MACHINE LEARNING BASED DISCRIMINATION FOR EXCITED STATE PROMOTED READOUT

TECHNICAL REPORT

Utkarsh Azad[*]
Center for Computational Natural Sciences and Bioinformatics,
Center for Quantum Science and Technology,
International Institute of Information Technology,
Hyderabad, TS 500032
`utkarsh.azad@research.iiit.ac.in`

Helena Zhang
IBM Quantum
Yorktown Heights, NY, 10598

August 12, 2022

ABSTRACT

A limiting factor for readout fidelity for superconducting qubits is the relaxation of the qubit to the ground state before the time needed for the resonator to reach its final target state. To reduce this effect and further improve the readout contrast, excited state promoted (ESP) readout. In this work, different classification algorithms like k-nearest neighbors, decision trees, support vector machines, etc., were used on the real data from five qubit IBMQ devices to measure their effectiveness to implement efficient single-qubit and multi-qubit discrimination methods. These methods were compared to deep neural networks such as feedforward neural networks based on their qubit-state-assignment fidelity performance, robustness to readout crosstalk, and training time.

Keywords Quantum Computing · Qubit Readout · Machine Learning

1 Introduction

Quantum computers are speculated to have a computational edge over their classical counterparts in solving problems in various areas such as quantum chemistry [1], molecular simulations [2], machine learning [3], etc. To achieve such an advantage, quantum hardware requires millions of good quality qubits. However, the present-day hardware, generally referred to as noisy intermediate-scale quantum (NISQ) hardware, contains not more than a hundred qubits, which severely limits their computational capabilities. The major problem in scaling up these devices is qubit control and readout difficulty. Hence, considerable work is required to retain and further improve these systems if we want to increase their size and complexity, particularly for combating errors at all stages of the computational pipeline: initialization, execution, and readout.

In this work, we present machine-learning-enabled qubit-state discrimination utilizing excite-state-promoted (ESP) readout [4], which a way to improve qubit readout fidelity in a scalable way by using extra levels of transmon, i.e., exciting the $|1\rangle$ state to the $|2\rangle$ state for readout. This essentially changes the discrimination problem from a two-state (Fig. 2a) system to a three-state (Fig. 2b) system. We evaluate the qubit-state discrimination performance of various machine learning models such as k-nearest neighbors (KNN), decision trees, Gaussian naive Bayes (GNB), linear and quadratic discriminant analysis (LDA and QDA), and a fully-connected neural network (FNN). To evaluate these different qubit-state discriminator techniques, we use ESP readout outputs from five of the five-qubit IBMQ hardware: *ibmq_rome*, *ibmq_bogota*, *ibmq_merlin*, *ibmq_belem* and *ibmq_quito*. We show that for these systems, we examine the qubit-state assignment performance using a confusion matrix and the cross-fidelity metric. We show that classifiers based on FNN and GNB perform out par LDA and QDA, which are standardly used for the multi-class classification such as multi-qubit discrimination.

---

[*] This work has been done as part of the Qiskit Advocate Mentorship Program (QAMP).

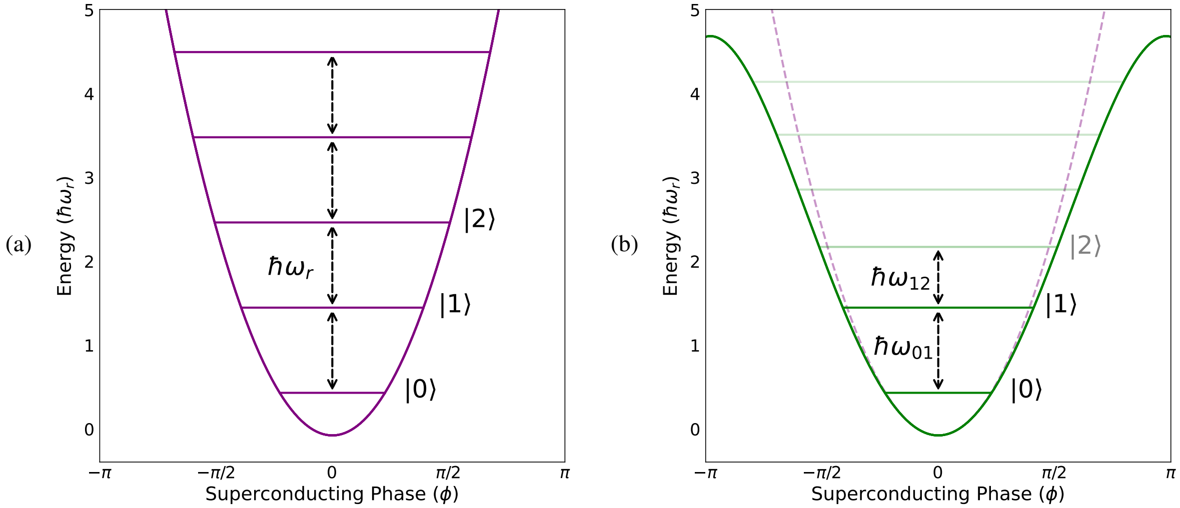

Figure 1: (a) Energy levels are equidistantly spaced by $\hbar\omega_r$ in the energy potential for the quantum harmonic oscillator (QHO). (b) Inclusion of Josephson inductance leads to non-equidistant energy levels as it changes the quadratic energy potential (dashed purple) into sinusoidal potential (solid green), i.e., by inclusion of higher order terms. The sufficient difference in the energy spacing $\hbar\omega_{01}$ and $\hbar\omega_{12}$ allows us to form computational basis by isolating two lowest energy levels $|0\rangle$ and $|1\rangle$ of the transmon. Both figures have been adapted from [5].

2 Theory of Qubits and their Readout

Over the past two decades, superconducting qubits have emerged as a leading quantum computing platform which has been used by various leading industries such as IBM [6], Google [7], Rigetti [8], etc. The basis of quantum hardware for all of them is a particular kind of superconducting qubit, popularly known as the transmon qubit, which is a superconducting circuit composed of Josephson junction and capacitor. The Hamiltonian of such a system can be described as:

$$H = 4E_C n^2 - E_J \cos(\phi), \tag{1}$$

where E_C, E_J denote the energies of the capacitor and Josephson junction present in the superconducting circuit, n is the reduced charge number operator and ϕ is the reduced flux across the Josephson junction with $\hbar = 1$.

2.1 Transmon qubits

In principle the variable ϕ can acquire a range of values, but in the regime where $\phi \to 0$, the system begins to behave as a transmon qubit. This allows us to approximate H (1) by performing a taylor expansion of the $E_J \cos(\phi)$ (ignoring constant terms):

$$\lim_{\phi\to 0} E_J \cos(\phi) \approx \frac{1}{2!}E_J\phi^2 - \frac{1}{4!}E_J\phi^4 + \mathcal{O}(\phi^6). \tag{2}$$

where the quadratic term ϕ^2 defines the standard quantum harmonic oscillator (1a), and the next subsequent higher-order terms contribute to anharmonicity in the system. This is important for the system for isolating the two lowest energy levels $|0\rangle$ and $|1\rangle$ and determining a computation basis (1b), which would not be possible in the case of the standard quantum harmonic oscillator due to presence of the equidistant energy levels.

We can further show that the system now resembles a Duffing oscillator with the Hamiltonian H_D by using the relations $n \sim (a - a^\dagger), \phi \sim (a + a^\dagger)$, where a and $a^\dagger$ are creation and annihilation operators for the qubit system:

$$H = \omega a^\dagger a + \frac{\alpha}{2} a^\dagger a^\dagger a a, \tag{3}$$

Here, ω corresponds to the ω_{01}, i.e., the excitation frequency from the ground state to the first excited energy state ($0 \rightarrow 1$), and α is the anharmonicity between the excitation frequencies ω_{01} and ω_{12}. By tuning the $|\alpha|$ to sufficiently large values, one can isolate the standard 2-dimensional subspace by suppressing leakage to the higher energy states (1).

2.2 Dispersive Readout

The ability to perform high fidelity readout of the qubit states is a crucial cornerstone of any quantum processor. The most common technique utilized in the circuit QED architecture is that of dispersive readout. In this method, the qubit (the quantum system) is entangled with an observable of a superconducting resonator (the probe), allowing us to gain information about the qubit state by interrogating the resonator - rather than directly interacting with the qubit. Therefore, readout performance depends on the signal-to-noise ratio of a microwave pulse tone sent to the resonator while minimizing the unwanted back-action on the qubit. In this regard, the qubit-resonator interaction is described by the Jaynes–Cummings Hamiltonian -

$$H_{JC} = \omega_r \left(a^\dagger a + \frac{1}{2} \right) + \frac{\omega_q}{2} \sigma_z + g(\sigma_+ a + \sigma_- a^+) \tag{4}$$

Here, ω_r and ω_q are the resonator and qubit frequencies, respectively. Whereas the transverse qubit-coupling resonator rate is shown by g and $a^\dagger(a)$ represents the creation (annihilation) operator for the qubits.

Now skipping some complex physics regarding how measurement is done experimentally, we rather focus on the readout event itself. It begins with a short microwave tone directed to the resonator at the resonator probe frequency ω_{RO} (the carrier frequency), which acquires the following form after interacting with the resonator:

$$s(t) = A_{RO} \cos(\omega_{RO} t + \theta_{RO}) = Re\Big\{ A_{RO} e^{j(\omega_{RO} t + \theta_{RO})} \Big\} \tag{5}$$

where A_{RO} and θ_{RO} are the qubit-state dependent amplitude and phase that we wish to measure and Re represents the real part of an expression. We can rewrite equation 5 in the "phasor" notation as follows:

$$s(t) = Re\left\{ \underbrace{A_{RO} e^{(\theta_{RO})}}_{phasor} e^{j(w_{RO} t)} \right\} \tag{6}$$

To perform qubit readout, we want to measure the "in-phase" component I and a "quadrature" component Q of the complex number represented by the phasor, to determine the amplitude A_{RO} and the phase θ_{RO}:

$$A_{RO} e^{(\theta_{RO})} = A_{RO} \cos(\theta_{RO}) + j A_{RO} \sin(\theta_{RO}) \equiv I + jQ \tag{7}$$

2.3 Discriminator

For performing the readout, we extract the I and Q components from the readout signal, which we can plot on an I-Q plane, similar to one shown in Fig. 2. It is seen that corresponding to the actual state of the qubit, the I-Q values tend to lie in specific clusters on the plane, each corresponding to one specific energy state. For example, in Fig. 2a, we see that two clusters marked by blue and red points correspond to the qubit states $|0\rangle$ and $|1\rangle$ states, respectively. Now, given the IQ data on the plane, we incorporate a discriminator (or a classifier) to find the boundaries of the cluster formed by each state, so that for the next incoming IQ output, we can predict the corresponding unknown state with sufficient confidence.

3 Excited State Promoted Readout

Excited State Promoted (ESP) readout is a way to improve qubit readout fidelity for superconducting qubits in a scalable way by using extra level of transmon, i.e., by exciting the $|1\rangle$ state to the $|2\rangle$ state for readout. This essentially changes the discrimination problem from a two-state system to a three-state system.

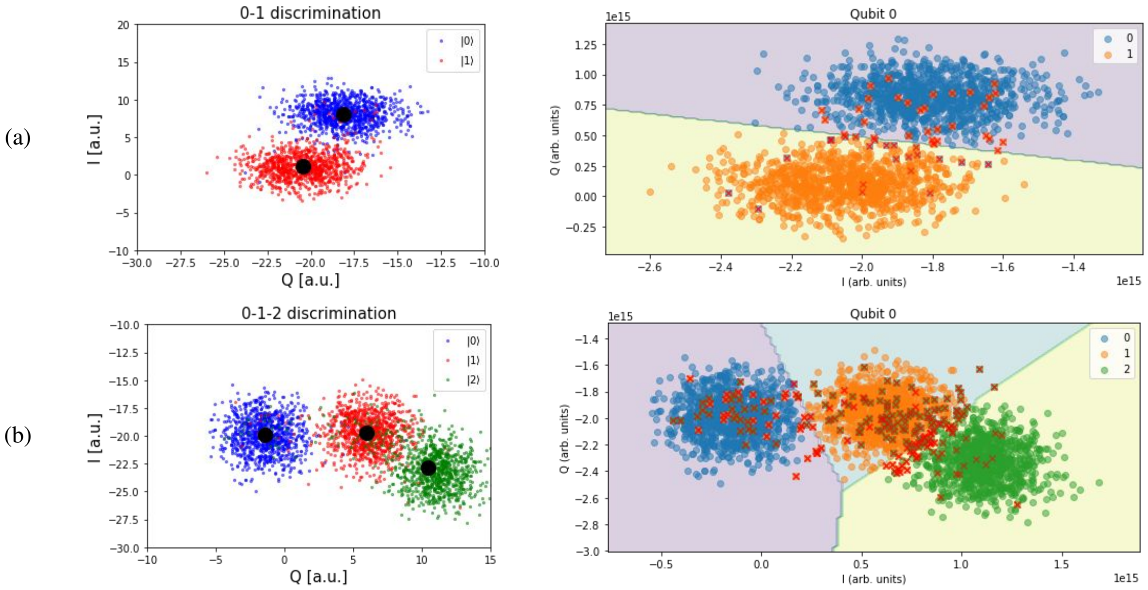

Figure 2: IQ plots for qubit states and their classification via a discriminator. The data presented here was acquired from *ibmq_armonk*.

3.1 Theory

In excited state promoted readout, we take advantage of higher excited states of the transmon by exciting the $|1\rangle$ state to $|2\rangle$ and then perform our measurements since $|2\rangle \rightarrow |0\rangle$ transition should be much more difficult. In order to do this, we first calibrate the frequencies amplitude of the π-pulse required for the $|0\rangle \rightarrow |1\rangle$ and the $|1\rangle \rightarrow |2\rangle$ transition using frequency spectroscopy and rabi experiment, respectively. Doing it for the former transition is straightforward and requires building only a gaussian wave packet. However, to assist the latter one, we use a sinusoidal sideband which allows us to change the local oscillator frequency without manually setting it.

3.2 Example

In Fig. 2, we present the ESP readout for the *ibmq_armonk* hardware, which is an open-access Canary r1.2 one-qubit hardware. The approximate ω_{01} frequency is 4.972 GHz, and the anharmonicity α is 347.19 MHz. We show the IQ plots for the $|0\rangle$-$|1\rangle$ state discrimination in Fig. 2a and the corresponding discrimination by the linear discriminant analysis (LDA). In the subsequent experiment, we use ESP and show the IQ plots in Fig. 2b. Notice, the overlap (marked by red crosses) between the $|0\rangle$ and $|1\rangle$ states decreases by exciting the $|1\rangle$ state to $|2\rangle$ state.

4 Machine Learning based discriminators

In principle, we can use various machine learning methods to classify between the system's different states. The first model that we look at is the k-nearest neighbor (KNN) classifier, which implements learning based on the k nearest neighbors of the given point for which a decision has to be taken. In our case, we use $k = 50$, and for determining proximity between the points, we calculate the "manhattan distance". The second model that we look at is the decision tree classifier (DTC), where the decision for the data is taken by continuously splitting it according to a certain parameter or set of rules. In our model, we use entropy to measure the quality of a split and restrict the depth of the decision tree to 20. The third model is the classifier based on the Gaussian Naive Bayes algorithm (GNB). The fourth and fifth models are the linear, and quadratic discriminant analysis (LDA and QDA) based classifiers. The final model we look at is a deep learning-based model called the fully connected neural network (FNN). Our FNN architecture (Fig. 3) is composed of three hidden layers (1st, 2nd, and 3rd layer consist of 1000, 500, and 300 nodes, respectively) that use ReLU activation functions, and the 3^N output layer has *softmax* activation. The network is trained (validation-training set ratio of 0.3) using the *Adam* optimizer with categorical cross-entropy as the loss function.

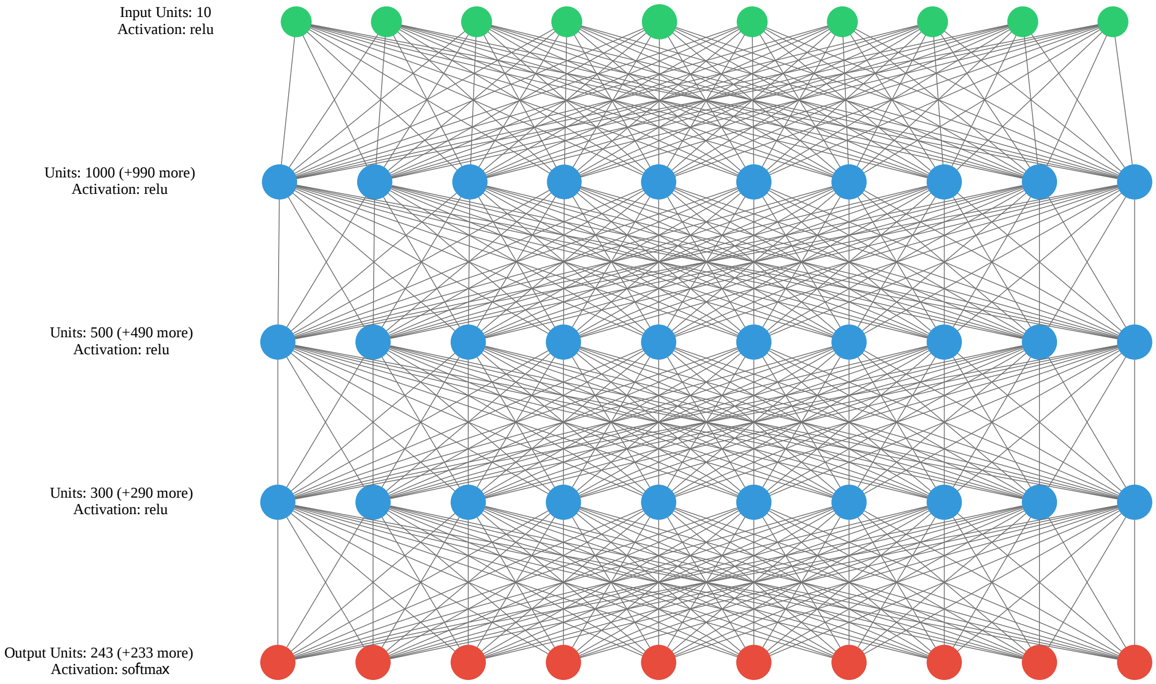

Figure 3: Fully connected Neural Network mdoel. It composes of three hidden layers of 1000, 500, and 300 nodes, respectively, that use ReLU activation functions. The 3^N output layer has *softmax* activation.

5 Results

In this section, we describe the details of our results for our five-qubit excited-state promoted (ESP) readout experiment results, comparing the performances of our six models described in the previous section: KNN, DTC, GNB, LDA, QDA and FNN, for the five 5-qubit IBMQ hardware: *ibmq_rome*, *imbq_bogota*, *ibmq_belem*, *ibmq_quito*, and *ibmq_manila*.

5.1 Data Accumulation and Preparation

To perform the excited-state promoted (ESP) readout experiments, we first performed frequency calibrations and amplitude calibration experiments for each individual qubit q_i on each given hardware to determine: ω_{01}, ω_{12}, A_{01} and A_{02}, for building the π_{01} and π_{12} pulses. In the next step, we build the pulses schedule for 3^N possible states, where $N = 5$ in our case. We show an example pulse schedule for preparing the state $|22102\rangle$ on *ibmq_bogota* in Fig. 6. We perform 2048 shot measurements for every such state to obtain the IQ data corresponding to each state and shot. This makes our overall data from given hardware of the size $(243, 2048, 5)$. We flatten this data for preparing test and train sets to $(243 \times 2048, 10)$, where 10 comes from splitting each of the $I + jQ$ data points into two separate numbers I and Q. We then perform outlier removal using an elliptic envelope strategy owing to the fact that individual distributions mainly follow a normalized gaussian distribution (Fig. 5). We then scale the data for individual qubits using *StandardScaler* method from the sk-learn library to impose uniformity in the data points. Finally, we split the data into test-train-validation set in the proportion $(50 : 30 : 20)$. Subsequently, since we are looking at supervised learning, we also prepare the label data for training our models. To do this, we label each state by the number represented by its corresponding bitstring. For example, $|22102\rangle$ would be $2 \times 3^0 + 0 \times 3^1 + 1 \times 3^2 + 2 \times 3^3 + 2 \times 3^4 = 227$. However, for FNN, we encode these integer labels as one-hot encoded binary vectors of size 243.

5.2 Comparision metrics

In order to compare the performances between different models, we compute the qubit-state-assignment fidelities $\mathcal{F}$, i.e., the measure of how many times the qubit predicted state for q_i matches with the correct state that it was in during measurement. For a qubit q_i we define it as follows:

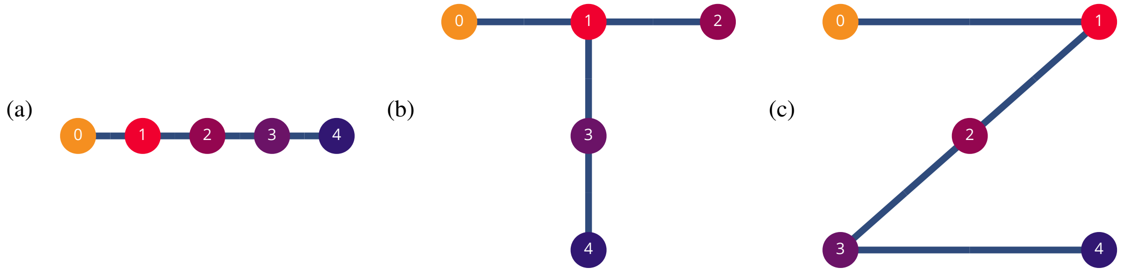

Figure 4: Topologies of the 5 qubit IBMQ hardware from which data was acquired: (a) *ibmq_rome* and *imbq_bogota*, (b) *ibmq_belem* and *ibmq_quito*, (c) *ibmq_manila*

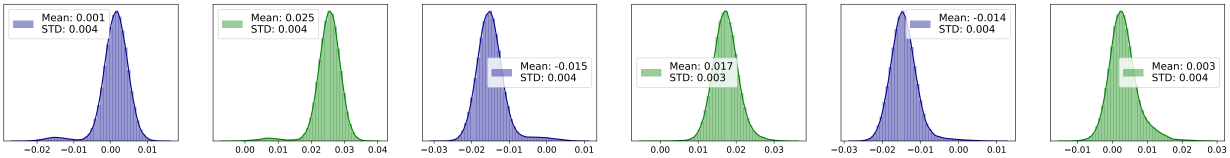

Figure 5: Probability distribution of IQ data obtained for the qubit q_4 from *imbq_belem* for states (a) $|0\rangle$, (b) $|1\rangle$ and (c) $|2\rangle$. The graphs in blue corresponds to the “in-phase” component I and the ones in green corresponds to the “quadrature” component Q.

$$\mathcal{F}_i = 1 - [P(0_i|\pi_{0\to1}) + P(0_i|\pi_{0\to2}) + P(1_i|\pi_{0\to0}) + P(1_i|\pi_{0\to2}) + P(2_i|\pi_{0\to0}) + P(2_i|\pi_{0\to1})]/6 \tag{8}$$

As you would notice, we are looking at subtracting the deviations (or mismatched) results from the total probability, i.e., 1. But the question then arises about the sources of these deviations. In principle, these deviations exist due to initialization errors, state transitions during the measurement, and readout crosstalk. The model that better learns about these deviations will attain a better assignment fidelity score.

Now, once we have defined $\mathcal{F}_i$ for each individual qubit q_i, we finally compare the discriminators using system-state-assignment fidelity $\mathcal{F}_{GM}$, which is described at the geometric mean all of the qubit-state-assignment fidelities,

$$\mathcal{F}_{GM} = (\mathcal{F}_1\mathcal{F}_2\mathcal{F}_3\mathcal{F}_4\mathcal{F}_5)^{1/5} \tag{9}$$

5.3 Key Observations

Now let us look at some key observations from our experiments below.

5.3.1 Single-qubit discrimination

For measuring the capabilities for single-qubit discrimination, for each qubit q_i on a given hardware, we look at the states where rest of the other states reside in their ground state, i.e., we look at the assignment fidelities for the states $|0\rangle^{\otimes i} \otimes |q_i\rangle \otimes |0\rangle^{\otimes (N-1)-i}$, where $|q_i\rangle \in \{|0\rangle, |1\rangle, |2\rangle\}$. We find that for the single-qubit discrimination task, almost all of the discriminating models (classifiers) performed almost similarly after hyperparameter tuning was performed. We expect that this could be attributed to ease of learning boundaries on a 2-D for the three normalized distributions with restricted overlaps.

5.3.2 Multi-qubit discrimination

For measuring the performance for multi-qubit discriminations, for each hardware, we prepared all of the qubits in all possible 3^N states. We then measured the individual $\mathcal{F}_i$ for each individual qubit and also the overall $\mathcal{F}_{GM}$ for each model. We present these results in the Table 1. We see that for except for *imbq_rome*, FNN could not satisfactorily outperform other machine learning models such as GNB and QDA. Overall, the trends show that even the multi-qubit discrimination performance for many models remains similar. This might be because either all of them could learn the

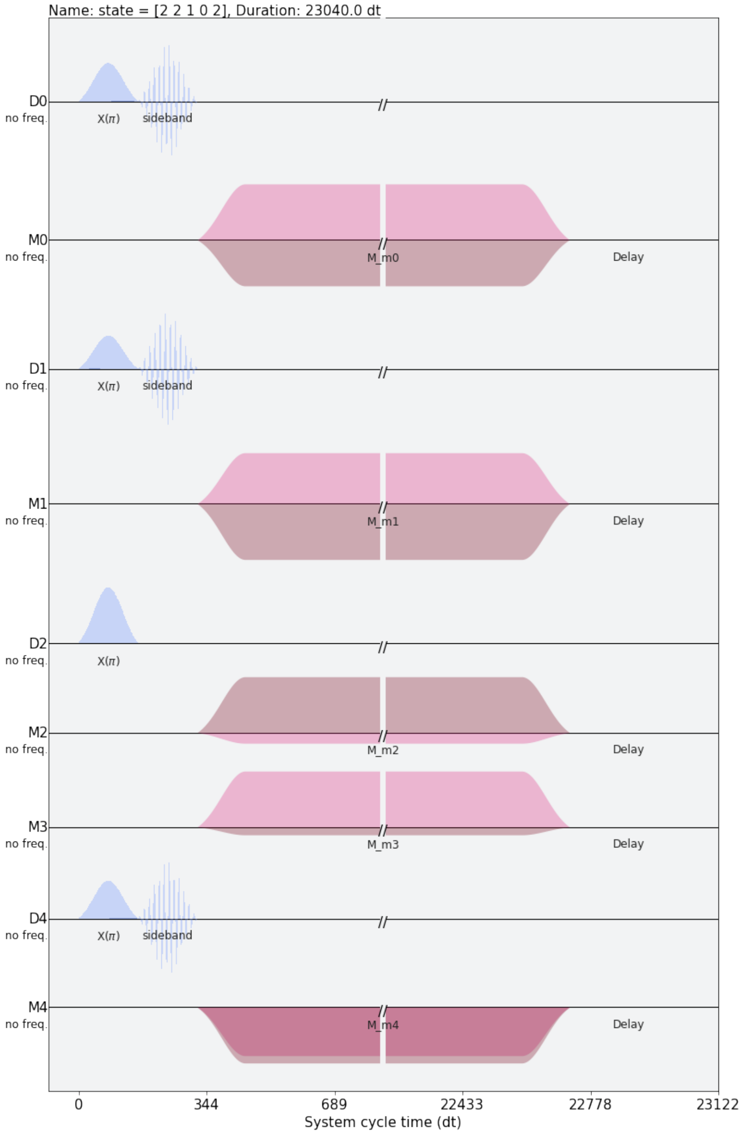

Figure 6: Pulse schedule for preparing state $|22102\rangle$ in *ibmq_bogota*

.

effect of crosstalk in the data or the overlaps present in-between state data points (due to noise) make it equally hard for them. It is seen that GNB might have a slight edge over the other classifiers due to the distribution followed by the IQ plots. In fact, we speculate that if the FNN could be designed to exploit this fact, its performance might be improved further. To further illustrate our thinking, we look at the confusion matrices for one of the hardwares in Fig. 7. We see that all of the states misassigned by each model have a lot of similarities.

5.3.3 Accuracy v/s Training time tradeoff

We fitted (or trained) all of our models for an explicitly similar amount of data points (243×1024). While DTC was easily the most expensive model when it comes to training time, the GNB came out to be quickest and that too with good single-qubit and multi-qubit state-assignment-fidelities. For FNN we saw that due to the presence of big density layers, the network was soon overfitting. However, at the same time, the big density layers is important as including some dropout might lead to loss of connection between different attributes of the data, which might have been essential, especially for this particular case of multi-class classification. For this reason, even the training time of FNN was competitive with the rest of the models, however, slower than when compared to GNB.

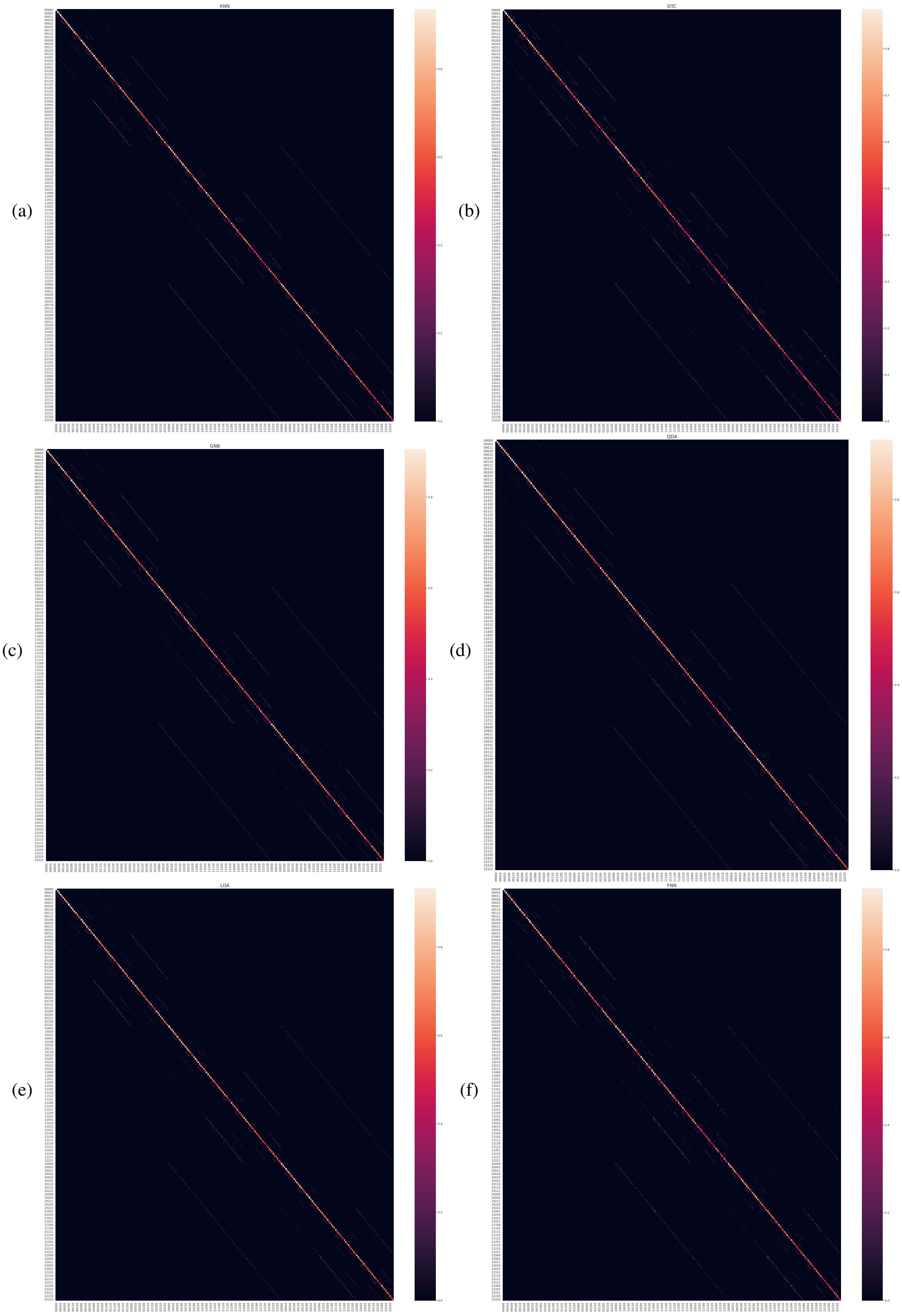

Figure 7: Confusion matrices for multi-qubit state assignment for *ibmq_bogota* for the machine learning models (a) KNN, (b) DTC, (c) GNB, (d) QDA, (e) LDA, and (f) FNN.

Table 1: Qubit-assignment fidelity comparison across hardware and classification methods.

Hardware		**KNN**	**DTC**	**GNB**	**QDA**	**LDA**	**FNN**
ibmq_rome	$\mathcal{F}_1$	0.913	0.891	0.918	0.919	0.918	0.921
	$\mathcal{F}_2$	0.918	0.914	0.925	0.927	0.926	0.926
	$\mathcal{F}_3$	0.975	0.963	0.978	0.978	0.978	0.979
	$\mathcal{F}_4$	0.940	0.936	0.943	0.945	0.943	0.945
	$\mathcal{F}_5$	0.924	0.910	0.930	0.932	0.930	0.930
	$\mathcal{F}_{GM}$	0.934	0.933	0.939	0.939	0.939	0.940
ibmq_bogota	$\mathcal{F}_1$	0.927	0.943	0.943	0.943	0.940	0.942
	$\mathcal{F}_2$	0.941	0.953	0.953	0.953	0.951	0.952
	$\mathcal{F}_3$	0.969	0.975	0.975	0.973	0.975	0.974
	$\mathcal{F}_4$	0.980	0.984	0.983	0.983	0.983	0.982
	$\mathcal{F}_5$	0.896	0.916	0.915	0.914	0.910	0.912
	$\mathcal{F}_{GM}$	0.937	0.890	0.943	0.946	0.946	0.944
ibmq_belem	$\mathcal{F}_1$	0.957	0.946	0.958	0.958	0.958	0.956
	$\mathcal{F}_2$	0.950	0.936	0.950	0.950	0.950	0.947
	$\mathcal{F}_3$	0.902	0.882	0.904	0.903	0.902	0.901
	$\mathcal{F}_4$	0.987	0.983	0.988	0.988	0.987	0.988
	$\mathcal{F}_5$	0.977	0.971	0.979	0.979	0.978	0.978
	$\mathcal{F}_{GM}$	0.954	0.943	0.955	0.955	0.955	0.954
ibmq_quito	$\mathcal{F}_1$	0.942	0.927	0.943	0.943	0.943	0.940
	$\mathcal{F}_2$	0.952	0.941	0.953	0.953	0.953	0.951
	$\mathcal{F}_3$	0.974	0.969	0.975	0.975	0.973	0.975
	$\mathcal{F}_4$	0.982	0.980	0.984	0.983	0.983	0.983
	$\mathcal{F}_5$	0.912	0.896	0.916	0.915	0.914	0.910
	$\mathcal{F}_{GM}$	0.952	0.942	0.954	0.954	0.953	0.952
ibmq_manila	$\mathcal{F}_1$	0.922	0.904	0.926	0.925	0.926	0.917
	$\mathcal{F}_2$	0.915	0.895	0.919	0.918	0.919	0.910
	$\mathcal{F}_3$	0.908	0.887	0.913	0.912	0.913	0.903
	$\mathcal{F}_4$	0.922	0.903	0.925	0.924	0.925	0.916
	$\mathcal{F}_5$	0.944	0.931	0.948	0.947	0.948	0.939
	$\mathcal{F}_{GM}$	0.922	0.904	0.926	0.925	0.926	0.917

References

[1] Yudong Cao, Jonathan Romero, Sukin Olson, Sim, Libor Veis, and et al. Aspuru-Guzik, Alán. Quantum chemistry in the age of quantum computing. *Chemical Reviews*, 119(19):10856–10915, 2019. PMID: 31469277.

[2] Xiao Yuan, Suguru Endo, Qi Zhao, Ying Li, and Simon C. Benjamin. Theory of variational quantum simulation. *Quantum*, 3:191, October 2019.

[3] Iris Cong, Soonwon Choi, and Mikhail D. Lukin. Quantum convolutional neural networks. *Nature*, 15(12):1273–1278, August 2019.

[4] François Mallet, Florian R. Ong, Agustin Palacios-Laloy, François Nguyen, Denis Bertet, and Daniel Esteve. Single-shot qubit readout in circuit quantum electrodynamics. *Nature Physics*, 5(11):791–795, November 2009.

[5] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver. A quantum engineer's guide to superconducting qubits. *Applied Physics Reviews*, 6(2):021318, June 2019.

[6] IBM Quantum. https://quantum-computing.ibm.com/, 2021.

[7] Cirq Developers. Cirq, March 2021.

[8] Robert S. Smith, Michael J. Curtis, and William J. Zeng. A Practical Quantum Instruction Set Architecture. *arXiv e-prints*, August 2016.