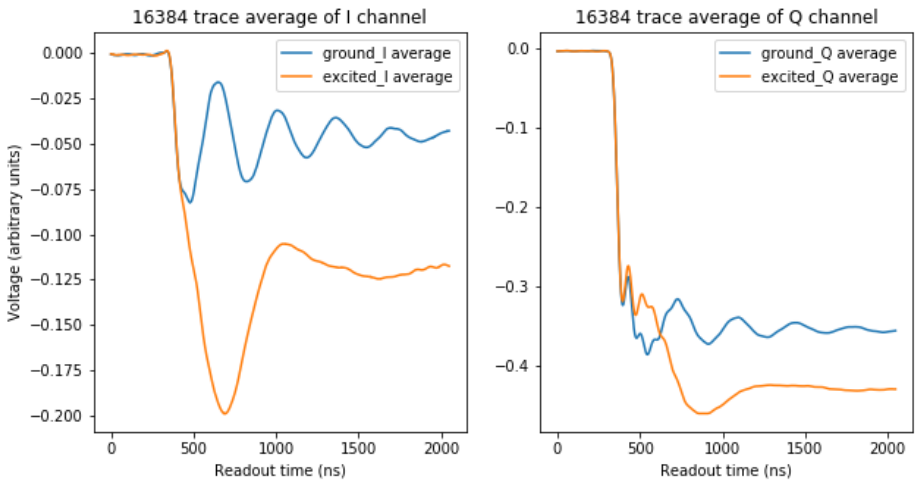

Figure 4: Example of the averaged voltage readout trace of the 16,384 ground and excited traces for a particular subdataset.

excited states. This stage carries essentially no physical information and is simply cropped from the data during preprocessing. The second stage is the integration stage, where the voltage starts to build towards the stable limit. The final stage is the stable stage, where the ground and excited voltage readouts are expected to saturate and separate from each other.

4 Methods

We first use traditional methods to generate a baseline to compare against. The baseline is a linear SVM and is applied to the average of the traces over timesteps across the I and Q channels. For our methods, we apply several standard binary classification models to cropped trace data without doing any averaging over timesteps. We describe our methods in further detail below.

4.1 Traditional Baseline

Traditional methods of qubit readout usually rely on using the saturation voltages, and hence require a long readout time. Specifically, the average voltage across timesteps is computed for each channel in each individual readout trace. Since there are two channels, I and Q, this gives 2-dimensional coordinates for each individual readout. Figure 5 shows a scatter plot of the ground state and excited state qubits in the IQ voltage space for a particular subdataset. The traditional method of threshold cutting in such a 2-dimensional space utilizes the linear support vector machine. Figure 5 shows the result of linear SVM and its accuracy on one of the subdatasets.

We performed the traditional method across all 1,116 subdatasets, and obtained a classification accuracy for each subdataset. Figure 6 shows the accuracy map for a 300 ns readout time. To better understand the performance of traditional method, we performed the same analysis for six different readout time windows. Figure 7 shows the dependence of readout fidelity on the readout power levels and readout time. It is expected that as the readout time window increases, the readout fidelity becomes better, since more information about the saturation voltages are being included. It is noteworthy that traditional method performance levelled off after about 600 ns readout time, after which not much fidelity improvement was observed.

4.2 Classification Methods

We apply a suite of standard binary classification models to predict the qubit readout directly from its trace. Note that in contrast to the method used in the traditional baseline, we do not average the traces across timesteps to generate a 2-dimensional feature. Instead, we simply crop the trace to

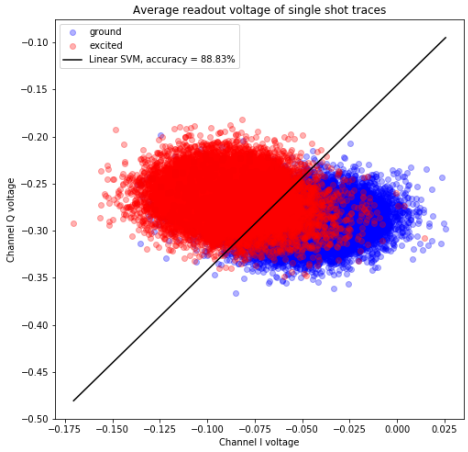

Figure 5: Voltage averages for each individual single shot trace. Each point in the plot corresponds to channel I and channel Q voltage coordinates of an individual single shot trace. The average of the single shot trace in this example are taken from the 300 ns readout time window.

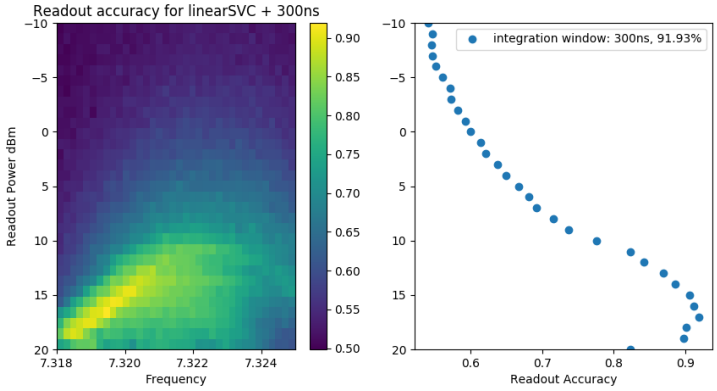

Figure 6: Each pixel in the frequency cross readout power landscape indicates the classification accuracy of using the traditional method on 300 ns readout time. The right panel shows the best accuracy we could achieve at each readout power level.

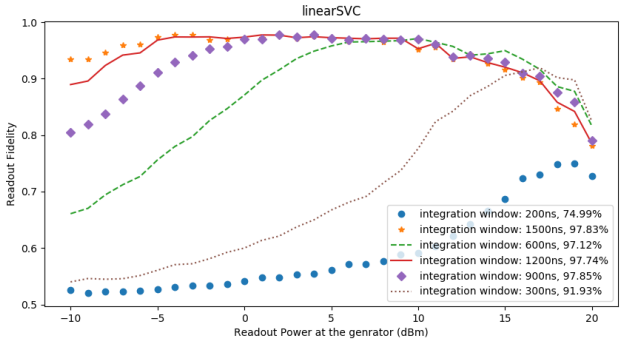

Figure 7: The readout fidelity obtained from traditional method as a function of the readout power. Each data point in the plot shows the best fidelity obtained at that power level across all sampling frequencies. The plot shows the trend for six different readout time windows.

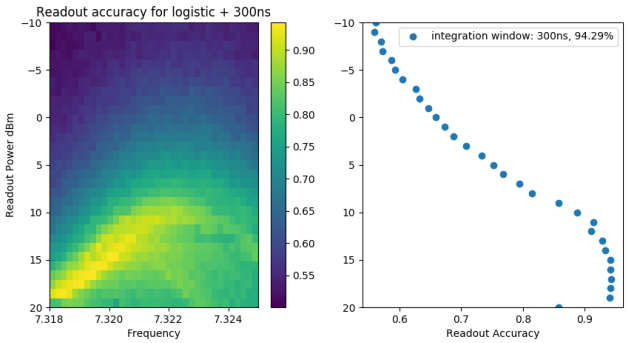

Figure 8: Accuracy map for logistic regression method and 300 ns of readout time.

simulate a shorter timestep, and use the cropped trace directly in our models. Since we are interested in minimizing the readout time of our methods to avoid qubit decoherence, we experiment with various-sized crops ranging from 200 ns to 1500 ns, which we call integration windows. Shorter integration windows give models with faster readout times.

We separately train models for each subdataset, and compare the best performance of each classifier across different power levels and different readout times. For each classifier, we tuned the hyperparameters using 5-fold cross validation using either the entire dataset or a representative selection of the subdatasets, depending on how long the classifier takes to run. Below, we list the classifiers that we used.

- **Logistic Regression.** We use the standard logistic regression setup for binary classification, with the L-BFGS optimizer.
- **Naive Bayes.** We use the binary classification version of Naive Bayes.
- **Random Forest.** We use standard random forest with 50 trees.
- **AdaBoost.** We use AdaBoost with 50 estimators.
- **Quadratic Discriminant Analysis.** We use QDA, which fits class conditional Gaussian densities to the data and uses Bayes rule to get a quadratic decision boundary.

5 Results

Since our data labels are balanced between the two classes, we measure the performance of our models with classification accuracy (readout fidelity). Note that we do not care about achieving high accuracy for every combination of frequency and readout power. Instead, we are searching over the space to find the best such combination.

Figure 8 shows the accuracy map landscape of using logistic regression method with 300 ns readout time window. Figure 9 shows the dependence on readout power levels for different classification models we used. With a 300 ns readout time window, our best method shows 2.36% improvement over the traditional baseline methods. It is noteworthy that all methods except QDA outperform the baseline method. This is likely because QDA assumes the underlying data follows a Gaussian distribution, which is not the case for our readout voltage traces.

Similar to the traditional baseline method, all methods show increasing performance with longer readout times. Figure 10 shows an example of logistic regression with various readout time windows. Upon comparison with Figure 7, we find that logistic regression performs better than the baseline across almost all readout times. We see less improvement over the baseline with longer readout times, as even the baseline method can achieve a relatively high fidelity if allowed a sufficiently long readout time. For example, with a very short readout time of 200 ns, the traditional baseline method achieves only 75% fidelity. Figure 11 shows the comparisons.

It is noteworthy that Figure 11 shows a clear increasing trend in fidelity as the readout power increases. This is expected as higher readout power means the hardware has a faster integration time

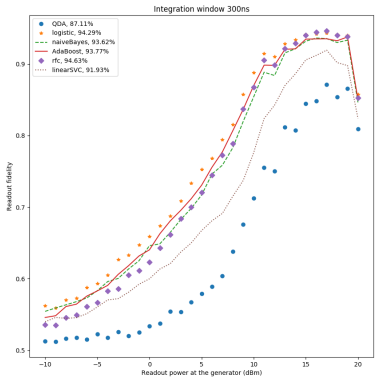

Figure 9: Readout fidelity of different classification methods as a function of different levels of readout power.

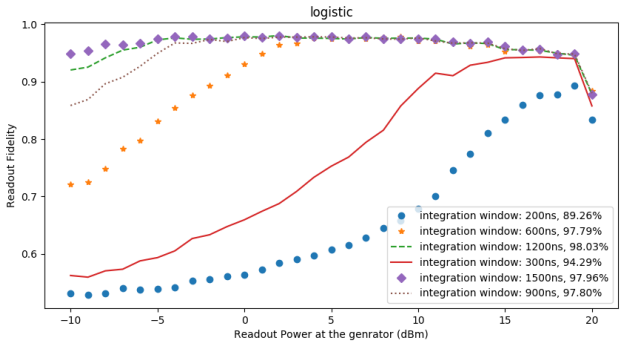

Figure 10: Best readout fidelity across all readout frequency as a function of readout power for logistic regression. Generally, longer readout time provides higher fidelity.

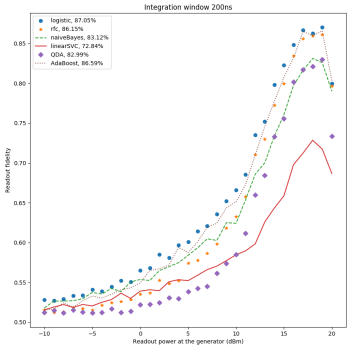

Figure 11: Readout fidelity of different classification methods as a function of different levels of readout power at 200 ns readout time. Our best method has about 15% fidelity improvement from the baseline.

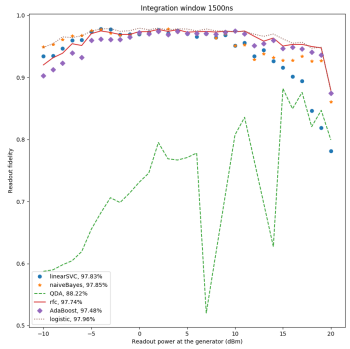

Figure 12: Readout fidelity of different classification methods as a function of different levels of readout power at 1500 ns readout time. The benefit of having higher readout power is not as significant anymore since longer readout time is used.

and hence the stable voltage is reached faster. Therefore, higher readout power compensates for a shorter readout time. However, this benefit diminishes as we include longer readout time. Figure 12 shows the fidelity trend at 1500 ns readout time. Clearly, many methods achieve similar fidelity across all readout power levels.

6 Conclusions

We improved over traditional baseline method in qubit readout state differentiation. Our methods achieve higher fidelity for the same readout time. This improvement is more significant as the readout time becomes shorter, which also enables us to do faster qubit readout.

7 Future Work

Classical machine learning techniques are able to improve both the readout accuracy and shorten the readout time required. However, such methods are not robust against changes in hardware configurations. Variations in the readout chain could introduce changes in their performance. We hence aim to fine tune neural network structures that provide even higher accuracy with shorter readout time required, while being robust against readout hardware changes. We are investigating the performance of different neural network structures on a range of subdatasets. Currently single layer neural networks provide higher accuracy than traditional classification methods across a small range of subdatasets we examined. However, multilayer neural network structures examined so far have not performed better than single layer neural network. Robust neural network models could then be deployed onto FPGA boards to integrate as part of the in situ qubit readout chain.

8 Acknowledgements

We would like to thank Zhaoqi Leng from Andrew Houck's group at Princeton University Department of Physics for providing the readout data and hardware experiment support.

References

[1] Peter W Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM review*, 41(2):303–332, 1999.

[2] Scott Aaronson and Lijie Chen. Complexity-theoretic foundations of quantum supremacy experiments. *arXiv preprint arXiv:1612.05903*, 2016.

[3] Sabrina S Hong, Alexander T Papageorge, Prasahnt Sivarajah, Genya Crossman, Nicolas Dider, Anthony M Polloreno, Eyob A Sete, Stefan W Turkowski, Marcus P da Silva, and

Blake R Johnson. Demonstration of a parametrically-activated entangling gate protected from flux noise. *arXiv preprint arXiv:1901.08035*, 2019.

[4] Ani Nersisyan, Stefano Poletto, Nasser Alidoust, Riccardo Manenti, Russ Renzas, Cat-Vu Bui, Kim Vu, Tyler Whyland, Yuvraj Mohan, Eyob A Sete, et al. Manufacturing low dissipation superconducting quantum processors. *arXiv preprint arXiv:1901.08042*, 2019.

[5] JA Schreier, Andrew A Houck, Jens Koch, David I Schuster, BR Johnson, JM Chow, Jay M Gambetta, J Majer, L Frunzio, Michel H Devoret, et al. Suppressing charge noise decoherence in superconducting charge qubits. *Physical Review B*, 77(18):180502, 2008.

[6] Colm A Ryan, Blake R Johnson, Jay M Gambetta, Jerry M Chow, Marcus P da Silva, Oliver E Dial, and Thomas A Ohki. Tomography via correlation of noisy measurement records. *Physical Review A*, 91(2):022118, 2015.

[7] AH Myerson, DJ Szwer, SC Webster, DTC Allcock, MJ Curtis, G Imreh, JA Sherman, DN Stacey, AM Steane, and DM Lucas. High-fidelity readout of trapped-ion qubits. *Physical Review Letters*, 100(20):200502, 2008.

[8] MD Reed, L DiCarlo, BR Johnson, L Sun, DI Schuster, L Frunzio, and RJ Schoelkopf. High-fidelity readout in circuit quantum electrodynamics using the jaynes-cummings nonlinearity. *Physical review letters*, 105(17):173601, 2010.

[9] Christopher E Langer. *High fidelity quantum information processing with trapped ions*. PhD thesis, 2006.

[10] Easwar Magesan, Jay M Gambetta, Antonio D Córcoles, and Jerry M Chow. Machine learning for discriminating quantum measurement trajectories and improving readout. *Physical review letters*, 114(20):200501, 2015.

[11] Alireza Seif, Kevin A Landsman, Norbert M Linke, Caroline Figgatt, Christopher Monroe, and Mohammad Hafezi. Machine learning assisted readout of trapped-ion qubits. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 51(17):174006, 2018.

[12] Zi-Han Ding, Jin-Ming Cui, Yun-Feng Huang, Chuan-Feng Li, Tao Tu, and Guang-Can Guo. Fast and high-fidelity readout of single trapped-ion qubit via machine learning methods. *arXiv preprint arXiv:1810.07997*, 2018.

[13] Neereja M Sundaresan, Rex Lundgren, Guanyu Zhu, Alexey V Gorshkov, and Andrew A Houck. Interacting qubit-photon bound states with superconducting circuits. *Physical Review X*, 9(1):011021, 2019.

A Additional Results Figures

Below we include additional result figures from our experiments.

A.1 Additional Naive Bayes Results

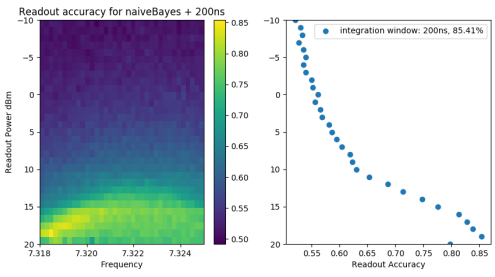

Figure 13: Accuracy map for Naive Bayes method and 200 ns of readout time.

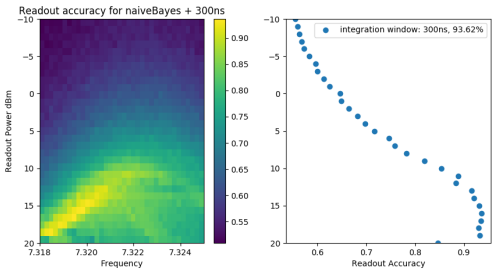

Figure 14: Accuracy map for Naive Bayes method and 300 ns of readout time.

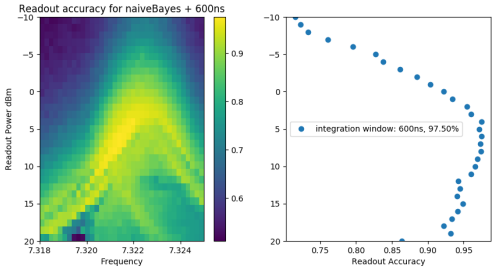

Figure 15: Accuracy map for Naive Bayes method and 600 ns of readout time.

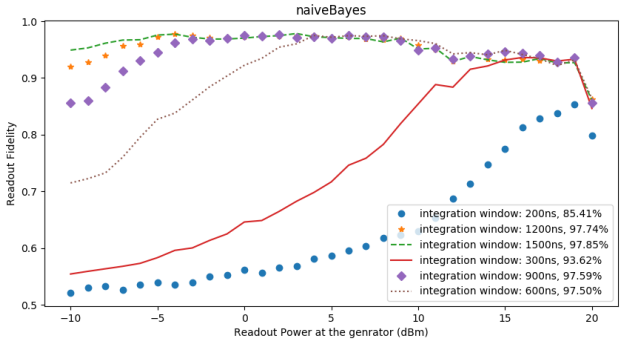

Figure 16: Best readout fidelity across all readout frequency as a function of readout power for Naive Bayes.

A.2 Additional Random Forest Results

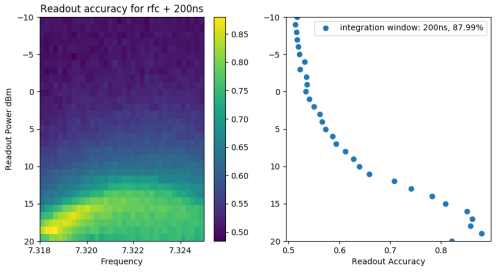

Figure 17: Accuracy map for random forest method and 200 ns of readout time.

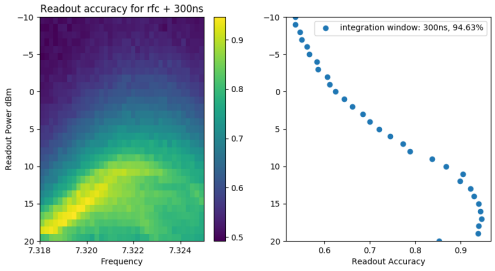

Figure 18: Accuracy map for random forest method and 300 ns of readout time.

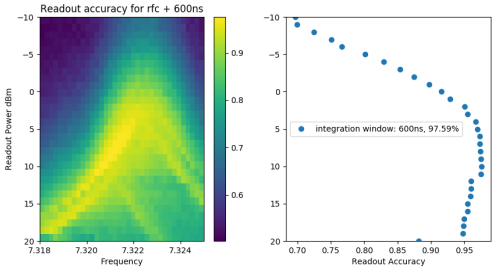

Figure 19: Accuracy map for random forest method and 600 ns of readout time.

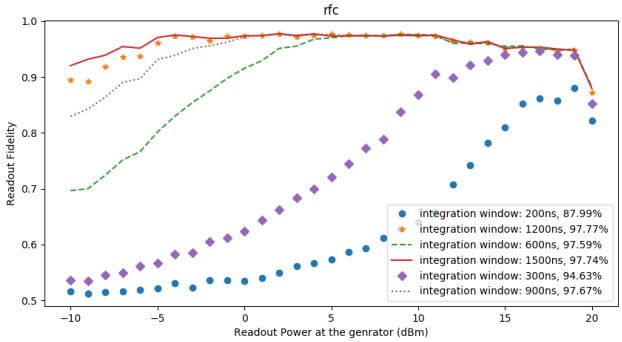

Figure 20: Best readout fidelity across all readout frequency as a function of readout power for random forest.

A.3 Additional AdaBoost Results

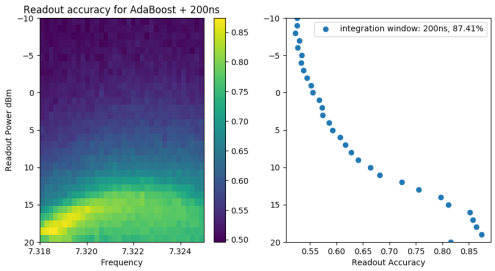

Figure 21: Accuracy map for AdaBoost method and 200 ns of readout time.

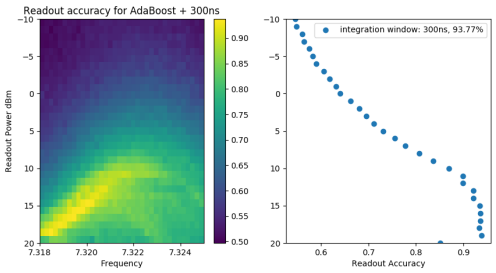

Figure 22: Accuracy map for AdaBoost method and 300 ns of readout time.

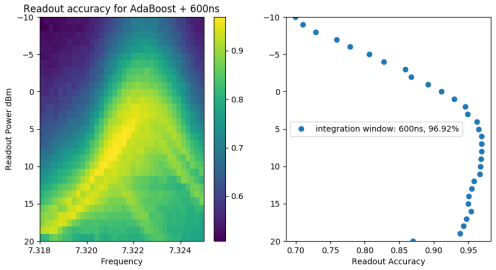

Figure 23: Accuracy map for AdaBoost method and 600 ns of readout time.

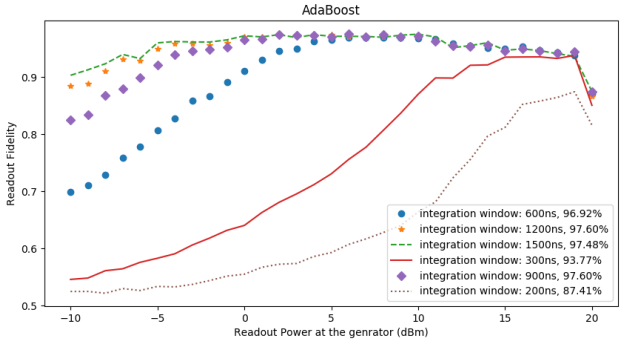

Figure 24: Best readout fidelity across all readout frequency as a function of readout power for AdaBoost.

A.4 Additional Quadratic Discriminant Analysis Results

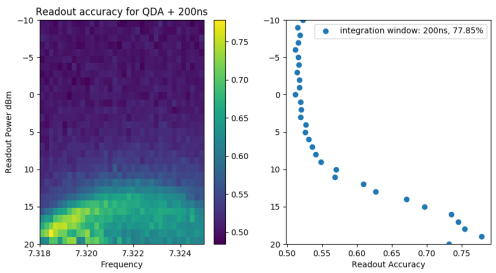

Figure 25: Accuracy map for QDA method and 200 ns of readout time.

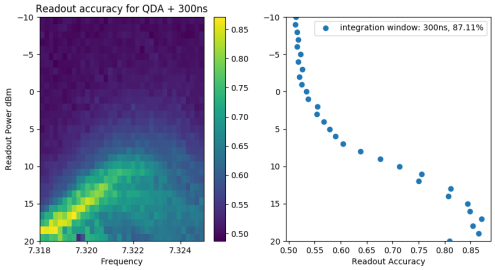

Figure 26: Accuracy map for QDA method and 300 ns of readout time.

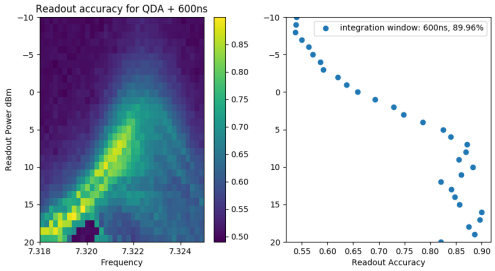

Figure 27: Accuracy map for QDA method and 600 ns of readout time.

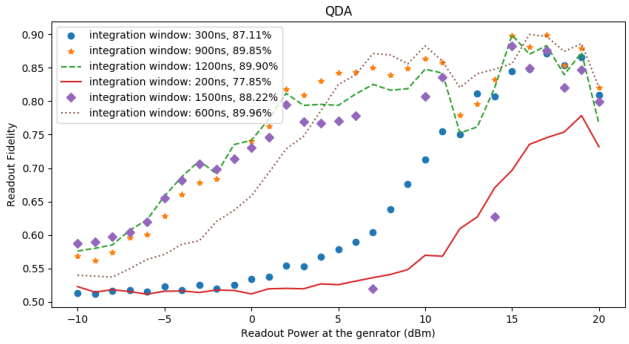

Figure 28: Best readout fidelity across all readout frequency as a function of readout power for QDA.