PHYSICALLY UNCLONABLE CHARACTERISTICS FOR VERIFICATION OF TRANSMON-BASED QUANTUM COMPUTERS

DISSERTATION

Leleia A. Hsia, Captain, USAF

AFIT-ENG-DS-21-S-007

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENG-DS-21-S-007

PHYSICALLY UNCLONABLE CHARACTERISTICS FOR VERIFICATION OF TRANSMON-BASED QUANTUM COMPUTERS

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy in Electrical Engineering

Leleia A. Hsia, B.S.E.E., M.S.E.E.

Captain, USAF

September 16, 2021

AFIT-ENG-DS-21-S-007

PHYSICALLY UNCLONABLE CHARACTERISTICS FOR VERIFICATION OF TRANSMON-BASED QUANTUM COMPUTERS

DISSERTATION

Leleia A. Hsia, B.S.E.E., M.S.E.E.
Captain, USAF

Committee Membership:

Laurence D. Merkle, Ph.D
Chair

Mary Y. Lanzerotti, Ph.D
Member

David E. Weeks, Ph.D
Member

AFIT-ENG-DS-21-S-007

Abstract

Future national security can be strengthened by verifying and securing the quantum computing supply chain. This dissertation proposes physically unclonable characteristics (PUCs), a method of quantum hardware verification inspired by classical physically unclonable functions, for future application to quantum processors implemented with transmon qubits. Qualitative and quantitative analysis is provided on the development of PUCs, including identifying qubit characteristics and qubit discrimination methods suitable for PUCs. Characteristics tested on IBM Quantum services include T1 and T2 coherence times, single-qubit and multi-qubit gate error rates, readout error rates, quantum process tomography metrics, and random benchmarking metrics. Results show that non-parametric qubit discrimination methods are best-suited for the characteristics tested, but require refinement before real-world implementation can be achieved.

Acknowledgements

Sincere appreciation is given first to God for providing me with the opportunity to attend the Air Force Institute of Technology, a world-class institution, and for sustaining me through the PhD program. I also would like to express my gratitude to Dr. Laurence D. Merkle for serving as my advisor and committee chair, and to Dr. Mary Y. Lanzerotti and Dr. David E. Weeks for serving as committee members on this research. I am grateful for their valuable time, expertise, and feedback. Finally, I would like to give a multitude of thanks to the leadership at AFIT and Eglin AFB, my co-authors and colleagues, and my family and friends for their support and encouragement.

This research was partially supported by the Information Directorate of the Air Force Research Laboratory (GFA8650-20-P-1707) and the AFRL IBM Q-Hub. I acknowledge the use of IBM Quantum services for this work. The views expressed are those of the author, and do not reflect the official policy or position of IBM or the IBM Quantum team.

Leleia A. Hsia

Table of Contents

Page

Page

List of Figures

List of Tables

Table Page

PHYSICALLY UNCLONABLE CHARACTERISTICS
FOR VERIFICATION OF TRANSMON-BASED QUANTUM COMPUTERS

I. Introduction

1.1 Motivation

Advanced computing capabilities that are being enabled by quantum computing will facilitate significant technological improvements across multiple industries, including machine learning for decision support and training, simulation for pharmaceuticals and materials, and optimization for logistics and financial services [82]. Most notably, quantum computing advancements will provide exciting possibilities for the Department of Defense (DoD) to provide additional capabilities for hardware verification in support of national security. In fact, the 2018 National Defense Strategy identifies advanced computing as a technology "that will ensure we will be able to fight and win the wars of the future" [116]. As a result, the DoD is investing in the advancement of quantum computing technologies, as indicated by R. Cohen's FY-20 Program Objective Memorandum [60].

Quantum computing, which has been described as a "disruptive technology" by Dr. Michael Hayduk[1], Deputy Director, Air Force Research Laboratory Information Directorate, would enable improvement of many warfighting applications, including secure communications and GPS-denied navigation [42, 73]. However, if an adversary can gain knowledge from unintended information leakage in quantum computing systems, sensitive communications may no longer be secure. Likewise, if an adversary

[1]Dr. Hayduk was a Division Chief within the Air Force Research Laboratory at the time of his comment.

can change the design or fabrication of quantum computing hardware implemented in navigation systems, secure missions may become compromised or thwarted through adversarial acquisition of asset location, inaccurate navigation calculations, or spoofing of location data.

It has been established that classical computing systems need to be hardware verified and secured as part of ensuring the systems' trustworthiness [40]. Counterfeit and tampered hardware can lead to unauthorized access to high-value corporate secrets and sensitive government networks. Such unauthorized access was recently observed in 2018 when it was reported that unauthorized Chinese microchips had affected server motherboards utilized by almost 30 companies, including a major bank, government contractors, Apple, and Amazon [138].

This work seeks to apply the concept of hardware verification and security to quantum computing systems as part of ensuring quantum computing systems' trustworthiness, especially in the current climate of supply chain vulnerability that jeopardizes technology leadership [130]. Early consideration as quantum systems are being created and developed will enable quantum hardware verification and security development to keep pace with quantum system development and to avoid pitfalls experienced with classical computing system development [125].

In this work, the phrase "hardware verification" refers to the process by which one ensures that a fabricated computing system is not counterfeit and is free of malicious insertions or modifications [98]. Although the phrase "hardware security" often encompasses the concept of "hardware verification" [66], in this work, the definition of "hardware security" is limited to the concept of ensuring a system (that is either assumed or verified to be free of malicious insertions or modifications) does not provide any unintended information to the adversary post-fabrication.

The main contribution of this work to the literature is the proposal of new ap-

proaches for the verification of transmon quantum computers and the initial development of physically unclonable characteristics (PUCs), the application of physically unclonable functions (PUFs) to transmon quantum computers. Results of this work show greatest potential for the one-dimensional Kolmogorov-Smirnov (KS) test, multi-dimensional KS test, and Jensen-Shannon divergence (JSD) test (with binning algorithms that do not assume normally distributed data) to function in PUCs for a variety of qubit characteristics.

1.2 Problem Description

It has been established that, like classical computers, quantum computers need to be hardware verified and secured [125]. Effective quantum hardware verification and security methods do not currently exist to my knowledge[2] [53], and the process of developing these methods is expansive in breadth. Many different types of quantum computer implementations are in development, which require different types of verification and security techniques to suit each implementation.

One proposed first step towards developing quantum hardware verification and security techniques is to investigate the techniques that exist for classical components and determine to what extent they apply to quantum components. It is possible that some techniques are directly applicable; some will need modification; and others will be completely inapplicable. Additionally, quantum computers may likely introduce uniquely quantum vulnerabilities, and entirely new techniques will need to be developed.

While the scope of the problem described above is appropriate for an entire community of researchers over many years, the main premise of this work is to begin the

[2]Chong has suggested some ideas for functional verification through simulation, but concludes that this type of verification is severely limited by the number of qubits, and there remains "a strong need for a methodology for verifying the correctness of both quantum software and hardware" [53].

process of applying classical hardware verification and security methods to quantum computers. The first step is to provide a framework for evaluating the extent to which a given classical method applies to a quantum computer with a given qubit implementation. Preliminary evaluations of potential applications and limitations are conducted within this framework for a subset of classical methods and for the silicon spin and transmon qubit implementations (Part A). Next, PUCs, the application of PUFs to transmon quantum computers, are developed (Part B).

The problem description for PUCs begins with the acknowledgment that PUFs do not directly apply to the transmon processor (more details in Chapter II), but they do follow a cause-and-effect paradigm, where the causes are fabrication variations, and the effect is inter-chip variation in performance parameters. The object of verification for PUFs is a function of performance characteristics (that is, bit value output from an arbiter is a function of path delay). Similarly, the causes for PUCs are fabrication variations and noise, and the effects are inter-qubit and intra-qubit performance variations. The objects of verification for PUCs are the performance characteristics themselves.

	PUFs	**PUCs**
Cause-and-Effect Paradigm	Cause = fabrication variations Effect = inter-chip variation in performance parameters	Cause = fabrication variations and noise Effect = inter-qubit and intra-qubit performance variations
Object of Identification	Function of performance characteristics (that is, bit value output from an arbiter is a function of path delay)	Performance characteristics

Table 1: Cause-and-effect Paradigm for PUFs and PUCs (reproduced from [94])

The role of PUFs is to contribute to supply chain security by serving as a technique to track and verify classical components [156]. An objective for PUCs is to serve in

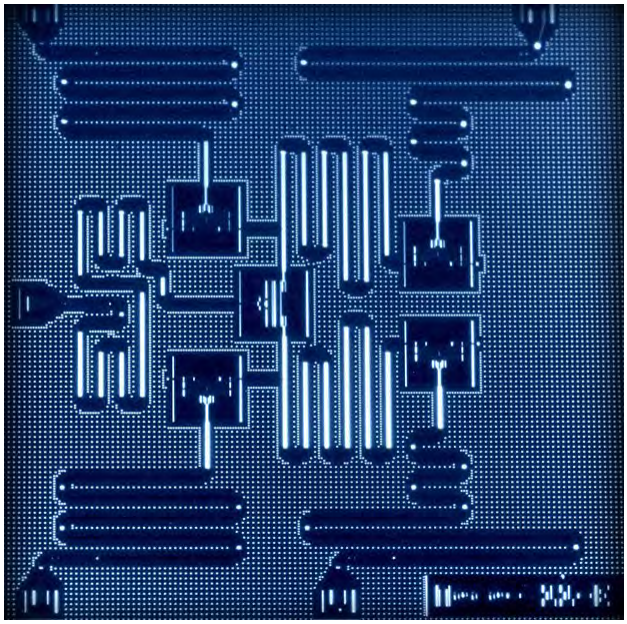


Figure 1: IBM superconducting quantum processor (reproduced from [63]). The image features a solid-state chip, where the five "squares" are the transmon qubits and the "squiggles" are waveguide resonators.

the same role for the quantum computing domain. Consider the situation shown in Fig. 2 in which a manufacturer fabricates a transmon processor, collects N_1 sets of characteristic data for each qubit (one set each day from Day 1 through Day N_1), and sends the processor and characteristic data to a customer at the next stage of the supply chain through separate channels. After the processor is delivered to the customer, the customer collects N_2 sets of characteristic data (one set each day from Day $N_1 + 1$ through Day $N_1 + N_2$). The customer then compares their data to the manufacturer's data using PUCs to ensure that a third party did not change or replace the processor.

Once PUCs develop to the point of real-world implementation, other use cases for PUCs can be considered. For example, once PUCs have advanced enough to account for the minor physical changes that occur during chip maintenance sessions, PUCs can function to verify the chip throughout the lifetime of the chip rather than just the initial deployment of a chip. For example, if a company owning a transmon-

based quantum computer were to hire a third party to perform maintenance on the computer, the owner could use PUCs to verify that the third party did not tamper with or replace the processor during the maintenance session.

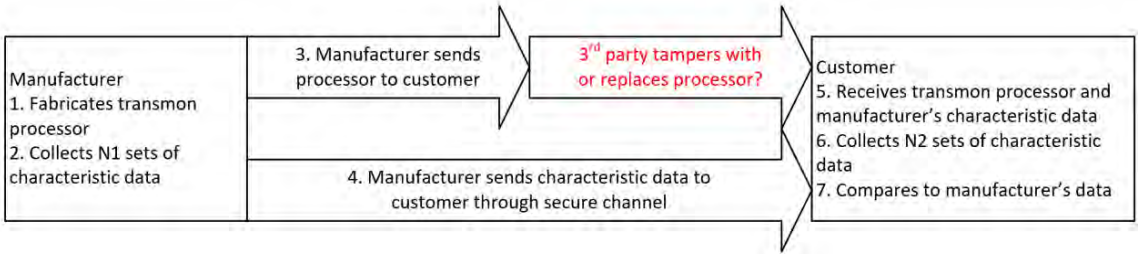


Figure 2: Concept diagram of use case scenario for PUCs proposed in this research

1.3 Research Objectives

The research is organized into Parts A and B. The research objective of Part A is to determine potential applications of current hardware verification and security techniques to emerging quantum computing technologies. This process requires a survey of the state-of-the-art in both areas, and the establishment of a framework for evaluating the relationships.

The research objective of Part B is to begin the process of developing and evaluating PUCs for transmon-based processors. The process of developing PUCs includes selecting and testing different characteristics (or combinations of characteristics) and selecting and testing qubit discrimination methods. PUCs are evaluated using the metrics of probability of detection (PD) and probability of false alarm (PFA).

1.3.1 Literature Review Questions and Research Questions

As few publications exist on the topic of hardware verification or security for quantum computing technologies (aside from those by Arapinis et al. [46], Chong [53], and Naveh et al. [125]), the central question for this research is as follows: How can

quantum computing technologies be hardware verified and secured? All literature review questions and research questions for Part A and Part B contribute to exploring this central question.

1.3.1.1 Part A – Determine Potential Applications

The literature review questions and research questions for Part A are intended a) to prompt a survey of the fields of quantum computing technologies and classical hardware verification/security techniques and b) to prompt consideration of the applications and limitations of relationships between the two fields. The questions are as follows:

- Literature Review Question A-1: What are current state-of-the-art hardware verification and security methods?
- Literature Review Question A-2: What are the current state-of-the-art quantum computing implementations?
- Research Question A-1: How might the techniques identified in Literature Review Question A-1 be applied to verification and security of the quantum computing systems identified in Literature Review Question A-2?

The scope of this document is limited to responding to the literature review and research questions above. However, the continued pursuit of developing quantum hardware verification and security requires additional research questions. These additional questions are:

- What vulnerabilities are inherent in quantum computing systems? Are any of these vulnerabilities unique? For example, are there vulnerabilities related to quantum noise [158] or the Aharonov-Bohm effect [144]?

- What new quantum hardware verification and security techniques can prevent the exploitation of these vulnerabilities?

1.3.1.2 Part B – Develop PUCs

The research questions for Part B are intended to guide the development of physically unclonable characteristics (PUCs). The questions are as follows:

- Literature Review Question B-1: What characteristics are used to describe individual qubits for transmon processors? What characteristics are used to describe sets of qubits for transmon processors?
- Literature Review Question B-2: What techniques exist to perform qubit discrimination?
- Research Question B-1: How can the underlying concept of the PUF be applied to a transmon-based quantum processor?
- Research Question B-2: Which characteristics identified in Literature Review Question B-1 are best suited for PUCs?
- Research Question B-3: Which techniques identified in Literature Review Question B-2 are most effective as qubit discrimination methods for PUCs?
- Research Question B-4: What obstacles exist in the development of PUCs and how can they be overcome?

The scope of this document includes responses to each of the literature review and research questions for Part B. However, the responses are not fully comprehensive, and additional answers to each of the questions require exploration beyond the scope of this document.

1.4 Contributions

The contributions of the research in this dissertation are listed below:

1. A foundation[3] for exploring the relationships (applications and limitations) between classical hardware verification and security techniques and quantum computing systems with differing qubit implementations (Part A).

2. A proposal to include radio-frequency (RF) distinct native attributes (DNA) (RF-DNA) [111] for future applications in transmon-based processors (Part A).

3. A proposal of physically unclonable characteristics (PUCs) for future application to transmon-based processors (Part B).

4. An evaluation of different qubit characteristics, discrimination methods, and data conditioning methods for application to PUCs (Part B).

1.5 Scope

As previously stated, the scope of the quantum hardware verification and security problem is appropriate for an entire community of researchers over many years. This section summarizes the scope (including the assumptions and limitations) of this research as a contribution to the community effort to verify and secure quantum hardware.

The scope of Part A is restricted to a high-level comprehensive (but not exhaustive) survey of emerging quantum computing implementations, classical hardware verification methods, and classical security methods. It additionally includes a preliminary analysis of the relationships between a subset of quantum computing implementations and a subset of classical hardware verification and security methods.

[3]This foundation is provided in the form of a list of metrics used in evaluation tables. Please refer to Sections 3.2.3 and 4.2.3 for details.

An in-depth analysis of the relationships is beyond the scope of this work, except for the relationships between PUFs, RF-DNA, and the transmon qubit. Furthermore, research into vulnerabilities introduced by quantum computing systems and the development of new quantum techniques to prevent the exploitation of these vulnerabilities fall outside the scope of this work.

The scope of Part B is limited to evaluating the suitability of a subset of characteristics and a subset of qubit discrimination methods for application to PUCs. Due to the unavailability of data provided by a manufacturer or customer as described in the use case, the scope is additionally limited by using data collected through the IBM Quantum Experience [41] to simulate a manufacturer's and customer's data. Furthermore, the following assumptions related to the use case apply to the analysis for Part B:

1. The manufacturer is trusted.

2. The manufacturer's characteristic data is trusted and secure through customer receipt.

3. The characteristic data is stored securely while in possession of both the customer and the manufacturer.

4. The future production process of transmon processors will be similar to the current production process of classical processors.

5. Differences between the manufacturer's testing conditions/control equipment and the customer's testing conditions/control equipment are negligible.

6. Only negligible physical changes to the transmon chip occur throughout the verification process.

1.6 Implementation

1.6.1 Part A – Determine Potential Applications

Implementation for Part A of the research involves three parts:

- Part A1 – Survey state-of-the-art classical hardware verification and security techniques
- Part A2 – Survey state-of-the-art quantum computing technologies
- Part A3 – Evaluate the relationships between a subset of hardware verification/security techniques and a subset of quantum computing technologies

Future implementation for Part A (not within the scope of this dissertation) will involve the evaluation of the remaining relationships between classical hardware verification/security techniques and qubit implementations. Furthermore, it will investigate and address verification/security measures for new vulnerabilities introduced by quantum computing systems.

1.6.2 Part B – Develop PUCs

Implementation for Part B, which encompasses the preliminary development of PUCs, also falls into three parts. In Part B1, best-fit distribution types are determined for the PUC characteristics of T1 coherence times, T2 coherence times, single-qubit gate error rates, multi-qubit gate error rates, readout error rates, and metrics obtained by quantum process tomography and randomized benchmarking.

In Part B2, outlier removal and data set truncation are tested as methods to reduce sampling errors. In Part B3, the t-test, one-dimensional KS test, JSD test, quadratic discriminant analysis (QDA) test, and two-dimensional KS test are evaluated as qubit discrimination methods for each of the characteristics.

1.7 Results

1.7.1 Part A – Determine Potential Applications

Results of this dissertation research have yielded an evaluation of the relationships between a subset of the hardware verification/security techniques (low power visual inspection, scanning electron microscopy (exterior), scanning electron microscopy (interior), x-ray fluorescence, gate-/transistor-level analysis, temperature cycling, DC power consumption test, functional verification, path delay fault test, PUF implementations, passive/active/reconfigurable hardware obfuscation, optical dissimulation, self-destruction, uniform timing for all operations, adding noise to the power line, power line filtering, power supply shielding, uniform power consumption, transistor- and gate-level secure logic, dynamic reconfiguration, temporal desynchronization, and spatial jamming) and a subset of the qubit implementations (silicon spin and transmon). The most notable result is the identification that RF-DNA and PUFs are applicable with modifications to transmon processors.

1.7.2 Part B – Develop PUCs

In Part B1, results indicate that the distribution types vary across each of the characteristics, with individual characteristics often having more than one "best-fit" distribution type. Although an initial analysis is performed for each characteristic to understand the results and/or identify interesting areas requiring additional research and analysis, the most important result from Part B1 is the identification that none of the characteristics are normally distributed.

Results for Part B2 show that data set truncation, which ensures that data sets only include data points collected between two consecutive maintenance sessions, eliminates most sampling errors and will be a continued data conditioning method. Outlier removal is observed in this section to have little impact on reducing sampling

error rates, and since the validity of outlier removal is uncertain, it is discontinued as a data conditioning method except for the outliers of multi-qubit gate error rate data points equal to 1, which are confirmed true outliers.

Part B3 results show that for single-qubit and multi-qubit gate error rates, the KS test has better PD but worse PFA results compared with the corresponding t-test results. For the readout error rate characteristic, the KS test provides better PD and PFA results than the t-test. Poor results in areas of balanced PD and PFA performance for JSD tests reveal the need for bin selection rules that do not assume normally distributed data. High PFA in conjunction with high PD values for QDA tests indicate that the QDA test is not appropriate for PUCs since QDA assumes multivariate normality, but the data are not normally distributed according to Part B1 results. The two-dimensional KS test results in overfitting for the majority of characteristic pairs, and underfitting for other characteristic pairs, indicating that the two-dimensional KS test requires more refining.

II. Background and Literature Review

This background and literature review chapter explores topics that are related to classical hardware verification and security techniques (Section 2.1.1), emerging quantum computing technologies (Section 2.1.3), transmon characteristics for physically unclonable characteristics (PUCs) (Section 2.2.1), and other related works (Section 2.2.4). Background information on topics related to sampling error rate reduction and qubit discrimination are deferred to Chapter III. The information is organized into subsections that correspond to the identified parts of this research.

2.1 Part A – Determine Potential Applications

2.1.1 Part A1 – Survey the State-of-the-Art Classical Hardware Verification And Security Techniques

2.1.1.1 State-of-the-Art Classical Hardware Verification Techniques

Hardware verification techniques implemented by Defense Advanced Research Projects Agency (DARPA) TRUST and similar programs [40] have taken on many different forms with the evolving sophistication of counterfeiting techniques. Beginning with visual inspection of parts, verification techniques have included DC bias testing and failure analysis, functional testing, physically unclonable function (PUF) implementations, gate- and transistor-level testing [39, 61, 74], topological constraints, [91, 92, 95], radio-frequency (RF) distinct native attributes (DNA) (RF-DNA) fingerprinting [58], and dynamic power analysis [68]. Many of the different forms of hardware verification techniques are organized into a taxonomy [86, 156] and summarized below.

Physical Inspection Techniques

- Incoming Inspection – Low power visual inspection, X-Ray imaging
- Exterior Tests – Blacktop testing, microblast analysis, package configuration and dimension analysis, scanning acoustic microscopy (SAM), scanning electron microscopy (SEM)
- Interior Tests – Optical inspection, wire pull, die shear (hermetic devices), ball shear, SAM, SEM
- Material Analysis – X-Ray fluorescence (XRF), Fourier transform infrared spectroscopy (FTIR), ion chromatography (IO), Raman spectroscopy, energy dispersive spectroscopy (EDS)
- Connectivity Analysis – Gate-/transistor-level analysis, topological constraints
- Miscellaneous - Temperature cycling, hermeticity testing

Electrical Inspection Techniques

- Parametric Tests – DC contact test, DC power consumption test, DC leakage test, DC output short current test, DC output drive current test, DC threshold test, DC bias test, AC propagation delay test, AC set up / hold time test, AC access time test, AC rise/fall time test
- Functional Tests – Functional verification, memory march test, microprocessor functional Fmax analysis
- Structural Tests – Stuck-at fault test, transition delay fault test, path delay fault test
- Burn-In Test

Identification Techniques

- Chip ID – PUF implementations, RF-DNA, hardware metering, secure split test (SST), combating die/IC recovery (CDIR), poly fuse-based technology for recording usage time, electronic chip ID (ECID)
- Package ID – DNA markings, nano-rods, magnetic PUFs.

A subset of the hardware verification techniques is described below. One technique from each category of inspections is selected to provide an introduction to the full scope of hardware verification tests.

Physical Inspection Techniques

Incoming Inspection

- Low Power Visual Inspection – Generally the first verification test performed, low-power visual inspection verifies that the component's packaging is undamaged and originates from the original component manufacturer (OCM). Additionally, it verifies that all shipping labels and documents are correct; all components are protected by the necessary electrostatic discharge (ESD) sensitive or moisture sensitive bags; all components in the lot are oriented in the same direction; all components are labeled with valid markings and date codes; and all components are free of significant visual anomalies (color variations, damaged leads, sanding/grinding markings, etc.) using up to 100X magnification [156].

Exterior Tests

- SEM – SEM is an imaging method that uses an electron column to generate a focused electron beam that scans the exterior surface of an electron component, generating an emission of scattered electrons and x-rays that an electron

detector uses to form an image with superfine resolution. Exterior SEM is used to inspect leads and the exterior of the component. Leads are checked for modification, incorrect metals, scratches, oxidation, and other anomalies. The exterior of the component is checked for sanding/grinding, ghost markings, burn marks, improper texture, oxidation, corrosion, contamination, and other packaging anomalies [156].

Interior Tests

- SEM – Using the same technique as exterior SEM, interior SEM is implemented after decapsulation of the electronic component to inspect the die for markings, reworked bond wires, damage, or other modifications [156].

Material Analysis

- X-Ray fluorescence (XRF) – XRF is a type of spectroscopy in which X-Ray emissions excite electrons from the material of the component, and the material emits photons when the electrons return to their ground state. The frequencies of the photon emissions are used to analyze the material composition of the component [156].

Gate-/Transistor-Level Analysis

- Gate-/Transistor-Level Analysis – This detailed level of analysis is a destructive process in which a fabricated chip is de-layered, the metallization and connectivity of the chip are captured photographically, images are converted to an electronic format, and netlists are generated algorithmically. The netlists are then compared to a golden standard to detect any deviations [92].

Electrical Inspections

Parametric Tests

- DC Power Consumption Test – As the name implies, the DC power consumption test measures the maximum power consumption under worst-case static (steady input logic) and dynamic (changing input logic) situations. Worst-case situations are created by increasing the ambient temperature to worst-case conditions and creating an open circuit on device outputs. The power consumption is determined by measuring the current drawn from the power supply at the specified power supply voltage [51].

Functional Tests

- Functional Verification – Functional verification can occur for both a component and a system of components. Functional verification for a (system of) component(s) determines that the (system of) component(s) produces the expected outputs for a given set of input sequences [86].

Structural Tests

- Path Delay Fault Test – A path-delay fault occurs when the delay of a combinational path exceeds a specified duration. In a path-delay fault test, an input is transitioned, and the output after a specified delay is measured to determine if the output matches expectations [51].

Identification

Chip ID

- PUF Implementation – PUFs (for example, arbiter PUF, ring oscillator PUF (RO-PUF), aging-resistant ring oscillator (ARO) PUF, and SRAM PUF [9]) are circuits that, through fabrication variations, serve as unique fingerprints for an integrated circuit (IC). In an arbiter PUF, two or more challenge bits are implemented simultaneously through a set of multiplexers, and the value of the output depends on which signal traveled the fastest through the circuit. This value is unique per IC [74].

Chip ID

- Radio-Frequency Distinct Native Attribute (RF-DNA) fingerprinting - Fingerprinting has been demonstrated for emitters [99], receivers [111], and integrated circuits [58, 59, 129]. It follows a fundamental process of collecting a burst of a passive or active RF emission for a class of data (for example, a log-periodic antenna (LPA) with open termination or an LPA with load termination), dividing the emission into R regions, dividing each region into sub-regions, and calculating statistics on each region that are used as features to form a fingerprint for device classification [58, 59, 99, 111, 129]. From each sub-region, the responses collected from the RF emission are the amplitude (a), phase (ϕ), and frequency (f). The variance (σ^2), skewness (γ), and kurtosis (κ) of each region's responses are calculated and concatenated into a single vector per region of a burst. The burst composite vector composed of the region vectors serves as a fingerprint for device classification, which is accomplished with multiple discriminant analysis with maximum likelihood (MDA-ML) [58, 59, 99, 111, 129].

2.1.2 State-Of-The-Art Classical Hardware Security Techniques

Two long-standing security vulnerabilities that have existed for classical computing hardware in defense applications include reverse engineering and side channel attacks (SCA). With regard to reverse engineering, some of the countermeasures include hardware obfuscation [66], optical dissimulation [85], and self-destruction [81].

Side-channel attacks exploit the side channels of classical computing systems, which include timing, power, electromagnetic, and acoustic [57]. Countermeasures for SCA exist at the transistor level, gate level, circuit level, and algorithmic level [115]. Only the hardware (transistor-, gate-, and circuit-level) countermeasures are considered in this work. Some of the hardware countermeasures for SCA include uniform timing for all operations [137], power line filtering [78], power supply shielding [57], EM shielding [157], transistor- and gate-level secure logic [157], dynamic reconfiguration [143], temporal desynchronization [110], and jamming [132]. Countermeasures for reverse engineering and SCA are described below.

2.1.2.1 Reverse Engineering Countermeasures

- Hardware Obfuscation – Three categories of hardware obfuscation exist: passive, active, and reconfigurable logic-based. In passive hardware obfuscation (also known as keyless hardware obfuscation), the design description is obfuscated and/or encrypted with cryptographic primitives. As an example, one type of passive obfuscation technique renames signals and reorganizes code to obscure the meaning and function. In active hardware obfuscation (also known as key-based hardware obfuscation), techniques are used to modify the functionality of a design. Key-based logic is inserted into combinational logic paths and/or finite state machines, and the user must provide the correct key to get correct results. Reconfigurable logic-based obfuscation techniques involve mak-

ing part of the chip design-reconfigurable to hide the details of the obfuscation methods during the manufacturing process [66, 142].

- Optical Dissimulation – Optical dissimulation techniques aim to make delayering a chip and analyzing the images difficult by adding dummy wires/interconnections/gates, scrambling the memories, and using custom cells [85].

- Self-destruction – Silicon wafers are manufactured such that they are either attached to or fabricated into tempered glass, which shatters explosively when broken. A heating element within the glass can be remotely triggered, causing the chip to shatter, physically destroying the chip [81].

2.1.2.2 SCA Countermeasures

- Uniform timing for all operations – By implementing uniform timing (same number of cycles) for all operations, variations in data-dependent instruction cycle counts are removed [137].

- Adding noise to the power line – Using a noise generator to add noise to the power line, the signal-to-noise ratio is decreased. Thus, more power traces are needed for power-analysis attacks, making the attacks less efficient [78].

- Power line filtering – Connecting a filter to a power line attenuates patterns in power consumption, making power-analysis attacks more difficult [78].

- Power supply shielding – Large shielding capacitors and other shielding devices are used to isolate the internal time-varying requirements of a circuit from easily probed terminals like the power supply [57].

- Uniform power consumption – One method to help ensure uniform power consumption involves matching all interconnect capacitances. This task is accom-

plished by implementing parallel interconnect routes on the same layer, and by shielding the routes on both sides with a power or ground line [157].

- Transistor- and gate-level secure logic – In addition to matching interconnect capacitances, transistor- and gate-level logic is made secure by implementing techniques such as: dual-rail pre-charge logic, current mode logic (the logical value is represented by current rather than voltage), gate-level randomization (each internal signal is masked by a mask bit), random switching logic and dual-rail random switching logic (a random bit changes every clock cycle), and masked dual-rail pre-charge logic (a mask signal changes every clock cycle and is shared between all gates) [120].

- Dynamic reconfiguration – For field-programmable gate array (FPGA) applications, configurable look-up tables can be integrated to change hardware internals of cipher implementations. As a result, different sets of circuits can operate at the same location at different times, which makes it difficult for an external attacker to predict circuit behavior [143].

- Temporal desynchronization – By implementing random process interrupts in hardware, random delays are inserted in the execution flow, and attackers are unable to synchronize power traces with events of interest [62].

- Spatial jamming – Used in combination with the data encryption standard (DES), spatial jamming has two parts. First, a random number generator creates an input signal for a multiplexer. The multiplexer then causes a cyclic permutation of the stored data based on the random number. Second, each bit is sampled and stored at multiple different (random) positions as determined by the random number generated [132].

2.1.3 Part A2 – Survey State-of-the-Art Quantum Computing Technologies

At present, a standard does not yet exist for the implementation of qubits in a quantum computing system, as there are several emerging technologies, including silicon spin qubits, ion traps, superconducting qubits (transmons), diamond vacancies, topological qubits, and quantum rings. In a silicon spin qubit implementation, qubits, which are defined by the spin states of a single electron, are embedded in silicon and are addressed and controlled by classical CMOS technology [57, 81]. In an ion trap implementation, "atomic clock" qubits are manipulated with laser light and read out through state-dependent fluorescence [108]. In a superconducting qubit implementation, the quantum state of a qubit is manipulated by electromagnetic pulses to control the magnetic flux, electric charge, or phase difference across a Josephson junction [56]. In a diamond vacancy implementation, the spin of free electrons associated with a dopant-vacancy "center" in a diamond's crystal lattice is used as a qubit [89]. In a topological qubit implementation, quantum information is encoded in topological properties of matter so that the qubits are protected from local perturbations [103]. In a quantum ring implementation, the spin of ring-shaped quantum dots placed in a magnetic field serve as qubits that are either controlled electrically or optically [167]. The silicon spin qubit and superconducting qubits are described in greater detail below.

2.1.3.1 Silicon Spin Qubit

Since the relationships between the silicon spin qubit implementation and various hardware verification techniques are discussed in this research [93], background information is provided about this type of implementation. In general, the qubit is implemented by an electron contained in either a single or double quantum dot [123].

Only the single quantum dot will be explored within the scope of this dissertation.

Single Quantum Dot The single quantum dot, which is connected through tunnel junctions to a gate, source, and drain, each with respective capacitances, prohibits electrical conduction through the Coulomb blockade phenomenon. The Coulomb blockade exists when the potential energies of the source and drain (μS and μD, respectively, in Figure 3) exist in the energy gap between N and $N+1$ electrons in the quantum dot, where the energy gap is equal to the charging energy ($\mathrm{E_C}$) plus the single electron energy spacing ($\Delta\epsilon$). The gate voltage ($\mathrm{V_G}$) can be adjusted to lower the energy gap below μS, so that electrons may enter and exit the quantum dot (Figure 4).

The concept of the Coulomb blockade is translated to the implementation of a quantum dot in a quantum computer (Figure 5). Whereas a source and drain exist in the Coulomb blockade model to enable the flow of electrons from source to quantum dot to drain, a quantum dot in a quantum computer only contains an electron reservoir, and electrons flow back and forth between the dot and the reservoir. In concept, the reservoir is both the source and drain, and there are electrodes attached to the reservoir to serve as gate voltages on the reservoir.

Measurement of a Single Quantum Dot Several methods have been developed to conduct measurement of single-qubit operations, including readout based on the Zeeman energy, readout based on singlet-triplet energy, and readout through the Pauli spin blockade. Readouts based on Zeeman energy [71, 123] and the Pauli spin blockade are discussed [162].

Readout Based on Zeeman Energy In a readout based on Zeeman energy [123], a large magnetic field (on the order of 10 T) is applied to align spin

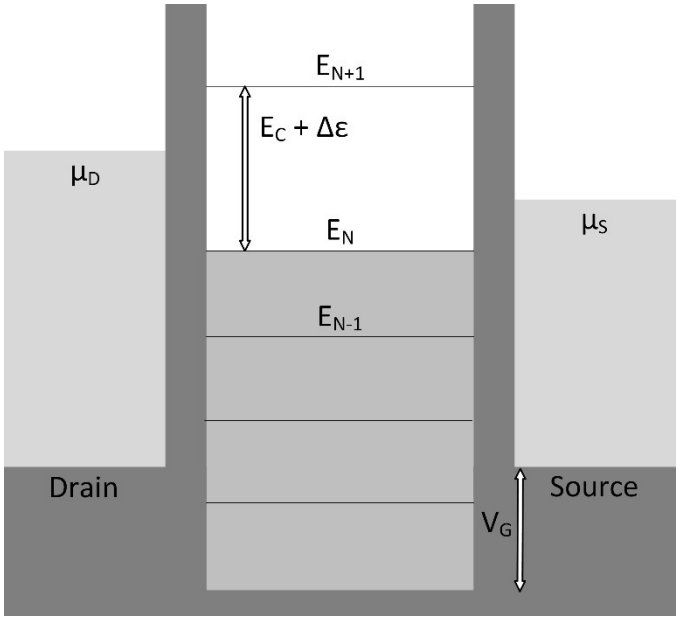


Figure 3: Diagram of energies in a Coulomb blockade (adapted from [123])

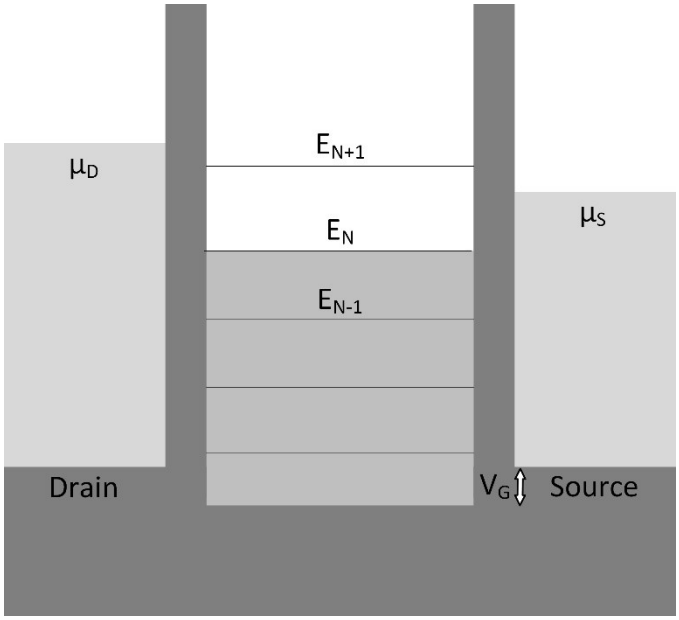


Figure 4: Gate voltage adjusted to allow an electron to enter quantum dot (adapted from [123])

directions (either up or down) and induce separate spin-up and spin-down energies through the Zeeman effect, such that the spin parallel to the magnetic field has a lower energy than the spin anti-parallel to the magnetic field. The temperature is lowered

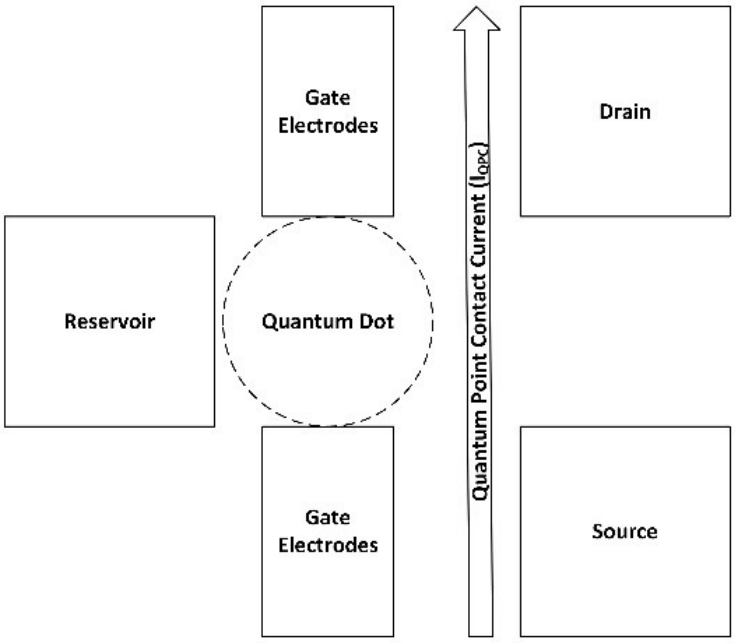


Figure 5: Concept diagram of a quantum dot in a quantum computer with readout based on Zeeman energy (adapted from [123])

to approximately 2K so that the thermal energy is below both the Zeeman splitting energy and the charging energy (E_C) to prevent thermal energy from changing the number of electrons in the quantum dot. Then, the gate voltage on the quantum dot is set to 0V and the gate voltages on the reservoir are adjusted so that the spin-up and spin-down energies of an electron in the quantum dot are higher than the Fermi energy of electrons in the reservoir. This process causes any electrons in the quantum dot to move to the reservoir. Next, a positive gate voltage is applied to the quantum dot to lower the spin-up and spin-down energies of the electron in the dot below the Fermi energy of the electrons in the reservoir. The gate voltages on the reservoir are adjusted to inject only a single electron from the reservoir into the quantum dot, in either the spin-up or spin-down state. To change the spin state of the electron, a microwave pulse at the electron's resonant frequency is applied to create an AC

magnetic field that drives the transitions between the spin-up and spin-down states of the quantum dot [163]. To read the spin state of the electron, the gate voltage on the quantum dot is lowered so that the Fermi energy of the reservoir is between the spin-up and spin-down energies of electrons in the dot. Thus, if the electron is in a spin-up state, it remains in the dot and no changes occur. However, if the electron is in a spin-down state, it escapes to the reservoir and is replaced by an electron in the spin-up state. The actions of an electron entering or leaving the quantum dot are observed by measuring a quantum point contact current, I_{QPC}, which is affected by the charge of the quantum dot.

Readout Based on Pauli Spin Blockade Single-spin readout based on the Pauli spin blockade uses the Pauli exclusion principle, which states that two fermions in a quantum system cannot exist in the same state. This type of readout [162] uses two neighboring quantum dots, with one dot operated as a reference dot, and the other operated as a measurement dot. The dots are initialized to a (0, 2) singlet ground state, meaning that one dot contains zero electrons, and the other dot is in a singlet ground state. A singlet ground state is an electronic state in which the electron spins are paired (one spin-up and the other spin-down) in the ground state. Next, the quantum dots are brought into a (1, 1) state, where each dot contains one electron. One quantum dot contains the spin-up electron, and the other contains the spin-down electron. Then, any operations are conducted with a microwave pulse.

To conduct measurement, the reference dot is placed in a spin-down state if it is not already in a spin-down state, and then an attempt is made to bring the dots back into a (0, 2) singlet ground state. If the attempt is successful, the qubit in the measurement dot is known to be in the spin-up state. If the attempt is unsuccessful, the quantum dots remain in the (1,1) state and the qubit in the measurement dot is known to be in the spin-down state. Observation of the quantum dot states (0, 2)

or (1, 1) is accomplished by using an RF circuit connected to the qubit control gate (circuitry that adjusts the voltages on the qubit). Since each qubit is capacitively coupled to its respective qubit control gate, a greater capacitive difference will exist between dots in a (0, 2) state versus a (1, 1) state. By connecting an RF circuit to each qubit gate, this capacitive difference can be observed by measuring the difference in reflected power in the RF circuit.

Qubit Entanglement for Spin Qubits Entanglement for spin qubits can be achieved through capacitive coupling [150]. The process of entanglement begins with two adjacent qubits that are capacitively coupled, but tunneling is suppressed. Since the (0,2) and (1,1) states have different charge configurations, one qubit imposes a different electric field on the other qubit and causes a shift in the precession frequency [23] of the other qubit, and vice versa. Evolving both qubits under this condition causes a capacitively mediated dipole-dipole coupling that generates an entangled state [150].

Computer Architecture for a Single Quantum Dot Qubit In order to implement a silicon spin qubit in a universal quantum computer, a scalable architecture must exist to implement, couple, address, and control multiple qubits for useful computations. One such architecture based entirely on complementary metal-oxide-semiconductor (CMOS) technology is proposed by Veldhorst et al. [162] and uses classical circuitry to address and control a silicon-based quantum processor. The proposed general architecture begins with a silicon wafer, which contains the classical circuitry, and an isotopically enriched silicon-28 layer, which contains the quantum processor. The silicon wafer with classical circuitry is subdivided into layers of bit lines, word lines, and gates (floating qubit gates and J-gates). The diagram in Figure 6 depicts how each component of the classical circuitry is connected to control the

qubits. The J-gates and qubit gates are arranged in a 2-dimensional array, such that each qubit gate is surrounded on each side by a J-gate (Figure 7). In Figure 6, the qubit gates $Q_{i,j}$ and $Q_{i-1,j}$ control the electron occupancy of each qubit and perform the spin qubit readout, and the $J_{i,j}$ gate couples qubits $Q_{i,j}$ and $Q_{i-1,j}$ and controls the qubit resonant frequency.

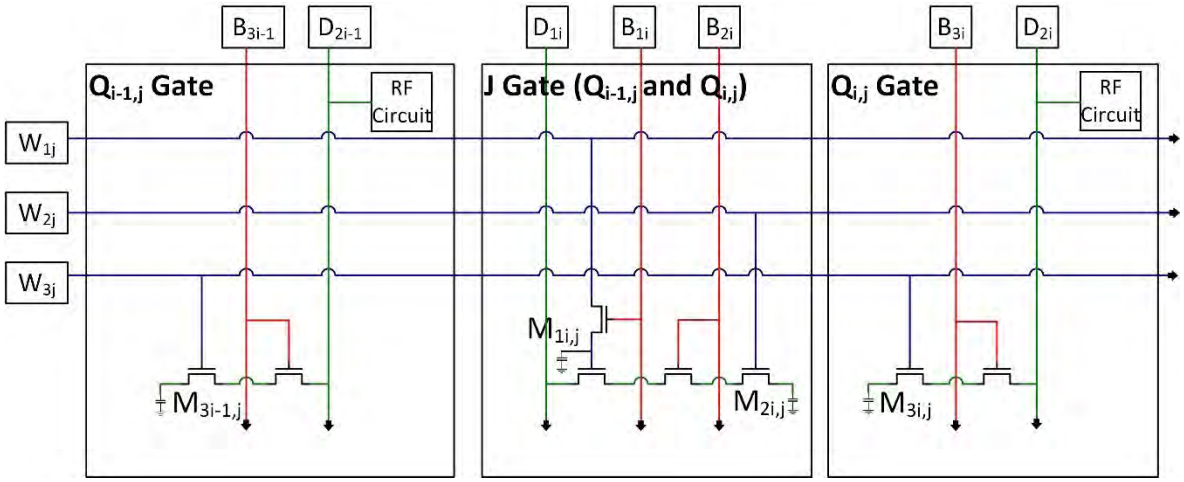


Figure 6: Qubit gate and J-gate circuitry (adapted from [162])

Qubit Gates Each qubit gate is connected to one word line, one bit line, and one data line. For example, the $Q_{i,j}$ gate in Figure 6 is connected to word line W3j, bit line B_{3i}, and data line D_{2i}. By applying a positive voltage on W_{3j} and B_{3i}, floating memory $M_{3i,j}$ is adjusted through a voltage on D_{2i} to either load an electron into or unload an electron from the quantum dot. The voltage on D_{2i} must be set so that the exchange coupling between qubits is negligible when the J-gate is not set to an "on" bias.

J-Gates The J-gate between $Q_{i-1,j}$ and $Q_{i,j}$ is connected to two word lines ($W_{1,j}$ and $W_{2,j}$) and two bit lines ($B_{1,i}$ and $B_{2,i}$). $W_{1,j}$ and $B_{1,i}$ enable individual qubit selection by adjusting the qubit resonant frequency by using the floating memory $M_{1i,j}$ to adjust the magnetic field. By applying an additional positive voltage to $W_{2,j}$ and

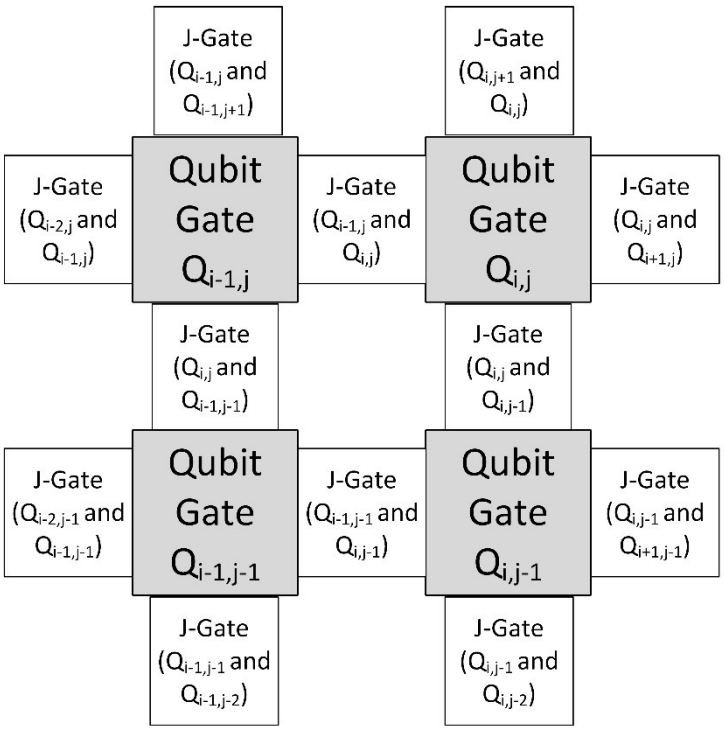


Figure 7: Arrangement of qubit gates and J-gates (adapted from [123])

$B_{2,i}$, the floating memory $M_{2i,j}$ is charged to create an electric field that couples qubits $Q_{i-1,j}$ and $Q_{i,j}$ ("on" bias) [136, **?**, 164]. For two-qubit operations, the J-gate is set to an "on" bias, and the microwave pulse is applied through the data line D_{1i}. The resultant oscillating magnetic field triggers the two-qubit operation. To conduct a one-qubit operation on $Q_{i,j}$, the resonant frequency of $Q_{i,j}$ is tuned to be distinct from the resonant frequency of all other qubits, and a microwave electron spin resonance (ESR) pulse is broadcast at that frequency.

Readout Readout based on the Pauli spin blockade is conducted with D_{2i}. The information provided by Veldhorst et al. [162] appears to present a discrepancy on the qubit initialization process. Section II of the document [162] indicates that the quantum dots are initialized by loading a single electron in each dot. However, readout based on the Pauli spin blockade begins by loading two electrons in a singlet

ground state into one quantum dot while leaving the neighboring dot empty. The remainder of the readout process in the document [162] is consistent with the process for readout based on the Pauli spin blockade.

2.1.3.2 Superconducting Qubit

The next type of qubit implementation to investigate is the superconducting qubit. Superconducting qubits are implemented by creating nonlinear resonator circuits with Josephson junctions [114]. Prior to discussing Josephson junctions, however, one must first explore foundational concepts.

Quantum LC Oscillators As the name suggests, quantum LC oscillators connect an inductor L and a capacitor C with superconducting wires to create discrete energy levels [77]. The LC oscillator is analogous to the linear harmonic oscillator [80], such that the inductance L, the inverse capacitance $1/C$, the flux of the inductor ϕ, and the charge on the capacitor Q are analogous to the mass m, force constant k, position x, and conjugate momentum, respectively [88, 141, 149]. As with the linear harmonic oscillator, the energy level spacing is the linear $\Delta E = \hbar\omega$, which is unsuitable for serving as a two-level qubit system. A non-linear LC circuit element like the Josephson junction is necessary for qubit applications.

Josephson Junctions Josephson junctions are composed of two superconducting electrodes with a thin insulator between them. They behave like a nonlinear LC circuit, where the tunnel junction behaves like a non-linear inductor (Josephson element), and the parallel electrodes behave like a capacitor in parallel [65, 114]. The inductive element facilitates Cooper pair tunneling, and the capacitive element localizes Cooper pairs [49, 114, 141].

Quantum Description of a Josephson Junction There are two main state bases and two main operators that are used to describe the behavior of a Josephson junction [104]: $|N\rangle$, $|\delta\rangle$, $\hat{N}$, and $e^{i\hat{\delta}}$. N is the integer number of Cooper pairs on the island, and δ is the phase difference across the junction. As a note, N can be negative, where $N < 0$ corresponds to a positive charge on the island; $N = 0$ corresponds to a neutral charge; and $N > 0$ corresponds to a negative charge. The operator $\hat{N} = \sum_n N |N\rangle \langle N|$, and intuitively $\hat{N} |N\rangle = N |N\rangle$. The operator $e^{i\hat{\delta}}$ is given by the equation[1] $e^{i\hat{\delta}} = \frac{1}{2\pi} \int_0^{2\pi} d\delta' e^{i\delta'} |\delta'\rangle \langle\delta'|$, and likewise, $e^{i\hat{\delta}} |\delta\rangle = e^{i\delta} |\delta\rangle$. The relationships between the two bases are defined by the equations

$$|\delta\rangle = \sum_{N=-\infty}^{\infty} e^{iN\delta} |N\rangle \tag{1}$$

and

$$|N\rangle = \int_0^{2\pi} d\delta e^{-iN\delta} |\delta\rangle . \tag{2}$$

Given these equations, one can calculate the phase operator acting on the $|N\rangle$ state as a Cooper pair tunneling from/to the island with the equation,

$$e^{\pm i\hat{\delta}} |N\rangle = |N \mp 1\rangle . \tag{3}$$

The phase operator can also be written as the expressions

$$e^{i\hat{\delta}} = \sum_N |N-1\rangle \langle N| \tag{4}$$

and

$$e^{-i\hat{\delta}} = \sum_N |N\rangle \langle N-1| . \tag{5}$$

[1]The notation differentiates δ', the variable of integration, from δ, the phase state as seen in the equation: $e^{i\hat{\delta}} |\varphi\rangle = \frac{1}{2\pi} \int_0^{2\pi} d\varphi' e^{i\varphi'} |\varphi'\rangle \langle\varphi'|\varphi\rangle$

Computer Architecture for a Single Quantum Dot Qubit The charge superconducting qubit, based on the Cooper pair box, is created by connecting a Josephson junction and a capacitor in series with superconducting wires, and a voltage source is placed across them (Figure 8). Other elements that exist in the circuit are a superconducting Cooper pair reservoir (located in the superconducting wire) and a superconducting island (located between the insulator of the Josephson junction and the insulator of the capacitor). The charge of the superconducting island is biased by the voltage source.

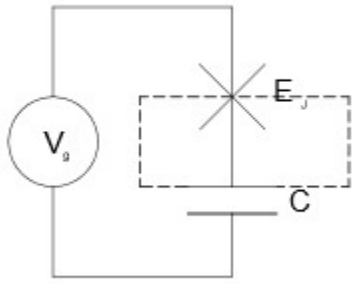


Figure 8: Diagram of a charge qubit circuit (reproduced from [77])

The Hamiltonian $\hat{H}$ of this circuit is[2]

$$\hat{H} = \frac{\left(\hat{Q} - Q_g\right)^2}{2\left(C_J + C_g\right) - E_J \cos\hat{\delta}}. \tag{6}$$

Note that $\cos(\hat{\delta}) = \frac{1}{2}(e^{i\hat{\delta}} + e^{-i\hat{\delta}})$. C_J is the capacitance of the Josephson junction, and C_g is the capacitance of the capacitor. Together, $C_J + C_g$ gives the total capacitance of the island. Q_g is the offset charge induced by the voltage on the gate capacitor. Written in terms of the operator $\hat{N}$, the Hamiltonian can be represented

[2]The first term of the equation comes from the equation for energy U stored in a capacitor, $U = \int_0^Q \frac{q}{C} dq = \frac{Q^2}{C}$

according to the expression

$$\hat{H} = E_C \left(\hat{N} - N_g\right)^2 - E_J \cos\hat{\delta}, \tag{7}$$

where E_c is the energy required to store a Cooper pair on the island, given by the equation

$$E_C = \frac{(2e)^2}{2\left(C_J + C_g\right)}, \tag{8}$$

and N_g is a unitless number related to the offset charge Q_g by the equation $N_g = \frac{Q_g}{2e} = C_g \frac{V_g}{2e}$.

To operate the circuit as a qubit, the value for N_g is set so that $N_g \approx \frac{1}{2}$. For $N_g = \frac{1}{2}$, the Hamiltonian is the same for $N = 0$ and $N = 1$, and a large energy difference exists between states where $N > 1$. Hence, a two-level system can be approximated. However, the degeneracy between $N = 0$ and $N = 1$ prevents this system from being useful as a qubit (except for when creating superposition states). Setting $N_g = \frac{1}{2} + \delta_g$ breaks the degeneracy while maintaining a two-level approximation [104].

The results of varying the ratio of E_J to E_C and varying N_g are displayed in Figure 9. For traditional charge qubits, when $E_J \ll E_C$ and $N_g \approx \frac{1}{2}$, the energies corresponding to $|0\rangle$ and $|1\rangle$ represent a qubit without degeneracy.

To perform operations with a charge qubit, N_g is set to 0 so that a large energy gap exists between $|0\rangle$ and $|1\rangle$ [104]. After a large amount of time has passed, there is a high probability that the system has relaxed into the ground state $|0\rangle$ based on thermodynamics. For fast single-qubit gates, N_g is instantaneously changed to $\frac{1}{2}$ at time t_0, such that $\psi(t_0) = |0\rangle$ and the Hamiltonian for the qubit in $|0\rangle$ becomes

$$\hat{H}\left(t_0\right) = -\left(\frac{E_J}{2}\right)\sigma_x. \tag{9}$$

The eigenstates for the new Hamiltonian are $|\pm\rangle$ and the probabilities of measuring

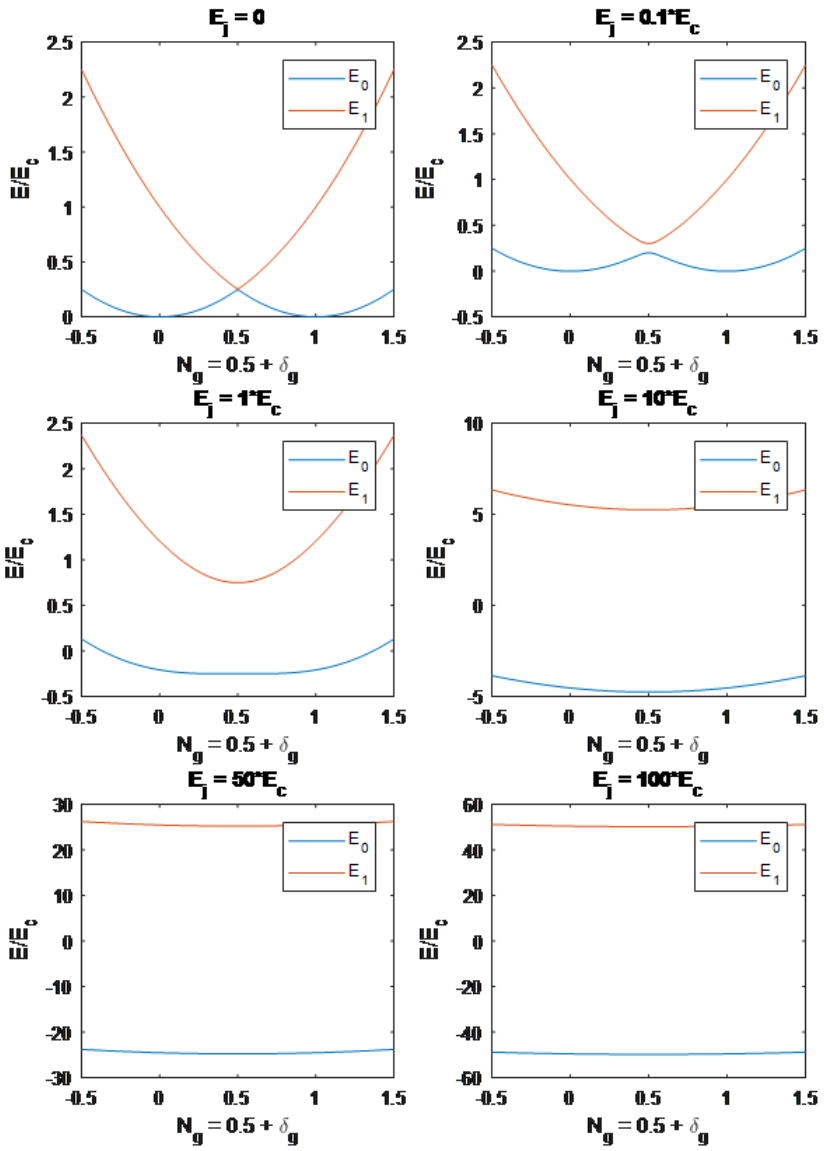


Figure 9: Plots of energy levels in the charge qubit vs. the offset charge as the ratio of E_J/E_C is increased

the system in states $|0\rangle$ and $|1\rangle$ at time t are given by

$$p_0(t) = \frac{1}{2}\left(1 + cos\left(\frac{E_J(t-t_0)}{\hbar}\right)\right) \tag{10}$$

and

$$\hat{H}(t_0) = -\left(\frac{E_J}{2}\right)\sigma_x. \tag{11}$$

After the desired delay, the qubit can be measured or N_g can be changed instantaneously again for the next operation.

Adiabatic (slow) single-qubit gates rely on the theory of adiabaticity, which states that "a quantum system that begins in the nondegenerate ground state of a time-dependent Hamiltonian will remain in the instantaneous ground state provided the Hamiltonian changes sufficiently slowly" [52]. N_g is changed very slowly to $\frac{1}{2}$, such that the final state will become $|+\rangle = \frac{1}{\sqrt{2}}\binom{1}{1}$ due to adiabaticity.

Transmission-line Shunted Plasma Oscillation Qubit (Transmon) Since the operation of the basic charge qubit depends on the value of N_g, its sensitivity to charge noise renders it difficult to implement. As a less charge-sensitive variation on the charge qubit, the transmon is operated where $E_J \gg E_C$. As seen in Figure 9, the energy spacing becomes independent of N_g, meaning that the qubit is no longer sensitive to the gate voltage, and the qubit is less sensitive to charge noise [77]. As a disadvantage, there is reduced anharmonicity in the energy levels, and methods like Derivative Removal Adiabatic Gate shaping are needed to correct for leakage into higher energy levels [79, 121].

Modifications to the basic charge qubit circuit are needed to implement a transmon qubit (depicted in Figure 10). The first modification is to the Cooper pair box. Instead of creating a Cooper pair box with one Josephson junction as in the basic charge qubit, two Josephson junctions are implemented in parallel as a super-

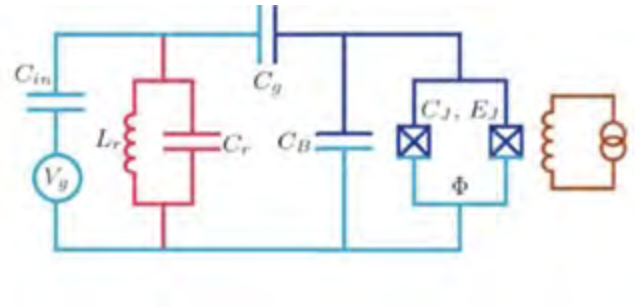


Figure 10: Transmon circuit schematic (reproduced from [101])

conducting quantum interference device (SQUID). While the SQUID behaves like a single Josephson junction from the perspective of the circuit, the SQUID allows for tuning of the Josephson energy by applying an external magnetic flux Φ [101].

The other modifications are in the form of additional circuit elements. The additional elements are: (1) a large shunting capacitor ($C_B \gg C_J$), (2) a resonator LC circuit (harmonic oscillator), (3) a drive circuit flux-coupled to the SQUID, and (4) a microwave drive circuit. The shunting capacitor is used to achieve the relationship $E_J \gg E_C$. Given that $E_C = \frac{(2e)^2}{2\left(C_J+C_B+C_g\right)}$, a large C_B causes E_C to become small [101].[3] The resonator LC circuit serves as the transmission line resonator (specifically, coplanar waveguide resonator) to apply microwave pulses to the qubit. The drive circuit creates the external flux through the SQUID to tune the qubit energy. The microwave drive circuit is capacitively coupled to the resonator to operate the qubit.

Operations on the qubit are achieved by applying a microwave pulse at the qubit resonant frequency. Readout of the qubit is achieved by applying a weak microwave pulse (just a few microwave photons) close to the resonant frequency [43, 101]. Since the microwave pulse is weak, amplification techniques such as using quantum-limited

[3]The author of this source uses e^2 instead of $(2e)^2$ (which is the convention used by other authors cited previously in this work) in the equation for E_C. However, the author recovers the factor of 4 for the Cooper pair in the Hamiltonian, which uses $4E_C$ instead of E_C (the convention used by other authors).

amplifiers (QLA) [43] or superconducting low-inductance undulator galvanometer (SLUG) amplifiers [109] are required. The frequency of the resonator experiences a dispersive shift dependent on the qubit state due to the AC Stark effect.[4] Thus, by measuring the phase or amplitude of the transmitted radiation of the microwave drive field, the state of the qubit can be inferred [101]. It is worth noting here that inter-qubit and intra-qubit variations occur in the operation and readout of transmon qubits due to fabrication variations, environmental noise, and temperature fluctuations [155].

2.2 Part B – Develop Physically Unclonable Characteristics (PUCs)

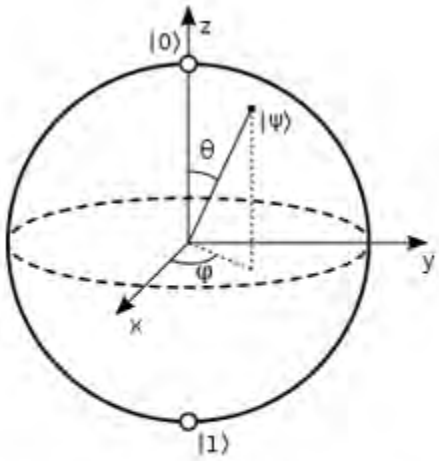
2.2.1 Part B1 – Determine Best-Fit Distribution Types for PUC Characteristics

The characteristics tested for implementation in PUCs in this work fall into three categories: IBM calibration data, quantum process tomography metrics, and randomized benchmarking metrics. For the first category of IBM calibration data, as a result of intra-qubit variations, IBM calibrates their transmon qubits on a near-daily basis and collects data on the characteristics of T1 and T2 coherence times, single-qubit and multi-qubit gate error rates, and readout error rates [41]. It has been observed that variation exists in the T1 and T2 coherence times and in the single-qubit gate, multi-qubit gate, and readout error rates of different qubits and links [155]. T1 and T2 coherence times vary significantly due to process variation, biasing, and temperature drifts. Gate errors occur when small perturbations or experimental conditions cause drifts in device characteristics, and temporal variation of the error rates are demonstrated.

[4]The AC Stark effect is associated with the displacement of atomic energy levels in the presence of an electromagnetic field [83].

2.2.1.1 T1 and T2 Coherence Times

The T1 coherence time refers to the time for a qubit that is in an excited state to decay to the ground state (that is, θ in Figure 11 decays to 0). The T2 coherence time refers to the time for the environment to cause a qubit phase error (that is, a change in φ) [155].



$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle$$

Figure 11: Bloch sphere and equation representing the state of an arbitrary qubit

2.2.1.2 Single-Qubit and Multi-Qubit Gate Error Rates

An example of a single-qubit gate error can be observed through the application of a quantum bit-flip gate to a qubit prepared in the ground state (Figure 12). In a quantum simulator, the output of the bit-flip gate is the excited state 100% of the time (Figure 13). However, in a noisy quantum computer (ibmq_poughkeepsie in Figure 14), the excited state is measured only 91.8% of the time. Part of this error is due to decoherence, part is due to gate error, and part is due to readout error.

Similarly, an example of a multi-qubit gate error can be observed through the application of a controlled-not (cx0_1) gate on two qubits prepared in the ground state.

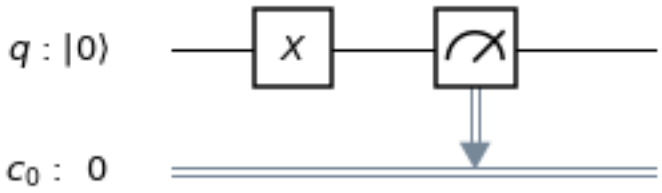


Figure 12: Quantum bit-flip circuit

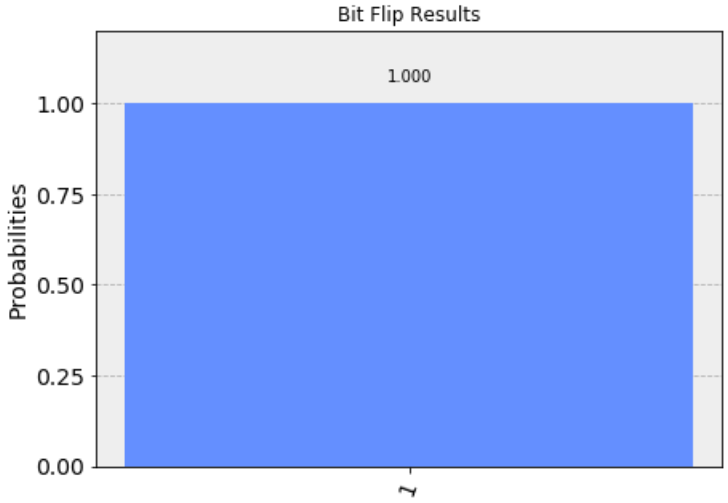


Figure 13: Results of quantum bit-flip gate on simulator

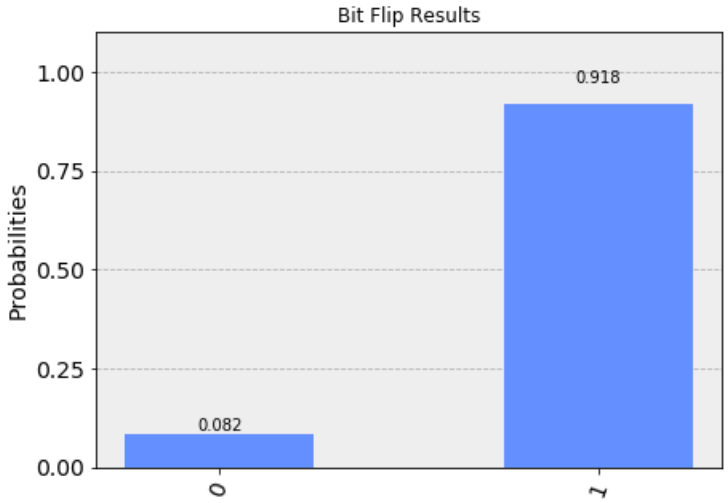


Figure 14: Results of quantum bit-flip gate on a ibmq_poughkeepsie qubit

Since the function of a controlled-not gate is to change the state of the target qubit only if the control qubit is in the excited state, the output of this circuit should result in both qubits remaining in the ground state, as shown by the simulation (Figure 16), but this output is obtained only 95.4% of the time on an ibmq_poughkeepsie qubit (Figure 17).

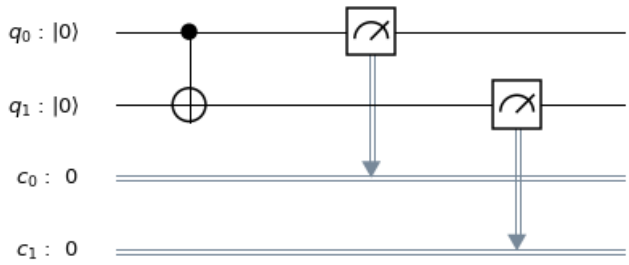


Figure 15: cx0_1 circuit

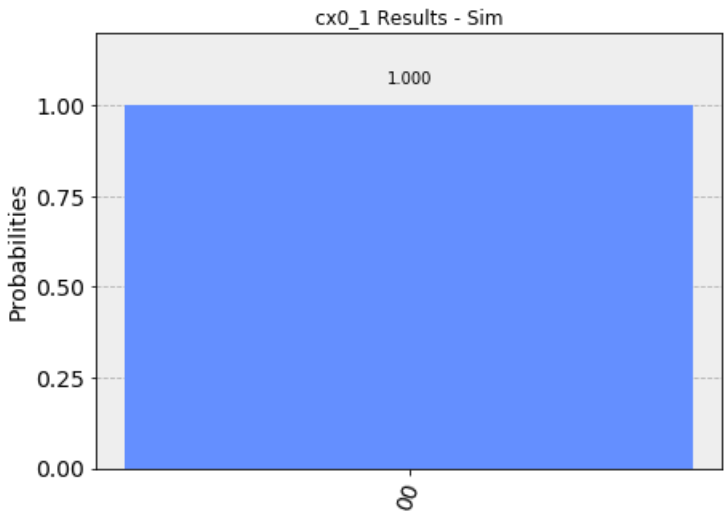


Figure 16: Results of cx0_1 circuit on simulator

2.2.1.3 Readout Error Rates

As the name implies, readout errors take place during the process of qubit readout. These errors are caused by transmission line noise (similar to a classical readout bit-

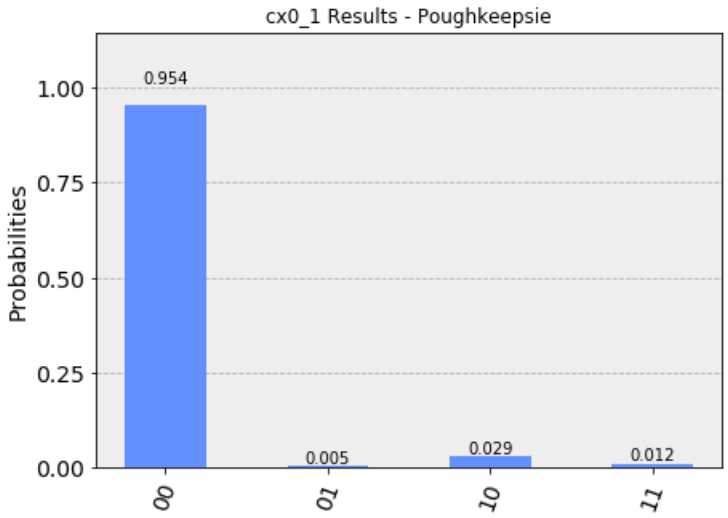


Figure 17: Results of cx0_1 circuit on ibmq_poughkeepsie

flip error) and by T1-readout errors, where the qubit decays from an excited state to the ground state during readout. Another source of readout errors is the small probability of measuring the opposite value due to the overlap in the measurement probability distributions corresponding to the ground and first excited states [122].

2.2.1.4 Other Calibration Data

Recently, IBM has included additional characteristics in their calibration data: anharmonicity, readout assignment error, probability of measuring $|0\rangle$ for a qubit prepared as a $|1\rangle$, probability of measuring $|1\rangle$ for a qubit prepared as a $|0\rangle$, and ID error. These characteristics are not investigated in this work, but should be evaluated for consideration in PUCs in future work.

2.2.1.5 Quantum Process Tomography

Quantum process tomography metrics differ from the calibration data. While the calibration data characteristics described individual qubits and physical links between connected qubits, quantum process tomography (QPT) provides information

for individual gates on individual (sets of) qubits [55]. In QPT, the gate output for a d-dimensional system is represented by the equation

$$\epsilon(\rho) = \sum_{m,n=0}^{d^2-1} \chi_{mn} B_m \rho B_n^\dagger, \tag{12}$$

where ρ is an arbitrary input state density matrix, B_n are operators that form a basis in the space of $d \times d$ matrices, and χ is the gate matrix. By evaluating the experimentally obtained input and output states, χ can be determined empirically. The experimental input and output states are obtained by preparing each of n qubits in the $|0\rangle$, $|1\rangle$, $|+\rangle$, and $|+i\rangle$ states, and then measuring the qubits in the X, Y, and Z bases [31]. By comparing $\chi_{experimental}$ with χ_{ideal}, the gate fidelity can be obtained. The gate fidelity is an ideal candidate to test as a characteristic for verification.

2.2.1.6 Random Benchmarking

In contrast to quantum process tomography, which describes individual gates for individual (sets of) qubits, randomized benchmarking (the third category) measures the average error rate of individual (sets of) qubits, up to an entire processor [135]. The process of randomized benchmarking involves performing random sequences of Clifford gates[5] [159] that, under ideal conditions, would return the qubits to their original state, and then measuring the qubits to determine if they have returned to the original state. The survival probability of the qubits returning to the original state is plotted against the sequence length and fit with non-linear least squares to the exponentially decaying model for sequence fidelity: $F_{seq} = A(\alpha^m) + B$. A and B are terms that absorb the state preparation and measurement errors, m is the sequence length of Clifford gates, and α is the growth factor that determines the

[5]Clifford gates have the unique property of transforming Pauli gates into other Pauli gates. This property makes it possible to easily compute a reversal element to return the qubits to their original state [30, 32].

error per Clifford (EPC). The EPC is calculated by $EPC = \frac{2^n-1}{2^n}(1-\alpha)$ where n is the number of qubits. There are also error limits determined for each of these parameters. The parameters and corresponding error limits can potentially serve as characteristics for transmon verification. The EPC characteristic, however, should be excluded as a characteristic since it does not provide unique information about the qubit or set of qubits beyond what is provided by the α parameter.

2.2.1.7 Other Characteristics

Other characteristics can be explored in the future for potential PUC implementation. One such characteristic is magnetic flux crosstalk. In an ideal transmon processor, control microwave pulses would only act on targeted qubits; however, this has not been achieved in implementation, causing crosstalk that reduces the fidelity of untargeted qubits. This crosstalk can be described as a measure of magnetic flux; specifically, the ratio $X_\Phi = d\Phi_A/d\Phi_B$, where a flux applied to qubit B results in a flux felt by qubit A [44].

Rigetti has developed two methods for directly measuring DC flux crosstalk: the resonator method and the qubit method. In the resonator method, when a DC flux is applied to qubit B, a shift occurs in the phase of the periodic response for the resonator of qubit A, and the magnitude of the phase shift is proportional to the flux crosstalk. In the qubit method, qubit A is flux biased such that the qubit's frequency response is a nearly linear function of applied flux. By measuring the frequency response of qubit A (f_A^{01}) at multiple values of Φ_A and at multiple values of Φ_B, the relation

$$\frac{d\Phi_A}{d\Phi_B} = \frac{\frac{df_A^{01}}{\Phi_A}}{\frac{df_A^{01}}{\Phi_B}}$$

can be used to determine the crosstalk ratio [44]. This crosstalk ratio is another ideal candidate to test as a characteristic for PUCs.

Beyond magnetic flux crosstalk, two other potential characteristics for PUCs are response curves generated from frequency sweeps and response curves generated from Rabi experiments for ground to excited state pulse amplitudes [6]. Both characteristics describe individual qubits.

2.2.1.8 Methods to Determine the Best-Fit Distribution

Please refer to Section 3.3.1 for a discussion that includes relevant background information on methods for determining the best-fit distribution types.

2.2.2 Part B2 – Reduce Sampling Errors

Please refer to Section 3.3.2 for a discussion that includes relevant background information on methods to reduce sampling errors.

2.2.3 Part B3 – Test Qubit Discrimination Methods

Please refer to Section 3.3.3 for a discussion that includes relevant background information on qubit discrimination methods.

2.2.4 Other Related Works

Works Related to Qubit Characteristic Distributions Other works have discussed the concept of variation in qubit performance parameters and have presented histograms to demonstrate the variation. A journal paper by Tannu and Qureshi presents distributions of T1 coherence times, T2 coherence times, single-qubit gate error rates, and multi-qubit gate error rates for all qubit and links of IBM-Q20 [155]. The figures display an envelope outlining the shape of the histograms, suggesting that a distribution model has been fit to the data in the histograms. However, the authors do not describe the generation of the envelopes or any distribution

models. The authors provide the means and standard deviations of the data in addition to the histograms, suggesting that they assume a normal distribution, although they do not state this explicitly.

Additionally, a journal paper by Burnett et al. presents a distribution of T1 coherence times, and the authors fit a Gaussian (a.k.a. normal) distribution model to the data points [50]. However, the authors neither evaluate the goodness of fit nor determine if the Gaussian distribution is the best-fit distribution.

Works Related to Quantum PUFs Two papers applying PUFs to quantum computers have been identified. The first paper on quantum PUFs was submitted to the International Association for Cryptologic Research Cryptology ePrint Archive and made available on arXiv, but withdrawn for currently unknown reasons [46]. The paper was then published on 15 June 2021 in *Quantum*, an open journal for quantum science [47]. While the dissertation work proposed in this document approaches the concept of quantum PUFs from the engineering perspective of developing methods of implementation (determining which characteristics to use, which qubit discrimination methods to use, etc.), the paper by Arapinis et al. approaches quantum PUFs (qPUFs) from the perspective of cryptographic theory. It presents a general quantum game-based framework for analyzing cryptographic primitives. Then, the authors use the framework to provide cryptographic definitions of qPUFs and state the robustness, uniqueness, and collision resistance requirements for qPUFs.

The second paper on quantum PUFs was accepted for publication in the *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* and made available on arXiv [131]. The article focuses not on the security of the supply chain from manufacturer to customer, but rather on the security of cloud-based access to quantum computers. The main risk of cloud-based quantum computing arises from the cloud service's allocation of access to the quantum computers. If the cloud service allocates

one device to two different users at the same time, cross-talk interference will reduce the reliability of the results. Results are also affected if the user requests access to a high-quality device (low gate error rates, for example), but the cloud service instead allocates access to a lower-quality device.

To combat these risks, the authors propose Basic Quantum PUFs (QuPUFs) and Resilient QuPUFs to obtain measurement results from a quantum circuit to verify unique device signatures. Basic QuPUFs create signatures from either the probability of measuring a 1 after conducting a Hadamard gate, or the probability of measuring a 1 after placing a qubit in an excited state and waiting for decoherence by implementing idle gates. Results demonstrated that the characteristics for Basic QuPUFs do not provide clear signatures. Resilient QuPUFs, on the other hand, show greater potential to provide unique signatures. Resilient QuPUFs based on the Hadamard gate have an added variable rotation gate, and Resilient QuPUFs based on decoherence have a variable number of idle gates. The authors sweep the rotation angle from 1 to 5 degrees in steps of 1 degree, and they sweep the number of idle gates from 100 to 500 in steps of 100. The means and standard deviations of the measured probabilities are reported for each step.[6] The probabilities for each step are converted into a signature per qubit, and the signatures per qubit are combined into a signature per chip.

Rather than determining probability of detection (PD) or probability of false alarm (PFA), the authors report the effectiveness of the QuPUFs in terms of inter-die Hamming Distance, which measures the change in response between all pairs of identical chips for the same challenges, and intra-die Hamming Distance, which measures the change in response of a single chip with respect to time under temporal variation. Ideal results are an inter-die HD of 50% and an intra-die HD of 0%. For the Hadamard-based Resilient QuPUF, experimentation achieved best results of a

[6]The paper does not mention any research demonstrating that the measured probabilities follow a normal distribution.

13.82% intra-die HD for ibmq_essex, 3.94% intra-die HD for ibmq_london, and 55.3% inter-die HD for ibmq_essex and ibmq_london. For the decoherence-based Resilient QuPUF, experimentation achieved best results of a 13% intra-die HD for ibmq_vigo, and the authors did not report inter-HD distance.

Works Related to Other Quantum Hardware Verification A blog post published by the Association for Computing Machinery Special Interest Group on Computer Architecture (ACM SIGARCH) discusses the need for quantum hardware verification. In particular, the author focuses on verification through simulation. The conclusion of the blog post is that verification through simulation is not an adequate long-term solution, since "we will always be bootstrapping from simulator to hardware, from hardware to larger hardware," which becomes "more challenging since each qubit we add to future machines makes the verification and simulation problem exponentially more difficult for current machines" [53].

2.3 Summary

This chapter discusses background information related to Part A of this research in Section 2.1. In particular, Sections 2.1.1 and 2.1.2 list and describe classical hardware verification and security techniques, and Section 2.1.3 lists emerging quantum computing technologies and describes the silicon spin qubit and transmon qubit in detail. Furthermore, Section 2.2.1 discusses information regarding calibration data, quantum process tomography metrics, and random benchmarking metrics as transmon characteristics for PUFs. Background information on topics related to sampling error rate reduction and qubit discrimination are deferred to Chapter III. Finally, Section 2.2.4 describes other works related to quantum PUFs and quantum hardware verification.

III. Methodology

3.1 Overview

This chapter describes the methodology of Parts A and B as introduced in Section 1.2. The methodology is organized into subsections, where each subsection corresponds to the different research parts identified in Section 1.6.

3.2 Part A – Determine Potential Applications

The goal of Part A is to determine potential applications of current hardware verification and security techniques to emerging quantum computing technologies. The methodology for Part A describes a survey of the state-of-the-art in both areas, and the establishment of a framework for evaluating the relationships.

3.2.1 Part A1 – Survey State-of-the-Art Classical Hardware Verification and Security Techniques

The methodology of Part A1 is simply to conduct a literature search on the various types of state-of-the-art classical hardware verification and security techniques. This literature search was reported in Section 2.1.1.

3.2.2 Part A2 – Survey State-of-the-Art Quantum Computing Technologies

Like with Part A1, the methodology of Part A2 is simply to conduct a literature search on the various types of state-of-the-art quantum computing technologies. This literature search was reported in Section 2.1.3. As a result of background research conducted for Part A2, the silicon spin qubit and the transmon are selected as quantum technologies considered in Part A3.

3.2.3 Part A3 – Evaluate the Relationships Between a Subset of Hardware Verification/Security Techniques and a Subset of Quantum Computing Technologies

In order to evaluate the relationships between classical hardware verification/security techniques and quantum computing technologies, one must first develop a framework. The following sets of labels are proposed in order to determine the applications and limitations between classical hardware verification/security techniques and quantum computer implementations:

- Destructive (D) / non-destructive (N) – Destructive hardware verification techniques are identified, since destructive techniques may not be appropriate for certain qubit implementations due to high costs.

- Trusted manufacturer required (TMR) / not required (TMNR) – For some of the hardware verification methods, like PUF implementation, a trusted manufacturer is required to provide correct component identification. As trusted manufacturers do not yet exist in this initial stage of quantum computer development, it is important to identify which methods require a trusted manufacturer.

- Object of verification/security exists (E) / does not exist (DNE) – Each of the hardware verification/security methods verifies/secures one or more specific aspects of a classical component. An equivalent aspect (or object of verification/security) may or may not exist in a quantum component.

- Modifications to hardware verification/security technique are required (R) / not required (NR) – For the relationships in which the object of verification/security exists for a quantum component, it might be necessary to modify the hardware verification/security technique in order to apply it to a quantum component.

In addition to developing a framework, a subset of techniques and quantum computer implementations must be selected for evaluation. The subset spans the breadth of the various categories of hardware verification and security techniques. The subset of classical hardware verification methods considered in this dissertation research is: low power visual inspection, scanning electron microscopy (exterior and interior), x-ray fluorescence, gate-/transistor-level analysis, temperature cycling, DC power consumption test, functional verification, path delay fault test, and PUF implementations. The subset of classical hardware security methods considered is: hardware obfuscation (passive, active, and reconfigurable), optical dissimulation, self-destruction, uniform timing for all operations, adding noise to the power line, power line filtering, power supply shielding, uniform power consumption, transistor- and gate-level secure logic, dynamic reconfiguration, temporal desynchronization, and spatial jamming. The subset of quantum computer implementations is: silicon spin qubit and transmon.

Chapter IV presents and discusses the results of applying the proposed labels in order to evaluate the relationship between the subset of verification/security techniques and quantum computer implementations.

3.3 Part B – Develop Physically Unclonable Characteristics

Although the work in Part B is divided into three sub-parts (B1, B2, and B3), it should be noted that these sub-parts were not executed in sequential order when performing the research. Probability of detection (PD) and probability of false alarm (PFA) results for calibration data characteristics were calculated with the t-test before discovering the need for sampling error rate reduction through data conditioning. Furthermore, the t-test, Jensen-Shannon divergence (JSD) test, and quadratic discriminant analysis (QDA) test were implemented on data-conditioned calibration data

characteristics before realizing the need to determine the best-fit distribution types for each characteristic. Since the sub-parts were not executed sequentially, data conditioning is already implemented in Part B1 although it is not discussed until Part B2. Also, some research in Part B3 implements parametric qubit discrimination methods although research in Part B1 shows that none of the characteristics are normally distributed. Despite the non-sequential execution, the research in this document is organized into Parts B1, B2, and B3 because information obtained from Parts B1 and B2 should influence future decisions made in Part B3.

3.3.1 Part B1 – Determine Best-Fit Distribution Types for PUC Characteristics

3.3.1.1 Part B1.1 – Collect Data

Naturally, the first step towards determining the best-fit distribution types for the characteristics is to collect the data for these characteristics. As previously noted, fabrication variations, environmental noise, and temperature fluctuations cause inter-qubit and intra-qubit variations. Transmon qubits require frequent calibrations because of these variations, and at least once per day, IBM calibrates the characteristics of T1 and T2 coherence times, single-qubit and multi-qubit gate error rates, and read-out error rates [41]. The screenshot in Figure 18 shows an example of the calibration data published by IBM on the IBM Quantum Experience website. The calibration data for each of the transmon processors is downloaded daily from the website and imported into a database using a custom program (Figure 19) designed in C# in Microsoft Visual Studio. The custom program also creates the MATLAB m-files that execute the data conditioning and PD/PFA/sampling error tests in Parts B2 and B3.

Additional characteristics for continued development of PUCs are metrics obtained by randomized benchmarking [55, 135] and quantum process tomography [55]. Ran-

Qubit	T1 (Âµs)	T2 (Âµs)	Frecuency	Readout e	Single-qubit	CNOT error rate
Q0	44.79331253	52.31846409	5.2863131	3.35E-02	9.79E-03	CX0_1: 5.721e-2, CX0_2: 5.214e-2
Q1	59.79923879	56.88998557	5.237848	2.48E-02	4.12E-03	CX1_2: 6.776e-2
Q2	70.67636957	60.64315893	5.0302726	1.32E-02	7.04E-03	
Q3	63.87970288	30.97101353	5.2960752	5.62E-02	1.03E-02	CX3_2: 6.859e-2, CX3_4: 3.961e-2
Q4	47.92653377	43.12488541	5.0846544	3.00E-02	5.32E-03	CX4_2: 5.931e-2

Figure 18: Screenshot of ibmq_yorktown calibration data for 25 August 2019 provided by IBM

domized benchmarking is performed for sets of one, two, and three qubits. Quantum process tomography is performed on the X gate (1 qubit), Hadamard gate (1), CX gate (2), SWAP gate (2), CCX gate (3), and CSWAP gate (3).

In order to collect data for metrics obtained by randomized benchmarking and quantum process tomography, programs executed through the IBM Quantum Experience submit quantum process tomography circuits to quantum processor backends and perform calculations on the circuit results to generate the data. While the calibration data is collected on a daily basis, the data from randomized benchmarking and quantum process tomography are collected on a near-daily basis. Due to the limitations of the research environment, usage of the quantum devices available through the IBM Quantum Experience is shared amongst users, meaning that circuit jobs must wait in a queue for execution. Furthermore, for two- and three-qubit randomized benchmarking and quantum process tomography circuits, the job size is so large that performing randomized benchmarking or quantum process tomography on just one set of qubits must be divided into several jobs, with each job occupying a space in the queue. Thus, the randomized benchmarking and quantum process tomography programs can take up to several days to execute to completion, particularly for three-qubit circuits.

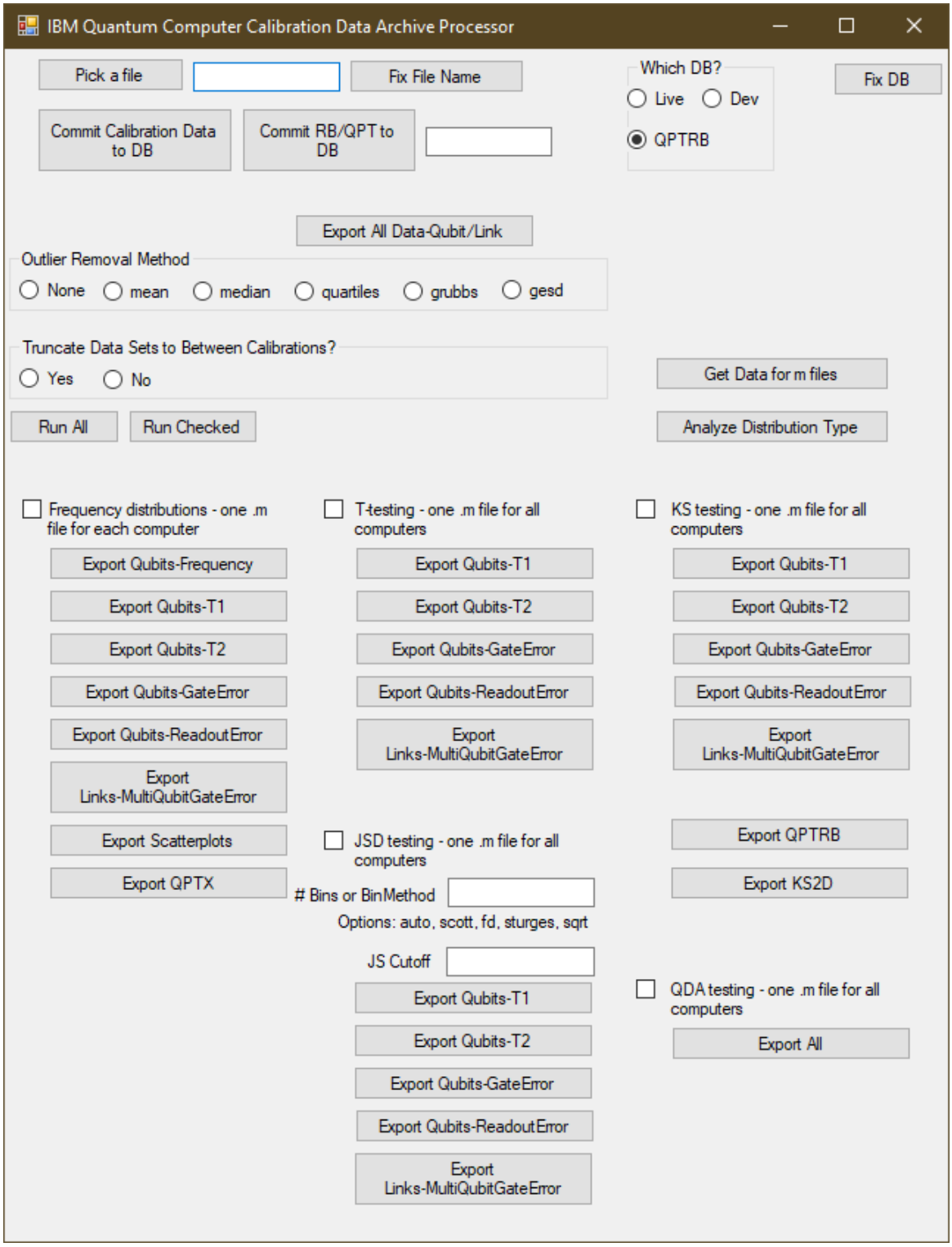


Figure 19: Graphical user interface of a custom program created to process characteristic data

3.3.1.2 Part B1.2 – Characterize Differences Between Individual Qubits and Links

After collecting the characteristic data, it is necessary to ensure the data differs from qubit to qubit and from link to link. Otherwise, the qubits and links cannot be individually identified. By comparing histograms, corresponding means and standard deviations, and chronological plots of each of the calibration characteristics for each of the qubits and links, the different behaviors for each qubit and link become apparent. While the means and standard deviations do not in general fully describe non-normally distributed data, they are sufficient to demonstrate that the distributions differ among qubits and among links. As an example, Table 2 shows the histograms and corresponding means and standard deviations of T1 data for Yorktown qubits Q00 and Q01, and Figure 20 shows the chronological plot of T1 data for Yorktown qubits Q00-Q04. Histograms and chronological plots for the set of calibration data for all of the Yorktown qubits are provided in Section 4.3.1.2.

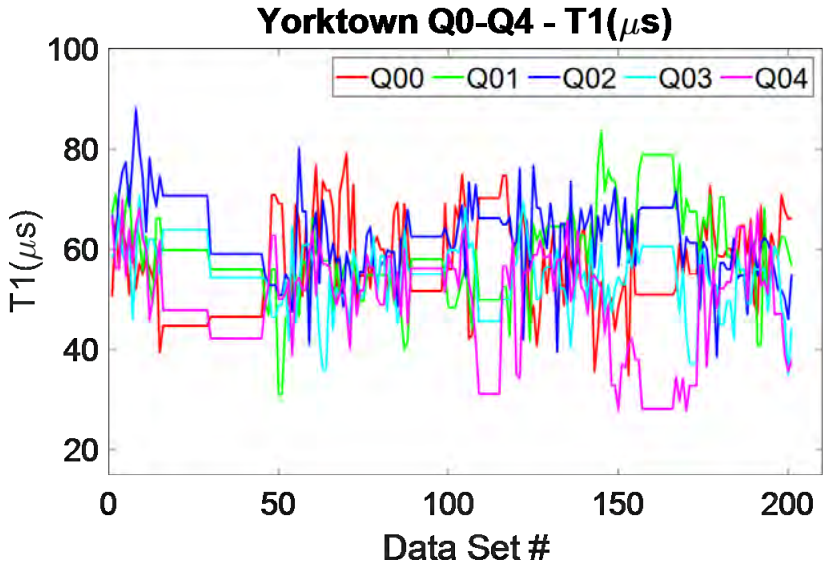


Figure 20: Chronological plot of T1 data for Yorktown qubits Q00-Q04

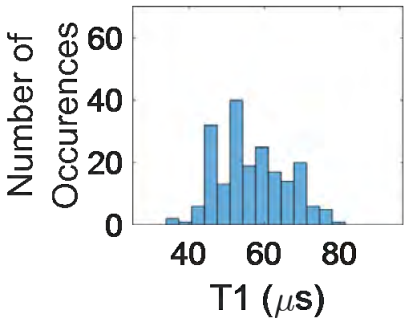
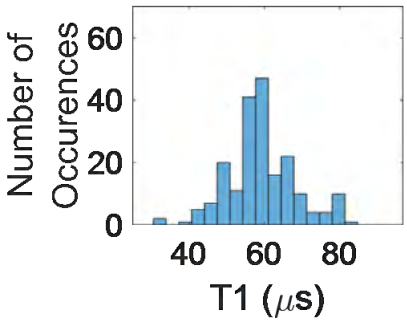
Qubit Number	Histogram	Mean	Standard Deviation
	T1		
Q00		56.56	9.28
Q01		59.10	9.00

Table 2: Yorktown Q00 and Q01 Histograms for T1 Data

3.3.1.3 Part B1.3 – Calculate p-values and the Negative Log-Likelihood, Akaike's Information Criterion, Corrected Akaike's Information Criterion, and Bayesian Information Criterion Metrics

As discussed in Chapter IV, determining the distribution types that best fit the characteristic data is necessary to guide future development of physically unclonable characteristics (PUCs). The main goal of Part B1.3 is to determine whether or not the characteristics follow normal distributions so that qubit discrimination methods can be selected accordingly in Part B3. The distribution types tested are the Birnbaum Saunders (2 parameters), Burr (3), exponential (1), extreme value (EV) (2), gamma (2), generalized extreme value (GEV) (3), generalized Pareto (3), half-normal (2), inverse Gaussian (2), logistic (2), loglogistic (2), lognormal (2), Nakagami (2), normal (2), Rayleigh (1), Rician (2), t location-scale (3), and Weibull (2). These distribution types are selected since they are readily available distribution types in MATLAB and cover a wide range of symmetric and skewed probability density shapes. In order to test the distribution types and determine the p-values, a distribution model is generated using maximum likelihood estimation for each characteristic, each distribution type, and each qubit/link. Then, the median of the p-values is taken over all qubits/links per characteristic and distribution type.

Two tests are performed to calculate the p-values: the Chi-square goodness-of-fit test and the Kolmogorov-Smirnov goodness-of-fit test. The Chi-square test requires binning of data, and the test statistic value depends on how the data is binned [1]. The main weakness of the Chi-square test is that an insufficient number of data points will cause the Chi-square method to fail. While the Kolmogorov-Smirnov (KS) test does not require data binning, it is more sensitive than the Chi-square test to data at the center of a distribution and less sensitive to data at the tails [22]. In this research,

distribution types are rejected for p-values below the significance level of $\alpha = 0.05$. In addition to the p-values, the negative log-likelihood (NLL), Akaike's information criterion (AIC), corrected AIC (AICc), and Bayesian information criterion (BIC) metrics are calculated to determine which of the distribution types provide the best fit.

For each of the four metrics (NLL, AIC, AICc, and BIC), the distribution model that provides the lowest value indicates the best fit to the data. NLL is simply the negative of log-likelihood, which is used to compare the likelihood of observing a given set of data points under different parameters [28, 67, 84]. NLL is calculated according to the expression,

$$\text{nll}(\theta|x) = -\sum_{i=1}^{n} \log f(\theta|x_i), \tag{13}$$

where n is the number of independent samples from the distribution $f(\theta|x)$. The benefit of presenting NLL instead of log-likelihood is that, like the AIC, AICc, and BIC, the lowest NLL value indicates the best fit (whereas the highest log-likelihood value indicates the best fit) [28]. AIC, AICc, and BIC are based on a goodness-of-fit term (here, NLL) and introduce a penalty to prevent overfitting [69]. AIC is calculated as

$$\text{AIC} = 2 \cdot \text{nll} + 2k, \tag{14}$$

where k is the number of independently adjusted parameters in the model. AIC has a strong negative bias when k increases relative to the number of data points n, which can lead to overfitting [45, 96]. AICc corrects AIC at small values of n and converges to AIC at large values of n. AICc is calculated as:

$$\text{AICc} = 2 \cdot \text{nll} + (2kn)/(n - k - 1). \tag{15}$$

BIC is often paired with AIC in model selection. BIC has a larger penalty term than

AIC, and as a result, prefers models with fewer parameters [145]. BIC is calculated as:

$$\text{BIC} = 2 \cdot \text{nll} + \ln(n) \cdot k. \tag{16}$$

Since values for the NLL, AIC, AICc, and BIC metrics have a meaningful interpretation only when comparing models for the same set of data, the process for generating and comparing the NLL, AIC, AICc, and BIC values differs slightly from the process for p-values. For each characteristic, each distribution type, and each qubit/link, a distribution model is generated with maximum likelihood estimation. From this point, the methodology follows two different paths. In the first path, for each metric, the distribution type with the lowest value is selected for each qubit/link. The distribution type selected for the greatest number of qubits/links is considered the best distribution type for the characteristic. In the second path, for each metric M, the metric's weight is determined with the equation [165]:

$$w_i(M) = \frac{\exp\left(-\frac{1}{2}(M_i - \min(M))\right)}{\sum_{j=1}^{J} \exp\left(-\frac{1}{2}(M_j - \min(M))\right)}, \tag{17}$$

where J is the total number of models tested for a set of data. The median of the weights is then taken across all qubits/links for the given distribution type. The distribution type with the highest median weight is considered the best fit for the characteristic.

3.3.2 Part B2 – Reduce Sampling Errors

As a workaround for the unavailability of data from a manufacturer and customer as proposed in the use case, the data collected in Part B1 is partitioned into two disjoint, chronologically-ordered sets. The first set represents the N_1 data sets provided by a manufacturer (from Day 1 to N_1), while the second set represents the N_2 data

sets collected by a customer (from Day $N_1 + 1$ to Day $N_1 + N_2$).[1] The partitioning repeats multiple times for $N_1 = 20, 40, 60, \ldots$ and $N_2 = 20, 40, 60, \ldots$. This variation in data set size reflects the numbers of data points per characteristic that a manufacturer and customer would need to collect to achieve a desired PD or PFA.

The fact that the data points are ordered chronologically, however, introduces a complication. In general, it is important to select samples randomly. Samples not chosen randomly are more prone to sampling bias leading to sampling errors, in which the sample test statistics (for example, the mean and variance) have a statistically significant difference from the population [39, 41]. However, the proposed use case makes random sample selection impossible. As a result, sampling bias can cause sampling errors that potentially reduce the accuracy of qubit and link discrimination. Part B2 evaluates the extent to which sampling errors are present and tests different methods to reduce sampling errors. To accomplish this, a multi-stage process for data conditioning is proposed. The data conditioning action taken on each paired set (the N_1 set and the N_2 set) changes with each stage.

- First stage - No data conditioning occurs in order to develop a baseline for sampling errors.

- Second stage – The presence of apparent outliers exists in Table 10 in Section 4.3.1.2. Outliers are removed during this stage. Five different methods for outlier removal are tested: quartiles, mean, median, Grubbs, and GESD. Each of these methods is a built-in MATLAB function [96]. The quartiles, mean, and median methods are expected to out-perform the Grubbs and GESD methods, as the Grubbs and GESD methods assume a normal distribution (results for Part B1 in Chapter IV show that the characteristics do not follow a normal

[1] Ideally, the data points are collected on a daily basis (once per 24-hour time period). However, as mentioned previously, some characteristic data points take longer than 24 hours to collect. As such, the term "Day" does not refer to a strict 24-hour time period.

distribution).

- Third stage – Multiple segments of "flat tops" appear in Table 10 in Section 4.3.1.2. These "flat tops" represent periods of time that the device is down for maintenance, and new characteristic data are not available. The maintenance process involves "manual recalibration of qubit control pulses, issues with the dilution refrigerator or the control electronics, software/configuration issues, etc." Additionally, "On rare occasions the maintenance may require a partial warmup of the dilution refrigerator housing the device, and some of the qubits' frequencies and/or coherence times may be slightly different after such an excursion due to microscopic changes in the materials that make up the qubits" [117]. As a result, the maintenance process potentially violates the assumptions that the differences between control equipment are negligible and that any physical changes to the chip are negligible. Thus, data sets need truncation to include only the points collected between two consecutive maintenance sessions.

- Fourth stage – Other potential sources of sampling errors could be detected through autocorrelations in the time series. By computing the autocorrelation time (if it exists) and changing the sampling rate to avoid autocorrelation (that is, sample every few days), sampling errors potentially can be reduced.

- Fifth stage – Another source of sampling errors could be caused by the effects of device aging. To determine the existence of any aging effects, a moving average can be calculated for each qubit/link and for each characteristic. The moving averages can be inspected to identify and quantify any aging trends. There exists a potential obstacle to obtaining useful results, however. Often, device aging is measured in terms of number of power cycles [64], time under voltage

stress [154], hours of burn-in treatment, or radiation dosage [76]. However, given the data currently available for this research, device aging can be measured only in number of days, which may prove useless since each qubit experiences a different amount of stress (that is, number of gate operations) per day.

3.3.3 Part B3 – Test Qubit Discrimination Methods

It is necessary to select and test different characteristics and different qubit discrimination methods to determine which are best suited for PUCs. Two metrics are calculated: probability of detection (PD) and probability of false alarm (PFA). These metrics are common in detector performance analysis and are implemented by Defense Advanced Research Projects Agency (DARPA) [61]. The following tests and algorithms are proposed as detectors: t-test [151], two-sample KS test [2, 22], Jensen-Shannon divergence (JSD) [107], quadratic discriminant analysis (QDA) [11, 67], and multiple discriminant analysis with maximum likelihood (MDA-ML) [90]. The methods are described below, and comparisons of the methods are provided in Tables 3 and 4.

3.3.3.1 Qubit Discrimination Methods

***t*-Test** As one of the most common types of hypothesis tests, the t-test was selected as a simple starting point. However, the t-test assumes a normal distribution, which is anticipated to cause poor qubit performance since a normal distribution of qubit characteristic data is not guaranteed. The t-test is conducted with significance levels of $\alpha = 0.001$ and $\alpha = 0.05$.

Kolmogorov-Smirnov Test The KS test was selected to remedy the limitations of the t-test, as it does not assume a normal distribution. The KS test determines if two samples come from the same population based upon the greatest

distance between two empirical cumulative distribution functions (ECDFs). The KS test statistic is

$$D = \max_x \left(\left| \hat{F}_1(x) - \hat{F}_2(x) \right| \right), \tag{18}$$

where $\hat{F}(x)$ is the proportion of values from a sample distribution less than or equal to x. Like the t-test, the KS test is conducted with significance levels of $\alpha = 0.001$ and $\alpha = 0.05$.

Jensen-Shannon Divergence While the Jensen-Shannon divergence (JSD) algorithm is primarily used in clustering and classification applications rather than same/not-same applications [10, 127, 160], it can also be applied to qubit discrimination. The algorithm was selected for its application to quantifying the distinguishability between two or more distributions and for the fact that it does not assume a normal distribution. The JSD measure is a symmetric version of the Kullback-Leibler divergence (KLD) measure for two distributions and is calculated as

$$\mathrm{JS}(p_1, p_2) = \frac{1}{2}\mathrm{KL}(p_1, m) + \frac{1}{2}\mathrm{KL}(p_2, m), \tag{19}$$

where $m = \frac{1}{2}(p_1 + p_2)$ and where $\mathrm{KL}(p_1, p_2) = \sum_{x \epsilon X} p_1(x) \ln \frac{p_1(x)}{p_2(x)}$ for discrete probability distributions and $\mathrm{KL}(p_1, p_2) = \int p_1(x) \ln \frac{p_1(x)}{p_2(x)} dx$ for continuous probability functions [87, 107, 140].

The calculation for the JSD measure returns a value between 0 (identical distributions) and 1 (distributions have no overlap). In order to make the same or not-same determination for the qubit, a cutoff point between 0 and 1 must be selected. Adjustable parameters for this test are the number of bins b and cutoff point c, where $0 < c < 1$.

Two different approaches are taken for bin count selection. The first approach is heuristic, where bin counts $b = 15$ and $b = 25$ are arbitrarily chosen and then increased

or decreased based on performance. The second approach is to apply the following binning algorithms available in MATLAB to determine the number of bins: Scott's rule, Freedman-Diaconis rule, Sturge's rule, and the square root rule [20]. Scott's rule and the Freedman-Diaconis rule are based on the standard deviation and interquartile range of the data set, respectively, while the simpler Sturges' rule and square root rule simply calculate the number of bins as a function of the number of elements in the data set. The square root rule and Sturge's rule are expected to perform well for small sample sizes and for characteristics that have non-skewed normally distributed data, but will otherwise cause oversmoothing [146]. Scott's rule minimizes the integrated mean square error and is expected to perform best for normal distributions since it is based on the standard deviation of a distribution [20, 147]. The Freedman-Diaconis rule is similar to Scott's rule but is expected to perform better for characteristics with many outliers in the data, as it is based on interquartile range rather than the standard deviation [20, 147].

The cutoff points for JSD are selected heuristically. As the number of bins decreases, the probabilities of detection and false alarm are expected to decrease, and as the cutoff point decreases, the probabilities of detection and false alarm are expected to increase. Chapter IV presents results for bin counts $b = 15$ and $b = 25$, for the binning algorithms of Scott's rule, Freedman-Diaconis rule, Sturge's rule, and the square root rule, and cutoff points $c = 0.05$, $c = 0.10$, $c = 0.15$, and $c = 0.20$.

As demonstrated in Chapter IV, optimal performance requires exploration of more complex rules, like cross-validation [148], since the data are not normally distributed. Further development of JSD as a qubit discriminator (and as a technique for determining sampling errors) should investigate methods of creating test statistics for the JSD test so that the statistical power of this test can be analyzed [5]. One method of creating test statistics involves comparing divergence scores of two known distri-

butions to a significance level α in order to determine the cutoff for testing a null hypothesis [126]. A method for unknown distributions proposes estimating a test statistic for the two-sample JSD test with a jackknife estimator. The test statistic then can be compared to chi-square statistics [153].

Quadratic Discriminant Analysis Quadratic discriminant analysis (QDA) is a generalization of linear discriminant analysis (LDA) and is advantageous for discriminating qubits represented by multiple characteristics at one time. QDA first models a class density as a multivariate Gaussian. The quadratic discriminant function is

$$g(x) = w_0 + \sum_{i=1}^{d} w_i x_i + \sum_{i=1}^{d} \sum_{j=1}^{d} w_{ij} x_i x_j, \tag{20}$$

where d represents the dimensionality of input vector x, and w represents elements in a weight vector [67]. The input vectors are unweighted in this implementation. Since the QDA models allow each class to have different means and covariance matrices, the outputs of QDA are quadratic decision boundaries [11].

QDA models predict the classifications of new observations. Three quantities are implemented in the calculation for the prediction: posterior probability, prior probability, and cost [29]. The equation for the prediction is

$$\hat{y} = \underset{y=1,\ldots,K}{\arg\min} \sum_{k=1}^{K} \hat{P}(k|x) C(y|k), \tag{21}$$

where $\hat{y}$ is the predicted classification, K is the number of classes, $\hat{P}(k \mid x)$ is the posterior probability of class k for observation x, and $C(y \mid k)$ is the user-defined cost of classifying an observation as y when its true class is k. The equation for posterior probability is $\hat{P}(k \mid x) = \frac{P(x|k)P(k)}{P(x)}$ where $P(x \mid k)$ is the density function of the

multivariate Gaussian given by the equation,

$$P(x \mid k) = \frac{1}{\left((2\pi)^d |\Sigma_k|\right)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_k)\Sigma_k^{-1}(x-\mu_k)^T\right), \tag{22}$$

and where Σ_k is the $d \times d$ covariance matrix for point x and μ_k is the $1 \times d$ mean.

Applying QDA for qubit discrimination has a few significant drawbacks. First, QDA models the classes with multivariate Gaussians, but the data are not guaranteed to follow a Gaussian distribution. Second, QDA makes classification decisions, not same/not-same decisions. Thus, QDA will assign all test data to the class with the best fit, even if it is a poor fit. This can potentially reduce the probability of detection, as a qubit is considered different in this test only if it receives an incorrect label.

For example, imagine a comparison between Processor A and Processor B, where the first qubit in each processor is called "Q00." Given a model for Processor A, if an observation from qubit Q00 from Processor B is labeled as "Q00," then the two qubits will be considered the same by the test, even though they are different. Q00 from Processor B will only be identified as different if given a label other than "Q00." Furthermore, QDA has a bias in performance towards larger processors, as processors with more qubit classes will improve the "fit" of the classifications.

Multiple Discriminant Analysis - Maximum Likelihood (MDA-ML) Like QDA, MDA-ML enables consideration of multiple characteristics at one time. Multiple discriminant analysis (MDA) is a generalization of the Fisher linear discriminant analysis (reduced-rank linear discriminant analysis) [90], which reduces dimensionality by finding the best separation of classes. Maximum likelihood (ML) estimation (which can use JSD as a goodness-of-fit measure [106]) is then applied for any type of distribution to determine the classification boundaries. An extensive discussion of MDA and ML is provided by Duda et al. [67]. While MDA-ML

is applied to the first four statistical moments of RF data regions as implemented in radio-frequency (RF) distinct native attributes (DNA) (RF-DNA), MDA-ML for PUCs would be applied to the raw data of the characteristics themselves, as too few data points exist to divide the characteristic data into enough regions for MDA-ML to be advantageous in such an application. MDA-ML is not tested in this work, but is recommended for future research.

Multi-Dimensional KS Test Experimentation with the above discrimination methods showed the most promising PD/PFA results for the KS test, as discussed in Chapter IV. For this reason, the multi-dimensional KS test is also selected for experimentation in this work. While the test statistic in the 1-dimensional KS test is the maximum absolute distance between the cumulative distribution functions (CDFs) of two datasets, the CDF is not well defined in more than one dimension. In the algorithm for the two-dimensional KS test proposed by Fasano and Franceschini [75], instead of comparing CDF distances, the two-dimensional KS test finds the maximum difference between the integrated probability in each of four natural quadrants around a given point. Thus, for each data point, the proportion of remaining data points in each quadrant is calculated for each sample, and the test statistic is the maximum difference between the proportions.

In a simple example to demonstrate the algorithm, suppose a data set contains three red triangles and three blue stars. First, the area around the first red triangle is divided into four quadrants (Figure 21). The upper left and lower left quadrants contain zero triangles and zero stars (0), with the absolute difference between proportions equal to zero. The upper right quadrant contains one of two triangles (0.5) and zero stars (0), with the absolute difference between proportions equal to 0.5. The lower right quadrant contains one of two triangles (0.5) and three of three stars (1.0), with the absolute difference between proportions equal to 0.5. This analysis is continued

for all remaining data points (Figures 22-26). The greatest absolute difference exists in the lower right quadrant of Figure 24, where there are two of two stars (1.0) and zero triangles (0), with an absolute difference equal to 1.0.

Once the maximum absolute difference is determined, the p-value can be calculated with the equations

$$p = 2 \sum_{j=1}^{\infty} (-1)^{j-1} \, e^{-2j^2\lambda^2} \tag{23}$$

and

$$\lambda = \left(\frac{\sqrt{N} \cdot D}{1 + \sqrt{1 - r^2 \left(0.25 - \frac{0.75}{\sqrt{N}} \right)}} \right), \text{where } N = \frac{N_1 N_2}{N_1 + N_2}, \tag{24}$$

r is the coefficient of correlation, and D is the maximum absolute difference.

This dissertation research adapts the two-dimensional KS test from code provided by Lau [105] and from code provided by Press, Teukolsky, Vetterling, and Flannery [133, 134], both of which are based on the above algorithm proposed by Fasano and Franceschini [75]. The algorithm can be expanded easily to additional dimensions, but is limited to two dimensions for this work.

Tables 3 and 4 compare each of the different qubit discrimination methods described in this section, and highlight the strengths and weaknesses of each method.

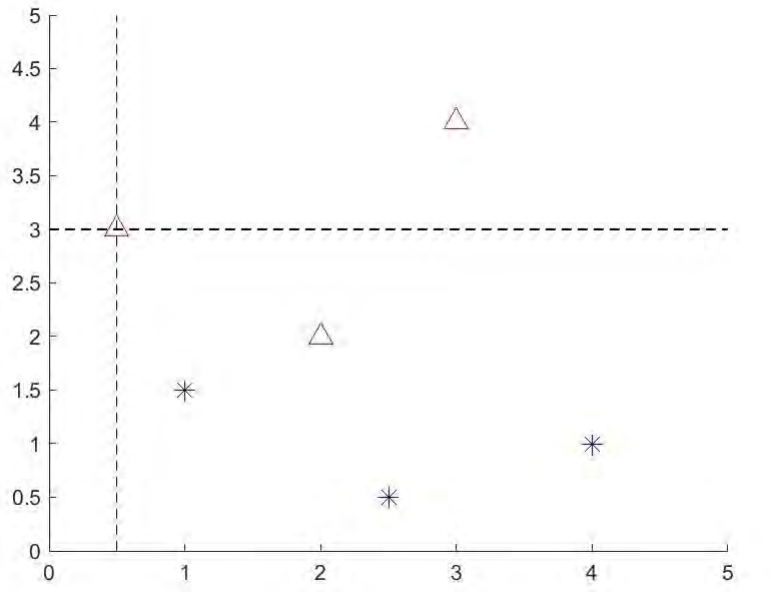


Figure 21: 2-D KS Test Algorithm - 1st Red Triangle

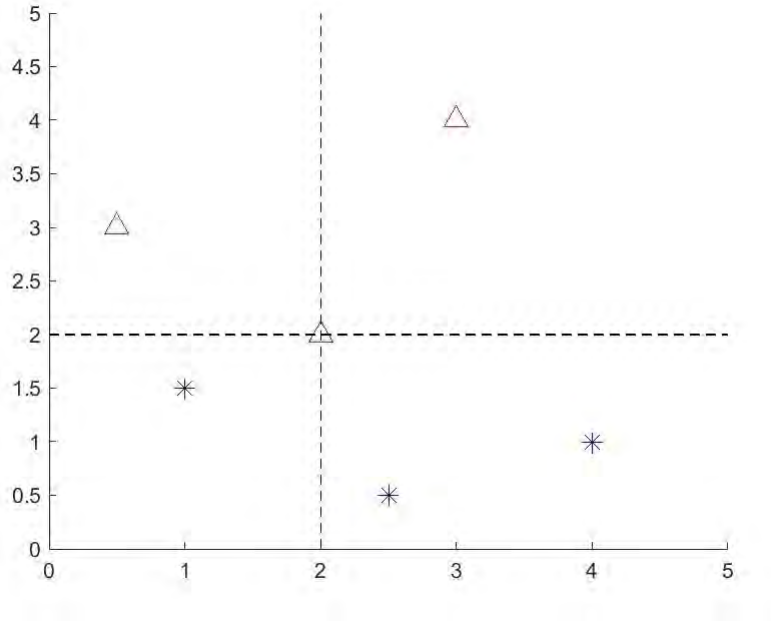


Figure 22: 2-D KS Test Algorithm - 2nd Red Triangle

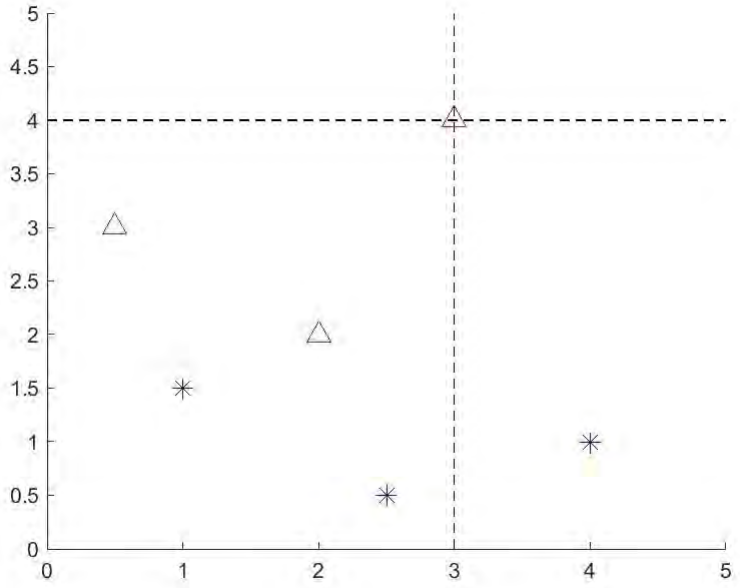


Figure 23: 2-D KS Test Algorithm - 3rd Red Triangle

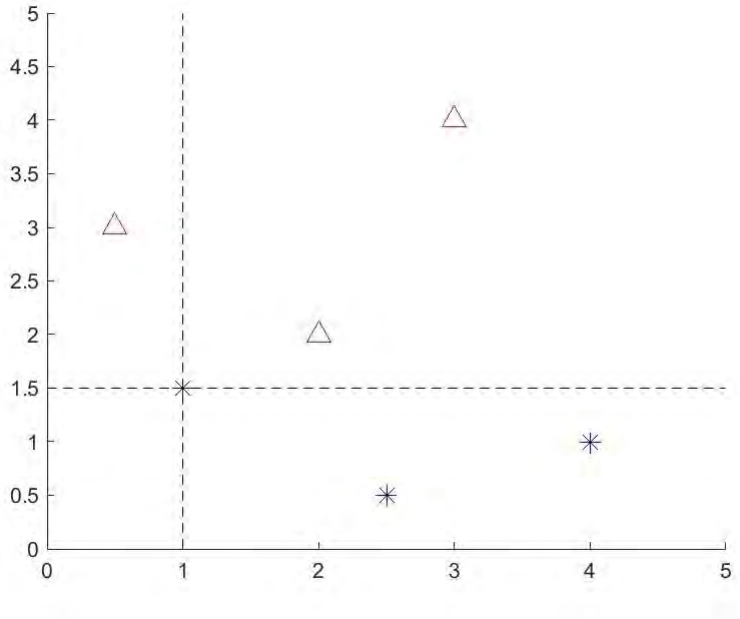


Figure 24: 2-D KS Test Algorithm - 1st Blue Star

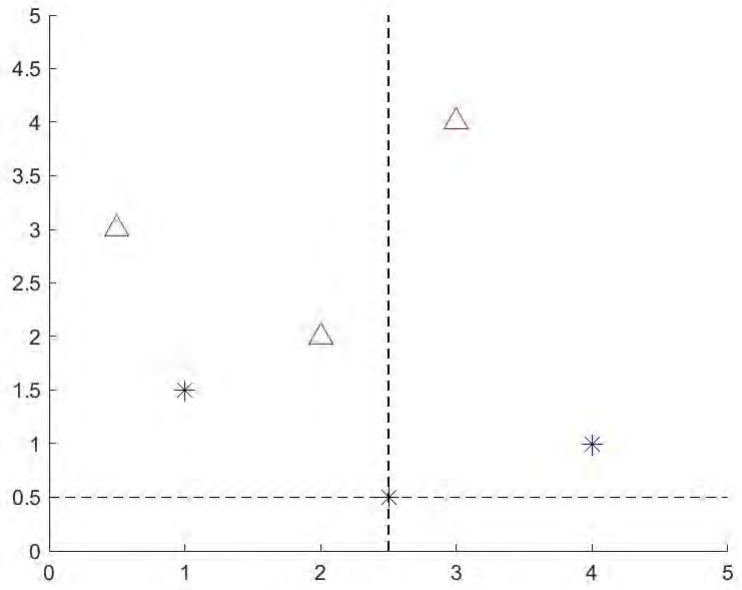


Figure 25: 2-D KS Test Algorithm - 2nd Blue Star

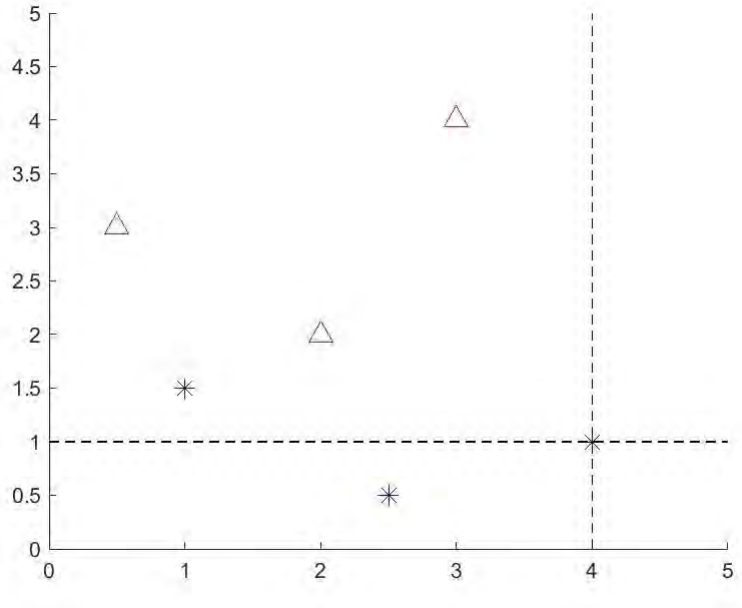


Figure 26: 2-D KS Test Algorithm - 3rd Blue Star

	***t*-Test**	**KS Test (1D)**	**KS Test (2+D)**	**Discrete JSD**	**QDA**	**MDA-ML**
Assumes normal distribution	Yes	No	No	No	Yes	No
# of characteristics simultaneously considered	1	1	2+	2+	2+	2+
Adjustable parameters	α	α	α	b, c	None	Currently unknown
Function	Compares two distributions based on mean, standard deviation, and number of data points	Measures the greatest distance between two cumulative distribution functions	Measures the greatest distance between the integrated probability in each of four natural quadrants around a given point	Measures the distinguishability between two probability distributions	Creates a quadratic classifier model by estimating the parameters of a Gaussian for each class; predicts classes of new data	Reduces dimensionality of data to find best class separation; determines classification boundaries

Table 3: Comparison of Various Methods for Qubit Discrimination

	Strengths	Weaknesses
t-Test	Computationally inexpensive	Assumes normal distribution
	Performs same/not-same decisions	Considers 1 PUC characteristic
1-D KS-Test	Does not assume normal distribution	Considers 1 PUC characteristic
	Performs same/not-same decisions	
2+-D KS-Test	Considers 2+ PUC characteristics simultaneously	
	Does not assume normal distribution	
	Performs same/not-same decisions	
Discrete JSD	Does not assume normal distribution	Considers 1 PUC characteristic
	Multiple adjustable parameters	
	Performs same/not-same decisions	
QDA	Considers 2+ PUC characteristics simultaneously	Assumes normal distribution
		Performs classification decisions
MDA-ML	Considers 2+ PUC characteristics simultaneously	Performs classification decisions
	Does not assume normal distribution	

Table 4: Strengths and Weaknesses for the Qubit Discrimination Problem

3.3.3.2 Calculating PD and PFA

Each of the qubit discrimination methods mentioned above is evaluated with the metrics of PD and PFA. PD and PFA are calculated for each of the discrimination methods, for each quantum processor characteristic, and for each data set size. Manufacturer qubits are compared against different customer qubits to calculate PD, and manufacturer qubits are compared against the same customer qubits to calculate PFA

(Figure 27). Each comparison produces a value of 1 (that is, the qubits are different) or a value of 0 (that is, the qubits are the same). PD and PFA are calculated by: $\frac{\#of1s}{\#ofcomparisons}$, where PD=1 and PFA=0 are desired.

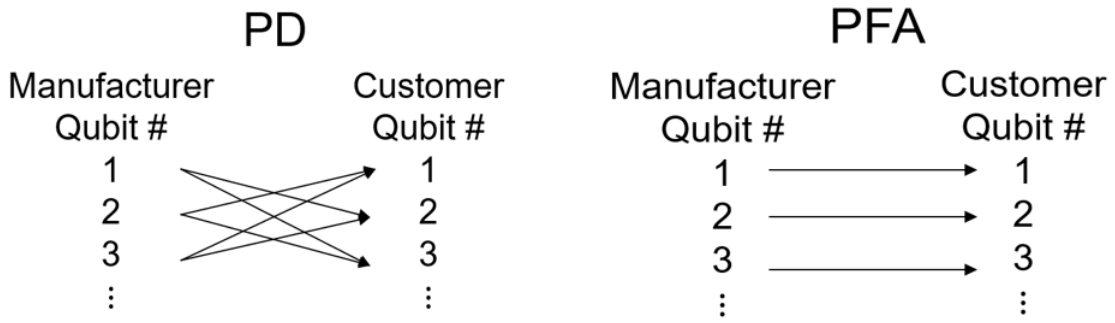


Figure 27: Qubit comparison methods for calculating PD and PFA

3.4 Summary

This chapter discusses the methodology for evaluating relationships between a subset of hardware verification/security techniques and a subset of quantum computing technologies in Part A. For Part B, this chapter describes the processes used for determining the best-fit distribution types for PUC characteristics, reducing sampling errors, and testing qubit discrimination methods.

IV. Results and Analysis

4.1 Overview

This chapter presents the results of Parts A and B as introduced in Chapter I. The results are organized into subsections, where each subsection corresponds to the identified parts of this research.

4.2 Part A – Determine Potential Applications

4.2.1 Part A1 – Survey State-of-the-Art Classical Hardware Verification and Security Techniques

Survey results for Part A1 are reported in Sections 2.1.1 and 2.1.2.

4.2.2 Part A2 – Survey State-of-the-Art Quantum Computing Technologies

Survey results for Part A2 are reported in Section 2.1.3.

4.2.3 Part A3 – Evaluate the Relationships Between a Subset of Hardware Verification/Security Techniques and a Subset of Quantum Computing Technologies

The relationships between classical hardware verification or security methods and the quantum computing technologies are evaluated using the four metrics proposed in Section 3.2.3. For Part A, the results should be described as preliminary, as further investigation into relationships identified as "DNE" may reveal underlying concepts inspired by the classical methods that are applicable to quantum processors.

4.2.3.1 Silicon Spin Qubit Results

The relationships between the subset of classical verification methods and the silicon spin qubit are evaluated in Table 5, and the relationships between classical security methods and the silicon spin qubit are evaluated in Table 6. Since the architecture for the silicon spin qubit implementation is divided into a classical layer and a qubit layer, relationships are evaluated only for the qubit layer. Also, since the relationships identified by "E" and "NR" require no change from classical relationships, they are excluded from discussion below.

- Gate-/Transistor-Level Analysis – In conventional circuits, operations are performed by sending an electrical signal through a sequence of logic gates. Gate-/transistor-level analysis verifies the connectivity of these logic gates. However, logic gates do not exist in the qubit layer. Thus, the object of verification does not exist.

- Functional Verification – For classical computers, the object of verification is the deterministic set of outputs for a given sequence of inputs. Since quantum computers are non-deterministic by nature, applying the functional verification technique requires modifications. In order to conduct functional verification for quantum computers, it is necessary to perform the same functional test multiple times to determine if the output probability distribution matches the theoretical expectation.

- Path Delay Fault Test – For classical components, the object of verification is the timing of a signal traveling through a CMOS logic path. Since CMOS logic gates do not exist in the qubit layer, the object of verification does not exist.

- PUF implementations – The object of verification in a classical component is the unique identity of a chip based on fabrication variations. While the unique

identity is an output from a set of multiplexers in a classical arbiter PUF, a different method of determining a unique identity is needed for application to transmon qubits.

- Hardware Obfuscation – For passive obfuscation, renaming signals is possible for the qubit layer, but no additional security would be achieved because this type of hardware obfuscation conceals the logic functions of combinatorial logic paths. For active obfuscation, the object of security does not exist because there are no combinatorial logic paths in the qubit layer. Reconfigurable logic-based obfuscation is inherent to the qubit layer as each gate operation reconfigures the logical design.

- Optical Dissimulation – The purpose of optical dissimulation is to conceal the logic functions of combinatorial logic paths through the use of dummy wires/interconnects/gates, scrambled memories, and custom cells. While it is possible to add dummy wires to the qubit layer, no additional security would be achieved because there are no combinatorial logic paths.

- Self-Destruction – While the concept of self-destruction fully applies to the qubit layer, the heating element may fail when contained in a near-absolute-zero refrigeration chamber. The method of self-destruction may require modification for the silicon spin processor. Additionally, the qubits are usually shielded from environmental perturbations, so the triggering mechanism may need modification to enable remote activation.

- Uniform timing for all operations – Gate operations for silicon spin qubits depend on the duration, pulse shape, amplitude, and frequency of the control microwave pulse. More investigation is required to determine if all gate operations can function with uniform timing.

- Adding noise to the power line – Because qubits are sensitive to noise, additional measures must be taken either to shield the qubit layer from the additional noise or to reduce the sensitivity of the qubits to the noise.

- Uniform power consumption – Gate operations for silicon spin qubits depend on the duration, pulse shape, amplitude, and frequency of the control microwave pulse. More investigation is required to determine if all gate operations can function with uniform power consumption.

- Transistor- and gate-level secure logic – Transistor- and gate-level secure logic techniques are not applicable as there are no transistors or CMOS gates in the qubit layer.

- Dynamic reconfiguration – While look-up tables are not part of the qubit layer, the concept of dynamic reconfiguration is inherent to the silicon spin processor, as different quantum circuits can operate at the same location at different times.

- Temporal desynchronization – Random process interrupts cannot be implemented in hardware for a silicon spin processor, but rather in software. However, the delays must be much shorter than decoherence times to avoid corrupting the calculations.

- Spatial jamming – This technique is not applicable to the silicon spin qubit because multiplexers do not exist in the qubit layer, and bits are not stored in the processor.

		Destructive (D) / Non-Destructive (ND)	Trusted Manufacturer Required (TMR) / Not Required (TMNR)	Object of Verification Exists (E) / Does Not Exist (DNE)	Mods. Required (R) / Not Required (NR)
Incoming Inspection	Low Power Visual Inspection	ND	TMR	E	NR
Exterior Tests	Scanning Electron Microscopy	ND	TMNR	E	NR
Interior Tests	Scanning Electron Microscopy	ND	TMNR	E	NR
Material Analysis	X-Ray Fluorescence (XRF)	ND	TMNR	E	NR
Gate-/transistor-level analysis	Gate-/transistor-level analysis	D	TMNR	DNE	–
Misc.	Temperature Cycling	D	TMNR	E	NR
Parametric Tests	DC Power Consumption Test	ND	TMNR	E	NR
Functional Tests	Functional Verification	ND	TMNR	E	R
Structural Tests	Path Delay Fault Test	ND	TMNR	DNE	–
Chip ID	PUF implementations	ND	TMR	E	R

Table 5: Relationships Between Hardware Verification Techniques and the Silicon Spin Qubit (Qubit Layer Only) (reproduced from [93])

		Destructive (D) / Non-Destructive (ND)	Trusted Manufacturer Required (TMR) / Not Required (TMNR)	Object of Verification Exists (E) / Does Not Exist (DNE)	Mods. Required (R) / Not Required (NR)
Hardware Obfuscation	Passive	ND	TMNR	DNE	–
Hardware Obfuscation	Active	ND	TMNR	DNE	–
Hardware Obfuscation	Reconfig.	ND	TMNR	E	NR
Optical Dissimulation		ND	TMNR	DNE	–
Self-Destruction		D	TMNR	E	R
Uniform timing for all operations		ND	TMNR	E	R
Adding noise to the power line		ND	TMNR	E	R
Power line filtering		ND	TMNR	E	NR
Power supply shielding		ND	TMNR	E	NR
Uniform power consumption		ND	TMNR	E	R
Transistor and gate-level secure logic		ND	TMNR	DNE	–
Dynamic reconfiguration		ND	TMNR	E	R
Temporal desynchronization		ND	TMNR	E	R
Spatial jamming		ND	TMNR	DNE	–

Table 6: Relationships Between Hardware Security Techniques and the Silicon Spin Qubit (Qubit Layer Only) (reproduced from [93])

4.2.3.2 Transmon Qubit Results

The relationships between classical verification methods and the transmon are listed in Table 7, and the relationships between classical security methods and the transmon are listed in Table 8. Since the architecture for the transmon is divided into a transmon processor and external classical control circuitry, relationships are only evaluated for the transmon processor (that is, the circuitry shown in Figure 10). It is assumed that any references to a "power line" refer to gate voltage V_g in Figure 10.

- Gate-/Transistor-Level Analysis – In conventional circuits, operations are performed by sending an electrical signal through a sequence of logic gates. Gate-/transistor-level analysis verifies the connectivity of these logic gates. However, logic gates do not exist in the transmon processor. Thus, the object of verification does not exist.

- Functional Verification – For classical computers, the object of verification is the deterministic set of outputs for a given sequence of inputs. Since quantum computers are non-deterministic by nature, modifications are required to apply the functional verification technique. In order to conduct functional verification for quantum computers, it is necessary to perform the same functional test multiple times to determine if the output probability distribution matches the theoretical expectation.

- Path Delay Fault Test – For classical components, the object of verification is the timing of a signal traveling through a CMOS logic path. Since CMOS logic gates do not exist in the transmon processor, the object of verification does not exist.

- PUF Implementations – The object of verification in a classical component is the unique identity of a chip based on fabrication variations. While the unique

identity is an output from a set of multiplexers in a classical arbiter PUF, a different method of determining a unique identity is necessary for application to transmon qubits.

- RF-DNA – RF-DNA for transmon processors can be implemented in an ideal and a non-ideal case. In the ideal case, one could collect a burst of active RF emission from the readout signal of a qubit (similar to RF-DNA for classical emitters). Then, the process of RF-DNA for transmon processors would follow the exact same process as for classical components. However, information about the measured frequency and phase of the readout signal is not available to users through the IBM Quantum Experience. Only information regarding the amplitude (pulse envelope) of the readout signal is made available to the user [118]. Thus, modifications are needed to implement RF-DNA for transmon processors in the non-ideal case. The following process is proposed for the non-ideal case:

  - Step A) Run S shots of an experiment, where each qubit is initialized to the ground state and the results of an experiment are Level 2 measurements of each qubit in the processor (that is, a 0 or a 1 per qubit) [118]. Each shot is analogous to a subregion, and each collection of S shots is analogous to a region. For example, in a two-qubit machine, each shot will result in a 00, 01, 10, or 11. Each region will have a distribution showing how many times a 00, 01, 10, or 11 was measured over the S shots. Thus, in each region, the mean, standard deviation, skewness, and kurtosis can be calculated for the distribution of the results from the S shots.
  - Step B) Repeat Step A R times so that R regions are obtained. The collection of R regions is analogous to a burst.
  - Step C) Repeat Step B B times so that B bursts are obtained.

- Step D) Generate burst fingerprints, and combine the burst fingerprints into a composite fingerprint for use in device classification with MDA-ML.

In the non-ideal case, the composite fingerprint is generated per device rather than per qubit per device.

Alternately, one could potentially collect RF emissions by placing a probe near the transmon processor to collect emissions from a single location (similar to RF-DNA for integrated circuits [58]) or by scanning the probe across the device to collect emissions from multiple locations [129]. Fingerprint generation and device classification would proceed as previously described. Also, scanning the probe across the device results in RF heat maps where the X and Y axes represent X and Y position, and the choice of color represents the power level. In this instance, image processing can potentially serve as another method of device discrimination. Before these forms of RF-DNA can be presented as viable options, however, the performance and spatial resolution of a probe inside a dilution refrigerator and the effect of a probe on transmon processor performance require investigation.

In both the ideal and non-ideal cases, additional vectors can be added to the fingerprint. Bursts can be collected for different types of experiments (for example, no gate on each qubit, Hadamard gate on each qubit, bit-flip gate on each qubit). Also, measurements can be obtained from qubits in higher energy states [6].

- Hardware Obfuscation – For passive obfuscation, renaming signals is possible for the transmon processor, but no additional security would be achieved since this type of hardware obfuscation conceals the logic functions of combinatorial logic paths. For active obfuscation, the object of security does not exist

because there are no combinatorial logic paths in the transmon processor. Reconfigurable logic-based obfuscation is inherent to the transmon processor as each gate operation reconfigures the logical design.

- Optical Dissimulation – The purpose of optical dissimulation is to conceal the logic functions of combinatorial logic paths through the use of dummy wires/interconnects/gates, scrambled memories, and custom cells. While it is possible to add dummy wires to the transmon processor, no additional security would be achieved because there are no combinatorial logic paths.

- Self-Destruction – While the concept of self-destruction fully applies to the transmon processor, the heating element may fail when contained in a near-absolute-zero refrigeration chamber. The method of self-destruction may require modification for the transmon processor. Additionally, the transmon is usually shielded from environmental perturbations, so the triggering mechanism may need modification to enable remote activation.

- Uniform timing for all operations – Gate operations for transmon qubits depend on the duration, pulse shape, amplitude, and frequency of the control microwave pulse. More investigation is required to determine if all gate operations can function with uniform timing.

- Adding noise to the power line – Since qubits are sensitive to noise, additional measures must be taken either to shield the transmon from the additional noise or reduce the sensitivity of the transmon to the noise.

- Uniform power consumption – Gate operations for transmon qubits depend on the duration, pulse shape, amplitude, and frequency of the control microwave pulse. More investigation is required to determine if all gate operations can function with uniform power consumption.

- Transistor- and gate-level secure logic – Transistor- and gate-level secure logic techniques are not applicable as transistors and CMOS gates do not exist in the transmon processor.

- Dynamic reconfiguration – While look-up tables are not part of a transmon processor, the concept of dynamic reconfiguration is inherent to the transmon processor, as different quantum circuits can operate at the same location at different times.

- Temporal desynchronization – Random process interrupts cannot be implemented in hardware for a transmon processor, but rather in software. However, the delays must be much shorter than decoherence times to avoid corrupting the calculations.

- Spatial jamming – This technique is not applicable to the transmon qubit because multiplexers do not exist in the processor and bits are not stored in the transmon processor.

		Destructive (D) / Non-Destructive (ND)	Trusted Manufacturer Required (TMR) / Not Required (TMNR)	Object of Verification Exists (E) / Does Not Exist (DNE)	Mods. Required (R) / Not Required (NR)
Incoming Inspection	Low Power Visual Inspection	ND	TMR	E	NR
Exterior Tests	SEM	ND	TMNR	E	NR
Interior Tests	SEM	ND	TMNR	E	NR
Material Analysis	XRF	ND	TMNR	E	NR
Connectivity analysis	Gate-/Transistor-Level Analysis	D	TMNR	DNE	–
Miscellaneous	Temperature Cycling	D	TMNR	E	NR
Parametric Tests	DC Power Consumption Test	ND	TMNR	E	NR
Functional Tests	Functional Verification	ND	TMNR	E	R
Structural Tests	Path Delay Fault Test	ND	TMNR	DNE	–
Chip ID	PUF implementations	ND	TMR	E	R
Chip ID	RF-DNA	ND	TMR	E	NR, R

Table 7: Relationships Between Hardware Verification Techniques and the Transmon

		Destructive (D) / Non-Destructive (ND)	Trusted Manufacturer Required (TMR) / Not Required (TMNR)	Object of Verification Exists (E) / Does Not Exist (DNE)	Mods. Required (R) / Not Required (NR)
Hardware Obfuscation	Passive	ND	TMNR	DNE	–
Hardware Obfuscation	Active	ND	TMNR	DNE	–
Hardware Obfuscation	Reconfig.	ND	TMNR	E	NR
Optical Dissimulation		ND	TMNR	DNE	–
Self-Destruction		D	TMNR	E	R
Uniform timing for all operations		ND	TMNR	E	R
Adding noise to the power line		ND	TMNR	E	R
Power line filtering		ND	TMNR	E	NR
Power supply shielding		ND	TMNR	E	NR
Uniform power consumption		ND	TMNR	E	R
Transistor and gate-level secure logic		ND	TMNR	DNE	–
Dynamic reconfiguration		ND	TMNR	E	R
Temporal desynchronization		ND	TMNR	E	R
Spatial jamming		ND	TMNR	DNE	–

Table 8: Relationships Between Hardware Security Techniques and the Transmon

4.3 Part B – Develop Physically Unclonable Characteristics (PUCs)

4.3.1 Part B1 – Determine Best-Fit Distribution Types for PUC Characteristics

4.3.1.1 Part B1.1 – Collect Data

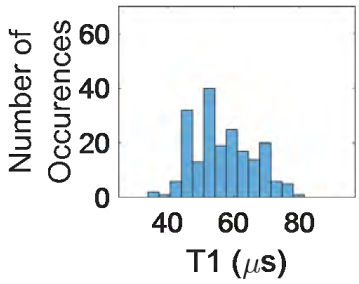
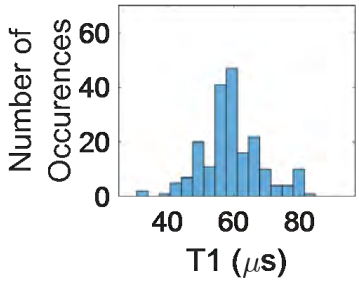
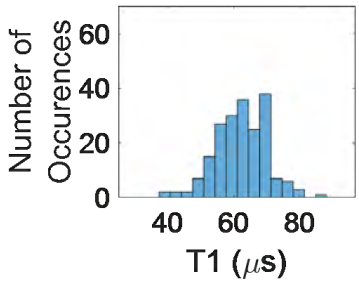
As of the time of writing this dissertation, characteristic data has been collected from the following IBM transmon processors since 4 December 2018:[1]

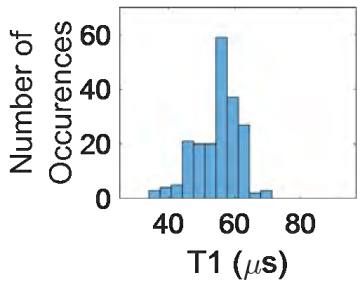
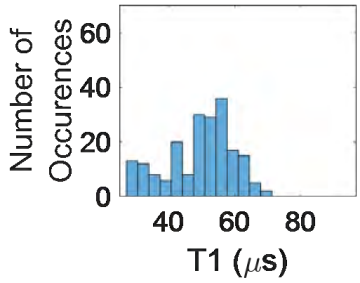
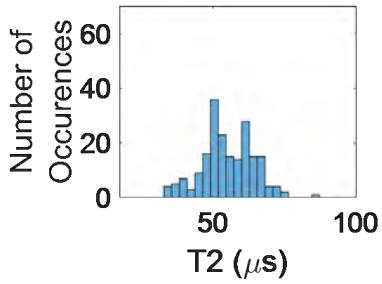
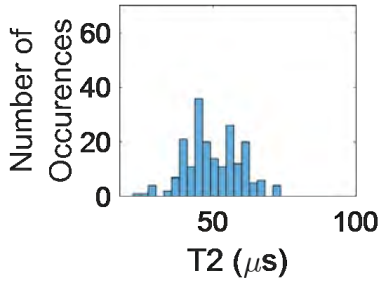
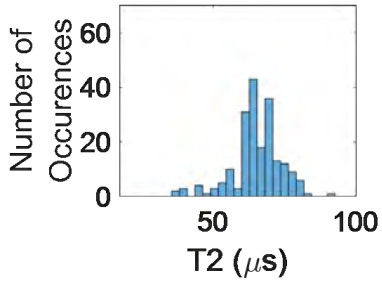
- ibmq_almaden (20 qubits)
- ibmq_armonk (1)
- ibmq_athens (5)
- ibmq_belem (5)
- ibmq_boeblingen (20)
- ibmq_bogota (5)
- ibmq_brooklyn (65)
- ibmq_burlington (5)
- ibmq_cambridge (28)
- ibmq_casablanca (7)
- ibmq_dublin (27)
- ibmq_essex (5)
- ibmq_guadalupe (16)
- ibmq_jakarta (7)
- ibmq_kolkata (27)
- ibmq_lima (5)
- ibmq_london (5)
- ibmq_manhattan (65)
- ibmq_manila (5)
- ibmq_melbourne (15)
- ibmq_montreal (27)
- ibmq_mumbai (27)
- ibmq_ourense (5)
- ibmq_paris (27)
- ibmq_poughkeepsie (20)
- ibmq_quito (5)
- ibmq_rochester (53)
- ibmq_rome (5)
- ibmq_santiago (5)
- ibmq_sydney (27)
- ibmq_tenerife (5)
- ibmq_tokyo (20)
- ibmq_toronto (27)
- ibmq_valencia (5)
- ibmq_vigo (5)
- ibmq_yorktown (5).

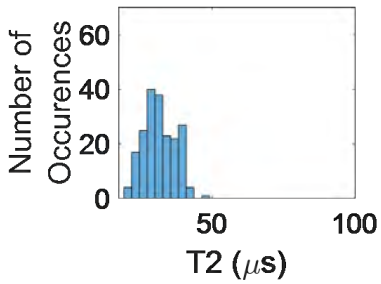
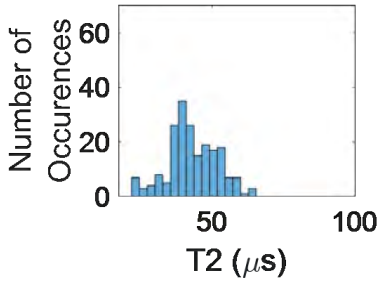
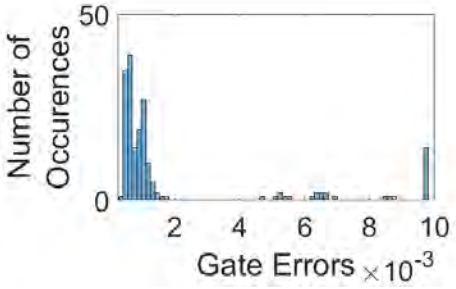
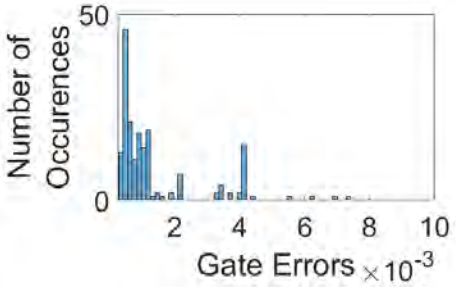
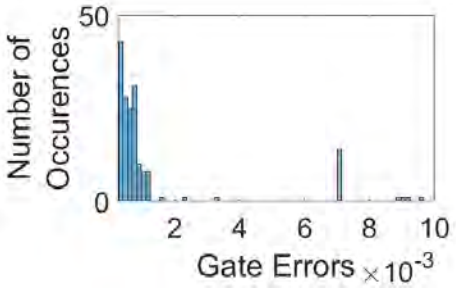
[1] Various processors have been commissioned or decommissioned since this date, so the collected data points for different processors span different segments of time.

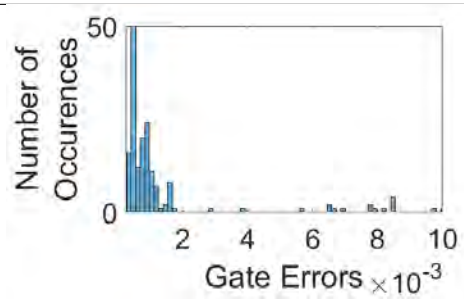
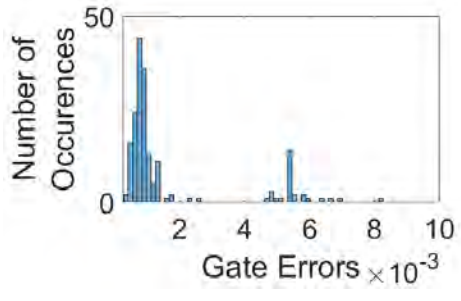
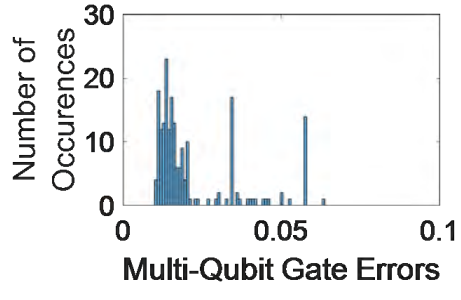
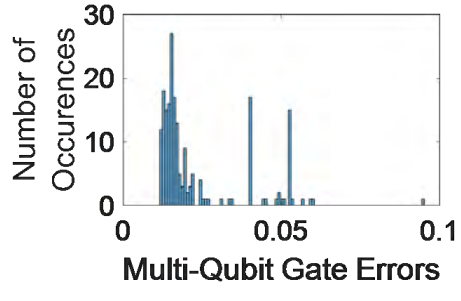
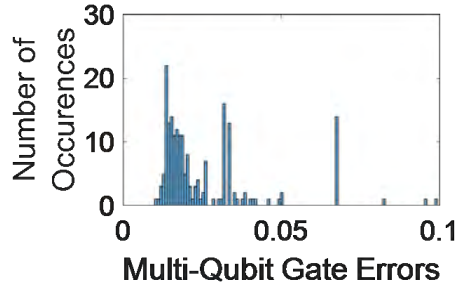
4.3.1.2 Part B1.2 – Characterize Difference Between Individual Qubits/Links

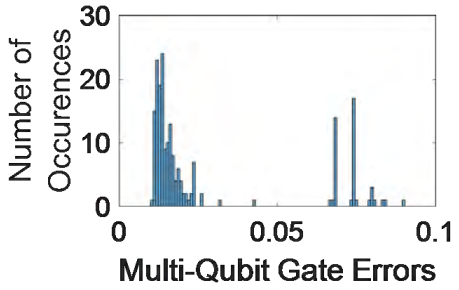
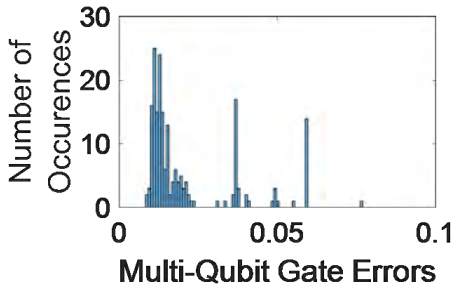
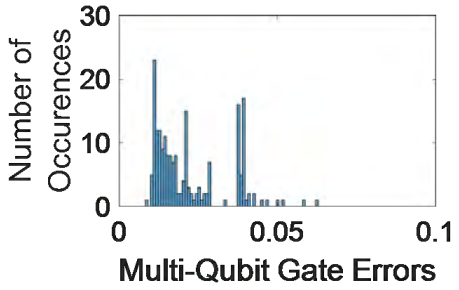
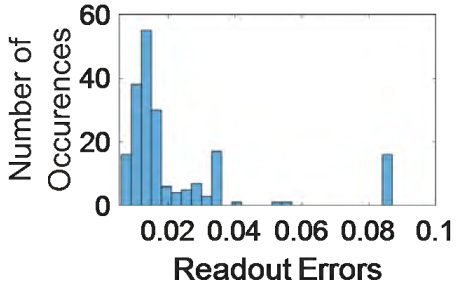
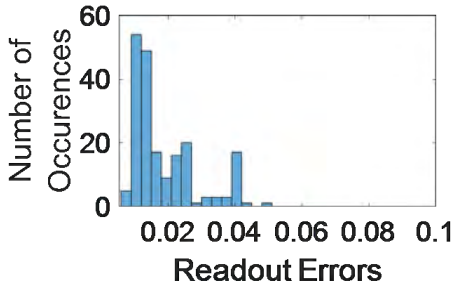
Table 9 presents histograms and corresponding means and standard deviations of each of the calibration characteristics for each ibmq_yorktown qubit. Table 10 presents each of the calibration characteristics for ibmq_yorktown qubits Q00-Q04 plotted in chronological order. By evaluating the histograms, the population parameters, and the chronological plots, it is easily observed that each qubit exhibits different behavior.

Qubit Number	Histogram	Mean	Standard Deviation
T1			
Q00		56.56	9.28
Q01		59.10	9.00
Q02		62.26	7.88

Q03		54.75	6.78
Q04		49.23	10.31
T2			
Q00		55.05	9.05
Q01		50.21	9.46
Q02		65.27	8.45

Q03		31.56	5.45
Q04		43.15	8.77
Single-Qubit Gate Error (Outliers above 10e-3 not shown)			
Q00		0.1610	0.5422
Q01		.1605	.5423
Q02		.1605	.5421

Q03		.1611	.5422
Q04		.1607	.5423
Multi-Qubit Gate Errors (Outliers Above 0.1 Not Shown)			
Q00-Q01		.0268	.0703
Q00-Q02		.0288	.0707
Q01-Q02		.0307	.0707

Q02-Q03		.0387	.1002
Q02-Q04		.0269	.0709
Q03-Q04		.0273	.0699
Readout Errors (Outliers above 0.1 not shown)			
Q00		.0246	.0358
Q1		.0213	.0343

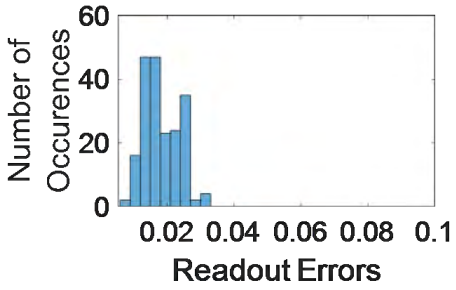
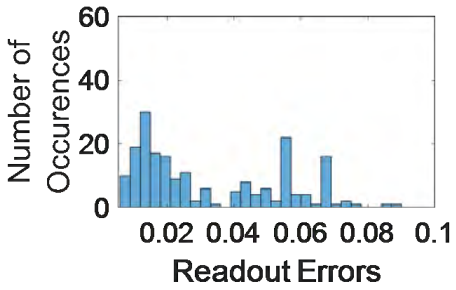
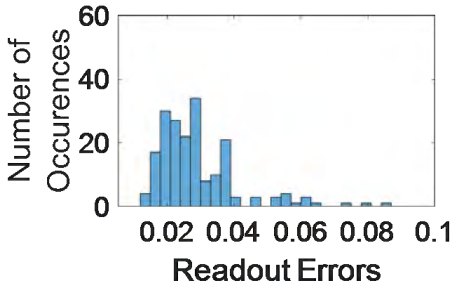
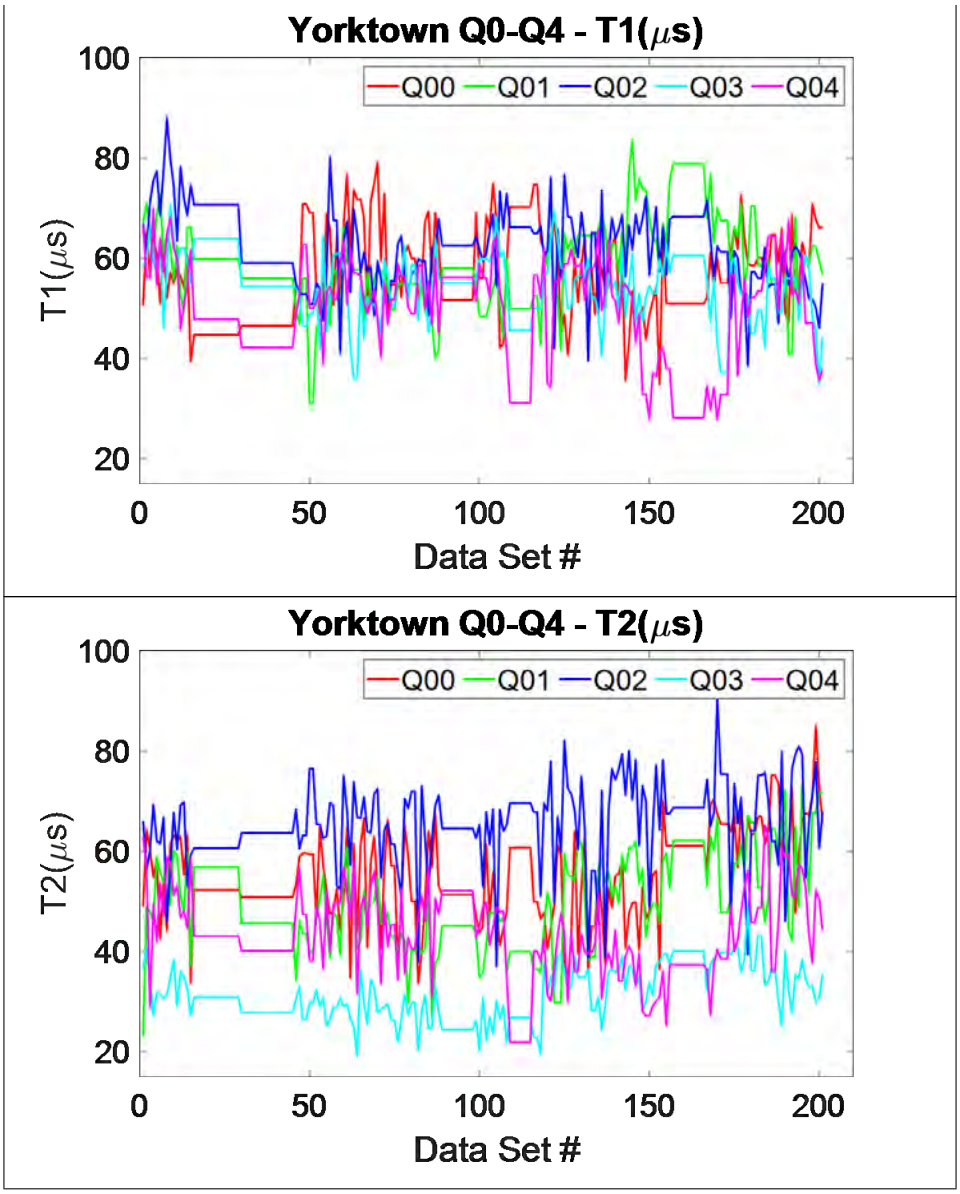
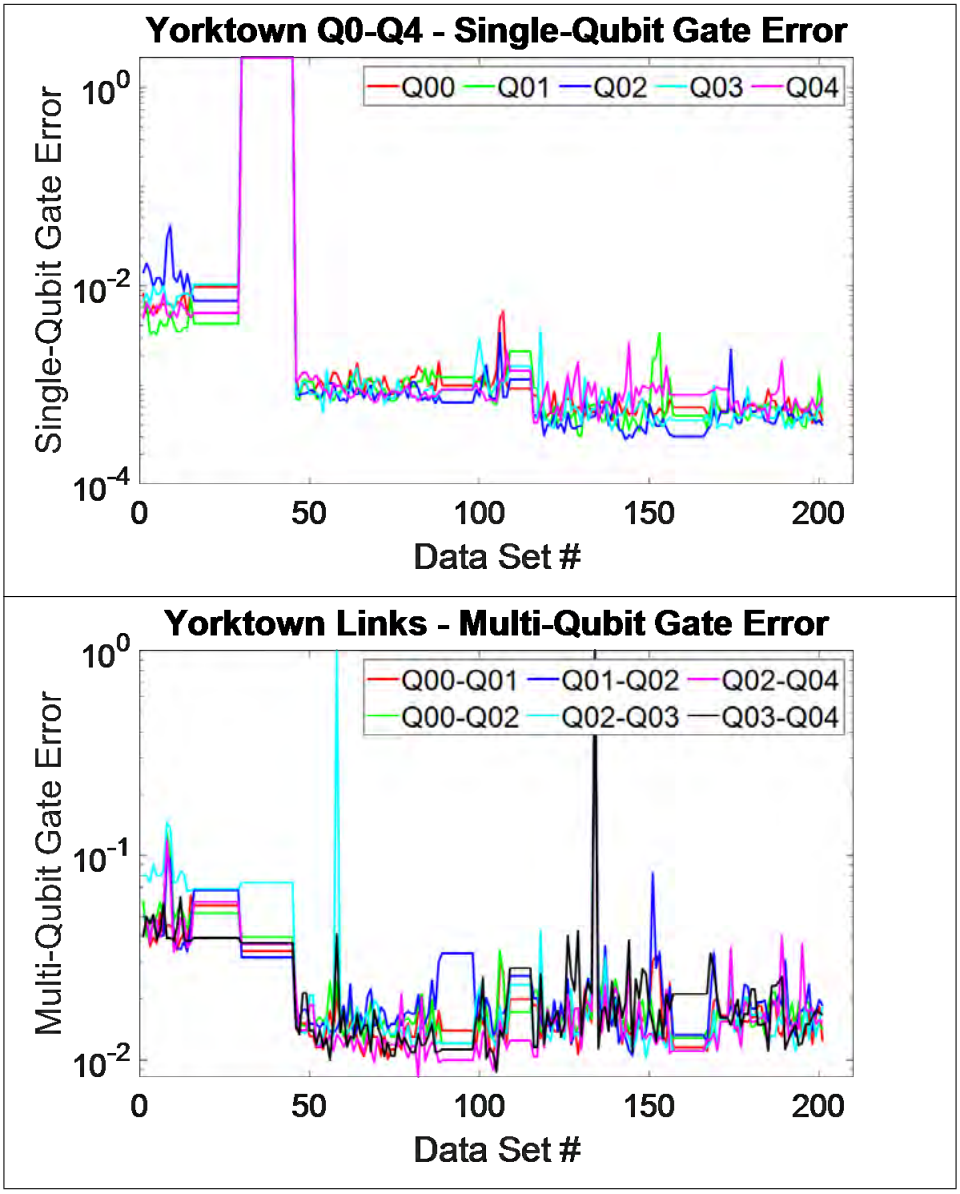
Q2		.0206	.0326
Q3		.0358	.0412
Q4		.0316	.0184

Table 9: Histograms and Corresponding Means and Standard Deviations of Each Calibration Characteristic for ibmq_yorktown Qubits (N=201)


Yorktown Q0-Q4 - T1(μs)
Q00 Q01 Q02 Q03 Q04
T1(μs)
Data Set #
Yorktown Q0-Q4 - T2(μs)
Q00 Q01 Q02 Q03 Q04
T2(μs)
Data Set #


Yorktown Q0-Q4 - Single-Qubit Gate Error
Q00 Q01 Q02 Q03 Q04
Single-Qubit Gate Error
10^{0}
10^{-2}
10^{-4}
0 50 100 150 200
Data Set #
Yorktown Links - Multi-Qubit Gate Error
Q00-Q01 Q01-Q02 Q02-Q04
Q00-Q02 Q02-Q03 Q03-Q04
Multi-Qubit Gate Error
10^{0}
10^{-1}
10^{-2}
0 50 100 150 200
Data Set #

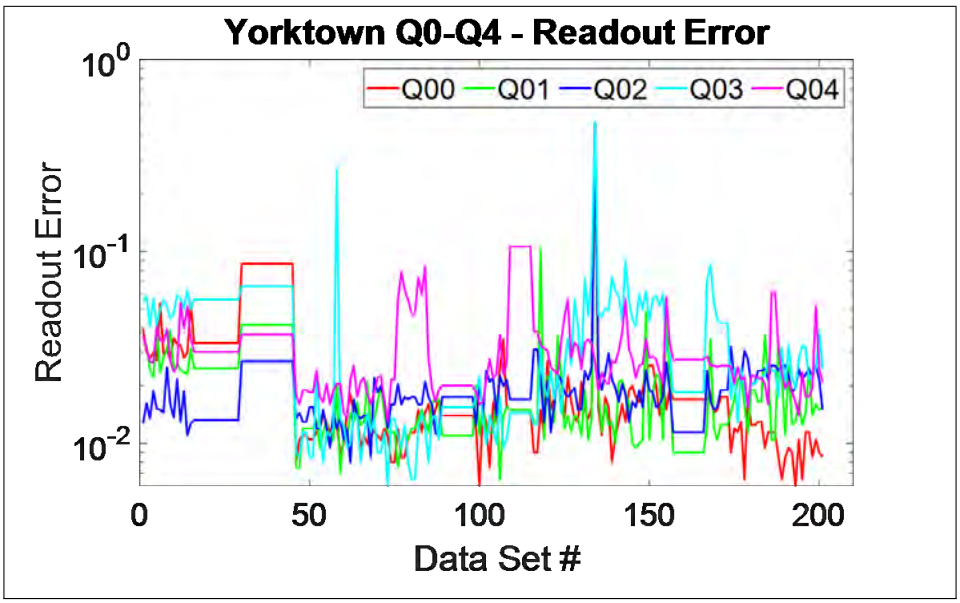


Table 10: Calibration Characteristic for ibmq_yorktown Qubits Plotted in Chronological Order

4.3.1.3 Part B1.3 – Calculate P-values and the Negative Log-Likelihood (NLL), Akaike's Information Criterion (AIC), Corrected AIC (AICc), and Bayesian Information Criterion (BIC) Metrics

In this section, p-values and the negative log-likelihood (NLL), Akaike's information criterion (AIC), corrected AIC (AICc), and Bayesian information criterion (BIC) metrics are determined for the calibration, quantum process tomography, and randomized benchmarking characteristics. It should be noted that data conditioning is not discussed until Part B2, but all metrics in this section are calculated for characteristics with data set truncation and with multi-qubit gate error rates equal to one removed from the data set.

The main purpose of this section is to determine whether or not the characteristics are normally distributed, as this determination influences the selection of qubit

discrimination methods. However, knowledge about the specific best-fit distribution types can provide interesting insight into the behavior of the quantum device. As such, an initial analysis is performed for each characteristic with the intent of understanding the results and/or identifying areas that require additional research and analysis.

P-Values P-value results for the Chi-square goodness-of-fit test and Kolmogorov-Smirnov (KS) test are shown in Table 11 for the calibration data, Table 12 for the quantum process tomography data, and Tables 13 through 24 for randomized benchmarking data. Cells highlighted in red represent distribution types rejected at the significance level of $\alpha = 0.05$. Cells that contain only a dash indicate that the Chi-square test, which requires data binning, failed to produce a result.

In Table 11, the exponential, generalized Pareto, half-normal, and Rayleigh distribution types are rejected for all calibration data characteristics. For the T1 coherence times, the Chi-square test additionally rejects the Birnbaum Saunders, inverse Gaussian, and lognormal distributions. For the T2 coherence times, the Chi-square test rejects the Birnbaum Saunders, gamma, inverse Gaussian, loglogistic, lognormal, and Nakagami distributions. For the single-qubit and multi-qubit gate error rate characteristics, the Chi-square test rejects the gamma, logistic, and t location-scale distributions, and both the Chi-square and KS tests reject the Nakagami, normal, Rician, and Weibull distributions. For readout error rates, the Chi-square test rejects the gamma, Nakagami, normal, Rician, and Weibull distributions. Results for the p-values are consistent with results for NLL, AIC, AICc, and BIC, since none of the distributions rejected by the p-values are selected by the NLL, AIC, AICc, and BIC, metrics.

	T1 Coherence Time		T2 Coherence Time		Single-Qubit Gate Error Rate		Multi-Qubit Gate Error Rate		Readout Error Rate	
	Chi2	KS	Chi2	KS	Chi2	KS	Chi2	KS	Chi2	KS
Birnbaum Saunders	0.032	0.275	0.003	0.094	0.050	0.084	0.051	0.120	0.055	0.204
Burr	0.254	0.849	0.097	0.743	0.151	0.784	0.097	0.771	0.159	0.780
Exponential	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Extreme Value	0.162	0.551	0.055	0.481	0.000	0.000	0.000	0.000	0.000	0.009
Gamma	0.092	0.419	0.011	0.218	0.027	0.053	0.030	0.072	0.040	0.157
Generalized Extreme Value	0.153	0.753	0.052	0.673	0.138	0.775	0.112	0.759	0.068	0.674
Generalized Pareto	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Half-Normal	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Inverse Gaussian	0.032	0.275	0.003	0.095	0.056	0.103	0.055	0.122	0.066	0.231
Logistic	0.239	0.780	0.147	0.756	0.037	0.130	0.041	0.157	0.094	0.378
Loglogistic	0.093	0.650	0.035	0.516	0.139	0.518	0.121	0.541	0.187	0.578
Lognormal	0.045	0.319	0.003	0.129	0.071	0.136	0.070	0.166	0.072	0.248
Nakagami	0.135	0.536	0.033	0.315	0.002	0.012	0.005	0.018	0.014	0.097
Normal	0.224	0.656	0.078	0.482	0.001	0.006	0.002	0.010	0.005	0.061
Rayleigh	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Rician	0.226	0.651	0.085	0.478	0.002	0.007	0.002	0.012	0.007	0.077
t Location-Scale	0.157	0.771	0.060	0.737	0.022	0.276	0.027	0.290	0.083	0.443
Weibull	0.280	0.694	0.127	0.561	0.001	0.013	0.000	0.009	0.004	0.088

Table 11: p-values for the Chi-Square Goodness-of-Fit Test and Kolmogorov-Smirnov Test for Each Calibration Data Characteristic. The red cells indicate distribution types rejected at the significance level of $\alpha = 0.05$.

Beginning with the quantum process tomography data and continuing for the randomized benchmarking data, p-value results for the Chi-square test become less reliable since there are fewer quantum process tomography (QPT) and randomized benchmarking (RB) data points compared to the calibration data. There are fewer data points for QPT and RB because a) data collection for these characteristics began later than the calibration data collection, and b) QPT and RB data points often take longer to collect, as explained in Section 3.3.1. Accordingly, the following discussion focuses on distribution types rejected by both the Chi-square test and the KS test.

In Table 12, for quantum process tomography on the X gate (QPTX), the Birnbaum Saunders, exponential, gamma, Generalized Pareto, half-normal, inverse Gaussian, lognormal, Nakagami, normal, Rayleigh, and Rician distributions are rejected by both the Chi-square and the KS tests. For quantum process tomography on the H gate (QPTH), only the inverse Gaussian distribution is rejected by both tests. It is possible others would also be rejected, but for several of the distribution types, there are too few data points for the Chi-square test to make a determination. For quantum process tomography on the CX gate (QPTCX) and the SWAP gate (QPTSWAP), the Birnbaum Saunders, exponential, gamma, half-normal, inverse Gaussian, lognormal, Nakagami, normal, Rayleigh, and Rician distributions are rejected by both tests. None of the distribution types are rejected by both tests for quantum process tomography on the CCX and CSWAP (QPTCCX and QPTCSWAP) gates, but the reason is likely because there are too few data points to make a determination with the Chi-square test in many cases.

	QPTX		QPTH		QPTCX		QPTSWAP		QPTCCX		QPTCSWAP^2	
	Chi2	KS	Chi2	KS	Chi2	KS	Chi2	KS	Chi2	KS	Chi2	KS
Birnbaum Saunders	0.001	0.020	0.007	0.056	0.016	0.000	0.004	0.015	–	0.011	–	0.794
Burr	0.337	0.865	0.172	0.903	0.051	0.104	0.059	0.328	–	0.835	–	–
Exponential	0.000	0.000	–	0.000	0.000	0.000	0.000	0.000	–	0.000	–	0.042
Extreme Value	0.077	0.337	0.069	0.327	0.138	0.183	0.116	0.205	0.175	0.536	–	0.825
Gamma	0.001	0.023	0.007	0.061	0.002	0.002	0.004	0.019	–	0.046	–	0.773
Generalized Extreme Value	0.031	0.271	0.065	0.480	0.019	0.333	0.006	0.396	–	0.531	–	0.777
Generalized Pareto	0.000	0.000	–	0.000	–	0.000	–	0.000	–	0.000	–	0.043
Half-Normal	0.000	0.000	–	0.000	0.000	0.000	0.000	0.000	–	0.000	–	0.028
Inverse Gaussian	0.000	0.000	0.000	0.005	0.019	0.005	0.005	0.017	–	0.010	–	0.770
Logistic	0.087	0.293	0.123	0.446	0.000	0.294	0.115	0.287	0.217	0.508	–	0.907
Loglogistic	0.072	0.255	0.116	0.371	0.184	0.237	0.074	0.360	0.234	0.193	–	0.878
Lognormal	0.001	0.021	0.004	0.061	0.013	0.007	0.004	0.041	–	0.022	–	0.836
Nakagami	0.002	0.025	0.010	0.067	0.010	0.003	0.003	0.024	0.012	0.058	–	0.770
Normal	0.002	0.029	0.014	0.080	0.000	0.008	0.001	0.036	0.040	0.127	–	0.840
Rayleigh	0.000	0.000	–	0.000	0.000	0.000	0.000	0.000	–	0.000	–	0.094
Rician	0.002	0.029	0.015	0.074	0.000	0.008	0.002	0.037	0.048	0.120	–	0.795
t Location-Scale	0.071	0.460	0.072	0.448	0.033	0.476	0.237	0.568	–	0.715	–	0.847
Weibull	0.058	0.302	0.107	0.259	0.123	0.097	0.053	0.199	0.202	0.198	–	0.803

Table 12: p-values for the Chi-Square Goodness-of-Fit Test and Kolmogorov-Smirnov Test for Each Quantum Process Tomography Characteristic. The red cells indicate distribution types rejected at the significance level of $\alpha = 0.05$.

In Tables 13 and 14, the p-values for one-qubit randomized benchmarking metrics with one seed of data (Seed 0) are presented. For fidelity, the Birnbaum Saunders, exponential, generalized Pareto, half-normal, inverse Gaussian, lognormal, Nakagami, and Rayleigh distribution types are rejected. For the a parameter, the exponential, extreme value, generalized Pareto, half-normal, normal, and Rayleigh distributions are rejected. It should be noted that the generalized Pareto distribution is rejected by both tests, but it is selected as the best-fit distribution by the negative log-likelihood and all three information criteria (AIC, AICc, and BIC). This discrepancy is discussed in Section 4.3.1.3. For the α parameter, both tests reject the exponential, gamma, generalized Pareto, half-normal, lognormal, Nakagami, normal, Rayleigh, and Rician distributions. For the b parameter, both tests reject the exponential, gamma, generalized Pareto, half-normal, Nakagami, normal, and Rayleigh distributions. For the err_a parameter, both tests reject the exponential, extreme value, half-normal, logistic, Nakagami, normal, Rayleigh, Rician, and t location-scale distributions. For the err_α parameter, both tests reject the exponential, extreme value, generalized Pareto, half-normal, and Rayleigh distributions. For the err_b parameter, both tests reject the exponential, extreme value, half-normal, logistic, Nakagami, normal, Rayleigh, Rician, and t location-scale distributions.

After calculating the one-qubit randomized benchmarking metrics with four additional seeds of data, fewer distribution types are rejected. In Tables 15 and 16, the p-values for one-qubit randomized benchmarking metrics with five seeds of data (Seed 4) are presented. For fidelity, only the inverse Gaussian distribution is rejected by both tests. For the a parameter, both tests reject the exponential, generalized Pareto, half-normal, and Rayleigh distributions. With five seeds of data, the generalized extreme value is selected as the best fit distribution instead of the generalized Pareto distribution for one seed of data, so the p-value tests do not reject the best-fit

distribution type as was the case for one seed of data. For the α parameter, both tests reject the generalized Pareto and Rayleigh distributions. For the b parameter, both tests reject the Birnbaum Saunders, gamma, generalized Pareto, inverse Gaussian, lognormal, and Rayleigh distributions. For the err_a parameter, both tests reject the extreme value, half-normal, normal, Rayleigh, and Rician distributions. For the err_α parameter, both tests reject the exponential, generalized Pareto, half-normal, and Rayleigh distributions. For the err_b parameter, both tests reject the extreme value, half-normal, normal, Rayleigh, and Rician distributions.

	Fid.		a		α		b	
	Chi2	KS	Chi2	KS	Chi2	KS	Chi2	KS
Birnbaum Saunders	0.002	0.044	0.033	0.154	0.027	0.019	-	0.007
Burr	0.122	0.847	0.199	0.787	0.440	0.793	-	0.875
Exponential	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Extreme Value	0.106	0.422	0.000	0.013	0.113	0.318	0.027	0.170
Gamma	0.002	0.051	0.004	0.097	0.026	0.019	0.000	0.000
Generalized Extreme Value	0.011	0.312	0.052	0.391	0.055	0.391	0.011	0.199
Generalized Pareto	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000
Half-Normal	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Inverse Gaussian	0.000	0.000	0.042	0.162	-	0.000	-	0.003
Logistic	0.077	0.347	0.001	0.144	0.124	0.103	0.000	0.097
Loglogistic	0.071	0.327	0.059	0.349	0.132	0.102	0.000	0.350
Lognormal	0.002	0.043	0.030	0.185	0.030	0.017	-	0.007
Nakagami	0.003	0.057	0.001	0.053	0.026	0.020	0.000	0.018
Normal	0.006	0.063	0.000	0.028	0.028	0.018	0.000	0.017
Rayleigh	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000
Rician	0.007	0.065	0.000	0.053	0.026	0.020	0.000	0.321
t Location-Scale	0.038	0.460	0.000	0.072	0.062	0.176	0.000	0.064
Weibull	0.105	0.429	0.000	0.054	0.111	0.324	0.000	0.239

Table 13: p-values for the Chi-Square Goodness-of-Fit Test and Kolmogorov-Smirnov Test for Each 1-Qubit Randomized Benchmarking Characteristic, Seed 0, Part I. The red cells indicate distribution types rejected at the significance level of $\alpha = 0.05$.

	err_a		*err_α*		*err_b*	
	Chi2	KS	Chi2	KS	Chi2	KS
Birnbaum Saunders	0.021	0.360	0.260	0.499	0.015	0.246
Burr	0.014	0.690	0.201	0.849	0.013	0.683
Exponential	0.008	0.010	0.000	0.000	0.006	0.013
Extreme Value	0.000	0.000	0.000	0.004	0.000	0.000
Gamma	0.008	0.116	0.194	0.343	0.007	0.130
Generalized Extreme Value	0.024	0.663	0.184	0.804	0.026	0.713
Generalized Pareto	0.020	0.258	0.000	0.000	0.010	0.220
Half-Normal	0.000	0.000	0.000	0.000	0.000	0.000
Inverse Gaussian	0.067	0.359	0.260	0.513	0.031	0.215
Logistic	0.001	0.005	0.157	0.449	0.002	0.008
Loglogistic	0.028	0.507	0.329	0.791	0.011	0.489
Lognormal	0.022	0.434	0.292	0.528	0.013	0.404
Nakagami	0.002	0.029	0.069	0.150	0.003	0.017
Normal	0.000	0.001	0.038	0.096	0.000	0.000
Rayleigh	0.000	0.000	0.000	0.000	0.000	0.000
Rician	0.000	0.000	0.072	0.206	0.000	0.000
t Location-Scale	0.000	0.003	0.116	0.668	0.000	0.005
Weibull	0.011	0.165	0.011	0.083	0.009	0.181

Table 14: p-values for the Chi-Square Goodness-of-Fit Test and Kolmogorov-Smirnov Test for Each 1-Qubit Randomized Benchmarking Characteristic, Seed 0, Part II. The red cells indicate distribution types rejected at the significance level of $\alpha = 0.05$.

For two-qubit randomized benchmarking metrics, both tests reject fewer distribution types compared to the one-qubit randomized benchmarking metrics. There are fewer data points collected for the two-qubit randomized benchmarking metrics compared to the one-qubit randomized benchmarking metrics, which prevents the Chi-square test from making a determination for many distribution types.

In Tables 17 and 18, the p-values for two-qubit randomized benchmarking metrics with one seed of data (Seed 0) are presented. For fidelity, none of the distribution types are rejected by both metrics. For the a parameter, both tests reject the Rayleigh distribution type. For the α parameter, both tests reject the Birnbaum Saunders, gamma, lognormal, Nakagami, normal, and Rician distribution types. For the b parameter, none of the distribution types are rejected by both tests. For the *err_a* and

	Fid.		a		α		b	
	Chi2	KS	Chi2	KS	Chi2	KS	Chi2	KS
Birnbaum Saunders	0.012	0.076	0.076	0.303	0.045	0.200	0.002	0.021
Burr	0.000	0.486	0.486	0.901	0.292	0.883	0.005	0.937
Exponential	-	0.000	0.000	0.000	-	0.000	-	0.000
Extreme Value	0.159	0.299	0.299	0.058	0.120	0.594	0.063	0.524
Gamma	0.013	0.079	0.079	0.243	0.045	0.200	0.002	0.027
Generalized Extreme Value	0.073	0.468	0.468	0.801	0.067	0.440	0.011	0.584
Generalized Pareto	-	0.000	0.000	0.003	0.000	0.000	0.000	0.000
Half-Normal	-	0.000	0.000	0.000	-	0.000	-	0.000
Inverse Gaussian	0.000	0.000	0.000	0.304	-	0.000	0.002	0.021
Logistic	0.102	0.283	0.283	0.476	0.184	0.652	0.013	0.450
Loglogistic	0.103	0.268	0.268	0.688	0.189	0.652	0.022	0.273
Lognormal	0.011	0.074	0.074	0.323	0.048	0.194	0.002	0.031
Nakagami	0.013	0.082	0.082	0.198	0.045	0.201	0.008	0.073
Normal	0.013	0.083	0.083	0.164	0.032	0.195	0.008	0.154
Rayleigh	-	0.000	0.000	0.003	0.000	0.000	0.000	0.003
Rician	0.014	0.084	0.084	0.171	0.032	0.201	0.026	0.150
t Location-Scale	0.034	0.422	0.422	0.446	0.324	0.738	0.024	0.409
Weibull	0.155	0.292	0.292	0.148	0.133	0.594	0.105	0.290

Table 15: p-values for the Chi-Square Goodness-of-Fit Test and Kolmogorov-Smirnov Test for Each 1-Qubit Randomized Benchmarking Characteristic, Seed 4, Part I. The red cells indicate distribution types rejected at the significance level of $\alpha = 0.05$.

err_b parameters, both tests reject the exponential, generalized Pareto, half-normal, Rayleigh, and t location-scale distributions. For the *err_α* parameter, both tests reject the extreme value distribution in addition to the distribution types rejected for the *err_a* and *err_b* parameters.

As observed with the one-qubit randomized benchmarking metrics, there are fewer distribution types rejected after the two-qubit randomized benchmarking metrics are calculated with four additional seeds of data. In Tables 19 and 20, the p-values for two-qubit randomized benchmarking metrics with five seeds of data (Seed 4) are

	err_a		err_α		err_b	
	Chi2	KS	Chi2	KS	Chi2	KS
Birnbaum Saunders	0.168	0.389	0.264	0.687	0.094	0.367
Burr	0.499	0.917	0.284	0.921	0.458	0.868
Exponential	0.078	0.108	0.000	0.000	0.079	0.091
Extreme Value	0.000	0.011	0.002	0.168	0.000	0.011
Gamma	0.131	0.344	0.318	0.660	0.161	0.361
Generalized Extreme Value	0.389	0.883	0.157	0.908	0.380	0.854
Generalized Pareto	0.029	0.416	0.000	0.000	0.028	0.386
Half-Normal	0.010	0.014	0.000	0.000	0.013	0.015
Inverse Gaussian	0.174	0.485	0.265	0.686	0.169	0.341
Logistic	0.143	0.161	0.437	0.764	0.114	0.173
Loglogistic	0.289	0.864	0.414	0.866	0.240	0.813
Lognormal	0.214	0.661	0.252	0.700	0.163	0.609
Nakagami	0.024	0.124	0.407	0.583	0.025	0.141
Normal	0.004	0.027	0.312	0.490	0.005	0.031
Rayleigh	0.000	0.000	0.000	0.000	0.000	0.000
Rician	0.000	0.000	0.339	0.554	0.000	0.000
t Location-Scale	0.060	0.214	0.313	0.772	0.084	0.211
Weibull	0.148	0.370	0.071	0.464	0.130	0.416

Table 16: p-values for the Chi-Square Goodness-of-Fit Test and Kolmogorov-Smirnov Test for Each 1-Qubit Randomized Benchmarking Characteristic, Seed 4, Part II. The red cells indicate distribution types rejected at the significance level of $\alpha = 0.05$.

presented. For fidelity and the a and α parameters, none of the distribution types are rejected by both tests. For the b parameter, both tests reject only the inverse Gaussian distribution. This rejection is extremely odd, since the normalized median value of the AICc and BIC metrics indicate that the inverse Gaussian is the best-fit distribution type. For the err_a parameter, both tests reject the exponential, extreme value, and half-normal distribution types. For the err_α parameter, both tests reject the exponential, extreme value, generalized Pareto, half-normal, Nakagami, normal, Rayleigh, and Rician distributions. For the err_b parameter, both tests reject the exponential, generalized Pareto, half-normal, and Rayleigh distributions.

Again, for three-qubit randomized benchmarking metrics, the Chi-square test cannot make a determination for many distribution types since there are fewer data

	Fid.		a		α		b	
	Chi2	KS	Chi2	KS	Chi2	KS	Chi2	KS
Birnbaum Saunders	0.033	0.439	0.145	0.468	0.003	0.045	0.023	0.561
Burr	0.115	0.884	0.216	0.913	-	0.728	0.188	0.885
Exponential	-	0.000	-	0.000	-	0.000	-	0.000
Extreme Value	0.320	0.690	0.019	0.587	0.306	0.299	0.025	0.310
Gamma	0.055	0.495	0.120	0.481	0.003	0.046	0.023	0.563
Generalized Extreme Value	0.243	0.778	0.100	0.760	0.106	0.342	0.049	0.754
Generalized Pareto	-	0.000	-	0.000	-	0.000	-	0.000
Half-Normal	-	0.000	-	0.000	-	0.000	-	0.000
Inverse Gaussian	0.033	0.436	0.068	0.366	-	0.000	0.000	0.059
Logistic	0.184	0.775	0.282	0.844	0.123	0.264	0.134	0.845
Loglogistic	0.116	0.716	0.271	0.829	0.119	0.258	0.109	0.844
Lognormal	0.033	0.473	0.137	0.473	0.002	0.045	0.021	0.563
Nakagami	0.071	0.540	0.123	0.502	0.003	0.047	0.028	0.558
Normal	0.091	0.584	0.121	0.518	0.003	0.047	0.030	0.574
Rayleigh	-	0.001	0.000	0.000	-	0.000	-	0.000
Rician	0.096	0.576	0.135	0.544	0.003	0.048	0.032	0.560
t Location-Scale	0.070	0.754	0.172	0.852	0.031	0.452	0.121	0.843
Weibull	0.270	0.700	0.030	0.638	0.284	0.291	0.015	0.352

Table 17: p-values for the Chi-Square Goodness-of-Fit Test and Kolmogorov-Smirnov Test for Each 2-Qubit Randomized Benchmarking Characteristic, Seed 0, Part I. The red cells indicate distribution types rejected at the significance level of $\alpha = 0.05$.

points.

In Tables 21 and 22, the p-values for three-qubit randomized benchmarking metrics with one seed of data (Seed 0) are presented. For fidelity, none of the distribution types are rejected by both metrics. For the a parameter, both tests reject the half-normal distribution. For the α parameter, both tests reject the Birnbaum Saunders, extreme value, gamma, inverse Gaussian, logistic, loglogistic, lognormal, Nakagami, normal, Rician, t location-scale, and Weibull distributions. For the b parameter, none of the distribution types are rejected by both tests. For the err_a and err_α param-

	err_a		*err_α*		*err_b*	
	Chi2	KS	Chi2	KS	Chi2	KS
Birnbaum Saunders	0.093	0.563	0.499	0.112	0.238	0.854
Burr	0.568	0.882	0.849	0.825	0.215	0.930
Exponential	0.000	0.000	0.000	0.000	0.000	0.000
Extreme Value	0.000	0.120	0.004	0.001	0.008	0.360
Gamma	0.068	0.515	0.343	0.079	0.252	0.847
Generalized Extreme Value	0.296	0.814	0.804	0.791	0.179	0.924
Generalized Pareto	0.000	0.001	0.000	0.003	0.000	0.005
Half-Normal	0.000	0.000	0.000	0.000	0.000	0.000
Inverse Gaussian	0.093	0.570	0.513	0.131	0.238	0.851
Logistic	0.122	0.666	0.449	0.093	0.288	0.871
Loglogistic	0.169	0.796	0.791	0.633	0.359	0.907
Lognormal	0.097	0.598	0.528	0.236	0.266	0.874
Nakagami	0.031	0.425	0.150	0.013	0.235	0.740
Normal	0.012	0.347	0.096	0.008	0.235	0.679
Rayleigh	0.000	0.002	0.000	0.000	0.000	0.011
Rician	0.012	0.383	0.206	0.000	0.257	0.687
t Location-Scale	0.000	0.003	0.116	0.668	0.000	0.005
Weibull	0.011	0.165	0.011	0.083	0.009	0.181

Table 18: p-values for the Chi-Square Goodness-of-Fit Test and Kolmogorov-Smirnov Test for Each 2-Qubit Randomized Benchmarking Characteristic, Seed 0, Part II. The red cells indicate distribution types rejected at the significance level of $\alpha = 0.05$.

eters, both tests reject the exponential, extreme value, half-normal, logistic, normal, Rayleigh, and Rician distributions. It is noteworthy that for the *err_a* parameter, the KS test rejects all distribution types, and for the *err_α* parameter, the KS test rejects all distribution types except for the Burr distribution. It is likely that these rejections are due to the small number of data points compared to one-qubit and two-qubit randomized benchmarking. For the *err_b* parameter, both tests reject the exponential, generalized Pareto, and half-normal distributions.

In Tables 23 and 24, the p-values for three-qubit randomized benchmarking metrics with five seeds of data (Seed 4) are presented. For fidelity, none of the distribution types are rejected by both tests. For the a parameter, both tests reject the extreme value, normal, Rayleigh, Rician, and Weibull distributions. For the α pa-

	Fid.		a		α		b	
	Chi2	KS	Chi2	KS	Chi2	KS	Chi2	KS
Birnbaum Saunders	0.020	0.244	0.007	0.204	0.001	0.057	0.018	0.335
Burr	0.041	0.907	0.106	0.835	0.138	0.806	0.163	0.841
Exponential	-	0.000	-	0.000	-	0.000	-	0.000
Extreme Value	0.326	0.694	0.019	0.551	0.133	0.341	0.009	0.224
Gamma	0.024	0.287	0.009	0.211	0.001	0.059	0.011	0.328
Generalized Extreme Value	0.206	0.740	0.043	0.672	0.049	0.362	0.026	0.710
Generalized Pareto	-	0.000	-	0.000	-	0.000	-	0.000
Half-Normal	-	0.000	-	0.000	-	0.000	-	0.000
Inverse Gaussian	0.020	0.225	0.007	0.065	-	-	0.000	0.003
Logistic	0.325	0.654	0.088	0.708	0.027	0.283	0.094	0.724
Loglogistic	0.148	0.546	0.114	0.630	0.023	0.280	0.084	0.746
Lognormal	0.019	0.255	0.007	0.205	0.001	0.062	0.028	0.332
Nakagami	0.039	0.306	0.012	0.210	0.001	0.060	0.012	0.320
Normal	0.055	0.357	0.011	0.211	0.001	0.063	0.014	0.305
Rayleigh	-	0.000	-	0.000	-	0.000	-	0.000
Rician	0.058	0.356	0.013	0.228	0.001	0.062	0.016	0.316
t Location-Scale	0.083	0.672	0.078	0.794	0.024	0.477	0.189	0.842
Weibull	0.318	0.624	0.017	0.499	0.129	0.328	0.006	0.182

Table 19: p-values for the Chi-Square Goodness-of-Fit Test and Kolmogorov-Smirnov Test for Each 2-Qubit Randomized Benchmarking Characteristic, Seed 4, Part I. The red cells indicate distribution types rejected at the significance level of $\alpha = 0.05$.

rameter, both tests reject the Birnbaum Saunders, extreme value, gamma, generalized extreme value, inverse Gaussian, logistic, loglogistic, lognormal, Nakagami, normal, Rician, and Weibull distributions. For the b parameter, none of the distribution types are rejected by both tests.

For the err_a and err_α parameters, the Chi-square test is able to make a determination for almost all distribution types. This is contrasted with results for Seed 0, where the Chi-square test was unable to make a determination for most of the distribution types. Since the number of data points is the same for Seed 0 as for

	err_a		*err_α*		*err_b*	
	Chi2	KS	Chi2	KS	Chi2	KS
Birnbaum Saunders	0.151	0.587	0.123	0.163	0.247	0.762
Burr	0.298	0.896	0.270	0.875	0.168	0.921
Exponential	0.000	0.007	0.000	0.000	0.000	0.000
Extreme Value	0.000	0.028	0.000	0.002	0.042	0.320
Gamma	0.147	0.446	0.058	0.114	0.249	0.767
Generalized Extreme Value	0.212	0.872	0.150	0.826	0.170	0.894
Generalized Pareto	0.000	0.067	0.000	0.009	0.000	0.004
Half-Normal	0.001	0.029	0.000	0.000	0.000	0.000
Inverse Gaussian	0.199	0.608	0.139	0.227	0.246	0.767
Logistic	0.070	0.468	0.146	0.129	0.255	0.854
Loglogistic	0.341	0.839	0.442	0.665	0.312	0.875
Lognormal	0.246	0.674	0.219	0.342	0.240	0.780
Nakagami	0.014	0.178	0.008	0.033	0.273	0.749
Normal	0.003	0.101	0.001	0.017	0.198	0.700
Rayleigh	0.015	0.121	0.000	0.000	0.000	0.011
Rician	0.013	0.126	0.000	0.000	0.221	0.722
t Location-Scale	0.081	0.614	0.283	0.353	0.194	0.843
Weibull	0.033	0.334	0.014	0.122	0.129	0.677

Table 20: p-values for the Chi-Square Goodness-of-Fit Test and Kolmogorov-Smirnov Test for Each 2-Qubit Randomized Benchmarking Characteristic, Seed 4, Part II. The red cells indicate distribution types rejected at the significance level of $\alpha = 0.05$.

Seed 4, this behavior suggests that the parameter error limits for a and α are closer in value, such that the binning function of the Chi-square test produces better results. For the *err_a* parameter, both tests reject the exponential, extreme value, generalized Pareto, half-normal, normal, Rayleigh, and Rician distributions. For the *err_α* parameter, both tests reject the exponential, extreme value, generalized Pareto, half-normal, Nakagami, normal, Rayleigh, and Rician distributions. For *err_b*, both tests reject the exponential, generalized Pareto, and half-normal distributions.

Negative Log-Likelihood (NLL), Akaike's Information Criterion (AIC), Corrected AIC (AICc), and Bayesian Information Criterion (BIC) The results for the NLL, AIC, AICc, and BIC, metrics are presented for each characteristic

	Fid.		a		α		b	
	Chi2	KS	Chi2	KS	Chi2	KS	Chi2	KS
Birnbaum Saunders	0.208	0.675	0.120	0.424	0.000	0.001	0.215	0.648
Burr	0.183	0.862	0.137	0.873	0.002	0.311	0.172	0.854
Exponential	-	0.000	-	0.000	-	0.000	-	0.000
Extreme Value	0.027	0.315	0.040	0.258	0.001	0.037	0.025	0.329
Gamma	0.188	0.663	0.223	0.529	0.000	0.002	0.210	0.662
Generalized Extreme Value	0.084	0.691	0.028	0.464	0.000	0.077	0.071	0.695
Generalized Pareto	-	0.000	0.077	0.000	-	0.000	-	0.000
Half-Normal	-	0.000	0.000	0.000	-	0.000	-	0.000
Inverse Gaussian	0.146	0.356	0.106	0.414	0.000	0.001	0.211	0.545
Logistic	0.304	0.839	0.153	0.713	0.000	0.036	0.340	0.853
Loglogistic	0.355	0.823	0.258	0.600	0.001	0.019	0.360	0.836
Lognormal	0.195	0.681	0.128	0.499	0.000	0.001	0.208	0.652
Nakagami	0.209	0.709	0.207	0.535	0.000	0.002	0.245	0.681
Normal	0.230	0.677	0.085	0.477	0.000	0.004	0.225	0.690
Rayleigh	-	0.000	0.107	0.004	-	0.000	-	0.000
Rician	0.249	0.685	0.133	0.498	0.000	0.004	0.245	0.691
t Location-Scale	0.200	0.842	0.050	0.718	0.000	0.049	0.260	0.837
Weibull	0.064	0.384	0.128	0.444	0.000	0.013	0.061	0.401

Table 21: p-values for the Chi-Square Goodness-of-Fit Test and Kolmogorov-Smirnov Test for Each 3-Qubit Randomized Benchmarking Characteristic, Seed 0, Part I. The red cells indicate distribution types rejected at the significance level of $\alpha = 0.05$.

in Tables 25 through 132. For each metric, three numbers are reported for each distribution type. The first number, the normalized average value, is the metric's weight averaged across all qubits/links for the given distribution type. The second number is the count of qubits/links that have the lowest metric value for the distribution type. The third number is the percentage corresponding to the count.

Calibration Characteristics For T1, the generalized extreme value (GEV) distribution is the best-fit distribution type as determined by NLL. The Weibull distribution is the best-fit distribution type as determined by AIC, AICc, and

	err_a		*err_α*		*err_b*	
	Chi2	KS	Chi2	KS	Chi2	KS
Birnbaum Saunders	-	0.000	-	0.000	0.277	0.616
Burr	-	0.022	-	0.090	0.316	0.937
Exponential	0.001	0.000	0.001	0.000	0.000	0.000
Extreme Value	0.000	0.000	0.000	0.000	0.014	0.233
Gamma	-	0.000	-	0.000	0.292	0.730
Generalized Extreme Value	-	0.001	-	0.002	0.281	0.900
Generalized Pareto	-	0.006	-	0.026	0.000	0.001
Half-Normal	0.000	0.000	0.000	0.000	0.000	0.000
Inverse Gaussian	-	0.000	-	0.000	0.263	0.611
Logistic	0.000	0.000	0.002	0.000	0.321	0.801
Loglogistic	-	0.010	-	0.036	0.398	0.882
Lognormal	-	0.001	-	0.002	0.292	0.679
Nakagami	-	0.000	-	0.000	0.357	0.664
Normal	0.000	0.000	0.000	0.000	0.272	0.675
Rayleigh	0.000	0.000	0.000	0.000	0.014	0.073
Rician	0.000	0.000	0.000	0.000	0.299	0.717
t Location-Scale	-	0.000	-	0.000	0.254	0.776
Weibull	-	0.000	-	0.000	0.271	0.714

Table 22: p-values for the Chi-Square Goodness-of-Fit Test and Kolmogorov-Smirnov Test for Each 3-Qubit Randomized Benchmarking Characteristic, Seed 0, Part II. The red cells indicate distribution types rejected at the significance level of $\alpha = 0.05$.

BIC, respectively. The similarities between the collected T1 data for ibmq_poughkeepsie Q00 and the corresponding GEV and Weibull distribution models can be observed in Table 27. The GEV distribution probability density function is given by the expression,

$$f(x|k,\mu,\sigma) = \frac{1}{\sigma}\exp\left(-\left(1+\frac{k(x-\mu)}{\sigma}\right)^{-\frac{1}{k}}\right)\left(1+\frac{k(x-\mu)}{\sigma}\right)^{-1-\frac{1}{k}}, \tag{25}$$

for $1+\frac{k(x-\mu)}{\sigma} > 0$, where k is the shape parameter, $\sigma \geq 0$ is the scale parameter, and μ is the location parameter [17, 18]. When $k < 0$ (Type III case of the GEV distribution), the Weibull distribution family is obtained [48]. For T1, 98.6% of all qubits are Type III. The Weibull distribution probability density function is given by

	Fid.		a		α		b	
	Chi2	KS	Chi2	KS	Chi2	KS	Chi2	KS
Birnbaum Saunders	0.118	0.440	0.002	0.088	0.000	0.000	0.068	0.439
Burr	0.171	0.820	0.072	0.574	0.319	0.693	0.136	0.825
Exponential	-	0.000	-	0.000	-	0.000	-	0.000
Extreme Value	0.019	0.280	0.000	0.008	0.000	0.032	0.013	0.304
Gamma	0.113	0.472	0.001	0.072	0.000	0.000	0.070	0.466
Generalized Extreme Value	0.057	0.583	0.002	0.191	0.008	0.035	0.054	0.540
Generalized Pareto	-	0.000	-	0.000	-	0.000	-	0.000
Half-Normal	-	0.000	-	0.000	-	0.000	-	0.000
Inverse Gaussian	0.035	0.150	0.002	0.086	0.000	0.000	0.035	0.125
Logistic	0.295	0.767	0.006	0.287	0.000	0.032	0.272	0.784
Loglogistic	0.247	0.752	0.016	0.346	0.001	0.012	0.237	0.787
Lognormal	0.110	0.430	0.001	0.090	0.000	0.000	0.070	0.429
Nakagami	0.133	0.498	0.000	0.056	0.000	0.001	0.084	0.482
Normal	0.131	0.487	0.000	0.040	0.000	0.001	0.101	0.486
Rayleigh	-	0.000	0.000	0.000	-	0.000	-	0.000
Rician	0.125	0.507	0.000	0.046	0.000	0.001	0.094	0.506
t Location-Scale	0.206	0.841	0.000	0.728	0.000	0.245	0.209	0.831
Weibull	0.034	0.350	0.000	0.035	0.000	0.010	0.026	0.366

Table 23: p-values for the Chi-Square Goodness-of-Fit Test and Kolmogorov-Smirnov Test for Each 3-Qubit Randomized Benchmarking Characteristic, Seed 4, Part I. The red cells indicate distribution types rejected at the significance level of $\alpha = 0.05$.

the expression,

$$f(x|a,b) = \frac{b}{a}\left(\frac{x}{a}\right)^{b-1} \exp\left(-\left(\frac{x}{a}\right)^{b}\right), \tag{26}$$

where $a > 0$ is the scale parameter and $b > 0$ is the shape parameter [36, 37]. Thus, the results observed from NLL, AIC, AICc, and BIC, are consistent with each other. Since AIC, AICc, and BIC penalize models with more parameters (the Weibull distribution used by MATLAB only has two parameters, while the GEV distribution has three parameters), it follows that AIC, AICc, and BIC indicate the Weibull distribution as a better distribution type than the GEV distribution.

	err_a		*err_α*		*err_b*	
	Chi2	KS	Chi2	KS	Chi2	KS
Birnbaum Saunders	0.435	0.000	0.209	0.000	0.233	0.658
Burr	0.178	0.054	0.414	0.287	0.191	0.890
Exponential	0.001	0.000	0.006	0.000	0.000	0.000
Extreme Value	0.000	0.000	0.000	0.000	0.004	0.134
Gamma	0.437	0.000	0.142	0.000	0.310	0.762
Generalized Extreme Value	0.141	0.000	0.427	0.021	0.164	0.825
Generalized Pareto	0.001	0.002	0.003	0.008	0.000	0.004
Half-Normal	0.000	0.000	0.000	0.000	0.000	0.000
Inverse Gaussian	0.446	0.000	0.469	0.000	0.230	0.669
Logistic	0.254	0.000	0.147	0.000	0.238	0.740
Loglogistic	0.389	0.028	0.548	0.048	0.321	0.870
Lognormal	0.427	0.000	0.515	0.001	0.264	0.699
Nakagami	0.305	0.000	0.019	0.000	0.279	0.715
Normal	0.000	0.000	0.000	0.000	0.185	0.592
Rayleigh	0.000	0.000	0.000	0.000	0.069	0.156
Rician	0.000	0.000	0.000	0.000	0.206	0.644
t Location-Scale	0.217	0.000	-	0.000	0.142	0.708
Weibull	0.319	0.000	0.071	0.000	0.209	0.667

Table 24: p-values for the Chi-Square Goodness-of-Fit Test and Kolmogorov-Smirnov Test for Each 3-Qubit Randomized Benchmarking Characteristic, Seed 4, Part II. The red cells indicate distribution types rejected at the significance level of $\alpha = 0.05$.

For T2, the GEV distribution is again selected as the best-fit distribution type as determined by NLL. The extreme value (EV) distribution and the Weibull distribution are selected as best-fit distribution types by equal AIC and BIC counts. The similarities between the collected T2 data for ibmq_poughkeepsie Q00 and the corresponding GEV, EV, and Weibull distribution models can be observed in Table 27. The EV distribution probability density function is given by the expression,

$$f(x|a,b) = \left(\frac{1}{\sigma}\right) \exp\left(\frac{x-\mu}{\sigma}\right) \exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right), \tag{27}$$

where μ is the location parameter, and $\sigma \geq 0$ is the scale parameter [12, 13]. The EV distribution has larger AIC, AICc, and BIC weights than the Weibull distribution.

However, the difference in weights is insignificant in this context since the Weibull distribution family is obtained when the data points are bounded below in the EV distribution (T2 data points are bounded below by zero) [4]. Since the EV distribution and Weibull distribution each have only two parameters (compared to the GEV distribution's three parameters), it follows that the information criteria favored the EV and Weibull distribution types almost equally.

The Weibull distribution is a lifetime distribution model used for reliability and failure times [3]. Nevertheless, it is initially surprising that the best-fit distribution type for the T1 and T2 characteristics is a Weibull distribution rather than the exponential distribution. T1 times are measured by repeatedly initializing a qubit to the excited state, and then taking measurements after an increasing time t to determine the survival probability of remaining in the excited state. The probability follows the exponential decay of $\exp(-t/T1)$. T2 times are measured using a Hahn echo experiment, and like the T1 times, the survival probability follows the exponential decay of $\exp(-t/T2)$ [6, 9]. Thus, when all measurements of a qubit's excited states occur within the short span of time (seconds) required to perform the calibration, the survival function is an exponential decay. The corresponding hazard function is constant, indicating that a qubit under nearly constant conditions has a constant rate of failure. For the data points represented in this work, however, the T1 coherence times are collected daily over spans of several weeks. The T1 values change daily, which suggests daily environmental noise variations, temperature variations, and (less likely) device variations (due to junction aging, for example [102, 161]). As a result of the daily variations, the probability that a prepared qubit would remain in the excited state over time resembles the survival function of a Weibull distribution rather than an exponential distribution. The Weibull distribution model can be compared to the exponential distribution model in Table 27. When the shape parameter of a

Weibull distribution is equal to 1, the exponential distribution with a constant hazard function is obtained. However, when the shape parameter of a Weibull distribution is greater than 1 (as was determined to be the case for all of the T1 and T2 distributions), the hazard rate increases over time, indicating that a qubit under changing qubit conditions becomes more likely to fail over time [166].

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.009	1	0.30%	0.001	4	1.40%	0.001	4	1.40%	0.001	4	1.40%
Burr	0.104	37	12.90%	0.058	1	0.30%	0.050	1	0.30%	0.024	0	0.00%
Exponential	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Extreme Value	0.041	30	10.50%	0.021	54	18.90%	0.023	56	19.60%	0.026	60	21.00%
Gamma	0.027	0	0.00%	0.008	10	3.50%	0.008	10	3.50%	0.010	10	3.50%
Generalized Extreme Value	0.125	151	52.80%	0.075	27	9.40%	0.065	23	8.00%	0.032	6	2.10%
Generalized Pareto	0.000	17	5.90%	0.000	14	4.90%	0.000	13	4.50%	0.000	13	4.50%
Half-Normal	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Inverse Gaussian	0.009	0	0.00%	0.001	4	1.40%	0.001	4	1.40%	0.001	5	1.70%
Logistic	0.055	7	2.40%	0.038	26	9.10%	0.039	27	9.40%	0.043	30	10.50%
Loglogistic	0.021	0	0.00%	0.005	4	1.40%	0.005	4	1.40%	0.006	4	1.40%
Lognormal	0.010	0	0.00%	0.001	2	0.70%	0.001	2	0.70%	0.002	2	0.70%
Nakagami	0.056	4	1.40%	0.036	34	11.90%	0.038	34	11.90%	0.046	33	11.50%
Normal	0.080	0	0.00%	0.091	8	2.80%	0.094	9	3.10%	0.099	11	3.80%
Rayleigh	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	1	0.30%
Rician	0.080	1	0.30%	0.089	20	7.00%	0.092	20	7.00%	0.100	20	7.00%
t Location Scale	0.086	23	8.00%	0.039	7	2.40%	0.034	6	2.10%	0.018	4	1.40%
Weibull	0.101	15	5.20%	0.114	71	24.80%	0.117	73	25.50%	0.134	83	29.00%

Table 25: NLL, AIC, AICc, and BIC Results for T1 coherence times. The green cells indicate the best-fit distribution type(s).

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.002	2	0.70%	0.000	4	1.40%	0.000	4	1.40%	0.000	4	1.40%
Burr	0.102	31	10.80%	0.052	6	2.10%	0.044	6	2.10%	0.025	4	1.40%
Exponential	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Extreme Value	0.054	38	13.30%	0.032	58	20.30%	0.035	63	22.00%	0.045	69	24.10%
Gamma	0.009	2	0.70%	0.001	5	1.70%	0.001	5	1.70%	0.001	5	1.70%
Generalized Extreme Value	0.124	105	36.70%	0.064	29	10.10%	0.055	24	8.40%	0.031	9	3.10%
Generalized Pareto	0.000	30	10.50%	0.000	24	8.40%	0.000	23	8.00%	0.000	21	7.30%
Half-Normal	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Inverse Gaussian	0.002	0	0.00%	0.000	7	2.40%	0.000	7	2.40%	0.000	7	2.40%
Logistic	0.056	15	5.20%	0.036	41	14.30%	0.038	41	14.30%	0.041	45	15.70%
Loglogistic	0.010	0	0.00%	0.001	4	1.40%	0.001	5	1.70%	0.001	6	2.10%
Lognormal	0.002	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Nakagami	0.025	2	0.70%	0.005	16	5.60%	0.006	16	5.60%	0.007	17	5.90%
Normal	0.059	0	0.00%	0.032	5	1.70%	0.033	5	1.70%	0.040	6	2.10%
Rayleigh	0.000	0	0.00%	0.000	2	0.70%	0.000	2	0.70%	0.000	2	0.70%
Rician	0.058	4	1.40%	0.030	11	3.80%	0.031	11	3.80%	0.038	12	4.20%
t Location Scale	0.087	38	13.30%	0.036	16	5.60%	0.029	15	5.20%	0.017	10	3.50%
Weibull	0.086	19	6.60%	0.076	58	20.30%	0.080	59	20.60%	0.089	69	24.10%

Table 26: NLL, AIC, AICc, and BIC Results for T2 coherence times. The green cells indicate the best-fit distribution type(s).

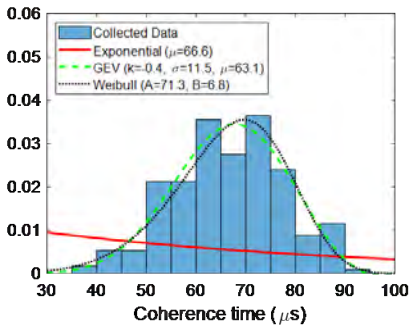
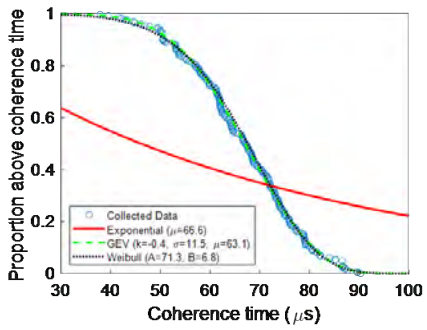
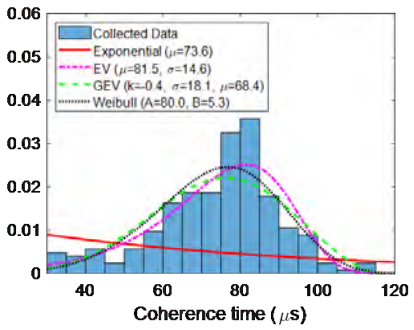
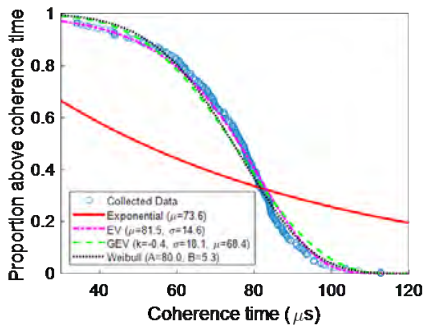
	Probability Density Function	Survival Function
T1		
T2		

Table 27: Distribution Model Analysis for ibmq_poughkeepsie Q00 T1 and T2 Coherence Times. In the left column, the probability density functions for the exponential, EV (for T2), GEV, and Weibull distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the exponential, EV (for T2), GEV, and Weibull distribution models are plotted on top of the proportion of collected data observations above a given coherence time.

Like T1 and T2 coherence times, qubit decoherence primarily drives single-qubit gate errors. For T1 and T2 coherence times, interactions with environmental noise dominate the decoherence. For single-qubit gate errors, however, the dominant source of decoherence is caused by interactions with the driving fields inducing leakage outside the qubit subspace and inducing phase shifts from the presence of other levels [54]. For multi-qubit gate errors, crosstalk is an additional source of decoherence [119].

Similar to T1 and T2 coherence times, the GEV distribution emerged as the best distribution type for both single-qubit gate error rates and multi-qubit gate error rates, as determined by all four metrics (see Tables 28 and 29). Unlike T1

and T2 coherence times, however, the shape parameter is positive for 87.1% of all the single-qubit gate error distributions and 87.0% of all the multi-qubit gate error distributions, indicating the Type II (Fréchet) case of the GEV distribution. The Fréchet distribution probability density function is given by the expression,

$$f(x|a,b) = ab(ax)^{b+1}\exp(-(ax)^{-b}), \tag{28}$$

where x is positive, $a > 0$ is the scale parameter, and $b > 0$ is the shape parameter [72]. The main application of the Fréchet distribution, also known as the Inverse Weibull distribution, is to model maximum values in a data set [14]. More specifically, in reliability engineering, "when a failure is related to the maximum value of a critical non-negative variable, this generative mechanism can be considered" [72]. In many reliability engineering cases, this "critical non-negative variable" is time. In the context of single-qubit and multi-qubit gate error rates, the error rates are determined by randomized benchmarking [32, 38, 112]. After a sequence of Clifford gates, the probability that a qubit returns to its initial state reduces exponentially as the number of gates in the sequence increases according to an exponential decay function $A\alpha^m + B$, where m is the number of Clifford gates. The single-qubit and multi-qubit gate error rates are extracted from the exponential decay. Thus, the concept of time in this context is measured by gate count, and the failure rate increases exponentially in relation to gate count. Because failure rate is a measure of the number of failures per gate count, failure rate and gate count have an inverse relationship. As such, each data point (again collected under changing conditions over spans of several weeks) represented in the distribution corresponds to an inverse "time" to failure. Whereas datapoints for T1 and T2 coherence times correspond to time to qubit failure and result in a Weibull distribution, datapoints for the single-qubit gate error rates correspond to an inverse gate count to failure and result in the Inverse Weibull

distribution. The similarities and differences between the collected single-qubit gate error rates for ibmq_poughkeepsie Q00, the collected multi-qubit gate error rates for ibmq_poughkeepsie link Q00-Q01, and the corresponding GEV (Type II) and Weibull distribution models can be observed in Table 30.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.006	4	1.40%	0.000	7	2.40%	0.000	7	2.40%	0.000	7	2.40%
Burr	0.358	96	33.60%	0.328	85	29.70%	0.325	83	29.00%	0.304	77	26.90%
Exponential	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Extreme Value	0.000	0	0.00%	0.000	1	0.30%	0.000	1	0.30%	0.000	1	0.30%
Gamma	0.001	1	0.30%	0.000	6	2.10%	0.000	7	2.40%	0.000	7	2.40%
Generalized Extreme Value	0.388	173	60.50%	0.407	140	49.00%	0.399	134	46.90%	0.369	123	43.00%
Generalized Pareto	0.000	6	2.10%	0.000	5	1.70%	0.000	4	1.40%	0.000	4	1.40%
Half-Normal	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Inverse Gaussian	0.008	2	0.70%	0.000	17	5.90%	0.000	21	7.30%	0.000	25	8.70%
Logistic	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Loglogistic	0.023	0	0.00%	0.003	16	5.60%	0.003	19	6.60%	0.010	30	10.50%
Lognormal	0.009	0	0.00%	0.000	1	0.30%	0.000	2	0.70%	0.001	4	1.40%
Nakagami	0.000	0	0.00%	0.000	3	1.00%	0.000	3	1.00%	0.000	3	1.00%
Normal	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Rayleigh	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Rician	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
t Location Scale	0.017	4	1.40%	0.001	4	1.40%	0.001	4	1.40%	0.001	4	1.40%
Weibull	0.000	0	0.00%	0.000	1	0.30%	0.000	1	0.30%	0.000	1	0.30%

Table 28: NLL, AIC, AICc, and BIC Results for Single-Qubit Gate Error Rates. The green cells indicate the best-fit distribution type(s).

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.004	2	0.30%	0.000	4	0.70%	0.000	4	0.70%	0.000	4	0.70%
Burr	0.330	182	31.10%	0.300	168	28.70%	0.297	158	27.00%	0.256	146	25.00%
Exponential	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Extreme Value	0.000	2	0.30%	0.000	2	0.30%	0.000	2	0.30%	0.000	3	0.50%
Gamma	0.001	0	0.00%	0.000	3	0.50%	0.000	3	0.50%	0.000	3	0.50%
Generalized Extreme Value	0.385	347	59.30%	0.417	284	48.50%	0.412	278	47.50%	0.344	256	43.80%
Generalized Pareto	0.000	13	2.20%	0.000	10	1.70%	0.000	10	1.70%	0.000	10	1.70%
Half-Normal	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Inverse Gaussian	0.005	5	0.90%	0.000	44	7.50%	0.000	46	7.90%	0.000	55	9.40%
Logistic	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Loglogistic	0.026	0	0.00%	0.004	27	4.60%	0.005	42	7.20%	0.009	64	10.90%
Lognormal	0.006	0	0.00%	0.000	6	1.00%	0.000	6	1.00%	0.001	6	1.00%
Nakagami	0.000	0	0.00%	0.000	2	0.30%	0.000	2	0.30%	0.000	2	0.30%
Normal	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Rayleigh	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	2	0.30%
Rician	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	2	0.30%
t Location Scale	0.020	31	5.30%	0.001	31	5.30%	0.001	30	5.10%	0.001	30	5.10%
Weibull	0.000	3	0.50%	0.000	4	0.70%	0.000	4	0.70%	0.000	2	0.30%

Table 29: NLL, AIC, AICc, and BIC Results for Multi-Qubit Gate Error Rates. The green cells indicate the best-fit distribution type(s).

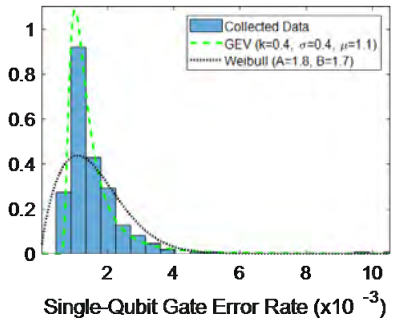
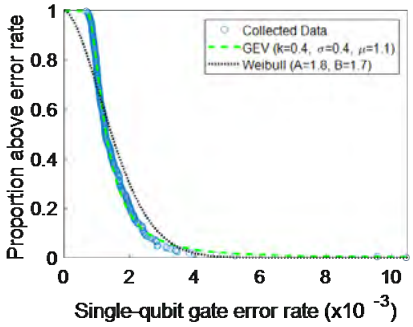
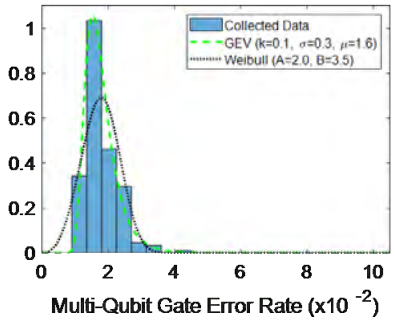
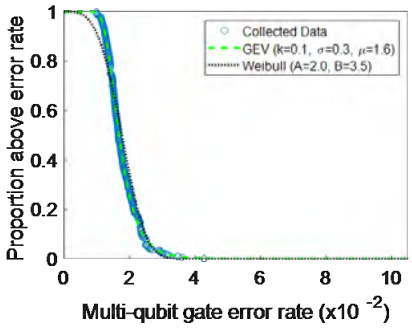
	Probability Density Function	Survival Function
Single-Qubit Gate Error Rate		
Multi-Qubit Gate Error Rate		

Table 30: Distribution Model Analysis for ibmq_poughkeepsie Q00 Single-Qubit Gate Error Rates and ibmq_poughkeepsie Q00-Q01 Multi-Qubit Gate Error Rates. In the left column, the probability density functions for the GEV and Weibull distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV and Weibull distribution models are plotted on top of the proportion of collected data observations above a given single-qubit gate error rate.

There are two dominant sources of readout errors. The first source is excited-state-to-ground-state (T1) decoherence events that occur during qubit measurement from environmental noise. Additional noise from the control equipment is filtered and attenuated to become negligible [79]. Because noise causes measurement uncertainty in distinguishing between the excited state and the ground state, readout signals are acquired for a long period of time to improve measurement contrast, but a long measurement time increases the chance of decoherence during measurement [70]. The

second source is the small probability of measuring the opposite value due to the fact that there is overlap in the measurement probability distributions corresponding to the ground and first excited states [122]. It should be noted that multilevel encoding has been proposed and tested as a method to nearly eradicate readout errors [70, 97], but it has not been implemented in the IBM systems available to users as of the time of writing this dissertation. While gate operations can induce qubit errors by interactions with the drive field, readout operations are conducted with a power low enough to avoid spurious qubit transitions, thus avoiding many of the types of errors incurred in measurements for gate error rates [113].

Results for readout errors are less straight-forward than results for coherence times and gate error rates. The two types of distributions with the best fit are the Burr distribution, as determined by the normalized average values for each metric, and the GEV distribution, as determined by the counts and percentages for each metric. The comparisons between the collected readout error rates for ibmq_poughkeepsie Q00 and the corresponding Burr and GEV distribution models can be observed in Table 32. The Burr distribution is a flexible distribution well suited for a large variety of applications. The Burr distribution probability density function is given by the expression,

$$f(x|a,c,k) = \frac{\frac{kc}{a}\left(\frac{x}{a}\right)^{c-1}}{\left(1+\left(\frac{x}{a}\right)^{c}\right)^{k+1}}, \tag{29}$$

where x is positive, $c > 0$ and $k > 0$ are shape parameters, and $a > 0$ is the scale parameter [8].

The results in Table 31 indicate that the GEV distribution is the best model for a greater number of qubits, but when the Burr distribution is the best model, it is a much better model than the GEV distribution model. Upon inspecting the Burr distribution models, $c > 1$ (unimodal) describes 100% of the distributions [139]. However, analyzing the GEV distribution models reveals a mixture of shape parameter

values. Analyzing the distributions for which the GEV distribution is the best fit, $k > 0$ (GEV Type II case – Fréchet distribution) is found for 52.1% of the distributions, and $k < 0$ (GEV Type III case – Weibull distribution) is found for 47.9% of the distributions.

A visual inspection of the parameter values for the Burr and GEV distribution types in Figure 28 reveal parameter clustering behaviors. For the Burr distribution, the densest clusters occur for approximately $3 \leq c \leq 30$, $0 \leq \alpha \leq 0.15$, and $0 \leq k \leq 3$. For the GEV distribution, the densest clusters occur for approximately $-0.4 \leq k \leq 0.6$, $0 \leq \sigma \leq 0.2$, and $0 \leq \mu \leq 0.1$. There are also a few surprising secondary clusters, particularly for the μ parameter of the GEV distribution. The secondary clusters are largely comprised of qubits for which the GEV distribution is rejected with $\alpha = 0.05$. Oddly, many of the data points for which the distribution type is rejected are also located within the primary clusters. The most significant observation is that the k-parameter cluster for the GEV distribution is roughly centered around $k = 0$, which makes interpretation of the results difficult since $k = 0$ is the key parameter value that determines whether the distribution is considered Type I, II, or III.

Much more analysis is required to fully understand and interpret these results. In particular, future study is required a) to understand the extent to which a GEV distribution with $k = -0.4$ is different from a GEV distribution with $k = 0.6$, b) to determine what qubit properties lead the Burr distribution to be a much better fit for the occasions that it is a better fit than the GEV distribution, c) to learn what causes the secondary clusters for the GEV μ parameter, and d) to understand why some datapoints within the primary clusters have a p-value that causes distribution type rejection. For the purposes of this paper, however, it is sufficient to identify that the readout error rates do not follow a normal distribution.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.033	6	2.10%	0.013	10	3.50%	0.014	11	3.80%	0.018	11	3.80%
Burr	0.229	93	32.50%	0.204	60	21.00%	0.189	60	21.00%	0.111	51	17.80%
Exponential	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Extreme Value	0.000	3	1.00%	0.000	8	2.80%	0.000	8	2.80%	0.000	8	2.80%
Gamma	0.018	2	0.70%	0.003	11	3.80%	0.003	11	3.80%	0.005	11	3.80%
Generalized Extreme Value	0.160	132	46.20%	0.099	68	23.80%	0.089	64	22.40%	0.048	55	19.20%
Generalized Pareto	0.000	11	3.80%	0.000	10	3.50%	0.000	9	3.10%	0.000	8	2.80%
Half-Normal	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Inverse Gaussian	0.038	3	1.00%	0.013	27	9.40%	0.013	30	10.50%	0.020	34	11.90%
Logistic	0.006	0	0.00%	0.000	5	1.70%	0.000	5	1.70%	0.000	6	2.10%
Loglogistic	0.056	0	0.00%	0.039	30	10.50%	0.041	31	10.80%	0.046	43	15.00%
Lognormal	0.038	0	0.00%	0.014	3	1.00%	0.017	3	1.00%	0.023	5	1.70%
Nakagami	0.004	1	0.30%	0.000	10	3.50%	0.000	10	3.50%	0.000	11	3.80%
Normal	0.001	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Rayleigh	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Rician	0.001	1	0.30%	0.000	10	3.50%	0.000	10	3.50%	0.000	11	3.80%
t Location Scale	0.048	33	11.50%	0.008	30	10.50%	0.007	30	0.00%	0.005	28	9.80%
Weibull	0.001	1	0.30%	0.000	4	1.40%	0.000	4	10.50%	0.000	4	1.40%

Table 31: NLL, AIC, AICc, and BIC Results for Readout Error Rates. The green cells indicate the best-fit distribution type(s).

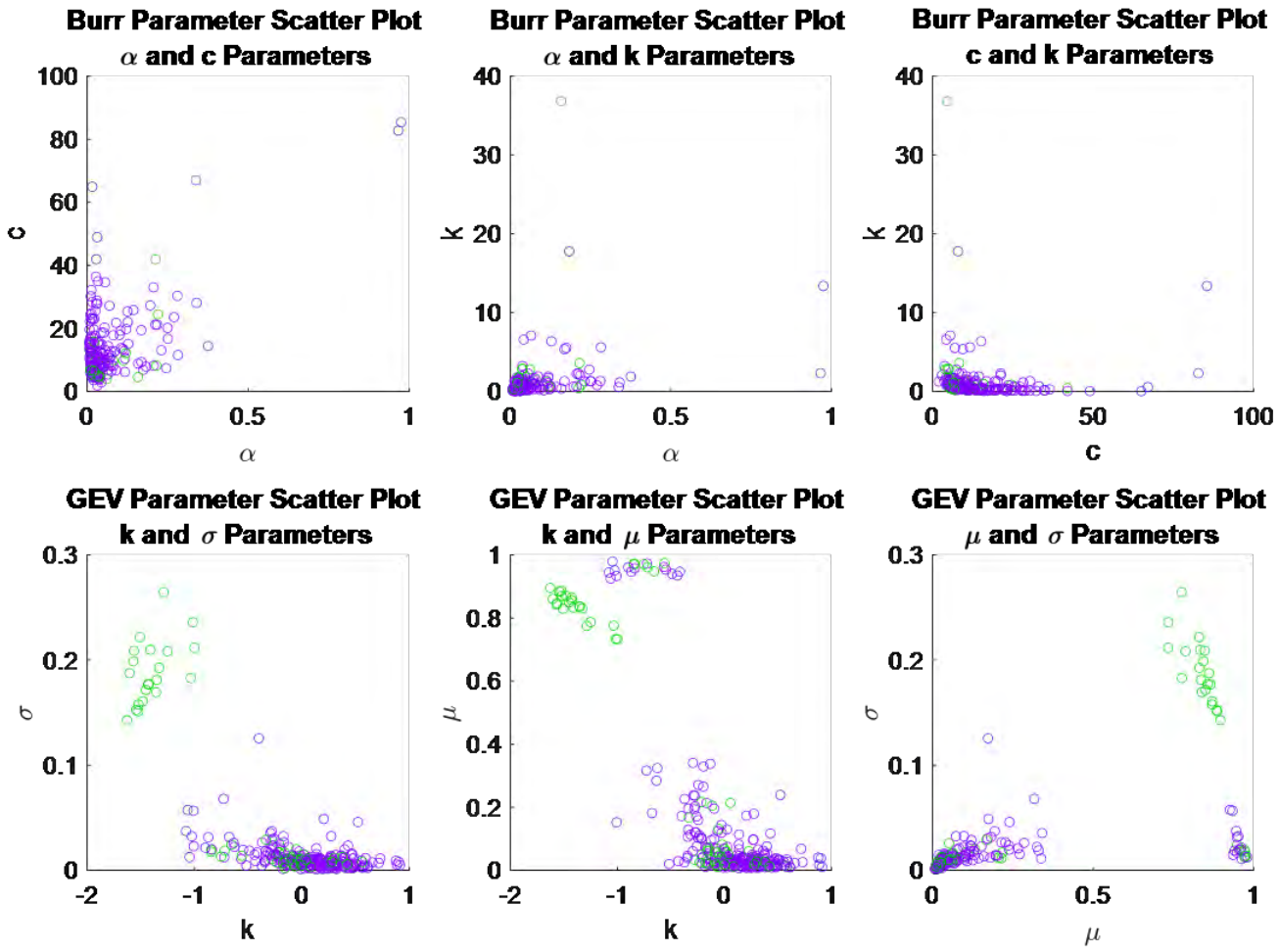


Figure 28: Burr and GEV distribution parameter scatter plots for readout error rates

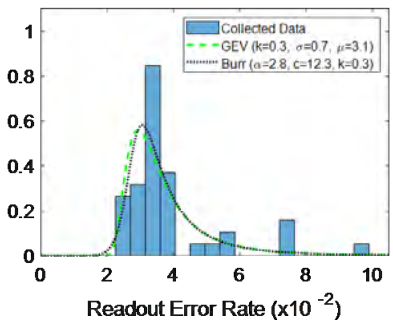
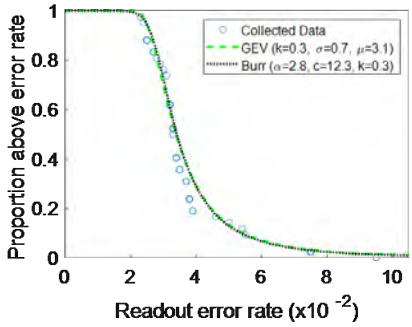
	Probability Density Function	Survival Function
Readout Error Rate		

Table 32: Distribution Model Analysis for ibmq_poughkeepsie Q00 Readout Error Rates. In the left column, the probability density functions for the GEV and Burr distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV and Burr distribution models are plotted on top of the proportion of collected data observations above a given readout error rate.

Quantum Process Tomography Performing distribution analysis on each of the quantum process tomography characteristics reveals many interesting results that are highly recommended for a future direction of study. The main purpose of this section is to determine whether or not the quantum process tomography characteristics are normally distributed, and results show that the characteristics are not normally distributed. Additionally, this section will give a brief description of the results for each characteristic along with comments about particularly interesting results and recommendations for how the characteristics should be further analyzed in future research.

For quantum process tomography on the X gate (QPTX) and on the Hadamard gate (QPTH), all metrics select the GEV distribution as the best-fit distribution (see Tables 33-36). Upon analyzing the GEV distribution, the shape parameter is positive for 100% of the qubits for QPTX and 97.8% of the qubits for QPTH, indicating the

Type II (Fréchet) case of the GEV distribution. It is not immediately clear why this distribution is the best fit for the quantum process tomography characteristics. Whereas the error rate characteristics from the calibration data are calculated in part by determining gate count to failure, QPT gate fidelity is not determined by any type of count to failure. It is simply a measure of the similarity between an ideal and an experimentally constructed matrix. Exploring other scenarios for which the Fréchet distribution is an applicable model may provide insight into why the Fréchet distribution is well-suited for QPTX and QPTH data.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Burr	0.286	7	9.2%	0.411	1	1.3%	0.405	0	0.0%	0.358	0	0.0%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.072	12	15.8%	0.047	14	18.4%	0.054	15	19.7%	0.124	16	21.1%
Gamma	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.422	36	47.4%	0.450	31	40.8%	0.444	30	39.5%	0.364	26	34.2%
Generalized Pareto	0.000	4	5.3%	0.000	4	5.3%	0.000	4	5.3%	0.000	4	5.3%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Logistic	0.004	0	0.0%	0.000	3	3.9%	0.000	4	5.3%	0.000	5	6.6%
Loglogistic	0.002	0	0.0%	0.000	2	2.6%	0.000	2	2.6%	0.000	2	2.6%
Lognormal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.000	0	0.0%	0.000	3	3.9%	0.000	3	3.9%	0.000	3	3.9%
t Location Scale	0.152	17	22.4%	0.065	14	18.4%	0.065	14	18.4%	0.060	13	17.1%
Weibull	0.061	0	0.0%	0.027	4	5.3%	0.032	4	5.3%	0.093	7	9.2%

Table 33: NLL, AIC, AICc, and BIC Results for QPTX. The green cells indicate the best-fit distribution type(s).

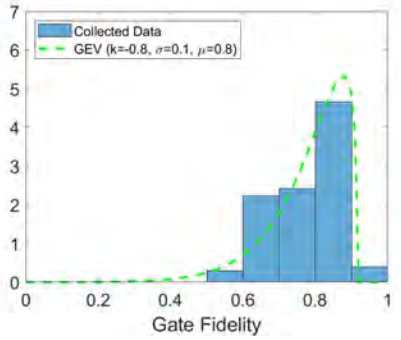
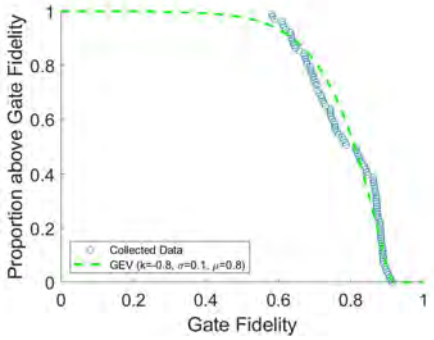
	Probability Density Function	Survival Function
QPTX		

Table 34: Distribution Model Analysis for ibmq_yorktown Q00 QPTX. In the left column, the probability density functions for the GEV distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival function for the GEV distribution model is plotted on top of the proportion of collected data observations above a given X-gate fidelity.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.003	0	0.0%	0.000	1	2.2%	0.000	1	2.2%	0.000	1	2.2%
Burr	0.203	6	13.0%	0.211	2	4.3%	0.208	1	2.2%	0.149	0	0.0%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.096	4	8.7%	0.051	6	13.0%	0.061	6	13.0%	0.185	7	15.2%
Gamma	0.005	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.429	26	56.5%	0.644	20	43.5%	0.623	20	43.5%	0.520	18	39.1%
Generalized Pareto	0.000	2	4.3%	0.000	2	4.3%	0.000	2	4.3%	0.000	2	4.3%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.003	0	0.0%	0.000	1	2.2%	0.000	1	2.2%	0.000	1	2.2%
Logistic	0.031	0	0.0%	0.006	5	10.9%	0.007	6	13.0%	0.012	6	13.0%
Loglogistic	0.016	0	0.0%	0.001	0	0.0%	0.001	0	0.0%	0.002	0	0.0%
Lognormal	0.003	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.007	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.009	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.001	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.009	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.001	1	2.2%
t Location Scale	0.120	8	17.4%	0.039	6	13.0%	0.041	5	10.9%	0.031	5	10.9%
Weibull	0.067	0	0.0%	0.048	3	6.5%	0.058	4	8.7%	0.099	5	10.9%

Table 35: NLL, AIC, AICc, and BIC Results for QPTH. The green cells indicate the best-fit distribution type(s).

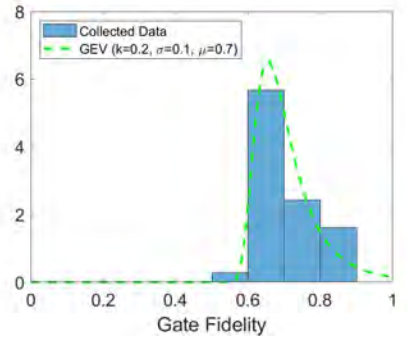
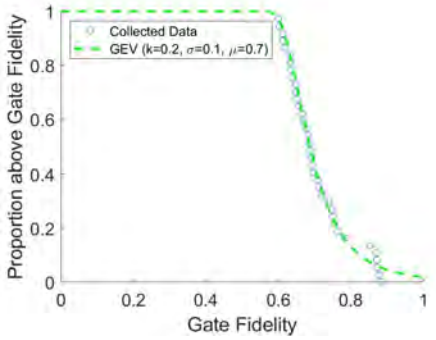
	Probability Density Function	Survival Function
QPTH		

Table 36: Distribution Model Analysis for ibmq_yorktown Q00 QPTH. In the left column, the probability density function for the GEV distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival function for the GEV distribution model is plotted on top of the proportion of collected data observations above a given H-gate fidelity.

For quantum process tomography on the CX gate (QPTCX) and the SWAP gate (QPTSWAP), the GEV distribution and the t location-scale distribution are identified as the best fit distribution types (see Tables 37–40). For both QPTCX and QPTSWAP, analysis of the GEV distribution reveals that 60% of the qubits are the Type III case (Weibull), and 40% are the Type II case (Frechét). It is unsurprising that the GEV results for QPTCX has more variation compared to QPTX and QPTH, as QPT on a two-qubit gate involves more sources of noise and error than QPT on a single-qubit gate. Further investigation should aim to determine whether the same qubits are identified as Type II or Type III for both QPTCX and QPTSWAP. It should also aim to determine which properties cause individual sets of qubits to follow a Type II versus a Type III distribution. An initial hypothesis is that the existence or non-existence of a direct physical link between two qubits might affect the Type II or Type III determination. Next steps should repeat this experiment, but separate the qubit sets into two groups — qubits with a direct physical link between them, and qubits without a direct physical link between them.

The t location-scale distribution is a general distribution that models data more prone to outliers than the normal distribution. The equation for the probability density function is

$$f\left(x|\mu,\sigma,\nu\right) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left[\frac{\nu+\left(\frac{x-\mu}{\sigma}\right)^2}{\nu}\right]^{\left(\frac{\nu+1}{2}\right)}, \tag{30}$$

where μ is the location parameter, $\sigma > 0$ is the scale parameter, and $\nu > 0$ is the shape parameter. The t location-scale distribution approaches the normal distribution as ν approaches infinity [34].

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Burr	0.216	8	20.0%	0.229	7	17.5%	0.245	7	17.5%	0.252	7	17.5%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.030	3	7.5%	0.014	8	20.0%	0.018	9	22.5%	0.040	11	27.5%
Gamma	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.378	12	30.0%	0.395	8	20.0%	0.387	8	20.0%	0.372	7	17.5%
Generalized Pareto	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Logistic	0.008	0	0.0%	0.001	3	7.5%	0.001	3	7.5%	0.002	4	10.0%
Loglogistic	0.005	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.001	0	0.0%
Lognormal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.346	17	42.5%	0.359	14	35.0%	0.346	13	32.5%	0.328	11	27.5%
Weibull	0.015	0	0.0%	0.002	0	0.0%	0.003	0	0.0%	0.006	0	0.0%

Table 37: NLL, AIC, AICc, and BIC Results for QPTCX. The green cells indicate the best-fit distribution type(s).

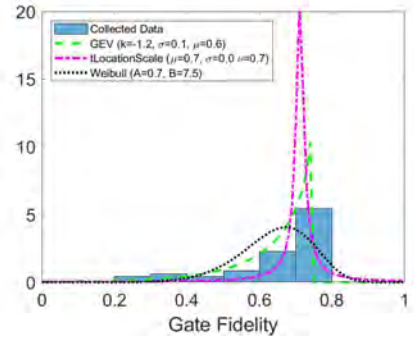
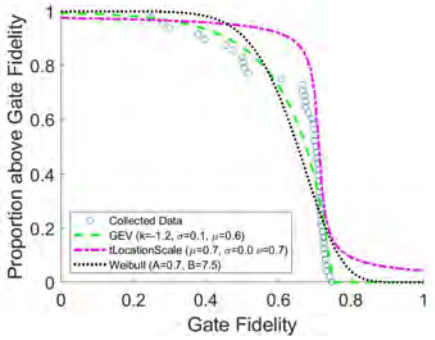
	Probability Density Function	Survival Function
QPTCX		

Table 38: Distribution Model Analysis for ibmq_yorktown Q00 QPTCX. In the left column, the probability density functions for the GEV, t location-scale, and Weibull distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, t location-scale, and Weibull distribution models are plotted on top of the proportion of collected data observations above a given CX-gate fidelity.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Burr	0.115	7	17.5%	0.036	7	17.5%	0.036	7	17.5%	0.040	7	17.5%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.081	4	10.0%	0.029	6	15.0%	0.035	7	17.5%	0.076	9	22.5%
Gamma	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.529	16	40.0%	0.828	14	35.0%	0.823	13	32.5%	0.745	11	27.5%
Generalized Pareto	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Logistic	0.005	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.006	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.001	0	0.0%
Lognormal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.210	13	32.5%	0.094	13	32.5%	0.091	13	32.5%	0.106	13	32.5%
Weibull	0.052	0	0.0%	0.012	0	0.0%	0.015	0	0.0%	0.033	0	0.0%

Table 39: NLL, AIC, AICc, and BIC Results for QPTSWAP. The green cells indicate the best-fit distribution type(s).

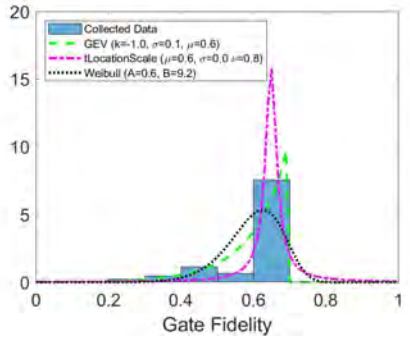
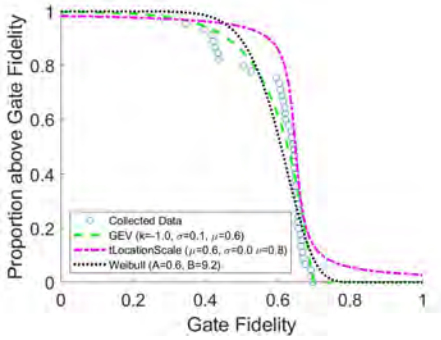
	Probability Density Function	Survival Function
QPTSWAP		

Table 40: Distribution Model Analysis for ibmq_yorktown Q00 QPTSWAP. In the left column, the probability density functions for the GEV, t location-scale, and Weibull distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, t location-scale, and Weibull distribution models are plotted on top of the proportion of collected data observations above a given SWAP-gate fidelity.

For quantum process tomography on the CCX gate (QPTCCX) and CSWAP gate (QPTCSWAP), the generalized Pareto distribution is identified as the best-fit distribution type (see Tables 41–44). The probability density function for the generalized Pareto distribution is given by the expression,

$$f(x|k,\sigma,\theta) = \left(\frac{1}{\sigma}\right)\left(1 + k\frac{x-0}{\sigma}\right)^{-1-\frac{1}{k}}, \tag{31}$$

where $k \neq 0$ is the shape parameter, σ is the scale parameter, and θ is the threshold parameter for $\theta < x$, when $k > 0$, or for $\theta < x < \theta - \sigma/k$ when $k < 0$.

When $k = 0$, the density is given by the expression,

$$f(x|0,\sigma,\theta) = \left(\frac{1}{\sigma}\right) e^{-\frac{x-\theta}{\sigma}}, \tag{32}$$

for $\theta < x$ [19].

The generalized Pareto distribution is equivalent to the exponential distribution

when $k = 0$ and $\theta = 0$, and it is equivalent to the Pareto distribution when $k > 0$ and $\theta = \sigma/k$. The generalized Pareto distribution models the tails of another distribution. When the tails decrease exponentially (for example, the normal distribution), the shape parameter of the generalized Pareto distribution is zero. When the tails decrease as a polynomial (for example, Student's t), the shape parameter is positive. When the tails are finite (for example, Beta), the shape parameter is negative [19].

For 100% of the sets of qubits tested for QPTCCX and QPTCSWAP, the shape parameter is negative. However, it is unlikely that this information is particularly meaningful. Referring back to the p-value table for QPTCCX and QPTCSWAP (Table 12), the KS test rejected the generalized Pareto distribution (the Chi-square test failed to produce a result). It is clear that the generalized Pareto distribution is not a good fit, albeit the best fit of the distribution types tested. Inspecting the probability density functions for QPTCCX and QPTCSWAP in Tables 42 and 44, respectively, possible bimodal behavior is observed. An initial hypothesis for the cause of this observed behavior is the existence of direct physical links between some sets of qubits but not others. Future research should aim to determine the cause for this behavior.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Burr	0.096	0	0.0%	0.027	0	0.0%	0.026	0	0.0%	0.025	0	0.0%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.066	1	5.6%	0.020	2	11.1%	0.029	3	16.7%	0.041	3	16.7%
Gamma	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.255	4	22.2%	0.220	3	16.7%	0.213	3	16.7%	0.215	3	16.7%
Generalized Pareto	0.508	9	50.0%	0.727	9	50.0%	0.725	8	44.4%	0.710	8	44.4%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Logistic	0.008	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.001	0	0.0%
Loglogistic	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Lognormal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.002	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.002	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.046	4	22.2%	0.004	4	22.2%	0.004	4	22.2%	0.005	4	22.2%
Weibull	0.016	0	0.0%	0.002	0	0.0%	0.003	0	0.0%	0.003	0	0.0%

Table 41: NLL, AIC, AICc, and BIC Results for QPTCCX. The green cells indicate the best-fit distribution type(s).

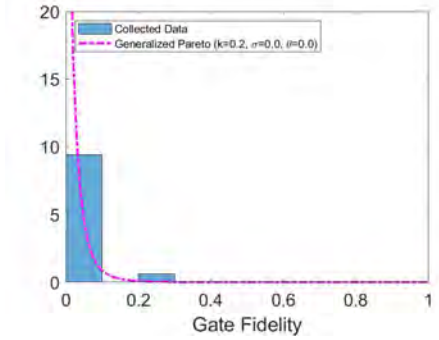
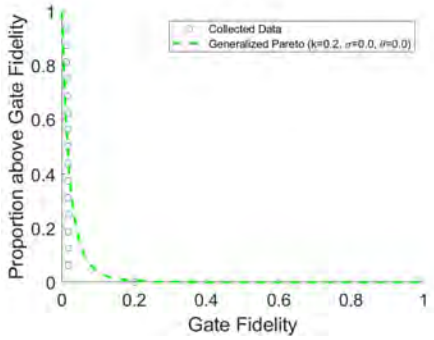
	Probability Density Function	Survival Function
QPTCCX		

Table 42: Distribution Model Analysis for ibmq_yorktown Q00_Q01_Q02 QPTCCX. In the left column, the probability density functions for the generalized Pareto distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival function for the generalized Pareto distribution model is plotted on top of the proportion of collected data observations above a given CCX-gate fidelity.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.008	0	0.0%	0.000	1	4.8%	0.092	1	4.8%	0.000	1	4.8%
Burr	–	0	0.0%	–	0	0.0%	–	0	0.0%	–	0	0.0%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.007	0	0.0%	0.000	0	0.0%	0.073	1	4.8%	0.000	0	0.0%
Gamma	0.008	0	0.0%	0.000	0	0.0%	0.092	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.059	3	14.3%	0.004	2	9.5%	0.001	0	0.0%	0.004	2	9.5%
Generalized Pareto	0.851	18	85.7%	0.994	18	85.7%	0.060	7	33.3%	0.994	18	85.7%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.008	0	0.0%	0.000	0	0.0%	0.092	7	33.3%	0.000	0	0.0%
Logistic	0.007	0	0.0%	0.000	0	0.0%	0.070	0	0.0%	0.000	0	0.0%
Loglogistic	0.007	0	0.0%	0.000	0	0.0%	0.075	0	0.0%	0.000	0	0.0%
Lognormal	0.008	0	0.0%	0.000	0	0.0%	0.087	0	0.0%	0.000	0	0.0%
Nakagami	0.008	0	0.0%	0.000	0	0.0%	0.095	1	4.8%	0.000	0	0.0%
Normal	0.007	0	0.0%	0.000	0	0.0%	0.085	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.007	1	4.8%	0.000	0	0.0%
Rician	0.008	0	0.0%	0.000	0	0.0%	0.090	0	0.0%	0.000	0	0.0%
t Location Scale	0.008	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Weibull	0.007	0	0.0%	0.000	0	0.0%	0.081	3	14.3%	0.000	0	0.0%

Table 43: NLL, AIC, AICc, and BIC Results for QPTCSWAP. The green cells indicate the best-fit distribution type(s).

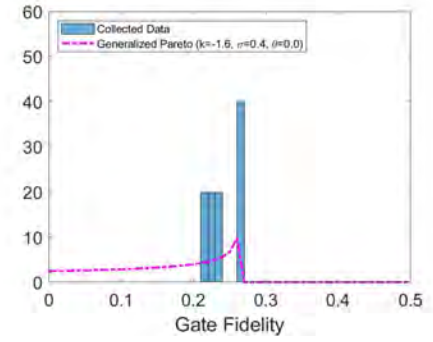
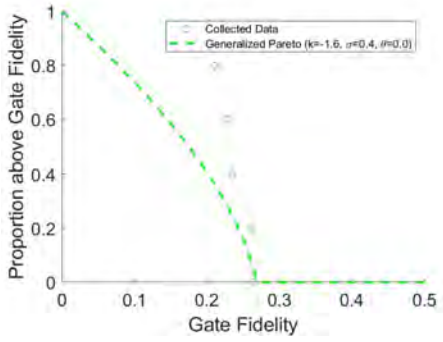
	Probability Density Function	Survival Function
QPTCSWAP		

Table 44: Distribution Model Analysis for ibmq_sydney Q00_Q01_Q02 QPTCSWAP. In the left column, the probability density functions for the generalized Pareto distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival function for the generalized Pareto distribution model is plotted on top of the proportion of collected data observations above a given CSWAP-gate fidelity.

Randomized Benchmarking As with quantum process tomography, distribution analysis on each of the randomized benchmarking characteristics uncovers interesting and surprising results for future investigation. Again, the main purpose of this section is to determine whether or not the randomized benchmarking characteristics are normally distributed. Results show that the randomized benchmarking characteristics are not normally distributed. A brief description of the results for each characteristic is provided along with notations for particularly interesting results and recommendations for future analysis.

For all the randomized benchmarking characteristics, significant investigation and consideration are needed in future work to understand why the characteristics are best represented by their respective best-fit distribution types. Many of the best-fit distribution types are related to reliability, survival rates, and failure rates. However, except for the fidelity and α characteristics, the analogous relationship between failure rates and the other randomized benchmarking characteristics is largely obscured. For

example, the a parameter does not directly describe survival/failure rate, but it just absorbs state preparation and measurement errors in determining the survival and failure rate.

For 1-Qubit Randomized Benchmarking Seed 0 Fidelity, the Burr, EV, GEV, and t location-scale distributions are identified as the best fit (see Tables 45 and 61). The k parameter of the GEV distribution is negative for 100% of the qubits sets, indicating the Type III Weibull distribution. It is interesting that the Burr distribution is indicated as a best-fit distribution, since MATLAB could not generate the Burr distribution for many of the qubits (ibmq_yorktown qubit Q03 is presented in Table 61 because MATLAB could not generate the Burr distribution for qubits Q00, Q01, or Q02). Future research should analyze individual qubits to determine which qubit properties cause some qubits to fit the Burr distribution very well, while the Burr distribution cannot be generated for others at all.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Burr	0.286	10	16.4%	0.407	1	1.6%	0.379	1	1.6%	0.211	0	0.0%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.146	4	6.6%	0.136	9	14.8%	0.150	11	18.0%	0.383	14	23.0%
Gamma	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.286	27	44.3%	0.215	22	36.1%	0.202	19	31.1%	0.126	16	26.2%
Generalized Pareto	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Logistic	0.015	0	0.0%	0.002	4	6.6%	0.002	4	6.6%	0.002	4	6.6%
Loglogistic	0.010	0	0.0%	0.001	0	0.0%	0.001	0	0.0%	0.001	0	0.0%
Lognormal	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.002	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.002	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.133	20	32.8%	0.074	19	31.1%	0.075	18	29.5%	0.054	18	29.5%
Weibull	0.115	0	0.0%	0.166	6	9.8%	0.191	8	13.1%	0.223	9	14.8%

Table 45: NLL, AIC, AICc, and BIC Results for 1-Qubit Randomized Benchmarking Seed 0 Fidelity. The green cells indicate the best-fit distribution type(s).

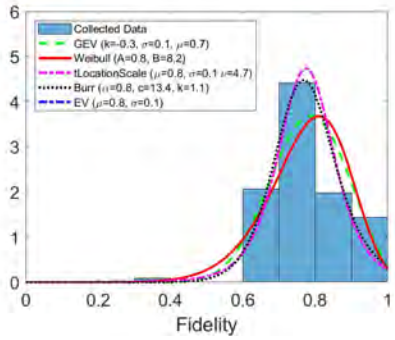
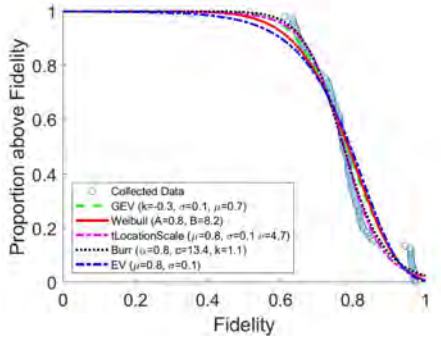
	Probability Density Function	Survival Function
Fidelity		

Table 46: Distribution Model Analysis for ibmq_yorktown Q03 1-Qubit Seed 0 Fidelity. In the left column, the probability density functions for the GEV, Weibull, t location-scale, Burr, and EV distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, Weibull, t location-scale, Burr, and EV distribution models are plotted on top of the proportion of collected data observations above a given fidelity.

For 1-Qubit Randomized Benchmarking Seed 0 a, the generalized Pareto distribution is identified as the best-fit distribution (see Tables 47–49). When referring to Table 13, however, both the Chi-square test and the KS test rejected the generalized Pareto distribution. Thus, although the generalized Pareto distribution is the best-fit distribution, it is not a good fit. While ibmq_yorktown Q00 in Table 48 does not provide any insight into why the generalized Pareto distribution is the best fit, evaluating ibmq_casablanca qubit Q00 does. In the probability distribution function in Table 49, clear visual evidence of a bimodal distribution exists for this qubit. At present, it is unclear what potentially causes the qubit's bimodal behavior. Given that only one qubit is involved, the presence or absence of physical links between qubits in qubit sets is not applicable. Future research should investigate what causes the bimodal behavior. Additionally, this experiment should be repeated for various types of bimodal distributions to determine the best-fit bimodal distribution types.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.022	0	0.0%	0.002	0	0.0%	0.002	0	0.0%	0.006	0	0.0%
Burr	0.210	8	13.1%	0.106	6	9.8%	0.096	6	9.8%	0.068	6	9.8%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Gamma	0.011	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.001	0	0.0%
Generalized Extreme Value	0.084	21	34.4%	0.019	13	21.3%	0.019	12	19.7%	0.009	10	16.4%
Generalized Pareto	0.608	30	49.2%	0.867	30	49.2%	0.876	30	49.2%	0.903	30	49.2%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.022	1	1.6%	0.002	6	9.8%	0.002	7	11.5%	0.006	9	14.8%
Logistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.015	0	0.0%	0.001	2	3.3%	0.001	2	3.3%	0.003	2	3.3%
Lognormal	0.023	0	0.0%	0.002	0	0.0%	0.003	0	0.0%	0.005	0	0.0%
Nakagami	0.002	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.001	1	1.6%	0.000	1	1.6%	0.000	1	1.6%	0.000	1	1.6%
Weibull	0.001	0	0.0%	0.000	3	4.9%	0.000	3	4.9%	0.000	3	4.9%

Table 47: NLL, AIC, AICc, and BIC Results for 1-Qubit Randomized Benchmarking Seed 0 a. The green cells indicate the best-fit distribution type(s).

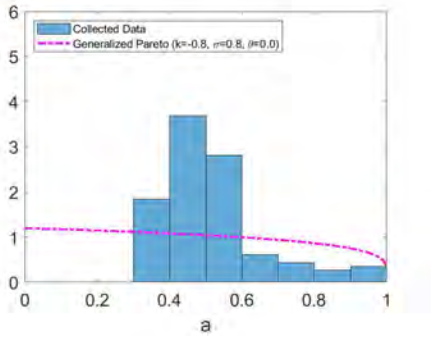
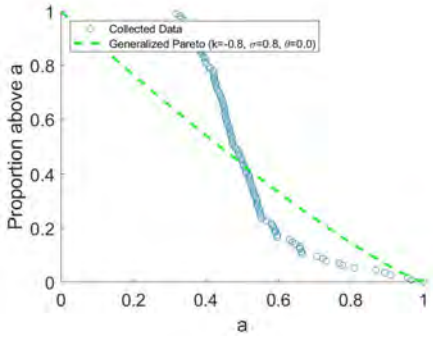
	Probability Density Function	Survival Function
a		

Table 48: Distribution Model Analysis for ibmq_yorktown Q00 1-Qubit Seed 0 a. In the left column, the probability density function for the generalized Pareto distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival function for the generalized Pareto distribution model is plotted on top of the proportion of collected data observations above a given a.

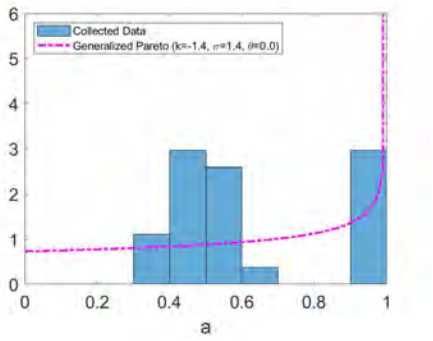
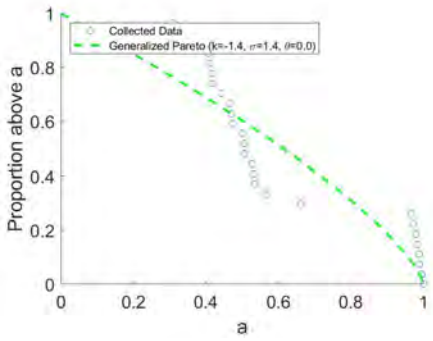
	Probability Density Function	Survival Function
a		

Table 49: Distribution Model Analysis for ibmq_casablanca Q00 1-Qubit Seed 0 a. In the left column, the probability density function for the generalized Pareto distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the generalized Pareto distribution model is plotted on top of the proportion of collected data observations above a given a.

For 1-Qubit Randomized Benchmarking Seed 0 α, the GEV distribution is identified as the best-fit distribution (see Tables 50 and 51). 98% of the qubits are Type III (Weibull). Looking at the probability density function for ibmq_yorktown qubit Q00 in Table 51, all values of α are very close to 1. Very little variability exists in

the α decay rate; as such, distinguishing between qubits based on this characteristic alone may prove difficult.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Burr	0.118	2	3.3%	0.081	0	0.0%	0.066	0	0.0%	0.048	0	0.0%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.003	1	1.6%	0.000	1	1.6%	0.000	1	1.6%	0.000	3	4.9%
Gamma	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.874	49	80.3%	0.919	47	77.0%	0.934	46	75.4%	0.952	43	70.5%
Generalized Pareto	0.000	1	1.6%	0.000	1	1.6%	0.000	1	1.6%	0.000	1	1.6%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Logistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Lognormal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.000	0	0.0%	0.000	1	1.6%	0.000	1	1.6%	0.000	2	3.3%
t Location Scale	0.002	8	13.1%	0.000	8	13.1%	0.000	8	13.1%	0.000	8	13.1%
Weibull	0.003	0	0.0%	0.000	3	4.9%	0.000	4	6.6%	0.000	4	6.6%

Table 50: NLL, AIC, AICc, and BIC Results for 1-Qubit Randomized Benchmarking Seed 0 α. The green cells indicate the best-fit distribution type(s).

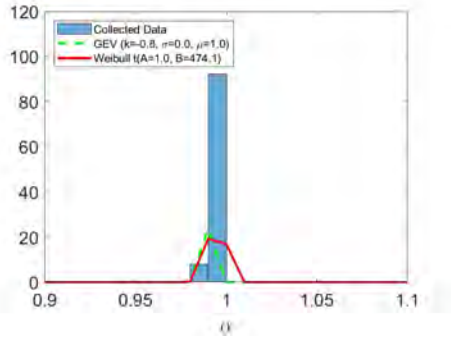
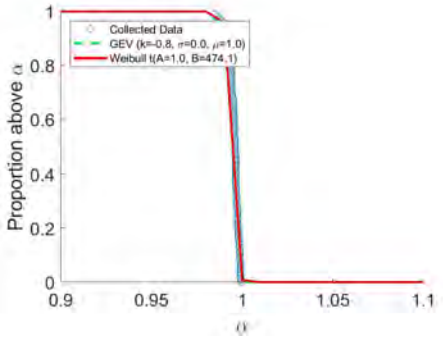
	Probability Density Function	Survival Function
α		

Table 51: Distribution Model Analysis for ibmq_yorktown Q00 1-Qubit Seed 0 α. In the left column, the probability density functions for the GEV and Weibull distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV and Weibull distribution models are plotted on top of the proportion of collected data observations above a given fidelity.

For 1-Qubit Randomized Benchmarking Seed 0 b, the GEV distribution is selected as the best-fit distribution (see Tables 52 and 53). 97% of the qubits are Type III (Weibull). However, MATLAB could not generate the Weibull distribution for many of the qubits, including ibmq_yorktown Q00 in Table 53, because many of the data points are equal to zero. For this reason, the NLL, AIC, etc. metrics for the Weibull distribution are very low although almost all the qubits are a Type III GEV distribution.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Burr	0.109	0	0.0%	0.071	0	0.0%	0.060	0	0.0%	0.052	0	0.0%
Exponential	0.000	2	3.3%	0.000	3	4.9%	0.000	3	4.9%	0.000	7	11.5%
Extreme Value	0.111	7	11.5%	0.071	10	16.4%	0.081	10	16.4%	0.238	14	23.0%
Gamma	0.002	0	0.0%	0.000	1	1.6%	0.000	1	1.6%	0.000	1	1.6%
Generalized Extreme Value	0.684	37	60.7%	0.806	32	52.5%	0.801	31	50.8%	0.635	27	44.3%
Generalized Pareto	0.002	2	3.3%	0.000	2	3.3%	0.000	2	3.3%	0.000	2	3.3%
Half-Normal	0.002	11	18.0%	0.000	11	18.0%	0.000	12	19.7%	0.000	8	13.1%
Inverse Gaussian	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Logistic	0.002	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.006	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.001	0	0.0%
Lognormal	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.003	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.002	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.005	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.011	2	3.3%	0.000	1	1.6%	0.000	1	1.6%	0.000	1	1.6%
Weibull	0.059	0	0.0%	0.052	1	1.6%	0.057	1	1.6%	0.073	1	1.6%

Table 52: NLL, AIC, AICc, and BIC Results for 1-Qubit Randomized Benchmarking Seed 0 b. The green cells indicate the best-fit distribution type(s).

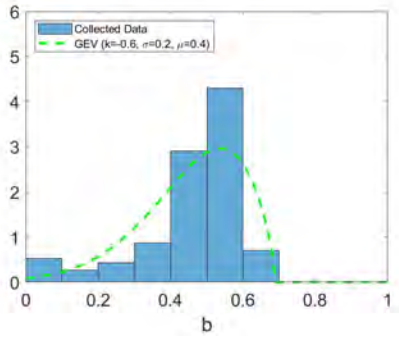
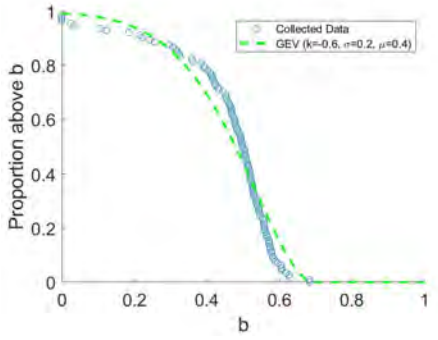
	Probability Density Function	Survival Function
b		

Table 53: Distribution Model Analysis for ibmq_yorktown Q00 1-Qubit Seed 0 b. In the left column, the probability density function for the GEV distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival function for the GEV distribution model is plotted on top of the proportion of collected data observations above a given b.

For 1-Qubit Randomized Benchmarking Seed 0 err_a, the Birnbaum-Saunders distribution is identified as the best-fit distribution type (see Tables 54 and 55). The Birnbaum-Saunders distribution is well-suited as a lifetime model for materials subject to cyclic patterns of stress and strain, where failure of the material usually results from the growth of a prominent flaw. The Birnbaum-Saunders distribution has the probability density function

$$f\left(x|\beta,\gamma\right) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{\left(\sqrt{\frac{x}{\beta}} - \sqrt{\frac{\beta}{x}}\right)^2}{2\gamma^2}\right]\left(\frac{\left(\sqrt{\frac{x}{\beta}} + \sqrt{\frac{\beta}{x}}\right)}{2\gamma x}\right), \tag{33}$$

where $\beta > 0$ is the scale parameter and $\gamma > 0$ is the shape parameter for $x > 0$. [7]

The relationship between the Birnbaum-Saunders distribution and the err_a parameter is not immediately clear, as the err_a parameter describes the error limits of the a parameter, which absorbs state preparation and measurement errors in randomized benchmarking. Instead, the α parameter has a more direct relationship to cyclic

patterns of stress and strain. α measures the decay rate, where each Clifford gate can be considered a cycle of stress. More investigation is necessary to determine the relationship between the err_a parameter and the Birnbaum-Saunders distribution.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.346	30	49.2%	0.662	29	47.5%	0.670	29	47.5%	0.662	29	47.5%
Burr	0.105	6	9.8%	0.031	4	6.6%	0.025	3	4.9%	0.012	3	4.9%
Exponential	0.005	0	0.0%	0.000	3	4.9%	0.000	3	4.9%	0.001	3	4.9%
Extreme Value	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Gamma	0.019	1	1.6%	0.001	0	0.0%	0.001	0	0.0%	0.001	0	0.0%
Generalized Extreme Value	0.137	9	14.8%	0.045	3	4.9%	0.040	3	4.9%	0.016	3	4.9%
Generalized Pareto	0.018	0	0.0%	0.001	0	0.0%	0.001	0	0.0%	0.000	0	0.0%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.198	11	18.0%	0.201	17	27.9%	0.203	17	27.9%	0.249	17	27.9%
Logistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.044	0	0.0%	0.012	1	1.6%	0.013	2	3.3%	0.012	2	3.3%
Lognormal	0.106	1	1.6%	0.043	1	1.6%	0.043	1	1.6%	0.042	1	1.6%
Nakagami	0.004	1	1.6%	0.000	1	1.6%	0.000	1	1.6%	0.000	1	1.6%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.000	1	1.6%	0.000	1	1.6%	0.000	1	1.6%	0.000	1	1.6%
Weibull	0.020	1	1.6%	0.003	1	1.6%	0.003	1	1.6%	0.003	1	1.6%

Table 54: NLL, AIC, AICc, and BIC Results for 1-Qubit Randomized Benchmarking Seed 0 *err_a*. The green cells indicate the best-fit distribution type(s).

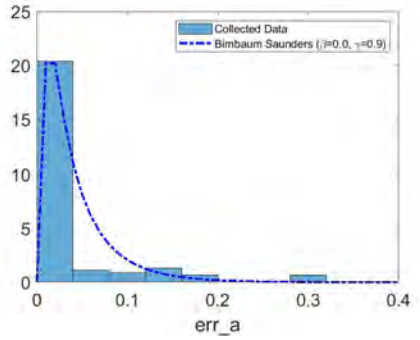
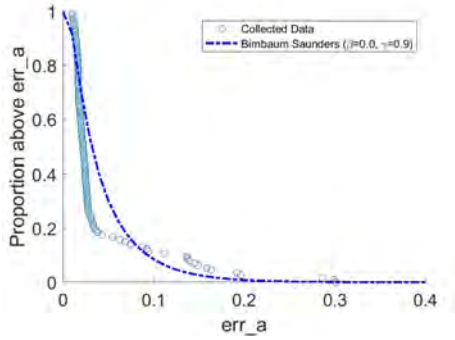
	Probability Density Function	Survival Function
err_a		

Table 55: Distribution Model Analysis for ibmq_yorktown Q00 1-Qubit Seed 0 *err_a*. In the left column, the probability density function for the Birnbaum Saunders distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival function for the Birnbaum Saunders distribution model is plotted on top of the proportion of collected data observations above a given *err_a*.

For 1-Qubit Randomized Benchmarking Seed 0 err_α, the Burr, GEV, and loglogistic distributions are identified as the best-fit distribution types (see Tables 56 and 57). For the GEV distribution, 72% of the qubits are Type II (Fréchet). The probability density function for the loglogistic distribution is given by the expression,

$$f(x|\mu,\sigma) = \frac{1}{\sigma}\frac{1}{x}\frac{e^z}{(1+e^z)^2}, \tag{34}$$

where $x \geq 0$ and $z = \frac{log(x)-\mu}{\sigma}$. $\mu > 0$ is the mean of logarithmic values and $\sigma > 0$ is the scale parameter of logarithmic values. The loglogistic distribution is related to the Burr distribution when the k parameter of the Burr distribution equals one. For the qubit analyzed in Table 57, the k parameter is somewhat close to one at $k = 0.6$.

The loglogistic distribution is often used in survival analysis to model events with an initial rate increase followed by a rate decrease [25]. Because err_α describes the error bounds on the decay rate α, it is not immediately clear how err_α is analogous to an event with an initial rate increase followed by a rate decrease. It is also presently

unclear how err_α is analogous to the failure of a critical non-negative variable of the Fréchet distribution. More investigation is required.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.065	0	0.0%	0.053	1	1.6%	0.056	1	1.6%	0.083	1	1.6%
Burr	0.270	20	32.8%	0.306	11	18.0%	0.280	11	18.0%	0.187	10	16.4%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Gamma	0.027	0	0.0%	0.009	0	0.0%	0.010	0	0.0%	0.014	0	0.0%
Generalized Extreme Value	0.262	31	50.8%	0.311	17	27.9%	0.287	17	27.9%	0.179	11	18.0%
Generalized Pareto	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.076	1	1.6%	0.069	6	9.8%	0.072	6	9.8%	0.127	10	16.4%
Logistic	0.005	0	0.0%	0.000	3	4.9%	0.000	3	4.9%	0.001	4	6.6%
Loglogistic	0.107	0	0.0%	0.142	10	16.4%	0.169	10	16.4%	0.244	13	21.3%
Lognormal	0.085	1	1.6%	0.076	2	3.3%	0.095	2	3.3%	0.144	2	3.3%
Nakagami	0.005	0	0.0%	0.000	1	1.6%	0.000	1	1.6%	0.000	1	1.6%
Normal	0.001	0	0.0%	0.000	2	3.3%	0.000	2	3.3%	0.000	2	3.3%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.006	0	0.0%	0.000	0	0.0%	0.001	0	0.0%	0.001	0	0.0%
t Location Scale	0.090	8	13.1%	0.033	8	13.1%	0.030	8	13.1%	0.020	7	11.5%
Weibull	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%

Table 56: NLL, AIC, AICc, and BIC Results for 1-Qubit Randomized Benchmarking Seed 0 err_α. The green cells indicate the best-fit distribution type(s).

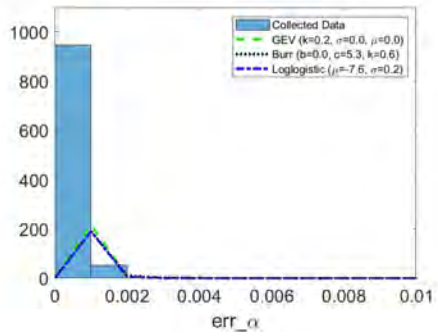
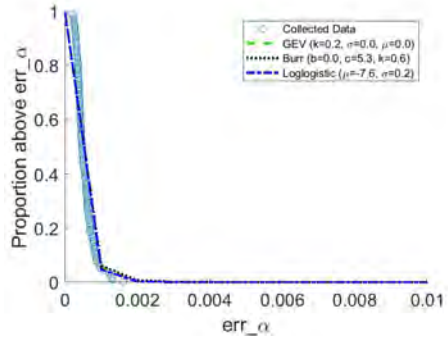
	Probability Density Function	Survival Function
err_α		

Table 57: Distribution Model Analysis for ibmq_yorktown Q00 1-Qubit Seed 0 err_α. In the left column, the probability density functions for the GEV, Burr, and loglogistic distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, Burr, and loglogistic distribution models are plotted on top of the proportion of collected data observations above a given err_α.

For 1-Qubit Randomized Benchmarking Seed 0 err_b, the Birnbaum-Saunders distribution is identified as the best-fit distribution (see Tables 58 and 59). Similar to the err_a parameter, the err_b parameter describes the error limits of the b parameter, which absorbs state preparation and measurement errors in randomized benchmarking. As a result, the relationship between the Birnbaum-Saunders distribution and the err_b parameter is presently unclear.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.331	31	50.8%	0.734	30	49.2%	0.740	30	49.2%	0.745	30	49.2%
Burr	0.120	8	13.1%	0.048	3	4.9%	0.043	3	4.9%	0.017	2	3.3%
Exponential	0.006	0	0.0%	0.000	3	4.9%	0.000	3	4.9%	0.001	4	6.6%
Extreme Value	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Gamma	0.032	3	4.9%	0.004	1	1.6%	0.004	1	1.6%	0.004	0	0.0%
Generalized Extreme Value	0.110	7	11.5%	0.028	5	8.2%	0.026	4	6.6%	0.010	4	6.6%
Generalized Pareto	0.032	1	1.6%	0.002	1	1.6%	0.001	1	1.6%	0.001	1	1.6%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.150	6	9.8%	0.076	8	13.1%	0.077	9	14.8%	0.094	10	16.4%
Logistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.060	0	0.0%	0.022	4	6.6%	0.022	4	6.6%	0.022	4	6.6%
Lognormal	0.123	2	3.3%	0.081	3	4.9%	0.081	3	4.9%	0.100	3	4.9%
Nakagami	0.006	3	4.9%	0.000	3	4.9%	0.000	3	4.9%	0.000	3	4.9%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Weibull	0.031	0	0.0%	0.005	0	0.0%	0.005	0	0.0%	0.005	0	0.0%

Table 58: NLL, AIC, AICc, and BIC Results for 1-Qubit Randomized Benchmarking Seed 0 *err_b*. The green cells indicate the best-fit distribution type(s).

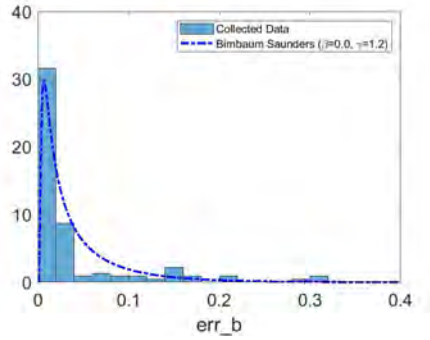
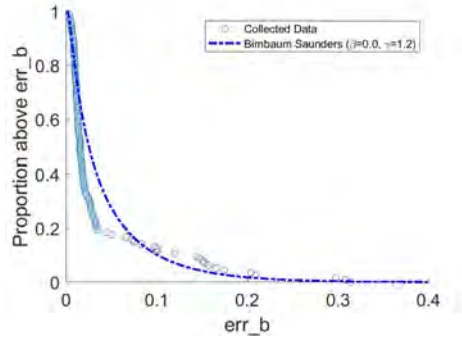
	Probability Density Function	Survival Function
err_b		

Table 59: Distribution Model Analysis for ibmq_yorktown Q00 1-Qubit Seed 0 *err_b*. In the left column, the probability density functions for the Birnbaum-Saunders distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the Birnbaum-Saunders distribution model is plotted on top of the proportion of collected data observations above a given *err_b*.

By comparison to Seed 0, results for 1-Qubit Randomized Benchmarking Seed 4 Fidelity are far more consistent. The GEV distribution is identified as the best-fit distribution type, with 100% of the qubits falling under the Type III case (Weibull) (see Tables 60 and 61). Given that the Weibull distribution is often used to model reliability data, it is unsurprising that fidelity, which describes the reliability of a sequence of Clifford gates, can be modeled by the Weibull distribution.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Burr	0.092	0	0.0%	0.031	0	0.0%	0.025	0	0.0%	0.022	0	0.0%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.069	7	15.6%	0.026	9	20.0%	0.029	10	22.2%	0.041	11	24.4%
Gamma	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.678	32	71.1%	0.914	26	57.8%	0.908	25	55.6%	0.888	24	53.3%
Generalized Pareto	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.000	0	0.0%	0.000	2	4.4%	0.000	2	4.4%	0.000	2	4.4%
Logistic	0.005	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.005	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Lognormal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.092	6	13.3%	0.010	6	13.3%	0.009	6	13.3%	0.011	6	13.3%
Weibull	0.054	0	0.0%	0.019	2	4.4%	0.027	2	4.4%	0.039	2	4.4%

Table 60: NLL, AIC, AICc, and BIC Results for 1-Qubit Randomized Benchmarking Seed 4 Fidelity. The green cells indicate the best-fit distribution type(s).

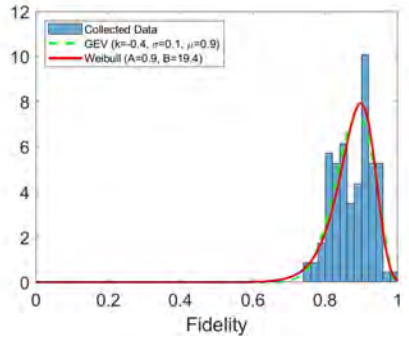
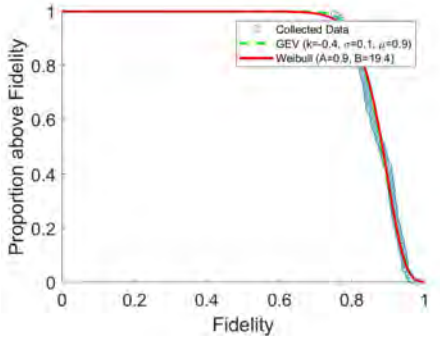
	Probability Density Function	Survival Function
Fidelity		

Table 61: Distribution Model Analysis for ibmq_yorktown Q00 1-Qubit Seed 4 Fidelity. In the left column, the probability density functions for the GEV and Weibull distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV and Weibull distribution models are plotted on top of the proportion of collected data observations above a given fidelity.

For 1-Qubit Randomized Benchmarking Seed 4 *a*, the bimodal behavior observed for Seed 0 appears to vanish (see Tables 62 and 63). This is somewhat surprising, as initial intuition would lead one to believe that the additional seeds of data would provide more definition to the two modes. Once the reason for the bimodal behavior observed with Seed 0 has been determined, the disappearance of the bimodal behavior for Seed 4 should be investigated.

Instead of the generalized Pareto distribution emerging as the best fit (albeit a terrible fit), the GEV distribution is identified as the best fit by all metrics except the normalized median value for negative log-likelihood. 70% of the qubits are the Type II (Fréchet) GEV distribution.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.068	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Burr	0.340	0	0.0%	0.031	0	0.0%	0.025	0	0.0%	0.022	0	0.0%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.000	7	15.6%	0.026	9	20.0%	0.029	10	22.2%	0.041	11	24.4%
Gamma	0.033	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.232	32	71.1%	0.914	26	57.8%	0.908	25	55.6%	0.888	24	53.3%
Generalized Pareto	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.069	0	0.0%	0.000	2	4.4%	0.000	2	4.4%	0.000	2	4.4%
Logistic	0.015	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.081	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Lognormal	0.068	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.019	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.007	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.010	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.049	6	13.3%	0.010	6	13.3%	0.009	6	13.3%	0.011	6	13.3%
Weibull	0.006	0	0.0%	0.019	2	4.4%	0.027	2	4.4%	0.039	2	4.4%

Table 62: NLL, AIC, AICc, and BIC Results for 1-Qubit Randomized Benchmarking Seed 4 a. The green cells indicate the best-fit distribution type(s).

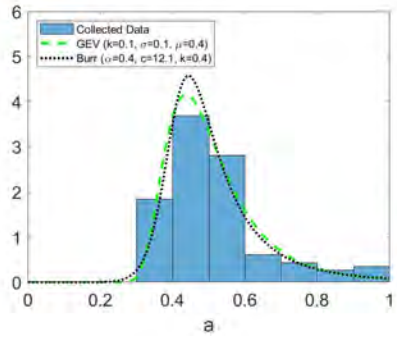
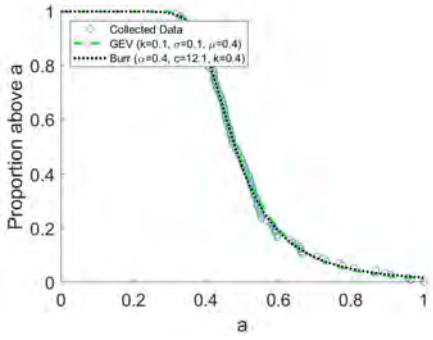
	Probability Density Function	Survival Function
a		

Table 63: Distribution Model Analysis for ibmq_yorktown Q00 1-Qubit Seed 4 a. In the left column, the probability density functions for the GEV and Burr distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV and Burr distribution models are plotted on top of the proportion of collected data observations above a given a.

For 1-Qubit Randomized Benchmarking Seed 4 α, results do not change significantly from Seed 0. The GEV distribution emerged again as the best-fit distribution, with 100% of the qubits falling under the Type III (Weibull) distribution (see Tables 64 and 65). The same is true for 1-Qubit Randomized Benchmarking Seed 4 b, that the results do not change significantly from Seed 0. All metrics except the BIC normalized median value identified the GEV distribution as the best-fit distribution (the EV distribution is selected by the BIC normalized median value), with 97% of the qubits falling under the Type III (Weibull) distribution (see Tables 66 and 67). Again, many of the data points are equal to zero, which prevents the MATLAB algorithm from generating the Weibull distribution.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.008	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Burr	0.109	2	4.4%	0.066	0	0.0%	0.058	0	0.0%	0.051	0	0.0%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.118	3	6.7%	0.104	7	15.6%	0.125	9	20.0%	0.163	9	20.0%
Gamma	0.008	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.434	32	71.1%	0.685	22	48.9%	0.648	20	44.4%	0.574	19	42.2%
Generalized Pareto	0.000	1	2.2%	0.000	1	2.2%	0.000	1	2.2%	0.000	1	2.2%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.008	0	0.0%	0.000	1	2.2%	0.000	1	2.2%	0.000	1	2.2%
Logistic	0.029	0	0.0%	0.004	1	2.2%	0.006	1	2.2%	0.009	1	2.2%
Loglogistic	0.029	0	0.0%	0.004	0	0.0%	0.006	0	0.0%	0.009	0	0.0%
Lognormal	0.008	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.008	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.008	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.008	0	0.0%	0.000	4	8.9%	0.000	4	8.9%	0.000	4	8.9%
t Location Scale	0.108	7	15.6%	0.033	7	15.6%	0.030	7	15.6%	0.029	7	15.6%
Weibull	0.118	0	0.0%	0.103	2	4.4%	0.126	2	4.4%	0.163	3	6.7%

Table 64: NLL, AIC, AICc, and BIC Results for 1-Qubit Randomized Benchmarking Seed 4 α. The green cells indicate the best-fit distribution type(s).

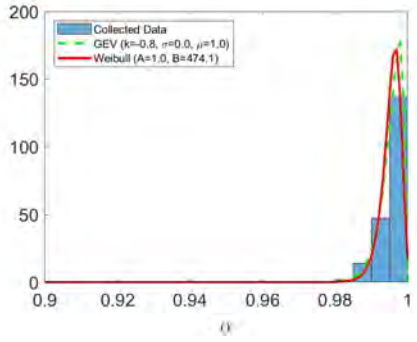
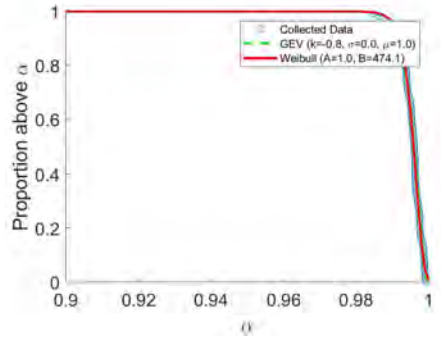
	Probability Density Function	Survival Function
α		

Table 65: Distribution Model Analysis for ibmq_yorktown Q00 1-Qubit Seed 4 α. In the left column, the probability density functions for the GEV and Weibull distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV and Weibull distribution models are plotted on top of the proportion of collected data observations above a given α.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.000	1	2.2%	0.000	1	2.2%	0.000	1	2.2%	0.000	1	2.2%
Burr	0.074	0	0.0%	0.052	0	0.0%	0.051	0	0.0%	0.039	0	0.0%
Exponential	0.000	0	0.0%	0.000	1	2.2%	0.000	1	2.2%	0.000	1	2.2%
Extreme Value	0.155	4	8.9%	0.208	10	22.2%	0.291	11	24.4%	0.413	13	28.9%
Gamma	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.425	26	57.8%	0.577	17	37.8%	0.507	16	35.6%	0.403	14	31.1%
Generalized Pareto	0.217	8	17.8%	0.137	8	17.8%	0.122	8	17.8%	0.108	8	17.8%
Half-Normal	0.000	4	8.9%	0.000	3	6.7%	0.000	3	6.7%	0.000	3	6.7%
Inverse Gaussian	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Logistic	0.022	0	0.0%	0.003	0	0.0%	0.004	0	0.0%	0.006	0	0.0%
Loglogistic	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Lognormal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.002	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.011	0	0.0%	0.001	1	2.2%	0.002	1	2.2%	0.002	1	2.2%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.007	0	0.0%	0.000	1	2.2%	0.001	1	2.2%	0.001	1	2.2%
t Location Scale	0.054	1	2.2%	0.010	1	2.2%	0.009	1	2.2%	0.006	1	2.2%
Weibull	0.029	1	2.2%	0.011	2	4.4%	0.013	2	4.4%	0.021	2	4.4%

Table 66: NLL, AIC, AICc, and BIC Results for 1-Qubit Randomized Benchmarking Seed 4 *b*. The green cells indicate the best-fit distribution type(s).

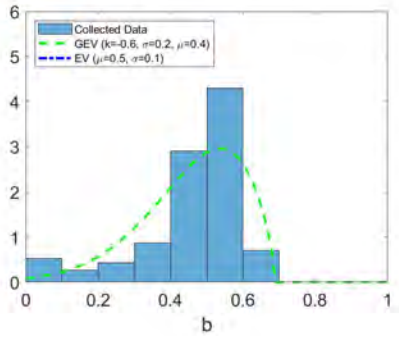
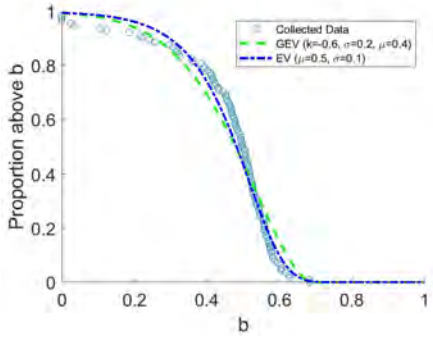
	Probability Density Function	Survival Function
b		

Table 67: Distribution Model Analysis for ibmq_yorktown Q00 1-Qubit Seed 4 b. In the left column, the probability density functions for the GEV and EV distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV and EV distribution models are plotted on top of the proportion of collected data observations above a given b.

For 1-Qubit Randomized Benchmarking Seed 4 err_a, results are far less consistent than Seed 0. This is surprising, since initial intuition would cause one to expect the additional data included in Seed 4 to lead to more consistent results than Seed 0. The Burr, inverse Gaussian, loglogistic, and lognormal distributions are selected as best-fit distributions (see Tables 68 and 69). The probability density function of the inverse Gaussian distribution is given by the expression,

$$f(x|\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left(-\frac{\lambda}{2\mu^2 x}(x-\mu)^2\right), \tag{35}$$

where $\mu > 0$ is the scale parameter, and $\lambda > 0$ is the shape parameter.

The inverse Gaussian distribution models non-negative positively skewed data [21]. Accordingly, the inverse Gaussian appropriately models err_a, which describes the error limit for the a parameter. The error limit is reported as non-negative data, and the majority of the error limit data points are close to zero, with few extending much beyond zero, as observed in the probability density function in Table 69.

The probability density function of the lognormal distribution is given by the expression,

$$f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\frac{-(\log(t)-\mu)^2}{2\sigma^2}\right), \tag{36}$$

where for $x > 0$, μ is the mean of logarithmic values, and $\sigma \geq 0$ is the standard deviation of logarithmic values. The logarithm of the lognormal distribution has a normal distribution [26]. Like the inverse Gaussian distribution, the lognormal distribution models non-negative positively skewed data.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.107	6	13.3%	0.153	6	13.3%	0.163	6	13.3%	0.163	5	11.1%
Burr	0.175	21	46.7%	0.143	4	8.9%	0.110	4	8.9%	0.085	3	6.7%
Exponential	0.022	0	0.0%	0.015	7	15.6%	0.019	7	15.6%	0.043	9	20.0%
Extreme Value	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Gamma	0.053	1	2.2%	0.037	2	4.4%	0.039	2	4.4%	0.039	1	2.2%
Generalized Extreme Value	0.170	7	15.6%	0.113	0	0.0%	0.093	0	0.0%	0.064	0	0.0%
Generalized Pareto	0.045	1	2.2%	0.009	1	2.2%	0.007	1	2.2%	0.005	1	2.2%
Half-Normal	0.005	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.119	5	11.1%	0.175	10	22.2%	0.183	10	22.2%	0.201	11	24.4%
Logistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.107	0	0.0%	0.143	11	24.4%	0.162	11	24.4%	0.170	11	24.4%
Lognormal	0.123	2	4.4%	0.176	1	2.2%	0.187	1	2.2%	0.192	1	2.2%
Nakagami	0.018	0	0.0%	0.004	0	0.0%	0.004	0	0.0%	0.004	0	0.0%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	1	2.2%	0.000	1	2.2%	0.000	2	4.4%
Rician	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.006	1	2.2%	0.000	1	2.2%	0.000	1	2.2%	0.000	1	2.2%
Weibull	0.049	1	2.2%	0.031	1	2.2%	0.033	1	2.2%	0.034	0	0.0%

Table 68: NLL, AIC, AICc, and BIC Results for 1-Qubit Randomized Benchmarking Seed 4 *err_a*. The green cells indicate the best-fit distribution type(s).

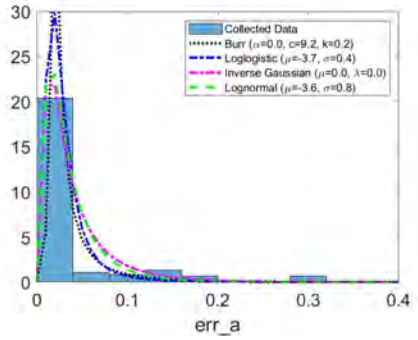
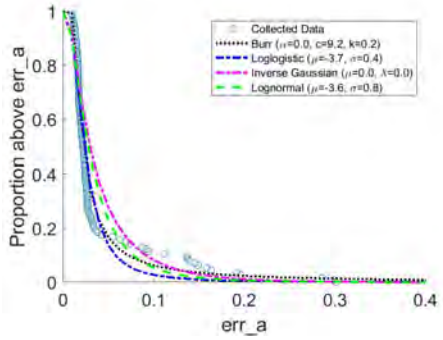
	Probability Density Function	Survival Function
err_a		

Table 69: Distribution Model Analysis for ibmq_yorktown Q00 1-Qubit Seed 4 *err_a*. In the left column, the probability density functions for the Burr, loglogistic, inverse Gaussian, and lognormal distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the Burr, loglogistic, inverse Gaussian, and lognormal distribution models are plotted on top of the proportion of collected data observations above a given *err_a*.

For 1-Qubit Randomized Benchmarking Seed 4 *err_α*, the results differ from Seed 0. The Birnbaum Saunders, GEV, and inverse Gaussian distributions are identified as the best-fit distribution types (see Tables 70 and 71). For the GEV distributions, 72% of the qubits are identified as Type II (Fréchet). The relationship between the *err_α* characteristic and the Birnbaum Saunders and Fréchet distributions are presently unclear. However, the probability distribution in Table 71 shows that the data points are non-negative and positively skewed. Thus, it makes sense that the inverse Gaussian distribution is selected as a best-fit distribution type.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.097	0	0.0%	0.156	3	6.7%	0.161	3	6.7%	0.170	3	6.7%
Burr	0.118	6	13.3%	0.073	6	13.3%	0.063	6	13.3%	0.040	6	13.3%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.001	0	0.0%	0.000	0	0.0%	0.000	1	2.2%	0.000	1	2.2%
Gamma	0.091	0	0.0%	0.115	0	0.0%	0.119	0	0.0%	0.130	0	0.0%
Generalized Extreme Value	0.153	30	66.7%	0.112	5	11.1%	0.098	5	11.1%	0.065	5	11.1%
Generalized Pareto	0.000	2	4.4%	0.000	2	4.4%	0.000	1	2.2%	0.000	1	2.2%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.097	1	2.2%	0.155	18	40.0%	0.160	18	40.0%	0.170	18	40.0%
Logistic	0.040	0	0.0%	0.020	0	0.0%	0.021	0	0.0%	0.024	0	0.0%
Loglogistic	0.083	0	0.0%	0.098	1	2.2%	0.104	1	2.2%	0.115	1	2.2%
Lognormal	0.096	0	0.0%	0.153	1	2.2%	0.157	1	2.2%	0.167	1	2.2%
Nakagami	0.061	1	2.2%	0.049	2	4.4%	0.050	2	4.4%	0.057	2	4.4%
Normal	0.033	0	0.0%	0.015	0	0.0%	0.016	0	0.0%	0.017	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.038	0	0.0%	0.020	2	4.4%	0.021	2	4.4%	0.023	2	4.4%
t Location Scale	0.074	5	11.1%	0.028	4	8.9%	0.025	4	8.9%	0.015	4	8.9%
Weibull	0.019	0	0.0%	0.005	1	2.2%	0.005	1	2.2%	0.006	1	2.2%

Table 70: NLL, AIC, AICc, and BIC Results for 1-Qubit Randomized Benchmarking Seed 4 err_α. The green cells indicate the best-fit distribution type(s).

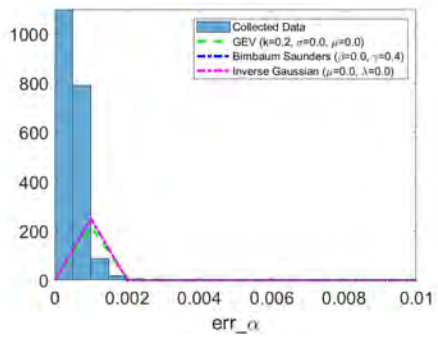
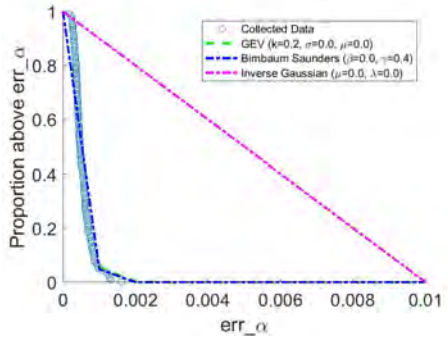
	Probability Density Function	Survival Function
err_α		

Table 71: Distribution Model Analysis for ibmq_yorktown Q00 1-Qubit Seed 4 err_α. In the left column, the probability density functions for the GEV, Birnbaum Saunders, and inverse Gaussian distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, Birnbaum Saunders, and inverse Gaussian distribution models are plotted on top of the proportion of collected data observations above a given err_α.

For 1-Qubit Randomized Benchmarking Seed 4 err_b, results differ from Seed 0. The Burr distribution and the loglogistic distributions are selected as the best-fit distribution types (see Tables 72 and 73). Both are well-suited for the non-negative positively skewed data observed in Table 73.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.109	8	17.8%	0.143	8	17.8%	0.157	8	17.8%	0.184	8	17.8%
Burr	0.188	20	44.4%	0.163	4	8.9%	0.128	4	8.9%	0.082	2	4.4%
Exponential	0.026	0	0.0%	0.018	8	17.8%	0.023	9	20.0%	0.046	9	20.0%
Extreme Value	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Gamma	0.071	2	4.4%	0.054	2	4.4%	0.058	2	4.4%	0.058	2	4.4%
Generalized Extreme Value	0.145	9	20.0%	0.105	0	0.0%	0.077	0	0.0%	0.060	0	0.0%
Generalized Pareto	0.045	1	2.2%	0.008	1	2.2%	0.007	1	2.2%	0.004	1	2.2%
Half-Normal	0.005	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.079	2	4.4%	0.091	6	13.3%	0.097	6	13.3%	0.097	7	15.6%
Logistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.118	0	0.0%	0.195	12	26.7%	0.211	12	26.7%	0.216	13	28.9%
Lognormal	0.124	2	4.4%	0.167	2	4.4%	0.181	1	2.2%	0.186	1	2.2%
Nakagami	0.024	0	0.0%	0.007	0	0.0%	0.008	0	0.0%	0.008	0	0.0%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	2	4.4%	0.000	2	4.4%	0.000	2	4.4%
Rician	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.005	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Weibull	0.059	1	2.2%	0.047	0	0.0%	0.052	0	0.0%	0.056	0	0.0%

Table 72: NLL, AIC, AICc, and BIC Results for 1-Qubit Randomized Benchmarking Seed 4 *err_b*. The green cells indicate the best-fit distribution type(s).

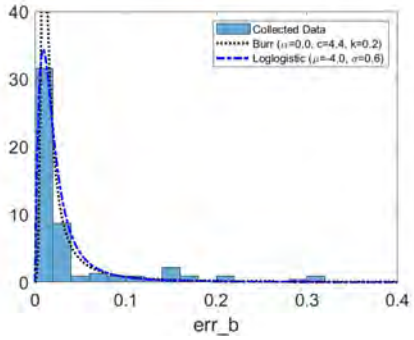
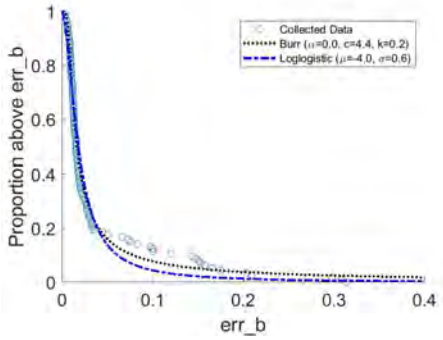
	Probability Density Function	Survival Function
err_b		

Table 73: Distribution Model Analysis for ibmq_yorktown Q00 1-Qubit Seed 4 *err_b*. In the left column, the probability density functions for the Burr and loglogistic distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the Burr and loglogistic distribution models are plotted on top of the proportion of collected data observations above a given *err_b*.

For 2-Qubit Randomized Benchmarking Seed 0 Fidelity, the generalized Pareto distribution is selected as the best-fit distribution (see Tables 74–76). However, observing the probability distribution function in Table 75, the generalized Pareto distribution fits the data poorly, which explains why the KS test p-value rejected this distribution type (the Chi-square test failed to make a determination) in Table 17. The generalized Pareto distribution usually indicates bimodal or multi-modal behavior in this work, which can be observed for the qubit pair ibmq_rome Q00_Q01 in Table 76.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.018	0	0.0%	0.004	1	0.5%	0.007	2	0.9%	0.006	2	0.9%
Burr	0.126	13	5.9%	0.135	1	0.5%	0.123	1	0.5%	0.088	1	0.5%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.073	12	5.5%	0.086	36	16.4%	0.134	40	18.2%	0.130	45	20.5%
Gamma	0.024	0	0.0%	0.007	1	0.5%	0.012	1	0.5%	0.010	1	0.5%
Generalized Extreme Value	0.156	77	35.0%	0.161	36	16.4%	0.126	31	14.1%	0.115	26	11.8%
Generalized Pareto	0.226	105	47.7%	0.379	104	47.3%	0.253	98	44.5%	0.338	100	45.5%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.018	0	0.0%	0.004	6	2.7%	0.007	10	4.5%	0.006	8	3.6%
Logistic	0.052	0	0.0%	0.036	6	2.7%	0.060	6	2.7%	0.050	6	2.7%
Loglogistic	0.030	0	0.0%	0.011	4	1.8%	0.018	5	2.3%	0.017	4	1.8%
Lognormal	0.018	0	0.0%	0.004	0	0.0%	0.007	0	0.0%	0.006	0	0.0%
Nakagami	0.030	0	0.0%	0.012	1	0.5%	0.019	1	0.5%	0.016	1	0.5%
Normal	0.040	0	0.0%	0.021	0	0.0%	0.035	0	0.0%	0.027	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.040	0	0.0%	0.021	8	3.6%	0.035	8	3.6%	0.027	8	3.6%
t Location Scale	0.071	12	5.5%	0.029	4	1.8%	0.025	2	0.9%	0.026	2	0.9%
Weibull	0.078	1	0.5%	0.091	12	5.5%	0.139	15	6.8%	0.137	16	7.3%

Table 74: NLL, AIC, AICc, and BIC Results for 2-Qubit Randomized Benchmarking Seed 0 Fidelity. The green cells indicate the best-fit distribution type(s).

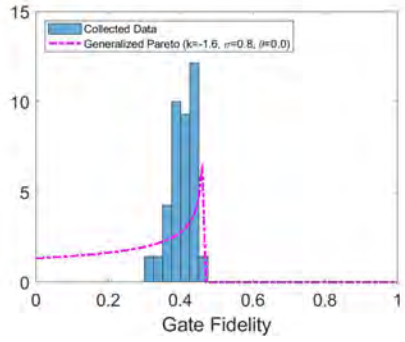
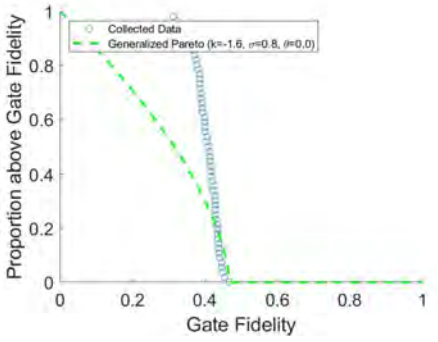
	Probability Density Function	Survival Function
Fidelity		

Table 75: Distribution Model Analysis for ibmq_yorktown Q00_Q01 2-Qubit Seed 0 Fidelity. In the left column, the probability density functions for the generalized Pareto distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival function for the generalized Pareto distribution model is plotted on top of the proportion of collected data observations above a given fidelity.

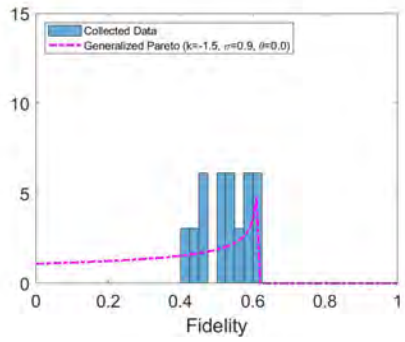
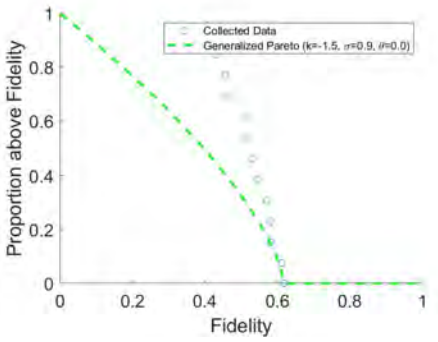
	Probability Density Function	Survival Function
Fidelity		

Table 76: Distribution Model Analysis for ibmq_rome Q00_Q01 2-Qubit Seed 0 Fidelity. In the left column, the probability density functions for the generalized Pareto distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival function for the generalized Pareto distribution model is plotted on top of the proportion of collected data observations above a given fidelity.

For 2-Qubit Randomized Benchmarking Seed 0 a, the EV, GEV, and logistic distributions are selected as the best-fit distribution types (see Tables 77 and 78). For the GEV distribution, 89% of the qubits are identified as Type III (Weibull). This

differs significantly from the distributions for 1-Qubit Randomized Benchmarking Seed 0 a, as the bimodal behavior appears to have vanished. Future work should investigate the reason for this change.

The probability density function for the logistic distribution is given by the expression,

$$f(x|\mu,\sigma) = \frac{\exp(\frac{x-\mu}{\sigma})}{\sigma\left(1+\exp\left(\frac{x-\mu}{\sigma}\right)\right)^2}, \tag{37}$$

where μ is the mean and $\sigma \geq 0$ is the scale parameter. Compared to the normal distribution, the logistic distribution has fatter tails and a higher kurtosis [24]. It is curious that the 2-Qubit Randomized Benchmarking Seed 0 a characteristic shifts towards a more normal distribution than the 1-qubit counterpart. Future work should investigate the reason for this shift.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.054	0	0.0%	0.052	2	1.9%	0.056	2	1.9%	0.062	2	1.9%
Burr	0.119	18	17.0%	0.108	7	6.6%	0.080	7	6.6%	0.059	6	5.7%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.056	2	1.9%	0.053	13	12.3%	0.058	15	14.2%	0.057	17	16.0%
Gamma	0.052	0	0.0%	0.056	2	1.9%	0.063	2	1.9%	0.067	2	1.9%
Generalized Extreme Value	0.128	59	55.7%	0.105	19	17.9%	0.087	14	13.2%	0.071	12	11.3%
Generalized Pareto	0.000	3	2.8%	0.000	3	2.8%	0.000	3	2.8%	0.000	3	2.8%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.054	0	0.0%	0.052	8	7.5%	0.056	10	9.4%	0.062	10	9.4%
Logistic	0.077	0	0.0%	0.104	7	6.6%	0.112	8	7.5%	0.118	9	8.5%
Loglogistic	0.074	0	0.0%	0.096	6	5.7%	0.104	7	6.6%	0.111	9	8.5%
Lognormal	0.054	0	0.0%	0.052	0	0.0%	0.055	0	0.0%	0.062	0	0.0%
Nakagami	0.058	0	0.0%	0.061	2	1.9%	0.066	2	1.9%	0.072	2	1.9%
Normal	0.061	0	0.0%	0.066	0	0.0%	0.072	0	0.0%	0.074	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.061	0	0.0%	0.066	10	9.4%	0.073	11	10.4%	0.075	11	10.4%
t Location Scale	0.094	23	21.7%	0.067	14	13.2%	0.050	12	11.3%	0.040	10	9.4%
Weibull	0.056	1	0.9%	0.064	13	12.3%	0.068	13	12.3%	0.069	13	12.3%

Table 77: NLL, AIC, AICc, and BIC Results for 2-Qubit Randomized Benchmarking Seed 0 a. The green cells indicate the best-fit distribution type(s).

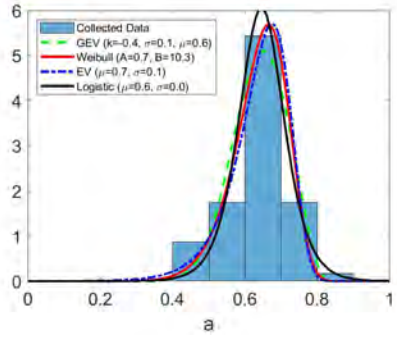
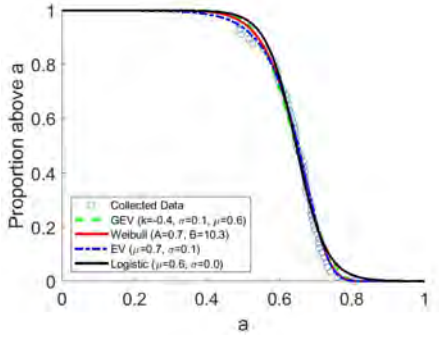
	Probability Density Function	Survival Function
a		

Table 78: Distribution Model Analysis for ibmq_yorktown Q00_Q01 2-Qubit Seed 0 a. In the left column, the probability density functions for the GEV, Weibull, EV, and logistic distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, Weibull, EV, and logistic distribution models are plotted on top of the proportion of collected data observations above a given a.

For 2-Qubit Randomized Benchmarking Seed 0 α, results are nearly the same as 1-Qubit Randomized Benchmarking Seed 0 and Seed 4 α. The GEV distribution is identified as the best-fit distribution type, and 100% of the qubit sets are Type III (Weibull) (see Tables 79 and 80).

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Burr	0.145	1	0.9%	0.115	0	0.0%	0.093	0	0.0%	0.091	0	0.0%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.059	4	3.8%	0.020	16	15.1%	0.029	18	17.0%	0.046	24	22.6%
Gamma	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.632	69	65.1%	0.805	60	56.6%	0.820	58	54.7%	0.802	52	49.1%
Generalized Pareto	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Logistic	0.002	0	0.0%	0.000	1	0.9%	0.000	1	0.9%	0.000	1	0.9%
Loglogistic	0.002	0	0.0%	0.000	2	1.9%	0.000	2	1.9%	0.000	2	1.9%
Lognormal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.000	0	0.0%	0.000	1	0.9%	0.000	1	0.9%	0.000	1	0.9%
t Location Scale	0.104	32	30.2%	0.044	23	21.7%	0.036	23	21.7%	0.032	22	20.8%
Weibull	0.056	0	0.0%	0.015	3	2.8%	0.022	3	2.8%	0.028	4	3.8%

Table 79: NLL, AIC, AICc, and BIC Results for 2-Qubit Randomized Benchmarking Seed 0 α. The green cells indicate the best-fit distribution type(s).

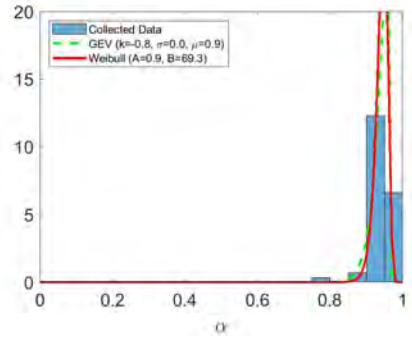
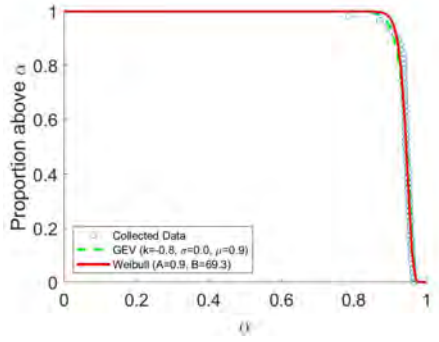
	Probability Density Function	Survival Function
α		

Table 80: Distribution Model Analysis for ibmq_yorktown Q00_Q01 2-Qubit Seed 0 α. In the left column, the probability density functions for the GEV and Weibull distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV and Weibull distribution models are plotted on top of the proportion of collected data observations above a given α.

Results for 2-Qubit Randomized Benchmarking Seed 0 b differ drastically from the 1-qubit results. The Burr, GEV, logistic, loglogistic, and t location-scale distributions are identified as best-fit distribution types. For the GEV distribution, 80% of the qubit sets are identified as Type III (Weibull). It is odd that the Burr distribution is identified as a best-fit distribution, since MATLAB could not generate the Burr distribution for many of the qubit pairs (in Table 82, ibmq_yorktown qubit set Q00_Q03 is presented because MATLAB could not generate the Burr distribution for Q00_Q01 and Q00_Q02). Future research should investigate what properties of the qubit sets lead to different best-fit distribution types.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.068	0	0.0%	0.080	2	1.9%	0.085	2	1.9%	0.090	2	1.9%
Burr	0.123	24	22.6%	0.091	7	6.6%	0.077	3	2.8%	0.054	3	2.8%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.020	5	4.7%	0.006	9	8.5%	0.006	9	8.5%	0.008	12	11.3%
Gamma	0.067	1	0.9%	0.082	1	0.9%	0.086	1	0.9%	0.093	1	0.9%
Generalized Extreme Value	0.120	51	48.1%	0.088	16	15.1%	0.071	12	11.3%	0.050	9	8.5%
Generalized Pareto	0.000	2	1.9%	0.000	2	1.9%	0.000	2	1.9%	0.000	2	1.9%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.068	0	0.0%	0.080	17	16.0%	0.085	18	17.0%	0.090	19	17.9%
Logistic	0.075	2	1.9%	0.098	4	3.8%	0.103	5	4.7%	0.111	5	4.7%
Loglogistic	0.075	0	0.0%	0.097	10	9.4%	0.103	16	15.1%	0.109	16	15.1%
Lognormal	0.067	0	0.0%	0.079	0	0.0%	0.085	0	0.0%	0.089	0	0.0%
Nakagami	0.067	0	0.0%	0.082	1	0.9%	0.085	1	0.9%	0.091	1	0.9%
Normal	0.066	0	0.0%	0.076	0	0.0%	0.078	0	0.0%	0.086	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.067	0	0.0%	0.076	9	8.5%	0.079	9	8.5%	0.087	9	8.5%
t Location Scale	0.094	21	19.8%	0.056	19	17.9%	0.047	17	16.0%	0.033	16	15.1%
Weibull	0.024	0	0.0%	0.009	9	8.5%	0.009	11	10.4%	0.010	11	10.4%

Table 81: NLL, AIC, AICc, and BIC Results for 2-Qubit Randomized Benchmarking Seed 0 b. The green cells indicate the best-fit distribution type(s).

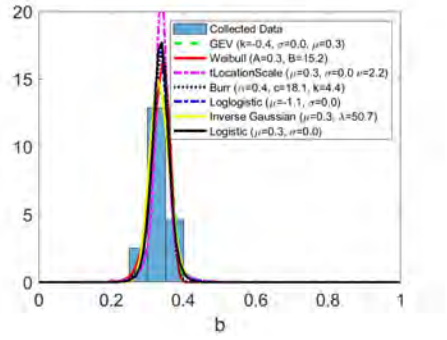
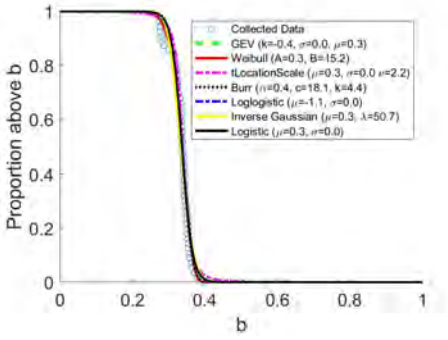
	Probability Density Function	Survival Function
b		

Table 82: Distribution Model Analysis for ibmq_yorktown Q00_Q03 2-Qubit Seed 0 b. In the left column, the probability density functions for the GEV, Weibull, t location-scale, Burr, loglogistic, inverse Gaussian, and logistic distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, Weibull, t location-scale, Burr, loglogistic, inverse Gaussian, and logistic distribution models are plotted on top of the proportion of collected data observations above a given b.

For 2-Qubit Randomized Benchmarking Seed 0 err_a, the Birnbaum Saunders, Burr, and GEV distributions are identified as the best-fit distribution types (see Tables 83 and 84). For the GEV distribution, 64% of the qubits are identified as Type II (Fréchet). As previously stated, more investigation is required to understand why these distributions best fit the err_a characteristic.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.088	1	0.9%	0.126	4	3.8%	0.141	4	3.8%	0.155	4	3.8%
Burr	0.190	25	23.6%	0.194	17	16.0%	0.155	15	14.2%	0.133	14	13.2%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Gamma	0.068	0	0.0%	0.065	4	3.8%	0.075	5	4.7%	0.085	5	4.7%
Generalized Extreme Value	0.186	62	58.5%	0.184	30	28.3%	0.162	25	23.6%	0.118	24	22.6%
Generalized Pareto	0.000	9	8.5%	0.000	6	5.7%	0.000	4	3.8%	0.000	4	3.8%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.088	1	0.9%	0.126	16	15.1%	0.140	19	17.9%	0.153	19	17.9%
Logistic	0.033	0	0.0%	0.016	2	1.9%	0.018	2	1.9%	0.020	2	1.9%
Loglogistic	0.077	0	0.0%	0.089	12	11.3%	0.099	16	15.1%	0.111	18	17.0%
Lognormal	0.088	0	0.0%	0.123	0	0.0%	0.136	0	0.0%	0.151	0	0.0%
Nakagami	0.039	1	0.9%	0.021	5	4.7%	0.024	5	4.7%	0.025	5	4.7%
Normal	0.019	0	0.0%	0.005	0	0.0%	0.006	0	0.0%	0.006	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.026	0	0.0%	0.009	4	3.8%	0.011	4	3.8%	0.012	4	3.8%
t Location Scale	0.078	7	6.6%	0.036	4	3.8%	0.029	4	3.8%	0.025	3	2.8%
Weibull	0.019	0	0.0%	0.005	2	1.9%	0.006	3	2.8%	0.006	4	3.8%

Table 83: NLL, AIC, AICc, and BIC Results for 2-Qubit Randomized Benchmarking Seed 0 *err_a*. The green cells indicate the best-fit distribution type(s).

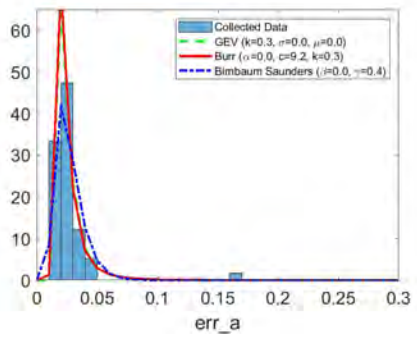
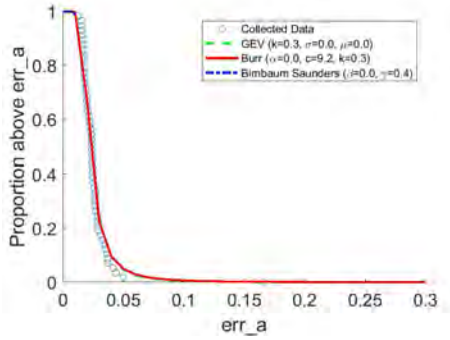
	Probability Density Function	Survival Function
err_a		

Table 84: Distribution Model Analysis for ibmq_yorktown Q00_Q01 2-Qubit Seed 0 *err_a*. In the left column, the probability density functions for the GEV, Burr, and Birnbaum Saunders distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, Burr, and Birnbaum Saunders distribution models are plotted on top of the proportion of collected data observations above a given *err_a*.

For 2-Qubit Randomized Benchmarking Seed 0 *err_α*, the GEV distribution is identified as the best-fit distribution type (see Tables 85 and 86). For the GEV distribution, 92% of the qubits are identified as Type II (Fréchet). More investigation is required to understand why the Type II GEV distribution best fits the 2-Qubit Randomized Benchmarking Seed 0 *err_α* characteristic.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.009	2	1.9%	0.000	2	1.9%	0.000	3	2.8%	0.001	2	1.9%
Burr	0.402	23	21.7%	0.427	21	19.8%	0.425	21	19.8%	0.436	20	18.9%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Gamma	0.001	0	0.0%	0.000	1	0.9%	0.000	1	0.9%	0.000	1	0.9%
Generalized Extreme Value	0.464	72	67.9%	0.548	58	54.7%	0.545	57	53.8%	0.520	56	52.8%
Generalized Pareto	0.000	5	4.7%	0.000	3	2.8%	0.000	2	1.9%	0.000	2	1.9%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.015	1	0.9%	0.001	15	14.2%	0.001	16	15.1%	0.002	16	15.1%
Logistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.055	0	0.0%	0.020	2	1.9%	0.025	2	1.9%	0.037	4	3.8%
Lognormal	0.015	0	0.0%	0.001	0	0.0%	0.001	0	0.0%	0.002	0	0.0%
Nakagami	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	1	0.9%	0.000	1	0.9%	0.000	2	1.9%
Rician	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.039	3	2.8%	0.003	3	2.8%	0.003	3	2.8%	0.003	3	2.8%
Weibull	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%

Table 85: NLL, AIC, AICc, and BIC Results for 2-Qubit Randomized Benchmarking Seed 0 err_α. The green cells indicate the best-fit distribution type(s).

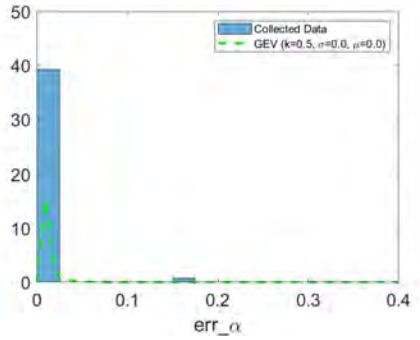
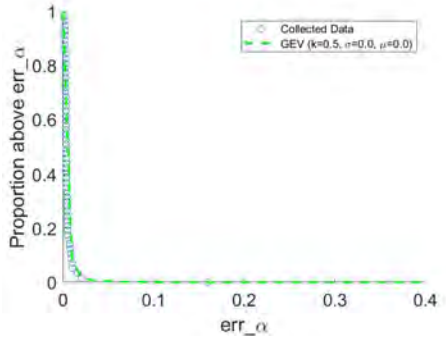
	Probability Density Function	Survival Function
err_α		

Table 86: Distribution Model Analysis for ibmq_yorktown Q00_Q01 2-Qubit Seed 0 err_α. In the left column, the probability density function for the GEV distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival function for the GEV distribution model is plotted on top of the proportion of collected data observations above a given err_α.

For 2-Qubit Randomized Benchmarking Seed 0 err_b, the Birnbaum Saunders, GEV, and inverse Gaussian distributions are selected as best fit distributions (see Tables 87 and 88). For the GEV distribution, 55% of the qubit sets are Type III (Weibull), and 45% are Type II (Fréchet). Future research should investigate what properties of the qubit sets cause such diversity in distribution types.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.099	6	5.7%	0.151	11	10.4%	0.155	11	10.4%	0.162	11	10.4%
Burr	0.095	19	17.9%	0.048	5	4.7%	0.039	5	4.7%	0.026	1	0.9%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.007	0	0.0%	0.001	0	0.0%	0.001	0	0.0%	0.001	0	0.0%
Gamma	0.087	4	3.8%	0.114	10	9.4%	0.120	10	9.4%	0.122	10	9.4%
Generalized Extreme Value	0.111	51	48.1%	0.070	9	8.5%	0.055	6	5.7%	0.039	4	3.8%
Generalized Pareto	0.003	16	15.1%	0.000	13	12.3%	0.000	13	12.3%	0.000	13	12.3%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.100	8	7.5%	0.150	27	25.5%	0.155	30	28.3%	0.162	32	30.2%
Logistic	0.045	0	0.0%	0.032	0	0.0%	0.034	0	0.0%	0.034	0	0.0%
Loglogistic	0.079	0	0.0%	0.094	14	13.2%	0.096	14	13.2%	0.099	18	17.0%
Lognormal	0.099	0	0.0%	0.146	1	0.9%	0.152	1	0.9%	0.160	1	0.9%
Nakagami	0.073	0	0.0%	0.078	6	5.7%	0.081	6	5.7%	0.084	6	5.7%
Normal	0.047	0	0.0%	0.032	0	0.0%	0.033	0	0.0%	0.033	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.052	0	0.0%	0.039	4	3.8%	0.040	4	3.8%	0.039	4	3.8%
t Location Scale	0.063	1	0.9%	0.022	1	0.9%	0.016	1	0.9%	0.012	1	0.9%
Weibull	0.040	1	0.9%	0.024	5	4.7%	0.025	5	4.7%	0.025	5	4.7%

Table 87: NLL, AIC, AICc, and BIC Results for 2-Qubit Randomized Benchmarking Seed 0 *err_b*. The green cells indicate the best-fit distribution type(s).

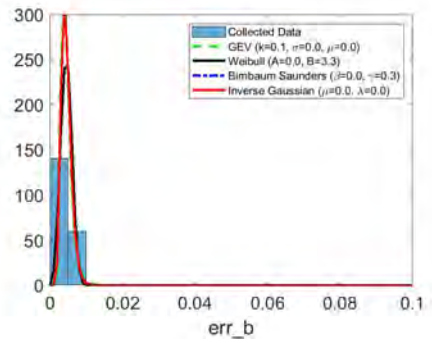
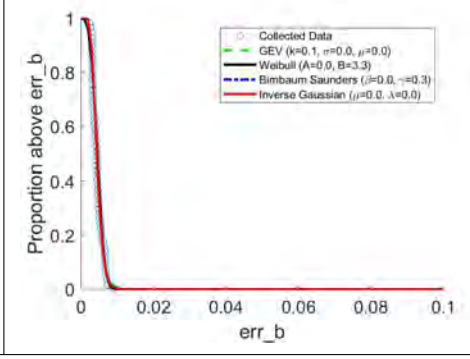
	Probability Density Function	Survival Function
err_b		

Table 88: Distribution Model Analysis for ibmq_yorktown Q00_Q01 2-Qubit Seed 0 *err_b*. In the left column, the probability density functions for the GEV, Weibull, Birnbaum Saunders, and inverse Gaussian distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, Weibull, Birnbaum Saunders, and inverse Gaussian distribution models are plotted on top of the proportion of collected data observations above a given *err_b*.

For 2-Qubit Randomized Benchmarking Seed 4 Fidelity, the EV, GEV, and generalized Pareto distributions are identified as the best-fit distributions (see Tables 89–91). For the GEV distribution, 98% of the qubits are Type III (Weibull), and for the generalized Pareto distribution, 100% of the qubit sets have a negative shape parameter. When the generalized Pareto distribution is indicated as a best-fit distribution, it often suggests bimodal behavior exists in some of the qubit sets. Possible bimodal behavior is observed in the probability density function for ibmq_casablanca Q00_Q02 in Table 91. Future research should investigate to determine if more bimodal behavior exists, and if so, investigate the qubit properties that cause this behavior.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.011	0	0.0%	0.001	0	0.0%	0.001	0	0.0%	0.002	0	0.0%
Burr	0.150	7	5.5%	0.176	1	0.8%	0.137	1	0.8%	0.105	1	0.8%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.125	13	10.2%	0.186	29	22.8%	0.241	32	25.2%	0.313	37	29.1%
Gamma	0.016	0	0.0%	0.003	0	0.0%	0.003	0	0.0%	0.004	0	0.0%
Generalized Extreme Value	0.297	61	48.0%	0.372	29	22.8%	0.318	25	19.7%	0.233	20	15.7%
Generalized Pareto	0.027	40	31.5%	0.003	40	31.5%	0.002	39	30.7%	0.002	39	30.7%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.011	0	0.0%	0.001	5	3.9%	0.001	5	3.9%	0.002	5	3.9%
Logistic	0.049	0	0.0%	0.027	2	1.6%	0.029	2	1.6%	0.036	2	1.6%
Loglogistic	0.025	0	0.0%	0.006	2	1.6%	0.007	2	1.6%	0.008	2	1.6%
Lognormal	0.011	0	0.0%	0.001	0	0.0%	0.001	0	0.0%	0.002	0	0.0%
Nakagami	0.024	0	0.0%	0.005	0	0.0%	0.007	0	0.0%	0.008	0	0.0%
Normal	0.031	0	0.0%	0.008	0	0.0%	0.010	0	0.0%	0.011	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.031	0	0.0%	0.008	5	3.9%	0.010	5	3.9%	0.011	5	3.9%
t Location Scale	0.082	5	3.9%	0.035	2	1.6%	0.030	2	1.6%	0.025	1	0.8%
Weibull	0.110	1	0.8%	0.167	12	9.4%	0.203	14	11.0%	0.238	15	11.8%

Table 89: NLL, AIC, AICc, and BIC Results for 2-Qubit Randomized Benchmarking Seed 4 Fidelity. The green cells indicate the best-fit distribution type(s).

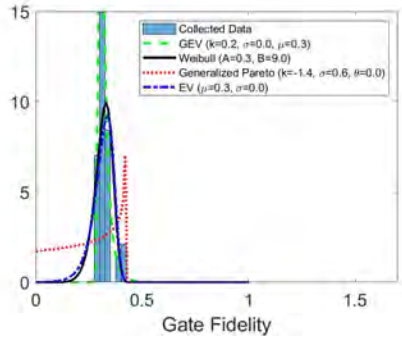
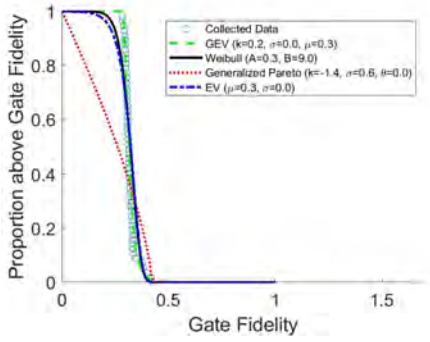
	Probability Density Function	Survival Function
Fidelity		

Table 90: Distribution Model Analysis for ibmq_yorktown Q00_Q01 2-Qubit Seed 4 Fidelity. In the left column, the probability density functions for the GEV, Weibull, generalized Pareto, and EV distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, Weibull, generalized Pareto, and EV distribution models are plotted on top of the proportion of collected data observations above a given fidelity.

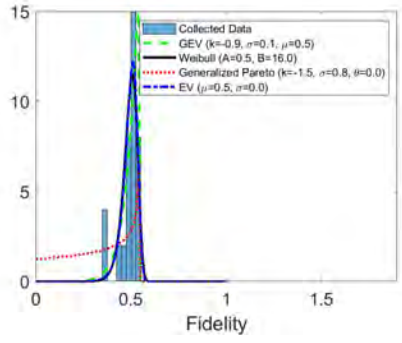
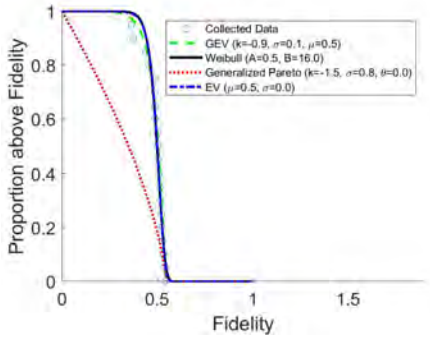
	Probability Density Function	Survival Function
Fidelity		

Table 91: Distribution Model Analysis for ibmq_casablanca Q00_Q02 2-Qubit Seed 4 Fidelity. In the left column, the probability density functions for the GEV, Weibull, generalized Pareto, and EV distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, Weibull, generalized Pareto, and EV distribution models are plotted on top of the proportion of collected data observations above a given fidelity.

For 2-Qubit Randomized Benchmarking Seed 4 a, the generalized Pareto distribution is identified as the best-fit distribution type, where the shape parameter is negative for 100% of the qubits (see Tables 92–94). Bimodal behavior is not observed

in Table 93, but what appears to be multi-modal behavior is observed in Table 94.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.022	0	0.0%	0.002	0	0.0%	0.002	0	0.0%	0.006	0	0.0%
Burr	0.210	8	13.1%	0.106	6	9.8%	0.096	6	9.8%	0.068	6	9.8%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Gamma	0.011	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.001	0	0.0%
Generalized Extreme Value	0.084	21	34.4%	0.019	13	21.3%	0.019	12	19.7%	0.009	10	16.4%
Generalized Pareto	0.608	30	49.2%	0.867	30	49.2%	0.876	30	49.2%	0.903	30	49.2%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.022	1	1.6%	0.002	6	9.8%	0.002	7	11.5%	0.006	9	14.8%
Logistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.015	0	0.0%	0.001	2	3.3%	0.001	2	3.3%	0.003	2	3.3%
Lognormal	0.023	0	0.0%	0.002	0	0.0%	0.003	0	0.0%	0.005	0	0.0%
Nakagami	0.002	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.001	1	1.6%	0.000	1	1.6%	0.000	1	1.6%	0.000	1	1.6%
Weibull	0.001	0	0.0%	0.000	3	4.9%	0.000	3	4.9%	0.000	3	4.9%

Table 92: NLL, AIC, AICc, and BIC Results for 2-Qubit Randomized Benchmarking Seed 4 a. The green cells indicate the best-fit distribution type(s).

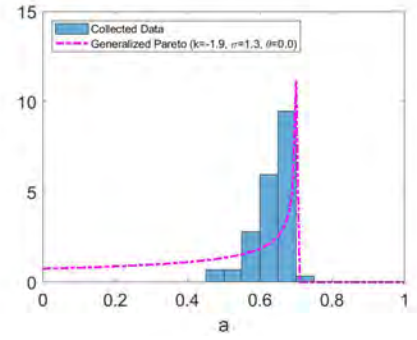
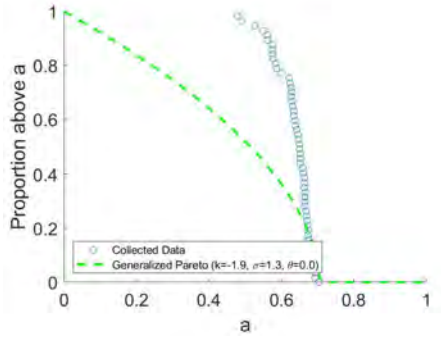
	Probability Density Function	Survival Function
a		

Table 93: Distribution Model Analysis for ibmq_yorktown Q00_Q01 2-Qubit Seed 4 a. In the left column, the probability density function for the generalized Pareto distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival function for the generalized Pareto distribution is plotted on top of the proportion of collected data observations above a given a.

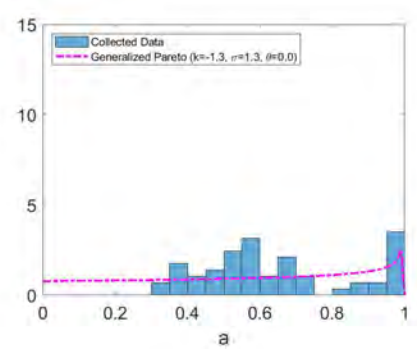
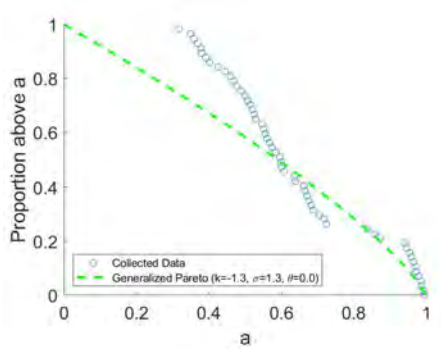
	Probability Density Function	Survival Function
a		

Table 94: Distribution Model Analysis for ibmq_yorktown Q00_Q03 2-Qubit Seed 4 a. In the left column, the probability density function for the generalized Pareto distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival function for the generalized Pareto distribution is plotted on top of the proportion of collected data observations above a given a.

For 2-Qubit Randomized Benchmarking Seed 4 α, results are very similar to all previous α results. The GEV distribution is identified as the best-fit distribution, and 100% of the qubits have a Type III (Weibull) distribution (see Tables 95 and 96).

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Burr	0.127	1	1.0%	0.074	0	0.0%	0.061	0	0.0%	0.058	0	0.0%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.053	9	8.6%	0.009	17	16.2%	0.014	18	17.1%	0.021	23	21.9%
Gamma	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.672	73	69.5%	0.887	64	61.0%	0.892	63	60.0%	0.885	58	55.2%
Generalized Pareto	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.000	0	0.0%	0.000	1	1.0%	0.000	1	1.0%	0.000	1	1.0%
Logistic	0.004	0	0.0%	0.000	1	1.0%	0.000	1	1.0%	0.000	2	1.9%
Loglogistic	0.003	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Lognormal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.000	0	0.0%	0.000	2	1.9%	0.000	2	1.9%	0.000	2	1.9%
t Location Scale	0.088	22	21.0%	0.020	18	17.1%	0.020	18	17.1%	0.017	17	16.2%
Weibull	0.050	0	0.0%	0.009	2	1.9%	0.013	2	1.9%	0.019	2	1.9%

Table 95: NLL, AIC, AICc, and BIC Results for 2-Qubit Randomized Benchmarking Seed 4 α. The green cells indicate the best-fit distribution type(s).

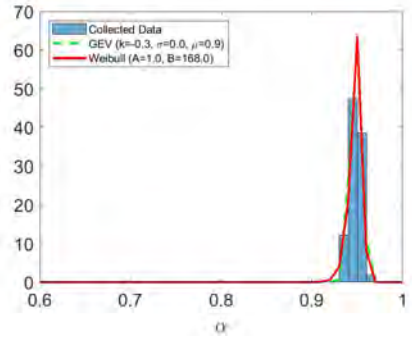
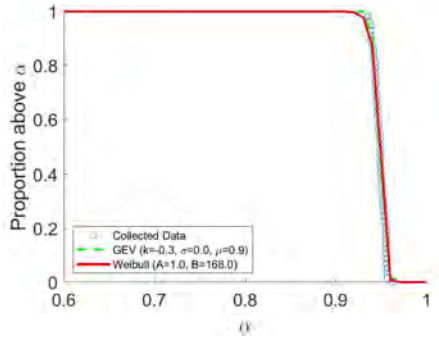
	Probability Density Function	Survival Function
α		

Table 96: Distribution Model Analysis for ibmq_yorktown Q00_Q01 2-Qubit Seed 4 α. In the left column, the probability density functions for the GEV and Weibull distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV and Weibull distribution models are plotted on top of the proportion of collected data observations above a given α.

For 2-Qubit Randomized Benchmarking Seed 4 b, the GEV distribution is identified as the best-fit distribution by all metrics except the BIC normalized median value, which identifies the loglogistic distribution as the best fit (see Tables 97 and 98). Compared to Seed 0, more of the metrics indicate the same best-fit distribution type, but within the GEV distribution, there is more variation in the qubits. For Seed 0, 80% of the qubits are Type III (Weibull), but for Seed 4, only 69% of the qubit pairs are Type III (Weibull). Future work should investigate the reason for this change. The data points in Table 98 are non-negative and positively skewed, so it is unsurprising that the loglogistic distribution fits the data well.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.053	0	0.00%	0.045	14	13.33%	0.057	14	13.33%	0.073	15	14.29%
Burr	0.161	13	12.38%	0.166	8	7.62%	0.135	7	6.67%	0.096	7	6.67%
Exponential	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
Extreme Value	0.026	8	7.62%	0.009	12	11.43%	0.011	12	11.43%	0.012	14	13.33%
Gamma	0.049	1	0.95%	0.039	3	2.86%	0.046	3	2.86%	0.07	3	2.86%
Generalized Extreme Value	0.204	65	61.90%	0.244	33	31.43%	0.197	30	28.57%	0.145	25	23.81%
Generalized Pareto	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
Half-Normal	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
Inverse Gaussian	0.001	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
Logistic	0.083	1	0.95%	0.11	5	4.76%	0.138	5	4.76%	0.134	5	4.76%
Loglogistic	0.089	0	0.00%	0.125	6	5.71%	0.153	8	7.62%	0.16	11	10.48%
Lognormal	0.053	0	0.00%	0.045	0	0.00%	0.057	0	0.00%	0.073	0	0.00%
Nakagami	0.049	0	0.00%	0.033	0	0.00%	0.039	0	0.00%	0.06	0	0.00%
Normal	0.044	0	0.00%	0.033	0	0.00%	0.037	0	0.00%	0.051	0	0.00%
Rayleigh	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
Rician	0.044	0	0.00%	0.033	8	7.62%	0.037	9	8.57%	0.051	9	8.57%
t Location Scale	0.121	17	16.19%	0.106	16	15.24%	0.081	16	15.24%	0.062	15	14.29%
Weibull	0.023	0	0.00%	0.011	0	0.00%	0.012	1	0.95%	0.011	1	0.95%

Table 97: NLL, AIC, AICc, and BIC Results for 2-Qubit Randomized Benchmarking Seed 4 *b*. The green cells indicate the best-fit distribution type(s).

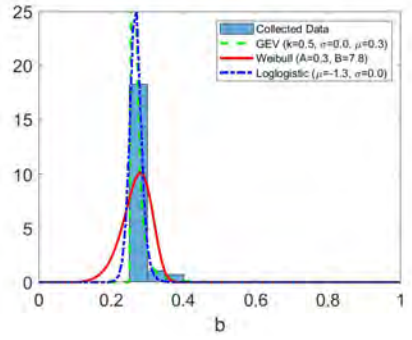
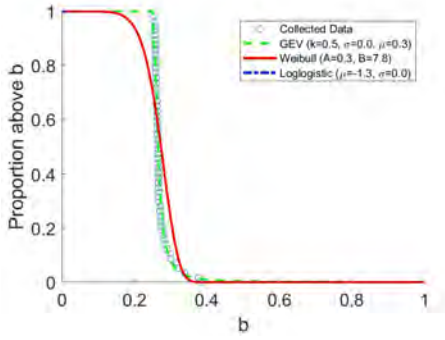
	Probability Density Function	Survival Function
b		

Table 98: Distribution Model Analysis for ibmq_yorktown Q00_Q01 2-Qubit Seed 4 *b*. In the left column, the probability density functions for the GEV, Weibull, and loglogistic distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, Weibull, and loglogistic distribution models are plotted on top of the proportion of collected data observations above a given *b*.

For 2-Qubit Randomized Benchmarking Seed 4 *err_a*, results are consistent with Seed 0, but show less variation. This meets expectations, as the greater amount of Clifford sequences used to produce Seed 4 results should lead to less variation in the fidelity equations. The Burr distribution, inverse Gaussian distribution, and loglogistic distributions are identified as best-fit distribution types (see Tables 99 and 100). Each of these distribution types are ideal for the non-negative, positively skewed data observed in Table 100.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.103	11	10.5%	0.130	11	10.5%	0.142	10	9.5%	0.155	3	2.9%
Burr	0.208	37	35.2%	0.176	12	11.4%	0.155	10	9.5%	0.085	6	5.7%
Exponential	0.001	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Gamma	0.053	5	4.8%	0.037	3	2.9%	0.038	3	2.9%	0.044	2	1.9%
Generalized Extreme Value	0.194	30	28.6%	0.155	20	19.0%	0.133	16	15.2%	0.092	14	13.3%
Generalized Pareto	0.004	7	6.7%	0.000	3	2.9%	0.000	2	1.9%	0.000	1	1.0%
Half-Normal	0.002	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.115	11	10.5%	0.164	17	16.2%	0.177	21	20.0%	0.206	23	21.9%
Logistic	0.005	0	0.0%	0.000	1	1.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.105	0	0.0%	0.153	28	26.7%	0.158	29	27.6%	0.197	32	30.5%
Lognormal	0.112	1	1.0%	0.164	1	1.0%	0.176	1	1.0%	0.194	1	1.0%
Nakagami	0.019	1	1.0%	0.004	2	1.9%	0.004	3	2.9%	0.005	3	2.9%
Normal	0.002	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.008	0	0.0%	0.002	5	4.8%	0.002	9	8.6%	0.006	19	18.1%
Rician	0.008	0	0.0%	0.001	0	0.0%	0.001	0	0.0%	0.001	0	0.0%
t Location Scale	0.032	2	1.9%	0.005	0	0.0%	0.004	0	0.0%	0.003	0	0.0%
Weibull	0.029	0	0.0%	0.009	2	1.9%	0.010	1	1.0%	0.011	1	1.0%

Table 99: NLL, AIC, AICc, and BIC Results for 2-Qubit Randomized Benchmarking Seed 4 *err_a*. The green cells indicate the best-fit distribution type(s).

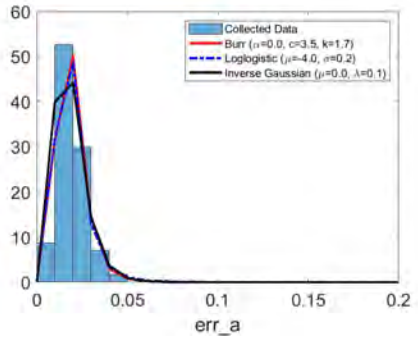
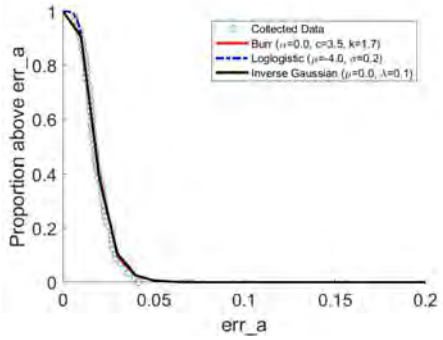
	Probability Density Function	Survival Function
err_a		

Table 100: Distribution Model Analysis for ibmq_yorktown Q00_Q01 2-Qubit Seed 4 *err_a*. In the left column, the probability density functions for the Burr, loglogistic, and inverse Gaussian distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the Burr, loglogistic, and inverse Gaussian distribution models are plotted on top of the proportion of collected data observations above a given *err_a*.

For 2-Qubit Randomized Benchmarking Seed 4 *err_a*, the GEV distribution is identified as the best fit by all metrics except the NLL normalized median value, which identified the Burr distribution as the best fit (see Tables 101 and 102). For the GEV distribution, 91% of the distributions are Type II (Fréchet). In the same manner as 2-Qubit Randomized Benchmarking Seed 4 *err_a*, results are consistent with Seed 0, but show less variation.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.040	3	2.9%	0.013	5	4.8%	0.016	5	4.8%	0.024	5	4.8%
Burr	0.366	33	31.4%	0.420	26	24.8%	0.405	22	21.0%	0.367	21	20.0%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Gamma	0.004	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.365	61	58.1%	0.440	50	47.6%	0.431	46	43.8%	0.403	42	40.0%
Generalized Pareto	0.000	3	2.9%	0.000	2	1.9%	0.000	2	1.9%	0.000	2	1.9%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.064	2	1.9%	0.031	12	11.4%	0.043	17	16.2%	0.072	19	18.1%
Logistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.072	0	0.0%	0.069	5	4.8%	0.075	8	7.6%	0.081	11	10.5%
Lognormal	0.052	0	0.0%	0.022	0	0.0%	0.025	0	0.0%	0.049	0	0.0%
Nakagami	0.000	0	0.0%	0.000	2	1.9%	0.000	2	1.9%	0.000	2	1.9%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.034	3	2.9%	0.005	3	2.9%	0.005	3	2.9%	0.004	3	2.9%
Weibull	0.002	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%

Table 101: NLL, AIC, AICc, and BIC Results for 2-Qubit Randomized Benchmarking Seed 4 err_α. The green cells indicate the best-fit distribution type(s).

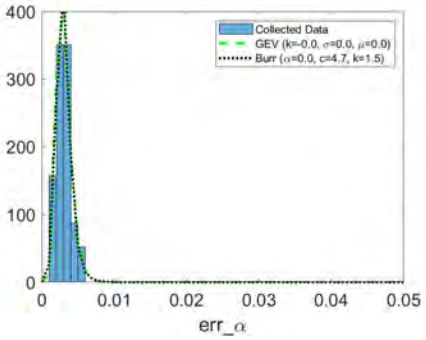
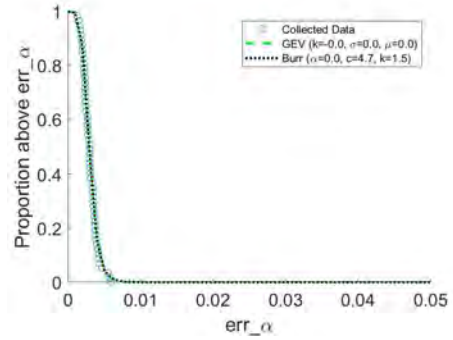
	Probability Density Function	Survival Function
err_α		

Table 102: Distribution Model Analysis for ibmq_yorktown Q00_Q01 2-Qubit Seed 4 err_α. In the left column, the probability density functions for the GEV and Burr distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV and Burr distribution models are plotted on top of the proportion of collected data observations above a given err_α.

For 2-Qubit Randomized Benchmarking Seed 4 err_b, results are similar to Seed 0, with the Birnbaum Saunders, GEV, and inverse Gaussian distributions identified as the best-fit distribution types (see Tables 103 and 104). For the GEV distributions, 59% of the qubit sets are Type III (Weibull) and 41% of the qubit sets are Type II (Fréchet).

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.094	4	3.8%	0.137	9	8.6%	0.145	9	8.6%	0.148	8	7.6%
Burr	0.096	16	15.2%	0.053	7	6.7%	0.041	6	5.7%	0.031	6	5.7%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.007	0	0.0%	0.001	0	0.0%	0.001	0	0.0%	0.001	0	0.0%
Gamma	0.088	1	1.0%	0.122	9	8.6%	0.125	9	8.6%	0.131	9	8.6%
Generalized Extreme Value	0.118	63	60.0%	0.080	12	11.4%	0.063	8	7.6%	0.041	6	5.7%
Generalized Pareto	0.001	10	9.5%	0.000	6	5.7%	0.000	6	5.7%	0.000	6	5.7%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.093	4	3.8%	0.135	26	24.8%	0.143	29	27.6%	0.151	29	27.6%
Logistic	0.047	1	1.0%	0.034	3	2.9%	0.035	3	2.9%	0.036	3	2.9%
Loglogistic	0.074	0	0.0%	0.080	10	9.5%	0.085	12	11.4%	0.089	12	11.4%
Lognormal	0.091	0	0.0%	0.133	0	0.0%	0.138	0	0.0%	0.142	2	1.9%
Nakagami	0.074	3	2.9%	0.089	16	15.2%	0.092	16	15.2%	0.094	16	15.2%
Normal	0.052	0	0.0%	0.039	0	0.0%	0.040	0	0.0%	0.044	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	1	1.0%
Rician	0.055	0	0.0%	0.048	4	3.8%	0.049	4	3.8%	0.051	4	3.8%
t Location Scale	0.068	2	1.9%	0.027	0	0.0%	0.021	0	0.0%	0.017	0	0.0%
Weibull	0.040	1	1.0%	0.022	3	2.9%	0.023	3	2.9%	0.025	3	2.9%

Table 103: NLL, AIC, AICc, and BIC Results for 2-Qubit Randomized Benchmarking Seed 4 *err_b*. The green cells indicate the best-fit distribution type(s).

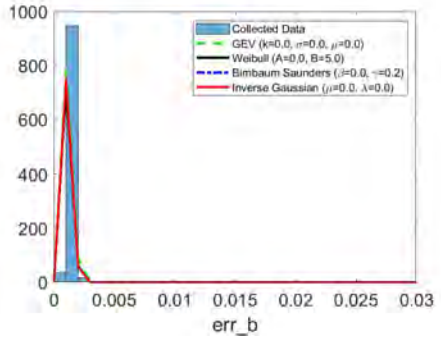
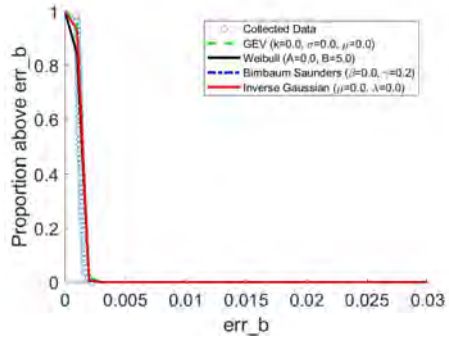
	Probability Density Function	Survival Function
err_b		

Table 104: Distribution Model Analysis for ibmq_yorktown Q00_Q01 2-Qubit Seed 4 *err_b*. In the left column, the probability density functions for the GEV, Weibull, Birnbaum Saunders, and inverse Gaussian distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, Weibull, Birnbaum Saunders, and inverse Gaussian distribution models are plotted on top of the proportion of collected data observations above a given *err_b*.

For 3-Qubit Randomized Benchmarking Seed 0 Fidelity, the NLL metric identify the GEV distribution as the best-fit distribution, and the other metrics identify the logistic distribution as the best-fit distribution (see Tables 105 and 106). For the GEV distribution, 100% of the qubit sets are Type III (Weibull). Results are surprisingly different from 2-Qubit Randomized Benchmarking Seed 0 Fidelity because the multi-modal behavior seems to vanish. An initial hypothesis for this shift in behavior is: so many sources of error are involved in 3-qubit gates that fidelity becomes low for all sets of qubits, regardless of physical configuration. Future research should investigate the correctness of this hypothesis.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.068	0	0.00%	0.073	1	1.80%	0.077	1	1.80%	0.082	1	1.80%
Burr	0.103	12	21.80%	0.068	1	1.80%	0.061	1	1.80%	0.029	1	1.80%
Exponential	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
Extreme Value	0.017	0	0.00%	0.005	0	0.00%	0.005	0	0.00%	0.005	1	1.80%
Gamma	0.073	0	0.00%	0.087	0	0.00%	0.088	0	0.00%	0.098	0	0.00%
Generalized Extreme Value	0.089	23	41.80%	0.048	5	9.10%	0.039	3	5.50%	0.022	2	3.60%
Generalized Pareto	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
Half-Normal	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
Inverse Gaussian	0.068	1	1.80%	0.073	7	12.70%	0.077	7	12.70%	0.082	7	12.70%
Logistic	0.081	6	10.90%	0.112	14	25.50%	0.115	14	25.50%	0.121	15	27.30%
Loglogistic	0.075	0	0.00%	0.084	10	18.20%	0.084	10	18.20%	0.092	10	18.20%
Lognormal	0.067	0	0.00%	0.072	0	0.00%	0.076	0	0.00%	0.082	0	0.00%
Nakagami	0.077	0	0.00%	0.102	2	3.60%	0.102	2	3.60%	0.113	2	3.60%
Normal	0.081	0	0.00%	0.107	0	0.00%	0.109	0	0.00%	0.117	0	0.00%
Rayleigh	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
Rician	0.081	1	1.80%	0.107	10	18.20%	0.109	11	20.00%	0.118	11	20.00%
t Location Scale	0.092	12	21.80%	0.05	3	5.50%	0.045	3	5.50%	0.026	1	1.80%
Weibull	0.028	0	0.00%	0.012	2	3.60%	0.012	3	5.50%	0.013	4	7.30%

Table 105: NLL, AIC, AICc, and BIC Results for 3-Qubit Randomized Benchmarking Seed 0 Fidelity. The green cells indicate the best-fit distribution type(s).

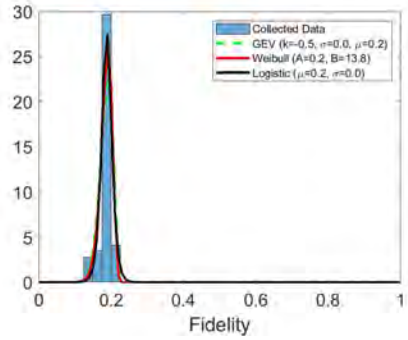
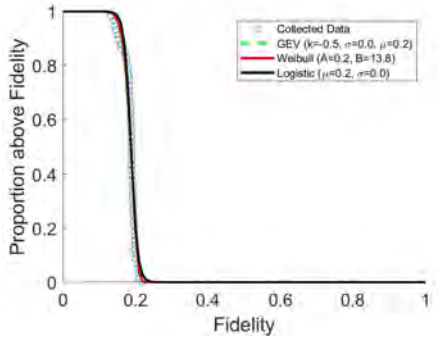
	Probability Density Function	Survival Function
Fidelity		

Table 106: Distribution Model Analysis for ibmq_yorktown Q00_Q01_Q02 3-Qubit Seed 0 Fidelity. In the left column, the probability density functions for the GEV, Weibull, and logistic distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, Weibull, and logistic distribution models are plotted on top of the proportion of collected data observations above a given Fidelity.

For 3-Qubit Randomized Benchmarking Seed 0 a, the generalized Pareto distribution is the best-fit distribution type (see Tables 107 and 108). The shape parameter is positive for 100% of the qubit sets. As with most generalized Pareto best-fit distributions in this work, bimodal behavior is observed in Table 108. Future research should investigate the reason for the bimodal behavior and repeat this experiment for bimodal distribution types.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.009	0	0.0%	0.000	1	1.8%	0.000	1	1.8%	0.001	1	1.8%
Burr	0.114	5	9.1%	0.059	1	1.8%	0.051	1	1.8%	0.036	1	1.8%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.002	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Gamma	0.021	0	0.0%	0.003	2	3.6%	0.003	2	3.6%	0.006	2	3.6%
Generalized Extreme Value	0.074	11	20.0%	0.013	0	0.0%	0.013	0	0.0%	0.009	0	0.0%
Generalized Pareto	0.579	31	56.4%	0.892	29	52.7%	0.897	29	52.7%	0.880	29	52.7%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.008	2	3.6%	0.000	4	7.3%	0.000	4	7.3%	0.001	4	7.3%
Logistic	0.017	2	3.6%	0.002	5	9.1%	0.003	5	9.1%	0.006	6	10.9%
Loglogistic	0.020	0	0.0%	0.003	3	5.5%	0.004	3	5.5%	0.009	3	5.5%
Lognormal	0.008	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.001	0	0.0%
Nakagami	0.032	0	0.0%	0.009	2	3.6%	0.010	2	3.6%	0.024	2	3.6%
Normal	0.025	0	0.0%	0.004	1	1.8%	0.005	1	1.8%	0.008	1	1.8%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.025	0	0.0%	0.004	1	1.8%	0.005	1	1.8%	0.008	1	1.8%
t Location Scale	0.043	4	7.3%	0.007	3	5.5%	0.006	3	5.5%	0.004	2	3.6%
Weibull	0.023	0	0.0%	0.002	3	5.5%	0.003	3	5.5%	0.008	3	5.5%

Table 107: NLL, AIC, AICc, and BIC Results for 3-Qubit Randomized Benchmarking Seed 0 a. The green cells indicate the best-fit distribution type(s).

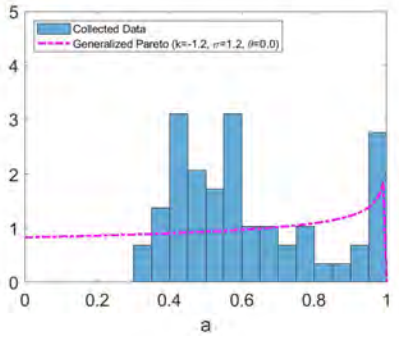
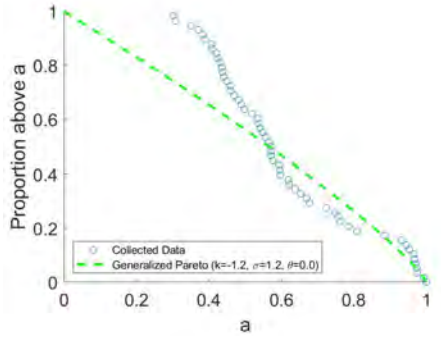
	Probability Density Function	Survival Function
a		

Table 108: Distribution Model Analysis for ibmq_yorktown Q00_Q01_Q02 3-Qubit Seed 0 a. In the left column, the probability density function for the generalized Pareto distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the generalized Pareto distribution model is plotted on top of the proportion of collected data observations above a given a.

For 3-Qubit Randomized Benchmarking Seed 0 α, the GEV distribution is identified as the best-fit distribution, where 100% of the qubit sets are Type II (Fréchet) (see Tables 109 and 110). Upon analyzing the probability density function in Table 110, more bimodal behavior is observed, which explains why the generalized Pareto distribution is the second-best-fit distribution. However, the fact that the generalized Pareto distribution is the second-best-fit distribution instead of the best-fit distribution indicates less bimodal or multi-modal behavior for this characteristic than other characteristics that exhibit such behavior. Future work should investigate the reason for reduced bimodal and multi-modal behavior.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Burr	0.200	1	1.8%	0.166	0	0.0%	0.153	0	0.0%	0.079	0	0.0%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.016	0	0.0%	0.001	0	0.0%	0.002	0	0.0%	0.003	0	0.0%
Gamma	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.635	27	49.1%	0.812	27	49.1%	0.825	27	49.1%	0.895	27	49.1%
Generalized Pareto	0.126	18	32.7%	0.020	18	32.7%	0.020	18	32.7%	0.022	18	32.7%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Logistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Lognormal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.016	9	16.4%	0.000	9	16.4%	0.000	9	16.4%	0.001	9	16.4%
Weibull	0.006	0	0.0%	0.000	1	1.8%	0.000	1	1.8%	0.000	1	1.8%

Table 109: NLL, AIC, AICc, and BIC Results for 3-Qubit Randomized Benchmarking Seed 0 α. The green cells indicate the best-fit distribution type(s).

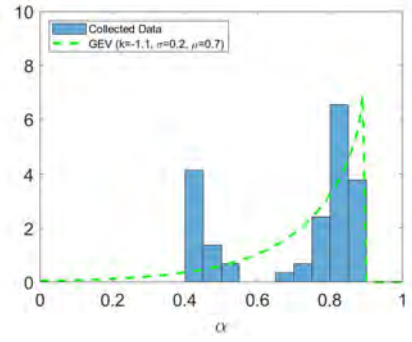
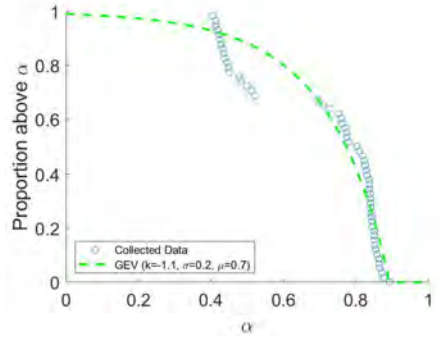
	Probability Density Function	Survival Function
α		

Table 110: Distribution Model Analysis for ibmq_yorktown Q00_Q01_Q02 3-Qubit Seed 0 α. In the left column, the probability density functions for the GEV distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival function for the GEV distribution model is plotted on top of the proportion of collected data observations above a given α.

For 3-Qubit Randomized Benchmarking Seed 0 b, the Burr, GEV, logistic, and Rician distribution models are identified as best-fit distributions (see Tables 111 and 112). The Rician probability density function is given by the expression,

$$f\left(x|s,\sigma\right) = I_0\left(\frac{xs}{\sigma^2}\right)\frac{x}{\sigma^2}e^{-\left(\frac{x^2+s^2}{2\sigma^2}\right)}, \tag{38}$$

where I_0 is the zero-order modified Bessel function of the first kind, $s \geq 0$ is the noncentrality parameter, and $\sigma > 0$ is the scale parameter $x > 0$. [33]. Rician distributions are often used to model signal scattering; it would make an interesting future study to determine what type of relationship exists between the b parameter and scattering from Clifford gate signals.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.065	0	0.0%	0.070	0	0.0%	0.072	0	0.0%	0.079	0	0.0%
Burr	0.103	11	20.0%	0.063	1	1.8%	0.057	1	1.8%	0.028	1	1.8%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.018	0	0.0%	0.006	0	0.0%	0.006	0	0.0%	0.006	1	1.8%
Gamma	0.073	0	0.0%	0.089	0	0.0%	0.091	0	0.0%	0.096	0	0.0%
Generalized Extreme Value	0.089	24	43.6%	0.051	5	9.1%	0.044	4	7.3%	0.023	2	3.6%
Generalized Pareto	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.065	1	1.8%	0.070	7	12.7%	0.072	7	12.7%	0.079	7	12.7%
Logistic	0.082	6	10.9%	0.104	14	25.5%	0.108	14	25.5%	0.109	14	25.5%
Loglogistic	0.074	0	0.0%	0.077	7	12.7%	0.078	7	12.7%	0.100	7	12.7%
Lognormal	0.065	0	0.0%	0.071	0	0.0%	0.072	0	0.0%	0.078	0	0.0%
Nakagami	0.077	0	0.0%	0.106	3	5.5%	0.108	3	5.5%	0.112	3	5.5%
Normal	0.080	0	0.0%	0.109	0	0.0%	0.112	0	0.0%	0.119	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.080	0	0.0%	0.110	13	23.6%	0.113	13	23.6%	0.120	14	25.5%
t Location Scale	0.092	13	23.6%	0.052	3	5.5%	0.046	3	5.5%	0.026	2	3.6%
Weibull	0.037	0	0.0%	0.023	2	3.6%	0.023	3	5.5%	0.024	4	7.3%

Table 111: NLL, AIC, AICc, and BIC Results for 3-Qubit Randomized Benchmarking Seed 0 b. The green cells indicate the best-fit distribution type(s).

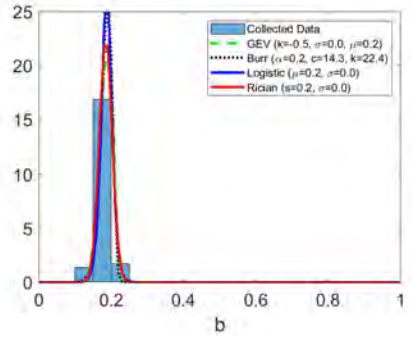
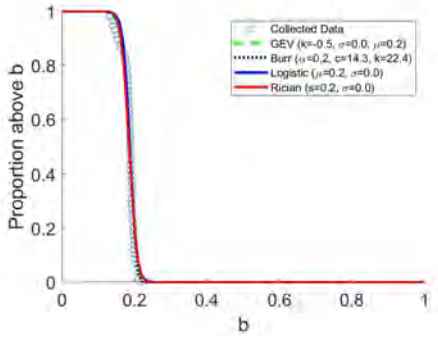
	Probability Density Function	Survival Function
b		

Table 112: Distribution Model Analysis for ibmq_yorktown Q00_Q01_Q02 3-Qubit Seed 0 b. In the left column, the probability density functions for the GEV, Burr, logistic, and Rician distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, Burr, logistic, and Rician distribution models are plotted on top of the proportion of collected data observations above a given b.

For 3-Qubit Randomized Benchmarking Seed 0 err_a and err_α, the Birnbaum Saunders and Burr distributions are identified as best-fit distributions(see Tables 113–116). Much investigation is needed for these characteristics. In the probability density functions in Tables 114 and 116, it can be seen that a) the Burr distribution is omitted because MATLAB could not fit the distribution to the data, and b) many of the data values seem unreasonable. Most data points are near zero, but several data points are several orders of magnitude larger. The data should be investigated to determine if these seemingly anomalous values are outliers or if they are actually characteristic of qubit set behavior.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.316	27	49.1%	0.358	28	50.9%	0.397	28	50.9%	0.411	28	50.9%
Burr	0.684	15	27.3%	0.642	15	27.3%	0.603	15	27.3%	0.589	15	27.3%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Gamma	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.000	8	14.5%	0.000	7	12.7%	0.000	7	12.7%	0.000	7	12.7%
Generalized Pareto	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.000	3	5.5%	0.000	3	5.5%	0.000	3	5.5%	0.000	3	5.5%
Logistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Lognormal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.000	2	3.6%	0.000	2	3.6%	0.000	2	3.6%	0.000	2	3.6%
Weibull	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%

Table 113: NLL, AIC, AICc, and BIC Results for 3-Qubit Randomized Benchmarking Seed 0 err_a. The green cells indicate the best-fit distribution type(s).

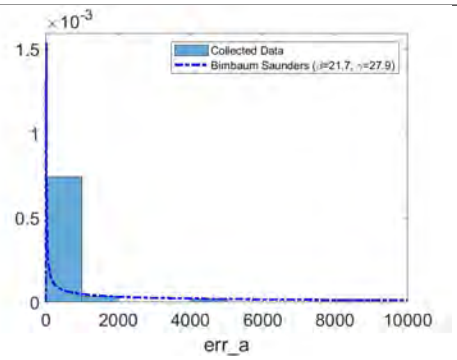
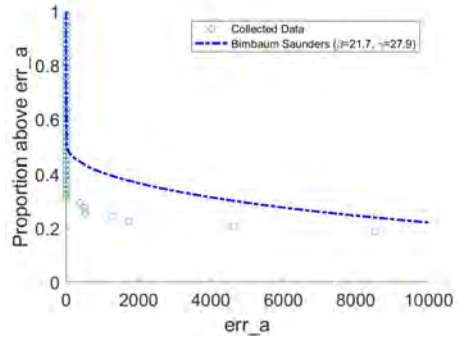
	Probability Density Function	Survival Function
err_a		

Table 114: Distribution Model Analysis for ibmq_yorktown Q00_Q01_Q02 3-Qubit Seed 0 *err_a*. In the left column, the probability density function for the Birnbaum Saunders distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the Birnbaum Saunders distribution model is plotted on top of the proportion of collected data observations above a given *err_a*.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.328	28	50.90%	0.380	28	50.90%	0.390	28	50.90%	0.398	28	50.90%
Burr	0.671	17	30.90%	0.620	17	30.90%	0.610	17	30.90%	0.602	17	30.90%
Exponential	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Extreme Value	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Gamma	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Generalized Extreme Value	0.000	9	16.40%	0.000	9	16.40%	0.000	9	16.40%	0.000	9	16.40%
Generalized Pareto	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Half-Normal	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Inverse Gaussian	0.001	0	0.00%	0.000	1	1.80%	0.000	1	1.80%	0.000	1	1.80%
Logistic	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Loglogistic	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Lognormal	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Nakagami	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Normal	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Rayleigh	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Rician	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
t Location Scale	0.000	1	1.80%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%
Weibull	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%	0.000	0	0.00%

Table 115: NLL, AIC, AICc, and BIC Results for 3-Qubit Randomized Benchmarking Seed 0 err_α. The green cells indicate the best-fit distribution type(s).

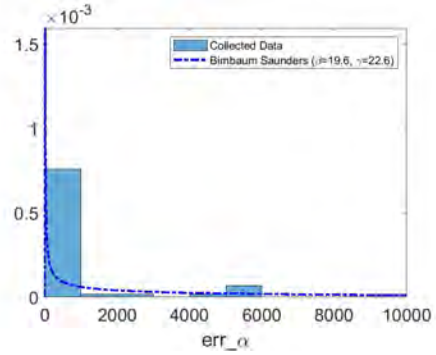
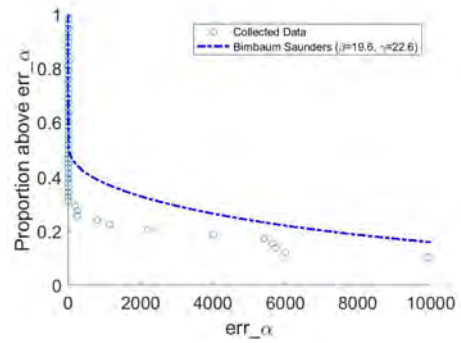
	Probability Density Function	Survival Function
err_α		

Table 116: Distribution Model Analysis for ibmq_yorktown Q00_Q01_Q02 3-Qubit Seed 0 err_α. In the left column, the probability density functions for the Birnbaum Saunders distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival function for the Birnbaum Saunders distribution model is plotted on top of the proportion of collected data observations above a given err_α.

For 3-Qubit Randomized Benchmarking Seed 0 err_b, the gamma, GEV, loglogistic, and Nakagami distributions are selected as the best-fit distributions (see Tables 117 and 118). For the GEV distribution, 56% of the qubit sets are Type III (Weibull), and 44% are Type II (Fréchet). The gamma distribution is commonly used when the data has a natural minimum of zero (the amount of error on the b parameter has a natural minimum of zero) [16]. The gamma probability density function is given by the expression,

$$f(x|a,b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{\frac{-x}{b}}, \tag{39}$$

where $a > 0$ is the shape parameter and $b > 0$ is the scale parameter [15].

The Nakagami distribution has the density function,

$$f(x|\mu,\omega) = 2\left(\frac{\mu}{\omega}\right)^{\mu} \frac{1}{\Gamma(\mu)} x^{2\mu-1} \exp\left(\frac{-\mu x^2}{\omega}\right), \tag{40}$$

where μ is the shape parameter and $\omega > 0$ is the scale parameter for $x > 0$ [27]. Nakagami distributions are often model signal scattering; it would make an interesting future study to determine what type of relationship exists between the err_b parameter and scattering from Clifford gate signals.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.074	1	1.8%	0.089	1	1.8%	0.093	1	1.8%	0.107	1	1.8%
Burr	0.137	15	27.3%	0.108	3	5.5%	0.097	3	5.5%	0.050	2	3.6%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.002	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Gamma	0.086	4	7.3%	0.113	7	12.7%	0.116	7	12.7%	0.132	7	12.7%
Generalized Extreme Value	0.138	17	30.9%	0.115	1	1.8%	0.106	1	1.8%	0.053	1	1.8%
Generalized Pareto	0.000	5	9.1%	0.000	4	7.3%	0.000	4	7.3%	0.000	3	5.5%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.074	3	5.5%	0.084	5	9.1%	0.086	5	9.1%	0.095	5	9.1%
Logistic	0.040	0	0.0%	0.027	1	1.8%	0.027	2	3.6%	0.031	2	3.6%
Loglogistic	0.067	1	1.8%	0.074	11	20.0%	0.076	11	20.0%	0.086	12	21.8%
Lognormal	0.076	0	0.0%	0.111	2	3.6%	0.113	2	3.6%	0.121	2	3.6%
Nakagami	0.082	4	7.3%	0.112	11	20.0%	0.114	11	20.0%	0.134	10	18.2%
Normal	0.044	0	0.0%	0.026	1	1.8%	0.027	1	1.8%	0.031	1	1.8%
Rayleigh	0.004	0	0.0%	0.001	0	0.0%	0.001	0	0.0%	0.002	1	1.8%
Rician	0.058	1	1.8%	0.055	3	5.5%	0.056	3	5.5%	0.068	3	5.5%
t Location Scale	0.057	2	3.6%	0.021	1	1.8%	0.018	0	0.0%	0.010	0	0.0%
Weibull	0.062	2	3.6%	0.065	4	7.3%	0.068	4	7.3%	0.081	5	9.1%

Table 117: NLL, AIC, AICc, and BIC Results for 3-Qubit Randomized Benchmarking Seed 0 *err_b*. The green cells indicate the best-fit distribution type(s).

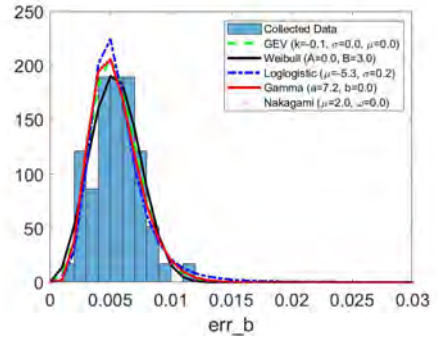
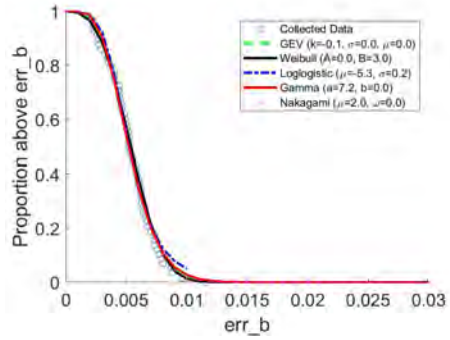
	Probability Density Function	Survival Function
err_b		

Table 118: Distribution Model Analysis for ibmq_yorktown Q00_Q01_Q02 3-Qubit Seed 0 *err_b*. In the left column, the probability density functions for the GEV, Weibull, loglogistic, gamma, and Nakagami distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, Weibull, loglogistic, gamma, and Nakagami distribution models are plotted on top of the proportion of collected data observations above a given *err_b*.

For 3-Qubit Randomized Benchmarking Seed 4 Fidelity, the Burr, GEV, logistic, loglogistic, and t location-scale distributions are identified as the best-fit distributions (see Tables 119 and 120). For the GEV distribution, 96% of the qubit sets are Type III (Weibull). Surprisingly, Seed 4 results are more varied than Seed 0 results. Future work should investigate the reason for this variation.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.051	0	0.0%	0.045	0	0.0%	0.047	0	0.0%	0.050	0	0.0%
Burr	0.131	8	14.8%	0.099	3	5.6%	0.089	3	5.6%	0.048	2	3.7%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.012	0	0.0%	0.003	1	1.9%	0.003	1	1.9%	0.003	2	3.7%
Gamma	0.061	0	0.0%	0.070	0	0.0%	0.072	0	0.0%	0.076	0	0.0%
Generalized Extreme Value	0.112	22	40.7%	0.081	5	9.3%	0.064	4	7.4%	0.040	0	0.0%
Generalized Pareto	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.051	0	0.0%	0.045	7	13.0%	0.047	8	14.8%	0.050	9	16.7%
Logistic	0.089	3	5.6%	0.126	11	20.4%	0.136	11	20.4%	0.163	14	25.9%
Loglogistic	0.094	0	0.0%	0.131	4	7.4%	0.139	4	7.4%	0.167	8	14.8%
Lognormal	0.051	0	0.0%	0.046	0	0.0%	0.047	0	0.0%	0.051	0	0.0%
Nakagami	0.063	0	0.0%	0.074	2	3.7%	0.081	2	3.7%	0.087	2	3.7%
Normal	0.070	0	0.0%	0.089	0	0.0%	0.093	0	0.0%	0.104	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.070	0	0.0%	0.090	4	7.4%	0.094	4	7.4%	0.105	5	9.3%
t Location Scale	0.119	21	38.9%	0.092	12	22.2%	0.078	12	22.2%	0.042	7	13.0%
Weibull	0.026	0	0.0%	0.010	5	9.3%	0.011	5	9.3%	0.013	5	9.3%

Table 119: NLL, AIC, AICc, and BIC Results for 3-Qubit Randomized Benchmarking Seed 4 Fidelity. The green cells indicate the best-fit distribution type(s).

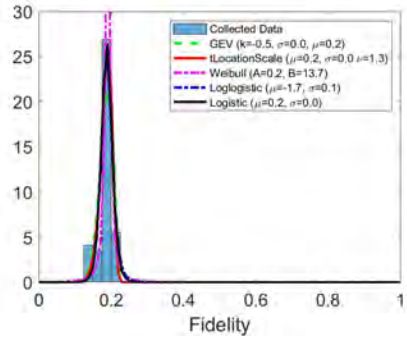
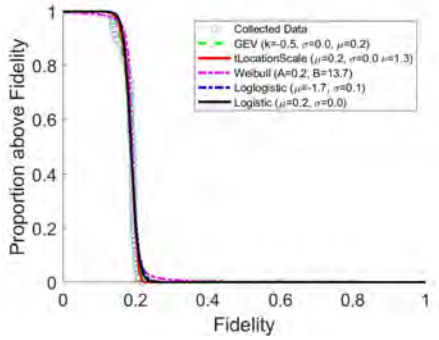
	Probability Density Function	Survival Function
Fidelity		

Table 120: Distribution Model Analysis for ibmq_yorktown Q00_Q01_Q02 3-Qubit Seed 4 Fidelity. In the left column, the probability density functions for the GEV, t location-scale, Weibull, loglogistic, and logistic distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, t location-scale, Weibull, loglogistic, and logistic distribution models are plotted on top of the proportion of collected data observations above a given fidelity.

For 3-Qubit Randomized Benchmarking Seed 4 a, the Burr and t location-scale distributions are identified as the best-fit distributions, and the generalized Pareto distribution is identified as a second-best-fit distribution (see Tables 121 and 122). Bimodal behavior observed in Table 122 explains why the generalized Pareto distribution emerged as a distribution with a good fit. However, the fact that the generalized Pareto distribution emerged as a second-best-fit rather than best-fit distribution suggests less bimodal or multi-modal behavior for this characteristic compared to Seed 0, which should be investigated in future work.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.007	0	0.0%	0.000	1	1.9%	0.000	1	1.9%	0.001	1	1.9%
Burr	0.471	9	16.7%	0.590	8	14.8%	0.603	8	14.8%	0.566	8	14.8%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.000	1	1.9%	0.000	4	7.4%	0.000	4	7.4%	0.000	4	7.4%
Gamma	0.007	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.001	0	0.0%
Generalized Extreme Value	0.099	10	18.5%	0.018	7	13.0%	0.020	7	13.0%	0.016	4	7.4%
Generalized Pareto	0.000	11	20.4%	0.000	11	20.4%	0.000	11	20.4%	0.000	11	20.4%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.007	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.001	3	5.6%
Logistic	0.047	0	0.0%	0.006	0	0.0%	0.008	0	0.0%	0.020	0	0.0%
Loglogistic	0.065	0	0.0%	0.024	1	1.9%	0.032	1	1.9%	0.085	2	3.7%
Lognormal	0.007	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.001	0	0.0%
Nakagami	0.006	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.001	0	0.0%
Normal	0.004	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.005	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.271	23	42.6%	0.361	19	35.2%	0.335	19	35.2%	0.310	18	33.3%
Weibull	0.003	0	0.0%	0.000	3	5.6%	0.000	3	5.6%	0.000	3	5.6%

Table 121: NLL, AIC, AICc, and BIC Results for 3-Qubit Randomized Benchmarking Seed 4 a. The green cells indicate the best-fit distribution type(s).

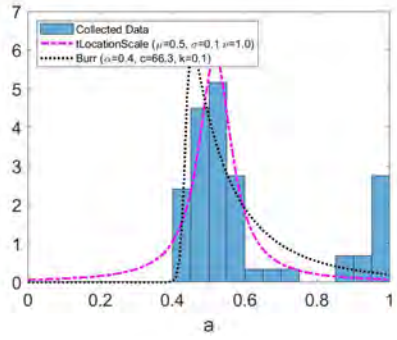
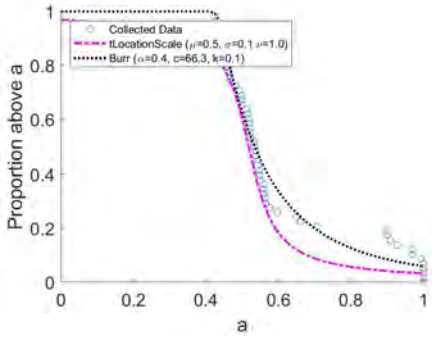
	Probability Density Function	Survival Function
a		

Table 122: Distribution Model Analysis for ibmq_yorktown Q00_Q01_Q02 3-Qubit Seed 4 a. In the left column, the probability density functions for the t location-scale and Burr distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the t location-scale and Burr distribution models are plotted on top of the proportion of collected data observations above a given a.

For 3-Qubit Randomized Benchmarking Seed 4 α, the GEV distribution is identified as the best-fit distribution (98% of the qubit sets are Type III (Weibull)), with the generalized Pareto distribution identified as the second-best-fit distribution (see Tables 123 and 124). Again, observed bimodal behavior in Table 124 should be explored in future research.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Burr	0.228	1	1.9%	0.220	0	0.0%	0.218	0	0.0%	0.129	0	0.0%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.007	1	1.9%	0.000	3	5.6%	0.000	4	7.4%	0.000	4	7.4%
Gamma	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.653	23	42.6%	0.765	21	38.9%	0.767	20	37.0%	0.854	19	35.2%
Generalized Pareto	0.013	16	29.6%	0.000	16	29.6%	0.000	16	29.6%	0.000	16	29.6%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Logistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Lognormal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.000	0	0.0%	0.000	0	0.0%	0.000	1	1.9%	0.000	1	1.9%
t Location Scale	0.097	13	24.1%	0.014	12	22.2%	0.014	11	20.4%	0.017	11	20.4%
Weibull	0.001	0	0.0%	0.000	2	3.7%	0.000	2	3.7%	0.000	3	5.6%

Table 123: NLL, AIC, AICc, and BIC Results for 3-Qubit Randomized Benchmarking Seed 4 α. The green cells indicate the best-fit distribution type(s).

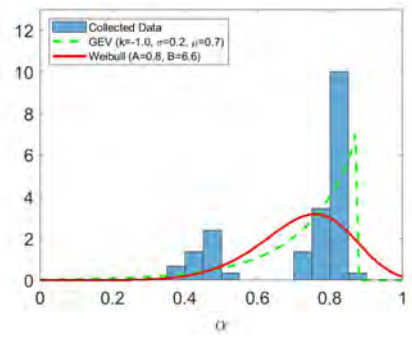
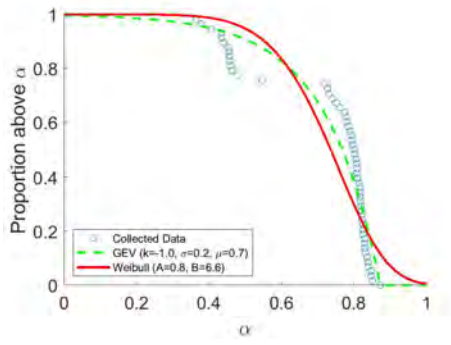
	Probability Density Function	Survival Function
α		

Table 124: Distribution Model Analysis for ibmq_yorktown Q00_Q01_Q02 3-Qubit Seed 4 α. In the left column, the probability density functions for the GEV and Weibull distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV and Weibull distribution models are plotted on top of the proportion of collected data observations above a given α.

For 3-Qubit Randomized Benchmarking Seed 4 b, the Burr, logistic, loglogistic, and t location-scale distributions are identified as the best-fit distributions (see Tables 125 and 126). All the best-fit distribution types are well-suited for non-negative, positively skewed data. However, for the qubit set ibmq_yorktown Q00_Q01_Q02 in Table 126, MATLAB could not fit the Burr distribution to the data. Future research should investigate what qubit set properties cause the Burr distribution to fit or not fit the data.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.048	0	0.0%	0.039	0	0.0%	0.041	0	0.0%	0.047	0	0.0%
Burr	0.132	10	18.5%	0.109	2	3.7%	0.096	2	3.7%	0.051	2	3.7%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.012	0	0.0%	0.002	1	1.9%	0.003	1	1.9%	0.003	2	3.7%
Gamma	0.058	0	0.0%	0.057	1	1.9%	0.060	1	1.9%	0.068	1	1.9%
Generalized Extreme Value	0.114	20	37.0%	0.079	4	7.4%	0.064	3	5.6%	0.041	0	0.0%
Generalized Pareto	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.048	0	0.0%	0.039	7	13.0%	0.041	8	14.8%	0.047	9	16.7%
Logistic	0.089	2	3.7%	0.117	10	18.5%	0.127	10	18.5%	0.166	13	24.1%
Loglogistic	0.090	0	0.0%	0.130	5	9.3%	0.141	5	9.3%	0.159	8	14.8%
Lognormal	0.048	0	0.0%	0.040	0	0.0%	0.042	0	0.0%	0.047	0	0.0%
Nakagami	0.066	0	0.0%	0.082	1	1.9%	0.086	1	1.9%	0.094	1	1.9%
Normal	0.070	0	0.0%	0.090	0	0.0%	0.096	0	0.0%	0.104	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.070	0	0.0%	0.091	4	7.4%	0.096	4	7.4%	0.105	5	9.3%
t Location Scale	0.131	22	40.7%	0.115	14	25.9%	0.097	14	25.9%	0.055	8	14.8%
Weibull	0.026	0	0.0%	0.010	5	9.3%	0.011	5	9.3%	0.013	5	9.3%

Table 125: NLL, AIC, AICc, and BIC Results for 3-Qubit Randomized Benchmarking Seed 4 *b*. The green cells indicate the best-fit distribution type(s).

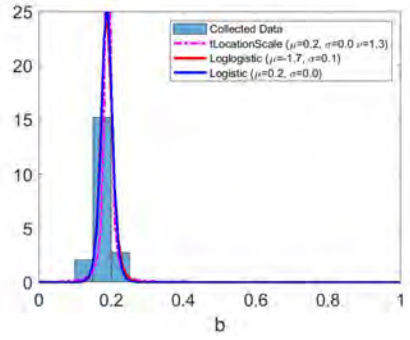
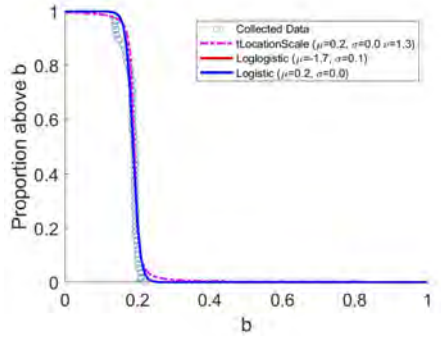
	Probability Density Function	Survival Function
b		

Table 126: Distribution Model Analysis for ibmq_yorktown Q00_Q01_Q02 3-Qubit Seed 4 b. In the left column, the probability density functions for the t location-scale, loglogistic, and logistic distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the t location-scale, loglogistic, and logistic distribution models are plotted on top of the proportion of collected data observations above a given b.

Results for 3-Qubit Randomized Benchmarking Seed 4 err_a and err_α are similar to Seed 0. The Birnbaum Saunders and Burr distributions are identified as best-fit distributions (see Tables 127–130). The probability density functions in Tables 128 and 130 show the same patterns as Seed 0, and the data points should be investigated to determine if these seemingly anomalous values are outliers or if they are actually characteristic of qubit set behavior.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.256	26	48.1%	0.326	26	48.1%	0.332	26	48.1%	0.351	26	48.1%
Burr	0.741	19	35.2%	0.674	16	29.6%	0.668	16	29.6%	0.649	16	29.6%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Gamma	0.000	2	3.7%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.002	4	7.4%	0.000	3	5.6%	0.000	3	5.6%	0.000	3	5.6%
Generalized Pareto	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.000	3	5.6%	0.000	3	5.6%	0.000	3	5.6%	0.000	3	5.6%
Logistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.000	0	0.0%	0.000	2	3.7%	0.000	2	3.7%	0.000	2	3.7%
Lognormal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Nakagami	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	4	7.4%	0.000	4	7.4%	0.000	4	7.4%
Rician	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Weibull	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%

Table 127: NLL, AIC, AICc, and BIC Results for 3-Qubit Randomized Benchmarking Seed 4 *err_a*. The green cells indicate the best-fit distribution type(s).

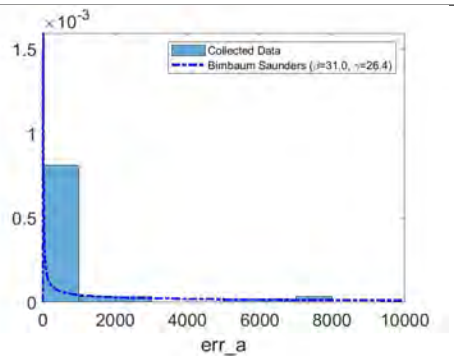
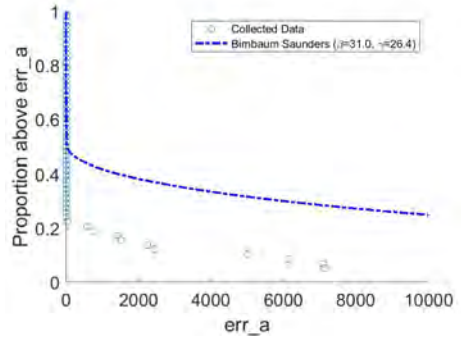
	Probability Density Function	Survival Function
err_a		

Table 128: Distribution Model Analysis for ibmq_yorktown Q00_Q01_Q02 3-Qubit Seed 4 *err_a*. In the left column, the probability density function for the Birnbaum Saunders distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival function for the Birnbaum Saunders distribution model is plotted on top of the proportion of collected data observations above a given *err_a*.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.176	24	44.4%	0.228	25	46.3%	0.231	25	46.3%	0.243	25	46.3%
Burr	0.677	17	31.5%	0.754	14	25.9%	0.752	14	25.9%	0.744	14	25.9%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Gamma	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Generalized Extreme Value	0.146	8	14.8%	0.018	6	11.1%	0.017	6	11.1%	0.013	6	11.1%
Generalized Pareto	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.000	3	5.6%	0.000	5	9.3%	0.000	5	9.3%	0.000	5	9.3%
Logistic	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Loglogistic	0.000	0	0.0%	0.000	3	5.6%	0.000	3	5.6%	0.000	3	5.6%
Lognormal	0.000	1	1.9%	0.000	1	1.9%	0.000	1	1.9%	0.000	1	1.9%
Nakagami	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rayleigh	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Rician	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
t Location Scale	0.000	1	1.9%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Weibull	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%

Table 129: NLL, AIC, AICc, and BIC Results for 3-Qubit Randomized Benchmarking Seed 4 err_α. The green cells indicate the best-fit distribution type(s).

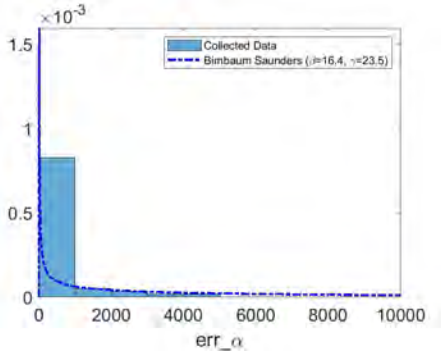
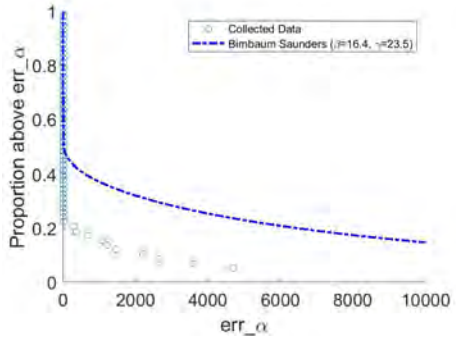
	Probability Density Function	Survival Function
err_α		

Table 130: Distribution Model Analysis for ibmq_yorktown Q00_Q01_Q02 3-Qubit Seed 4 err_α. In the left column, the probability density function for the Birnbaum Saunders distribution model is plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival function for the Birnbaum Saunders distribution model is plotted on top of the proportion of collected data observations above a given err_α.

For 3-Qubit Randomized Benchmarking Seed 4 err_b, the Burr, gamma, and GEV distributions are identified as best-fit distributions (see Tables 131 and 132). For the GEV distributions, 65% of the qubit sets are Type III (Weibull). Seed 4 results differ significantly from Seed 0 results. Comparing the histograms from Table 118 for Seed 0 and Table 132 for Seed 4, the err_b distribution for Seed 4 is narrower and closer to zero than the distribution for Seed 0, which meets expectations for five seeds of data versus one seed of data.

	Negative Log-likelihood			AIC			AICc			BIC		
	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%	Norm. Med. Value	Count	%
Birnbaum Saunders	0.075	4	7.4%	0.090	4	7.4%	0.091	4	7.4%	0.100	4	7.4%
Burr	0.125	16	29.6%	0.092	0	0.0%	0.082	0	0.0%	0.041	0	0.0%
Exponential	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Extreme Value	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	1	1.9%
Gamma	0.100	4	7.4%	0.158	14	25.9%	0.162	14	25.9%	0.172	14	25.9%
Generalized Extreme Value	0.137	16	29.6%	0.098	1	1.9%	0.087	1	1.9%	0.044	0	0.0%
Generalized Pareto	0.000	3	5.6%	0.000	2	3.7%	0.000	2	3.7%	0.000	2	3.7%
Half-Normal	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%	0.000	0	0.0%
Inverse Gaussian	0.070	3	5.6%	0.078	5	9.3%	0.080	5	9.3%	0.086	5	9.3%
Logistic	0.031	0	0.0%	0.013	0	0.0%	0.014	0	0.0%	0.017	0	0.0%
Loglogistic	0.078	0	0.0%	0.091	12	22.2%	0.093	12	22.2%	0.093	12	22.2%
Lognormal	0.081	0	0.0%	0.108	1	1.9%	0.112	1	1.9%	0.128	1	1.9%
Nakagami	0.088	3	5.6%	0.121	5	9.3%	0.124	5	9.3%	0.139	5	9.3%
Normal	0.039	0	0.0%	0.025	0	0.0%	0.026	0	0.0%	0.029	0	0.0%
Rayleigh	0.009	0	0.0%	0.003	2	3.7%	0.003	3	5.6%	0.009	3	5.6%
Rician	0.055	0	0.0%	0.047	2	3.7%	0.049	2	3.7%	0.057	2	3.7%
t Location Scale	0.051	1	1.9%	0.014	1	1.9%	0.012	1	1.9%	0.008	1	1.9%
Weibull	0.060	4	7.4%	0.063	5	9.3%	0.068	4	7.4%	0.077	4	7.4%

Table 131: NLL, AIC, AICc, and BIC Results for 3-Qubit Randomized Benchmarking Seed 4 *err_b*. The green cells indicate the best-fit distribution type(s).

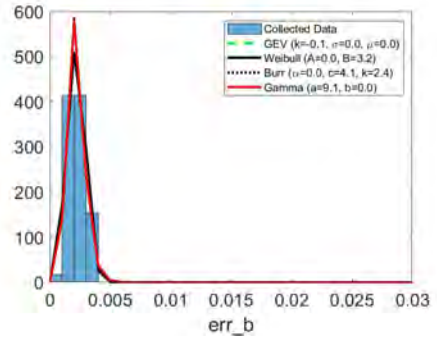
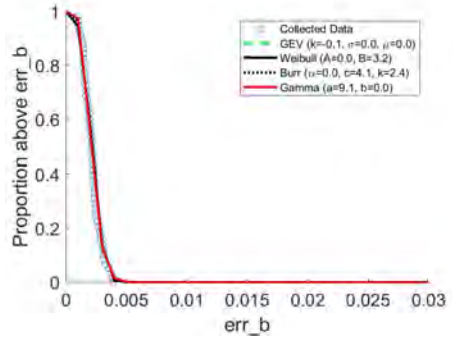
	Probability Density Function	Survival Function
err_b		

Table 132: Distribution Model Analysis for ibmq_yorktown Q00_Q01_Q02 3-Qubit Seed 4 *err_b*. In the left column, the probability density functions for the GEV, Weibull, Burr, and gamma distribution models are plotted on top of the histogram of the collected data. The histogram is normalized so that the area of each bar is the relative number of observations. In the right column, the survival functions for the GEV, Weibull, Burr, and gamma distribution models are plotted on top of the proportion of collected data observations above a given *err_b*.

4.3.2 Part B2 – Reduce Sampling Errors

4.3.2.1 *t*-test/*F*-test vs. KS Test for Calculating Sampling Errors

Tables 134 through 141 show the sampling error rate results from calculating sampling errors with the t-test/F-test and the KS test with different types of data conditioning. On the vertical axis in each image, the size of N1 increases from 20, 40, ... 480. On the horizontal axis, the size of N2 increases from 20, 40, ... 480. Given that none of the characteristics in Part B1 follow a normal distribution, the KS test is anticipated to outperform the t-test/F-test for calculating sampling errors. Results provided in this section are in accordance with this prediction.

For no data conditioning in the first stage (Table 133), the sampling error rates are nearly equivalent between the t-test/F-test and the KS test for T1 and T2 sampling error rates. For single-qubit, multi-qubit, and readout error rates, the sampling error rates are slightly reduced for the KS test. For data conditioning with median outlier

removal (Table 136) in the second stage, the sampling error rates are nearly equivalent between the t-test/F-test and the KS test for all characteristics.

For data conditioning with data set truncation (Table 137) and data set conditioning with both data set truncation and median outlier removal (Table 139), the KS tests show slightly lower sampling error rates for T1 and T2 tests than the t-test/F-test, and markedly lower sampling error rates for single-qubit, multi-qubit, and readout error rates. This pattern of observations suggests that the data for T1 and T2 coherence times more closely follow a normal distribution than the data for single-qubit, multi-qubit, and readout error rates. Results for Part B1 confirm that while T1 and T2 coherence times do have distributions closer to normal than the error rates, the normal distribution is not the best-fit distribution type for any of the characteristics. Given the results for both the sampling error rates and the best-fit distribution types, it is clear that the KS test is the superior method for calculating sampling error rates.

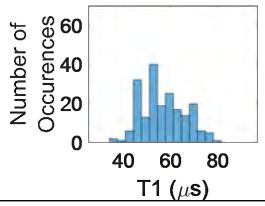
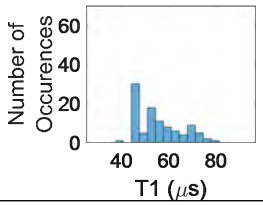
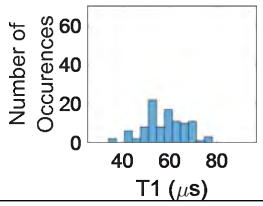
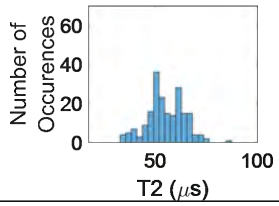
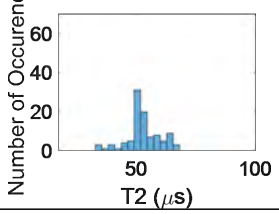
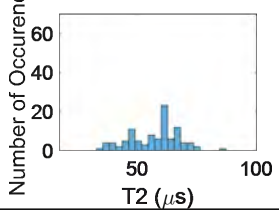
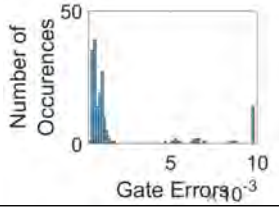
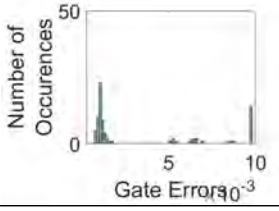
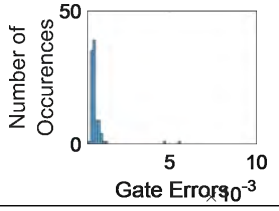
4.3.2.2 Data Conditioning Methods

Results for the data conditioning methods are also displayed in Tables 134 through 141. Since results demonstrate that the KS test outperforms the t-test/F-test for calculating sampling error rates, discussion is provided only for the KS test.

4.3.2.3 First Stage – No Data Conditioning

For the first stage, no conditioning occurs in order to form a baseline for sampling errors. The histograms and corresponding means and standard deviations for the population, N1 sample, and N2 sample (sample sizes here are selected as half of the population) are presented in Table 133 for each of the calibration characteristics for ibmq_yorktown qubit Q00. As none of the characteristics are normally distributed,

a change in the mean or standard deviation of a characteristic is sufficient but not necessary to conclude that the distribution changes in an stage of data conditioning. The baseline sampling errors are shown in Table 134.

	Population	N1 (DP 1-100)	N2 (DP 101-201)
T1s			
			
Mean	56.56	55.46	57.64
St. Dev.	9.28	9.58	8.89
T2s			
			
Mean	55.05	52.96	57.12
St. Dev.	9.05	7.03	10.30
Single-Qubit Gate Error (Outliers above 10x10-3 not shown)			
			
Mean	0.161	0.323	0.001
St. Dev.	0.542	0.736	0.001
Multi-Qubit Gate Error (Outliers above 0.1 not shown)			

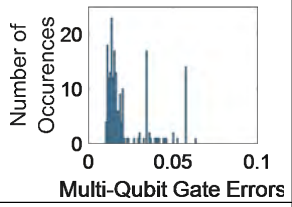
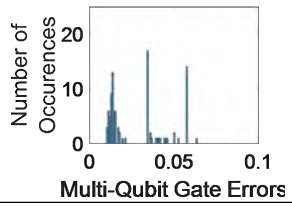
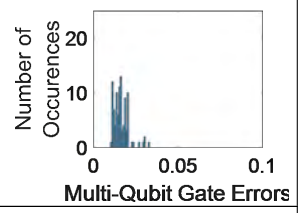
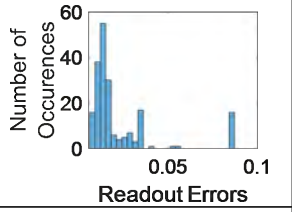
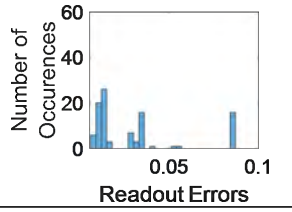
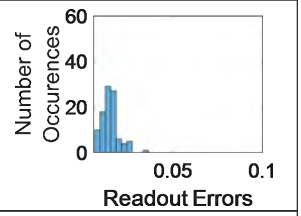
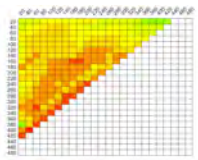
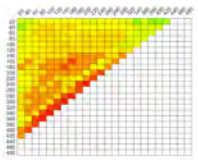
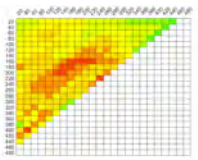
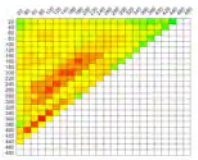
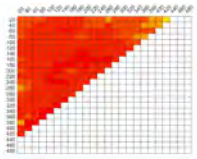
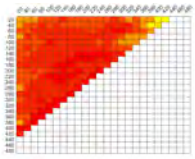
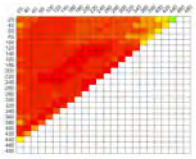
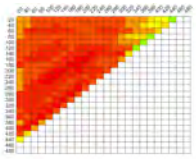
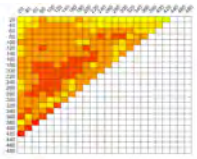
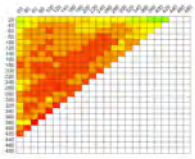
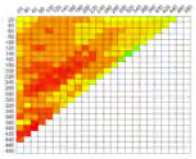
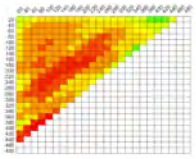
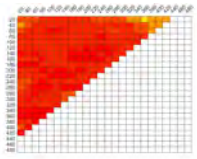
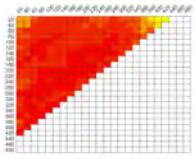
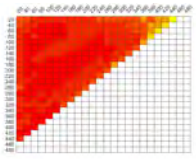
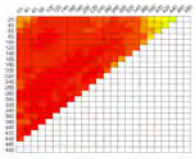
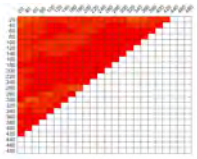
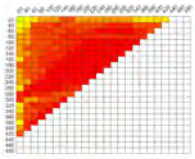
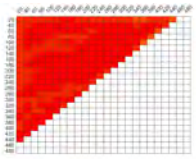
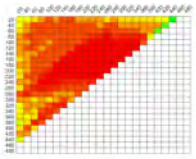
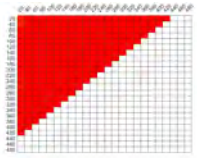
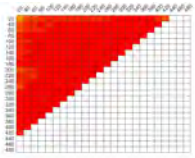
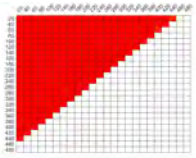
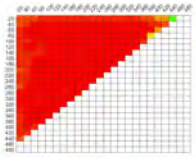
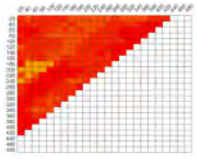
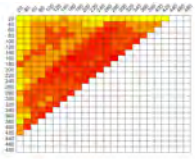
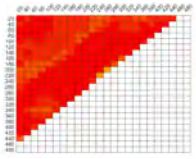
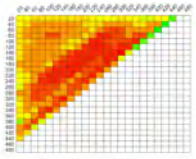
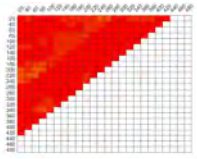
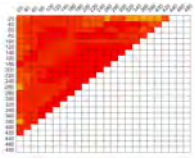
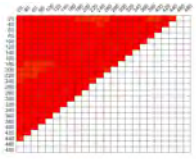
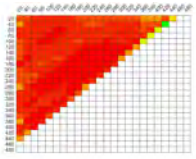
			
Mean	.027	.028	.026
St. Dev.	.070	.017	.098
Readout Error			
			
Mean	.025	.030	.019
St. Dev.	.036	.027	.042

Table 133: ibmq_yorktown Q00 Histograms and Corresponding Means and Standard Deviations for the Population, N1 Sample, and N2 Sample of Each Characteristics; No Data Conditioning

	PFA		PD	
	Color Scale for PFA and PD Sampling Error Rate 			
	t-test/F-test	KS Test	t-test/F-test	KS Test
	T1			
$\alpha = 0.001$				

$\alpha = 0.05$				
	T2			
$\alpha = 0.001$				
$\alpha = 0.05$				
	Single-Qubit Gate Error Rate			
$\alpha = 0.001$				
$\alpha = 0.05$				
	Multi-Qubit Gate Error Rate			
$\alpha = 0.001$				
$\alpha = 0.05$				
	Readout Error Rate			

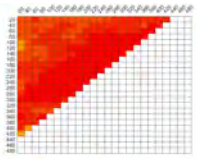
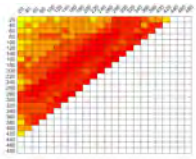
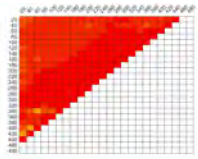
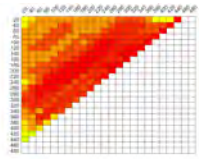
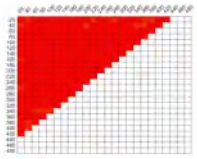
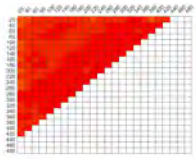
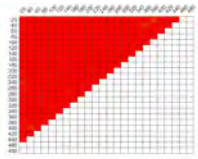
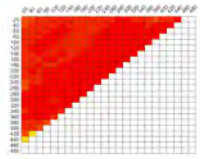
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 134: Comparison of t-test/F-test vs. KS Test for Determining if a Sampling Error Exists; No Data Conditioning

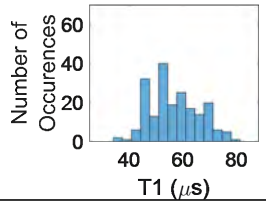
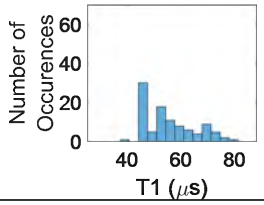
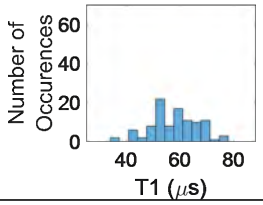
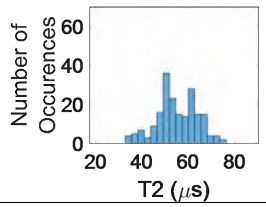
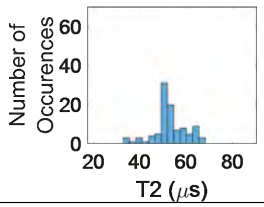
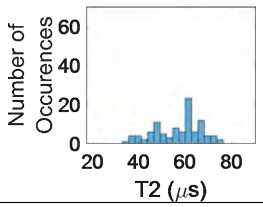
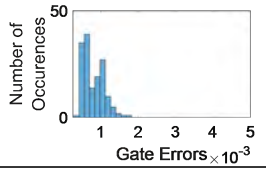
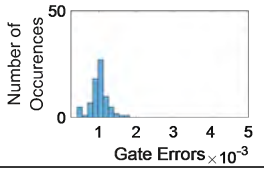
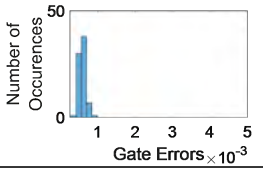
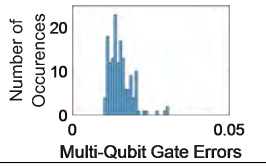
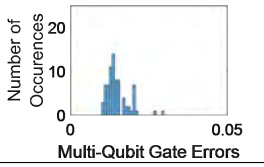
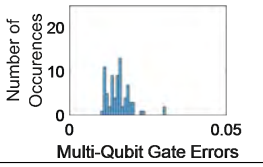
4.3.2.4 Second Stage – Outlier Removal

For the second stage, outliers are removed from the data set using five different outlier removal methods: quartiles, mean, median, generalized extreme Studentized deviate (GESD), and Grubbs. Only the results for the median outlier method, which has a slightly greater effect than the other methods, are presented in the main text. Comparative sampling error rate results for the five outlier removal methods (calculated with the t-test/F-test) are provided in the appendix.

The histograms and corresponding means and standard deviations for the population, N1 sample, and N2 sample (sample sizes here are selected as half of the population) with outliers removed using the median method are presented in Table 135. The removal of outliers had the greatest effect on changing the population and sample parameters of the single-qubit gate error, multi-qubit gate error, and readout error characteristics.

It can be seen in Table 136 that median outlier removal without any data set truncation provides some improvement in sampling error rates for each of the characteristics. In particular, the greatest improvement occurred for the error rate char-

acteristics.

	Population	N1 (DP 1-100)	N2 (DP 101-201)
T1s			
			
Mean	56.56	55.46	57.64
St. Dev.	9.28	9.57	8.89
T2s			
			
Mean	54.90	52.96	56.71
St. Dev.	8.82	7.03	10.0
Single-Qubit Gate Error			
			
Mean	7.98e-4	1.00e-3	5.89e-4
St. Dev.	2.74e-4	2.36e-4	8.73e-5
Multi-Qubit Gate Error			
			
Mean	1.53e-2	1.48e-2	1.58e-2
St. Dev.	3.67e-2	3.50e-2	3.79e-2

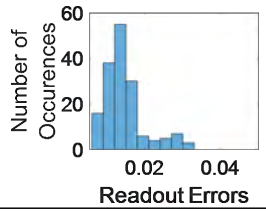
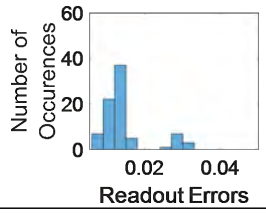
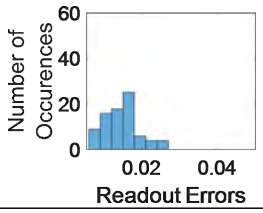
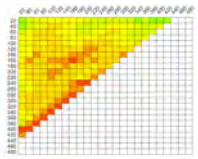
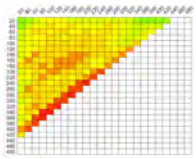
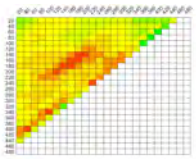
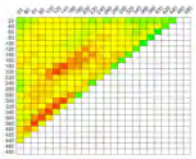
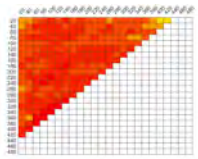
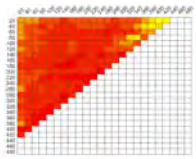
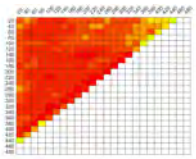
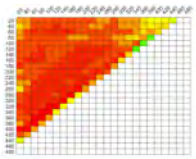
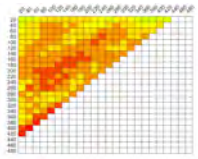
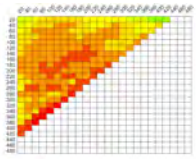
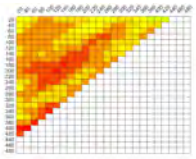
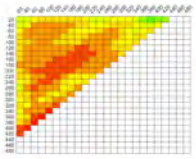
Readout Error			
			
Mean	1.47e-2	1.47e-2	1.47e-2
St. Dev.	5.43e-2	6.14e-2	4.64e-3

Table 135: ibmq_yorktown Q00 Histograms and Corresponding Means and Standard Deviations for the Population, N1 Sample, and N2 Sample of Each Calibration Characteristic; Outliers Removed with Median Method

	PFA		PD	
	Color Scale for PFA and PD Sampling Error Rate 			
	t-test/F-test	KS Test	t-test/F-test	KS Test
	T1			
$\alpha = 0.001$				
$\alpha = 0.05$				
	T2			
$\alpha = 0.001$				

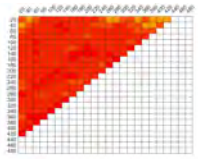
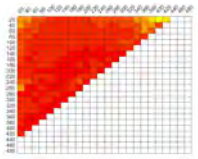
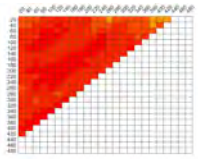
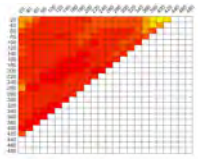
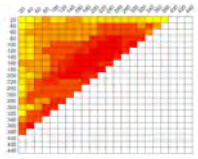
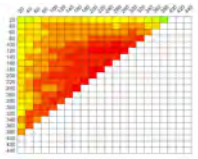
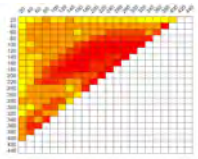
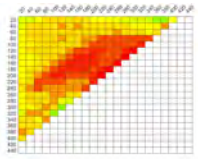
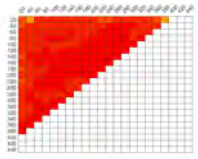
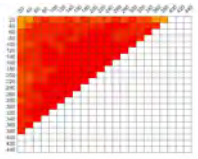
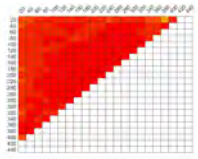
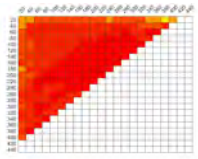
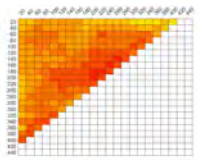
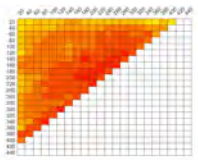
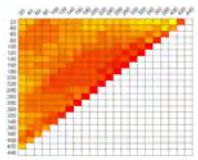
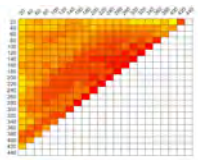
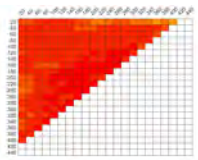
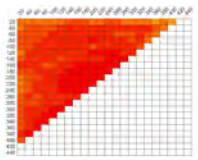
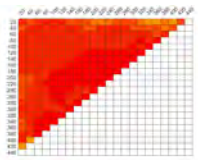
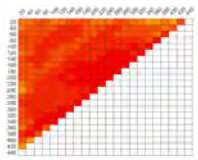
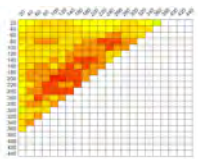
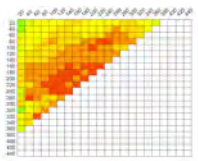
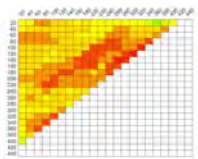
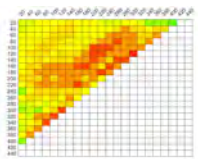
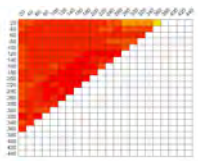
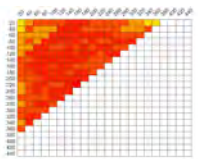
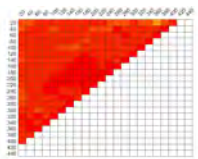
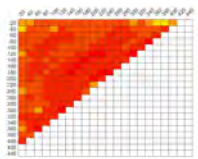
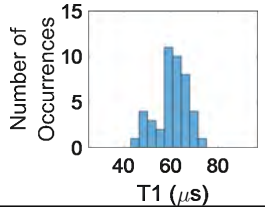
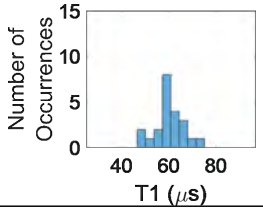
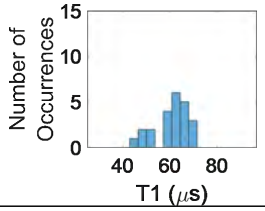
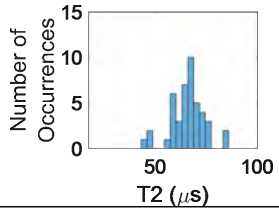
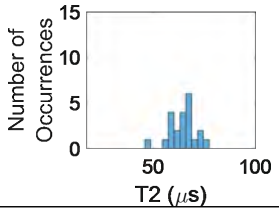
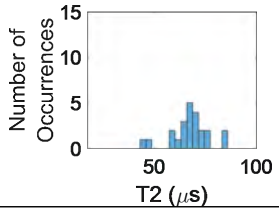
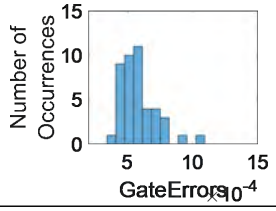
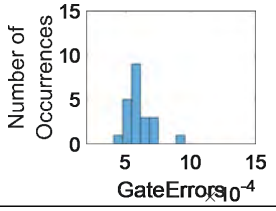
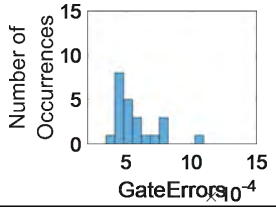
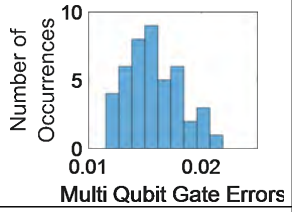
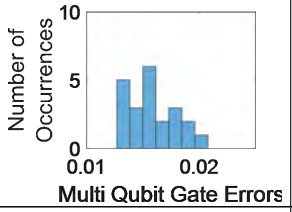
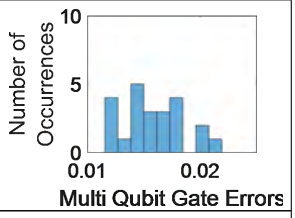
$\alpha = 0.05$				
	Single-Qubit Gate Error Rate			
$\alpha = 0.001$				
$\alpha = 0.05$				
	Multi-Qubit Gate Error Rate			
$\alpha = 0.001$				
$\alpha = 0.05$				
	Readout Error Rate			
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 136: Comparison of t-test/F-test vs. KS Test for Determining if a Sampling Error Exists; Median Outlier Removal

4.3.2.5 Third Stage – Data Set Truncation

Compared to outlier removal, data set truncation has a much greater effect on changing the population and sample parameters and on reducing sampling errors. The histograms and corresponding means and standard deviations for the population, N1 sample, and N2 sample with data set truncation are presented in Table 137. The sampling error rates for data set truncation are presented in Table 138. For T1 and T2 coherence times and for single-qubit and readout error rates, the sampling error rates are markedly reduced, in many cases to near zero, for $\alpha = 0.001$. While sampling error rates are reduced significantly for multi-qubit error rates, sampling errors are still prevalent.

Adding median outlier removal to the truncated data sets (Table 139) has a small effect in reducing sampling error rates for each of the characteristics. However, it is unclear at this point whether data points identified as outliers are actually outliers or are indeed data points characteristic of the qubit's behavior. Thus, it is uncertain whether or not it is valid to remove the outliers from the data set. As such, probability of detection (PD)/probability of false alarm (PFA) analysis and determination of the best-fit distribution types will proceed for data set truncation without outlier removal, but with one exception. For the multi-qubit gate error rates, it was determined that the data points equal to 1 are appropriate to remove from the data set. Such datapoints are not characteristic of the quantum processor, but rather, "it is the automated calibration not fitting properly," and the calibration "just needs to be re-run" [124]. Thus, removing multi-qubit gate error rates equal to 1 from the truncated data set is the only outlier removal presented in the remainder of this work. By removing the multi-qubit error rates equal to 1 from the truncated data set (Table 141), the sampling error rates improve very slightly compared to Table 137, particularly for PFA.

	Population	N1 (DP 1-100)	N2 (DP 101-201)
T1s			
			
Mean	60.52	59.90	61.00
St. Dev.	6.60	6.03	7.08
T2s			
			
Mean	65.81	64.14	67.48
St. Dev.	8.15	6.22	9.34
Single-Qubit Gate Error			
			
Mean	5.81e-4	5.99e-4	5.62e-4
St. Dev.	1.37e-4	9.54e-5	1.65e-4
Multi-Qubit Gate Error			
			
Mean	1.58e-2	1.59e-2	1.58e-2

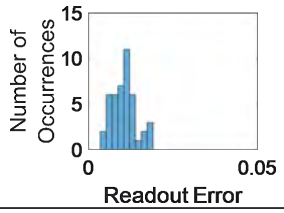
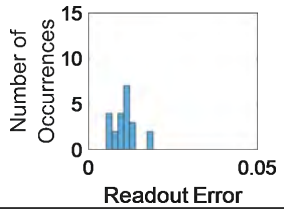
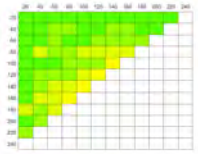
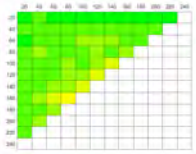
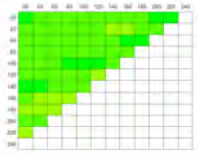
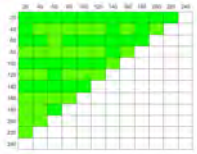
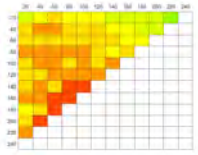
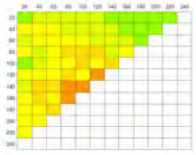
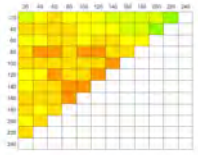
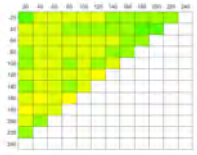
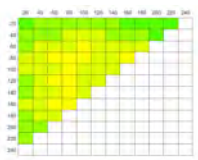
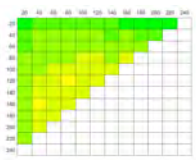
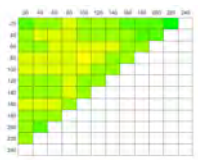
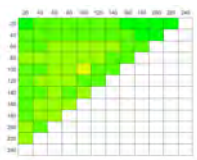
St. Dev.	2.31e-3	1.99e-3	2.60e-3
Readout Error			
			
Mean	1.06e-2	1.06e-2	1.06e-2
St. Dev.	3.46e-3	3.12e-3	3.77e-3

Table 137: ibmq_yorktown Q00 Histograms and Corresponding Means and Standard Deviations for the Population, N1 Sample, and N2 Sample of Each Calibration Characteristic; Data Set Truncation, No Outlier Removal

	PFA		PD	
	Color Scale for PFA and PD Sampling Error Rate 			
	t-test/F-test	KS Test	t-test/F-test	KS Test
	T1			
$\alpha = 0.001$				
$\alpha = 0.05$				
	T2			
$\alpha = 0.001$				

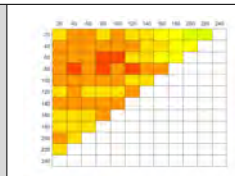
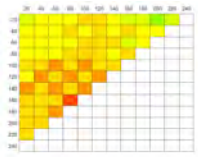
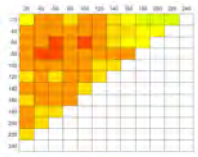
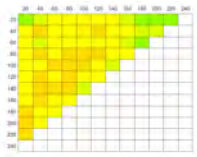
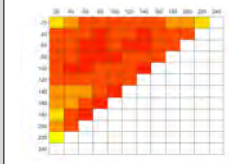
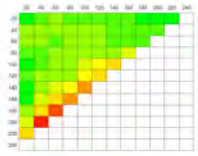
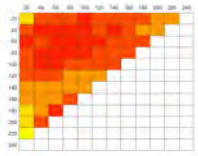
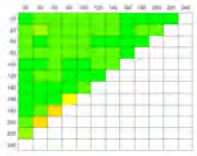
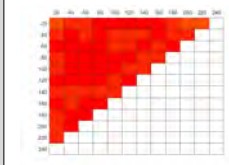
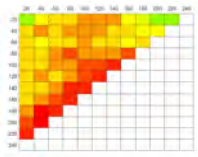
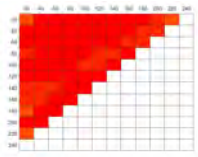
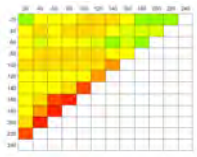
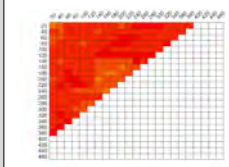
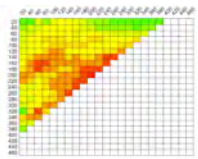
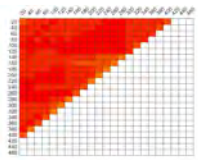
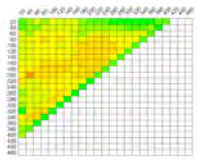
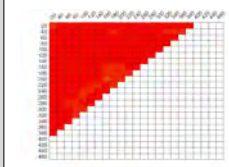
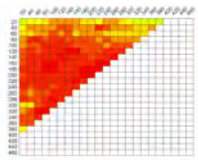
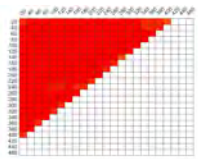
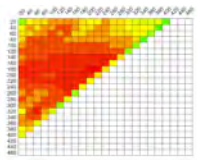
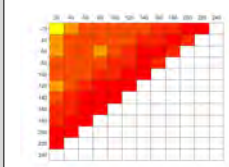
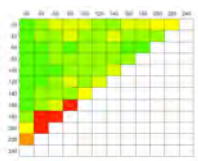
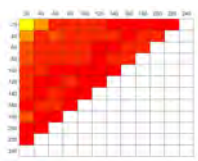
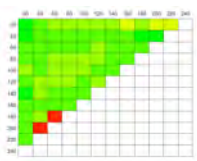
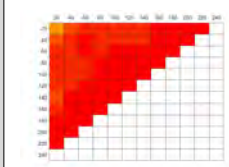
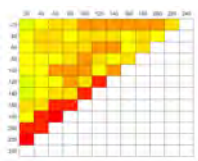
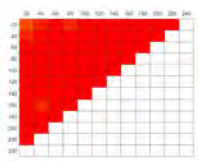
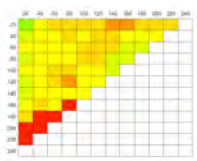
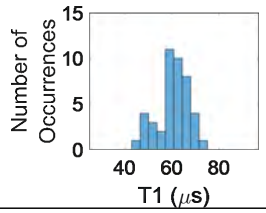
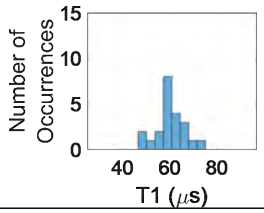
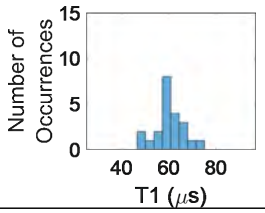
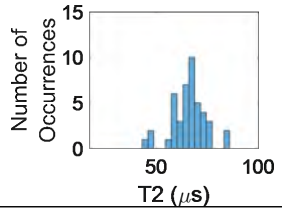
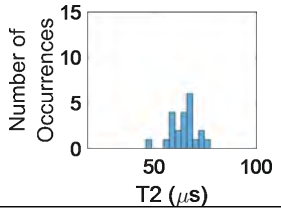
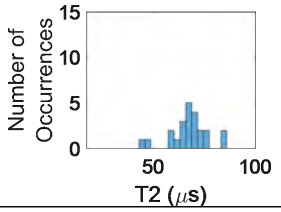
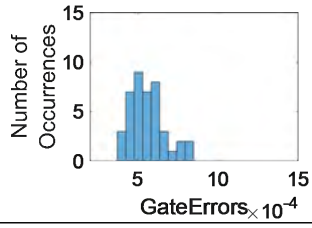
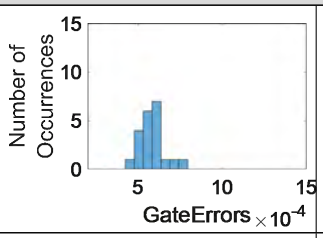
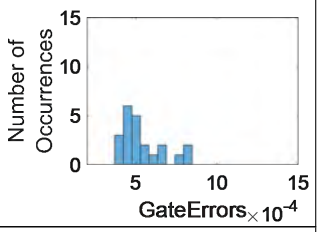
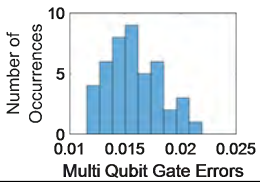
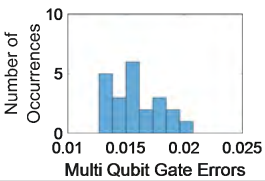
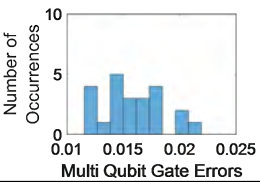
$\alpha = 0.05$				
	Single-Qubit Gate Error Rate			
$\alpha = 0.001$				
$\alpha = 0.05$				
	Multi-Qubit Gate Error Rate			
$\alpha = 0.001$				
$\alpha = 0.05$				
	Readout Error Rate			
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 138: Comparison of t-test/F-test vs. KS Test for Determining if a Sampling Error Exists; Data Set Truncation, No Outlier Removal

	Population	N1 (DP 1-100)	N2 (DP 101-201)
T1s			
			
Mean	60.52	59.90	61.00
St. Dev.	6.60	6.03	7.08
T2s			
			
Mean	65.81	64.14	67.48
St. Dev.	8.15	6.22	9.34
Single-Qubit Gate Error			
			
Mean	5.62e-4	5.85e-4	5.40e-4
St. Dev.	1.06e-4	6.93e-5	1.28e-4
Multi-Qubit Gate Error			
			
Mean	1.58e-2	1.59e-2	1.58e-2
St. Dev.	2.31e-3	1.99e-3	2.60e-3

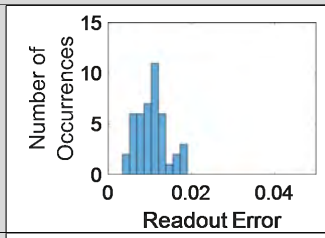
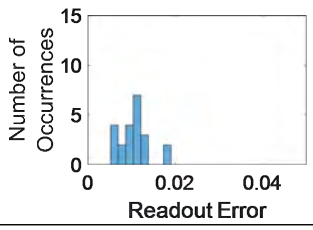
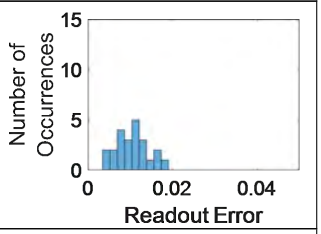
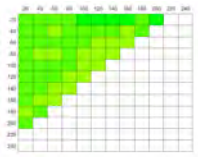
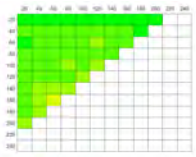
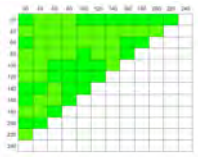
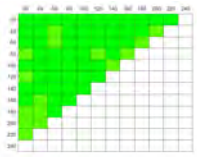
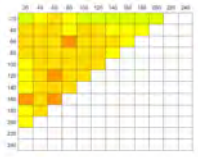
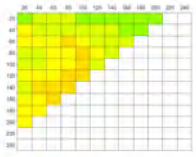
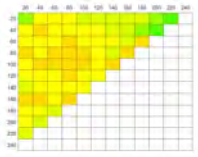
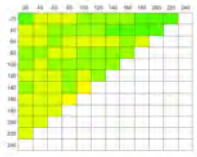
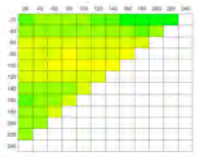
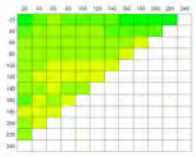
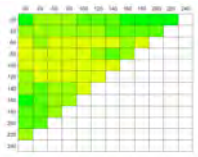
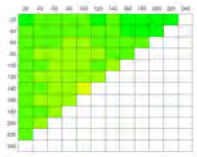
Readout Error			
			
Mean	1.06e-2	1.06e-2	1.06e-2
St. Dev.	3.46e-3	3.12e-3	3.77e-3

Table 139: ibmq_yorktown Q00 Histograms and Corresponding Means and Standard Deviations for the Population, N1 Sample, and N2 Sample of Each Calibration Characteristic; Data Set Truncation with Median Outlier Removal

	PFA		PD	
	Color Scale for PFA and PD Sampling Error Rate			
				
	t-test/F-test	KS Test	t-test/F-test	KS Test
	T1			
$\alpha = 0.001$				
$\alpha = 0.05$				
	T2			
$\alpha = 0.001$				

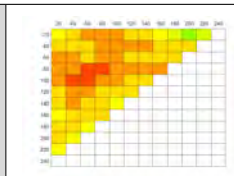
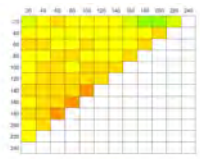
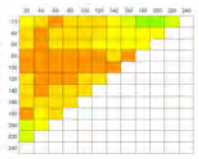
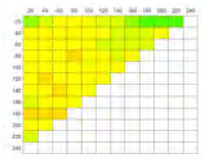
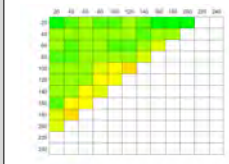
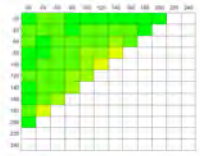
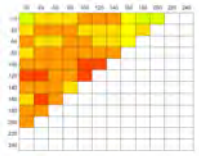
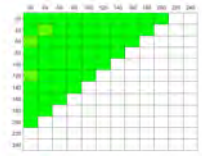
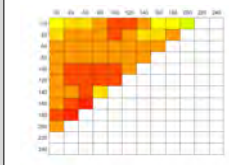
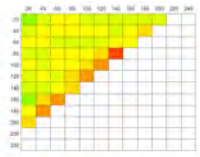
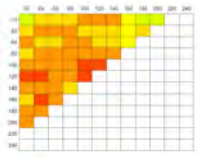
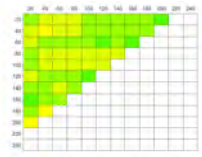
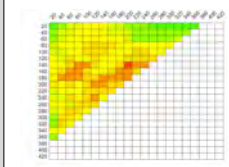
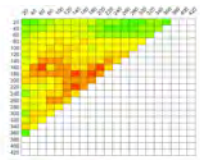
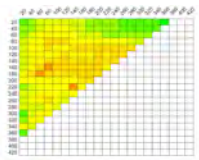
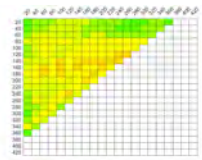
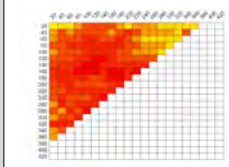
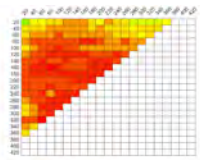
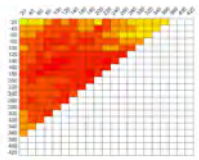
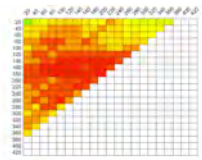
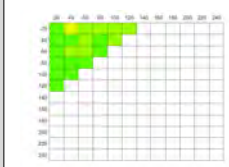
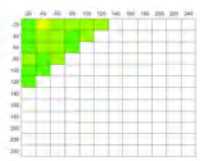
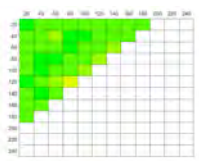
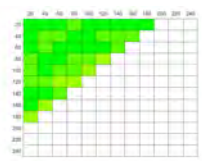
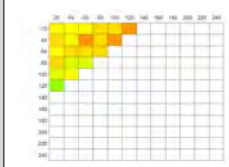
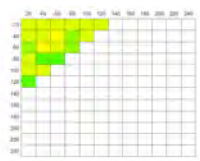
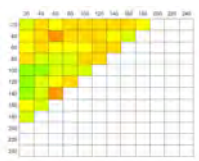
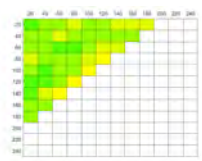
$\alpha = 0.05$				
	Single-Qubit Gate Error Rate			
$\alpha = 0.001$				
$\alpha = 0.05$				
	Multi-Qubit Gate Error Rate			
$\alpha = 0.001$				
$\alpha = 0.05$				
	Readout Error Rate			
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 140: Comparison of t-test/F-test vs. KS Test for Determining if a Sampling Error Exists; Data Set Truncation with Median Outlier Removal

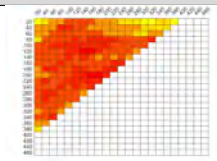
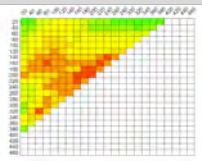
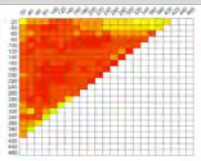
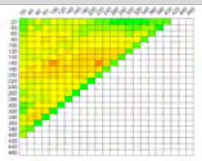
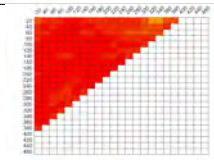
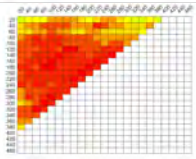
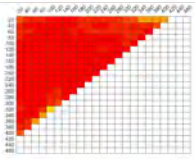
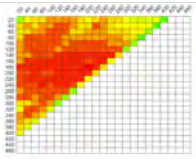
	PFA		KS Test	PD
	Color Scale for PFA and PD Sampling Error Rate			
	0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1			
	t-test/F-test	KS Test	t-test/F-test	KS Test
	Multi-Qubit Gate Error Rate			
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 141: Comparison of t-test/F-test vs. KS Test for Determining if a Sampling Error Exists; Data Set Truncation with Multi-Qubit Gate Error = 1 Removed

4.3.3 Part B3 – Test Qubit Discrimination Methods

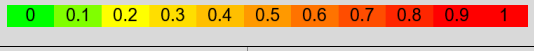
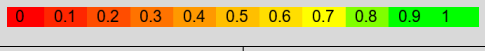
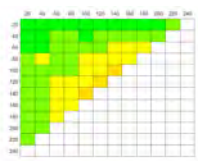
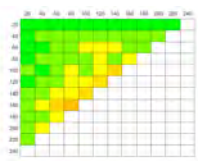
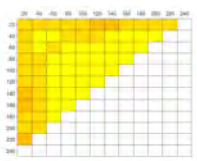
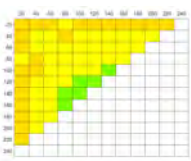
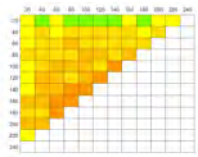
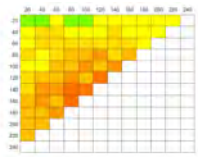
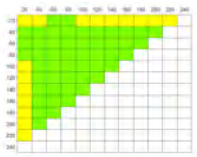
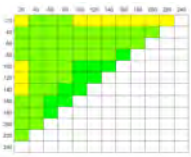
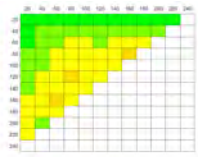
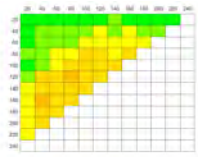
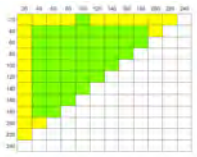
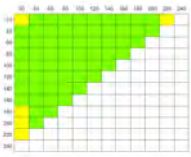
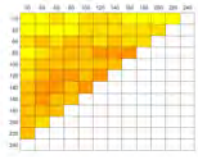
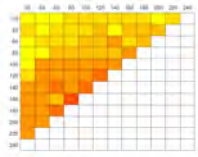
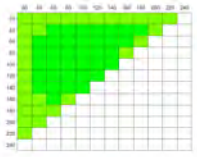
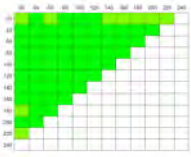
4.3.3.1 Qubit Discrimination Methods

t-Test and Kolmogorov-Smirnov Test

Calibration Characteristics For the calibration characteristics, PD and PFA results for the t-test and KS test are displayed in Table 142 for data set truncation with multi-qubit gate error rates equal to 1 removed from the data set. The tables show that for T1 and T2 coherence times, the KS test results are similar to the t-test results for both PFA and PD. For single-qubit and multi-qubit gate error rates, the KS test has better PD results, but worse PFA results than the t-test. For the readout error rate characteristic, the KS test provides better PD and PFA results than the t-test. None of the characteristics stand out for simultaneously having exceptionally low PFA results and high PD results for significance levels $\alpha = 0.001$ and $\alpha = 0.05$. However, experimentation with different significance levels may improve

PD and PFA results in the following areas:

- T1 between $\alpha = 0.001$ and $\alpha = 0.05$
- T2 between $\alpha = 0.001$ and $\alpha = 0.05$ for low N1 values
- Single-Qubit Gate Error Rate between $\alpha = 0.001$ and $\alpha = 0.05$
- Readout Error Rate between $\alpha = 0.001$ and $\alpha = 0.05$

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	t-test/F-test	KS Test	t-test/F-test	KS Test
	T1			
$\alpha = 0.001$				
$\alpha = 0.05$				
	T2			
$\alpha = 0.001$				
$\alpha = 0.05$				
	Single-Qubit Gate Error Rate			

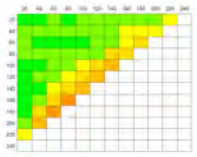
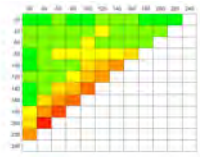
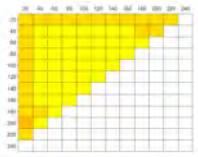
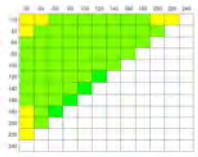
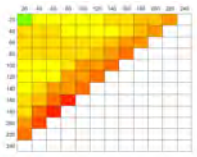
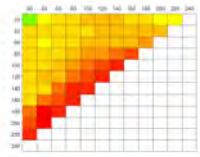
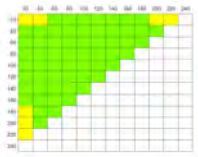
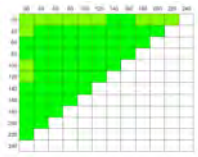
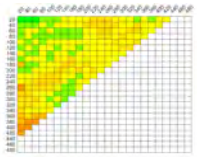
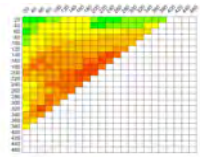
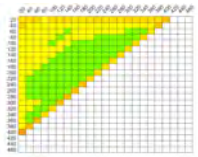
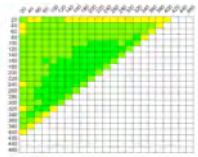
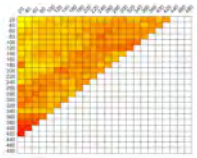
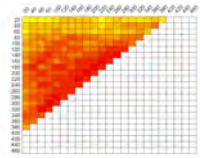
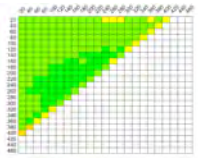
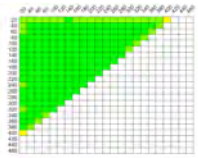
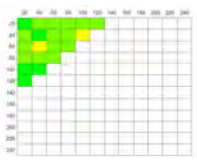
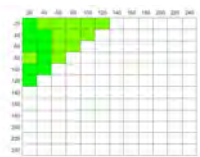
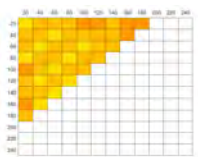
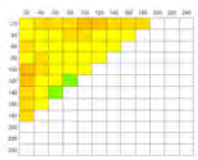
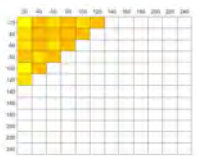
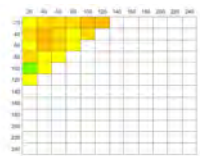
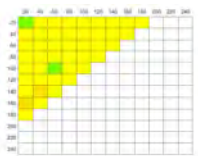
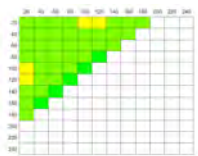
$\alpha = 0.001$				
$\alpha = 0.05$				
	Multi-Qubit Gate Error Rate			
$\alpha = 0.001$				
$\alpha = 0.05$				
	Readout Error Rate			
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 142: PD/PFA Results for t-Test and KS test with Data Set Truncation and Multi-Qubit Gate Error = 1 Removed from the Data Set for Calibration Characteristics

Quantum Process Tomography and Randomized Benchmarking

Characteristics Tables 143 through 149 show the KS test results for the quantum

process tomography and randomized benchmarking metrics. As a note, the graphs show increments of five data points instead of twenty data points because less data has been collected for quantum process tomography and randomized benchmarking compared to the calibration data. Not enough data points were collected for QPTCSWAP to generate results. This reveals one of the limitations of this work: since data is collected through the IBM Quantum Experience, access to the quantum devices is shared with many users. As such, it is not always possible to execute lengthy circuits like the QPTCSWAP circuit on a frequent enough basis to generate sufficient data. Future research for the QPTCSWAP circuit and other circuits of similar complexity should take place on processors not shared with other users.

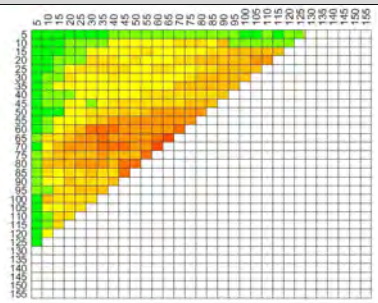
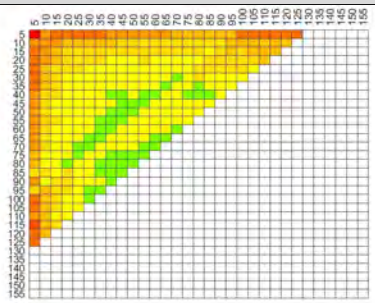
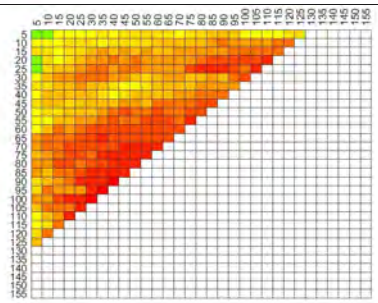
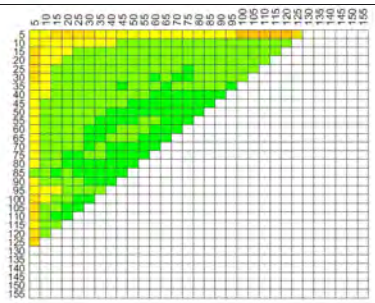
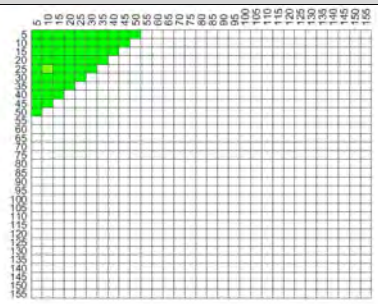
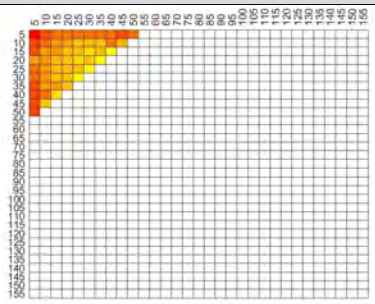
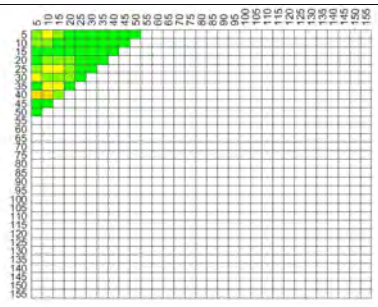
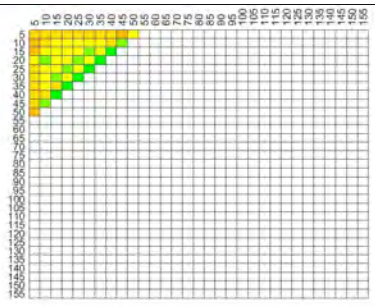
Upon visual inspection, the most promising characteristics that show the possibility for areas with balanced PD and PFA performance that meet minimally reasonable standards (PD above 0.8 and PFA below 0.2) are observed for the following characteristics:

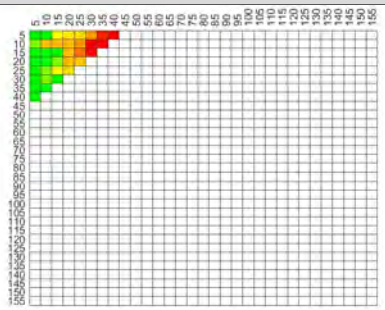
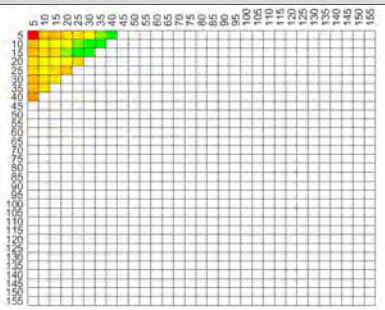
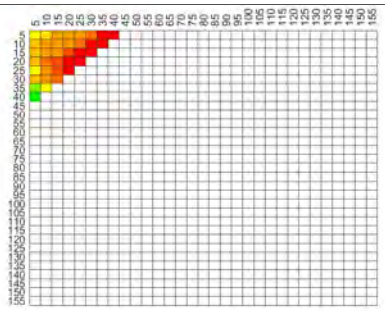
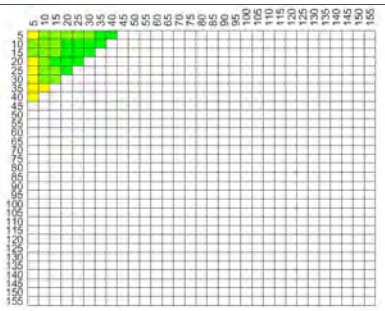
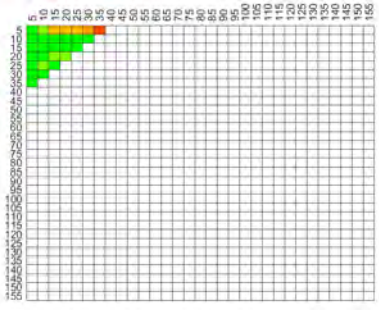
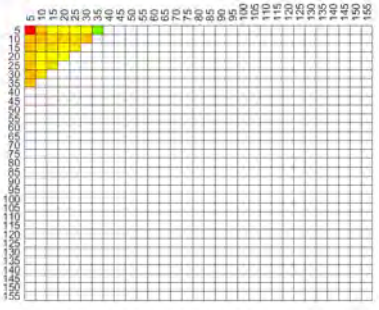
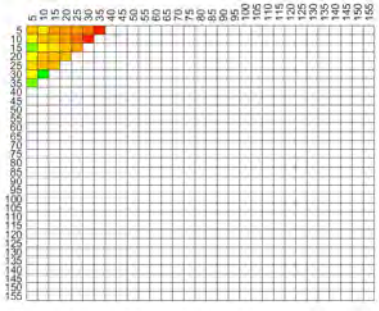
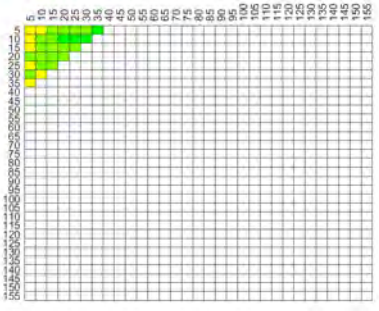
- QPTH slightly above $\alpha = 0.05$
- QPTCX between $\alpha = 0.001$ and $\alpha = 0.05$ for low values of N2
- QPTSWAP between $\alpha = 0.001$ and $\alpha = 0.05$
- QPTCCX between $\alpha = 0.001$ and $\alpha = 0.05$
- RB1 Seed 0 Fidelity between $\alpha = 0.001$ and $\alpha = 0.05$
- RB1 Seed 4 Fidelity between $\alpha = 0.001$ and $\alpha = 0.05$
- RB2 Seed 0 Fidelity between $\alpha = 0.001$ and $\alpha = 0.05$
- RB2 Seed 0 α slightly above $\alpha = 0.05$
- RB2 Seed 0 b slightly above $\alpha = 0.05$

- RB2 Seed 0 err_b slightly above $\alpha = 0.05$
- RB2 Seed 4 a between $\alpha = 0.001$ and $\alpha = 0.05$
- RB2 Seed 4 α slightly above $\alpha = 0.05$
- RB2 Seed 4 b slightly above $\alpha = 0.05$
- RB2 Seed 0 err_α slightly above $\alpha = 0.05$
- RB3 Seed 0 Fidelity slightly above $\alpha = 0.05$
- RB3 Seed 4 Fidelity slightly above $\alpha = 0.05$

In general, quantum process tomography (i.e. fidelity of a gate matrix) and fidelity from randomized benchmarking are reasonable candidates for further experimentation with different significance levels. This is likely because the determination of fidelity incorporates many characteristics that make a qubit (or set of qubits) unique: T1 and T2 coherence times, gate errors, and state preparation and measurement errors. Characteristics from two-qubit randomized benchmarking are also recommended for further experimentation. Two-qubit randomized benchmarking incorporates errors from two-qubit interactions that do not exist in one-qubit randomized benchmarking, but it is less heavily affected by these errors by comparison to three-qubit randomized benchmarking.

Further experimentation with the different significance levels identified above are anticipated to produce reasonably good results. However, for all characteristics tested (calibration, quantum process tomography, and randomized benchmarking characteristics), it appears that achieving excellent PD and PFA performance (PD above 0.9 and PFA below 0.1) might require exploring other qubit discrimination methods. Additionally, exploring other qubit characteristics could lead to improved results.

	PFA	PD
	Color Scale for PFA Tests	Color Scale for PD Tests
		
	QPTX	
$\alpha = 0.001$		
$\alpha = 0.05$		
	QPTH	
$\alpha = 0.001$		
$\alpha = 0.05$		

	QPTCX	
$\alpha = 0.001$		
$\alpha = 0.05$		
	QPTSWAP	
$\alpha = 0.001$		
$\alpha = 0.05$		
	QPTCCX	

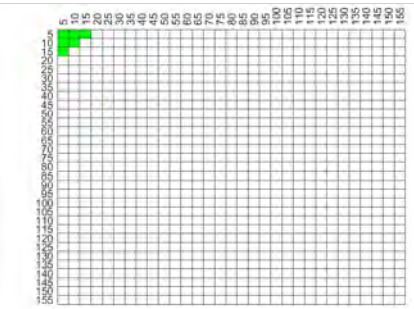
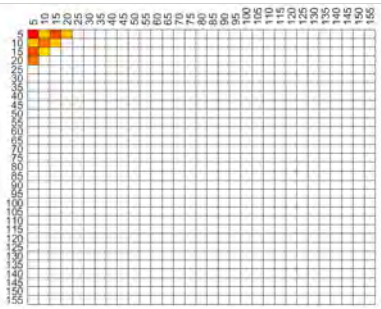
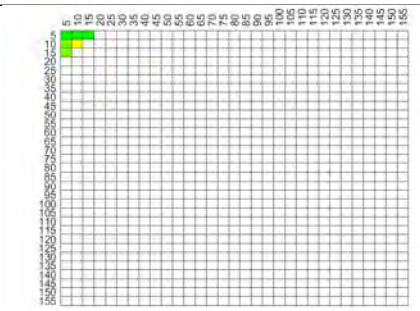
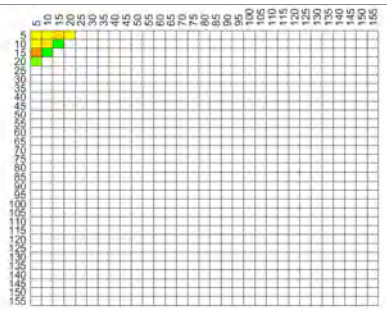
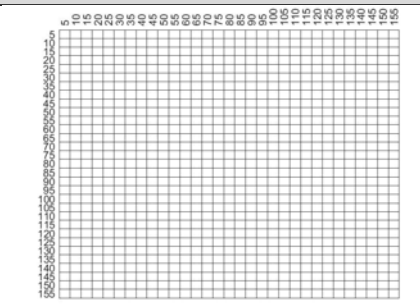
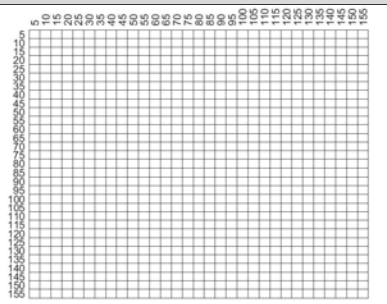
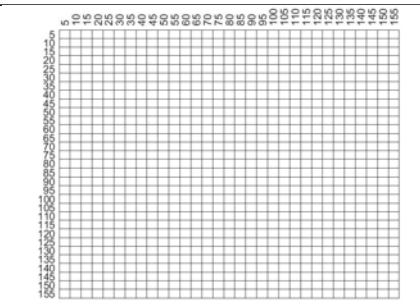
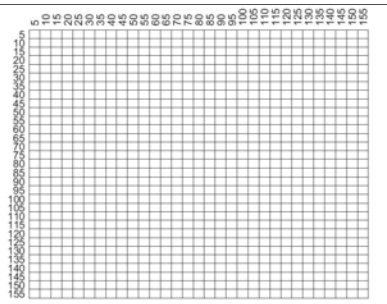
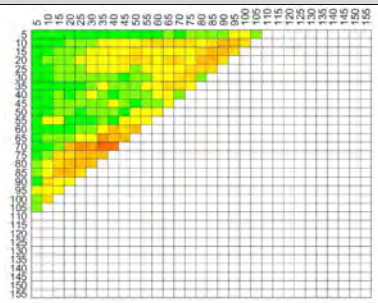
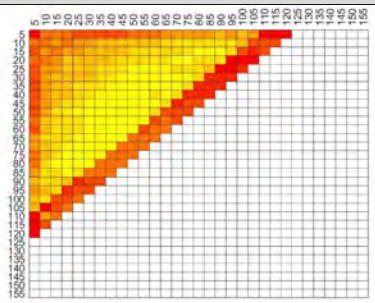
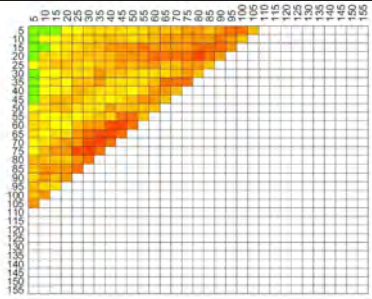
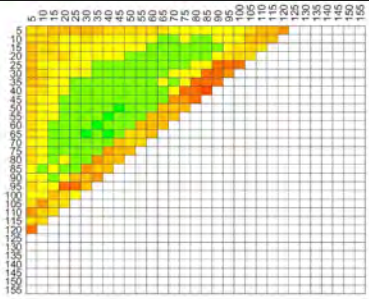
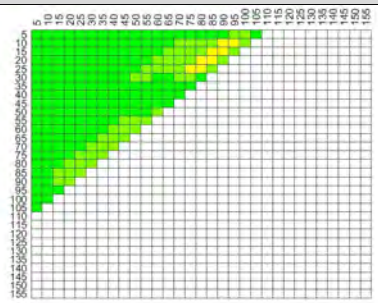
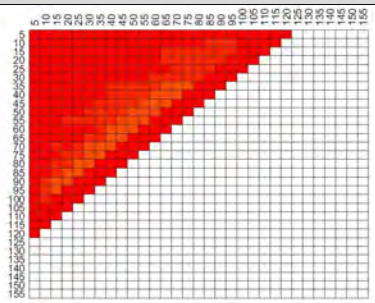
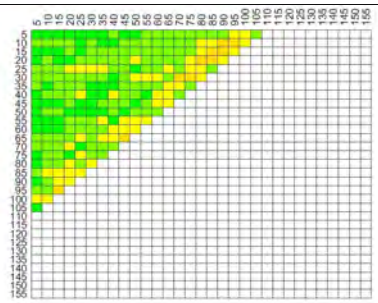
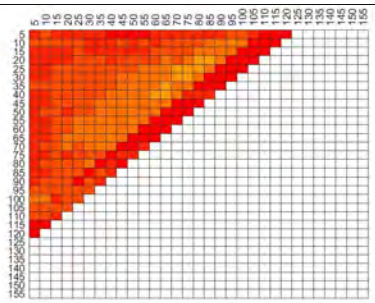
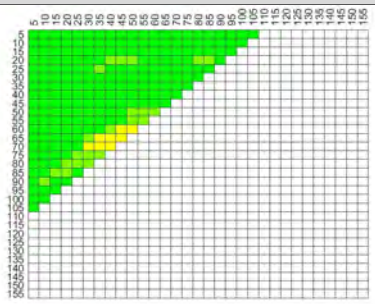
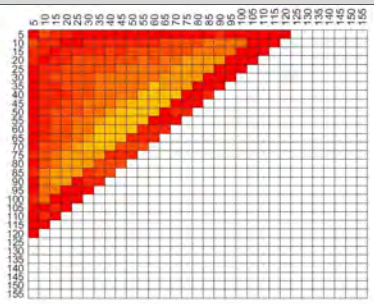
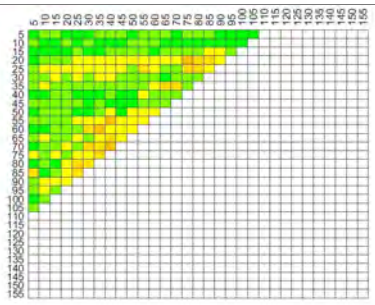
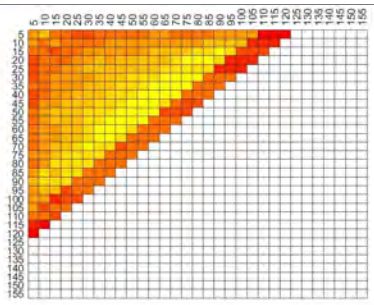
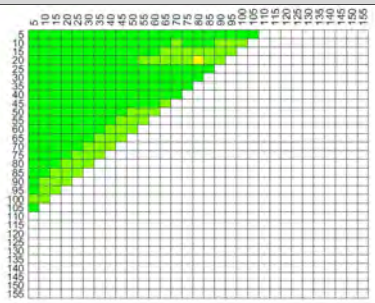
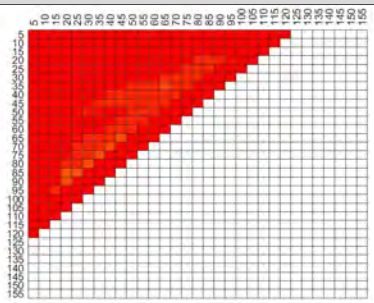
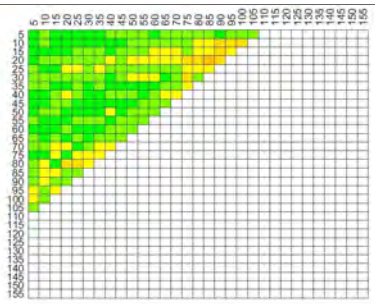
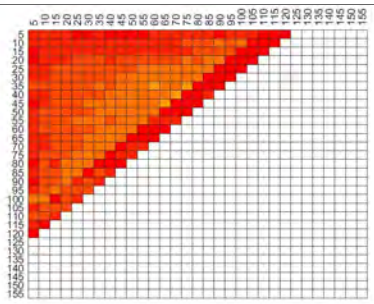
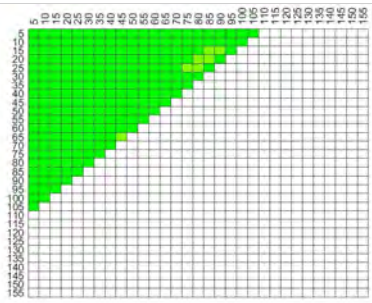
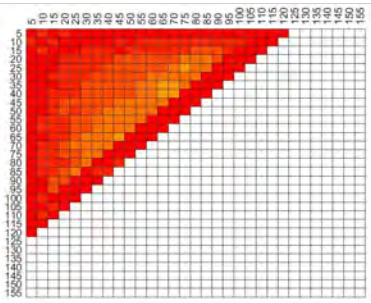
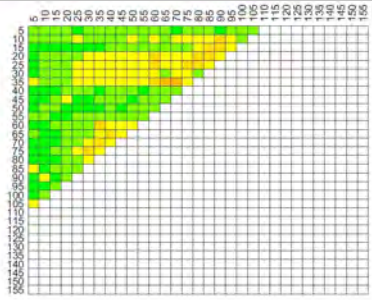
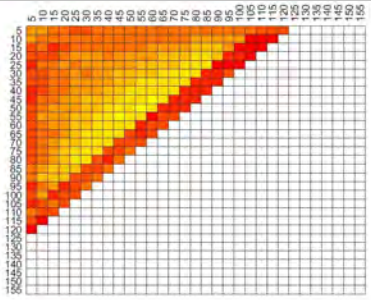
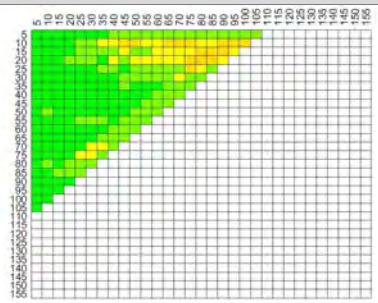
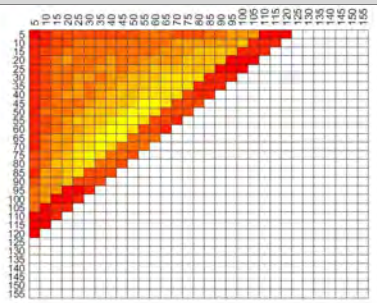
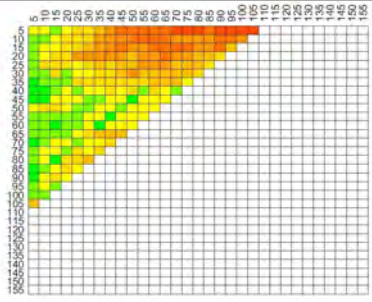
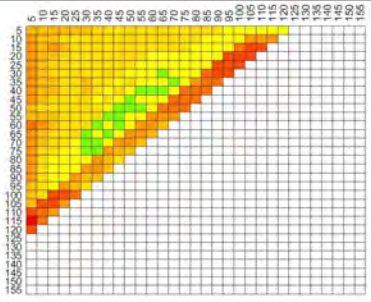
$\alpha = 0.001$		
$\alpha = 0.05$		
	QPTCSWAP	
$\alpha = 0.001$		
$\alpha = 0.05$		

Table 143: PD/PFA Results for KS Test with Data Set Truncation for Quantum Process Tomography Metrics

	PFA	PD
	Color Scale for PFA Tests	Color Scale for PD Tests
		
	RB1 – Seed 0 – Fidelity	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB1 – Seed 0 – a	
$\alpha = 0.001$		
$\alpha = 0.05$		

	RB1 – Seed 0 – α	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB1 – Seed 0 – b	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB1 – Seed 0 – err_a	

$\alpha = 0.001$		
$\alpha = 0.05$		
	RB1 – Seed 0 – err_α	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB1 – Seed 0 – err_b	

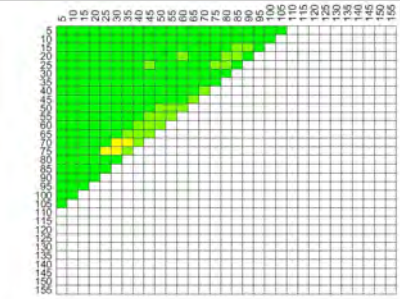
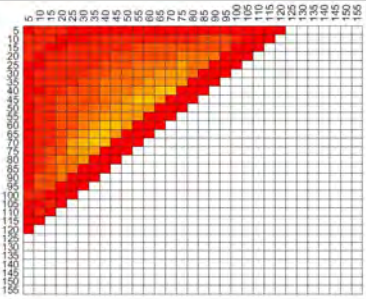
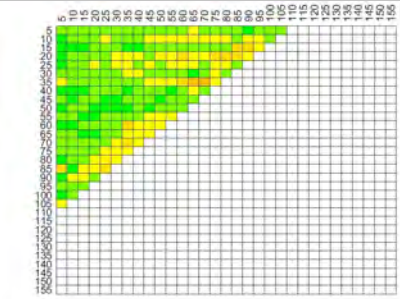
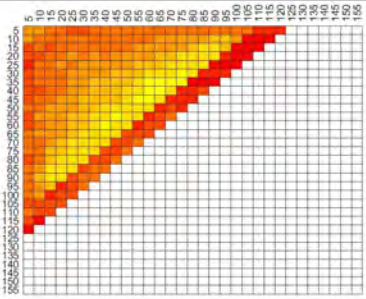
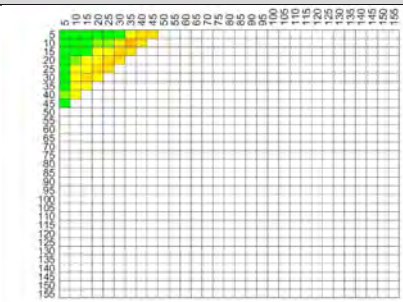
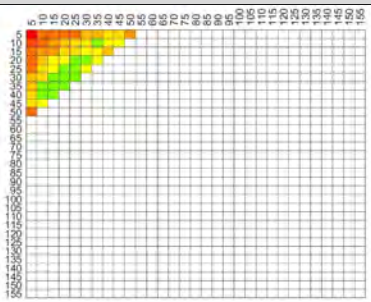
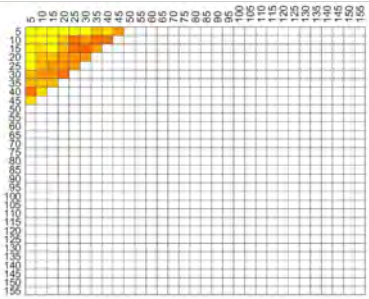
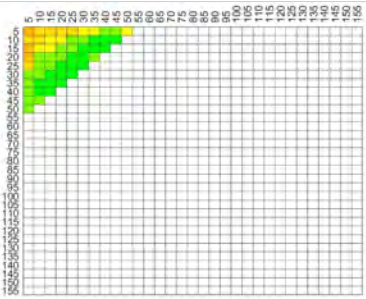
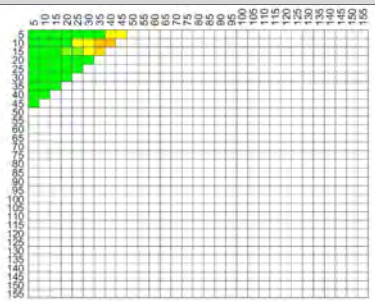
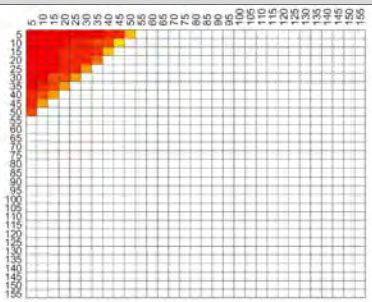
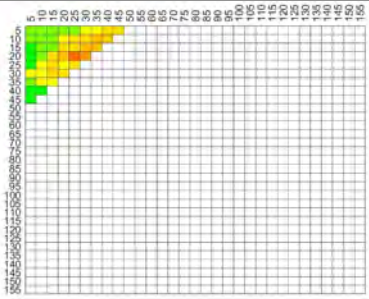
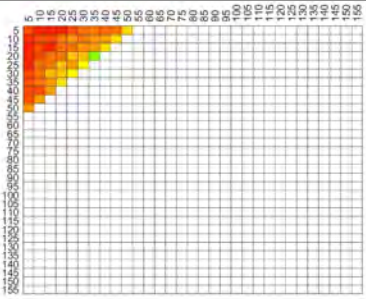
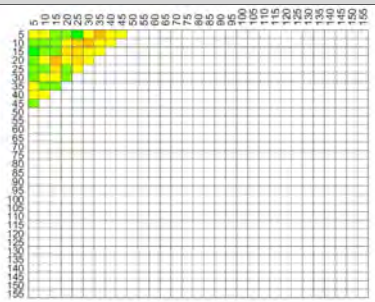
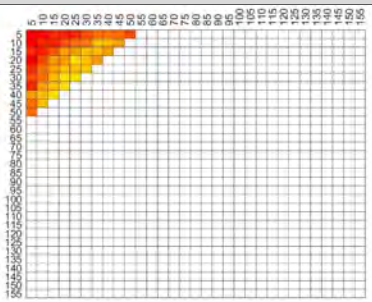
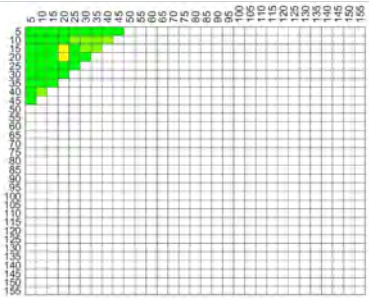
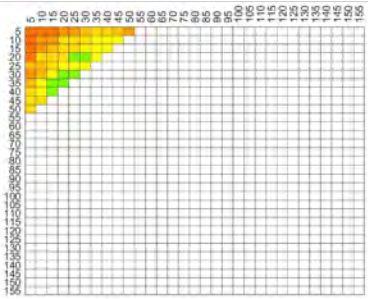
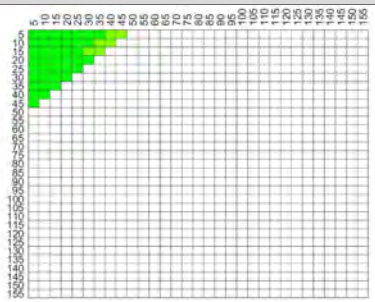
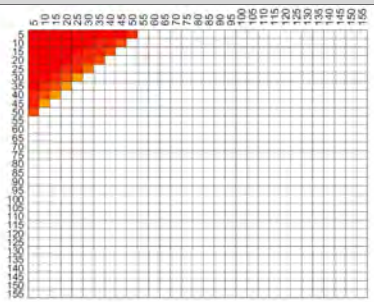
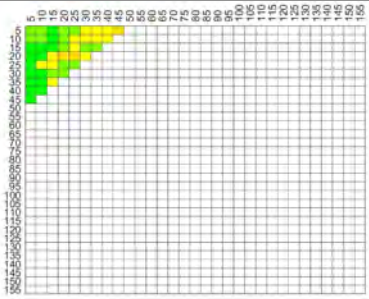
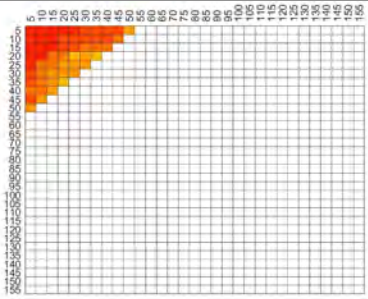
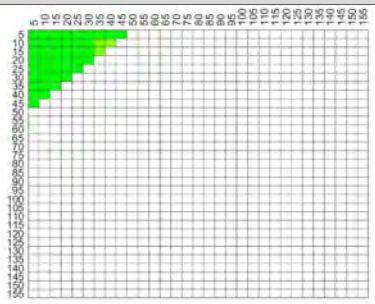
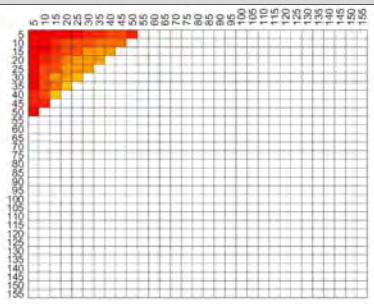
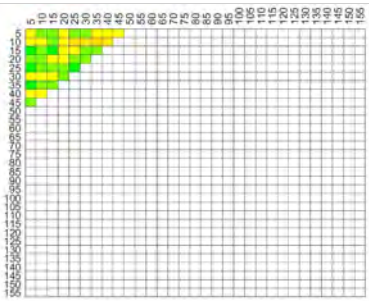
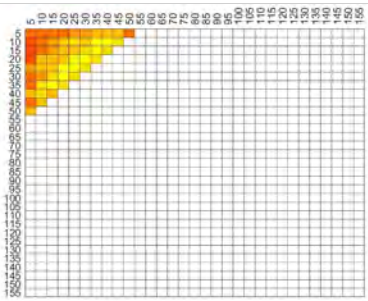
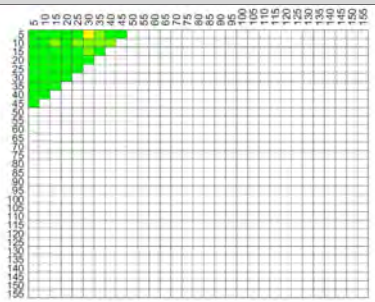
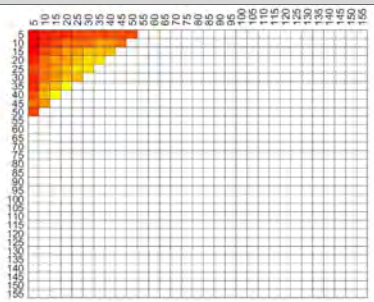
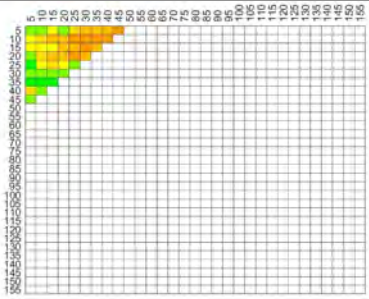
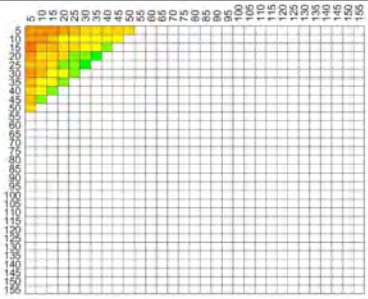
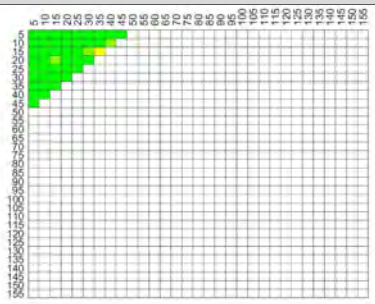
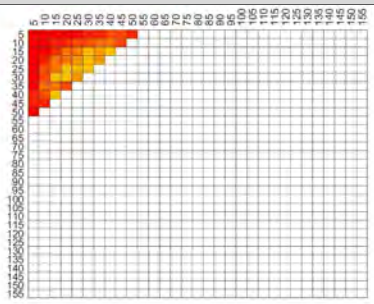
$\alpha = 0.001$		
$\alpha = 0.05$		

Table 144: PD/PFA Results for KS Test with Data Set Truncation for 1-Qubit Randomized Benchmarking Metrics, Seed 0

	PFA	PD
	Color Scale for PFA Tests	Color Scale for PD Tests
		
	RB1 – Seed 4 – Fidelity	
$\alpha = 0.001$		

$\alpha = 0.05$		
	RB1 – Seed 4 – a	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB1 – Seed 4 – α	
$\alpha = 0.001$		

$\alpha = 0.05$		
	RB1 – Seed 4 – b	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB1 – Seed 4 – *err_a*	
$\alpha = 0.001$		

$\alpha = 0.05$		
	RB1 – Seed 4 – err_α	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB1 – Seed 4 – err_b	
$\alpha = 0.001$		

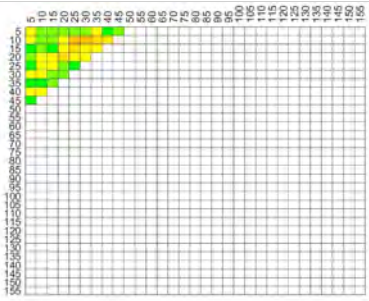
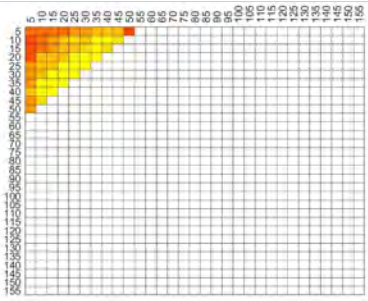
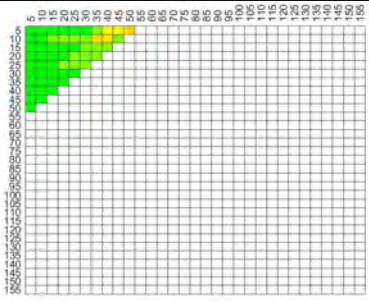
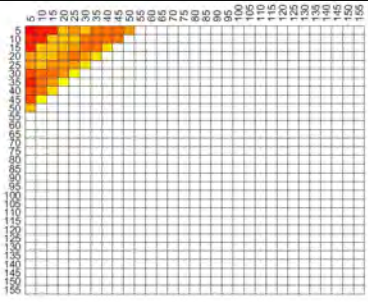
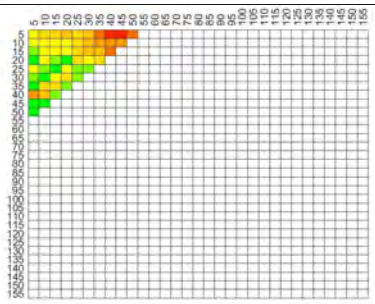
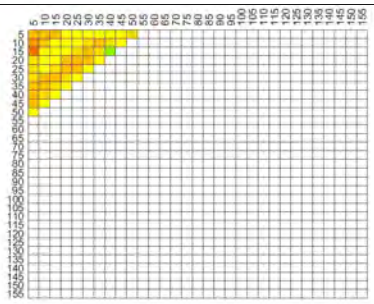
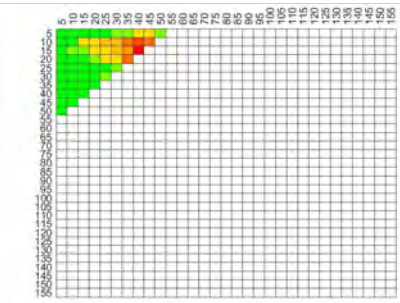
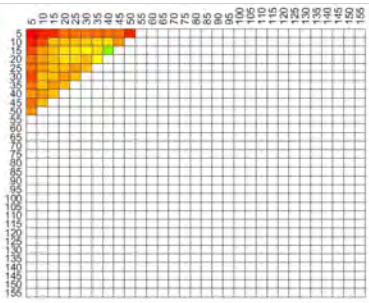
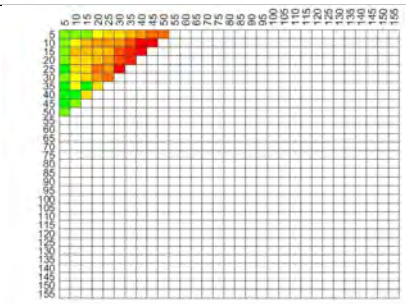
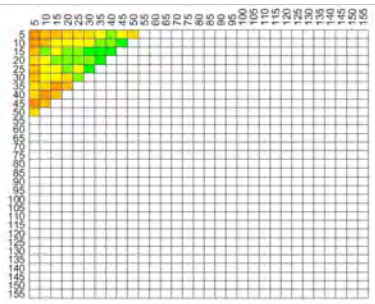
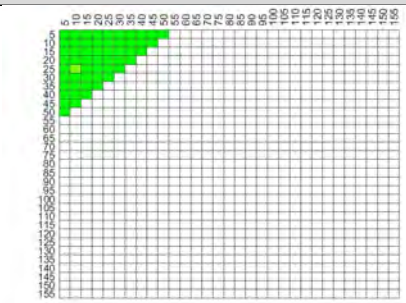
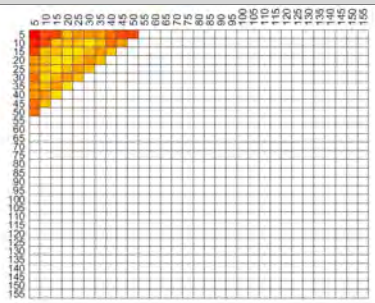
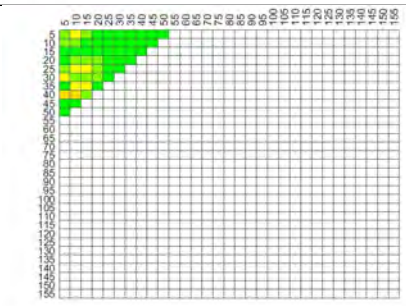
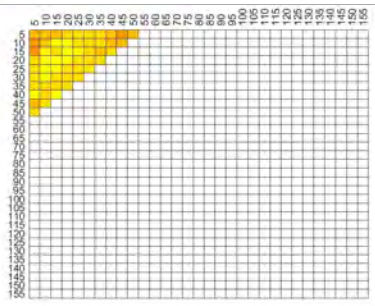
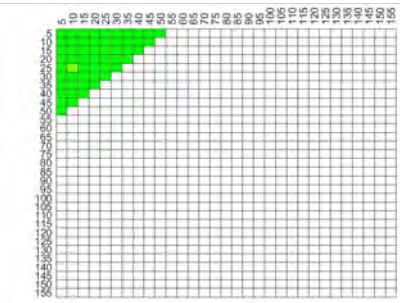
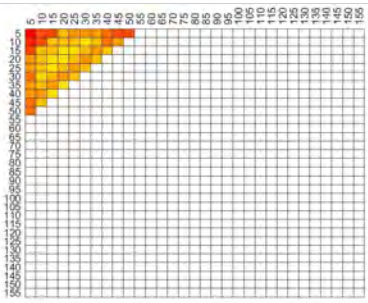
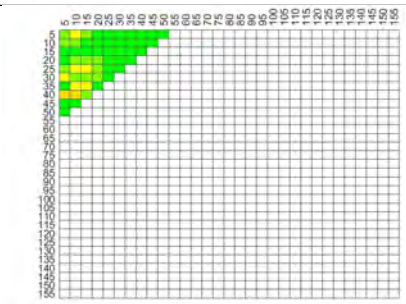
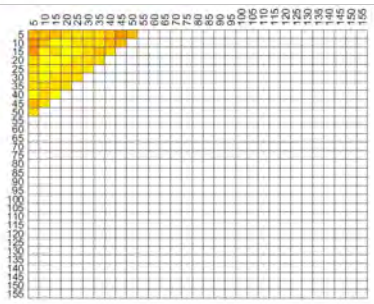
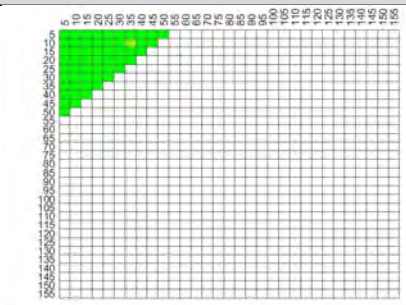
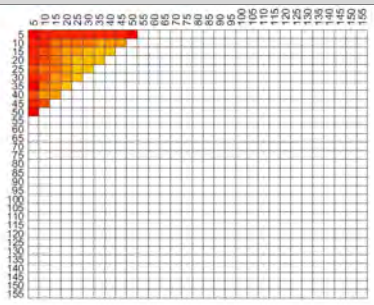
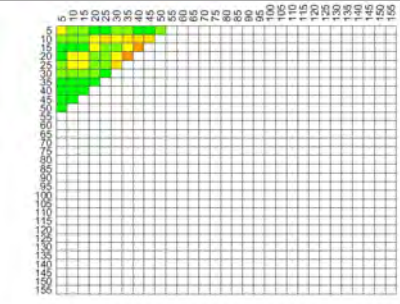
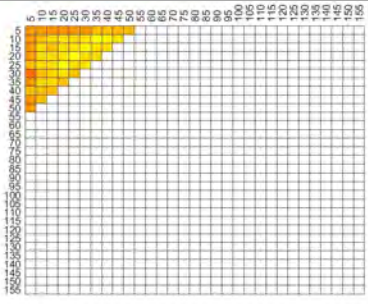
$\alpha = 0.05$		

Table 145: PD/PFA Results for KS Test with Data Set Truncation for 1-Qubit Randomized Benchmarking Metrics, Seed 4

	PFA	PD
	Color Scale for PFA Tests	Color Scale for PD Tests
		
	RB2 – Seed 0 – Fidelity	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB2 – Seed 0 – a	

$\alpha = 0.001$		
$\alpha = 0.05$		
	RB2 – Seed 0 – α	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB2 – Seed 0 – b	

$\alpha = 0.001$		
$\alpha = 0.05$		
	RB2 – Seed 0 – err_a	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB2 – Seed 0 – err_α	

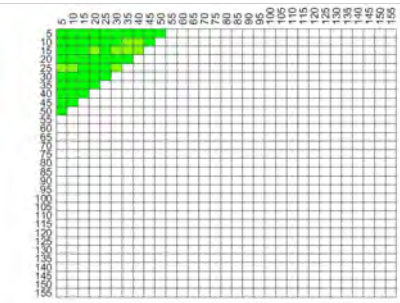
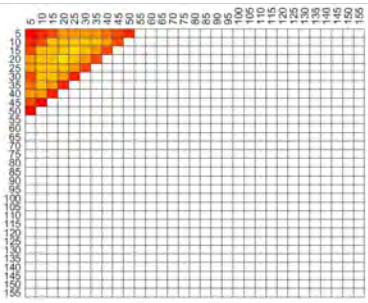
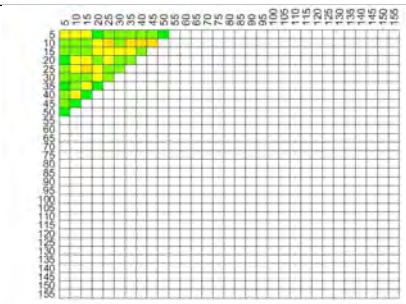
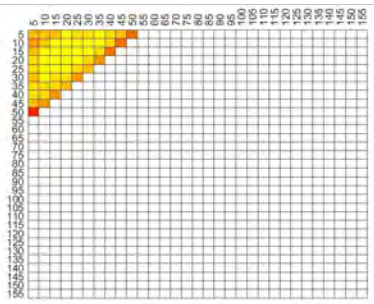
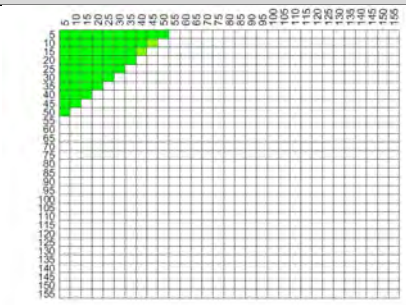
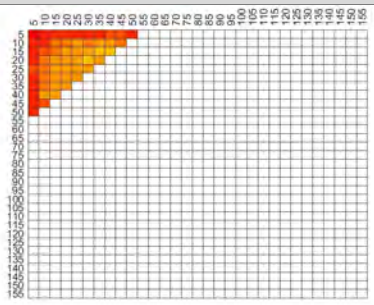
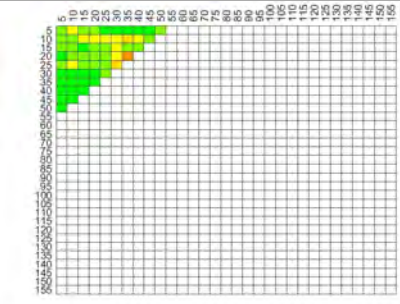
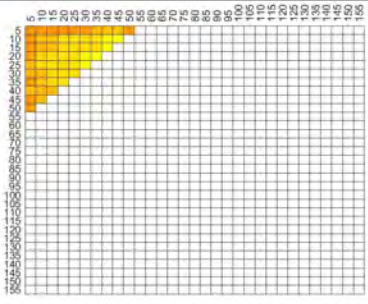
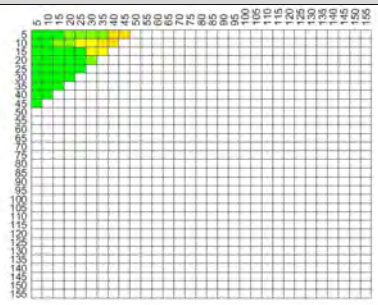
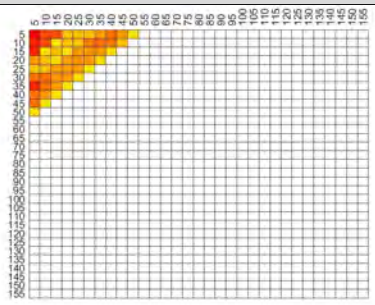
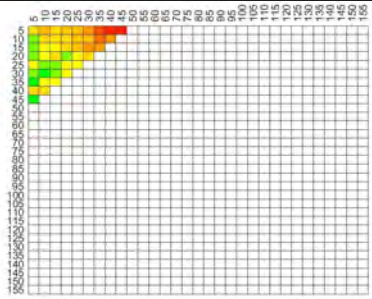
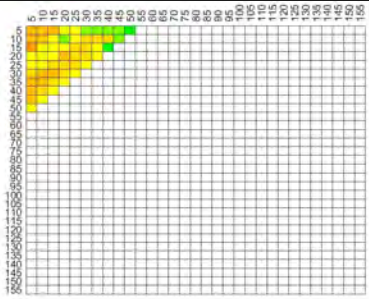
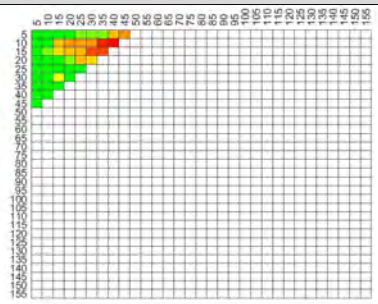
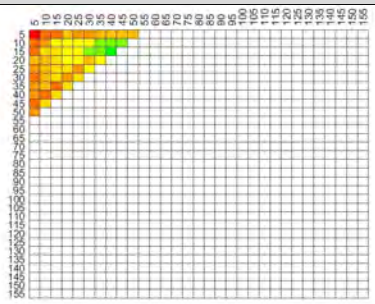
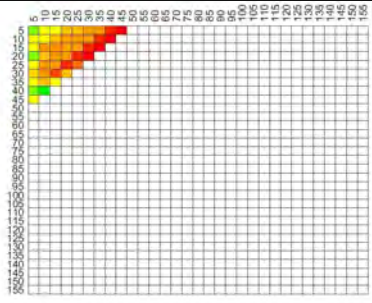
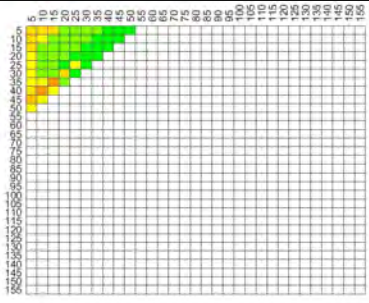
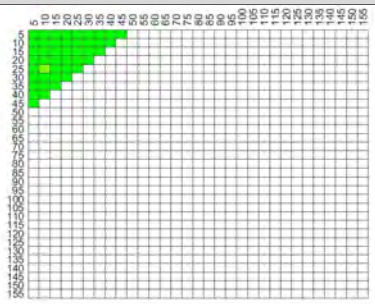
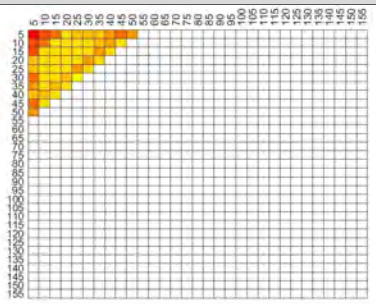
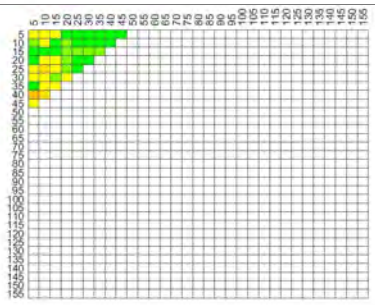
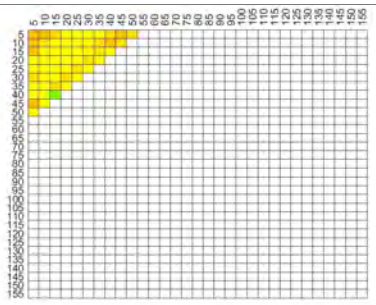
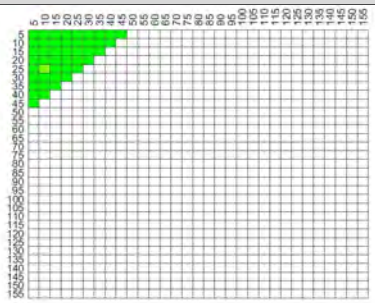
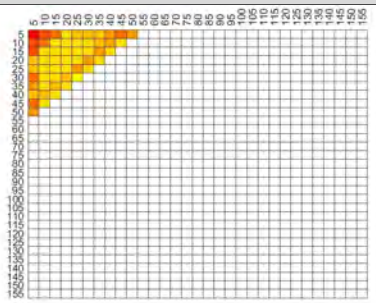
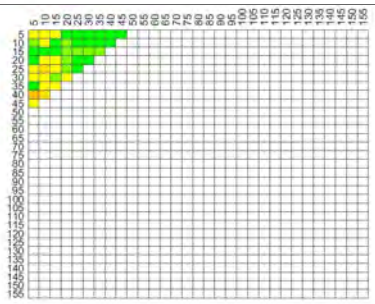
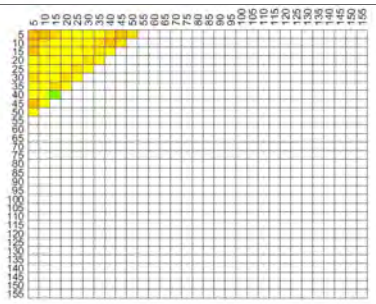
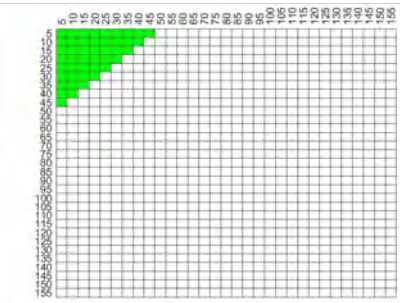
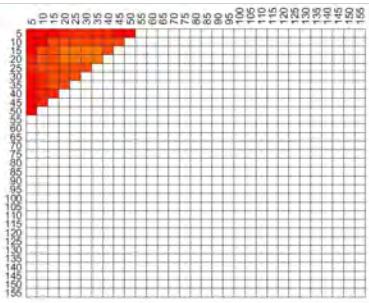
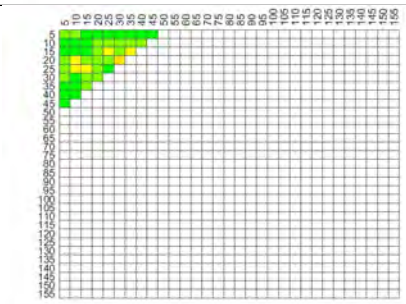
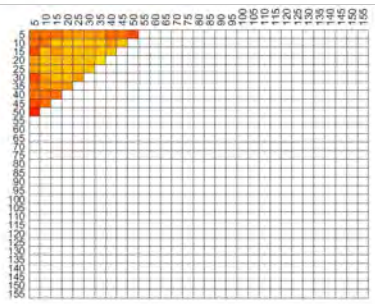
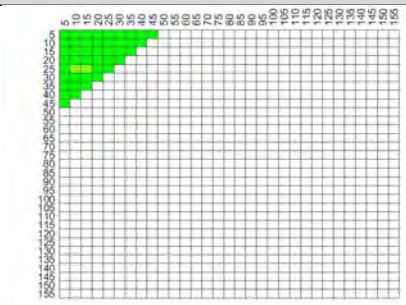
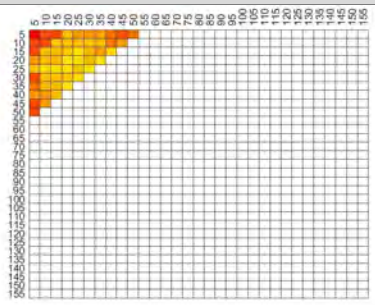
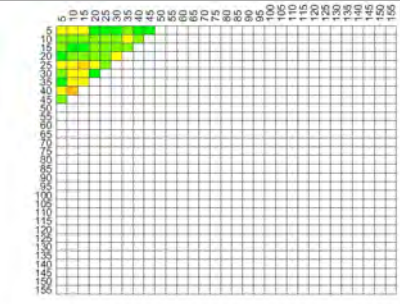
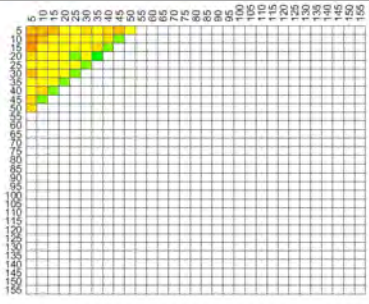
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB2 – Seed 0 – *err_b*	
$\alpha = 0.001$		
$\alpha = 0.05$		

Table 146: PD/PFA Results for KS Test with Data Set Truncation for 2-Qubit Randomized Benchmarking Metrics, Seed 0

	PFA	PD
	Color Scale for PFA Tests	Color Scale for PD Tests
		
	RB2 – Seed 4 – Fidelity	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB2 – Seed 4 – a	
$\alpha = 0.001$		
$\alpha = 0.05$		

	RB2 – Seed 4 – α	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB2 – Seed 4 – b	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB2 – Seed 4 – *err_a*	

$\alpha = 0.001$		
$\alpha = 0.05$		
	RB2 – Seed 4 – err_α	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB2 – Seed 4 – err_b	

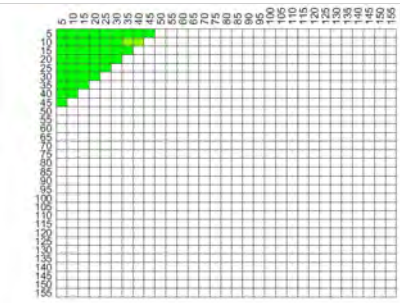
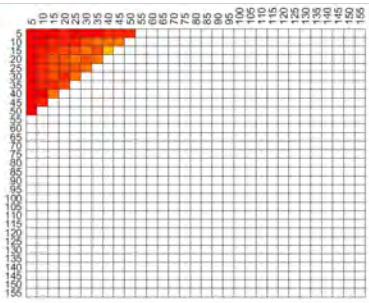
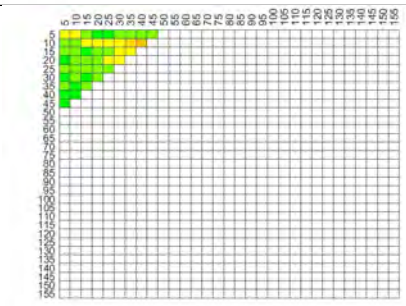
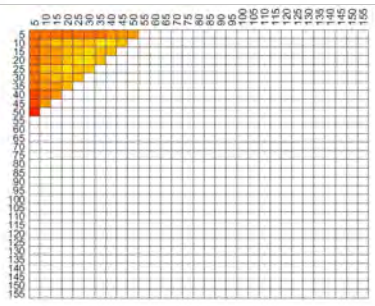
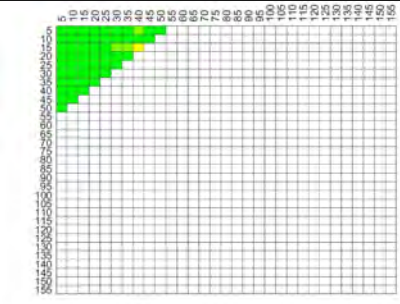
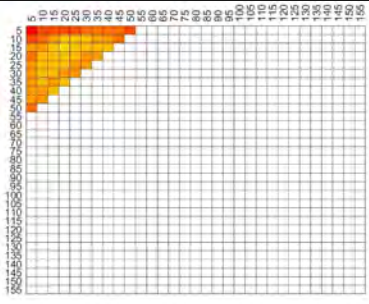
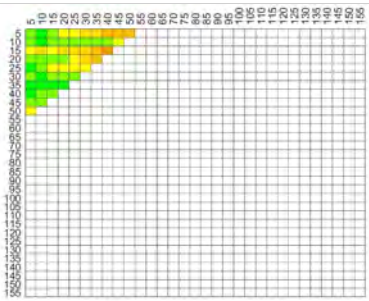
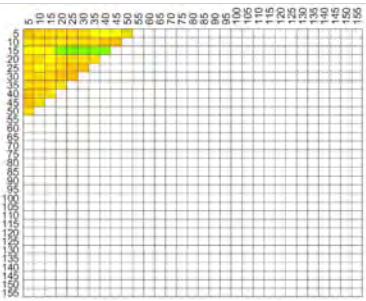
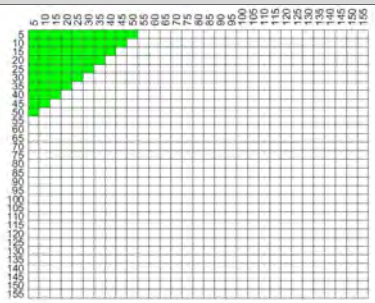
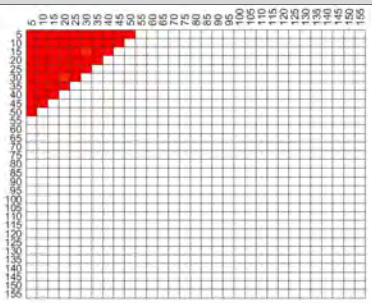
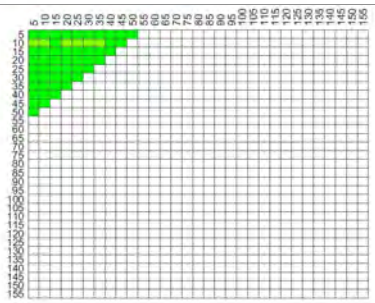
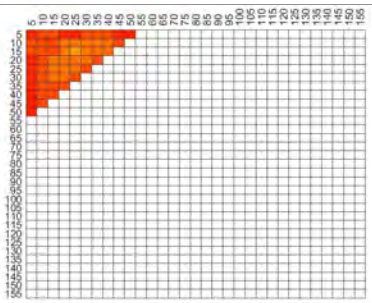
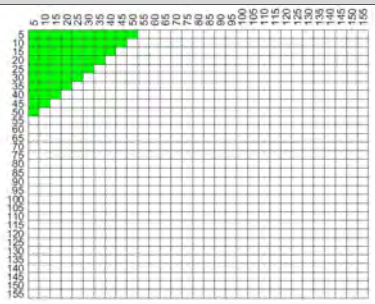
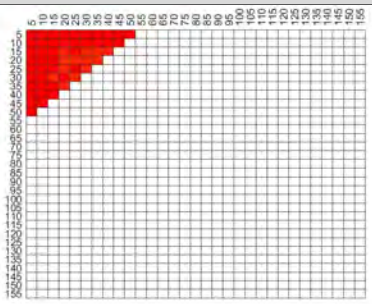
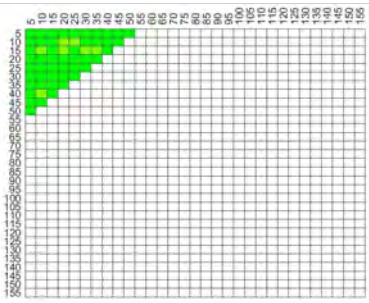
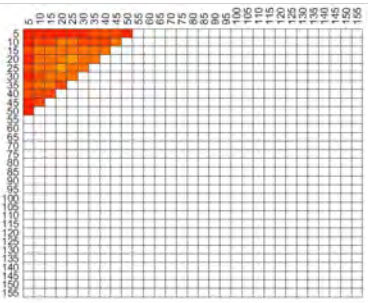
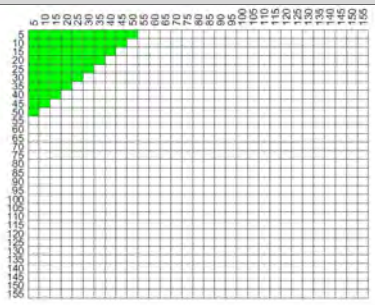
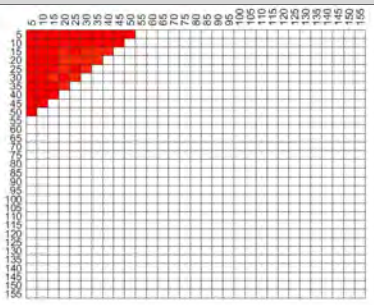
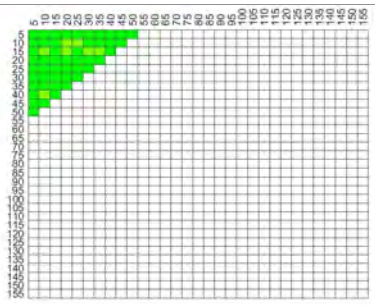
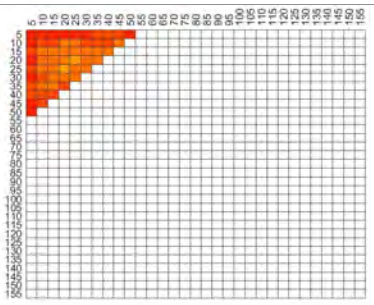
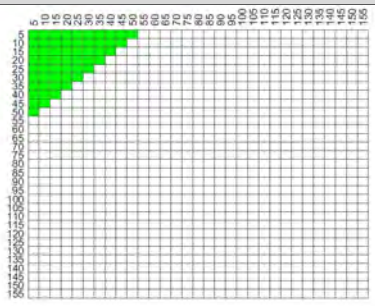
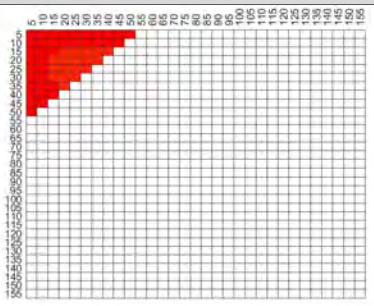
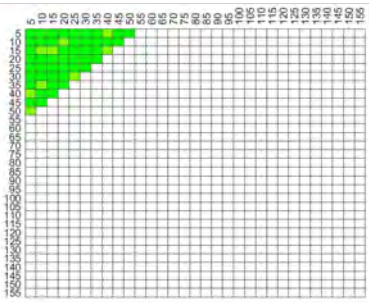
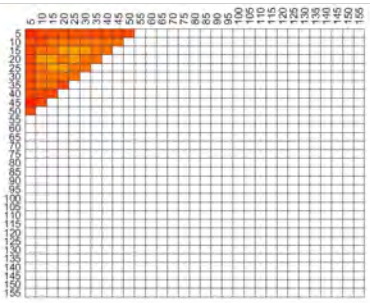
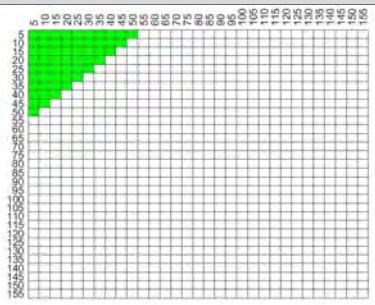
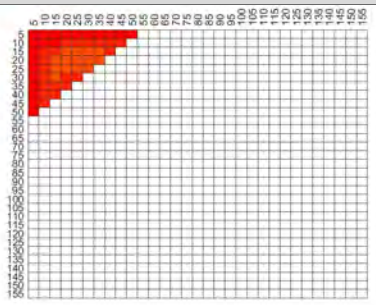
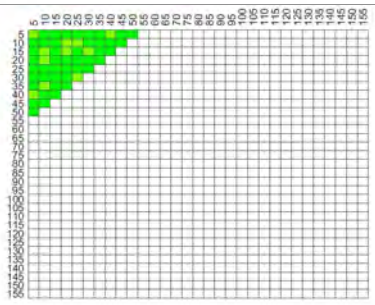
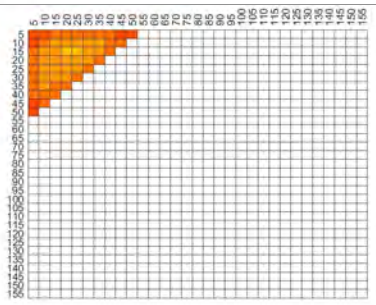
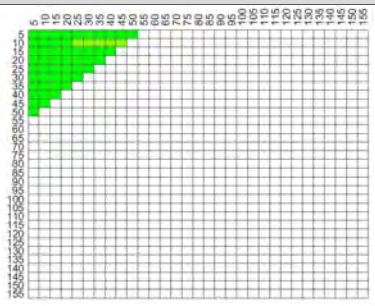
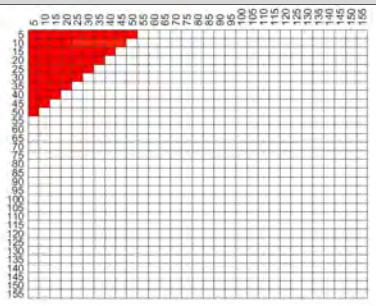
$\alpha = 0.001$		
$\alpha = 0.05$		

Table 147: PD/PFA Results for KS Test with Data Set Truncation for 2-Qubit Randomized Benchmarking Metrics, Seed 4

	PFA	PD
	Color Scale for PFA Tests	Color Scale for PD Tests
		
	RB3 – Seed 0 – Fidelity	
$\alpha = 0.001$		

$\alpha = 0.05$		
	RB3 – Seed 0 – a	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB3 – Seed 0 – α	
$\alpha = 0.001$		

$\alpha = 0.05$		
	RB3 – Seed 0 – b	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB3 – Seed 0 – *err_a*	
$\alpha = 0.001$		

$\alpha = 0.05$		
	RB3 – Seed 0 – err_α	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB3 – Seed 0 – err_b	
$\alpha = 0.001$		

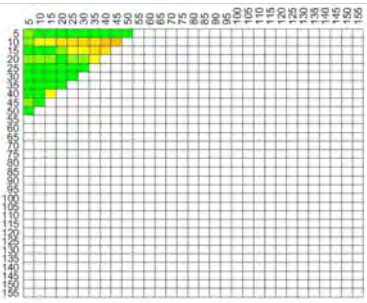
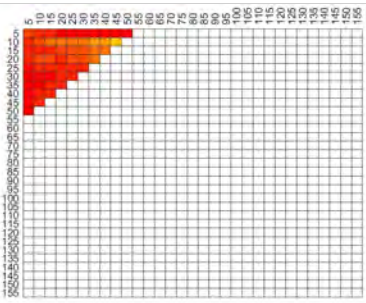
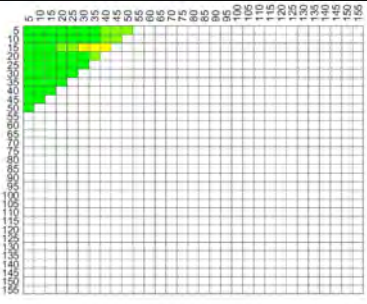
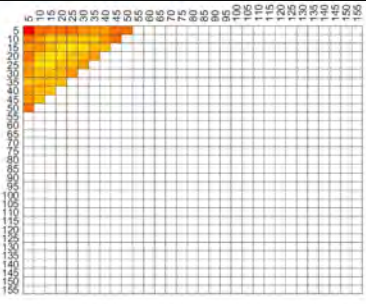
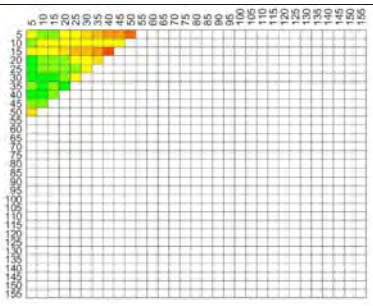
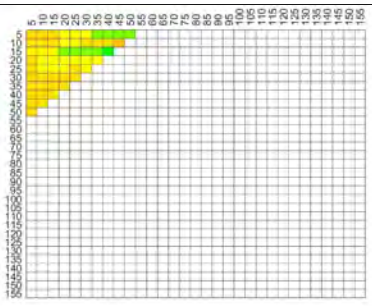
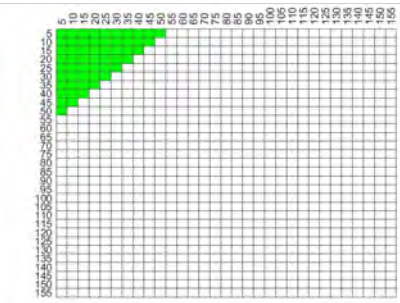
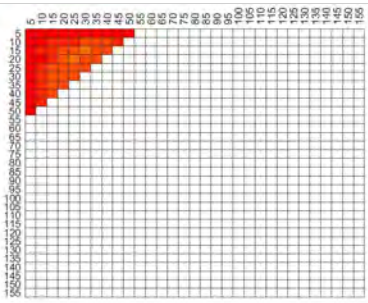
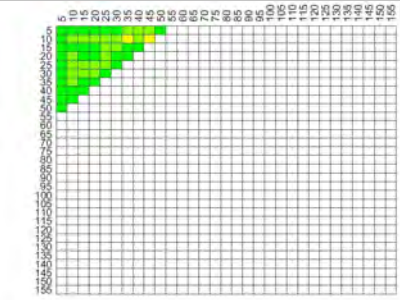
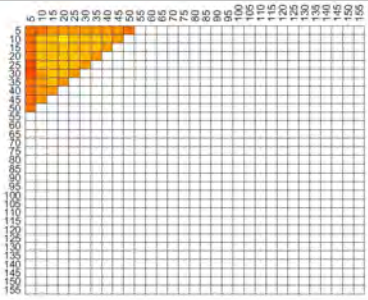
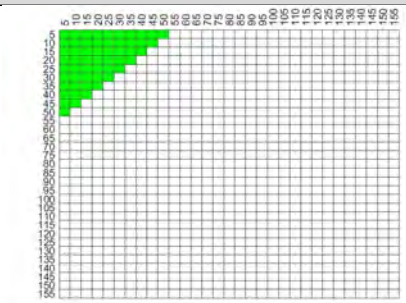
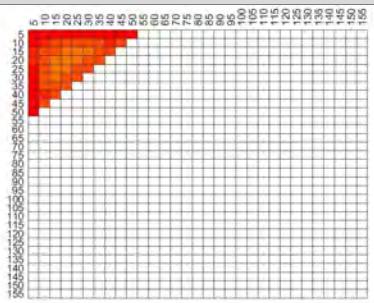
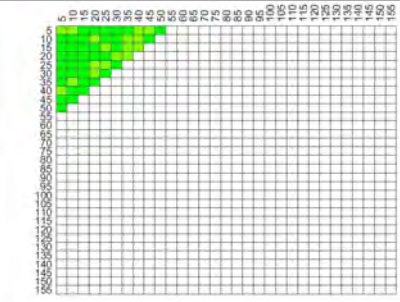
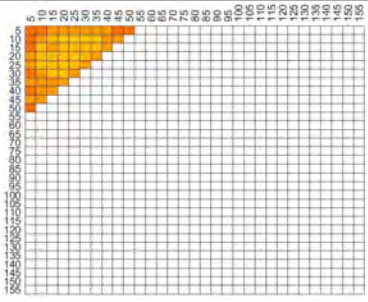
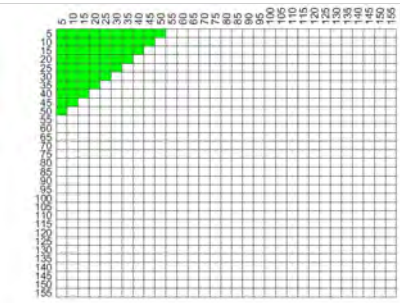
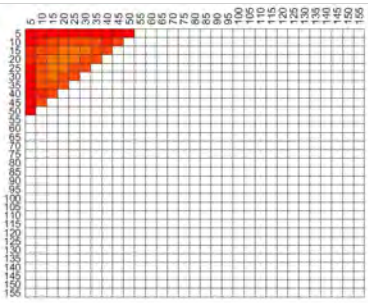
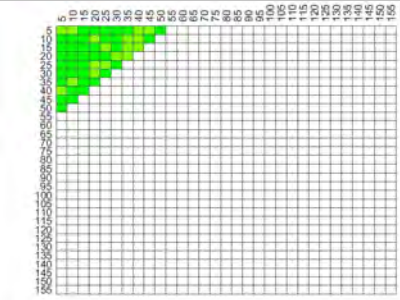
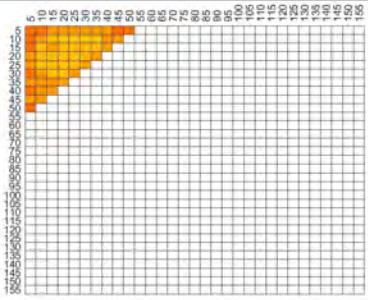
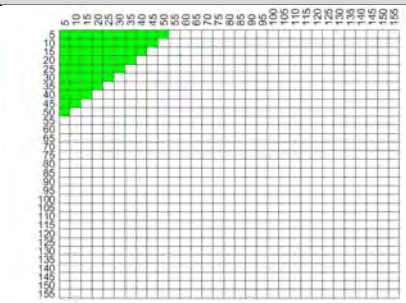
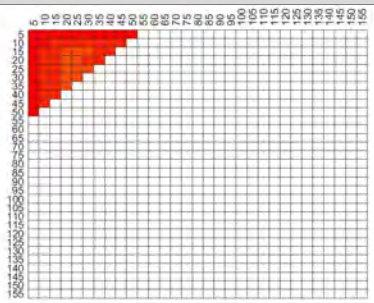
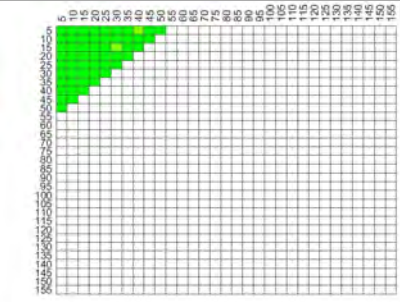
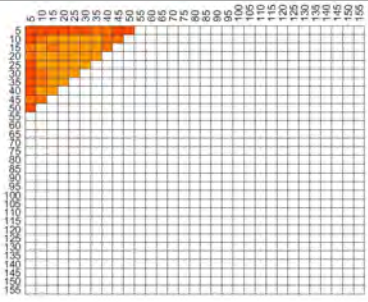
$\alpha = 0.05$		

Table 148: PD/PFA Results for KS Test with Data Set Truncation for 3-Qubit Randomized Benchmarking Metrics, Seed 0

	PFA	PD
	Color Scale for PFA Tests	Color Scale for PD Tests
		
	RB3 – Seed 4 – Fidelity	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB3 – Seed 4 – a	

$\alpha = 0.001$		
$\alpha = 0.05$		
	RB3 – Seed 4 – α	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB3 – Seed 4 – b	

$\alpha = 0.001$		
$\alpha = 0.05$		
	RB3 – Seed 4 – err_a	
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB3 – Seed 4 – err_α	

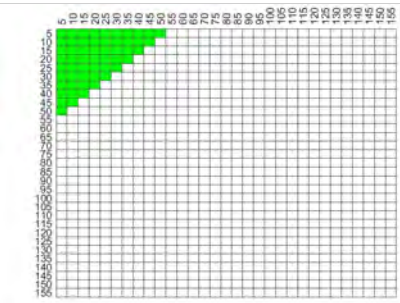
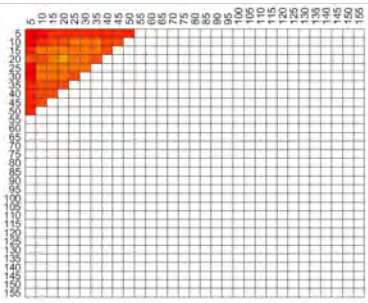
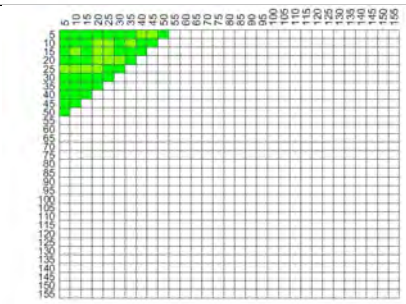
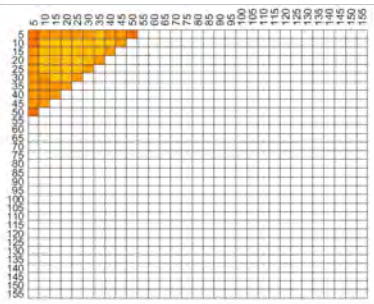
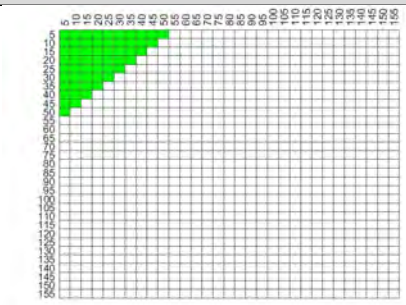
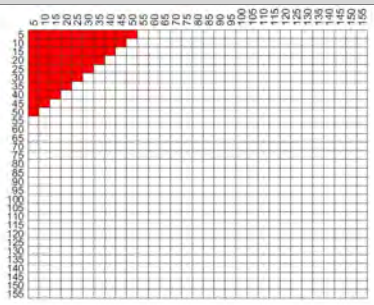
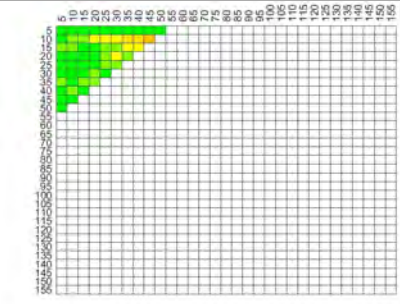
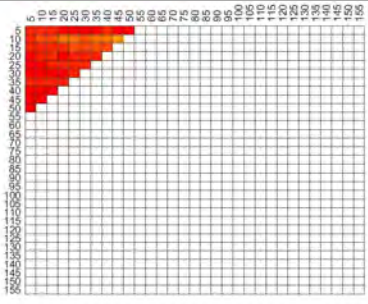
$\alpha = 0.001$		
$\alpha = 0.05$		
	RB3 – Seed 4 – *err_b*	
$\alpha = 0.001$		
$\alpha = 0.05$		

Table 149: PD/PFA Results for KS Test with Data Set Truncation for 3-Qubit Randomized Benchmarking Metrics, Seed 4

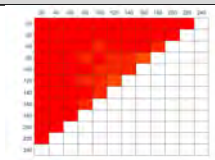
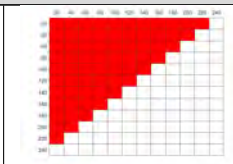
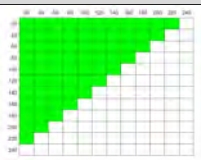
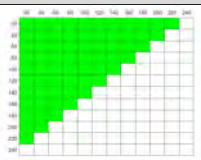
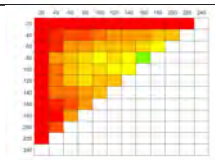
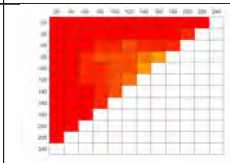
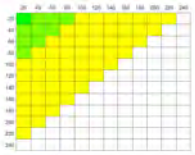
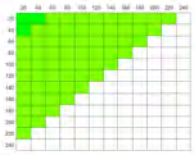
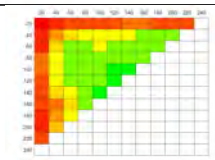
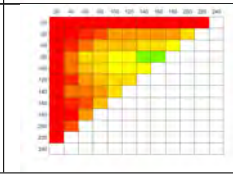
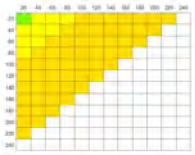
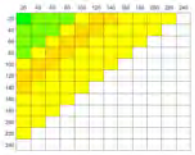
Jensen-Shannon Divergence It is premature to directly compare the Jensen-Shannon divergence (JSD) test to the t-test and KS test because the cutoff points for the JSD test have not been correlated with significance values.[3] Nevertheless, trends within the JSD test for different bin count/bin selection methods and cutoff points can be discussed. Results for the JSD test with data set truncation and multi-qubit gate error rates equal to 1 removed from the data set are presented in Tables 150-152. Table 150 shows results for bin counts $b = 15$ and $b = 25$. Tables 151 and 152 show PFA and PD results, respectively, for the bin selection methods of Scott's rule, Freedman-Diaconis rule, Sturge's rule, and the square root rule.

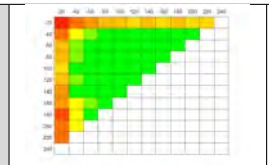
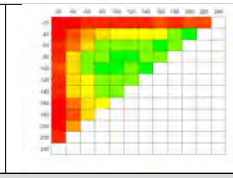
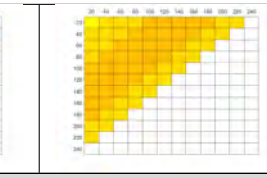
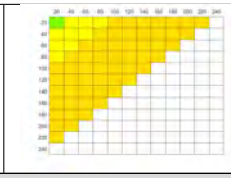
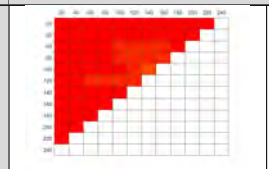
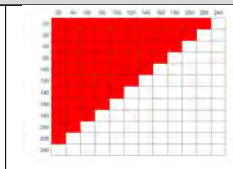
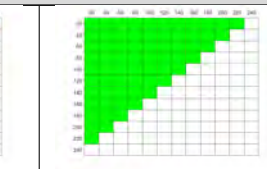
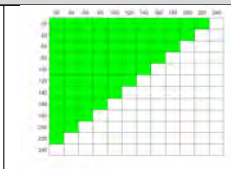
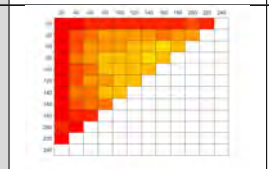
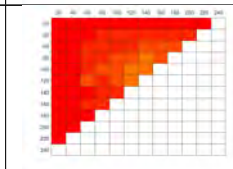
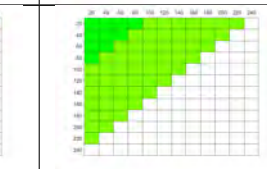
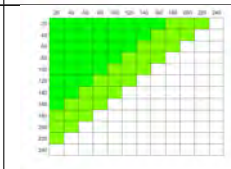
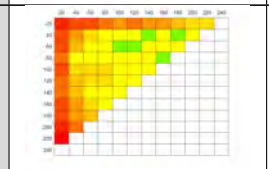
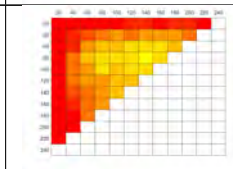
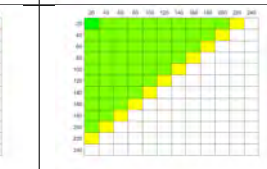
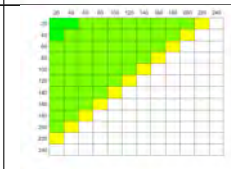
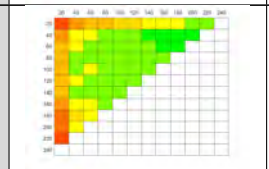
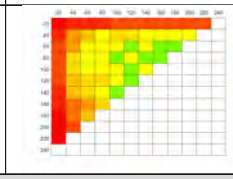
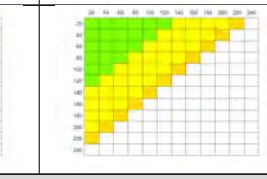
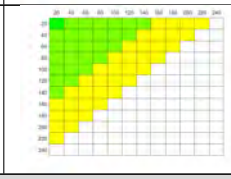
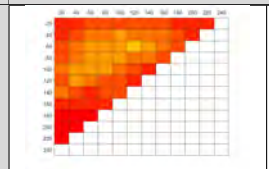
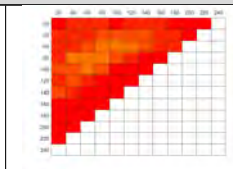
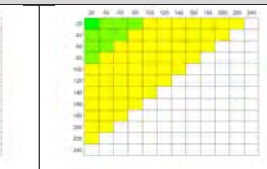
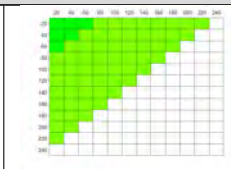
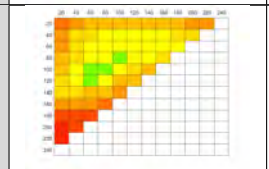
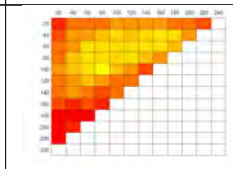
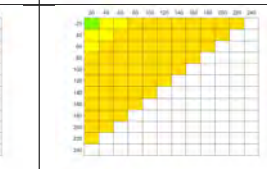
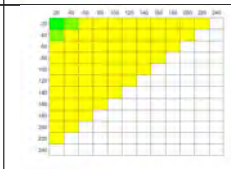
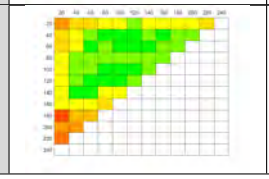
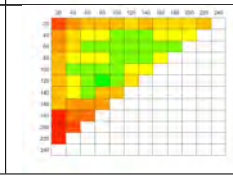
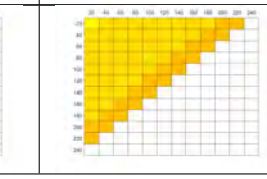
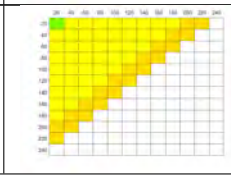
As anticipated, Table 150 shows that PFA improves with increasing cutoff values c and decreasing number of bins b, and PD improves with decreasing c and increasing b. Cutoff point $c = 0.05$ is too low to provide any useful information, as both PFA and PD values are very high for each of these tests. However, balanced performance between PFA and PD seems to occur between $c = 0.10$ and $c = 0.15$ for T1 with 15 bins and single-qubit gate error rate with 15 and 25 bins, around $c = 0.15$ for T1 with 25 bins and multi-qubit gate error rate and readout error rate with 15 and 25 bins, and around $c = 0.20$ for T2 with 15 and 25 bins. Although there are areas where performance between PFA and PD seems balanced, the PFA and PD rates in these balanced areas are still undesirable.

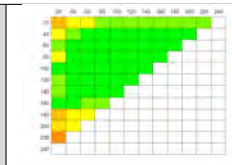
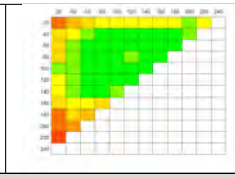
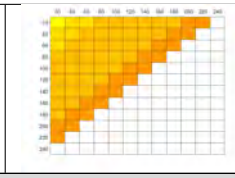
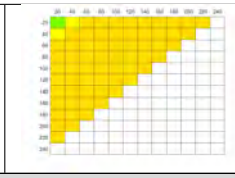
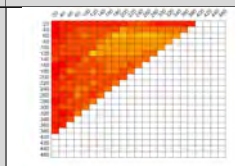
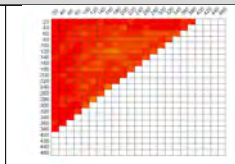
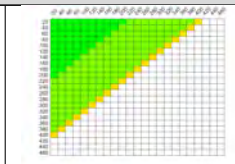
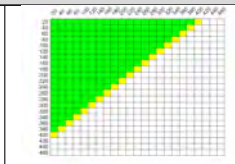
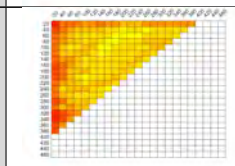
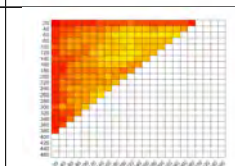
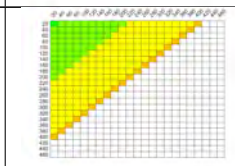
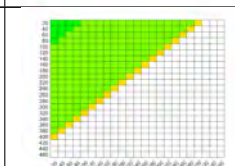
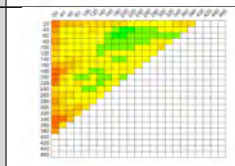
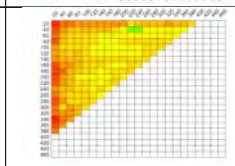
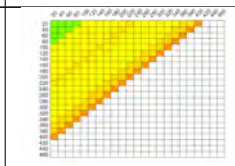
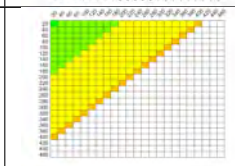
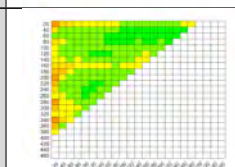
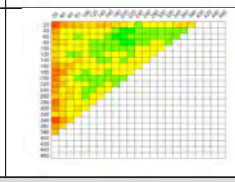
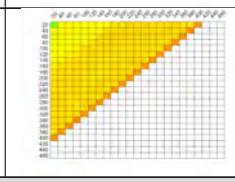
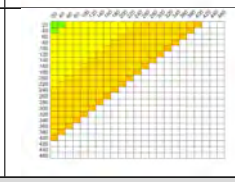
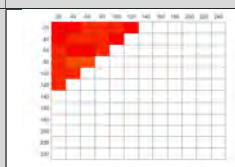
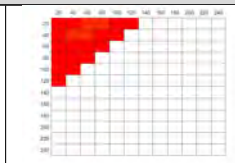
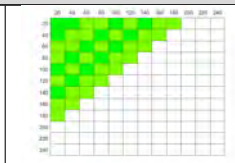
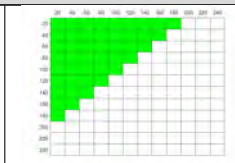
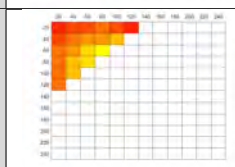
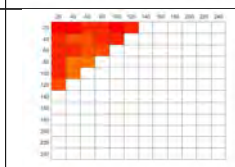
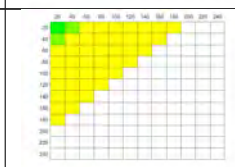
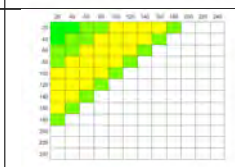
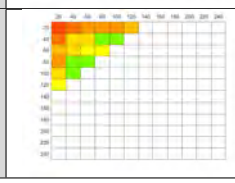
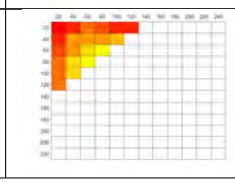
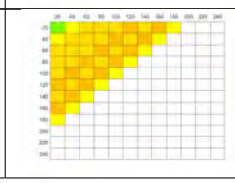
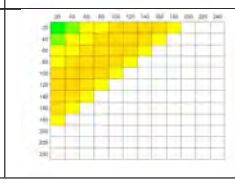
In Table 151, Scott's rule and Sturge's rule result in the lowest PFA rates for each of the five characteristics. In Table 152, the Freedman-Diaconis rule and the square root rule result in the highest PD rates for T1 and T2, and the Freedman-Diaconis rule results in in the highest PD rates for each of the error rate characteristics. For all of the binning rules, PFA and PD performance is most balanced between $c = 0.10$

[3]The JSD test is primarily applied to classification decisions rather than same-not-same decisions. For this reason, a standardized method for creating test statistics has not been created. However, emergent methods for creating test statistics have been proposed in recent literature as identified in Section 3.3.3.

and $c = 0.15$ for T1 and readout error rate, between $c = 0.15$ and $c = 0.20$ for T2, between $c = 0.05$ and $c = 0.10$ for single-qubit gate error rate, and around $c = 0.15$ for multi-qubit gate error rate. While most of the PFA rates are undesirably high and PD rates undesirably low, PFA and PD results are somewhat more acceptable for T2 characteristics with the Scott's and Sturge's rules for cutoff $c = 0.15$, approximately where $40 < N1, N2 < 100$. Nevertheless, the results are not ideal, suggesting that the tested bin selection rules, which are best suited for normal distributions, may not be appropriate for the data. Since it has been determined in Part B1 that the normal distribution is not the best-fit distribution type for any of the characteristics, more complex bin selection rules like cross-validation will be required to improve PFA and PD results for the JSD tests.

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
	0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1		0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1	
	b = 15	b = 25	b = 15	b = 25
	T1			
c = 0.05				
c = 0.10				
c = 0.15				

c = 0.20				
	T2			
c = 0.05				
c = 0.10				
c = 0.15				
c = 0.20				
	Single-Qubit Gate Error Rate			
c = 0.05				
c = 0.10				
c = 0.15				

c = 0.20				
	Multi-Qubit Gate Error Rate			
c = 0.05				
c = 0.10				
c = 0.15				
c = 0.20				
	Readout Error Rate			
c = 0.05				
c = 0.10				
c = 0.15				

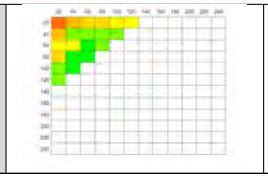
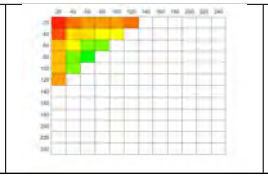
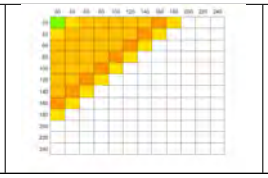
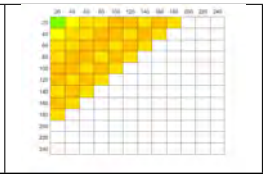
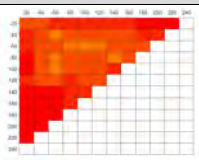
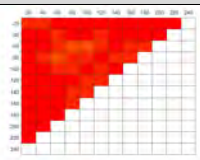
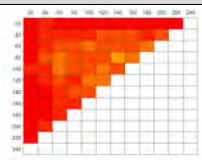
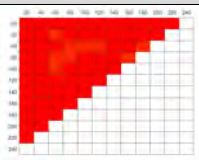
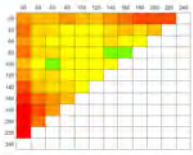
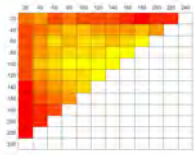
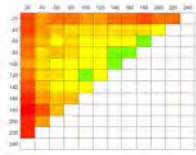
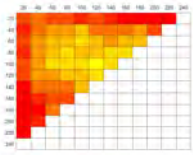
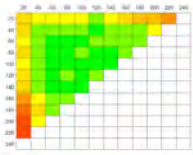
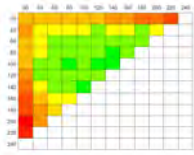
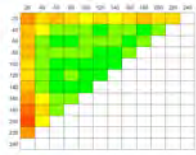
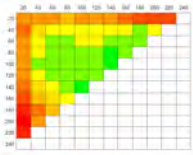
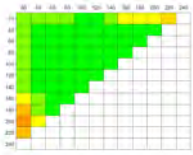
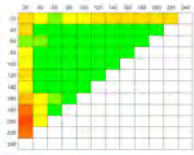
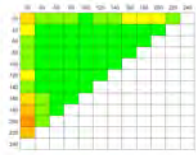
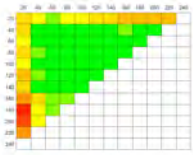
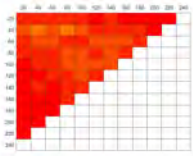
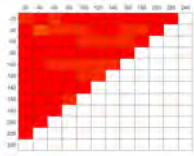
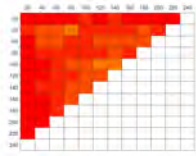
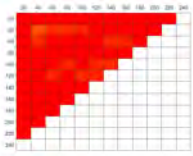
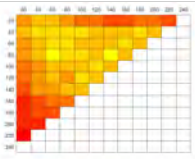
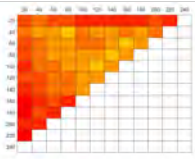
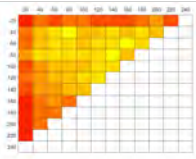
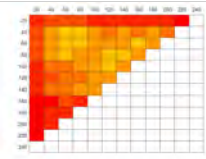
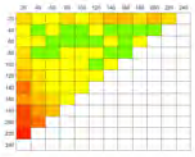
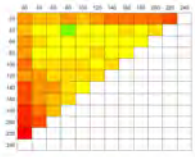
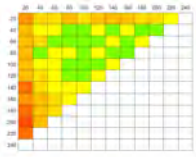
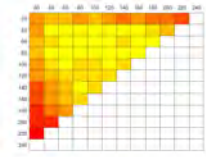
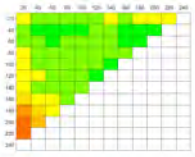
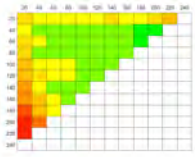
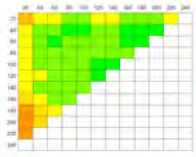
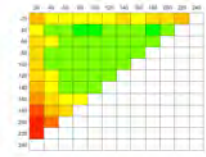
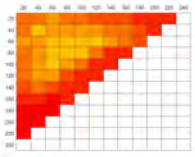
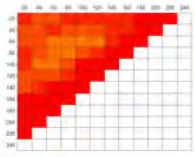
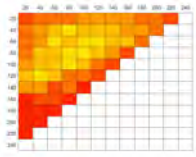
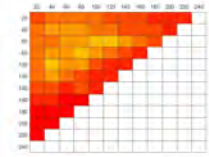
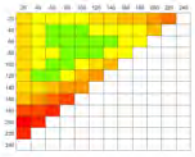
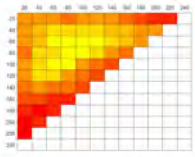
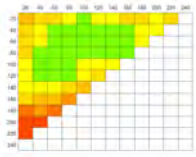
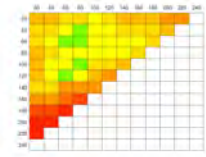
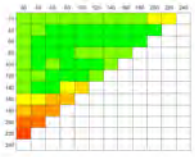
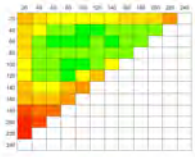
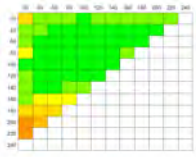
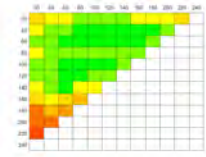
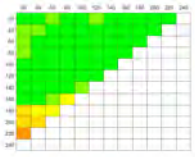
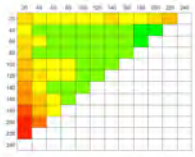
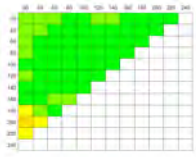
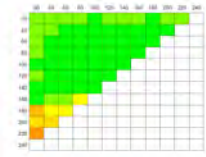
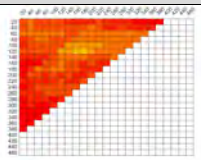
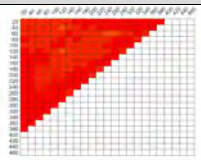
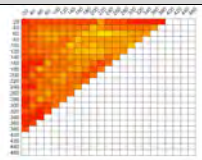
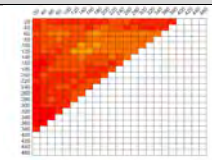
c = 0.20				

Table 150: PD/PFA Results for JSD Test with Data Set Truncation and Multi-qubit Gate Error = 1 Removed from the Data Set

	PFA			
	Color Scale for PFA Tests			
	0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1			
	Scott's Rule	F-D Rule	Sturge's Rule	Square Root Rule
	T1			
c = 0.05				
c = 0.10				
c = 0.15				
c = 0.20				
	T2			
c = 0.05				

c = 0.10				
c = 0.15				
c = 0.20				
	Single-Qubit Gate Error Rate			
c = 0.05				
c = 0.10				
c = 0.15				
c = 0.20				
	Multi-Qubit Gate Error Rate			
c = 0.05				

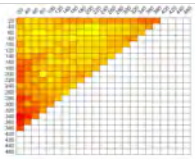
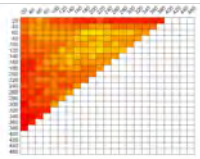
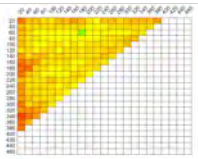
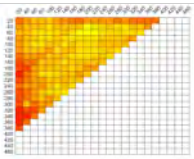
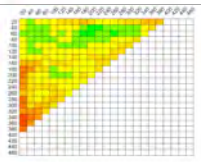
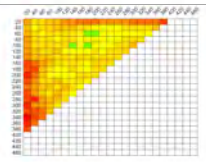
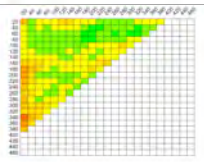
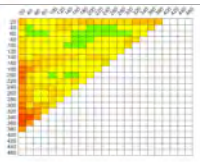
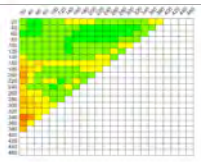
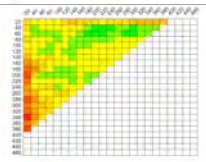
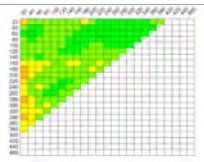
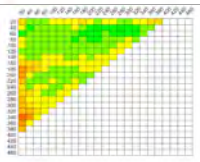
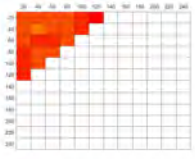
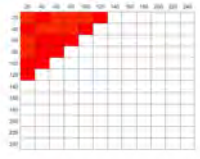
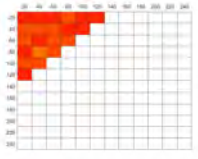
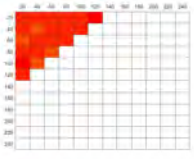
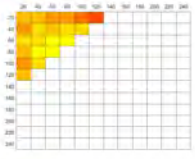
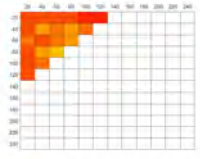
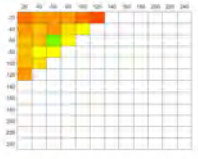
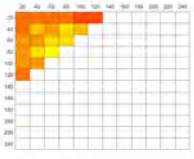
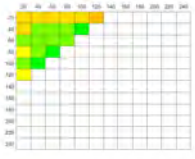
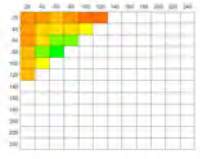
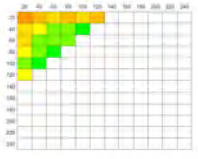
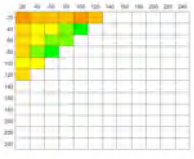
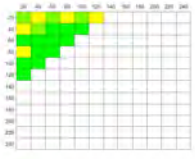
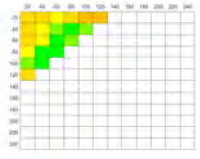
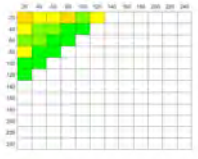
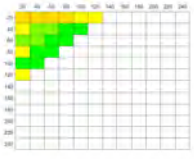
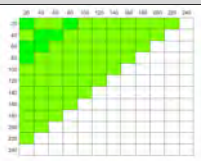
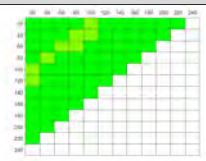
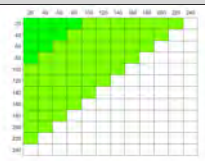
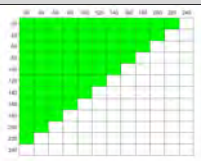
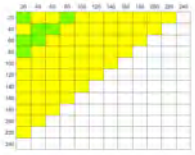
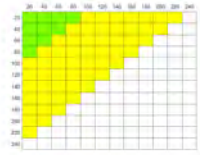
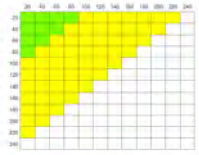
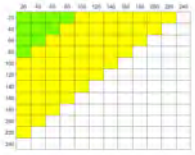
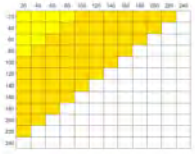
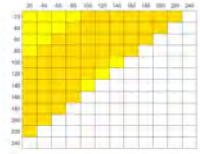
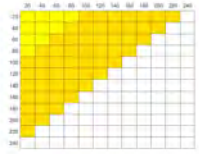
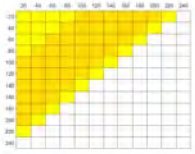
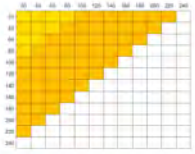
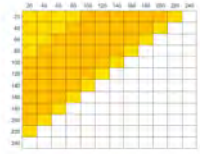
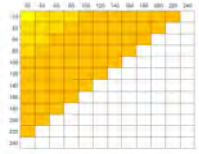
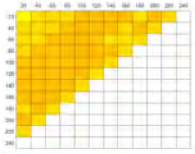
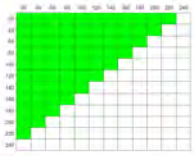
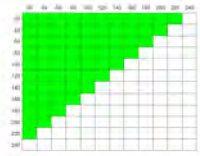
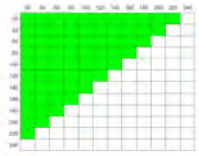
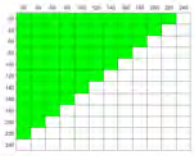
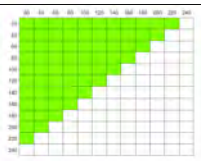
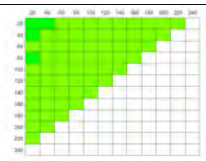
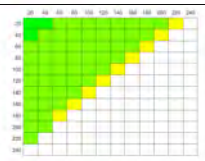
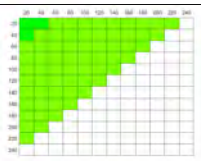
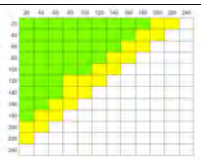
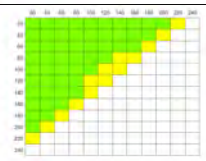
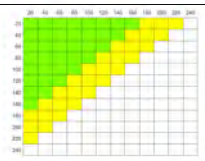
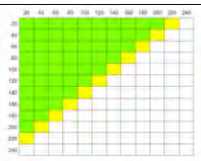
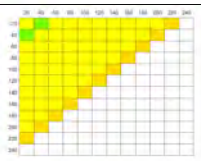
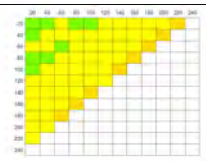
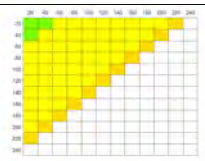
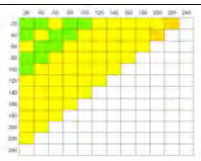
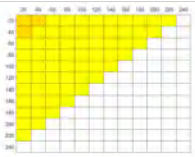
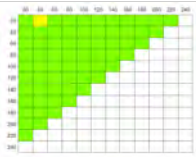
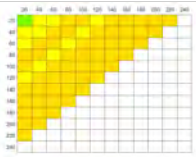
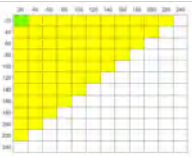
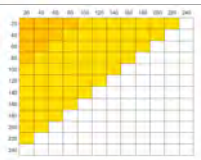
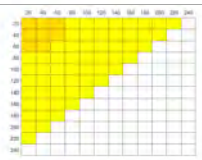
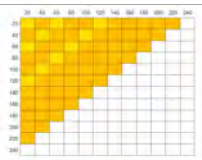
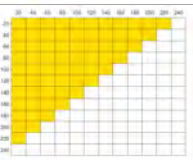
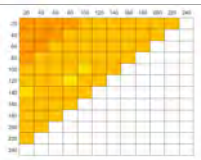
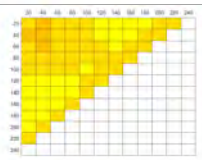
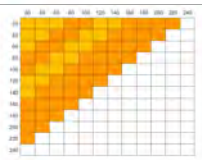
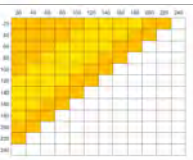
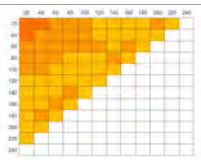
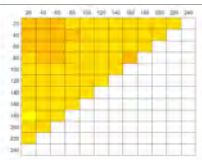
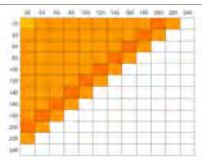
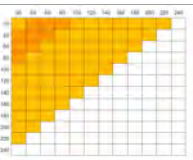
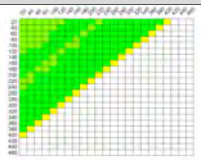
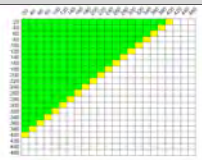
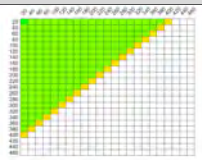
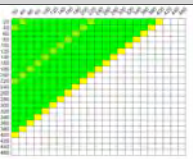
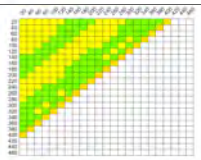
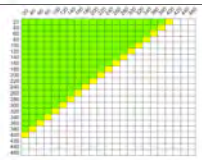
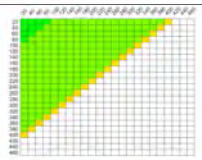
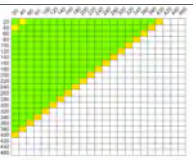
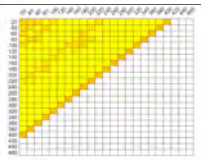
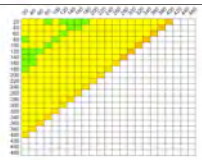
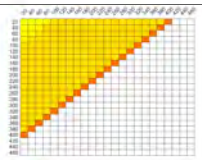
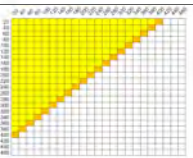
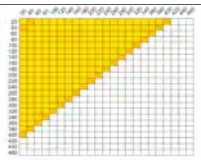
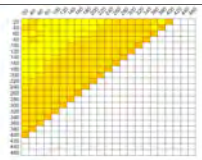
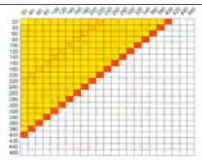
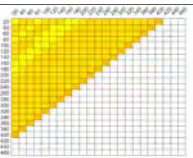
c = 0.10				
c = 0.15				
c = 0.20				
	Readout Error Rate			
c = 0.05				
c = 0.10				
c = 0.15				
c = 0.20				

Table 151: PFA Results for JSD Test with Data Set Truncation and Multi-Qubit Gate Error = 1 Removed from the Data Set, Number of Bins b Selected for Each Comparison by Binning Algorithms

	PD
	Color Scale for PD Tests
	

	Scott's Rule	F-D Rule	Sturge's Rule	Square Root Rule
	T1			
c = 0.05				
c = 0.10				
c = 0.15				
c = 0.20				
	T2			
c = 0.05				
c = 0.10				
c = 0.15				
c = 0.20				
	Single-Qubit Gate Error Rate			

c = 0.05				
c = 0.10				
c = 0.15				
c = 0.20				
	Multi-Qubit Gate Error Rate			
c = 0.05				
c = 0.10				
c = 0.15				
c = 0.20				
	Readout Error Rate			

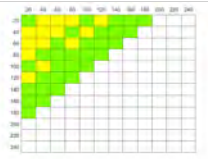
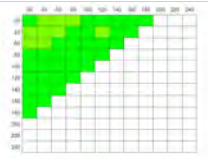
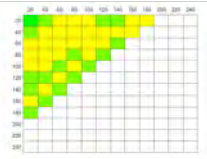
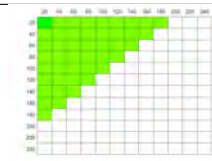
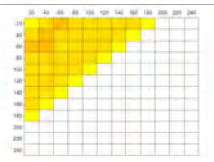
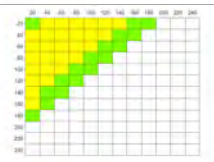
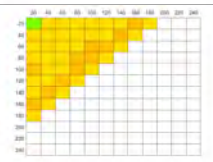
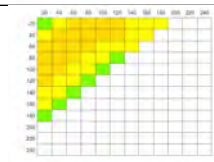
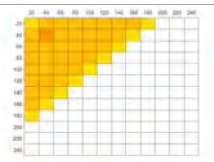
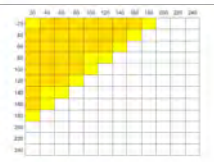
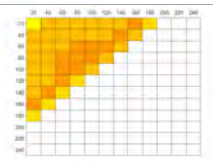
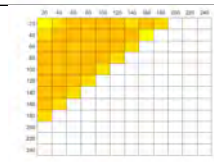
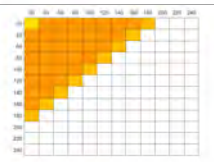
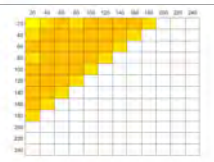
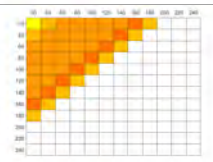
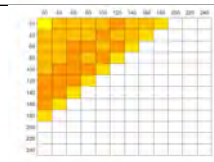
c = 0.05				
c = 0.10				
c = 0.15				
c = 0.20				

Table 152: PD Results for JSD Test with Data Set Truncation and Multi-Qubit Gate Error = 1 Removed from the Data Set, Number of Bins b Selected for Each Comparison by Binning Algorithms

Quadratic Discriminant Analysis Results for quadratic discriminant analysis (QDA) in Table 153 show that PFA is moderately high for all data set sizes, and PD is somewhat higher than PFA for all data set sizes. Observing high PFA values in conjunction with the high PD values serves as evidence of overfitting. As overfitting often occurs when a model is too complex (for example, when a model has too many parameters), conducting a principal component analysis (PCA) or multiple discriminant analysis (MDA) technique to reduce the dimensionality could improve results [67]. Another possible reason for the undesirable results is that QDA assumes multivariate normality, but the data are not normally distributed. As such, QDA is not recommended for further development with the current characteristics.

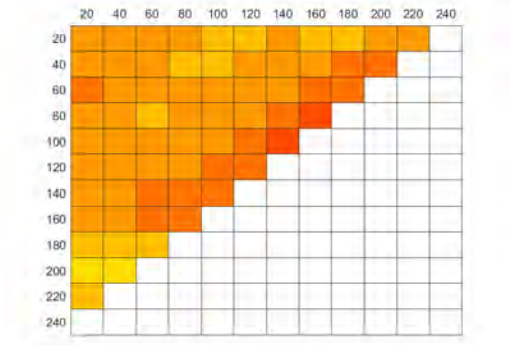
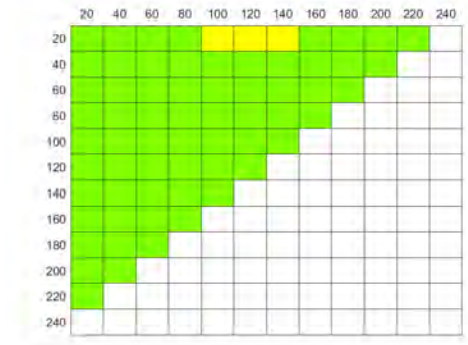
PFA	PD
Color Scale for PFA Tests	Color Scale for PD Tests
	

Table 153: PD/PFA Results for QDA with Data Set Truncation

4.3.4 Two-Dimensional Kolmogorov-Smirnov Test

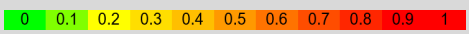
Results for the two-dimensional KS tests are presented in Tables 154-181. In most cases for one-qubit and two-qubit two-dimensional KS tests, both PD and PFA are very high, which indicates overfitting. Implementing dimensionality reduction techniques can possibly improve results. As an example, principal component analysis or linear discriminant analysis could find the best separation of qubit (set) data, and then the one-dimensional KS test would serve as the qubit discrimination method.

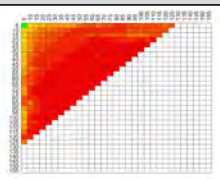
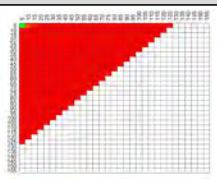
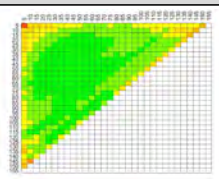
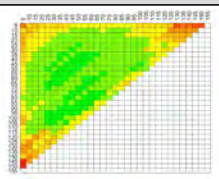
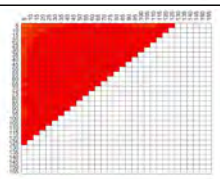
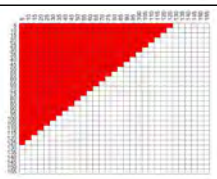
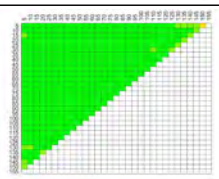
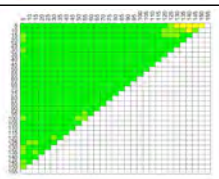
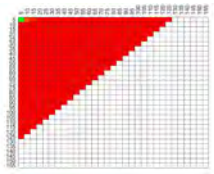
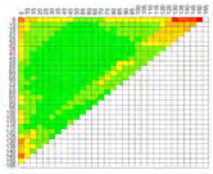
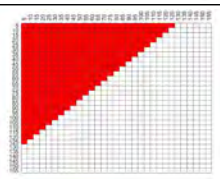
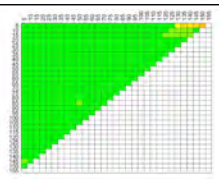
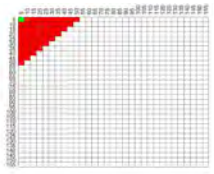
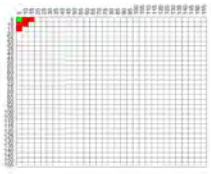
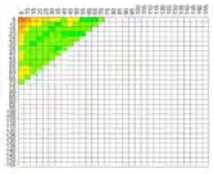
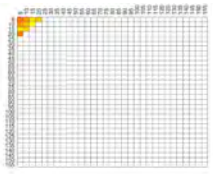
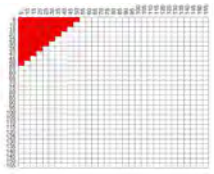
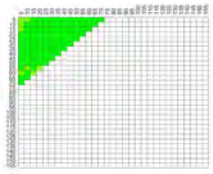
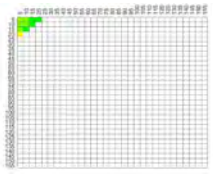
Results for most of the three-qubit tests are, in simple terms, disappointing. PD is low while PFA is high. It is likely that too many sources of noise and error are introduced in three-qubit randomized benchmarking for the characteristic pairs to be useful for qubit set discrimination in most cases. There are a few interesting areas worth noting, however, where this general trend does not apply. Unique results occur in Table 164 where the two characteristics are the 1-Qubit Randomized Benchmarking, Seed 4, *err_a* parameter (RB1 Seed 4 *err_a*) and *err_b* parameter (RB1 Seed 4 *err_b*), in Table 173 where the two characteristics are the 2-Qubit Randomized

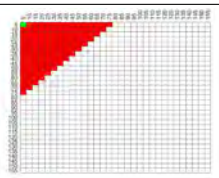
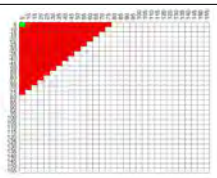
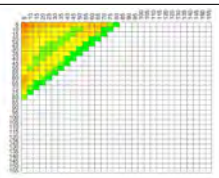
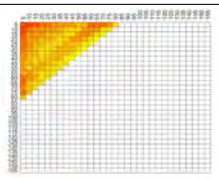
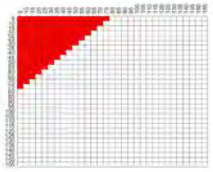
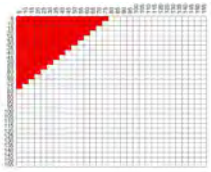
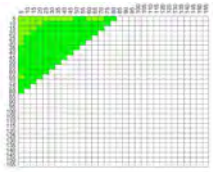
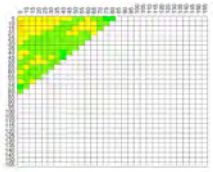
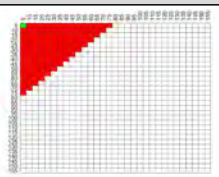
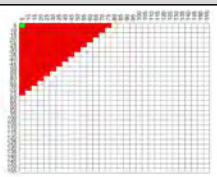
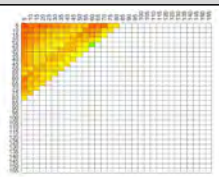
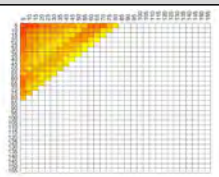
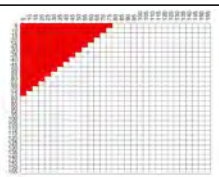
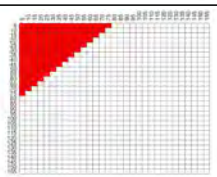
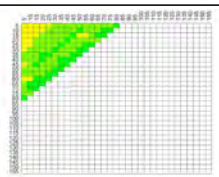
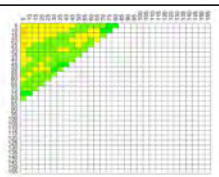
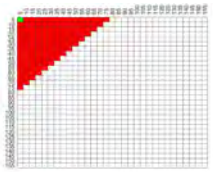
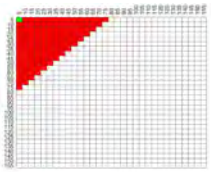
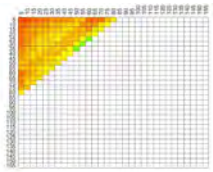
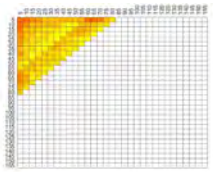
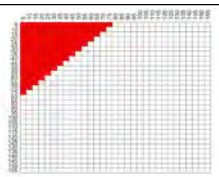
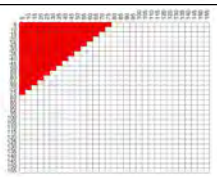
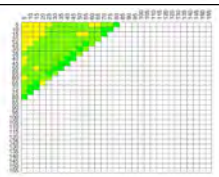
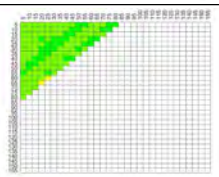
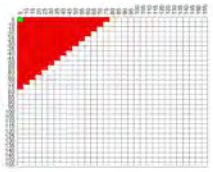
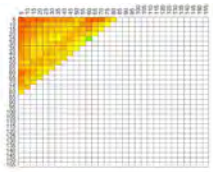
Benchmarking, Seed 4 err_a parameter (RB2 Seed 4 err_a) and err_α parameter (RB2 Seed 4 err_α), in Table 176 where the two characteristics are the 3-Qubit Randomized Benchmarking, Seed 4 Fidelity (RB3 Seed 4 Fidelity) and b parameter (RB3 Seed 4 b), and in Table 180 where the two characteristics are the 3-Qubit Randomized Benchmarking, Seed 4 err_a parameter (RB3 Seed 4 err_a) and err_α parameter (RB3 Seed 4 err_α).

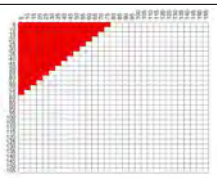
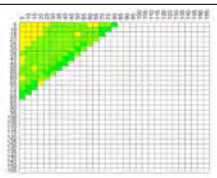
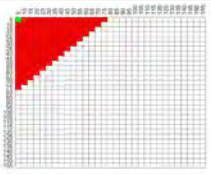
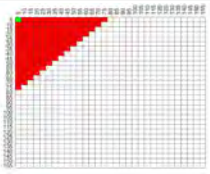
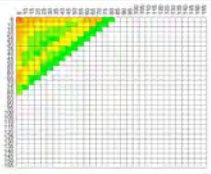
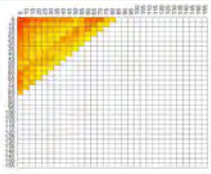
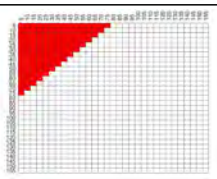
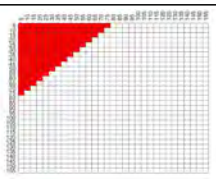
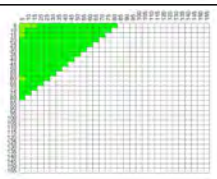
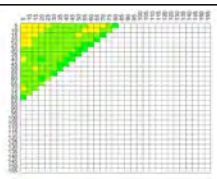
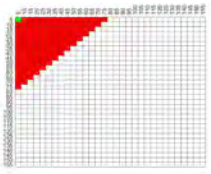
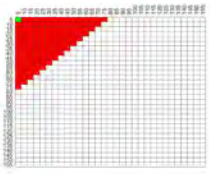
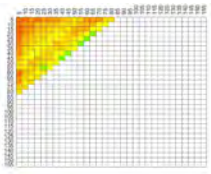
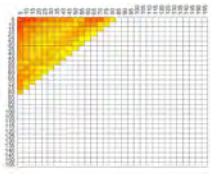
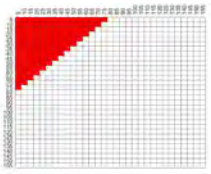
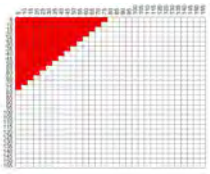
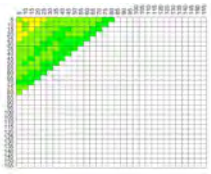
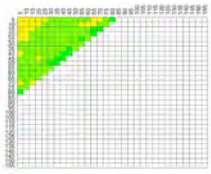
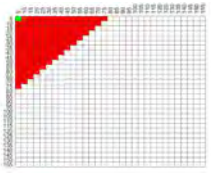
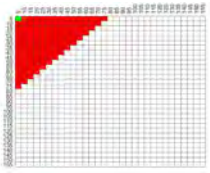
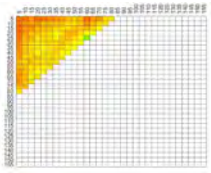
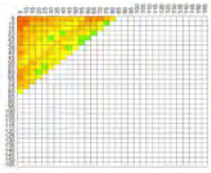
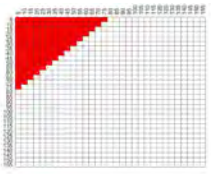
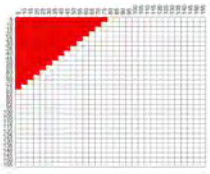
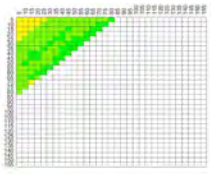
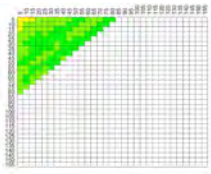
With these four characteristic pairs, the PD and PFA values are both low, which indicates underfitting. Thus, these combinations of characteristics make good candidates for experimentation with 3-dimensional KS testing. A third characteristic could provide enough specificity per qubit to maintain low PFA while improving PD.

It is curious that the characteristic pairs that indicate underfitting rather than overfitting change as the number of qubits is increased for randomized benchmarking. In the one-qubit and two-qubit cases, all characteristics describe error ranges on the fidelity equation parameters. In the one-qubit case, the error ranges are for parameters that absorb state preparation and measurement errors. In the two-qubit case, the error ranges are for the a parameter, which absorbs state preparation and measurement errors, and the α parameter, which describes the decay rate. In the three-qubit case, however, one of the characteristic pairs does not describe the error ranges at all, but rather the fidelity and the b parameter. It would make an interesting future study to investigate what changes occur in the 1-to-2-to-3-qubit randomized benchmarking processes that cause the shift in behavior.

	PFA	PD
	Color Scale for PFA Tests	Color Scale for PD Tests
		
	T1	

	T2	Single-Qubit Gate Error	T2	Single-Qubit Gate Error
$\alpha = 0.001$				
$\alpha = 0.05$				
	Readout Error		Readout Error	
$\alpha = 0.001$				
$\alpha = 0.05$				
	QPTX	QPTH	QPTX	QPTH
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 0 Fid.	RB1 Seed 0 a	RB1 Seed 0 Fid.	RB1 Seed 0 a

$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 0 α	RB1 Seed 0 b	RB1 Seed 0 α	RB1 Seed 0 b
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 0 err_a	RB1 Seed 0 err_α	RB1 Seed 0 err_a	RB1 Seed 0 err_α
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 0 err_b		RB1 Seed 0 err_b	
$\alpha = 0.001$				

$\alpha = 0.05$				
	RB1 Seed 4 Fid.	RB1 Seed 4 a	RB1 Seed 4 Fid.	RB1 Seed 4 a
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 α	RB1 Seed 4 b	RB1 Seed 4 α	RB1 Seed 4 b
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 *err_a*	RB1 Seed 4 *err_α*	RB1 Seed 4 *err_a*	RB1 Seed 4 *err_α*
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 *err_b*		RB1 Seed 4 *err_b*	

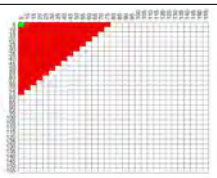
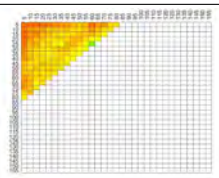
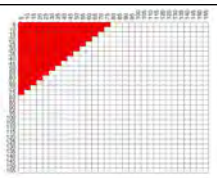
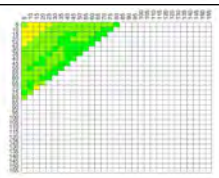
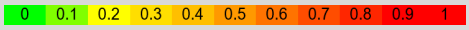
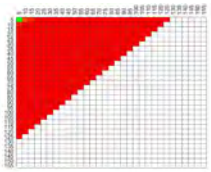
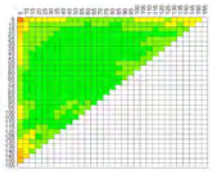
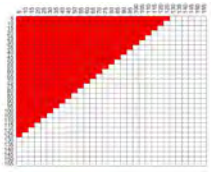
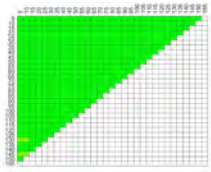
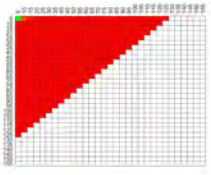
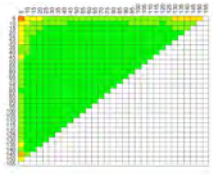
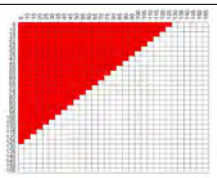
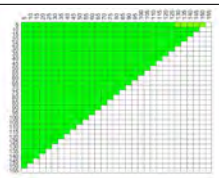
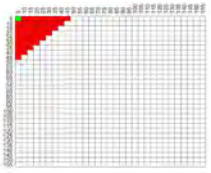
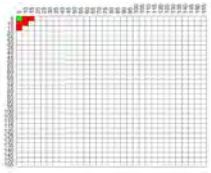
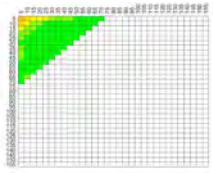
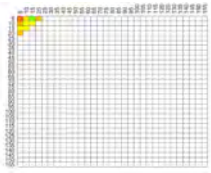
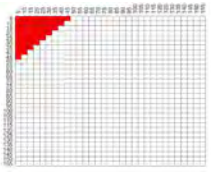
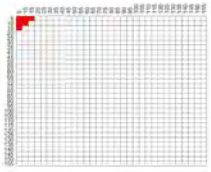
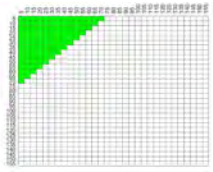
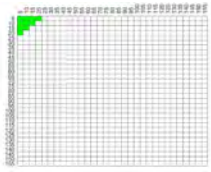
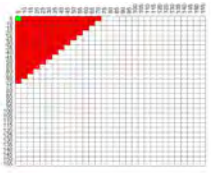
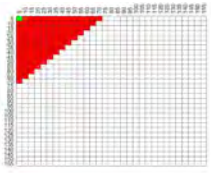
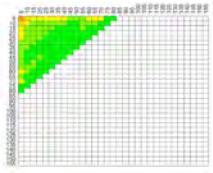
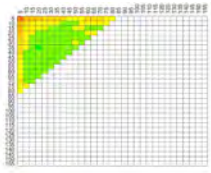
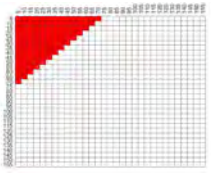
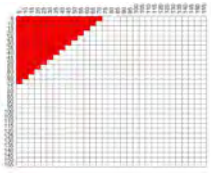
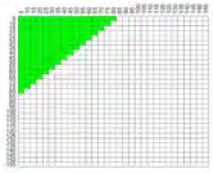
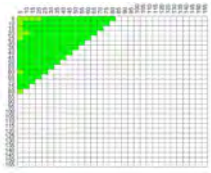
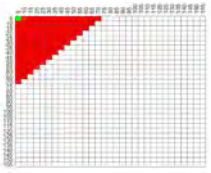
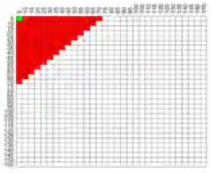
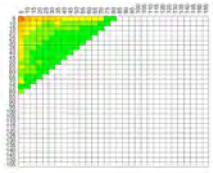
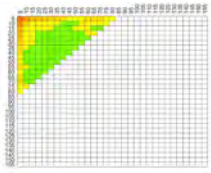
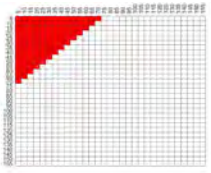
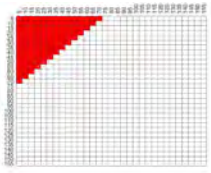
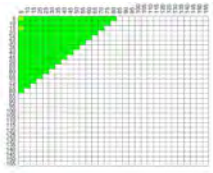
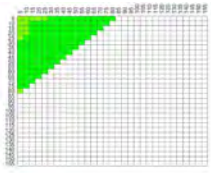
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 154: PD/PFA Results for 2-Dimensional KS Test, Single-Qubit, Part I

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	T2			
		Single-Qubit Gate Error		Single-Qubit Gate Error
$\alpha = 0.001$				
$\alpha = 0.05$				
	Readout Error		Readout Error	
$\alpha = 0.001$				

$\alpha = 0.05$				
	QPTX	QPTH	QPTX	QPTH
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 Fid.	RB1 Seed 4 a	RB1 Seed 4 Fid.	RB1 Seed 4 a
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 α	RB1 Seed 4 b	RB1 Seed 4 α	RB1 Seed 4 b
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 err_a	RB1 Seed 4 err_α	RB1 Seed 4 err_a	RB1 Seed 4 err_α

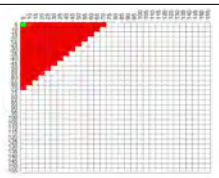
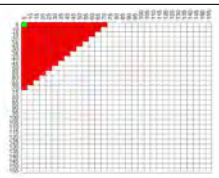
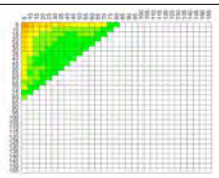
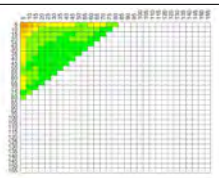
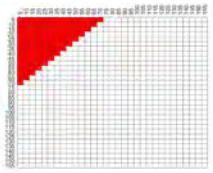
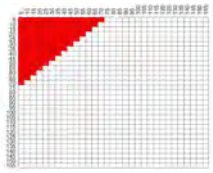
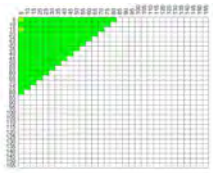
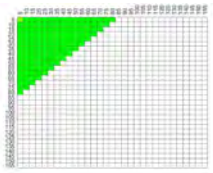
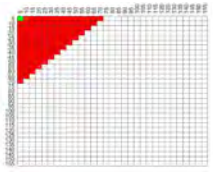
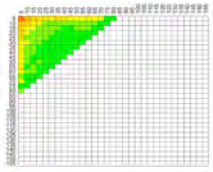
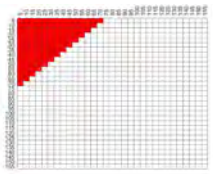
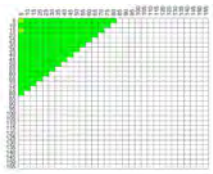
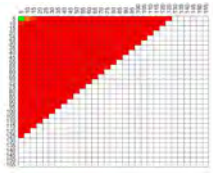
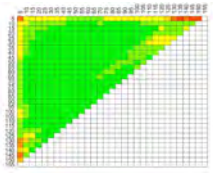
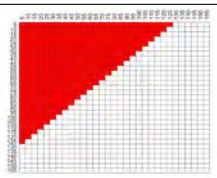
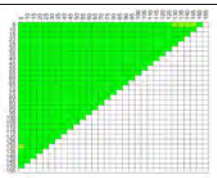
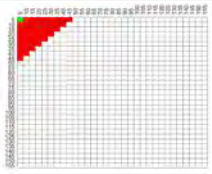
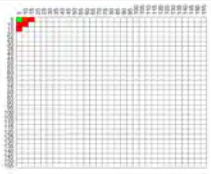
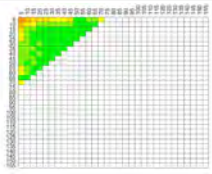
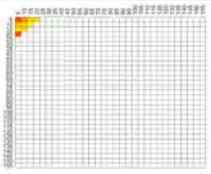
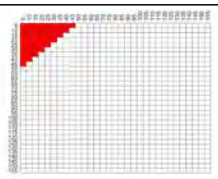
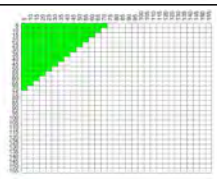
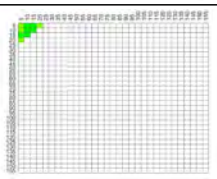
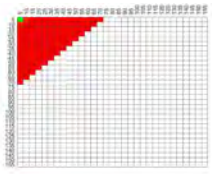
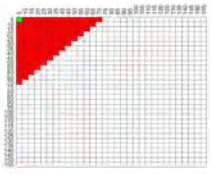
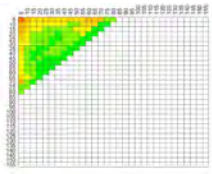
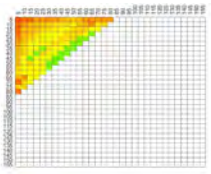
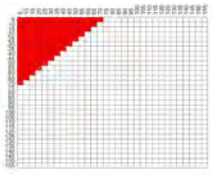
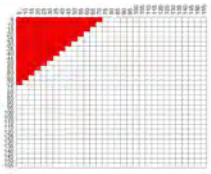
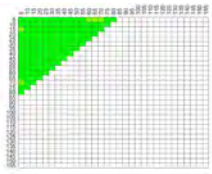
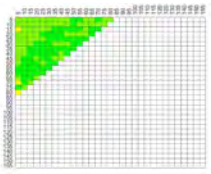
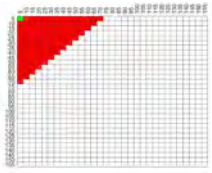
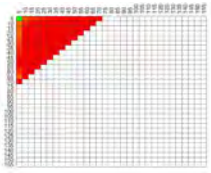
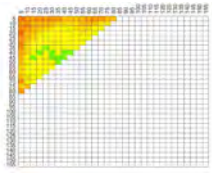
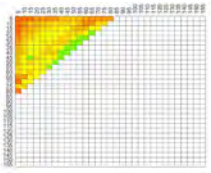
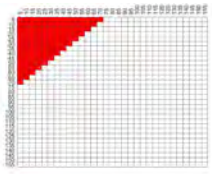
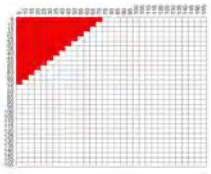
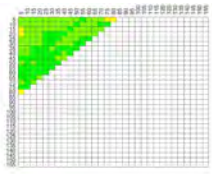
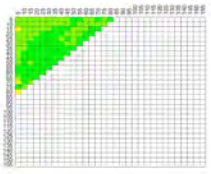
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 *err_b*		RB1 Seed 4 *err_b*	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 155: PD/PFA Results for 2-Dimensional KS Test, Single-Qubit, Part II

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	Single-Qubit Gate Error			
		Readout Error		Readout Error
$\alpha = 0.001$				

$\alpha = 0.05$				
	QPTX	QPTH	QPTX	QPTH
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 Fid.	RB1 Seed 4 a	RB1 Seed 4 Fid.	RB1 Seed 4 a
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 α	RB1 Seed 4 b	RB1 Seed 4 α	RB1 Seed 4 b
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 err_a	RB1 Seed 4 err_α	RB1 Seed 4 err_a	RB1 Seed 4 err_α

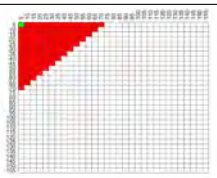
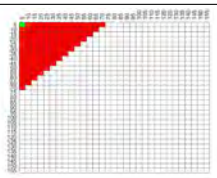
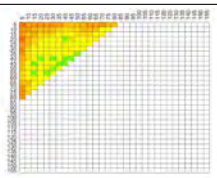
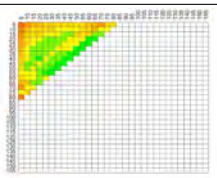
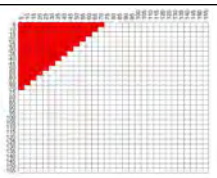
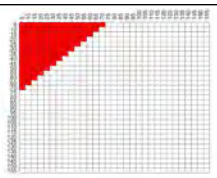
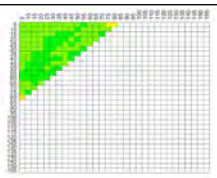
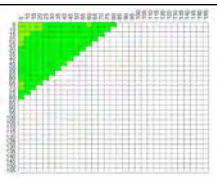
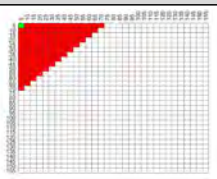
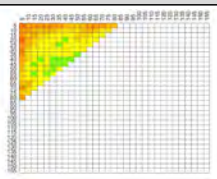
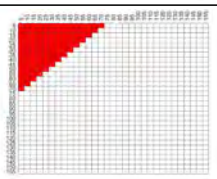
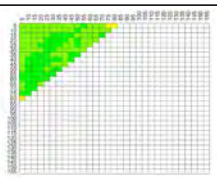
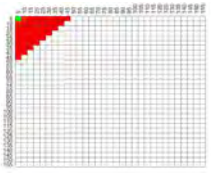
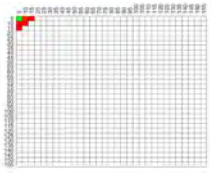
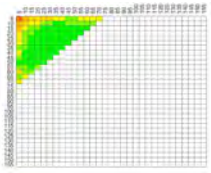
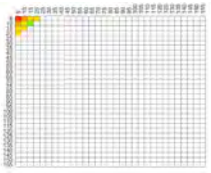
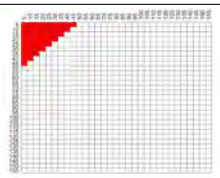
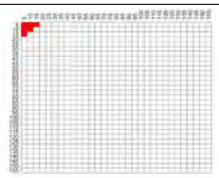
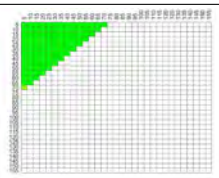
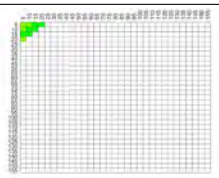
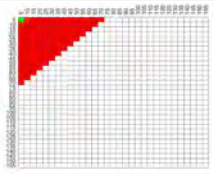
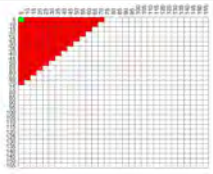
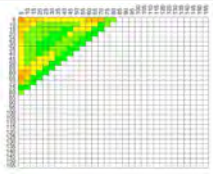
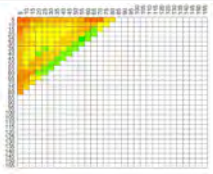
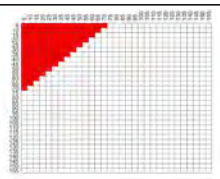
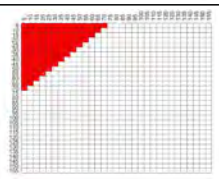
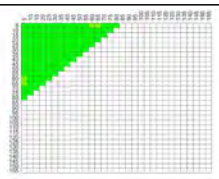
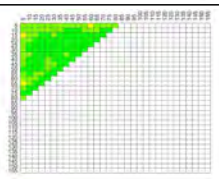
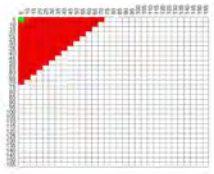
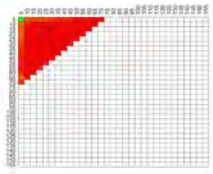
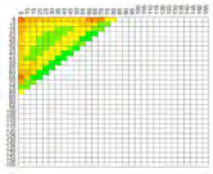
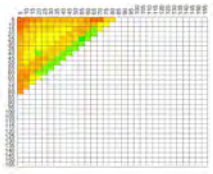
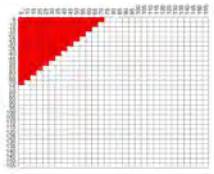
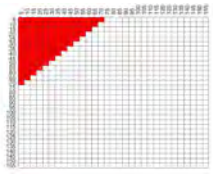
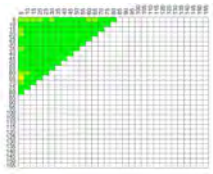
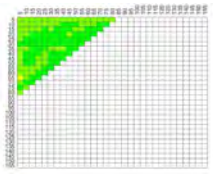
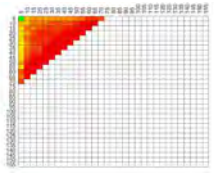
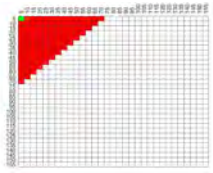
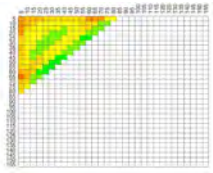
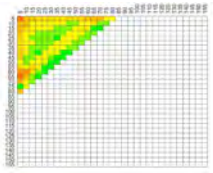
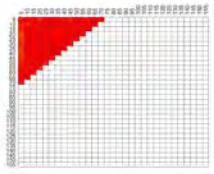
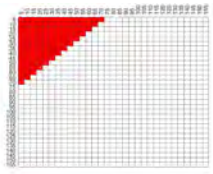
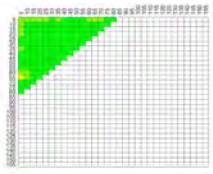
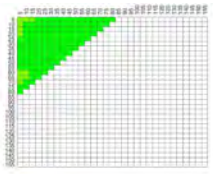
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 *err_b*		RB1 Seed 4 *err_b*	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 156: PD/PFA Results for 2-Dimensional KS Test, Single-Qubit, Part III

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	Readout Error			
	QPTX	QPTH	QPTX	QPTH
$\alpha = 0.001$				

$\alpha = 0.05$				
	RB1 Seed 4 Fid.	RB1 Seed 4 a	RB1 Seed 4 Fid.	RB1 Seed 4 a
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 α	RB1 Seed 4 b	RB1 Seed 4 α	RB1 Seed 4 b
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 *err_a*	RB1 Seed 4 *err_*α	RB1 Seed 4 *err_a*	RB1 Seed 4 *err_*α
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 *err_b*		RB1 Seed 4 *err_b*	

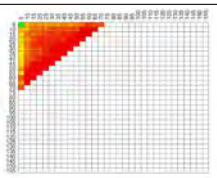
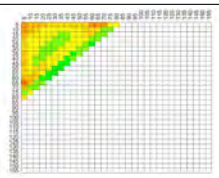
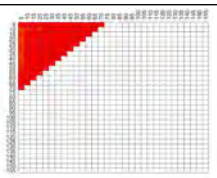
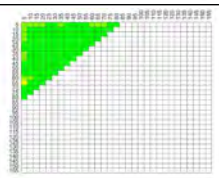
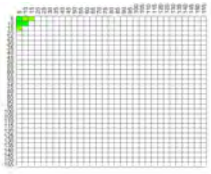
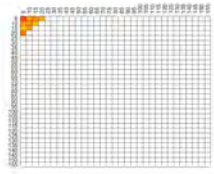
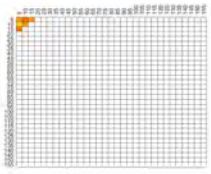
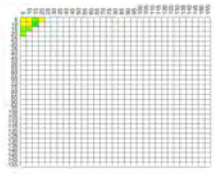
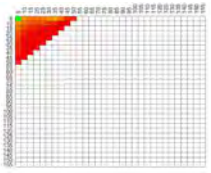
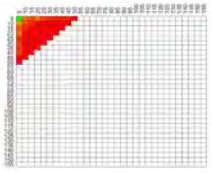
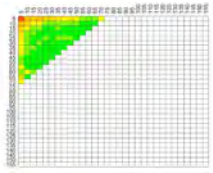
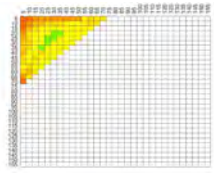
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 157: PD/PFA Results for 2-Dimensional KS Test, Single-Qubit, Part IV

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	QPTX			
		QPTH		QPTH
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 Fid.	RB1 Seed 4 a	RB1 Seed 4 Fid.	RB1 Seed 4 a
$\alpha = 0.001$				

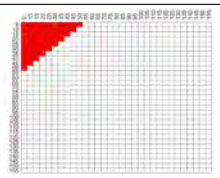
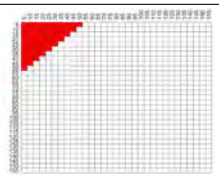
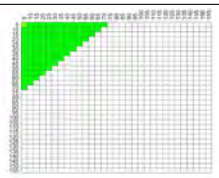
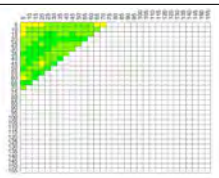
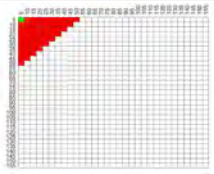
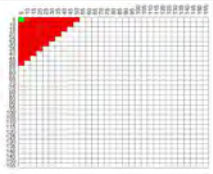
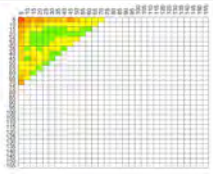
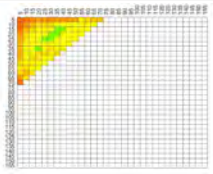
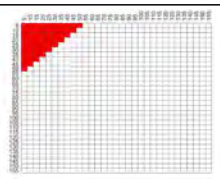
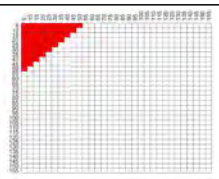
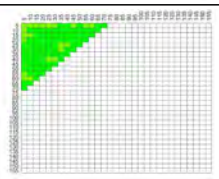
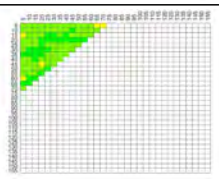
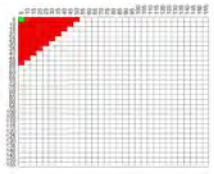
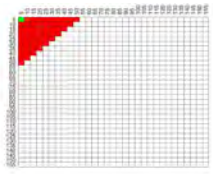
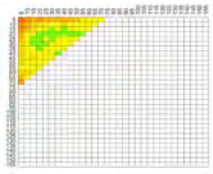
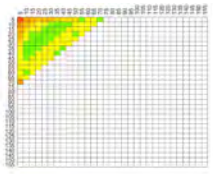
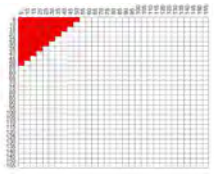
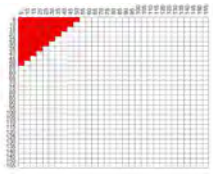
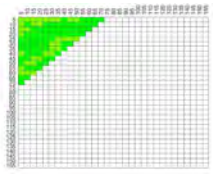
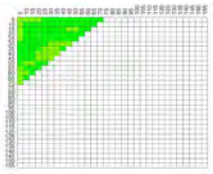
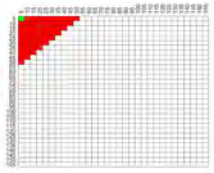
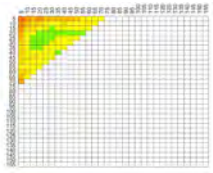
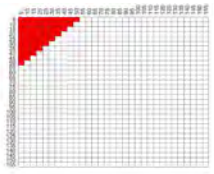
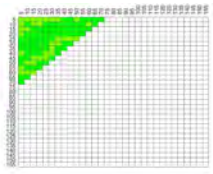
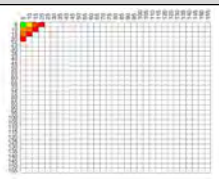
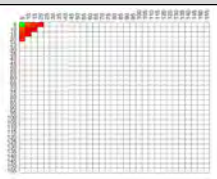
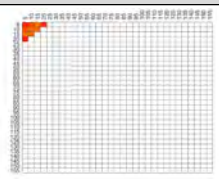
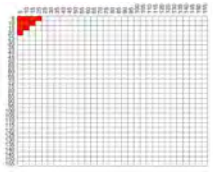
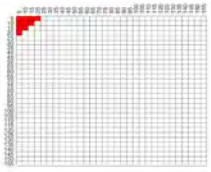
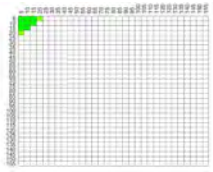
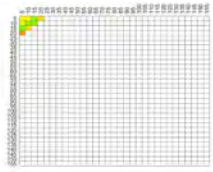
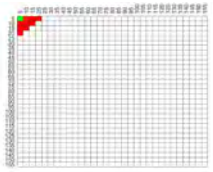
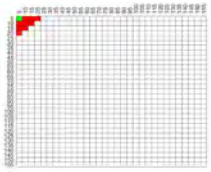
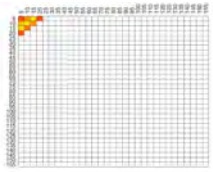
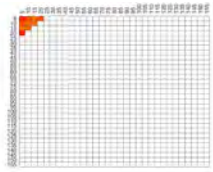
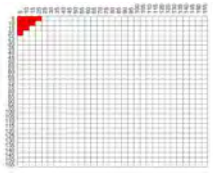
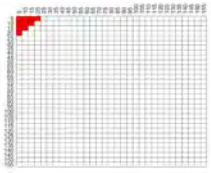
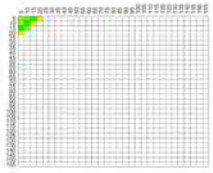
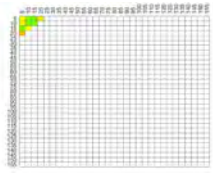
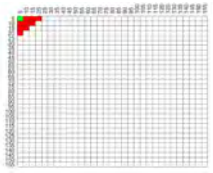
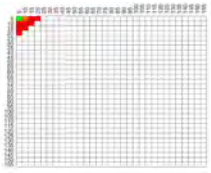
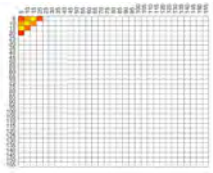
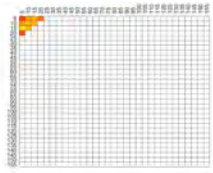
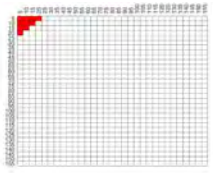
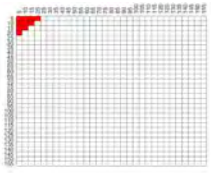
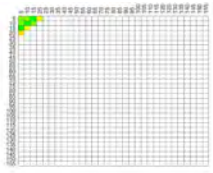
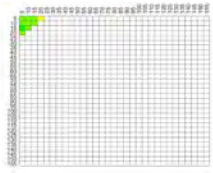
$\alpha = 0.05$				
	RB1 Seed 4 α	RB1 Seed 4 b	RB1 Seed 4 α	RB1 Seed 4 b
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 err_a	RB1 Seed 4 err_α	RB1 Seed 4 err_a	RB1 Seed 4 err_α
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 err_b		RB1 Seed 4 err_b	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 158: PD/PFA Results for 2-Dimensional KS Test, Single-Qubit, Part V

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	QPTH			
	RB1 Seed 4 Fid.	RB1 Seed 4 a	RB1 Seed 4 Fid.	RB1 Seed 4 a
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 α	RB1 Seed 4 b	RB1 Seed 4 α	RB1 Seed 4 b
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 *err_a*	RB1 Seed 4 *err_*α	RB1 Seed 4 *err_a*	RB1 Seed 4 *err_*α
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 *err_b*		RB1 Seed 4 *err_b*	

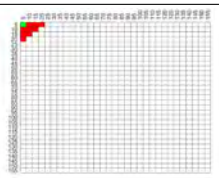
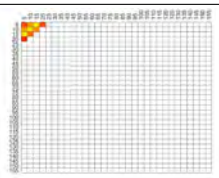
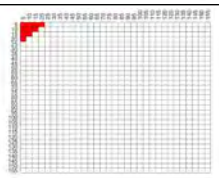
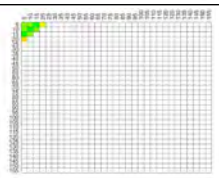
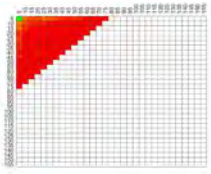
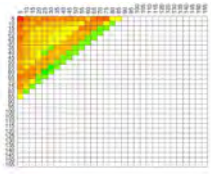
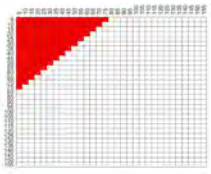
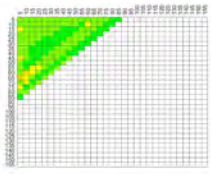
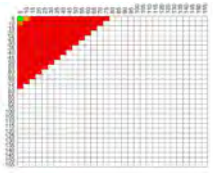
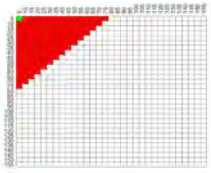
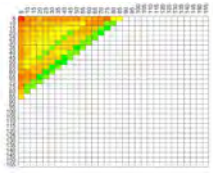
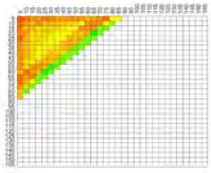
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 159: PD/PFA Results for 2-Dimensional KS Test, Single-Qubit, Part VI

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	RB1 Seed 4 Fidelity			
		RB1 Seed 4 a		RB1 Seed 4 a
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 α	RB1 Seed 4 b	RB1 Seed 4 α	RB1 Seed 4 b
$\alpha = 0.001$				

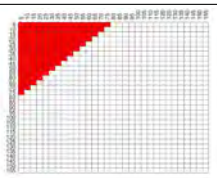
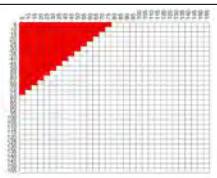
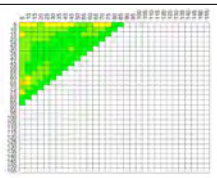
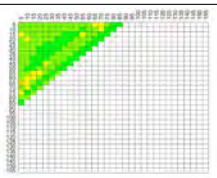
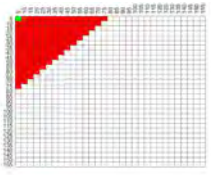
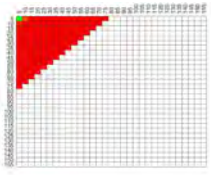
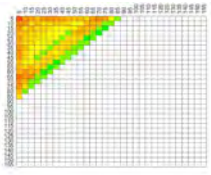
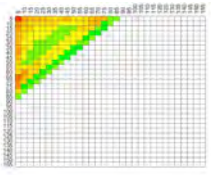
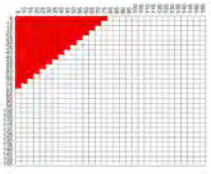
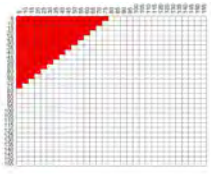
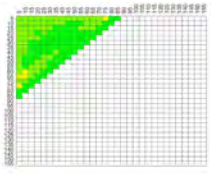
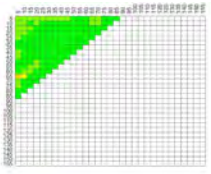
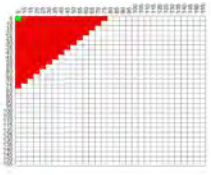
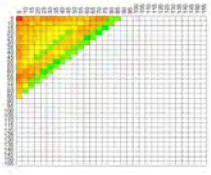
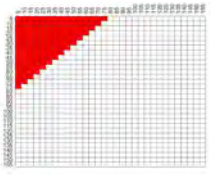
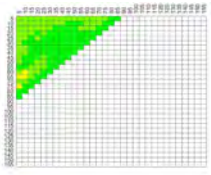
$\alpha = 0.05$				
	RB1 Seed 4 *err_a*	RB1 Seed 4 *err_α*	RB1 Seed 4 *err_a*	RB1 Seed 4 *err_α*
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 *err_b*		RB1 Seed 4 *err_b*	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 160: PD/PFA Results for 2-Dimensional KS Test, Single-Qubit, Part VII

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	RB1 Seed 4 a			
	RB1 Seed 4 α	RB1 Seed 4 b	RB1 Seed 4 α	RB1 Seed 4 b

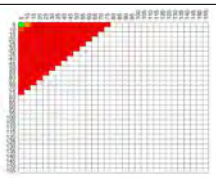
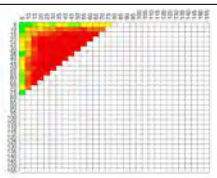
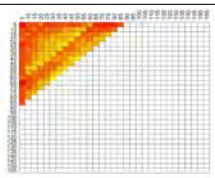
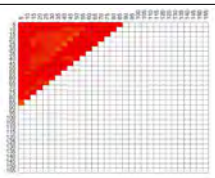
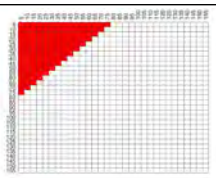
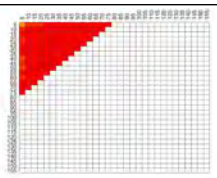
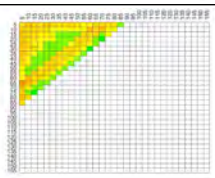
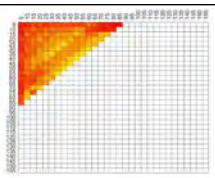
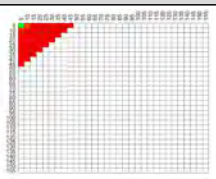
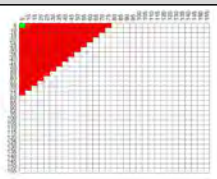
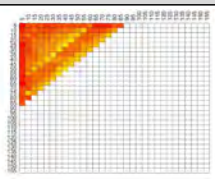
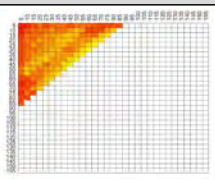
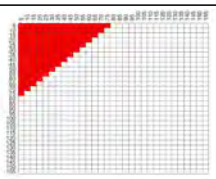
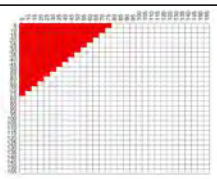
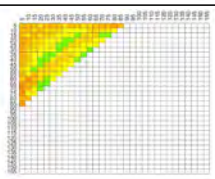
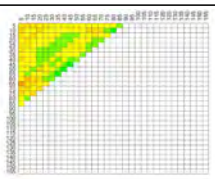
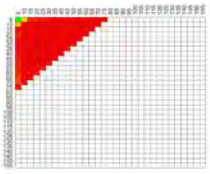
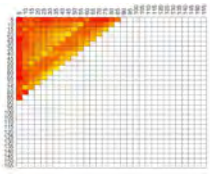
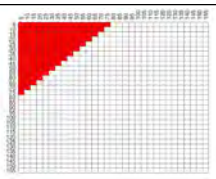
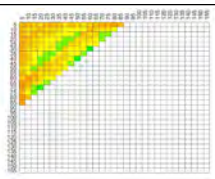
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 *err_a*	RB1 Seed 4 *err_α*	RB1 Seed 4 *err_a*	RB1 Seed 4 *err_α*
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 *err_b*		RB1 Seed 4 *err_b*	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 161: PD/PFA Results for 2-Dimensional KS Test, Single-Qubit, Part VIII

	PFA	PD
	Color Scale for PFA Tests	Color Scale for PD Tests
		

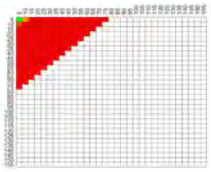
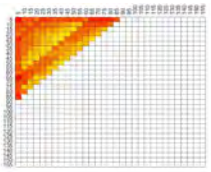
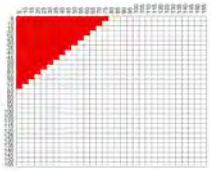
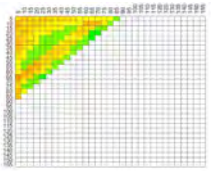
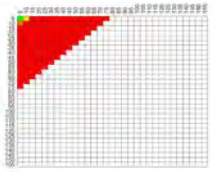
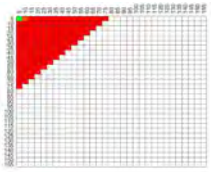
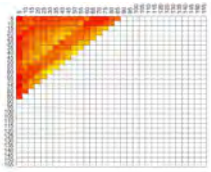
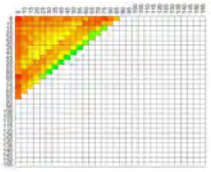
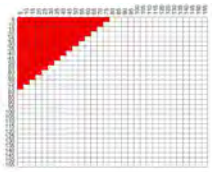
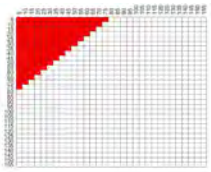
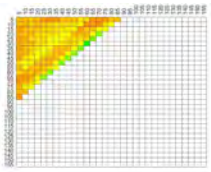
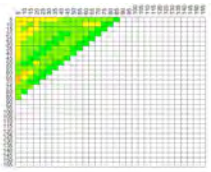
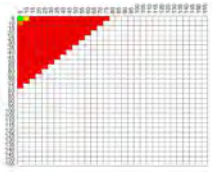
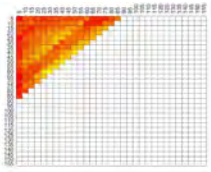
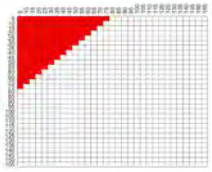
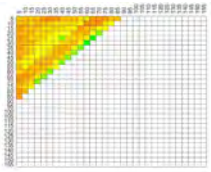
	RB1 Seed 4 α			
		RB1 Seed 4 b		RB1 Seed 4 b
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 err_a	RB1 Seed 4 err_α	RB1 Seed 4 err_a	RB1 Seed 4 err_α
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 err_b		RB1 Seed 4 err_b	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 162: PD/PFA Results for 2-Dimensional KS Test, Single-Qubit, Part IX

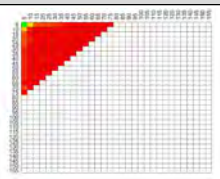
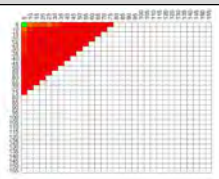
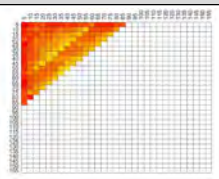
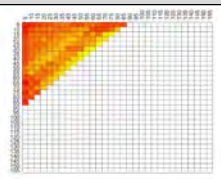
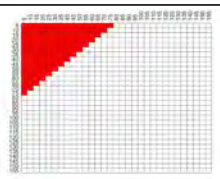
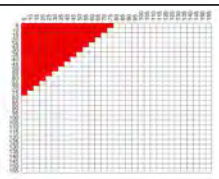
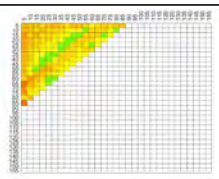
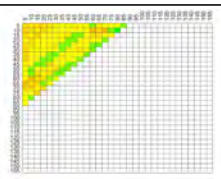
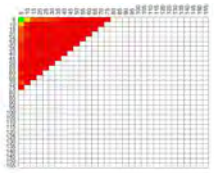
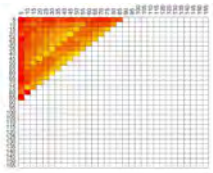
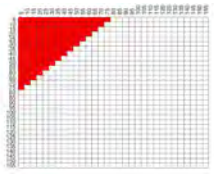
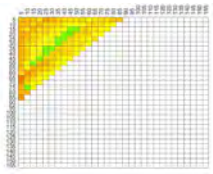
	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	RB1 Seed 4 b			
	RB1 Seed 4 err_a	RB1 Seed 4 err_α	RB1 Seed 4 err_a	RB1 Seed 4 err_α
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 err_b		RB1 Seed 4 err_b	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 163: PD/PFA Results for 2-Dimensional KS Test, Single-Qubit, Part X

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	RB1 Seed 4 err_a			
		RB1 Seed 4 err_α		RB1 Seed 4 err_α

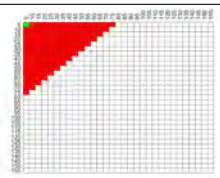
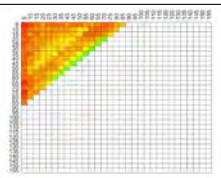
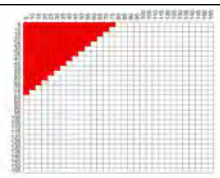
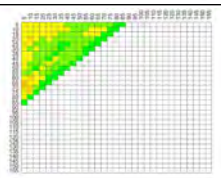
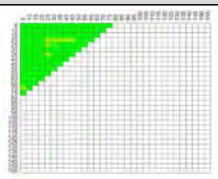
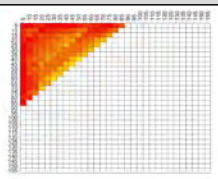
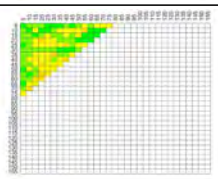
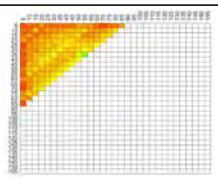
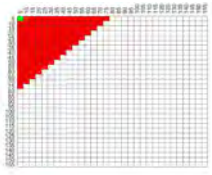
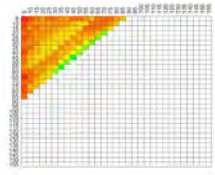
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB1 Seed 4 *err_b*		RB1 Seed 4 *err_b*	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 164: PD/PFA Results for 2-Dimensional KS Test, Single-Qubit, Part XI

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	RB1 Seed 4 *err_α*			
	RB1 Seed 4 *err_b*		RB1 Seed 4 *err_b*	
$\alpha = 0.001$				

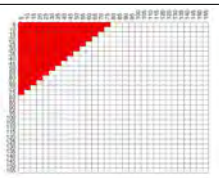
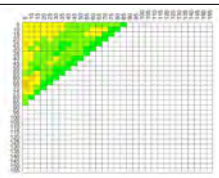
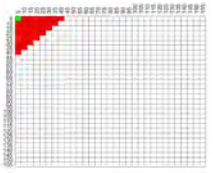
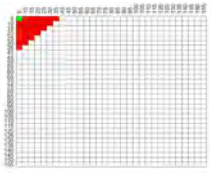
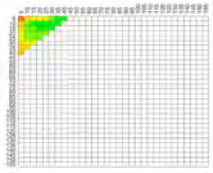
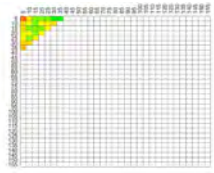
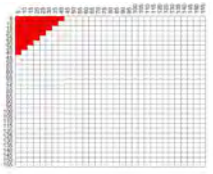
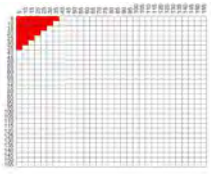
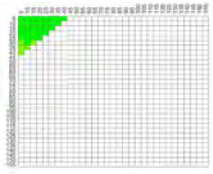
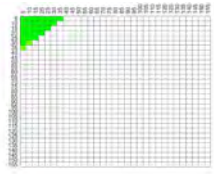
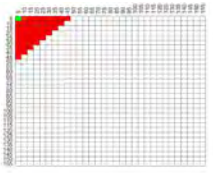
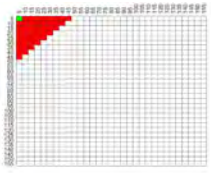
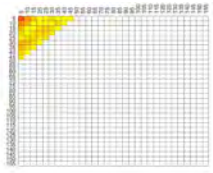
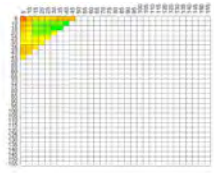
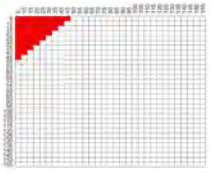
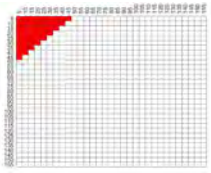
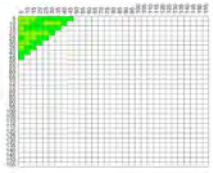
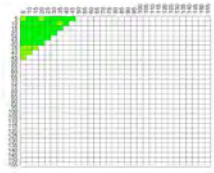
$\alpha = 0.05$				

Table 165: PD/PFA Results for 2-Dimensional KS Test, Single-Qubit, Part XII

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	Multi-Qubit Gate Error			
	QPTCX	QPTSWAP	QPTCX	QPTSWAP
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB2 Seed 4 Fid.	RB2 Seed 4 a	RB2 Seed 4 Fid.	RB2 Seed 4 a
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB2 Seed 4 α	RB2 Seed 4 b	RB2 Seed 4 α	RB2 Seed 4 b

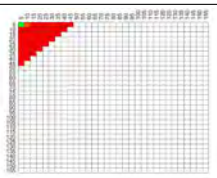
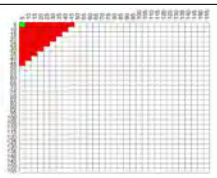
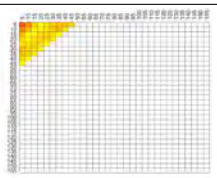
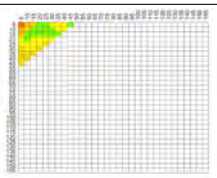
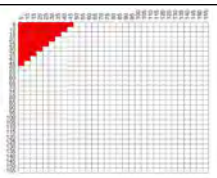
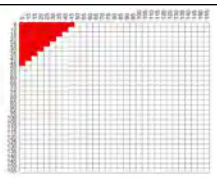
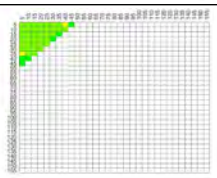
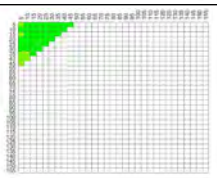
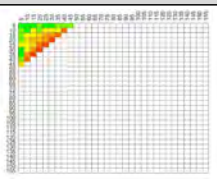
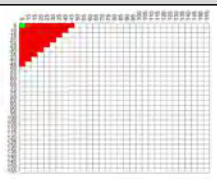
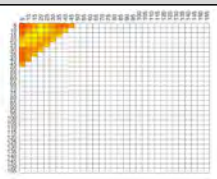
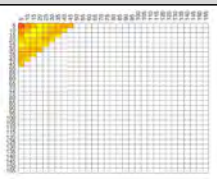
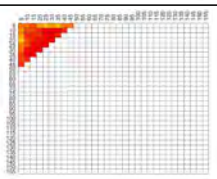
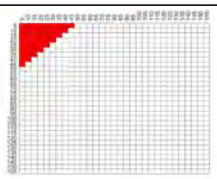
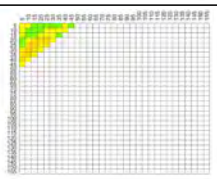
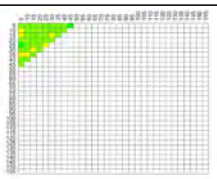
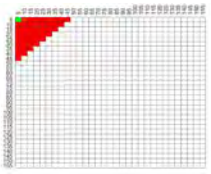
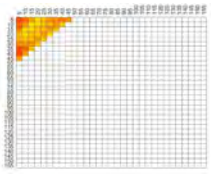
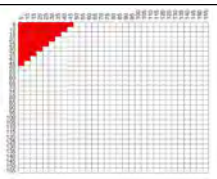
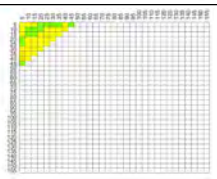
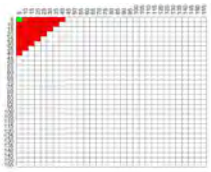
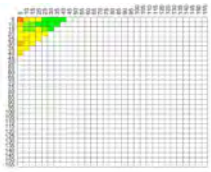
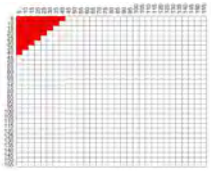
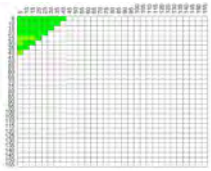
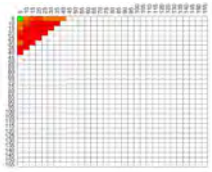
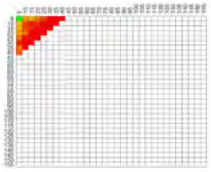
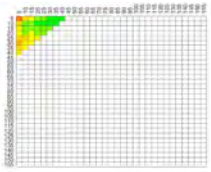
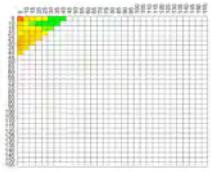
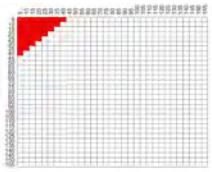
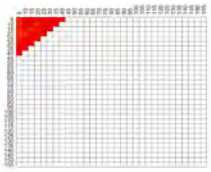
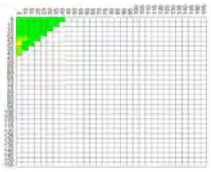
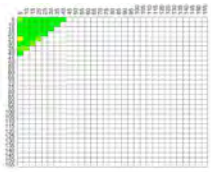
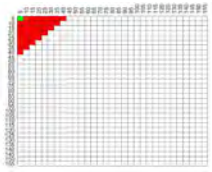
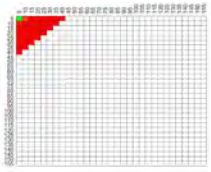
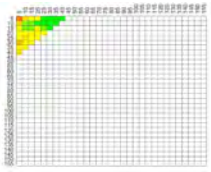
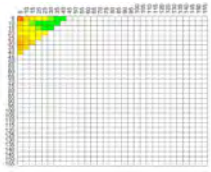
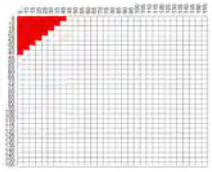
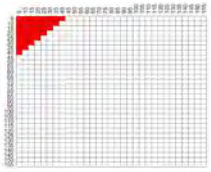
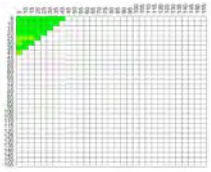
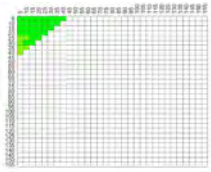
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB2 Seed 4 err_a	RB2 Seed 4 err_α	RB2 Seed 4 err_a	RB2 Seed 4 err_α
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB2 Seed 4 err_b		RB2 Seed 4 err_b	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 166: PD/PFA Results for 2-Dimensional KS Test, Two-Qubit, Part I

	PFA	PD
	Color Scale for PFA Tests	Color Scale for PD Tests
		

	QPTCX			
		QPTSWAP		QPTSWAP
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB2 Seed 4 Fid.	RB2 Seed 4 a	RB2 Seed 4 Fid.	RB2 Seed 4 a
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB2 Seed 4 α	RB2 Seed 4 b	RB2 Seed 4 α	RB2 Seed 4 b
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB2 Seed 4 err_a	RB2 Seed 4 err_α	RB2 Seed 4 err_a	RB2 Seed 4 err_α

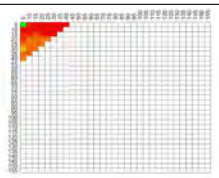
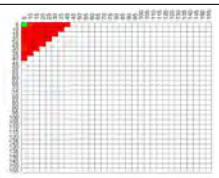
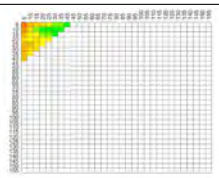
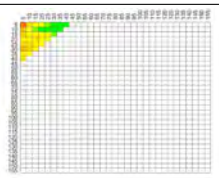
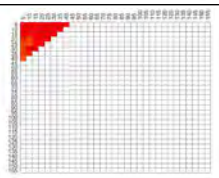
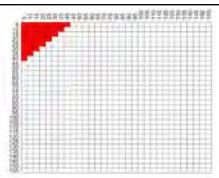
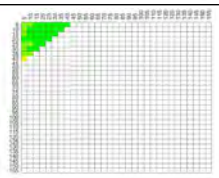
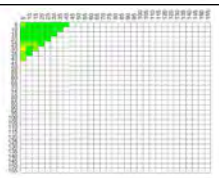
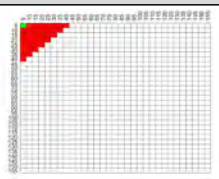
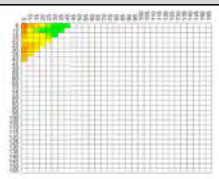
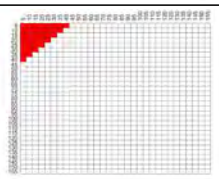
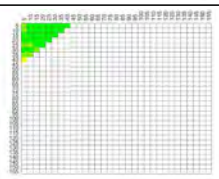
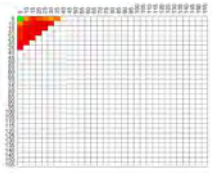
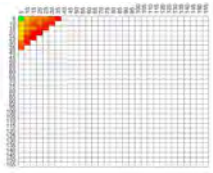
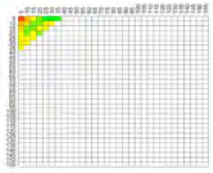
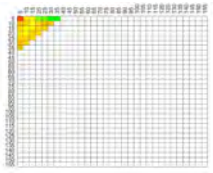
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB2 Seed 4 *err_b*		RB2 Seed 4 *err_b*	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 167: PD/PFA Results for 2-Dimensional KS Test, Two-Qubit, Part II

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	QPTSWAP			
	RB2 Seed 4 Fid.	RB2 Seed 4 a	RB2 Seed 4 Fid.	RB2 Seed 4 a
$\alpha = 0.001$				

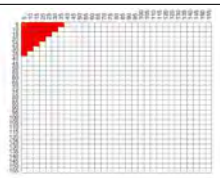
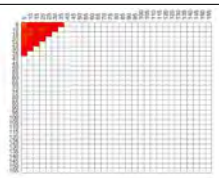
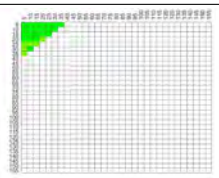
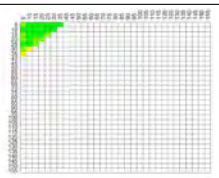
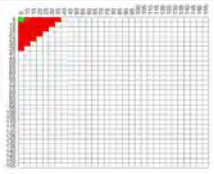
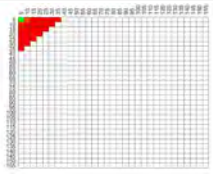
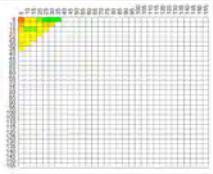
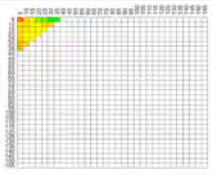
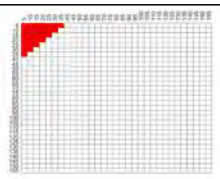
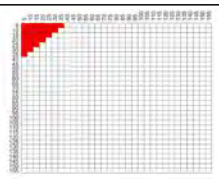
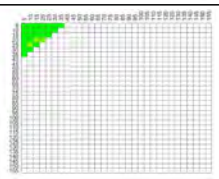
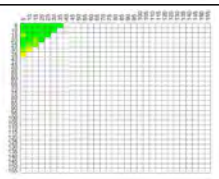
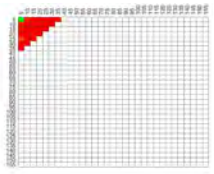
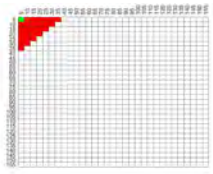
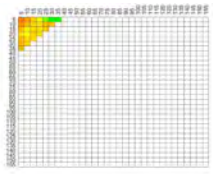
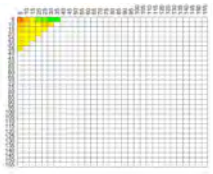
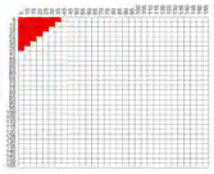
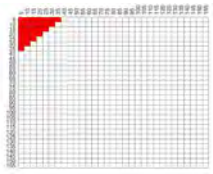
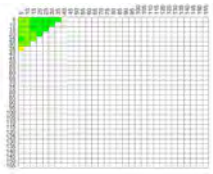
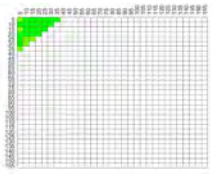
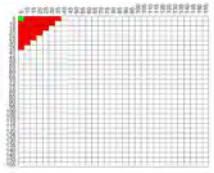
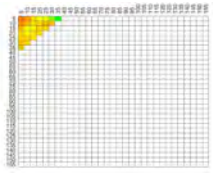
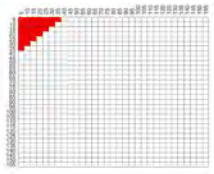
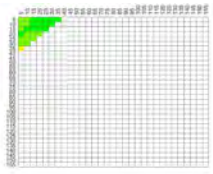
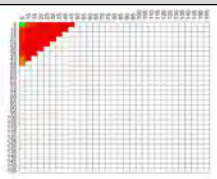
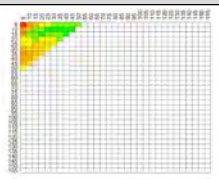
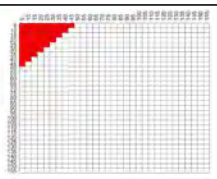
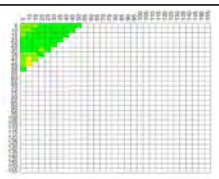
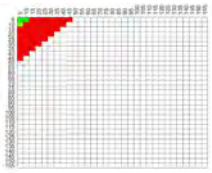
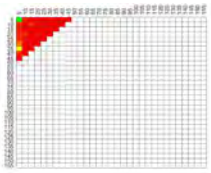
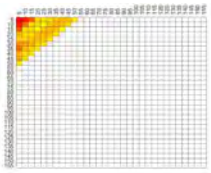
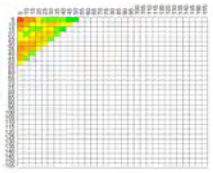
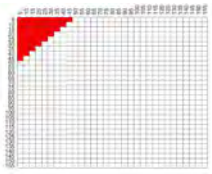
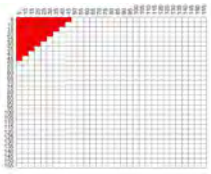
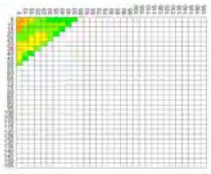
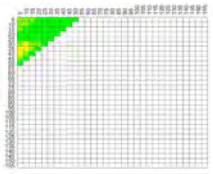
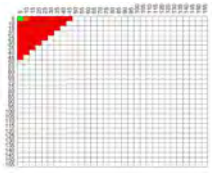
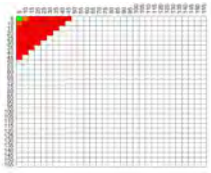
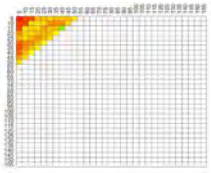
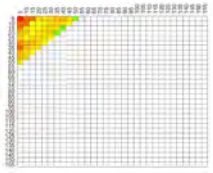
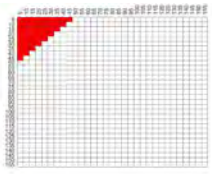
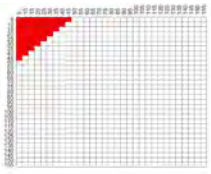
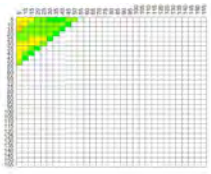
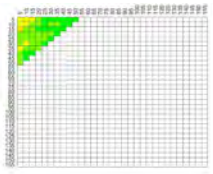
$\alpha = 0.05$				
	RB2 Seed 4 α	RB2 Seed 4 b	RB2 Seed 4 α	RB2 Seed 4 b
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB2 Seed 4 err_a	RB2 Seed 4 err_α	RB2 Seed 4 err_a	RB2 Seed 4 err_α
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB2 Seed 4 err_b		RB2 Seed 4 err_b	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 168: PD/PFA Results for 2-Dimensional KS Test, Two-Qubit, Part III

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	RB2 Seed 4 Fidelity			
		RB2 Seed 4 a		RB2 Seed 4 a
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB2 Seed 4 α	RB2 Seed 4 b	RB2 Seed 4 α	RB2 Seed 4 b
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB2 Seed 4 *err_a*	RB2 Seed 4 *err_α*	RB2 Seed 4 *err_a*	RB2 Seed 4 *err_α*
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB2 Seed 4 *err_b*		RB2 Seed 4 *err_b*	

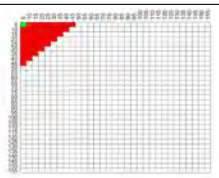
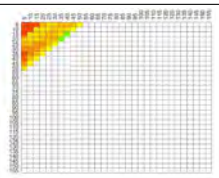
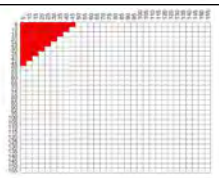
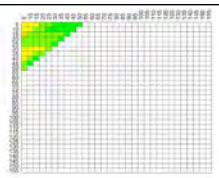
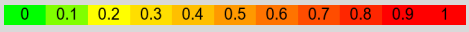
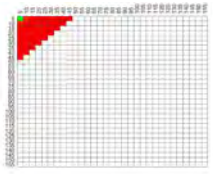
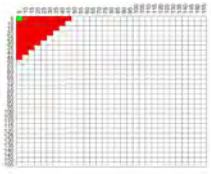
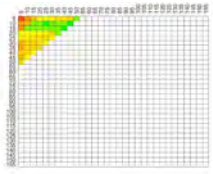
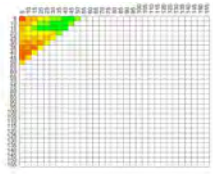
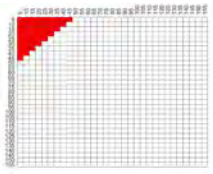
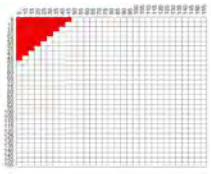
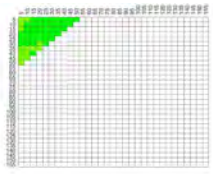
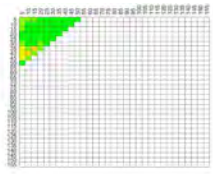
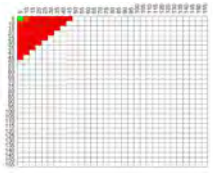
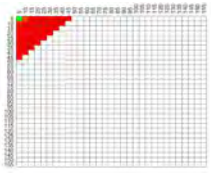
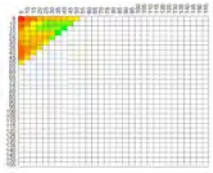
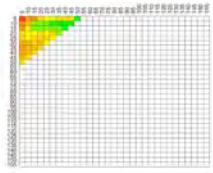
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 169: PD/PFA Results for 2-Dimensional KS Test, Two-Qubit, Part IV

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	RB2 Seed 4 a			
	RB2 Seed 4 α	RB2 Seed 4 b	RB2 Seed 4 α	RB2 Seed 4 b
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB2 Seed 4 err_a	RB2 Seed 4 err_α	RB2 Seed 4 err_a	RB2 Seed 4 err_α
$\alpha = 0.001$				

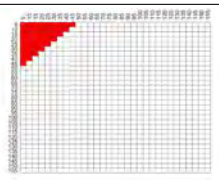
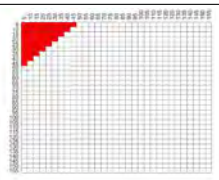
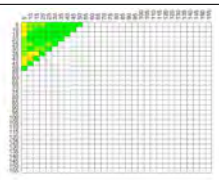
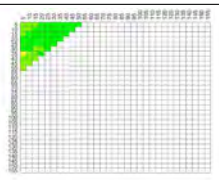
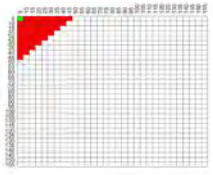
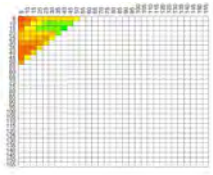
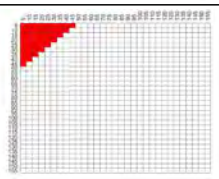
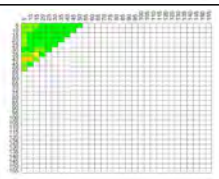
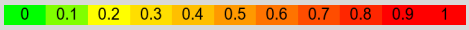
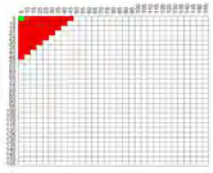
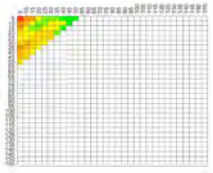
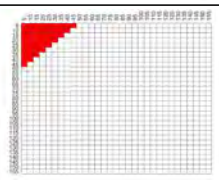
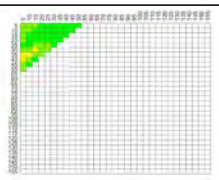
$\alpha = 0.05$				
	RB2 Seed 4 *err_b*		RB2 Seed 4 *err_b*	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 170: PD/PFA Results for 2-Dimensional KS Test, Two-Qubit, Part V

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	RB2 Seed 4 α			
		RB2 Seed 4 *b*		RB2 Seed 4 *b*
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB2 Seed 4 *err_a*	RB2 Seed 4 *err_α*	RB2 Seed 4 *err_a*	RB2 Seed 4 *err_α*

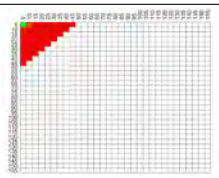
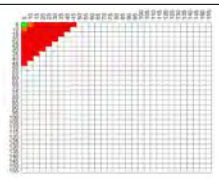
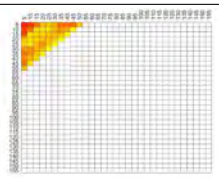
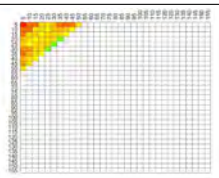
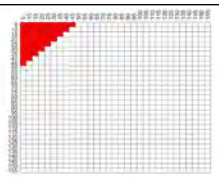
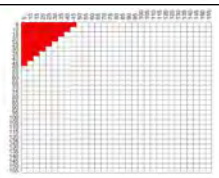
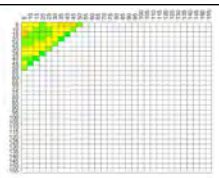
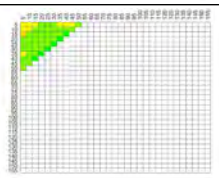
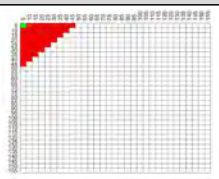
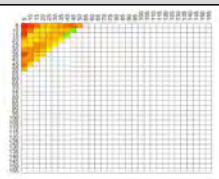
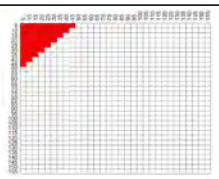
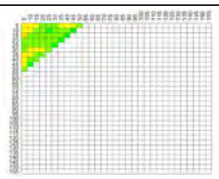
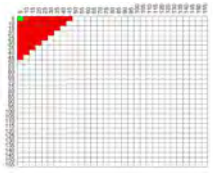
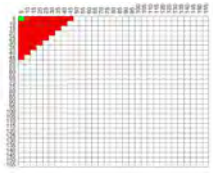
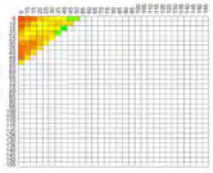
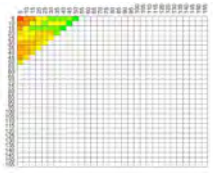
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB2 Seed 4 *err_b*		RB2 Seed 4 *err_b*	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 171: PD/PFA Results for 2-Dimensional KS Test, Two-Qubit, Part VI

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	RB2 Seed 4 b			
	RB2 Seed 4 *err_a*	RB2 Seed 4 *err_α*	RB2 Seed 4 *err_a*	RB2 Seed 4 *err_α*
$\alpha = 0.001$				

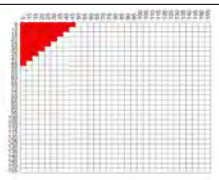
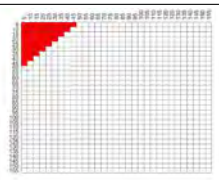
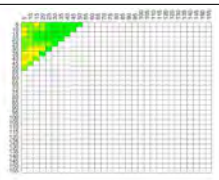
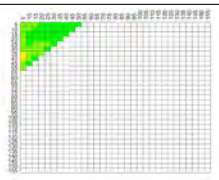
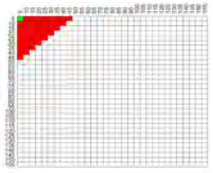
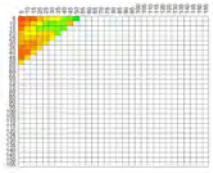
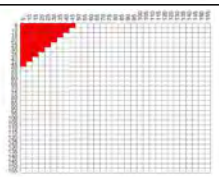
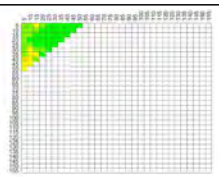
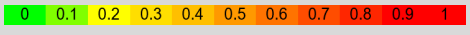
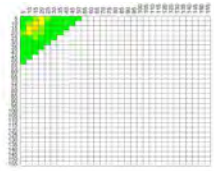
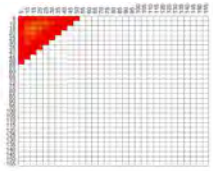
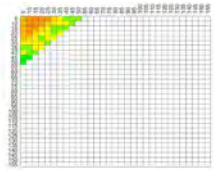
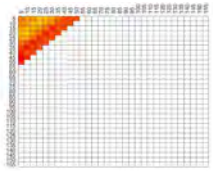
$\alpha = 0.05$				
	RB2 Seed 4 *err_b*		RB2 Seed 4 *err_b*	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 172: PD/PFA Results for 2-Dimensional KS Test, Two-Qubit, Part VII

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	RB2 Seed 4 *err_a*			
		RB2 Seed 4 *err_α*		RB2 Seed 4 *err_α*
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB2 Seed 4 *err_b*		RB2 Seed 4 *err_b*	

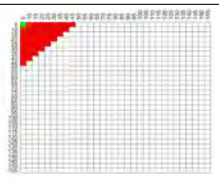
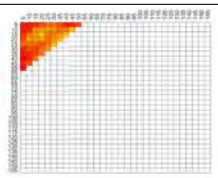
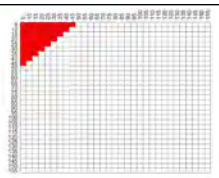
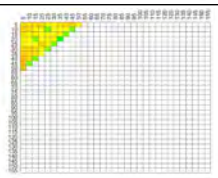
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 173: PD/PFA Results for 2-Dimensional KS Test, Two-Qubit, Part VIII

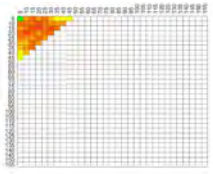
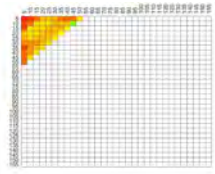
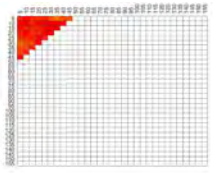
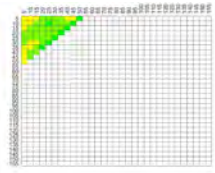
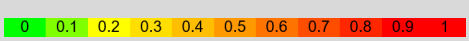
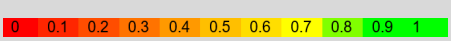
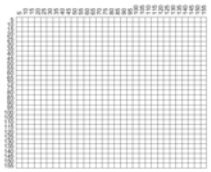
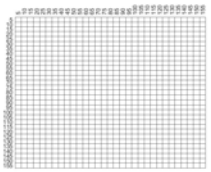
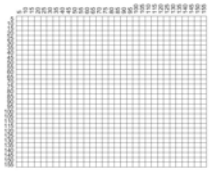
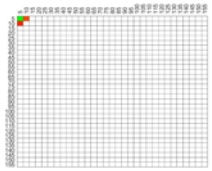
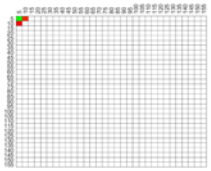
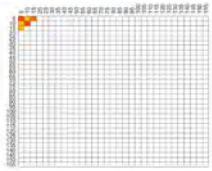
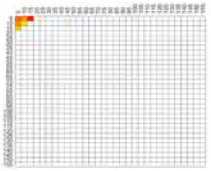
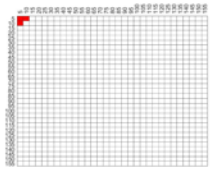
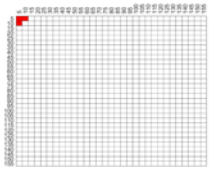
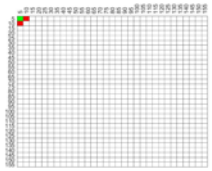
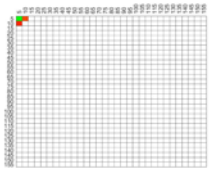
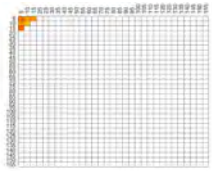
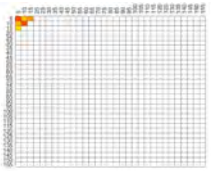
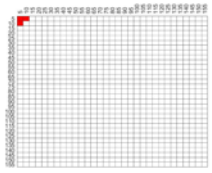
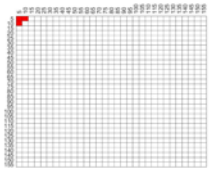
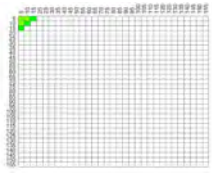
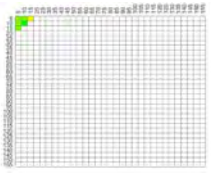
	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	RB2 Seed 4 err_α			
	RB2 Seed 4 err_b		RB2 Seed 4 err_b	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 174: PD/PFA Results for 2-Dimensional KS Test, Two-Qubit, Part IX

	PFA	PD
	Color Scale for PFA Tests	Color Scale for PD Tests
		
	QPTCCX	

		QPTCSWAP[4]		QPTCSWAP
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB3 Seed 4 Fid.	RB3 Seed 4 a	RB3 Seed 4 Fid.	RB3 Seed 4 a
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB3 Seed 4 α	RB3 Seed 4 b	RB3 Seed 4 α	RB3 Seed 4 b
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB3 Seed 4 err_a	RB3 Seed 4 err_α	RB3 Seed 4 err_a	RB3 Seed 4 err_α

[4]Not enough data points were collected to have a meaningful analysis for QPTCSWAP.

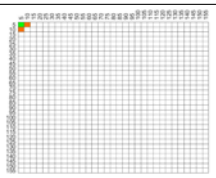
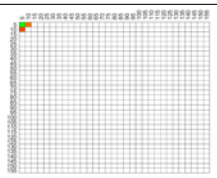
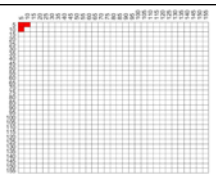
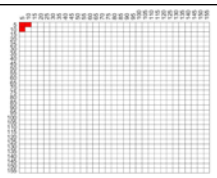
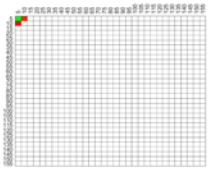
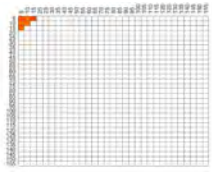
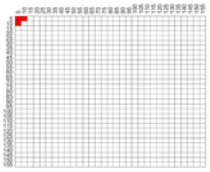
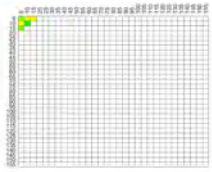
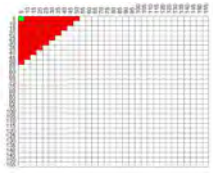
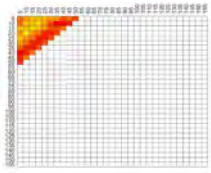
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB3 Seed 4 *err_b*		RB3 Seed 4 *err_b*	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 175: PD/PFA Results for 2-Dimensional KS Test, Three-Qubit, Part I

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	RB3 Seed 4 Fidelity			
		RB3 Seed 4 a		RB3 Seed 4 a
$\alpha = 0.001$				

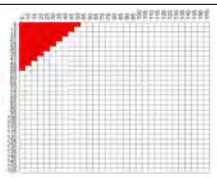
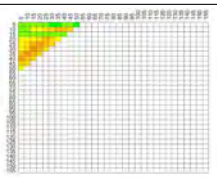
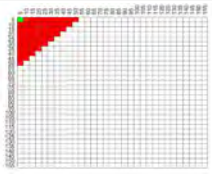
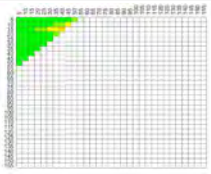
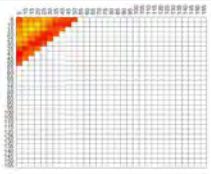
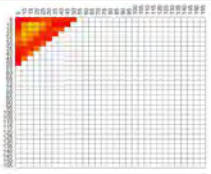
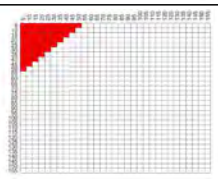
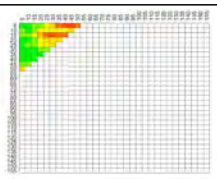
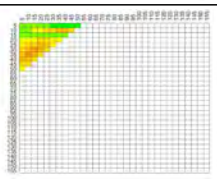
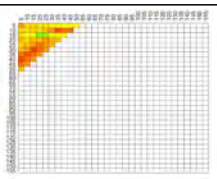
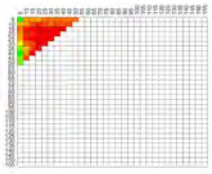
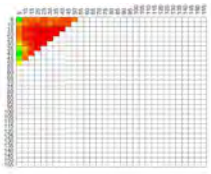
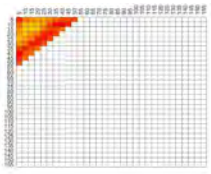
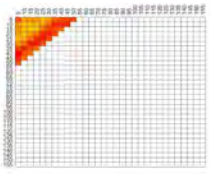
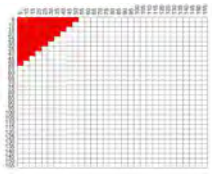
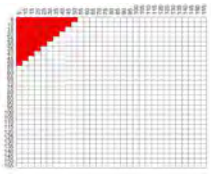
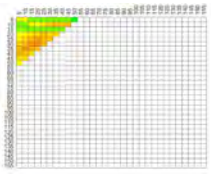
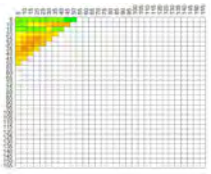
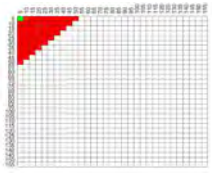
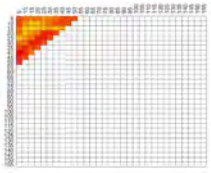
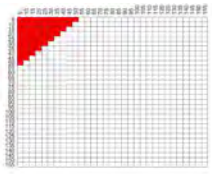
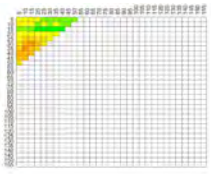
$\alpha = 0.05$				
	RB3 Seed 4 α	RB3 Seed 4 b	RB3 Seed 4 α	RB3 Seed 4 b
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB3 Seed 4 err_a	RB3 Seed 4 err_α	RB3 Seed 4 err_a	RB3 Seed 4 err_α
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB3 Seed 4 err_b		RB3 Seed 4 err_b	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 176: PD/PFA Results for 2-Dimensional KS Test, Three-Qubit, Part II

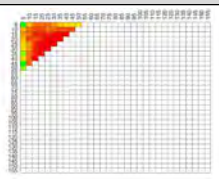
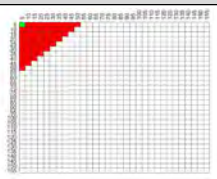
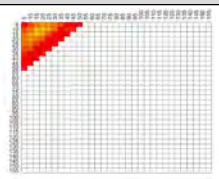
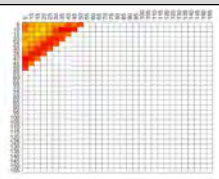
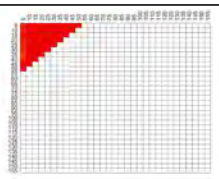
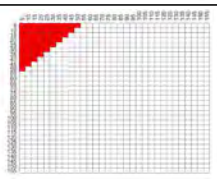
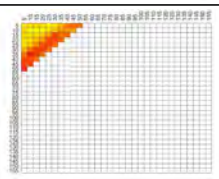
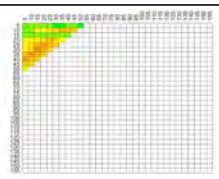
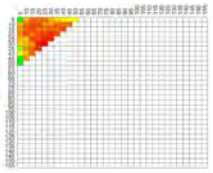
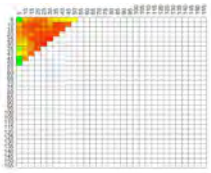
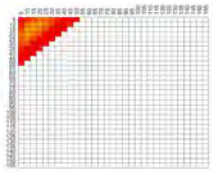
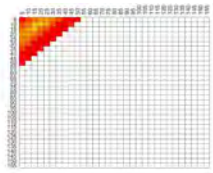
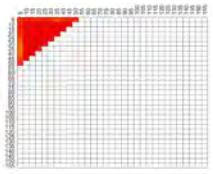
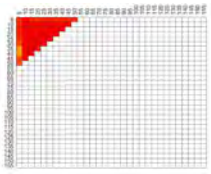
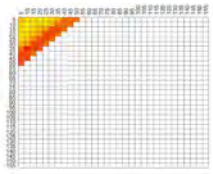
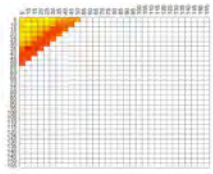
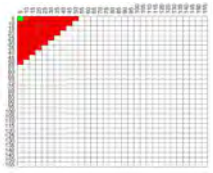
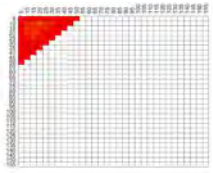
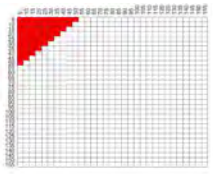
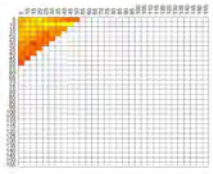
	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	RB3 Seed 4 a			
	RB3 Seed 4 α	RB3 Seed 4 b	RB3 Seed 4 α	RB3 Seed 4 b
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB3 Seed 4 *err_a*	RB3 Seed 4 *err_*α	RB3 Seed 4 *err_a*	RB3 Seed 4 *err_*α
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB3 Seed 4 *err_b*		RB3 Seed 4 *err_b*	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 177: PD/PFA Results for 2-Dimensional KS Test, Three-Qubit, Part III

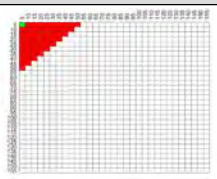
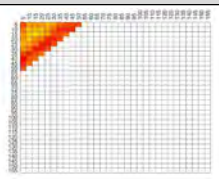
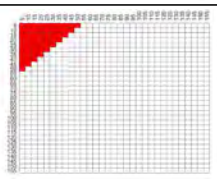
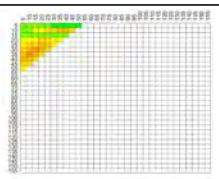
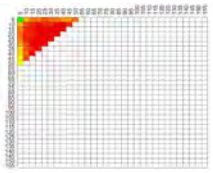
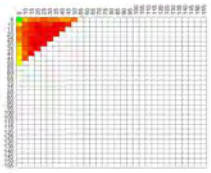
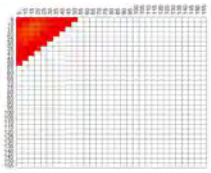
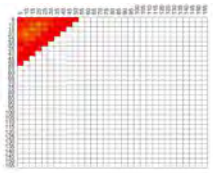
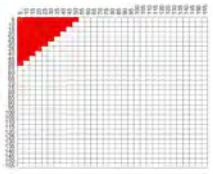
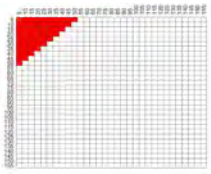
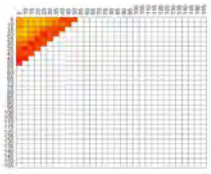
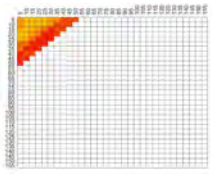
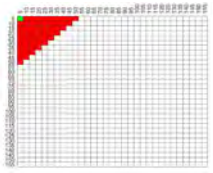
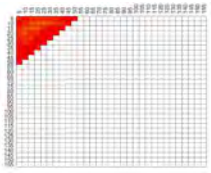
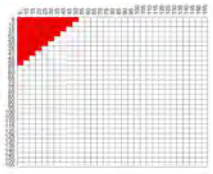
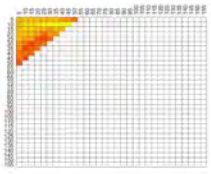
	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	RB3 Seed 4 α			
		RB3 Seed 4 b		RB3 Seed 4 b
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB3 Seed 4 *err_a*	RB3 Seed 4 *err_α*	RB3 Seed 4 *err_a*	RB3 Seed 4 *err_α*
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB3 Seed 4 *err_b*		RB3 Seed 4 *err_b*	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 178: PD/PFA Results for 2-Dimensional KS Test, Three-Qubit, Part IV

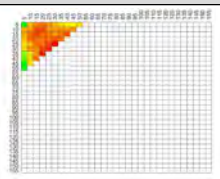
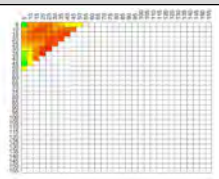
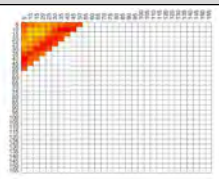
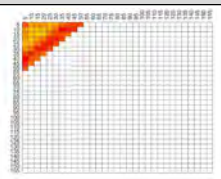
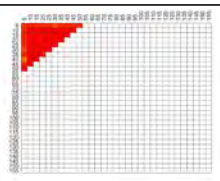
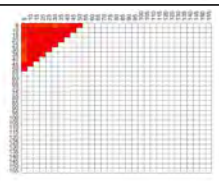
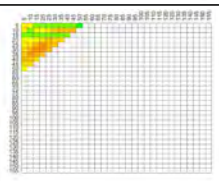
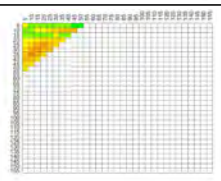
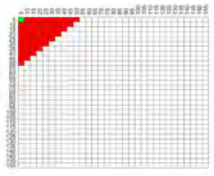
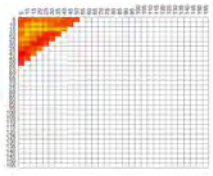
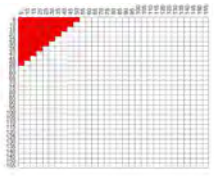
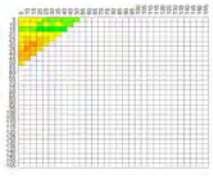
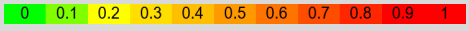
	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	RB3 Seed 4 b			
	RB3 Seed 4 *err_a*	RB3 Seed 4 *err_α*	RB3 Seed 4 *err_a*	RB3 Seed 4 *err_α*
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB3 Seed 4 *err_b*		RB3 Seed 4 *err_b*	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 179: PD/PFA Results for 2-Dimensional KS Test, Three-Qubit, Part V

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	RB3 Seed 4 *err_a*			
		RB3 Seed 4 *err_α*		RB3 Seed 4 *err_α*

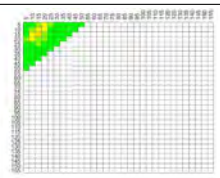
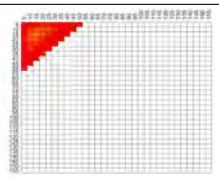
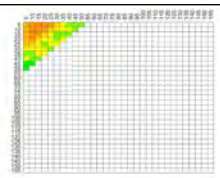
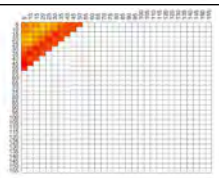
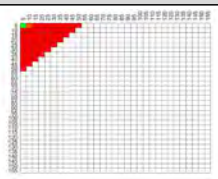
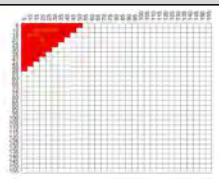
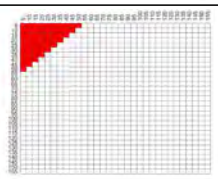
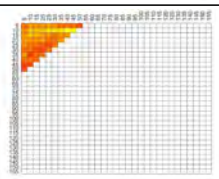
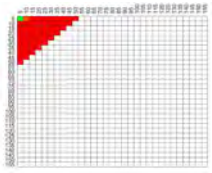
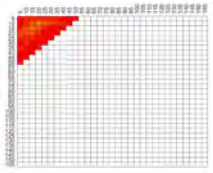
$\alpha = 0.001$				
$\alpha = 0.05$				
	RB3 Seed 4 *err_b*		RB3 Seed 4 *err_b*	
$\alpha = 0.001$				
$\alpha = 0.05$				

Table 180: PD/PFA Results for 2-Dimensional KS Test, Three-Qubit, Part VI

	PFA		PD	
	Color Scale for PFA Tests		Color Scale for PD Tests	
				
	RB3 Seed 4 *err_α*			
	RB3 Seed 4 *err_b*		RB3 Seed 4 *err_b*	
$\alpha = 0.001$				

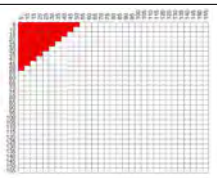
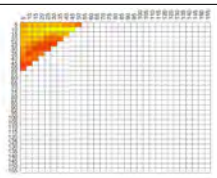
α = 0.05				

Table 181: PD/PFA Results for 2-Dimensional KS Test, Three-Qubit, Part VII

4.4 Summary

This chapter presents an analysis of the relationships between a subset of hardware verification/security techniques and a subset of quantum computing technologies in Part A. In Part B1, best-fit distribution types are determined for each of the physically unclonable characteristic (PUC) characteristics, where the main conclusion is that none of the characteristics are normally distributed. In Part B2, sampling error rates for the calibration data characteristics are evaluated and successfully mitigated with data set truncation combined with removing multi-qubit gate error rate data points equal to one. In Part B3, multiple qubit discrimination tests are evaluated for each of the characteristics. Best results are achieved for the non-parametric tests, with refining of the JSD and the two-dimensional KS test recommended for future development.

V. Conclusions

The goal of the work presented in this dissertation document is to apply the concepts of hardware verification and security to quantum computing systems as part of ensuring quantum computing systems' trustworthiness. Part A of the research surveys the state-of-the-art in current classical hardware verification and security techniques and emerging quantum computing technologies, and it also evaluates a subset of relationships between hardware verification/security techniques and emerging quantum computing technologies. Part B of the research focuses on the development of physically unclonable characteristics (PUCs) (physically unclonable functions (PUFs) applied to transmon qubits).

5.1 Findings for Literature Review and Research Questions

It is helpful to understand the accomplishments attained by the research reported in this dissertation by providing findings related to the literature review and research questions introduced in Chapter I.

- Literature Review Question A-1: What are current state-of-the-art hardware verification and security methods?
    - Findings: Current state-of-the-art hardware verification methods include physical inspection techniques (incoming inspection, exterior tests, interior tests, material analysis, connectivity analysis, and miscellaneous), electrical inspection techniques (parametric tests, functional tests, structural tests, and burn-in test), and identification techniques (chip ID and package ID) (Section 2.1.1). Current state-of-the-art security methods include reverse engineering countermeasures (hardware obfuscation, optical dissimulation, and self-destruction)

and side channel attacks (SCA) countermeasures (uniform timing for all operations, power line filtering, power supply shielding, EM shielding, transistor- and gate-level secure logic, dynamic reconfiguration, temporal desynchronization, and jamming) (Section 2.1.2).

- Literature Review Question A-2: What are the current state-of-the-art quantum computing implementations?

  - Findings: Current state-of-the-art quantum computing implementations include diamond vacancy, ion trap, quantum ring, silicon spin qubit, superconducting qubit (transmon), and topological qubit (Section 2.1.3). At present, most industry progress has been achieved on developing transmon qubits.

- Research Question A-1: How might the techniques identified in Literature Review Question A-1 be applied to verification and security of the quantum computing systems identified in Literature Review Question A-2?

  - Findings: First, a framework is developed in this research to evaluate applicability of classical hardware verification and security techniques to quantum computing systems (Section 3.2.3). Next, a subset of techniques is evaluated against the silicon spin and transmon qubits using this framework (Section 4.2.3).

- Literature Review Question B-1: What characteristics are used to describe individual qubits for transmon processors? What characteristics are used to describe sets of qubits for transmon processors?

  - Findings: The characteristics of T1 and T2 coherence times, single-qubit gate error rates, readout error rates, and response curves describe individual qubits (Sections 2.2.1.1, 2.2.1.2, 2.2.1.3, and 5.6). Quantum process tomography

can describe single-qubits and sets of qubits (Section 2.2.1.5). Multi-qubit gate error rates describe sets of two qubits (Section 2.2.1.2). Magnetic flux crosstalk can describe sets of two or more qubits (Section 2.2.1.7). Randomized benchmarking can describe a single qubit up to an entire processor (Section 2.2.1.6).

- Literature Review Question B-2: What techniques exist to perform qubit discrimination?

  - Findings: Techniques that can potentially perform qubit discrimination include the t-test, one-dimensional and multi-dimensional Kolmogorov-Smirnov (KS) test, Jensen-Shannon divergence (JSD) test, quadratic discriminant analysis (QDA) test, and multiple discriminant analysis with maximum likelihood (MDA-ML) test (Section 4.3.3.1).

- Research Question B-1: How can the underlying concept of the PUF be applied to a transmon-based quantum processor?

  - Findings: While PUFs do not directly apply to the transmon processor, they follow a cause-and-effect paradigm that directly translates to the transmon processor (Section 1.2).

- Research Question B-2: Which characteristics identified in Literature Review Question B-1 are best suited for PUCs?

  - Findings: It is not currently possible to determine which of the characteristics are best suited for PUCs. The qubit discrimination methods need more refining before making a determination (Section 4.3.3).

- Research Question B-3: Which techniques identified in Literature Review Question B-2 are most effective qubit discrimination methods for PUCs?

- Findings: Presently, the one-dimensional and multi-dimensional KS tests and the JSD test show the greatest potential for PUCs since they are non-parametric methods. More development for the multi-dimensional KS test and the JSD test are required before determining the most effective method (Section 4.3.3).

- Research Question B-4: What obstacles exist in the development of PUCs and how can they be overcome?

  - Findings: A major obstacle in the development of PUCs arose from sampling errors. Of the data conditioning methods tested to reduce sampling errors, results show that data set truncation eliminates most sampling errors when the sampling errors are determined by the KS test with $\alpha = 0.001$. For this reason, there are no current plans to pursue data conditioning further (Section 4.3.2.5).

5.2 Research Contributions

The primary contributions of the research detailed in this document are:

1. A foundation for exploring the relationships (applications and limitations) between classical hardware verification and security techniques and QC systems with differing qubit implementations (Part A)

2. A proposal of radio-frequency (RF) distinct native attributes (DNA) (RF-DNA) for future application to transmon-based processors (Part A)

3. A proposal of physically unclonable characteristics (PUCs) for future application to transmon-based processors (Part B)

4. An evaluation of different qubit characteristics, discrimination methods, and data conditioning methods for application to PUCs (Part B).

5.3 Part A – Conclusion

This dissertation document reports a survey of the state-of-the-art in current classical hardware verification and security techniques and emerging quantum computing technologies. Additionally, a framework for evaluating the applications and limitations of these techniques to quantum computing technologies is proposed and demonstrated for a subset of techniques and quantum technologies. Emerging from this work are the discoveries that the concept of RF-DNA can be applied to transmon processors, and that the concept of PUFs can be applied to transmon processors in a method called PUCs. This method of applying PUFs to transmon processors is developed in Part B.

5.4 Part A – Future work

The expansive breadth of classical hardware verification and security methods make it so that the application of these methods to quantum computers requires an entire community of researchers. Future work should involve evaluating relationships between the full set of HW verification and security techniques listed in Sections 2.1.1 and 2.1.2 and the qubit implementations listed in Section 2.1.3. In particular, RF-DNA fingerprinting for quantum processors should be developed. A few steps towards evaluating the relationships between the full set of HW verification and security techniques are identified in Chapter IV and presented in the following list:

- Determine if all gate operations on silicon spin qubits can function with uniform timing.
- Determine if all gate operations on transmon qubits can function with uniform timing.

- Determine if all gate operations on silicon spin qubits can function with uniform power consumption.
- Determine if all gate operations on transmon qubits can function with uniform power consumption.
- Evaluate the performance and spatial resolution of a probe inside a dilution refrigerator for RF-DNA applied to transmon processors.
- Evaluate the effect of a probe on transmon processor performance for RF-DNA applied to transmon processors.

Lastly, future research should identify any vulnerabilities unique to quantum computing systems and develop new hardware verification and security techniques to prevent the exploitation of these vulnerabilities.

5.5 Part B – Conclusion

Development of PUCs has progressed to include the characteristics of calibration data (T1 and T2 coherence times and single-qubit, multi-qubit, and readout error), quantum process tomography metrics (for the X, H, CX, SWAP, CCX, and CSWAP gates), and randomized benchmarking metrics (1-qubit, 2-qubit, and 3-qubit metrics, with one seed of data and with five seeds of data). Furthermore, development has included the t-test, one-dimensional KS test, two-dimensional KS test, JSD, and QDA as discrimination methods.

In Part B1, best-fit distribution types are determined for each of the PUC characteristics. The best-fit distribution types vary across each of the characteristics, with individual characteristics often having more than one best-fit distribution. An initial analysis is performed for each characteristic to understand the results and identify interesting areas requiring additional research and analysis. The main goal of this

section is to determine whether the characteristics are normally distributed. Thus, the most important conclusion from the results is that none of the characteristics are normally distributed.

In Part B2, sampling error rates for the calibration data characteristics are evaluated using the t-test/F-test and the KS test, with the KS test unsurprisingly demonstrating better performance for the non-normally distributed data. Causes for sampling errors are identified, and mitigation is explored with multiple data conditioning methods. The most effective mitigation method is data set truncation combined with removing only the outliers of multi-qubit gate error rate data points equal to one. This mitigation method eliminates most sampling errors and is a continued data conditioning method in Part B3.

In Part B3, the t-test, one-dimensional KS test, JSD, QDA, and two-dimensional KS test are tested as qubit discrimination methods for each of the characteristics. Best results overall are observed for the one-dimensional KS test. Implementation of JSD and the two-dimensional KS test requires more refining, and the t-test and QDA methods produce poor results due to the assumption of Gaussian distributions.

5.6 Part B – Future work

A tremendous amount of future work is identified for the development of PUCs. For Part B1, the majority of future work can be categorized into five areas:

1. The new calibration data of anharmonicity, probability of measuring $|0\rangle$ for a qubit prepared as a $|1\rangle$, probability of measuring $|1\rangle$ for a qubit prepared as a $|0\rangle$, and ID error should be collected and evaluated.

2. Most characteristics have more than one best-fit distribution. Additional research should be conducted to determine which qubits or sets of qubits lead to

which types of distributions, and what physical properties are shared by qubits (qubit sets) of common distributions.

3. For several of the randomized benchmarking characteristics, the analogous relationship is largely obscured between the characteristics and the primary type of data modeled by some of the best-fit distributions (for example, failure rates or signal scattering). Future research should give additional considerations to understanding these analogous relationships.

4. Many of the qubits (qubit sets) appear to display bimodal or even multi-modal behavior for some of the characteristics. A hypothesis for the cause of such behavior observed in two- or three-qubit sets is the existence or non-existence of direct physical links between qubits in the sets. Future work should test this hypothesis, and determine if there is a different cause. Similarly, future work should determine the cause for the appearance or disappearance of multi-modal behavior as the number of qubits and number of seeds increases or decreases.

5. Seemingly anomalous data points should be investigated to determine if they are outliers or if they are actually characteristic of qubit set behavior.

For Part B2, the data conditioning method identified is presently sufficient for reducing sampling error rates, and no further research is currently recommended. For Part B3, the KS test should be further developed for quantum process tomography (QPT) and randomized benchmarking (RB) characteristics by experimenting with different significance levels as identified in Section 4.3.3.1. Development for JSD should continue by implementing bin selection rules that do not assume a normal distribution (for example, cross-validation) to improve probability of false alarm (PFA) and probability of detection (PD) results. Future work for the two-dimensional KS test should involve conducting dimensionality reduction to improve results for characteristic pairs

that demonstrate overfitting. For the characteristic pairs that demonstrate underfitting, experimentation with the three-dimensional KS test is recommended for future work. MDA-ML is additionally recommended as a qubit discrimination method to test in PUCs.

There are more recommendations for future work beyond the three sub-parts identified in Part B of this dissertation. First, additional types of characteristics should be tested for suitability in PUCs. For example, magnetic flux cross-talk between qubits, resonance response curves, and Josephson junction I-V characteristics are potential characteristics at operating temperature [9, 44, 152]. Reflected power in transmission lines and Josephson junction resistance [100, 128], are potential room-temperature characteristics. Second, complications arising from real-world implementation of the PUC use case scenario should be considered, and mitigation methods for the complications should be developed. For example, the noise environment for the manufacturer will likely differ from the noise environment for the customer. If possible, developing a noise profile for each and determining how qubit behavior changes in response to the different noise profiles would greatly improve the ability to implement PUCs in a real-world scenario. Third, data collection should take place on dedicated devices (that is, access to the devices is not shared amongst other users). In this way, data can be collected in a more timely manner. Also, with a dedicated device, the number and type of operations conducted on a device can be easily tracked and used for characterizing device aging.

5.7 Heilmeier Catechism

This section concludes the document and summarizes the work with direct responses to the questions posed in the Heilmeier Catechism, a set of questions created by George H. Heilmeier, a former Defense Advanced Research Projects Agency

(DARPA) director (1975-1977), to assist with the evaluation of proposed research programs [35]:

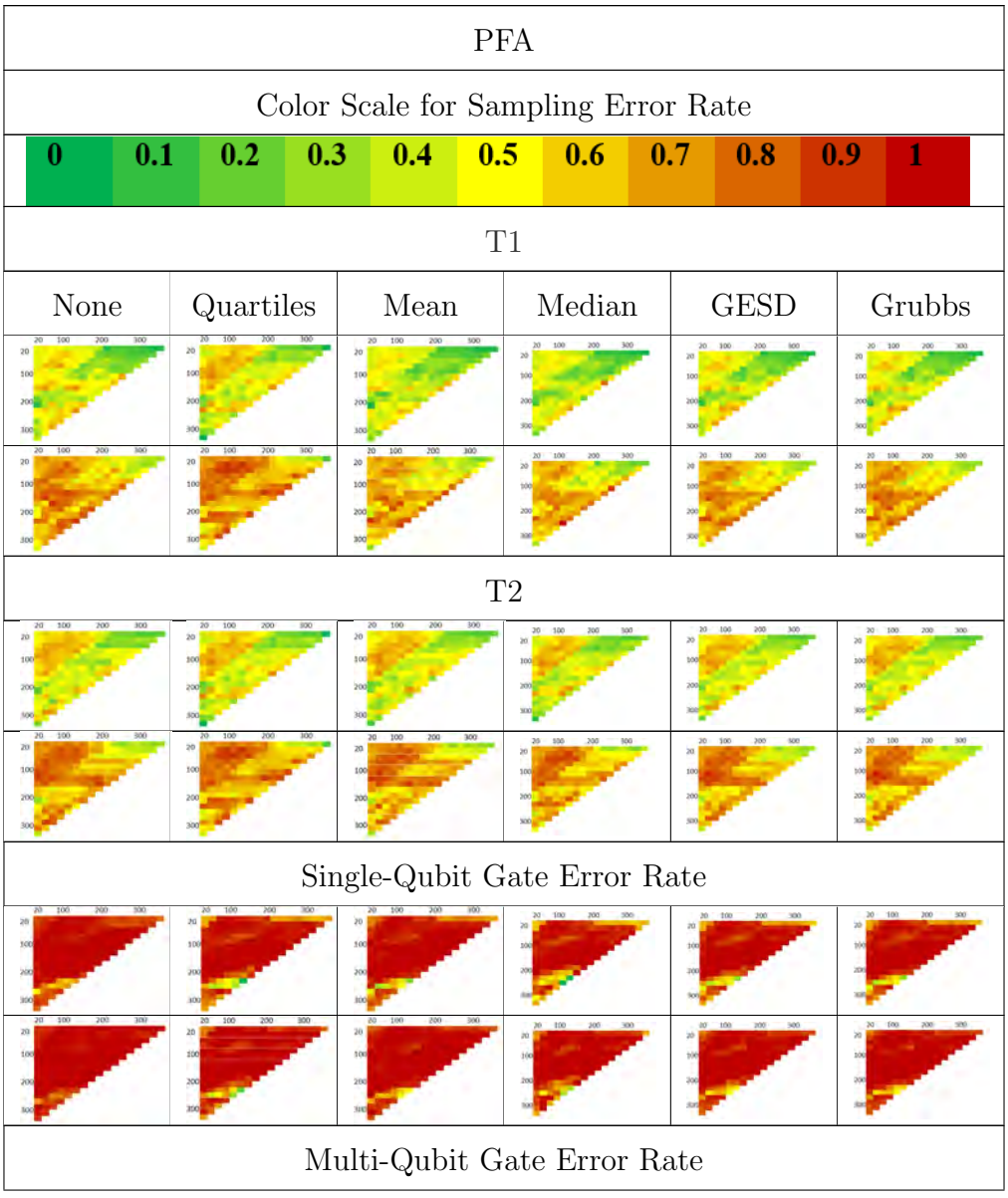
- "What are you trying to do? Articulate your objectives using absolutely no jargon."
  - Response: This work aims to apply the concepts of hardware verification and security to quantum computing systems.
- "How is it done today, and what are the limits of current practice?"
  - Response: To my knowledge, hardware verification and security is not currently implemented for quantum computers, and it is limited to classical devices only.
- "What is new in your approach and why do you think it will be successful?"
  - Response: The work in this document presents a novel framework for considering the application of classical hardware verification and security techniques to quantum computing technologies. Additionally, the development of PUCs is novel. It seems likely that PUCs will succeed due to promising preliminary results for the reduction of sampling errors and due to the large number of different characteristics and qubit discrimination methods available for exploration.
- "Who cares? If you are successful, what difference will it make?"
  - Response: The results of this work are of primary interest to parties who require trustworthy electronics (for example, the Department of Defense) or whose economy will benefit from combating counterfeit electronics. If this work succeeds, it will contribute to the ability of quantum hardware verification and

security development to keep pace with quantum computing system development and to avoid pitfalls experienced with classical computing system development.

- "What are the risks?"
  - Response: The risks for this work lie in the current assumptions and scope limitations. Data gathered from real-world manufacturers and customers may present complications and challenges that may prove difficult to overcome.
- "How much will it cost?"
  - Response: At the current stage of development, all resources required for this research are provided free of cost by the Information Directorate of the Air Force Research Laboratory and the AFRL IBM Q-Hub.
- "How long will it take?"
  - Response: The work detailed within the scope of the dissertation required three years to complete. The broader goal of developing methods for quantum hardware verification and security, however, is an ongoing challenge that will last for several decades to come.
- "What are the mid-term and final 'exams' to check for success?"
  - Response: It is difficult to establish mid-term and final "exams" for verification and security of quantum computers since it is a vast area still largely unexplored by the greater scientific community. However, a potential mid-term goal is to implement at least one quantum verification/security method in a real-world context. The final goal is to implement enough quantum verification/security methods to eliminate counterfeiting and tampering with quantum devices.

Appendix A. Comparison of Outlier Removal Methods on Sampling Error Rates

This section of the appendix presents comparative sampling error rate results for the five outlier removal methods (calculated with the$t$-test/f-test). The five outlier removal methods are: quartiles, mean, median, GESD, and Grubbs. Results show that median outlier removal is the most effective of the five methods for reducing sampling error rates.



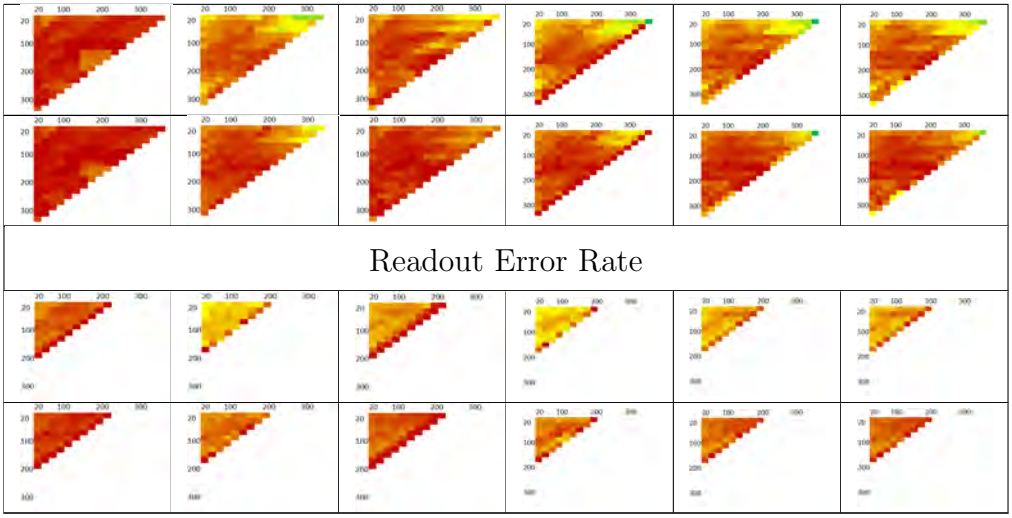
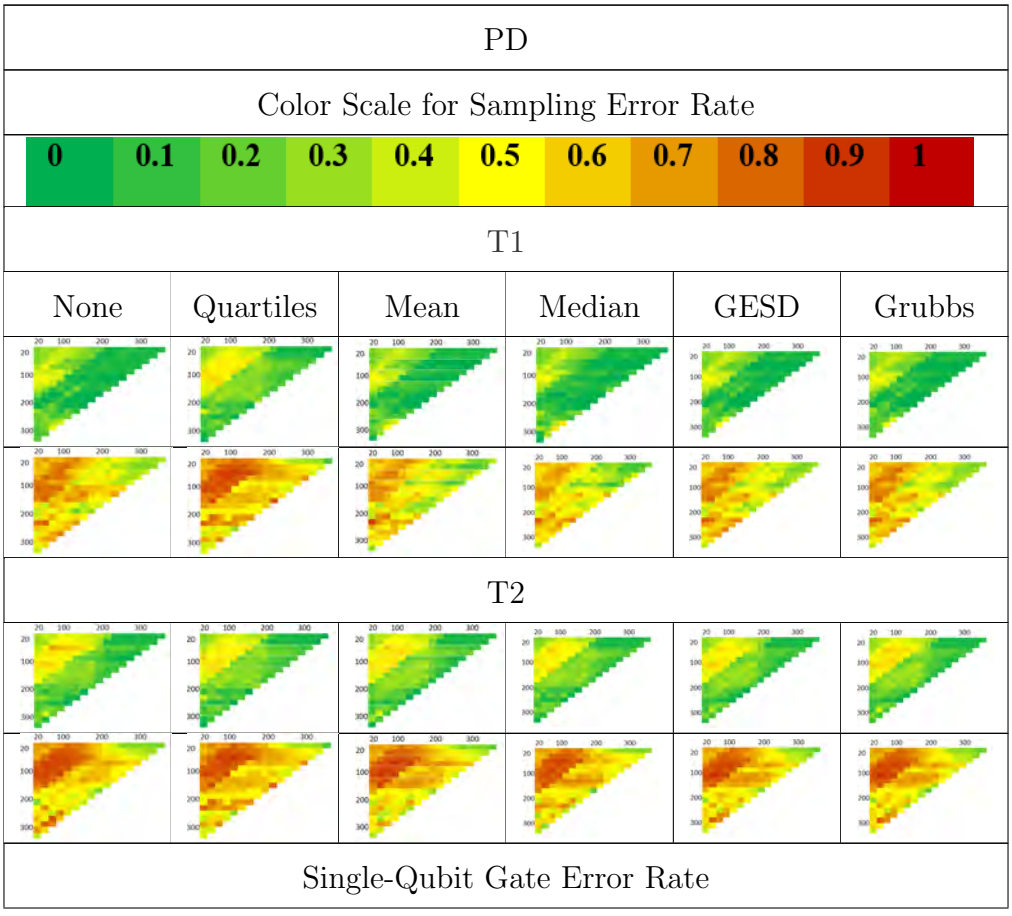


Table 182: Comparison of Outlier Removal Methods on Sampling Error Rates (t-Test/f-Test Only) for PFA Tests



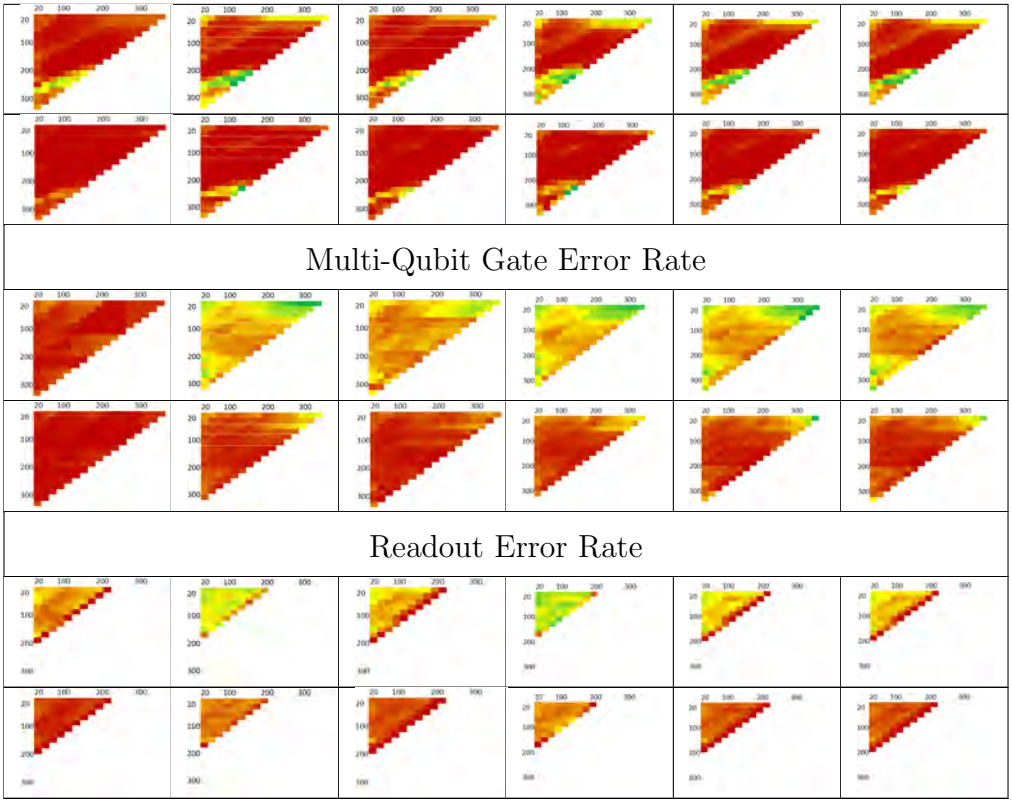


Table 183: Comparison of Outlier Removal Methods on Sampling Error Rates (t-Test/f-Test Only) for PD Tests

Bibliography

1. "1.3.5.15. Chi-Square Goodness-of-Fit Test," in *NIST/SEMATECH e-Handbook of Statistical Methods*, Accessed 1 January 2021. [Online]. Available: https://www.itl.nist.gov/div898/handbook/eda/section3/eda35f.htm

2. "1.3.5.16. Kolmogorov-Smirnov Goodness-of-Fit Test," in *NIST/SEMATECH e-Handbook of Statistical Methods*, Accessed 1 January 2021. [Online]. Available: https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm

3. "8.1.6. What are the basic lifetime distribution models used for non-repairable populations?" in *NIST/SEMATECH e-Handbook of Statistical Methods*, Accessed 1 January 2021. [Online]. Available: https://www.itl.nist.gov/div898/handbook/apr/section1/apr16.htm

4. "8.1.6.3. Extreme value distributions," in *NIST/SEMATECH e-Handbook of Statistical Methods*, Accessed 26 September 2018. [Online]. Available: https://www.itl.nist.gov/div898/handbook/apr/section1/apr163.htm

5. "A Gentle Introduction to Statistical Power and Power Analysis in Python," Accessed 1 May 2020. [Online]. Available: https://machinelearningmastery.com/statistical-power-and-power-analysis-in-python/

6. "Accessing Higher Energy States," Accessed 3 May 2020. [Online]. Available: https://qiskit.org/textbook/ch-quantum-hardware/accessing_higher_energy_states.html

7. "Birnbaum-Saunders Distribution - MATLAB & Simulink," Accessed 7 July 2021. [Online]. Available: https://www.mathworks.com/help/stats/birnbaum-saunders-distribution.html

8. “Burr Type XII Distribution - MATLAB & Simulink,” Accessed 7 July 2021. [Online]. Available: https://www.mathworks.com/help/stats/burr-type-xii-distribution.html

9. “Calibrating Qubits with Qiskit Pulse,” Accessed 25 January 2021. [Online]. Available: https://qiskit.org/textbook/ch-quantum-hardware/calibrating-qubits-pulse.html#3.2-Determining-0-vs-1-

10. “Cluster Analysis,” Accessed 1 May 2020. [Online]. Available: http://www.bici.eu/biss2008/courses/files/saitta/Clustering.pdf

11. “Creating Discriminant Analysis Model - MATLAB & Simulink,” Accessed 1 May 2020. [Online]. Available: https://www.mathworks.com/help/stats/creating-discriminant-analysis-model.html

12. “Extreme Value Distribution - MATLAB & Simulink,” Accessed 26 January 2021. [Online]. Available: https://www.mathworks.com/help/stats/extreme-value-distribution.html

13. “Extreme value probability distribution object - MATLAB,” Accessed 26 January 2021. [Online]. Available: https://www.mathworks.com/help/stats/prob.extremevaluedistribution.html

14. “Fréchet Distribution: Definition, Examples - Statistics How To,” Accessed 26 January 2021. [Online]. Available: https://www.statisticshowto.com/frechet-distribution/

15. “Gamma Distribution - MATLAB & Simulink,” Accessed 12 July 2021. [Online]. Available: https://www.mathworks.com/help/stats/gamma-distribution.html?searchHighlight=gammadistribution&s_tid=srchtitle

16. “Gamma Distribution: Definition, PDF, Finding in Excel - Statistics How To,” Accessed 12 July 2021. [Online]. Available: https://www.statisticshowto.com/gamma-distribution/

17. “Generalized Extreme Value Distribution - MATLAB & Simulink,” Accessed 26 January 2021. [Online]. Available: https://www.mathworks.com/help/stats/generalized-extreme-value-distribution.html

18. “Generalized extreme value probability distribution object - MATLAB,” Accessed 26 January 2021. [Online]. Available: https://www.mathworks.com/help/stats/prob.generalizedextremevaluedistribution.html?searchHighlight=generalizedextremevaluedistribution&s_tid=srchtitle

19. “Generalized Pareto Distribution - MATLAB & Simulink,” Accessed 14 July 2021. [Online]. Available: https://www.mathworks.com/help/stats/generalized-pareto-distribution.html

20. “Histogram bin counts - MATLAB histcounts,” Accessed 23 July 2020. [Online]. Available: https://www.mathworks.com/help/matlab/ref/histcounts.html#buijikm-1-edges

21. “Inverse Gaussian Distribution - MATLAB & Simulink,” Accessed 14 July 2021. [Online]. Available: https://www.mathworks.com/help/stats/inverse-gaussian-distribution.html

22. “Kolmogorov-Smirnov 2-Sample Goodness of Fit Test,” Accessed 11 April 2020. [Online]. Available: https://www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/ks2samp.htm

23. “Larmor precession,” Accessed 11 April 2020. [Online]. Available: http://hyperphysics.phy-astr.gsu.edu/hbase/magnetic/larmor.html

24. "Logistic Distribution - MATLAB & Simulink," Accessed 12 July 2021. [Online]. Available: https://www.mathworks.com/help/stats/logistic-distribution.html?searchHighlight=logistic&s_tid=srchtitle

25. "Loglogistic Distribution - MATLAB & Simulink," Accessed 11 July 2021. [Online]. Available: https://www.mathworks.com/help/stats/loglogistic-distribution.html?searchHighlight=Loglogisticdistribution&s_tid=srchtitle

26. "Lognormal Distribution - MATLAB & Simulink," Accessed 14 July 2021. [Online]. Available: https://www.mathworks.com/help/stats/lognormal-distribution.html

27. "Nakagami Distribution - MATLAB & Simulink," Accessed 12 July 2021. [Online]. Available: https://www.mathworks.com/help/stats/nakagami-distribution.html?searchHighlight=nakagami&s_tid=srchtitle

28. "Negative Loglikelihood Functions - MATLAB & Simulink," Accessed 29 December 2020. [Online]. Available: https://www.mathworks.com/help/stats/negative-loglikelihood-functions.html

29. "Prediction Using Discriminant Analysis Models - MATLAB & Simulink," Accessed 2 May 2020. [Online]. Available: https://www.mathworks.com/help/stats/prediction-using-discriminant-analysis-models.html

30. "Proving Universality," Accessed 10 June 2021. [Online]. Available: https://qiskit.org/textbook/ch-gates/proving-universality.html#big-red

31. "qiskit.ignis.verification.process_tomography_circuits — Qiskit 0.28.0 documentation," Accessed 20 July 2021. [Online]. Available: https://qiskit.org/documentation/stubs/qiskit.ignis.verification.process_tomography_circuits.html

32. "Randomized Benchmarking — Qiskit 0.23.2 documentation," Accessed 19 January 2021. [Online]. Available: https://qiskit.org/documentation/tutorials/noise/4_randomized_benchmarking.html

33. "Rician Distribution - MATLAB & Simulink," Accessed 12 July 2021. [Online]. Available: https://www.mathworks.com/help/stats/rician-distribution.html?searchHighlight=rician&s_tid=srchtitle

34. "t Location-Scale Distribution - MATLAB & Simulink," Accessed 11 July 2021. [Online]. Available: https://www.mathworks.com/help/stats/t-location-scale-distribution.html

35. "The Heilmeier Catechism," Accessed 19 June 2020. [Online]. Available: https://www.darpa.mil/work-with-us/heilmeier-catechism

36. "Weibull Distribution - MATLAB & Simulink," Accessed 26 January 2021. [Online]. Available: https://www.mathworks.com/help/stats/weibull-distribution.html

37. "Weibull probability distribution object - MATLAB," Accessed 26 January 2021. [Online]. Available: https://www.mathworks.com/help/stats/prob.weibulldistribution.html

38. "What does the error rate on the IBMQ website mean? - Quantum Computing Stack Exchange," Accessed 19 January 2021. [Online]. Available: https://quantumcomputing.stackexchange.com/questions/9073/what-does-the-error-rate-on-the-ibmq-website-mean

39. "Want to See Some "Fake" Microelectronics?" in *DARPA MTO Exposition*. NAVSEA Warfare Centers, 2014.

40. "A DARPA Approach to Trusted Microelectronics," pp. 1–2, 2018, Accessed 06 January 2019. [Online]. Available: https://www.darpa.mil/attachments/ExecutiveSummary_FINAL.PDF

41. "IBM Quantum," 2021, Accessed 3 August 2021. [Online]. Available: https://quantum-computing.ibm.com/

42. "Innovare Advancement Center to present International Quantum Information Science Workshop," 2021, Accessed 24 August 2021. [Online]. Available: https://romesentinel.com/stories/innovare-advancement-center-to-present-international-quantum-information-science-workshop,116799

43. B. Abdo, "Rising above the noise: quantum-limited amplifiers empower the readout of IBM Quantum systems — IBM Research Blog," 2020, Accessed 11 April 2020. [Online]. Available: https://www.ibm.com/blogs/research/2020/01/quantum-limited-amplifiers/

44. D. M. Abrams, N. Didier, S. A. Caldwell, B. R. Johnson, and C. A. Ryan, "Methods for Measuring Magnetic Flux Crosstalk between Tunable Transmons," *Physical Review Applied*, vol. 12, no. 6, p. 064022, Dec 2019.

45. H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

46. M. Arapinis, M. Delavar, M. Doosti, and E. Kashefi, "Quantum Physical Unclonable Functions: Possibilities and Impossibilities," 2019, Accessed 24 August 2021. [Online]. Available: http://arxiv.org/abs/1910.02126

47. ——, "Quantum physical unclonable functions: Possibilities and impossibilities," *Quantum*, vol. 5, pp. 1–32, 2021.

48. T. G. Bali, "The generalized extreme value distribution," *Economics Letters*, vol. 79, no. 3, pp. 423–427, Jun 2003.

49. J. Bardeen, L. N. Cooper, and J. R. Schrieffer, "Theory of Superconductivity," *Physical Review*, vol. 108, no. 5, pp. 1175–1204, Dec 1957.

50. J. J. Burnett, A. Bengtsson, M. Scigliuzzo, D. Niepce, M. Kudra, P. Delsing, and J. Bylander, "Decoherence benchmarking of superconducting qubits," *npj Quantum Information*, vol. 5, no. 1, 2019.

51. M. L. Bushnell and V. D. Agrawal, *Essentials of Electronic Testing for Digital, Memory and Mixed-Signal VLSI.* Kluwer Academic Publishers, 2002.

52. A. Childs, "Lecture 18: The quantum adiabatic theorem," Accessed 22 May 2019. [Online]. Available: https://www.math.uwaterloo.ca/~amchilds/teaching/w08/l18.pdf

53. F. Chong, "ACM SIGARCH Blog- Verifying Quantum Software and Hardware," 2018, Accessed 24 July 2021. [Online]. Available: https://cccblog.org/2018/06/19/acm-sigarch-blog-verifying-quantum-software-and-hardware/

54. J. M. Chow, L. Dicarlo, J. M. Gambetta, F. Motzoi, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, "Optimized driving of superconducting artificial atoms for improved single-qubit gates," *RAPID COMMUNICATIONS PHYSICAL REVIEW A*, vol. 82, p. 40305, 2010.

55. J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, "Randomized benchmarking and process tomography for gate errors in a solid-state qubit," *Physical Review Letters*, vol. 102, no. 9, p. 090502, Mar 2009.

56. J. Clarke and F. K. Wilhelm, "Superconducting quantum bits," *Nature*, vol. 453, no. 7198, pp. 1031–1042, Jun 2008.

57. W. E. Cobb, "Exploitation of Unintentional Information Leakage From Integrated Circuits," Ph.D. dissertation, Air Force Institute of Technology, 2011, Accessed 26 September 2018. [Online]. Available: http://www.dtic.mil/dtic/tr/fulltext/u2/a550584.pdf

58. W. E. Cobb, E. D. Laspe, R. O. Baldwin, M. A. Temple, and Y. C. Kim, "Intrinsic physical-layer authentication of integrated circuits," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 1 PART 1, pp. 14–24, Feb 2012.

59. W. E. Cobb, M. A. Temple, R. O. Baldwin, E. W. Garcia, and E. D. Laspe, "Intrinsic Physical Layer Authentication of Integrated Circuits," 2015, Accessed 24 Aug 2021. [Online]. Available: https://www.freepatentsonline.com/9036891.html

60. R. Cohen, "Air Force's FY-20 POM shifts more money toward quantum computing, autonomy," *Inside Defense*, 2018, Accessed 26 September 2018. [Online]. Available: https://insidedefense.com/insider/air-forces-fy-20-pom-shifts-more-money-toward-quantum-computing-autonomy

61. D. Collins, "DARPA "TRUST in IC's" Effort," 2007, Accessed 26 September 2018. [Online]. Available: http://www.acq.osd.mil/dsb/reports/2005-02-HPMS_Report_Final.pdf

62. J. S. Coron and I. Kizhvatov, "An efficient method for random delay generation in embedded software," in *Lecture Notes in Computer Science (including*

subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 5747 LNCS. Springer, Berlin, Heidelberg, 2009, pp. 156–170.

63. R. Courtland, "IBM Puts a Quantum Processor in the Cloud," *IEEE Spectrum*, May 2016, Accessed 24 August 2021. [Online]. Available: https://spectrum.ieee.org/tech-talk/computing/hardware/ibm-puts-a-quantum-processor-in-the-cloud

64. P. Davari, O. Kristensen, and F. Iannuzzo, "Investigation of acoustic emission as a non-invasive method for detection of power semiconductor aging," *Microelectronics Reliability*, vol. 88-90, pp. 545–549, 2018.

65. M. H. Devoret, A. Wallraff, and J. M. Martinis, "Superconducting Qubits: A Short Review," 2004, Accessed 24 June 2019. [Online]. Available: http://arxiv.org/abs/cond-mat/0411174

66. J. Dofe, Q. Yu, H. Wang, and E. Salman, "Hardware Security Threats and Potential Countermeasures in Emerging 3D ICs," in *Proceedings of the 26th edition on Great Lakes Symposium on VLSI - GLSVLSI '16*. New York, New York, USA: ACM Press, 2016, pp. 69–74.

67. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*, 2nd ed. Wiley-Interscience, 2001.

68. J. A. Durrum and R. A. Kolker, "Bullet Background Paper on Dynamic Power Analysis (DPA)," 2018.

69. J. J. Dziak, D. L. Coffman, S. T. Lanza, R. Li, and L. S. Jermiin, "Sensitivity and specificity of information criteria," *Briefings in Bioinformatics*, vol. 21, no. 2, pp. 553–565, Mar 2020.

70. S. S. Elder, C. S. Wang, P. Reinhold, C. T. Hann, K. S. Chou, B. J. Lester, S. Rosenblum, L. Frunzio, L. Jiang, and R. J. Schoelkopf, "High-Fidelity Measurement of Qubits Encoded in Multilevel Superconducting Circuits," *Physical Review X*, vol. 10, 2020.

71. J. M. Elzerman, R. Hanson, L. H. Willems Van Beveren, B. Witkamp, L. M. K. Vandersypen, and L. P. Kouwenhoven, "Single-shot read-out of an individual electron spin in a quantum dot," 2004, Accessed 25 November 2018. [Online]. Available: https://arxiv.org/pdf/cond-mat/0411232.pdf

72. P. Erto, "The Inverse Weibull Survival Distribution and its Proper Application," May 2013, Accessed 19 January 2021. [Online]. Available: http://arxiv.org/abs/1305.6909

73. S. Erwin, "Pentagon sees quantum computing as key weapon for war in space," 2018, Accessed 26 September 2018. [Online]. Available: https://spacenews.com/pentagon-sees-quantum-computing-as-key-weapon-for-war-in-space/

74. D. Evans, "Understanding and Mitigating Supply Chain Risks for Computing and Communications (or: Who's Driving Your Missiles?)," Alexandria, VA, 2009, Accessed 24 August 2021. [Online]. Available: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiA7qu0kMryAhVYCTQIHRYBCw0QFnoECAYQAQ&url=https%3A%2F%2Fwww.cs.virginia.edu%2F~evans%2Ftalks%2Fdssg.pptx&usg=AOvVaw1vytJ5HQA9kUjjPCLXl-24

75. G. Fasano and A. Franceschini, "A multidimensional version of the Kolmogorov–Smirnov test," *Monthly Notices of the Royal Astronomical Society*, vol. 225, no. 1, pp. 155–170, Mar 1987.

76. D. M. Fleetwood, M. P. Rodgers, L. Tsetseris, X. J. Zhou, I. Batyrev, S. Wang, R. D. Schrimpf, and S. T. Pantelides, "Effects of device aging on microelectronics radiation response and reliability," in *2006 25th International Conference on Microelectronics, MIEL 2006 - Proceedings*, 2006, pp. 89–98.

77. A. I. Gabilondo, E. R. Ortega, and E. S. Villanueva, "Quantum Computation With Superconductors," Ph.D. dissertation, Universidad del Pais Vasco, 2016, Accessed 11 April 2019. [Online]. Available: https://www.qutisgroup.com/wp-content/uploads/2014/10/Amaia_TFG.pdf

78. H. Gamaarachchi and H. Ganegoda, "Power Analysis Based Side Channel Attack," p. 91, 2018, Accessed 25 August 2021. [Online]. Available: http://arxiv.org/abs/1801.00932

79. J. M. Gambetta, J. M. Chow, and M. Steffen, "Building logical qubits in a superconducting quantum computing system," *npj Quantum Information*, vol. 3, no. 1, pp. 1–7, Dec 2017.

80. A. Ghatak, S. Lokanathan, A. Ghatak, and S. Lokanathan, "Linear Harmonic Oscillator: I Solution of the Schrödinger Equation and Relationship with the Classical Oscillator," in *Quantum Mechanics: Theory and Applications*. Dordrecht: Springer Netherlands, 2013, pp. 159–196.

81. T. Ghose, "This Chip Will Self-Destruct in 5 Seconds," 2015, Accessed 26 September 2018. [Online]. Available: https://www.livescience.com/52397-self-destructing-chip-secures-data.html

82. D. Gil, J. Mantas, R. Sutor, L. Kesterson-Townes, F. Flöther, and C. Schnabel, "Coming soon to your business—Quantum computing," Accessed 7 July 2020. [Online]. Available: https://www.ibm.com/downloads/cas/OV1V0NLX

83. L. Gil and G. L. Lippi, "Self-induced ac Stark shift in lasers," *Physical Review A*, vol. 83, no. 4, p. 043840, Apr 2011.

84. S. Glen, "Likelihood Function: Overview / Simple Definition," Dec 2017, Accessed 31 December 2020. [Online]. Available: https://www.statisticshowto.com/likelihood-function/

85. S. Guilley, J.-L. Danger, R. Nguyen, and P. Nguyen, "System-Level Methods to Prevent Reverse-Engineering, Cloning, and Trojan Insertion System-Level Methods to Prevent Reverse-Engineering, Cloning, and Trojan Insertion," *Communications in Computer and Information Science*, vol. 285, pp. 433–438, 2012.

86. U. Guin, D. Dimase, and M. Tehranipoor, "Counterfeit Integrated Circuits: Detection, Avoidance, and the Challenges Ahead," *J Electron Test*, vol. 30, pp. 9–23, 2014.

87. J. Han. Kullback-Leibler Divergence. Accessed 2 May 2020. [Online]. Available: http://hanj.cs.illinois.edu/cs412/bk3/KL-divergence.pdf

88. K. Hansen, "Raising and lowering operators," Accessed 18 April 2019. [Online]. Available: http://physics.gu.se/~klavs/FYP310/RaisLow.pdf

89. L. Hardesty, "Toward mass-producible quantum computers," 2017, Accessed 27 September 2018. [Online]. Available: http://news.mit.edu/2017/toward-mass-producible-quantum-computers-0526

90. T. Hastie, R. Tibshirani, and J. Friedman, *Springer Series in Statistics The Elements of Statistical Learning - Data Mining, Inference, and Prediction*, 2009, vol. 2nd.

91. L. Hsia, G. Vernizzi, M. Lanzerotti, and D. Langley, "Sampling iso-functional signal switches in library circuits for microelectronics verification with topo-

logical constraints," in *Proceedings of the IEEE National Aerospace Electronics Conference, NAECON*, 2017.

92. L. Hsia, G. Vernizzi, M. Lanzerotti, D. Langley, M. Seery, and L. Orlando, "Topological constraints of gate-level circuits obtained through standard cell recognition (SCR)," in *Proceedings of the IEEE National Aerospace Electronics Conference, NAECON*, 2015.

93. L. A. Hsia, L. D. Merkle, G. Vernizzi, M. Y. Lanzerotti, and D. Langley, "On Classical Hardware Verification and Security Techniques for Quantum Computing Systems," in *Government Microcircuit Applications & Critical Technology Conference 2019*, 2019, pp. 399–404.

94. L. A. Hsia, L. D. Merkle, D. E. Weeks, G. Vernizzi, M. Y. Lanzerotti, and D. Langley, "Physically Unclonable Characteristics for Verification of Transmon-Based Quantum Computers," in *Government Microcircuit Applications & Critical Technology Conference 2020*, 2020.

95. L. A. Hsia, G. Vernizzi, M. Y. Lanzerotti, and D. Langley, "Building a Library for Microelectronics Verification with Topological Constraints," in *42nd Annual Government Microcircuit Applications & Critical Technology Conference*, 2017.

96. C. M. Hurvich and C. L. Tsai, "Regression and time series model selection in small samples," *Biometrika*, vol. 76, no. 2, pp. 297–307, 1989.

97. P. Jurcevic, A. Javadi-Abhari, L. S. Bishop, I. Lauer, D. F. Bogorin, M. Brink, L. Capelluto, O. Günlük, T. Itoko, N. Kanazawa, A. Kandala, G. A. Keefe, K. Krsulich, W. Landers, E. P. Lewandowski, D. T. McClure, G. Nannicini, A. Narasgond, H. M. Nayfeh, E. Pritchett, M. B. Rothwell, S. Srinivasan, N. Sundaresan, C. Wang, K. X. Wei, C. J. Wood, J. B. Yau, E. J. Zhang, O. E. Dial,

J. M. Chow, and J. M. Gambetta, "Demonstration of quantum volume 64 on a superconducting quantum computing system," *Quantum Science and Technology*, vol. 6, no. 2, 2021.

98. A. G. Kimura, S. B. Bibyk, B. P. Dupaix, M. J. Casto, and G. L. Creech, "Metrics for Analyzing Quantifiable Differentiation of Designs with Varying Integrity for Hardware Assurance," Accessed 26 September 2018. [Online]. Available: http://www.dtic.mil/dtic/tr/fulltext/u2/1041893.pdf

99. R. W. Klein, M. A. Temple, and M. J. Mendenhall, "Application of wavelet-based RF fingerprinting to enhance wireless network security," *Journal of Communications and Networks*, vol. 11, no. 6, pp. 544–555, 2009.

100. A. Kleinsasser, T. Chui, B. Bumble, and E. Ladizinsky, "Critical current density and temperature dependence of Nb-Al oxide-Nb junction resistance and implications for room temperature characterization," *IEEE Transactions on Applied Superconductivity*, vol. 23, no. 3, 2013.

101. J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, "Charge-insensitive qubit design derived from the Cooper pair box," *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 76, no. 4, Feb 2007.

102. P. J. Koppinen, L. M. Väistö, and I. J. Maasilta, "Complete stabilization and improvement of the characteristics of tunnel junctions by thermal annealing," *Applied Physics Letters*, vol. 90, no. 5, p. 53503, 2007.

103. V. T. Lahtinen and J. K. Pachos, "A short introduction to topological quantum computation," *SciPost Phys*, vol. 3, p. 21, 2017.

104. N. K. Langford, "Circuit QED-Lecture Notes," Accessed 15 May 2019. [Online]. Available: https://pure.royalholloway.ac.uk/portal/files/16799519/langfordnk_circuitqed_lecturenotes_rhul.pdf

105. B. Lau, "2-d Kolmorogov-Smirnov test, n-d energy test, Hotelling T^2 test," 2021, Accessed 28 April 2021. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/48309-2-d-kolmorogov-smirnov-test-n-d-energy-test-hotelling-t-2-test

106. M. Levene and A. Kononovicius, "Empirical survival Jensen-Shannon divergence as a goodness-of-Fit measure for maximum likelihood estimation and curve fitting," *Communications in Statistics: Simulation and Computation*, 2019.

107. J. Lin, "Divergence Measures Based on the Shannon Entropy," *IEEE Transactions on Information Theory*, vol. 37, 1991.

108. N. M. Linke, D. Maslov, M. Roetteler, S. Debnath, C. Figgatt, K. A. Landsman, K. Wright, and C. Monroe, "Comparing the architectures of the first programmable quantum computers," in *2017 Conference on Lasers and Electro-Optics Europe & European Quantum Electronics Conference (CLEO/Europe-EQEC)*. IEEE, Jun 2017, pp. 1–1.

109. Y. Liu, S. J. Srinivasan, D. Hover, S. Zhu, R. McDermott, and A. A. Houck, "High fidelity readout of a transmon qubit using a superconducting low-inductance undulatory galvanometer microwave amplifier," *New Journal of Physics*, vol. 16, 2014.

110. Y. Lu, M. P. O'Neill, and J. V. McCanny, "FPGA implementation and analysis of random delay insertion countermeasure against DPA," in *Proceedings of*

the 2008 International Conference on Field-Programmable Technology, ICFPT 2008. IEEE, Dec 2008, pp. 201–208.

111. M. W. Lukacs, A. J. Zeqolari, P. J. Collins, and M. A. Temple, ""RF-DNA" Fingerprinting for Antenna Classification," *IEEE Antennas and Wireless Propagation Letters*, vol. 14, pp. 1455–1458, Mar 2015.

112. E. Magesan, J. M. Gambetta, and J. Emerson, "Characterizing quantum gates via randomized benchmarking," *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 85, no. 4, p. 042311, Apr 2012.

113. F. Mallet, F. R. Ong, A. Palacios-Laloy, F. Nguyen, P. Bertet, D. Vion, and D. Esteve, "Single-shot qubit readout in circuit quantum electrodynamics," *Nature Physics*, vol. 5, no. 11, pp. 791–795, Sep 2009.

114. J. M. Martinis and K. Osborne, "Superconducting Qubits and the Physics of Josephson Junctions," Feb 2004, Accessed 11 April 2019. [Online]. Available: http://arxiv.org/abs/cond-mat/0402415

115. H. Marzouqi, K. Salah, M. Al-Qutayri, and M. C. Y. Yeun, "A unified countermeasure against side channel attacks on cryptographic RFID," in *2011 International Conference for Internet Technology and Secured Transactions*, December 2011, pp. 13–18.

116. J. Mattis, "2018 National Defense Strategy of the United States Summary," p. 11, 2018, Accessed 26 September 2018. [Online]. Available: https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf

117. D. T. McClure, "Slack — ibm-q-systems — Qiskit," Accessed 8 April 2020. [Online]. Available: https://app.slack.com/client/T7RSPHKK2/C8GLMQ1U5/thread/C8GLMQ1U5-1586186668.050900

118. D. C. McKay, T. Alexander, L. Bello, M. J. Biercuk, L. Bishop, J. Chen, J. M. Chow, A. D. Córcoles, D. Egger, S. Filipp, J. Gomez, M. Hush, A. Javadi-Abhari, D. Moreda, P. Nation, B. Paulovicks, E. Winston, C. J. Wood, J. Wootton, and J. M. Gambetta, "Qiskit Backend Specifications for OpenQASM and OpenPulse Experiments," 2018.

119. D. C. McKay, S. Sheldon, J. A. Smolin, J. M. Chow, and J. M. Gambetta, "Three-Qubit Randomized Benchmarking," *Physical Review Letters*, vol. 122, no. 20, p. 200502, May 2019.

120. A. Moradi, "Secure Logic Styles ECRYPT II summer school on Design and Security of Cryptographic Algorithms and Devices," 2011, Accessed 6 June 2019. [Online]. Available: https://www.cosic.esat.kuleuven.be/ecrypt/courses/albena11/slides/amir_moradi_secure_logic_styles.pdf

121. F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, "Simple Pulses for Elimination of Leakage in Weakly Nonlinear Qubits," *Physical Review Letters*, vol. 103, no. 11, p. 110501, Sep 2009.

122. B. Nachman, M. Urbanek, W. A. de Jong, and C. W. Bauer, "Unfolding quantum computer readout noise," *npj Quantum Information*, vol. 6, no. 1, pp. 1–7, Dec 2020.

123. M. Nakahara and T. Ohmi, *Quantum computing : from linear algebra to physical realizations*. CRC Press, 2008.

124. P. Nation, "Slack — ibm-q-systems — Qiskit," Accessed 24 December 2020. [Online]. Available: https://app.slack.com/client/T7RSPHKK2/C8GLMQ1U5/thread/C8GLMQ1U5-1607612146.489500

125. Y. Naveh, E. Kashefi, J. R. Wootton, and K. Bertels, "Theoretical and practical aspects of verification of quantum computers," in *Proceedings of the 2018 Design, Automation and Test in Europe Conference and Exhibition, DATE 2018*. IEEE, Mar 2018, pp. 721–730.

126. H. V. Nguyen and J. Vreeken, "Non-parametric Jensen-Shannon divergence," in *Lecture Notes in Computer Science*, vol. 9285. Springer Verlag, 2015, pp. 173–189.

127. F. Nielsen, "On the Jensen–Shannon Symmetrization of Distances Relying on Abstract Means," *Entropy*, vol. 21, no. 5, p. 485, May 2019.

128. M. J. O'Hara and K. K. Berggren, "Room-temperature testing for high critical-current-density Josephson junctions," *IEEE Transactions on Applied Superconductivity*, vol. 10, no. 4, pp. 1669–1672, Dec 2000.

129. S. O'Neill, "Radio Frequency-Based Device Discrimination of Mixed-Signal Integrated Circuits and Counterfeit Detection," Ph.D. dissertation, Air Force Institute of Technology, 2017.

130. A. Peterson, "New Microelectronics Initiative Sets Up Possible Funding Infusion," *FYI: Science Policy News from AIP*, 2021, Accessed 24 February 2021. [Online]. Available: https://www.aip.org/fyi/2021/new-microelectronics-initiative-sets-possible-funding-infusion?utm_medium=email&utm_source=FYI&dm_i=1ZJN,7952P,E29DQU,TELBU,1

131. K. Phalak, A. Ash-Saki, M. Alam, R. O. Topaloglu, and S. Ghosh, "Quantum PUF for Security and Trust in Quantum Computing," Accessed 11 August 2021. [Online]. Available: https://arxiv.org/pdf/2104.06244.pdf

132. F. Poucheret, L. Barthe, P. Benoit, L. Torres, P. Maurine, M. Robert, and F. Poucheret, "Spatial EM Jamming: a Countermeasure Against EM Analysis?" in *Proceedings of the 2010 18th IEEE/IFIP International Conference on VLSI and System-on-Chip, VLSI-SoC 2010 (2010) 105-110*, 2010, p. 544358.

133. W. H. Press and S. A. Teukolsky, "Kolmogorov-Smirnov Test for Two-Dimensional Data," *Computers in Physics*, vol. 2, no. 4, p. 74, 1988.

134. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "14.7 Do Two-Dimensional Distributions Differ?" in *Numerical Recipes in C: The Art of Scientific Computing*, second edition ed. Cambridge University Press, 1992, ch. Chapter 14, pp. 645–649.

135. T. Proctor, K. Rudinger, K. Young, M. Sarovar, and R. Blume-Kohout, "What Randomized Benchmarking Actually Measures," *Physical Review Letters*, vol. 119, no. 13, p. 130502, Sep 2017.

136. R. Rahman, S. H. Park, T. B. Boykin, G. Klimeck, S. Rogge, and L. C. L. Hollenberg, "Gate-induced g-factor control and dimensional transition for donors in multivalley semiconductors," *Physical Review B - Condensed Matter and Materials Physics*, vol. 80, no. 15, 2009.

137. A. Raucher, "Protecting Against Side-Channel Attacks with an Ultra-Low Power Processor," Accessed 26 September 2018. [Online]. Available: https://www.synopsys.com/designware-ip/technical-bulletin/protecting-against-side-channel.html

138. J. Robertson and M. Riley, "The Big Hack: How China Used a Tiny Chip to Infiltrate U.S. Companies," 2018, Accessed 15 July 2020. [Online]. Available: https://www.bloomberg.com/news/features/2018-10-04/the-big-hack-how-china-used-a-tiny-chip-to-infiltrate-america-s-top-companies

139. R. N. Rodriguez, "A Guide to the Burr Type XII Distributions," *Biometrika*, vol. 64, no. 1, p. 129, Apr 1977.

140. T. Rosa dos Reis, "Measuring the statistical similarity between two samples using Jensen-Shannon and Kullback-Leibler divergences," Accessed 2 May 2020. [Online]. Available: https://medium.com/datalab-log/measuring-the-statistical-similarity-between-two-samples-using-jensen-shannon-and-kullback-leibler-8d05af514b15

141. A. C. Rose-Innes and E. H. Rhoderick, *Introduction to superconductivity*. Pergamon Press, 1978.

142. F. Saqib and J. Plusquellic, "VLSI Test and Hardware Security Background for Hardware Obfuscation," in *Hardware Protection through Obfuscation*. Cham: Springer International Publishing, 2017, pp. 33–68.

143. P. Sasdrich, A. Moradi, O. Mischke, and T. Güneysu, "Achieving Side-Channel Protection with Dynamic Logic Reconfiguration on Modern FPGAs," in *IEEE International Symposium on Hardware Oriented Security and Trust*, 2015.

144. H. Schomerus and J. P. Robinson, "Entanglement between static and flying qubits in an Aharonov–Bohm double electrometer," *New Journal of Physics*, vol. 9, no. 3, pp. 67–67, Mar 2007.

145. G. Schwarz, "Estimating the Dimension of a Model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

146. D. W. Scott, "Sturges' rule," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 1, no. 3, pp. 303–306, Nov 2009.

147. ——, "Scott's rule," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 497–502, Jul 2010.

148. ——, *Multivariate Density Estimation*, ser. Wiley Series in Probability and Statistics. Wiley, Mar 2015.

149. M. Shapiro, "Derivation of the coordinate-momentum commutation relations from canonical invariance," *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 74, no. 4, p. 042104, Oct 2006.

150. M. D. Shulman, O. E. Dial, S. P. Harvey, H. Bluhm, V. Umansky, and A. Yacoby, "Demonstration of entanglement of electrostatically coupled singlet-triplet qubits," *Science*, vol. 336, no. 6078, pp. 202–205, 2012.

151. D. Siegle, "T Test," Accessed 11 April 2020. [Online]. Available: https://researchbasics.education.uconn.edu/t-test/

152. R. W. Simmonds, D. A. Hite, R. McDermott, M. Steffen, K. B. Cooper, K. M. Lang, J. M. Martinis, and D. P. Pappas, "Josephson junction materials research using phase qubits," in *Quantum Computing in Solid State Systems*, 2006, pp. 86–94.

153. A. M. Stewart, Z. Zhang, J. Jiang, E. Christou, and C. Depken, "Jensen-Shannon Divergence: Estimation and Hypothesis Testing," 2019, Accessed 1 May 2020. [Online]. Available: https://math.uncc.edu/sites/math.uncc.edu/files/fields/preprint_archive/paper/2019_06.pdf

154. Y. Su, L. Li, M. Liu, D. Geng, Y. Gong, G. Yang, X. Chuai, Y. Zhao, X. Shi, L. Zhang, and N. Lu, "Dynamic Time Evolutionary Aging Analysis for Device-

Circuit Lifetime Estimation of Thin-Film Transistors," *IEEE Electron Device Letters*, vol. 40, no. 9, pp. 1439–1442, Jul 2019.

155. S. S. Tannu and M. K. Qureshi, "Not All Qubits Are Created Equal: A Case for Variability-Aware Policies for NISQ-Era Quantum Computers," in *International Conference on Architectural Support for Programming Languages and Operating Systems - ASPLOS*, 2019, pp. 987–999.

156. M. Tehranipoor, U. Guin, and D. Forte, *Counterfeit Integrated Circuits: Detection and Avoidance*, 2015.

157. K. Tiri and I. Verbauwhede, "A VLSI design flow for secure side-channel attack resistant ICs," *Proceedings -Design, Automation and Test in Europe, DATE '05*, vol. 2005, pp. 58–63, 2005.

158. N. V. Tkachenko, *Optical measurements*. Elsevier Science, 2006.

159. J. Tolar, "On Clifford groups in quantum computing," in *Journal of Physics: Conference Series*, vol. 1071, no. 1, 2018.

160. T. Uchiyama, "Information-Theoretic Clustering and Algorithms," in *Advances in Statistical Methodologies and Their Application to Real Problems*, 2017.

161. L. R. Vale and R. H. Ono, "Long term stability of YBCO-based Josephson junctions," *IEEE Transactions on Applied Superconductivity*, vol. 9, no. 2 PART 3, pp. 3382–3385, 1999.

162. M. Veldhorst, H. G. J. Eenink, C. H. Yang, and A. S. Dzurak, "Silicon CMOS architecture for a spin-based quantum computer," *Nature Communications*, vol. 8, no. 1, p. 1766, Dec 2017.

163. M. Veldhorst, J. C. Hwang, C. H. Yang, A. W. Leenstra, B. De Ronde, J. P. Dehollain, J. T. Muhonen, F. E. Hudson, K. M. Itoh, A. Morello, and A. S. Dzurak, "An addressable quantum dot qubit with fault-tolerant control-fidelity," *Nature Nanotechnology*, vol. 9, no. 12, pp. 981–985, 2014.

164. M. Veldhorst, C. H. Yang, J. C. Hwang, W. Huang, J. P. Dehollain, J. T. Muhonen, S. Simmons, A. Laucht, F. E. Hudson, K. M. Itoh, A. Morello, and A. S. Dzurak, "A two-qubit logic gate in silicon," *Nature*, vol. 526, no. 7573, pp. 410–414, 2015.

165. E. J. Wagenmakers and S. Farrell, "AIC model selection using Akaike weights," pp. 192–196, 2004.

166. W. R. Wessels, "Use of the Weibull versus exponential to model part reliability," in *2007 Proceedings - Annual Reliability and Maintainability Symposium, RAMS*, 2007, pp. 131–135.

167. E. Zipper, M. Kurpas, J. Sadowski, and M. M. Maśka, "Semiconductor quantum ring as a solid-state spin qubit," *Journal of Physics: Condensed Matter*, vol. 23, no. 11, p. 115302, Nov 2011.

Acronyms

ACM SIGARCH Association for Computing Machinery Special Interest Group on Computer Architecture. 48

AIC Akaike's information criterion. 58, 59, 97, 98, 102, 110, 111, 113, 114, 152

AICc corrected AIC. 58, 59, 97, 98, 102, 106, 110, 111, 113, 114

ARO aging-resistant ring oscillator. 19

BIC Bayesian information criterion. 58, 59, 97, 98, 102, 106, 110, 112, 113, 114, 166, 199

CDF cumulative distribution function. 67

CDIR combating die/IC recovery. 16

DARPA Defense Advanced Research Projects Agency. 14, 62, 339

DES data encryption standard. 22

DoD Department of Defense. 1

ECDF empirical cumulative distribution function. 63

ECID electronic chip ID. 16

EPC error per Clifford. 44

ESD electrostatic discharge. 16

ESR electron spin resonance. 30

EV extreme value. 57, 114, 115, 145, 166, 179, 192, 194

FPGA field-programmable gate array. 22

GESD generalized extreme Studentized deviate. 241

GEV generalized extreme value. 57, 111, 112, 113, 114, 115, 119, 120, 121, 124, 125, 126, 129, 130, 132, 134, 137, 139, 145, 147, 149, 152, 154, 157, 160, 162, 164, 166, 168, 170, 173, 175, 179, 182, 184, 186, 188, 190, 192, 194, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 219, 222, 224, 226, 228, 234, 236

IC integrated circuit. 19

JSD Jensen-Shannon divergence. 3, 11, 13, 51, 62, 63, 64, 65, 66, 72, 282, 283, 330, 333, 334, 336, 337, 338

KLD Kullback-Leibler divergence. 63

KS Kolmogorov-Smirnov. 3, 11, 13, 57, 62, 63, 67, 68, 98, 100, 108, 140, 147, 177, 236, 237, 252, 254, 282, 292, 330, 333, 334, 336, 337, 338, 339

LDA linear discriminant analysis. 65

LPA log-periodic antenna. 19

MDA multiple discriminant analysis. 66, 291

MDA-ML multiple discriminant analysis with maximum likelihood. 19, 62, 66, 67, 333, 339

ML maximum likelihood. 66

NLL negative log-likelihood. 58, 59, 97, 98, 110, 111, 113, 114, 152, 203, 207

OCM original component manufacturer. 16

PCA principal component analysis. 291

PD probability of detection. 6, 13, 47, 51, 52, 60, 62, 67, 73, 74, 245, 252, 253, 255, 256, 282, 283, 291, 292, 293, 338

PFA probability of false alarm. 6, 13, 47, 51, 52, 60, 62, 67, 73, 74, 245, 252, 253, 255, 256, 282, 283, 291, 292, 293, 338

PUC physically unclonable characteristic. xi, 3, 4, 5, 6, 8, 11, 13, 14, 38, 42, 44, 45, 57, 62, 67, 74, 330, 331, 333, 334, 335, 336, 337, 339, 340

PUF physically unclonable function. xi, 3, 4, 8, 10, 12, 14, 16, 19, 46, 48, 331, 333, 335

QDA quadratic discriminant analysis. 11, 13, 51, 62, 65, 66, 72, 291, 333, 336, 337

QLA quantum-limited amplifiers. 37

QPT quantum process tomography. 42, 43, 100, 130, 134, 338

RB randomized benchmarking. 100, 338

RF-DNA radio-frequency (RF) distinct native attributes (DNA). 9, 10, 12, 14, 16, 67, 334, 335, 336

RO-PUF ring oscillator PUF. 19

SAM scanning acoustic microscopy. 15

SCA side channel attacks. 20, 21, 332

SEM scanning electron microscopy. 15, 16, 17

SLUG superconducting low-inductance undulator galvanometer. 38

SQUID superconducting quantum interference device. 36, 37

SST secure split test. 16

XRF x-Ray fluorescence. 17

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
16–09–2021	Dissertation	Aug 2018 — Sep 2021

4. TITLE AND SUBTITLE

PHYSICALLY UNCLONABLE CHARACTERISTICS FOR VERIFICATION OF TRANSMON-BASED QUANTUM COMPUTERS

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Hsia, Leleia A, Capt

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENG-DS-21-S-007

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFRL/RITQ
Ms. Laura Wessing
525 Brooks Road
Rome, NY 13441
COMM: 315-330-2937

10. SPONSOR/MONITOR'S ACRONYM(S)

AFRL/RITQ

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Future national security can be strengthened by verifying and securing the quantum computing supply chain. This dissertation proposes physically unclonable characteristics (PUCs), a method of quantum hardware verification inspired by classical physically unclonable functions, for future application to quantum processors implemented with transmon qubits. Qualitative and quantitative analysis is provided on the development of PUCs, including identifying qubit characteristics and qubit discrimination methods suitable for PUCs. Characteristics tested on IBM Quantum services include T1 and T2 coherence times, single-qubit and multi-qubit gate error rates, readout error rates, quantum process tomography metrics, and random benchmarking metrics. Results show that non-parametric qubit discrimination methods are best-suited for the characteristics tested, but require refinement before real-world implementation can be achieved.

15. SUBJECT TERMS

Hardware verification, Physically unclonable characteristics, Physically unclonable functions, Quantum computing, Transmon

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Laurence D. Merkle, AFIT/ENG
U	U	U	UU	396	19b. TELEPHONE NUMBER *(include area code)* (937) 255-6565; laurence.merkle@afit.edu