Efficient Quantum Measurement using Cumulative Distribution Functions

Zachery Utt
Depertment of Computer & Information Science & Engineering
University of Florida, Gainesville, Florida, USA

***Abstract*—Quantum measurement is one of the critical steps in quantum computing that determines the probabilities associated with qubit states after conducting several circuit executions and measurements. As a mesoscopic quantum system, real quantum computers are prone to noise. Therefore, a major challenge in quantum measurement is correct interpretation of noisy results from a quantum computer. While there are promising classification based solutions, they either produce incorrect results (misclassify) or require many measurements (expensive). In this paper, we present an efficient technique to estimate quantum state through analysis of probability distributions of post-measurement data. Specifically, we estimate the quantum state using cumulative distribution functions to compare the measured distribution of a sample with the distributions of basis states $|0\rangle$ and $|1\rangle$. Our experimental results demonstrate a drastic reduction (88%) in single qubit readout error. We also show significant reductions in error (12%) when used to boost existing multi-qubit discriminator models, and demonstrate efficacy on higher dimensional quantum systems.**

***Index Terms*—Quantum Computing, quantum measurement, error mitigation, statistical learning**

I. Introduction

Quantum computing is expected to significantly outperform classical computing on many hard problems due to quantum mechanical effects such as entanglement and superposition [1], [2]. While a classical computer can only be in one possible state at a time, a quantum computer can be in an arbitrary combination of states at the same time. Unfortunately, quantum computing also introduces a significant level of noise and uncertainty compared to classical computing [3]–[5]. Moreover, the result of measuring a quantum computer forces the arbitrary state to one known state with some probability, which requires several executions to identify the final output.

There are various sources of noise (errors) in quantum computers including (a) initial state preparation, (b) actual computation, and (c) measurement of results. In this paper, we specifically focus on mitigation of quantum measurement errors. In order to enable noise-resilient quantum computing, it is crucial to mitigate measurement errors that ranges from 5% to 30% in today's machines [6]. Modern quantum computers approach measurement by coupling sensitive equipment (which introduces noise) with statistical techniques that infer properties of the quantum state. This analysis is performed on classical computers using a large amount of quantum data produced by repeatedly measuring quantum circuit output across many iterations (shots) to converge to a correct solution. The statistical model used for mapping a quantum measurement to its inferred quantum state is known as a *qubit discriminator*.

This work was partially supported by the NSF grant CCF-1908131.

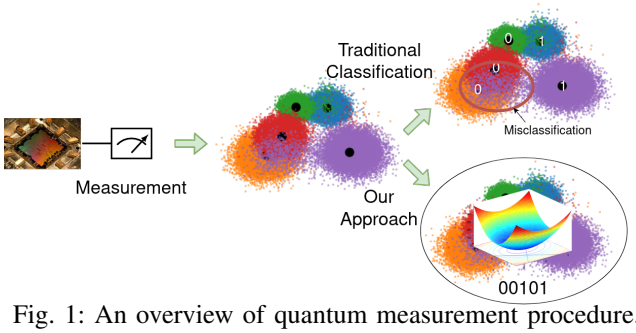


Fig. 1: An overview of quantum measurement procedure.

Figure 1 shows an overview of the measurement procedure in quantum computers. In popular physical realizations of quantum computers, such as superconducting transmons, quantum measurement devices represent a collapsed qubit as a 2-component vector – the in-phase and quadrature components (IQ) of an observed wave transmitted through the resonator [7]. The quantum measurement device returns sets of measurement results, such as IQ points. A classifier is then used to label each point as belonging to either $|0\rangle$ or $|1\rangle$. Our approach uses the measurement data to form a cumulative distribution function, which is compared to that of the training distributions using convex optimization. Such an approach does not exclude hidden statistical properties that may be present in the measurement data.

Quantum measurement error arises due to noisy measurement readings as well as classification errors caused by imperfect discriminators. Hence, the accuracy of the quantum computer is contingent on the performance of the qubit discriminator. We propose an efficient classification technique to improve the measurement accuracy. Specifically, this paper makes the following major contributions.

- We propose a framework for mitigation of quantum measurement errors using cumulative distribution functions to accurately classify quantum measurements.
- Experimental evaluation demonstrates the effectiveness of our model in terms of non-linearity, statistical consistency, and versatility compared to state-of-the-art qubit discriminator approaches.

This paper is organized as follows. Section II surveys related efforts. Section III describes our proposed framework. Section V presents the experimental results. Finally, Section VI concludes the paper.

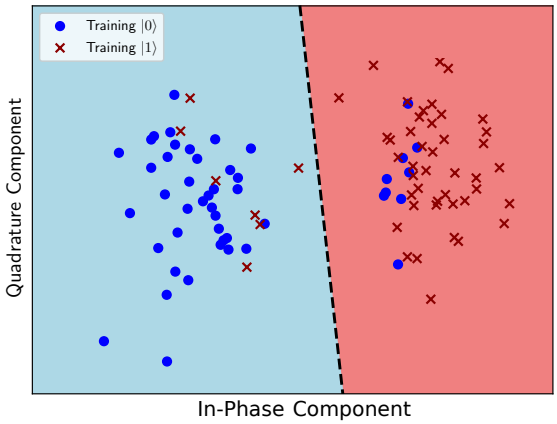


Fig. 2: An overview of linear quantum state discrimination in the IQ space. The model is trained by configuring the quantum computer to output known samples of basis states $|0\rangle$ and $|1\rangle$, which are measured and used to divide the measurement space with a hyperplane. When performing inference, a measurement that is located above the hyperplane will generate a prediction of $|1\rangle$; and a measurement located below the hyperplane will generate a prediction of $|0\rangle$. Training data was sampled from *ibm_quito*, and the hyperplane was computed using *scikit* learn.

II. Background and Related Work

In this section, we describe state-of-the art methods for qubit state discrimination and identify areas for improvement.

A. Related Work

Machine learning techniques are widely used for qubit state discrimination [8]–[19]. Linear Discriminant Analysis (LDA) is one such model commonly used for qubit state discrimination. A hyperplane is selected to partition the IQ vector space into regions of $|0\rangle$ and $|1\rangle$ based on the measured IQ outputs from the training data, which the model assumes follows a Gaussian distribution. Figure 2 visualizes this method, and highlights the difficulty of the classification task. Due to noise in quantum systems (imperfect measurement devices, environmental contamination, and qubit crosstalk), sampled data contains high variance and may appear in the partition; leading to misclassifications.

Other machine learning models such as k-nerset neighbors (kNN) [10], [11], deep neural networks (DNN) [11]–[18], and support vector machines (SVM) [12] have been used with quantum IQ data to partition the measurement space into regions of $|0\rangle$ and $|1\rangle$. Some of these models also consider the effects of quantum "crosstalk" – a phenomena where unwanted interactions among qubits can be predicted and accounted for post-readout. While these methods offer alternative ways to partition; *they each implement the same inference workflow by mapping each qubit measurement to a single location within the partition space.* An overall qubit state is obtained by analyzing the frequencies associated with each prediction class. Beyond variations in the partitioning method, further improvements have been obtained by enabling models to tag samples as "inconclusive" [12], [19], and discarding such samples from processing. Some of these methods can be extended to classify higher energy states [18].

B. Limitations of State-of-the-Art Approaches

The existing quantum measurement classification methods have the following fundamental limitations.

- The existing models operate by partitioning the IQ space into regions corresponding to each basis state. Regardless of the partitioning method used, the measurement space is inherently noisy, leading to classifications shown in Figure 2
- The existing methods map a single IQ measurement tuple into a single quantum state. Since only a single measurement tuple is used for prediction, valuable statistical information encoded within the distribution of test data is neglected.

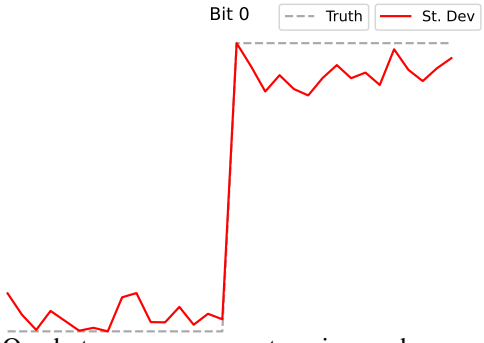


Fig. 3: Quadrature measurement variance shown on IBM Quito for the $|0\rangle$ and $|1\rangle$ basis states.

  For example, using data collected from the IBM Quito quantum computer, the $|1\rangle$ basis state contains higher variance than the $|0\rangle$ state [6], and thus sample variance – a distribution property – contains information about bitstate unused by state-of-the-art methods.
- Existing methods assume properties of quantum data (Gaussian) and partition boundaries (linear, quadratic). Since quantum measurements diverge from ideal distributions, such assumptions may introduce bias.
- Some current methods operate by discarding data deemed "inconclusive". We believe a method that quantifies uncertainty without discarding data can outperform these techniques.
- Most qubit discriminator models are difficult to effectively boost (combine with other models). While ensemble techniques exist, such methods require significantly more computation and have not to date demonstrated superior results for quantum discrimination.
- It is difficult to quantify or guarantee convergence with many state of the art qubit discriminator methods. Quantum circuits are often sampled for tens of thousands of iterations since no stochastic framework exists to bound the error associated with classification.

In practice, many quantum engineers evaluate the accuracy of the discriminator via the use of a test set. Sampling is repeated until the discriminator reaches an accuracy threshold on the test set, requiring potentially thousands of additional quantum samples.

We propose a sophisticated quantum discriminator that overcomes these limitations, guarantees convergence, and as an additional benefit, can produce an estimate for the number of samples needed to attain convergence within a threshold without requiring a holdout (testing) set.

III. Distribution-based Classification for Mitigating Quantum Measurement Errors

The goal of measurement classification is to take the results of measuring qubits in a quantum register (a collection of IQ points) and correctly identify the corresponding bitstring labels. For example, after measuring the quantum state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, the classifier should provide bitstrings "00" and "11", each occurring with equal probability. Traditionally, classifiers are trained to partition the IQ space, as shown in Figure 2. We choose to use linear discriminant analysis [lda] and k-nearest neighbors [kNN] as baselines for this work, due to their prevalence in the community and widespread use in open source libraries, such as Qiskit, and high performance among other discrimination methods [11] on IBM's quantum machines.

Rather than providing yet another approach to partition the IQ space, we propose an entirely novel distribution-based classification workflow that overcomes the shortcomings outlined in Section II. Unlike previous methods which produce classifications for every measurement shot, our method directly estimates the probabilities of $|0\rangle$ and $|1\rangle$ in one task.

In this section, we first outline the use of cumulative distribution functions (CDF) for classification. We then show an example of classification on a single qubit using CDFs. Finally, we incorporate our CDF approach to existing classification techniques, as highlighted in Figure 5 – providing the advantages of CDF while also being scalable even for a large number of qubits.

A. Classification using Cumulative Distribution Functions

It is a well established fact that the Cumulative Distribution Function (CDF) uniquely characterizes a probability distribution. Since a qubit exists in a superposition of states $|0\rangle$ and $|1\rangle$, it follows that the qubit exists in a mixed distribution of basis states $|0\rangle$ and $|1\rangle$. Thus, its unique CDF can be decomposed into a linear (convex) combination of $|0\rangle$ and $|1\rangle$ CDFs, where the weights associated with the constituent $|0\rangle$ and $|1\rangle$ CDFs directly represent the true proportion of measurements that collapsed into each of the $|0\rangle$ and $|1\rangle$ states. This decomposition for an arbitrary qubit superposition is depicted in Figure 4.

Properties like randomness, state stability, and variance are inherently accounted for in the CDF. Additionally, by analyzing the CDF of the qubit as a whole, we avoid the need to classify each shot individually to arrive at a state estimate. Instead, we analyze properties of the qubit distribution, the $|0\rangle$ distribution, and the $|1\rangle$ distribution. Next, we discuss methods for estimating each of these CDFs and the procedure used for the decomposition.

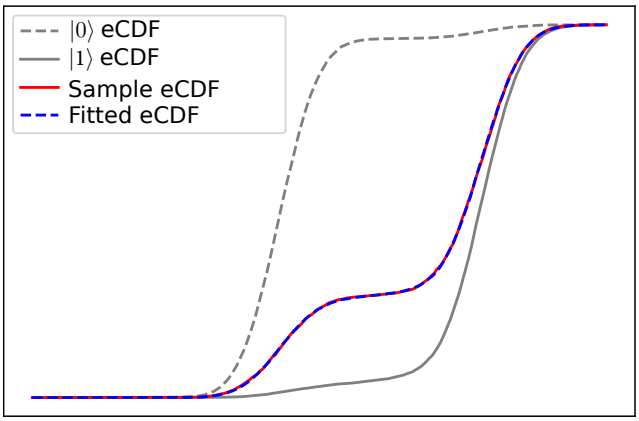


Fig. 4: Visualization of the distribution approach to single qubit state discrimination. The training $|0\rangle$ and $|1\rangle$ eCDFs are shown in gray. The sample qubit eCDF is shown in blue, and the fitted eCDF estimate is shown in red. The fitted eCDF is obtained by creating a convex combination of both gray curves with weights selected to follow the blue curve as closely as possible, as outlined in Section III-B. The coefficients used for the combination are the estimate for the quibt's state. Data obtained from ibm_quito.

B. Empirical CDF based Single Qubit Classification

We begin by preparing a training data set of size k for both $|0\rangle$ and $|1\rangle$ quantum states. Each state is prepared, measured, and tagged with the associated $|0\rangle$ or $|1\rangle$ label. Since each measurement consists of both the in-phase and quadrature components, both the $|0\rangle$ and $|1\rangle$ data sets have dimensions (k by 2). We then construct a test distribution of dimensions (k by 2) by placing the qubit into a random mixed distribution of $|0\rangle$ with frequency α and $|1\rangle$ with frequency $1-\alpha$, where α is randomly selected. We will evaluate our method's ability to reproduce α given the test data, the training data, and the training labels. Next, we study the distribution of both the $|0\rangle$ and $|1\rangle$ training data sets. We estimate the CDF of both sets by computing the empirical CDF (eCDF). The empirical CDF is a consistent and unbiased estimator that converges absolutely to the true CDF. Moreover, as an additional benefit, the Dvoretzky–Kiefer–Wolfowitz inequality [20] provides a closed form error bound for each of the eCDF estimators as a function of the sample size. The eCDF is computed by finding the proportion of values in the data set less than or equal to x, given by

$$\hat{P}(X \leq x) = \frac{1}{k} \sum_{i=1}^{k} I(t_i \leq x) \tag{1}$$

where t_i runs through each element of the set. Here, $I(q)$ is the indicator function, which is given as 1 if q is true, and 0 otherwise. We implement binary search to compute this sum in $\log(k)$ complexity. We then perform a linear interpolation to transform these staircase-like empirical CDFs to smooth estimates — a technique useful for small data sets.

Algorithm 1 eCDF computation with binary Search

input: x: the value to evaluate the eCDF at,
array: the sorted array of quantum measurement data
output: the interpolated eCDF estimate for x

procedure ECDF($x, array$)
  if $x \geq array[len(array) - 1]$ **then**
    return 1
  else if $x \leq array[0]$ **then**
    return 0
  end if
  $upper \leftarrow$ *binarySearch*(x, $array$)
  $lower \leftarrow upper - 1$
  $difference \leftarrow array[upper] - array[lower]$
  return $lower + ((x - array[lower])/difference)$
end procedure

Algorithm 1 describes the procedure for computing the eCDF for x given a sorted array of either in-phase or quadrature quantum measurements. Using this method, we compute individual eCDFs for the in-phase and quadrature component for each basis state. Similarly, we compute eCDFs for the in-phase and quadrature component for the sample.

We use least squares regression to obtain a value for $\hat{\alpha}$, under the constraint $0 \leq \hat{\alpha} \leq 1$, such that we minimize

$$\left\| (\hat{\alpha} \cdot \hat{F}_{|0\rangle,\text{quad}}(x) + (1-\hat{\alpha}) \cdot \hat{F}_{|1\rangle,\text{quad}}) - \hat{F}_{\text{sample,quad}} \right\| + \left\| (\hat{\alpha} \cdot \hat{F}_{|0\rangle,\text{in}}(x) + (1-\hat{\alpha}) \cdot \hat{F}_{|1\rangle,\text{in}}) - \hat{F}_{\text{sample,in}} \right\|$$

where $\hat{F}(x)$ denotes an eCDF estimate, *in* denotes the in-phase dataset component, and *quad* denotes the quadrature component of the quantum measurement dataset. We use $\sqrt{\hat{\alpha}}$ and $\sqrt{1-\hat{\alpha}}$ as the estimates for the qubit's state. Overall, $\hat{\alpha}$ is computed in $k\log(k)$ time complexity.

C. Empirical CDF-based Multi-Qubit Classification

In the above section, we demonstrated how an eCDF single qubit discriminator can be used *instead* of a traditional discriminator. This is possible since the eCDF discriminator fully constrained all two basis states. In this section, we show how the method can be employed *with* existing methods for quantum computers with more than one qubit.

With the single qubit eCDF estimation method, it is possible to effectively decompose the mixed distribution of a single qubit into known distributions of $|0\rangle$ and $|1\rangle$ states using estimation and regression techniques. On a machine of n qubits, this method generates n constraints on the position

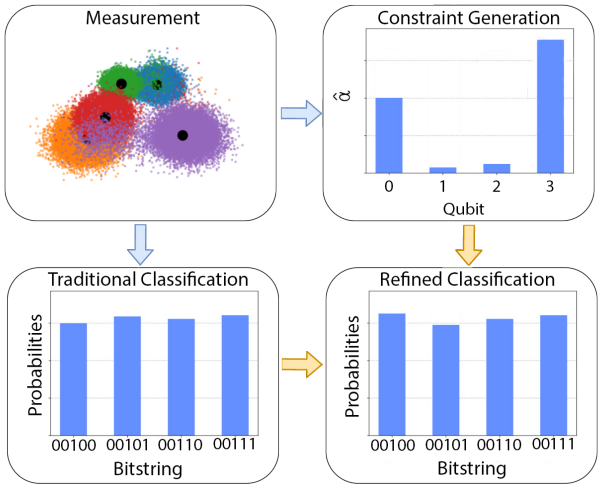


Fig. 5: A high-level overview of the multi-qubit measurement classification procedure that consists of three stages. The first stage (*Traditional Classification*) uses existing methods to produce an estimate. The second stage (*Constraint Generation*) applies the eCDF method to each qubit to generate constraints. The final stage applies the constraints to the existing estimate to produce a new estimate (*Refined Classification*).

space of 2^n basis states. To illustrate this point, suppose we have a quantum computer with $n = 2$ bits, and it is estimated from the above method that qubit 0 decomposes into $|0\rangle$ with frequency α_0 and that qubit 1 decomposes into $|0\rangle$ with frequency α_1. From this, we have the following constraints (X indicates a "don't care" bit that can take any value): (1) The frequencies of states $|X0\rangle$ given by $|00\rangle$ and $|10\rangle$ sum to α_0, (2) the frequencies of states $|0X\rangle$ given by $|00\rangle$ and $|01\rangle$ sum to α_1, and (3) the frequencies of states $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$ sum to 1.

For a computer with n qubits, the solution space contains 2^n unique basis states, $n+1$ constraints, with $2^n - (n+1)$ remaining free variables. For time complexity purposes, a qubit discriminator can not typically constrain all 2^n states. We demonstrate how the constraints can improve performance and enhance state of the art classification methods. To underscore this claim, we employ a linear discriminator to classify quantum measurements and measure classification performance before and after the constraints are applied. Our workflow is highlighted in Figure 5.

We begin by creating a training distribution for each qubit in the $|0\rangle$ and $|1\rangle$ quantum states in the same manner as the single qubit method. We then generate a sample convex label vector $\boldsymbol{\alpha}$ by sampling a random proportion of measurements from each of the 2^n states. For example, if $\boldsymbol{\alpha} = \begin{bmatrix}0.25 & 0.75 & 0 & \dots & 0\end{bmatrix}$, then 25% of our test distribution would be sampled from the 00000 state, and 75% of our test distribution would be sampled from the 00001 state.

Next, we employ the existing state-of-the-art linear dis-

criminant analysis [lda] method to produce an estimate for the qubit's state, given as $\hat{\boldsymbol{\alpha}}_1$. This is done by first training n traditional models on each qubit's training set. In the traditional manner, each IQ pair in the test set is classified independently as $|0\rangle$ or $|1\rangle$, producing an estimate bitstring. This is repeated for every qubit string in the test set and the frequencies are computed to generate $\hat{\boldsymbol{\alpha}}_1$.

Finally, we generate the constraints and produce a refined estimate $\hat{\boldsymbol{\alpha}}_2$ that adheres to each constraint. As described in the previous section, we create eCDF estimates for each qubit's $|0\rangle$, $|1\rangle$, and test distributions. We compile these constraints into $\boldsymbol{\beta}$, a vector of length n which, for each qubit, independently estimates the proportion of that qubit's test distribution measured in the $|0\rangle$ state.

Due to the presence of free variables, there are many possible candidates which adhere to all $\boldsymbol{\beta}$ constraints. Rather than considering all of them, we define the refined estimate $\hat{\boldsymbol{\alpha}}_2$ as the distribution closest to $\hat{\boldsymbol{\alpha}}_1$ that adheres to all β constraints. In other words, we update the estimate $\hat{\boldsymbol{\alpha}}_1$ to satisfy the marginal probabilities given by the constraints $\boldsymbol{\beta}$ while minimizing $\|\hat{\boldsymbol{\alpha}}_1 - \hat{\boldsymbol{\alpha}}_2\|$.

D. Empirical CDF based Single Qudit Classification

In this section, we show how the eCDF classification method can be extended for quantum systems with more than 2 basis states. Some existing classification methods, like linear discriminant analysis, support vector machines, and binary logistic regression do not naturally extend beyond 2 prediction categories. While strategies like "one-vs-rest" can be used to train binary models for each category; such methods are associated with higher training costs, lower performance, and greater susceptibility to class imbalance.

Rather than training multiple models (as done in "one-vs-rest"), we take the approach of designing a multi-class solution for qubit state discrimination. Let us suppose our qudit has c basis states. Then, our convex optimization operation can be expressed as:

$$\left\| \sum_{i=1}^{c-1} (\hat{\gamma}_i \cdot \hat{F}_{|i\rangle,\text{quad}}(x)) + (1 - \sum_{i=1}^{c-1} (\hat{\gamma}_i) \cdot \hat{F}_{|c\rangle,\text{quad}}) - \hat{F}_{\text{sample,quad}} \right\| + \left\| \sum_{i=1}^{c-1} (\hat{\gamma}_i \cdot \hat{F}_{|i\rangle,\text{in}}(x)) + (1 - \sum_{i=1}^{c-1} (\hat{\gamma}_i) \cdot \hat{F}_{|c\rangle,\text{in}}) - \hat{F}_{\text{sample,in}} \right\|$$

IV. Error Bounds for eCDF based State Discrimination

Even after a machine learning model has been trained, the reliable performance of traditional models has proven challenging. Typically, model performance is measured by computing the accuracy of the model on a set of predictors; where the prediction of the model is compared against known labels. We identify the following issues with this approach

- Machine learning models are susceptible to over-fitting. Indeed, when a model is trained on a particular dataset, it is quite often the case that the model identified nonessential features of the training set (such as random noise) which will produce optimistic and inaccurate estimations of the true performance of the model on data outside of the training set.
- Reliable accuracy measurements rely on holdout sets. Often, a large percentage of the data is not used for training purposes and is "set aside" for evaluation purposes. While this solves the optimism due to over-fitting issue, data collection is expensive; and thus, a considerable cost is spent on data which will not directly improve the performance of the model.
- Holdout sets can be misleading or biased. The use of a holdout set for model evaluation purposes relies on the assumption that the holdout set is fully representative of the population. In reality, the act of partitioning data into "training" and "holdout" sets may introduce bias and may present misguided estimates of the true performance of the model if the data selected for the holdout set skew towards different features or categories.

For these reasons, it is of tremendous value that our method can provide an accurate performance estimate without the use of a holdout set.

A. Error Bounds

Our goal in this section is to produce a confidence interval for α such that a given a confidence-level p and a number of shots k, we compute a sufficient upper and lower estimate for α.

Lemma: Let $1-p$ be a confidence level, and let $\{\hat{F}_n(x)\}$ be a sequence of eCDFs. Then, $\forall x, 0 \leq i \leq n$

$$P(\hat{F}_i(x) - \epsilon_i \leq F_i(x) \leq \hat{F}_i(x) + \epsilon_i) \geq 1 - p$$

where

$$\epsilon_i = \sqrt{\frac{\ln \frac{2}{n \cdot p}}{2 \cdot k_i}}$$

Proof: This is derived from previous work by Dvoretzky, Kiefer, Wolfowitz, and Massart (later enhanced by Kolmogorov and Smirnov) which proved for a *single* distribution with a true cCDF $F(x)$, and an eCDF $\hat{F}(x)$, a $1-p$ confidence interval is given by

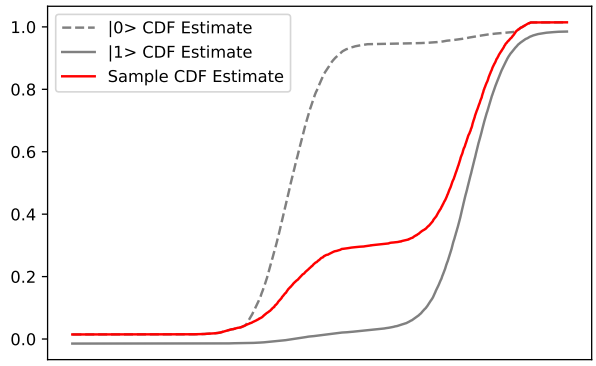
$$P(\hat{F}(x) - \epsilon \leq F(x) \leq \hat{F}(x) + \epsilon) \geq 1 - p$$

where $\epsilon = \sqrt{\frac{\ln \frac{2}{p}}{2 \cdot n}}$.

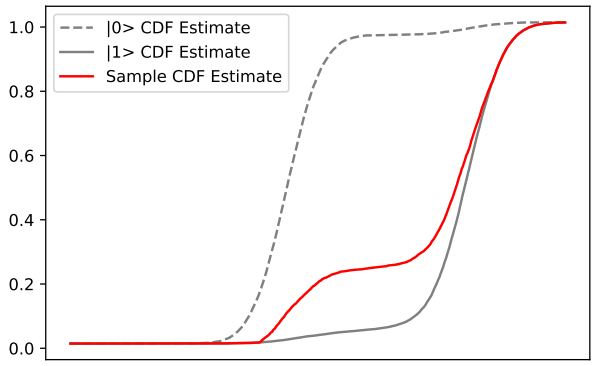
In our method, we are separately computing n empirical CDFs. We can apply the Bonferroni adjustment method

$$p_{\text{family}} = \frac{p}{n}$$

to create a *family*-wide confidence interval. This allows us to extend our expression for a single confidence interval to a sequence of confidence intervals in which we are $1-p$ confident that *all* confidence intervals in the sequence are correct. ■



(a) Curve selection so α is minimized



(b) Curve selection so α is maximized

Fig. 6: Selection of curves from the CDF space to maximize and minimize α. In this example, we have a 99% confidence interval for α given as $(0.186, 0.316)$ with a point estimate $\hat{\alpha} = 0.249$.

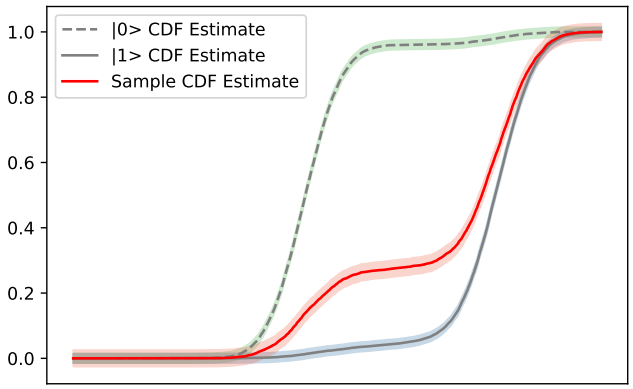


Fig. 7: Visualization of the 99% confidence interval for the family of eCDFs. If many samples were taken, and all 3 eCDF confidence intervals were generated for each sample, we would find for at least 99% of samples, all 3 eCDF intervals correctly contain the true CDF.

B. Confidence Interval Construction

In light of these three error bounds, we view each bound as a space of CDFs, where all three spaces contain their respective true CDF with $1 - p_{\text{total}}$ confidence, as shown in Figure 7. Since we view the sample CDF as a sum of CDFs from both the $|0\rangle$ and $|1\rangle$ weighted by α, we can attain bounds on α by exploring the maximum and minimum values such a weight can take while staying within the bounds.

From its construction, to attain the lowest possible value for α, we would select the $|0\rangle$ curve from the family of $|0\rangle$ curves to be as similar (from a least squares perspective) as possible to the sample curve. We would likewise select the $|1\rangle$ curve to be as far from the sample curve as possible.

We will make a simplifying assumption which is consistent with the physical design of the quantum computer and the data observed. We will assume that the true $|0\rangle$ CDF is greater than the true $|1\rangle$ CDF for all x.

Since the sample CDF is a weighted convex sum of the true $|0\rangle$ and $|1\rangle$ CDF, it lies between between both curves. As shown in Figure 6, the maximum α occurs when $\hat{F}_{|0\rangle}(x)$ is maximized; and $\hat{F}_{|1\rangle}(x)$ is minimized while staying as close as possible to $\hat{F}_{\text{sample}}(x)$.

V. Experiments

This section demonstrates the effectiveness of our proposed quantum measurement methods compared to the state-of-the-art approaches. We first outline our experimental setup for single qubit architectures. Next, we demonstrate our method's effectiveness on higher dimensional qudit architectures. Lastly, we show how our method can be implemented on machines with multiple qubits.

A. Experimental Setup

We use ibm_quito, a 5-qubit machine, to initialize states and perform measurements. Quantum circuits and measurements are performed to output $|0\rangle$ and $|1\rangle$ basis states for each qubit. Each measurement is performed 20,000 times (shots), thereby obtaining 20,000 samples of IQ measurements in each of the $|0\rangle$ and $|1\rangle$ quantum states. Data was partitioned into a training and testing set. We evaluate the effectiveness of our proposed methods compared to each state-of-the-art approach implemented in Qiskit [21]. We use Scipy's optimizations library to perform all necessary minimization using the "Nelder - Mead" method.

B. Classification Results for Single Qubits

For a single test, we first shuffle the entirety of the experiment dataset. The set is then partitioned into training and testing data. We then generate 1,000 random values for α, each of which lies between 0 and 1. For each value of α, we then build a mixed testing dataset of size 5,000 composed of α% randomly selected values from the testing data of $|0\rangle$ and $(1 - \alpha)$% randomly selected values from the testing data of $|1\rangle$. The model is then evaluated on how well it can reconstruct the value of α. We evaluate the mean absolute error (MAE) as the absolute difference between the measurement and truth, given as $|\hat{\alpha} - \alpha|$.

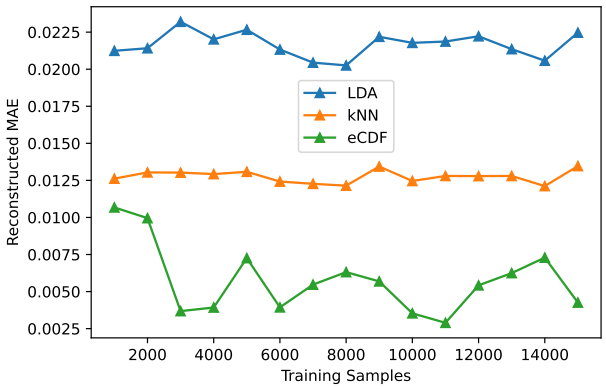


Fig. 8: Mean Absolute Error (MAE) for single qubit reconstruction using two traditional models (k-nearest-neighbhors [knn] and linear discriminant analysis [lda]) and our proposed eCDF discriminator method. The data was shuffled across 120 iterations; and for each iteration, 1000 random distributions were generated, reconstructed, and evaluated as described above.

Figure 8 shows the resulting Mean Absolute Error (MAE) of using 2 traditional discriminator versus our eCDF-based method. The eCDF model attained a lower error at all training sizes and greatly diminished variance as the sample size enlarged. This demonstrates that our proposed approach (eCDF) outperforms traditional discriminators.

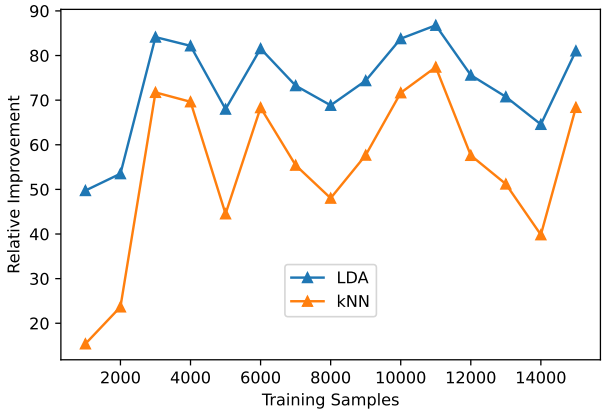


Fig. 9: Percent improvement for single qubit Mean Absolute Error (MAE) of proposed eCDF method compared with two traditional models (k-nearest-neighbors [knn] and linear discriminant analysis [lda]) and our proposed eCDF discriminator method.

Figure 9 shows the relationship between number of samples and the MAE provided by our proposed approach (CDF) as well as existing approach (linear discriminant analysis, k-nearest neighbors). It highlights two important points: (1) increasing training samples improves the performance at a disproportionately higher rate than the baseline, and (2) our proposed solution significantly outperforms (up to 88%) state-of-the-art, and attains significant improvements above the baseline at each training size. Figure 10 shows that our approach attains much lower variance than its machine learning based counterparts.

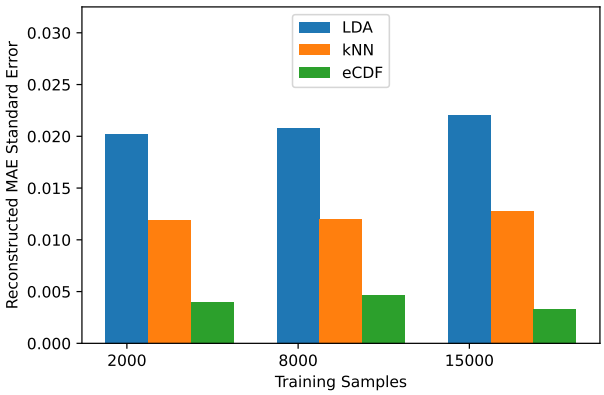


Fig. 10: Standard Error in single qubit Mean Absolute Error (MAE) of proposed eCDF method compared with two traditional models (k-nearest-neighbors [knn] and linear discriminant analysis [lda]) and our proposed eCDF discriminator method.

C. Classification Results for Single Qutrits

Using the same IBM Quito computer, we constructed a training and test set for a single qutrit, with basis states $|0\rangle, |1\rangle$, and $|2\rangle$. We compare our eCDF qutrit state discriminator with a k-nearest neighbors (kNN) discriminator.

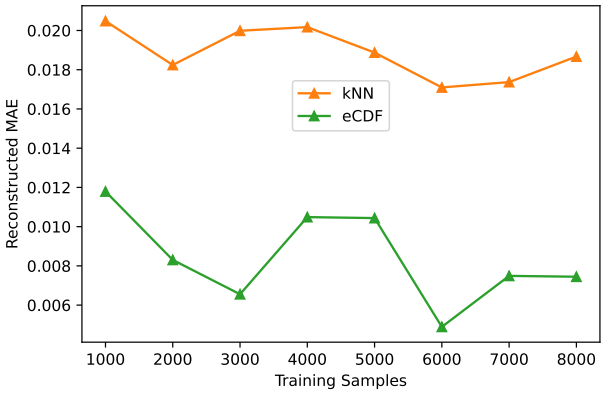


Fig. 11: Mean Absolute Error (MAE) for single qutrit reconstruction k-nearest-neighbors [knn] and and our proposed eCDF discriminator method. The data was shuffled across 60 iterations; and for each iteration, 1000 random distributions were generated, reconstructed, and evaluated as described above.

We see that the eCDF discriminator attains lower error at all training samples. Moreover, we see the eCDF discriminator attains lower variance, and thus performs more consistently than its machine learning counterpart on qutrit measurement data.

D. Classification Results for Multiple Qubits

We begin by configuring the quantum computer to output each of 2^5 basis states for the 5 qubit machine. Similarly, each measurement is performed $k = 20,000$ times. It should be noted, however, that for the purposes of evaluating the

method, we consider all 2^n states as candidates for output of the quantum computer. In practice, our method does not require enumeration of all 2^n states; and only considers a maximum of $\min(2^n, k)$ states.

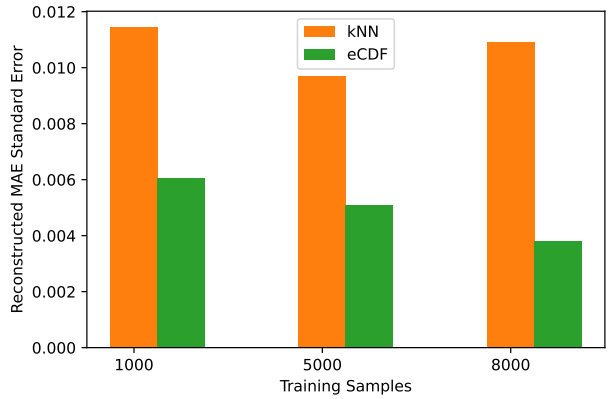


Fig. 12: Standard Error for single qutrit Mean Absolute Error (MAE) of proposed eCDF method compared an existing model (k-nearest-neighbors [knn]) and our proposed eCDF discriminator method.

For a single test, we shuffle the dataset and partition the experimental data into training and testing datasets. We generate a test vector by producing a random convex vector $\boldsymbol{\alpha}$ of size 2^n. We construct a test dataset of size 5,000 by randomly sampling $\alpha_i\%$ values from the i^{th} basis state.

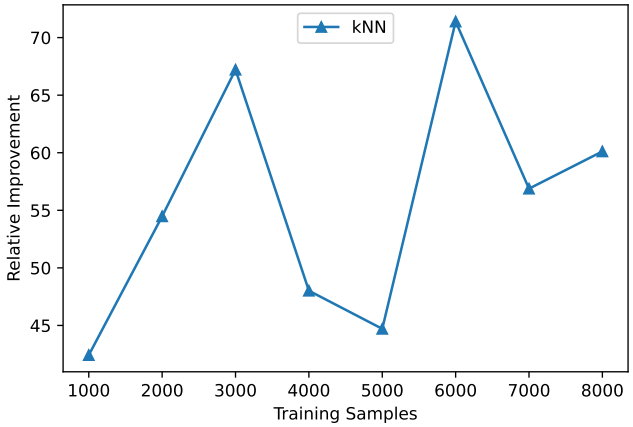


Fig. 13: Percent improvement in Mean Absolute Error (MAE) of proposed eCDF method compared with k-nearest-neighbors [knn] using our proposed eCDF discriminator method on single qutrit data.

The pipeline is evaluated on how well it can reconstruct the value of α. We compute the mean absolute error as the value $\frac{\|\boldsymbol{\alpha}-\hat{\boldsymbol{\alpha}}_2\|^2}{2^n}$, which we compare to the baseline mean absolute error given by $\frac{\|\boldsymbol{\alpha}-\hat{\boldsymbol{\alpha}}_1\|^2}{2^n}$.

As shown in Figure 14, the qubit discriminator pipeline with our proposed eCDF model outperforms the traditional discriminator at all training sizes. Therefore, it is beneficial to combine eCDF with traditional models.

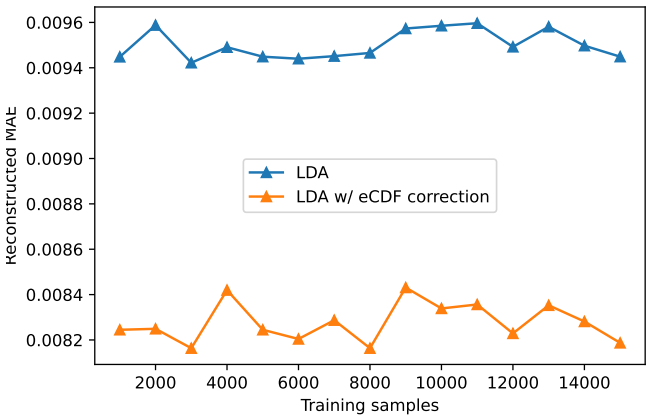


Fig. 14: Mean Absolute Error (MAE) for reconstruction using linear discriminant analysis [lda] and the proposed estimation pipeline with eCDF correction. The data was shuffled across; and for each iteration 100 random distributions were generated, reconstructed, and evaluated as described above.

VI. Conclusion

Quantum measurement classification is fundamental to a successful execution of any quantum algorithm. Measurement classification includes several nuances, such as inherent physical error, as well as randomness associated with measured data. In this work, we have introduced a new qubit classifier model that is able to outperform current state-of-the-art machine learning discriminators. The model's performance improvement is achieved through its novel statistical distribution viewpoint, which enables the model to capture important features while diminishing effects of noise and bias associated with individual measurement. Specifically, our proposed eCDF technique significantly outperforms (up to 88% for single qubits) state-of-the-art in single qubit classification accuracy, with similar improvements in higher dimensional architectures (up to 70% for single qutrits. We showed that this method offers a fundamental improvement (up to 12%) to state of the art multi-qubit classification methods by building a qubit discriminator pipeline that first performs any standard qubit discriminator method, followed by an eCDF qubit correction stage. In addition to superior performance, our eCDF method can generate confidence intervals for qubit state without requiring a holdout set.

As demand for quantum computing increases, techniques that can attain convergence with fewer measurements enable quantum providers to trade off valuable quantum computer resources with processing performed on classical computers. This work opens a path to building robust, yet simple, measurement classifiers based on fundamental statistical principles. It invites quantum engineers to engage with quantum data at the distribution level and provides a framework to add independent qubit distribution insights into existing quantum classification workflows.

VII. Acknowledgements

I would like to thank my faculty mentor, Dr. Prabhat Mishra, for working with me on this project. I would also like to recognize the contributions of Daniel Volya, a PhD student, in generating the quantum data for this work.

References

[1] A. Zulehner and R. Wille, *Introducing Design Automation for Quantum Computing*. Springer, 2020.

[2] D. Volya and P. Mishra, "Quantum spectral clustering of mixed graphs," in *ACM/IEEE Design Automation Conference (DAC)*, 2021, pp. 463–468.

[3] ——, "Impact of noise on quantum algorithms in noisy intermediate-scale quantum systems," in *ICCD*, 2020.

[4] ——, "Modeling of noisy quantum circuits using random matrix theory," in *ICCD*, 2022, pp. 132–138.

[5] ——, "Quantum data compression for efficient generation of control pulses," in *Asia and South Pacific Design Automation Conference (ASPDAC)*, 2023.

[6] S. Tannu and M. Qureshi, "Mitigating measurement errors in quantum computers by exploiting state-dependent bias," in *MICRO*, 2019.

[7] R. Bianchetti *et al.*, "Control and tomography of a three level superconducting artificial atom," *Physical Review Letters*, 2010.

[8] L. A. Martinez, Y. J. Rosen, and J. L. DuBois, "Improving qubit readout with hidden markov models," *Phys. Rev. A*, vol. 102, p. 062426, Dec 2020. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.102.062426

[9] U. Azad and H. Zhang, "Machine learning based discrimination for excited state promoted readout," in *2022 IEEE/ACM 7th Symposium on Edge Computing (SEC)*, 2022, pp. 362–367.

[10] D. Quiroga, P. Date, and R. C. Pooser, "Discriminating Quantum States with Quantum Machine Learning," *ICRC*, pp. 56–63, 2021.

[11] U. Azad and H. Zhang, "Machine learning based discrimination for excited state promoted readout," Tech. Rep., 02 2022.

[12] B. Lienhard *et al.*, "Deep-neural-network discrimination of multiplexed superconducting-qubit states," *Phys. Rev. Applied*, 2022.

[13] P. Duan *et al.*, "Mitigating crosstalk-induced qubit readout error with shallow-neural-network discrimination," *Phys. Rev. Applied*, 2021.

[14] J. Kim, B. Oh, Y. Chong, E. Hwang, and D. K. Park, "Quantum readout error mitigation via deep learning," *New Journal of Physics*, vol. 24, no. 7, p. 073009, jul 2022.

[15] É . Genois *et al.*, "Quantum-tailored machine-learning characterization of a superconducting qubit," *PRX Quantum*, vol. 2, no. 4, 2021.

[16] P. Duan *et al.*, "Mitigating crosstalk-induced qubit readout error with shallow-neural-network discrimination," *Physical Review Applied*, vol. 16, no. 2, aug 2021.

[17] R. Navarathna *et al.*, "Neural networks for on-the-fly single-shot state classification," *Applied Physics Letters*, 2021.

[18] E. Magesan, J. M. Gambetta, A. Córcoles, and J. M. Chow, "Machine learning for discriminating quantum measurement trajectories and improving readout," *Physical Review Letters*, vol. 114, no. 20, 2015.

[19] D. Gavinsky *et al.*, "Bounded-error quantum state identification and exponential separations in communication complexity," in *ACM Symposium on Theory of Computing*, 2006, p. 594–603.

[20] A. Dvoretzky *et al.*, "Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator," *The Annals of Mathematical Statistics, 27 (3)*, pp. 642 – 669, 1956.

[21] H. Abraham *et al.*, "Qiskit: An open-source framework for quantum computing," 2019.