Measuring NISQ Gate-Based Qubit Stability Using a 1+1 Field Theory and Cycle Benchmarking

Kübra Yeter-Aydeniz ‡

Physics Division and Computational Sciences and Engineering Division
Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

Zachary Parks

Department of Computer Science, North Carolina State University
Raleigh, North Carolina 27695

Aadithya Nair

Department of Computer Science, North Carolina State University
Raleigh, North Carolina 27695

Erik Gustafson Σ

Department of Physics and Astronomy, The University of Iowa, Iowa City, IA 52242

Alexander F. Kemper

Department of Physics, North Carolina State University, Raleigh, North Carolina 27695

Raphael C. Pooser

Computational Sciences and Engineering Division, Oak Ridge National Laboratory
Oak Ridge, TN 37831, USA

Yannick Meurice

Department of Physics and Astronomy, The University of Iowa, Iowa City, IA 52242

Patrick Dreher ‖

Electrical and Computer Engineering and Department of Physics
North Carolina State University, Raleigh, NC 27695, USA

‡ current address: Emerging Engineering and Physical Sciences Department.
The MITRE Corporation, 7515 Colshire Drive, McLean, VA 22102-7539, USA.
Σ current address: Theoretical Physics Division, Fermi National Accelerator Lab, Batavia, IL 60510
‖ Corresponding Author: email: padreher@ncsu.edu

Abstract. Some of the most problematic issues that limit the implementation of applications on Noisy Intermediate Scale Quantum (NISQ) machines are the adverse impacts of both incoherent and coherent errors. We conducted an in-depth study of coherent errors on a quantum hardware platform using a transverse field Ising model Hamiltonian as a sample user application. We report here on the results from these computations using several error mitigation protocols that profile these errors and provide an indication of the qubit stability. Through a detailed set of measurements we identify inter-day and intra-day qubit calibration drift and the impacts of quantum circuit placement on groups of qubits in different physical locations on the processor. This paper also discusses how these measurements can provide a better understanding of these types of errors and how they may improve efforts to validate the accuracy of quantum computations. ¶

1. Introduction

Today researchers and application developers have access to the first generation of Noisy Intermediate Scale Quantum (NISQ) [1] quantum computing (QC) hardware platforms. This has opened opportunities for users to begin exploring how to reformulate existing algorithms designed for digital computers onto quantum computing hardware platforms. These machines also offer an operational environment where new algorithms specially optimized for quantum computers can be developed. These new capabilities can now provide the quantum computing user community with an "on-ramp" into the world of quantum computing to explore problems in the near future that, up to this point, have been inaccessible using even the most powerful digital high performance computers.

Although current quantum computing hardware capabilities now allow some basic applications to be successfully implemented, at the present time QC platforms utilizing various hardware architectures can only offer a range from a few qubits up to a few hundred noisy qubits over the next three to five years. These noisy qubits degrade the fidelity of the quantum computations. This limitation on the number of qubits excludes any possibility of constructing "logical qubits" to compensate and correct for the errors introduced by the noise in the quantum system. These NISQ machines can only maintain a coherent quantum state for short periods of time. These times are usually expressed in standard metrics of T1 (relaxation time) and T2 (dephasing time) and two qubit CNOT error rates.

Understanding these different types of errors and developing methods to mitigate them is a critical area of research essential for advancing Quantum Information Science. The types of errors that may occur when running algorithms on quantum hardware can be categorized as incoherent or coherent errors.

Incoherent errors are due to uncontrolled interactions between qubits and the environment that result in the decoherence of the overall quantum state. One particular example of a technique to mitigate decoherence is to scale up the level of noise by introducing pairs of CNOT gates, measure the output in each instance, and then extrapolating the signal output back to a zero-noise limit. This has been extensively studied over the past few years. [2, 3, 4].

Coherent errors are mainly caused by systematic errors in the control of the qubits. This type of error results from over or under rotation in qubit control pulses that can result in calibration errors, drift in the qubit properties, or crosstalk where a nearby qubit is impacted by the coherent rotations directed at a different qubit. These errors are quantified through norm-based error-metrics such as the diamond distance [5] and total variation distance [6] that are defined through a rigorous mathematical representation of unitary operators rotated through an angle θ relative to an intended target state.

Developing a full characterization of these errors on quantum hardware platforms is a challenge because doing so requires quantum process tomography [7, 8, 9, 10, 11] or gate set tomography [12, 13, 14], both of which require resources that grow exponentially with the number of qubits. Alternatives to such an exhaustive option

are process fidelity measurements such as Randomized Benchmarking (RB) [15, 16, 17], Cycle Benchmarking (CB) [18] and Randomized Compiling (RC) [5]. Both CB and RC are better able to characterize the cross talk errors in the quantum circuit [19, 1, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30].

Using the IBM Quantum `ibmq_boeblingen` processor, we did an initial examination of the processor's re-calibration data and backend properties. There were many seemingly large random fluctuations occurring in several of the single qubit gate properties from day-to-day. Some of these results are illustrated in Table C1 in Appendix C. For example, qubit 6 shows some extreme fluctuations in the measurements for T_1, T_2, the IBM Quantum backend readout error and the values for the basis gates U_2 and U_3 between the morning of January 24th and 29th. Other single qubits also showed similar fluctuations in their backend properties.

Because it is known that, in general, two qubit gates contribute a substantially larger error on QC hardware platforms than do single qubit gates, our group conducted an in-depth stability analysis of these two-qubit gates on `ibmq_boeblingen` by measuring the process infidelities using CB and comparing them to the results obtained from RB measurements based on the IBM quantum processor qubit re-calibrations. Because of its wide applicability in quantum field theories and many-body interactions [31, 32, 30, 33, 34, 2, 35, 36, 37, 3, 38, 39, 40, 41, 42, 43, 44, 10, 45, 46, 47, 48, 40, 49, 50] we used two-qubit gates in the Transverse Field Ising Model (TFIM) for our study of the two-qubit gate error properties.

This paper is organized as follows. In Sec. 2 we describe the Transverse Field Ising Model used to study the error characterization in an IBM superconducting transmon platform. Sec. 3 discusses the methodology and conditions that were implemented using this model and the types of data that were collected. Sec. 4 summarizes the results of the computations using this data that illustrate inter-day and intra-day quantum hardware processor calibration drift and other measures and tests of stability of the TFIM circuits. Using the results from these measurements, a discussion of some of the implications of these results is presented. Finally, Sec. 5 summarizes our observations and and discusses next steps.

2. Physics Model

We study the transverse field Ising model (TFIM) with open boundary conditions that has the system Hamiltonian

$$H = -J \sum_{i=1}^{N_s-1} \hat{X}_i \hat{X}_{i+1} - h_T \sum_{i=1}^{N_s} \hat{Z}_i. \tag{1}$$

The operators $\hat{X}_i$ and $\hat{Z}_i$ correspond to the Pauli matrices $\hat{\sigma}^x$ and $\hat{\sigma}^z$ respectively. N_s corresponds to the number of sites in the model. J is the nearest neighbor (hopping) coupling and controls the movement of the spins and creation of spin pairs, while h_T is the on-site energy. This model has been used in a variety of contexts related to quantum

computing [31, 32, 30, 33, 34, 2, 35, 36, 37, 3, 38, 39, 40, 41, 42, 43, 44, 10, 45, 46, 47, 48, 49, 50].

In the following, we will study the case with $N_s = 4$, $J = 0.02$ and $h_T = 1$. This choice of parameters provides a simple particle picture and has been used in recent quantum computing studies [51, 52, 53].

The system can be evolved in time using the complex exponential of the Hamiltonian:

$$\hat{U}(t) = e^{-it\hat{H}}. \tag{2}$$

Following Refs. [54, 53, 52], the Trotter approximation is applied to the evolution operator with the explicit form:

$$\hat{U}(t; N) = \left(\hat{U}_1(t/N; h_t)\hat{U}_2(t/N; J) \right)^N + \mathcal{O}(t^2/N) \tag{3}$$

where N is the number of Trotter steps to be implemented, $(\delta t = \frac{t}{N})$ is the Trotter step size. For the values of J and h_t chosen here, a time of approximately $t \sim 100$ ($Jt \sim 2$) is needed in order to observe changes of the occupations that can be interpreted as the motion of a particle across the size of the system.

In our experiments, we used the exponentials of the 1- and 2-body operators in the Hamiltonian

$$\hat{U}_1(\delta t; h_t) = e^{-ih_T \delta t \sum_{i=1}^4 \hat{Z}_i}, \tag{4}$$

and

$$\hat{U}_2(\delta t; J) = e^{-iJ\delta t \sum_{i=1}^3 \hat{X}_i \hat{X}_{i+1}}. \tag{5}$$

We chose a Trotter step $\delta t = 10$ which allows us to reach significant changes using five to ten steps. Notice that $\delta t = 10$ is much larger than what would be required to control the error of one Trotter step with an accuracy proportional δt^2 or δt^3 for an improved Trotter approximation. Instead, it was noticed [55, 52] that for this nonlinear regime, the error grows linearly with a small coefficient. Recent work has shown that the standard error bounds are overly pessimistic [56, 57].

The operators defined in Eqs. 4 and 5 can be expressed as a combination of the two quantum circuit elements (1-qubit and 2-qubit gates) shown in Fig. 1.

3. Methodology

The Transverse Field Ising model (TFIM) Hamiltonian offered a model for studying the error characterization properties of multiple two-qubit CNOT gates. To model the TFIM Hamiltonian in Eq. (1), we selected two quantum circuit diagrams of the TFIM (Fig. 1) that represent two choices or orderings for two-qubit components of this application (labelled as Circuit 1 and Circuit 2). These two quantum circuits are equivalent combinations because the gates representing the different terms in the TFIM Hamiltonian commute with each other. Both result in time evolution of a state with $\mathcal{U} = e^{-iH\delta t}$ using Trotterization where δt is the time interval for one Trotter step. Each
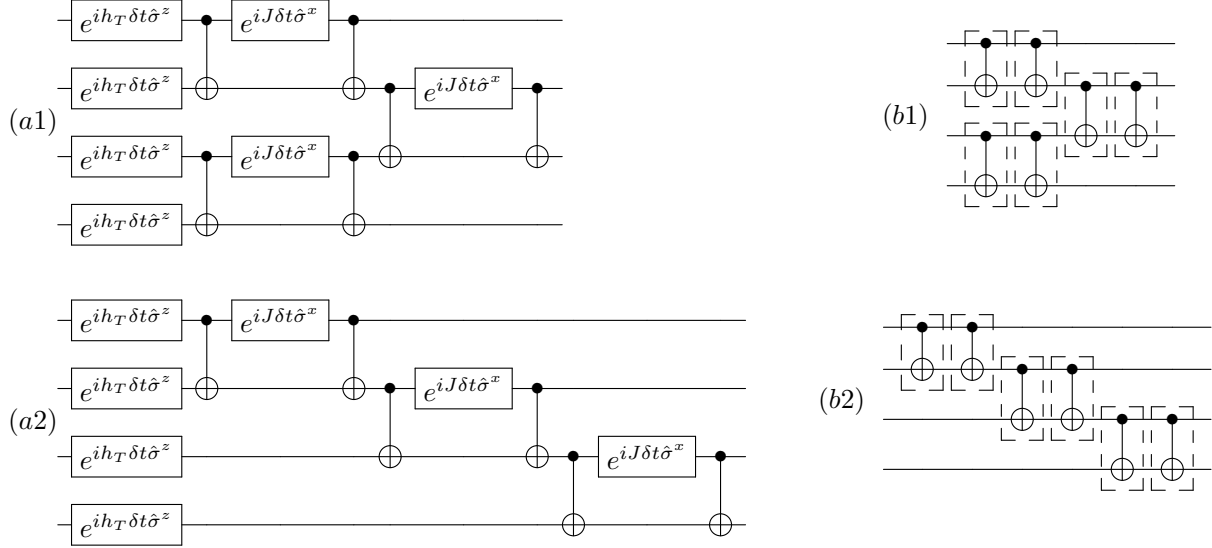
Figure 1: ($a1$,$a2$): The quantum circuits for one Trotter step of the time evolution with the open boundary condition Ising model Hamiltonian. We define the quantum circuit in ($a1$) and ($a2$) as Circuit 1 and Circuit 2 respectively.
($b1$,$b2$): The CNOT hard cycles used at each Trotter step for Circuit 1 ($b1$) and Circuit 2 ($b2$). The quantum circuits were drawn using the Q-circuit package [58].

circuit has three sets of CNOT gates. We focused on measuring CNOT gate performance because the error rates for two qubit gates are substantially higher than single qubit errors.

We selected the 20-qubit `ibmq_boeblingen` hardware platform (Fig. 2) in order to study the stability of this application on a superconducting transmon device. Because the two-qubit gates in quantum circuits are a major source of the errors generated in a quantum computer, this project focused on studying the behavior of the combinations of CNOT gates under various conditions.

The performance characteristics of the 2-qubit gates on the `ibmq_boeblingen` processor representing the CNOT gates for the TFIM Hamiltonian were explored using Cycle Benchmarking (CB) [18] and the Quantum Capacity (QCAP). We chose this protocol because CB is scalable for estimating the effect of all global and local error mechanisms that occur across multi-qubit quantum processors when a clock cycle of operations is applied to a specific quantum register. To calculate the process infidelity of the cycles of interest through CB and the QCAP bound we utilized the True-Q software by Keysight Technologies [+]. True-Q is a software tool that provides methods to calibrate and optimize the performance of quantum devices. Appendix B summarizes CB protocol in general and how it was implemented within the True-Q software in more detail.

[+] https://trueq.quantumbenchmark.com/index.html

Within the qubit layout on `ibmq_boeblingen` we selected three separate groups of qubits as shown in Fig. 2 (Layouts 1, 2, and 3) to study the error characterization due to TFIM Trotterization. Within each layout, an exhaustive combination of pairs of adjacent qubits was identified and labeled as Cycle 1 through 4 in order to measure the specific error characteristics of each qubit pair associated with each of the CNOT gates in the TFIM circuit. Each row in the table in Fig. 2 corresponds to the specific CNOT combinations for that specific layout. For example, Layout 1 measurements included all

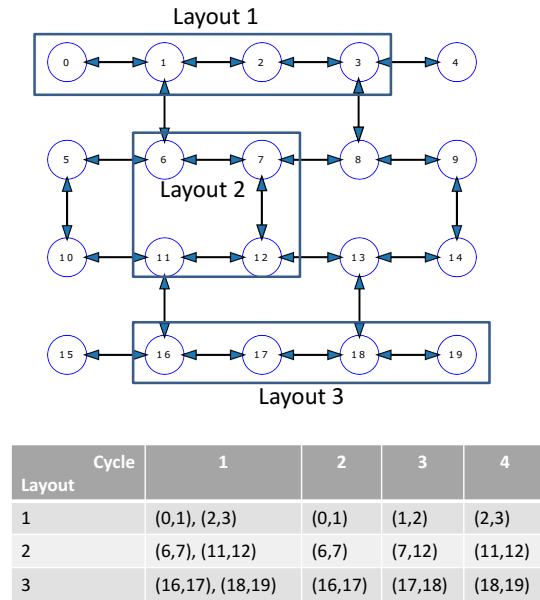| Cycle Layout	1	2	3	4
1	(0,1), (2,3)	(0,1)	(1,2)	(2,3)
2	(6,7), (11,12)	(6,7)	(7,12)	(11,12)
3	(16,17), (18,19)	(16,17)	(17,18)	(18,19)

Figure 2: Top: Qubit layout for the `ibmq_boeblingen` quantum computer. Three different groups of qubits were selected to run the set of cycle benchmarking and TFIM computations described in this paper. Layout 1 refers to qubits [0, 1, 2, 3], Layout 2 refers to qubits [6, 7, 12, 11] and Layout 3 refers to qubits [16, 17, 18, 19]. The exhaustive set of four different paired CNOTs (four different cycles) were used to calculate the process infidelities on each Layout on the `ibmq_boeblingen`.processor

We ran a series of computations at three different preset time periods each day. Fig. 3 shows the daily sequence of calibrations and measurements. These computations and measurements were done over an 8 consecutive day time period in January 2021 while `ibmq_boeblingen` was in a dedicated reservation mode of operation. A dedicated reservation mode removes the hardware platform from general use and guarantees a user full dedicated access to the IBM hardware platform. This mode of operation allowed the qubit calibrations, computations and measurements to run uninterrupted, undisturbed by other users accessing these qubits during this time.

Each day there was a block of dedicated reserved time on `ibmq_boeblingen` from 4 am - 10 am and again from 3 pm - 11pm. Each morning at 4 am IBM did a full

re-calibration of all qubits and a 2 qubit re-calibration at 6 pm. After the 4 am and 6 pm IBM re-calibrations were completed we recorded `ibmq_boeblingen`'s backend properties. Appendix A summarizes the daily IBM re-calibration schedule and procedures.

At 4 a.m. each morning `ibmq_boeblingen` was placed in a dedicated reservation mode. Qubit re-calibrations, computations and "morning" measurements were done from 6 am to 10 am. The machine was then opened for general use from 10 am to 3 pm. At 3 pm the machine was closed to other users and again placed in dedicated reservation mode. A set of "afternoon" measurements were made on the qubits from 3 pm to 6 pm. The "afternoon" measurements were deliberately scheduled when the machine was closed to all other users but prior to the night IBM re-calibration. At 6 pm IBM did a two-qubit re-calibration. After the re-calibration was finished another set of backend properties were recorded and then a full set of "night" measurements were made from 8 pm - 11 pm.

Running these experiments three different times on each day (morning, afternoon, night) for each of the three different Layouts with two different Circuits over an eight day consecutive time period gave inter-day and intra-day measurements of processor performance for each layout plus measurements based on qubit choice and circuit structure that could then be compared.

The measurement were done using cycle benchmarking (CB) and randomized
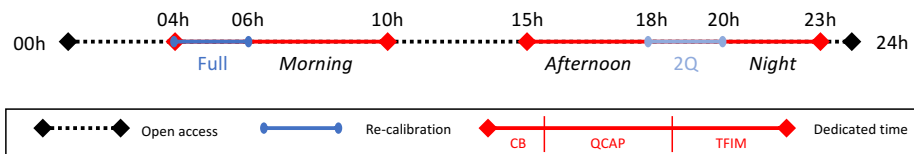
Figure 3: Daily timeline for calibrations, computations and measurements. The red lines denote the dedicated reservation time periods. The blue lines denote the time periods when IBM performed a full 1 and 2 qubit re-calibration each morning and a two-qubit (2Q) only re-calibration each night.

4. Results

This section reports on the measurements obtained from the cycle benchmarking and TFIM computations on the IBM `ibmq_boeblingen` processor during the 8 day time window in January 2021. The results reported here are unique in that the project had access to a very generous level of reserved run-time both in terms of dedicated access to the machine and length of contiguous dedicated time available for running detailed benchmarking computations. Having this level of reserved time assured that the IBM re-calibrations were always performed on a quiescent machine and that the "morning", "afternoon" and "night" CB, QCAP and TFIM computations fully ran without any other user's jobs being interleaved between the individual computations.

These types of environmental factors assured an optimal set of run-time conditions for measuring the RB, CB and TFIM quantities. Appendix B summarized the process infidelity and the QCAP bound computational procedure for analyzing the data. Using these procedures, the analysis of the data focused on four areas: inter-day and intra-day calibration drift of the qubits (Sec. 4.1 and 4.2), dependencies on qubit layouts for concurrent calculations (Sec. 4.3) and impacts from different circuit structure choices (Sec. 4.4).

4.1. Inter-day and Intra-Day Qubit Hardware Performance

Inter-day qubit drift was detected from analyzing data collected during the consecutive 8 day running period. An example illustrating this drift is seen by examining the data collected from Circuit 1, Layout 2 on January 24th and January 29th. For both the January 24th and January 29th data, the Pauli infidelities for each hard cycle are calculated for each Pauli decay term from the morning run on Layout 2 (qubits [6, 7, 12,11]) and plotted in Fig 4.

The RB two qubit error rates (r) that were recorded in the IBM backend properties after the completion of the IBM re-calibrations for those days were also included. The average error rate, r in standard RB assumes that the noise is gate-independent and $r \approx \epsilon$ where ϵ is the average gate-set infidelity. The average error rate can be expressed in terms of RB decay rate p as,

$$r \equiv \frac{d-1}{d}(1-p) \tag{6}$$

where $d = 2^n$ (n is the number of qubits) [59]. In this paper, we report the process infidelity, e_F, as a measure of infidelity for coherent errors. For completeness we convert average gate set fidelity, r, to the process infidelity, e_F using,

$$e_F = r\frac{d+1}{d} \tag{7}$$

from [60]. The overall process infidelity from the CB calculations and the RB measured process infidelities were then plotted versus the exhaustive set of two-qubit CNOT cycles

for Layout 2 measured on both the morning of January 24th and January 29th as shown in Fig. 5 **(a)**. The two-qubit error measurements are listed in Table C2 in Appendix C.

The QCAP bound as a function of number Trotter steps was then calculated using CB. Both the QCAP_{CB} and QCAP_{RB} results were plotted in Fig. 5 **(b)**. After this QCAP data was collected the TFIM Trotterization on Circuit 1 was run while the `ibmq_boeblingen` processor while the processor continued to be in dedicated mode. Fig. 5**(c)** shows the occupation numbers on the first site $\langle \hat{n}_1(t) \rangle$ as a function of Trotter step/time calculated as a function of time from the morning runs of layout 2 (qubits [6, 7, 12, 11]) on January 24, 2021 and January 29, 2021 compared to exact Trotter approximation.
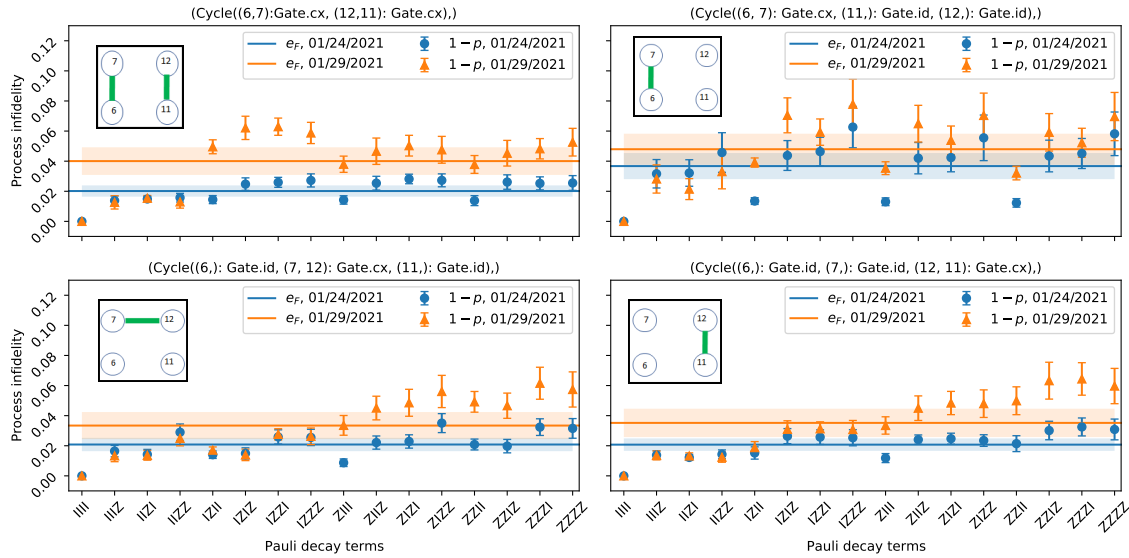

Figure 4: The Pauli infidelities for each hard cycle calculated for each Pauli decay term from morning run on qubits [6, 7, 12, 11] (layout 2) on days 01/24/2021 (blue lines and data points) and 01/29/2021 (orange lines and data points). The upper left graph is Cycle 1, the upper right is Cycle 2, lower left is Cycle 3 and the lower right is Cycle 4. The total process infidelity for each of the 4 different cycles is graphed on each plot as a blue and orange solid lines along with the shaded error bands for that day (24th in blue and 29th in orange). The shaded regions show the error on the process infidelity and the error bars on the markers show the statistical errors on Pauli decay terms.

Intra-day qubit drift was also detected from data collected during the consecutive 8-day running period. A similar procedure used to detect the inter-day qubit drift was also applied when analyzing data checking for intra-day drift. Examples illustrating this drift are seen by examining the data collected from the morning, afternoon and night runs for both January 27th and January 30th using Layout 2.

The RB average process infidelity was recorded based the reported backend property of the two-qubit error rate after the IBM full processor qubit re-calibration. The RB error rate was converted to an RB process infidelity in a similar procedure as was done
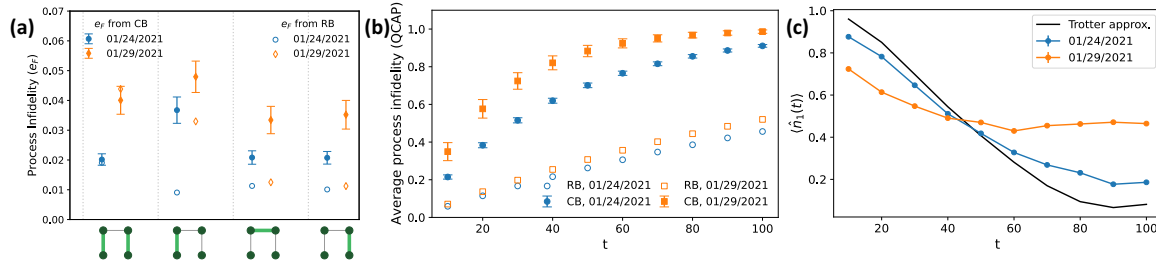
Figure 5: **Inter-day drift in error as characterized by RB and CB. (a)** The process infidelities calculated using RB and CB such that the two-qubit CNOT cycles 1, 2, 3, and 4 are identified along the horizontal axis from left to right, **(b)** The QCAP bound as a function of evolution time calculated using RB (QCAP$_{\text{RB}}$) and CB (QCAP$_{\text{CB}}$), **(c)** The particle number in site 1 calculated as a function of evolution time from morning run of layout 2 (qubits [6, 7, 12, 11]) on days 01/24/2021 and 01/29/2021 compared to exact Trotter approximation.

for the inter-day RB data.

The CB process infidelities for those days were also computed. The individual process infidelities for each CNOT pair and the overall process infidelity are shown in Fig. 6 thru Fig. 9.

The overall process infidelity from the CB calculations and the RB measured process infidelities were plotted versus the exhaustive set of two-qubit CNOT cycles for Layout 2 measured on the morning, afternoon and night of January 27th and January 30th. Figs. 10 **(a)** shows the CB versus RB process infidelity data and analysis and Fig. 10 **(b)** shows the graph of the QCAP$_{\text{CB}}$ and the QCAP$_{\text{RB}}$ values versus evolution time for the January 27th data. Similarly Figs. 10 **(c)** shows the CB versus RB process infidelity results and Figs. 10 **(d)** shows the graph of the QCAP$_{\text{CB}}$ and the QCAP$_{RB}$ values versus evolution time for the January 30th data.

4.2. Analysis of Inter-day and Intra-day Calibration Drift

We observed both inter-day and intra-day calibration drifts on the `ibmq_boeblingen` processor. The analysis of the inter-day data shows that both the RB and CB measurements indicate that the processor is substantially drifting from day to day despite both the daily IBM RB re-calibrations and even intra-day (morning and night) IBM re-calibrations of the processor. This inability of the processor to accurately and faithfully reproduce results within error bars from day-to-day measurements or even measurements recorded at different times within the same day on a totally quiet machine that is free from interference from other users has serious implications as to how users process and interpret their results from quantum computing computations run on these hardware platforms. Some of these implications include:
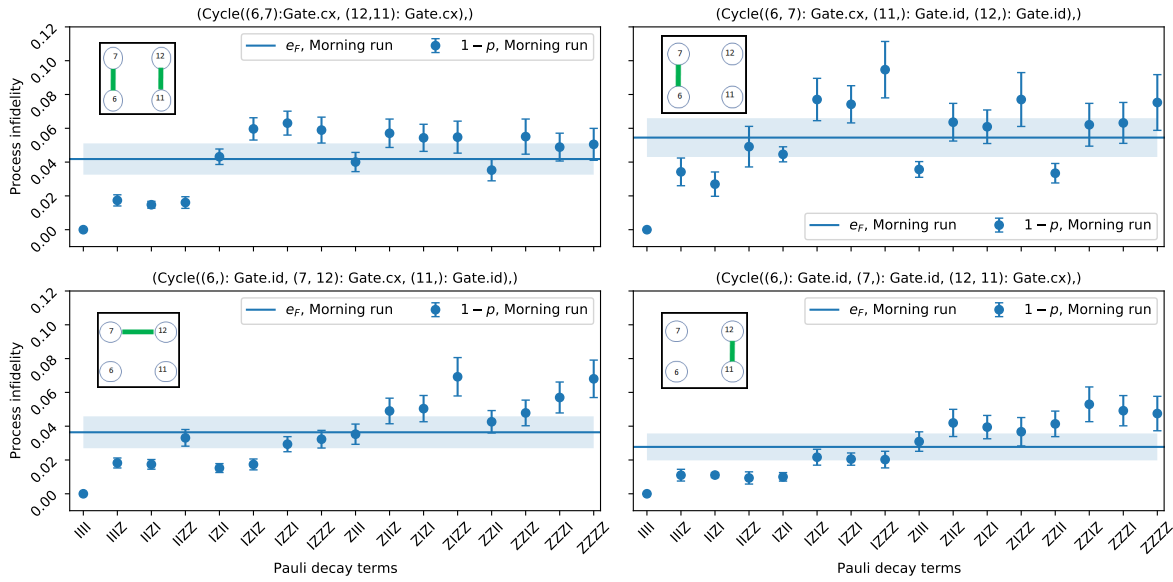
Figure 6: The Pauli infidelities for each hard cycle calculated for each Pauli decay term from morning run of layout 2 (qubits [6, 7, 12, 11]) on January 27, 2021. The error bars on each of the markers show the statistical errors on the Pauli decay terms. The total process infidelity for each of the 4 different cycles are plotted as blue solid lines along with the shaded error bands. The shaded regions show the error on the average process infidelity.
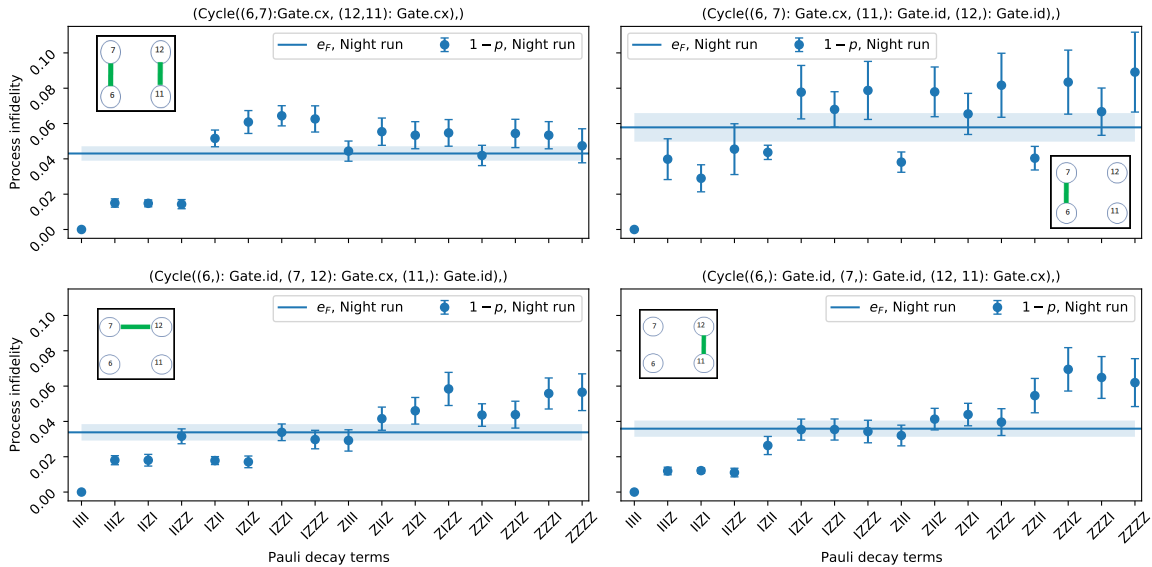

Figure 7: The Pauli infidelities for each hard cycle calculated for each Pauli decay term from night run of layout 2 (qubits [6, 7, 12, 11]) on January 27, 2021. The error bars on each of the markers show the statistical errors on the Pauli decay terms. The total process infidelity for each of the 4 different cycles are plotted as blue solid lines along with the shaded error bands. The shaded regions show the error on the average process infidelity

Figure 8: The Pauli infidelities for each hard cycle calculated for each Pauli decay term from morning run of layout 2 (qubits [6, 7, 12, 11]) on 01/30/2021. The error bars on each of the markers show the statistical errors on the Pauli decay terms. The total process infidelity for each of the 4 different cycles are plotted as blue solid lines along with the shaded error bands. The shaded regions show the error on the average process infidelity.

Figure 9: The Pauli infidelities for each hard cycle calculated for each Pauli decay term from night run of layout 2 (qubits [6, 7, 12, 11]) on day 01/30/2021. The error bars on each of the markers show the statistical errors on the Pauli decay terms. The total process infidelity for each of the 4 different cycles are plotted as blue solid lines along with the shaded error bands. The shaded regions show the error on the average process infidelity.
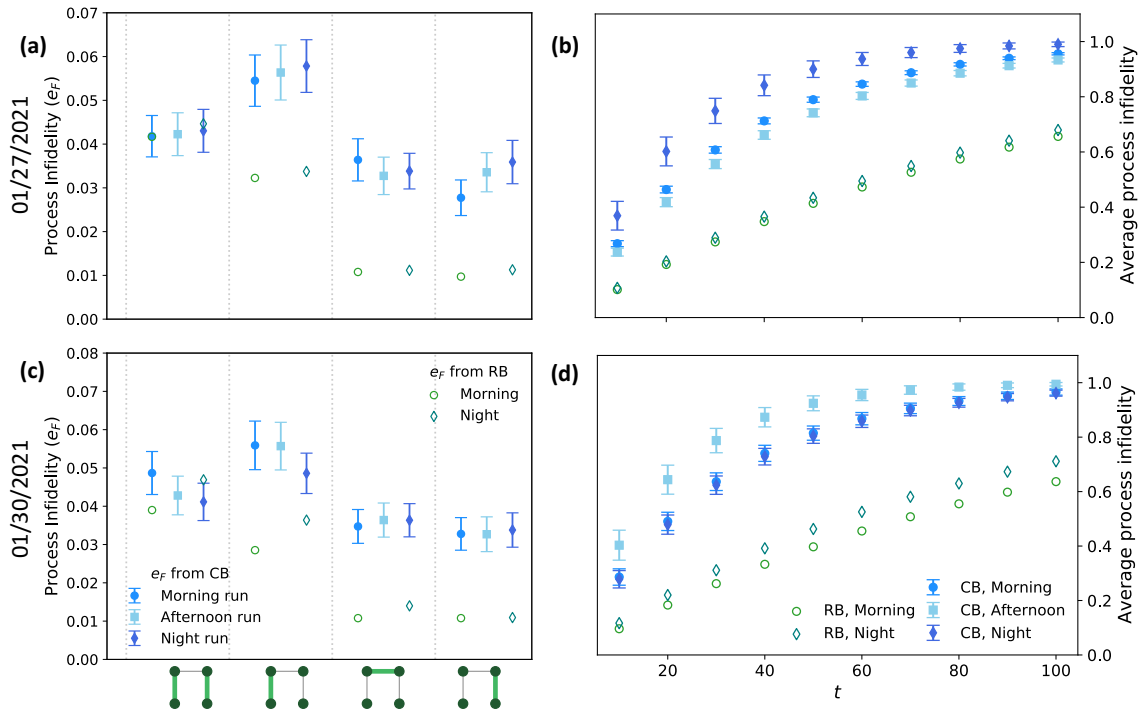
Figure 10: **Error characterization via CB vs. RB.** The process infidelities for cycles 1, 2, 3, and 4 calculated using RB and cycle CB from morning, afternoon and night runs of layout 2 [qubits 6, 7, 12, 11] on **(a)** January 27, 2021, **(c)** January 30, 2021. The QCAP bound as a function of evolution time calculated using RB ($QCAP_{RB}$) and CB ($QCAP_{CB}$) from morning, afternoon and night runs of layout 2 [qubits 6, 7, 12, 11] on **(b)** January 27, 2021, **(d)** January 30, 2021.

- Users cannot necessarily assume that the results from applications run on the same qubits on the same processor at the same time each day but on different days can simply be combined together. A similar concern is raised even when trying to aggregate results run at different times on the same day. Data recorded when the processor is drifting at these levels have potential statistics implications for users who try to simply aggregate their results from different runs and perform statistical fits on their aggregated data.

- The daily IBM one and two qubit re-calibrations of the processor consistently under reported the total magnitude of the qubit process infidelities. This under-reporting using RB data inflates the magnitude of the signal to noise ratio for data recorded on these processors. This leads to over-estimates of the total circuit coherence times and has direct implications for the signal to noise ratio justifying where to place the cutoff for the total number of measurements that can be included in a user's data analysis.

- As the QCAP values increase from zero to one, it represents a deterioration in the ability of the circuit to faithfully produce a correct set of measurements at each

Trotter step when run on this specific quantum computing hardware platform at two different dates and times. A QCAP value of greater than 0.5 is usually a qualitative indicator of the deterioration of a processor to faithfully and consistently reproduce results from a specific circuit versus circuit depth (number of Trotter steps).

(i) The two-qubit gates in a circuit are the main source of the errors in the output data. In terms of circuit depth (number of Trotter steps), the QCAP results using randomized benchmarking for the process infidelities show a far more optimistic scenario as to the number of data points that can actually be justified for inclusion in the user's data analysis as compared to the number of data points that can be justified based on the $QCAP_{CB}$ results. This has potential implications for analysis of the user's data and implications as to the accuracy of the results.

(ii) QCAP graphs are a measurement of the stability of the processor over a time interval. Comparison of QCAP bounds measured at different time intervals from applications run on the same qubits on the same processor at the same time each day but on different days or at different times on the same day should have similar results that fall within each graph's error bars. Figure 5 and Figure 10 clearly show that this is not the case and consequently results from these measurements cannot be simply combined because that processor is not even stable against drift in the time interval when these measurement were recorded.

4.3. Qubit Dependencies for Concurrent Calculations

Dependencies on qubit selection and their impact on concurrent calculations on a quantum processor were also investigated. To generate the data for these measurements identical copies of Circuit 1 were loaded onto qubits located in different physical areas of the hardware platform (Layout 1 and Layout 2) on a specific date and time during the consecutive 8 day running period.

An example showing this spatial qubit dependency can be seen from an examination of the January 27th morning run data for Circuit 1 on both Layout 1 (qubits [0, 1, 2, 3]) and Layout 2 (qubits [6, 7, 12, 11]). Fig. 11 shows the QCAP bound as a function of evolution time calculated using randomized benchmarking ($QCAP_{RB}$) and cycle benchmarking ($QCAP_{CB}$) from the morning run of circuit 1 on both layout 1 and layout 2 on January 27, 2021.

Several observations can be deduced from these graphs.

(i) The error bars from the Layout 1 and Layout 2 $QCAP_{CB}$ measurements do not overlap, indicating the reproducibility of Circuit 1 outputs run on different subsets of qubits in non-overlapping areas on the processor are not fully consistent. The figures also show the length of the time evolution that can indicate valid circuit measurements varies depending on the subset of qubits chosen to run the circuit.

(ii) QCAP values greater than .5 indicate that the circuit results are impacted by noise and that the results from the computations are less reliable. The $QCAP_{RB}$ shows a substantial difference when compared to $QCAP_{CB}$ measurements. $QCAP_{RB}$ will overestimate the length of the evolution time that the circuit will be able to produce useful data as compared to the length of time indicated from the $QCAP_{CB}$ measurements.

(iii) The $QCAP_{RB}$ measurements are not sensitive to the impacts of coherent errors and crosstalk affecting the results from Circuit 1 running on the different qubit Layouts.

These results indicate that simple parallel processing procedures borrowed from digital computing methods cannot be simply translated and implemented on today NISQ based quantum hardware platforms. The data clearly show that the individual circuit computations run on Layout 1 and Layout 2 cannot be directly or simply aggregated together because their results cannot be duplicated within the QCAP bound error bars for the computations from each circuit.
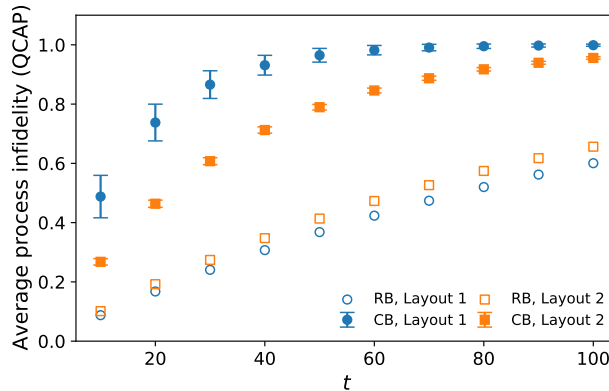

Figure 11: The figure shows the QCAP bound as a function of evolution time calculated using randomized benchmarking ($QCAP_{RB}$) and cycle benchmarking ($QCAP_{CB}$) from the morning run of circuit 1 on both layout 1 and layout 2 on January 27, 2021.

4.4. Circuit Structure Dependencies

A final set of experiments was performed to test the idea that the accumulation of coherent errors strongly depends on the details of the specific circuit layout being implemented. To demonstrate this effect we selected two equivalent sets of CNOT gates as shown in Fig. 1 (Circuit 1 and Circuit 2). Although both circuits equivalently represent the TFIM Hamiltonian, the number of steps per cycle for circuit 2 is greater than that for circuit 1 due to the difference in the layout of the CNOT gates.

The purpose of using two different layouts that produce the same physics was to investigate the quantum circuit structure dependencies. Data from the morning runs of layout 2 (qubits [6, 7, 12, 11]) on days 01/24/2021 and 01/29/2021 for Circuit 1 and Circuit 2 were selected for analysis. The CB $QCAP_{CB}$ bound as a function of evolution

time was calculated using only the set of CNOT gates for Circuit 1 and Circuit 2. The $QCAP_{RB}$ was computed based on the two-qubit backend properties published after the completion of the IBM re-calibrations. The measured values for both the $QCAP_{CB}$ and $QCAP_{RB}$ bounds for the CNOT hard cycles in circuit 1 and circuit 2 are plotted as a function of evolution time in Fig. 12.

These results are interesting when comparing these two different types of measurements. For both the January 24th and 29th morning run for Layout 2 data, the $QCAP_{CB}$ bound increases more rapidly than the $QCAP_{RB}$. There are 3 observations that can be deduced from examining Fig. 12.

(i) The $QCAP_{RB}$ gives a far too optimistic indication of the time evolution that can be considered for accumulating valid circuit measurements when compared to $QCAP_{CB}$. This type of result would not be detected if only considering individual CNOT gates and measuring their process infidelity using RB because such a result only depends on the number of CNOT gates and not on the details of the CNOT gate layout itself. The individual CNOT gates measured through randomized benchmarking $QCAP_{RB}$ calculated from Eq. (B.1) do not depend on the CNOT layout in the circuit, i.e. they only depend on the number of CNOT gates in the circuit. CB measurements are a better indicator to show how the ordering of the gates in the circuit impact the overall coherence times.

(ii) For CB results, the figure shows that the QCAP bound for circuit 2 deteriorates faster than circuit 1 because circuit 2 has a greater number of CNOTs for each cycle. This is a totally expected result due to the longer depth of Circuit 2 compared to Circuit 1.

(iii) When comparing the $QCAP_{CB}$ measurements between morning runs on Layout 2 for January 24th and 29th the error bars on these values do not overlap indicating that there is an inter-day CNOT qubit drift occurring and that the measured results from each Circuit are not within statistical consistency from one day to the next.

5. Observations and Next Steps

The results from the computations and measurements reported here illustrate the fragility of today's NISQ platforms. Several observations from our work here can be readily identified.

The pattern of errors documented in this research are arising from an inability to sufficiently control the coherent and systematic errors from each individual quantum circuit that arises when running on a specific quantum hardware platform. The two-qubit quantum gates (such as the CNOT) produce error rates that are usually an order of magnitude larger than single qubit gates.

The coherent errors arising from the under or over rotation of the qubits in response to control pulses significantly contribute to the total qubit error. This complexity in characterizing the qubit hardware errors indicates that it is not possible to characterize
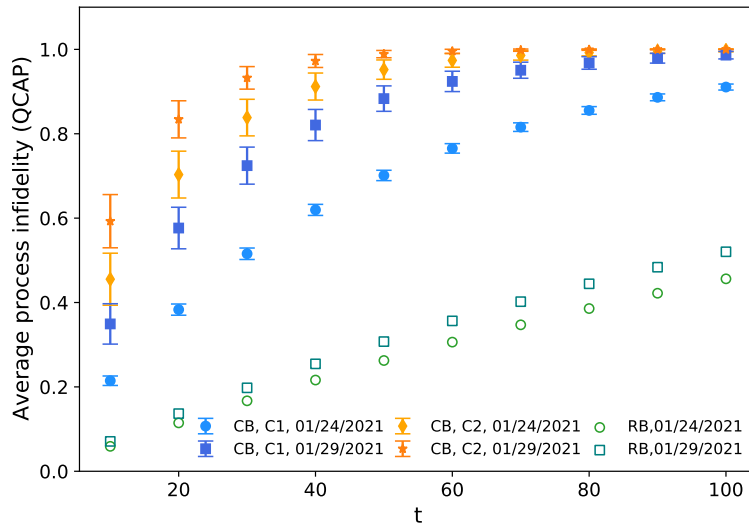
Figure 12: **Comparing two Trotter decomposition that gives the same physics.**
The QCAP bound as a function of evolution time calculated using RB (QCAP_{RB}) and
CB (QCAP_{CB}) from morning run of layout 2 (qubits [6, 7, 12, 11]) on days 01/24/2021
and 01/29/2021 for Circuit 1 and Circuit 2 with only CNOT gates used as hard cycles.
The plotted error bars only show the statistical error.

multi-qubit errors by a single parameter. These hardware error rates on a quantum
processor are a complex mix of the choice of qubits and choice of gates, the order
or sequence that the gates are being applied in the circuit and the local direct gate
combinations and surrounding spectator qubits.

Consequences of these coherent errors appear when running benchmark type
computations within a controlled hardware environment. These instabilities appear
as both inter-day and intra-day qubit drift compromising the consistency of quantum
computation results as well as physical qubit idiosyncrasies dependent on which physical
susbset set of qubits is used for the computations.

At the present time the number of qubits used for these computations is quite
modest and so the accuracy of the quantum computing results can always be
checked with digital computers. However, as the number of qubits used in quantum
computations over the next several years exceeds a few hundred, the hardware will
still not have a sufficient number of qubits to support logical error correction for these
computations. In addition, a direct verification of the quantum computer's output will
be far beyond the reach of any digital computer.

If the user community is to have any level of confidence in the published results from
running quantum circuits using several hundred qubits, there must be a demonstrated
ability to control the noise and errors on these systems. These mitigation strategies must
demonstrate that the measurements from applications run on these quantum hardware
platforms are sufficiently stable so that the results can be trusted with a high level of
confidence. Without such supporting evidence, these computations will be overshadowed

by the possibility that the data is tainted with some of the instabilities as documented in this report and will be considered unreliable.

We recognize that `ibmq_boeblingen` is an older generation quantum processor. IBM is well aware of these noise and coherent error issues and they have a large complex hardware and software effort underway to address these challenges [2],[61], [62], [63], [64], [65]. Although there has been progress with both the hardware and software since the release of the `ibmq_boeblingen` processor, the error mitigation protocols and methodologies summarized here are still applicable.

IBM has also recently introduced the IBM Qiskit Runtime Program to help streamline computations and minimize intra-day drift. In addition, other groups are investigating additional methods for improving the reliability of quantum processing computations [66, 67, 68, 69, 70] and comparisons among methods would be desirable.

This group has also begun several other in-depth follow-on projects to address other aspects concerning the integrity and validity of NISQ computations through robust characterization and stability analysis protocols. These include a detailed analysis of the impacts of coherent errors on magnon spectra measurements and real-time phase shifts. The results from these ongoing projects will be reported in future publications.

6. Acknowledgments

7. Author Contributions

P.D. led the team, designed the operational implementation, coordinated with IBM Research for dedicated access to the IBM quantum hardware platforms and was a major contributing author for this paper. KYA ran the cycle benchmarking simulations and constructed the CB and QCAP graphs and made substantive contributions to the text. ZP captured the daily IBM backend property data. ZP also ran the TFIM simulations and wrote the analysis software for post-processing and plotting the TFIM data. EG provided the orignal TFIM code and the physics model circuits. AN built the tables in the paper. AFK, RP and YM edited and reviewed the document.

8. Competing interests

The authors declare no competing interests

[1] Preskill J 2018 *Quantum* **2** 79 ISSN 2521-327X URL https://doi.org/10.22331/q-2018-08-06-79

[2] Temme K, Bravyi S and Gambetta J M 2017 *Phys. Rev. Lett.* **119**(18) 180509 URL https://link.aps.org/doi/10.1103/PhysRevLett.119.180509

[3] Li Y and Benjamin S C 2017 *Phys. Rev. X* **7**(2) 021050 URL https://link.aps.org/doi/10.1103/PhysRevX.7.021050

[4] Giurgica-Tiron T, Hindy Y, LaRose R, Mari A and Zeng W J 2020 *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)* URL http://dx.doi.org/10.1109/QCE49297.2020.00045

[5] Wallman J J and Emerson J 2016 *Phys. Rev. A* **94**(5) 052325 URL https://link.aps.org/doi/10.1103/PhysRevA.94.052325

[6] Hashim A, Naik R K, Morvan A, Ville J L, Mitchell B, Kreikebaum J M, Davis M, Smith E, Iancu C, O'Brien K P, Hincks I, Wallman J J, Emerson J and Siddiqi I 2021 (*Preprint* 2010.00215)

[7] Mohseni M, Rezakhani A T and Lidar D A 2008 *Physical Review A* **77** ISSN 1094-1622 URL http://dx.doi.org/10.1103/PhysRevA.77.032322

[8] Flammia S T, Gross D, Liu Y K and Eisert J 2012 **14** 095022 URL https://doi.org/10.1088/1367-2630/14/9/095022

[9] Chuang I L and Nielsen M A 1997 *Journal of Modern Optics* **44** 2455–2467 ISSN 1362-3044 URL http://dx.doi.org/10.1080/09500349708231894

[10] Mostame S and Schützhold R 2008 *Phys. Rev. Lett.* **101**(22) 220501 URL https://link.aps.org/doi/10.1103/PhysRevLett.101.220501

[11] Branderhorst M P A, Nunn J, Walmsley I A and Kosut R L 2009 **11** 115010 URL https://doi.org/10.1088/1367-2630/11/11/115010

[12] Merkel S T, Gambetta J M, Smolin J A, Poletto S, Córcoles A D, Johnson B R, Ryan C A and Steffen M 2013 *Physical Review A* **87** ISSN 1094-1622 URL http://dx.doi.org/10.1103/PhysRevA.87.062119

[13] Blume-Kohout R, Gamble J K, Nielsen E, Mizrahi J, Sterk J D and Maunz P 2013 Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit (*Preprint* 1310.4492)

[14] Scholten T L and Blume-Kohout R 2018 *New Journal of Physics* **20** 023050 ISSN 1367-2630 URL http://dx.doi.org/10.1088/1367-2630/aaa7e2

[15] Emerson J, Alicki R and Życzkowski K 2005 *Journal of Optics B: Quantum and Semiclassical Optics* **7** S347–S352 URL https://doi.org/10.1088/1464-4266/7/10/021

[16] Dankert C, Cleve R, Emerson J and Livine E 2009 *Physical Review A* **80** ISSN 1094-1622 URL http://dx.doi.org/10.1103/PhysRevA.80.012304

[17] Magesan E, Blume-Kohout R and Emerson J 2011 *Physical Review A* **84** ISSN 1094-1622 URL http://dx.doi.org/10.1103/PhysRevA.84.012309

[18] Erhard A, Wallman J J, Postler L, Meth M, Stricker R, Martinez E A, Schindler P, Monz T, Emerson J and Blatt R 2019 *Nature Communications* **10** ISSN 2041-1723 URL http://dx.doi.org/10.1038/s41467-019-13068-7

[19] Gong M, Chen M C, Zheng Y, Wang S, Zha C, Deng H, Yan Z, Rong H, Wu Y, Li S, Chen F, Zhao Y, Liang F, Lin J, Xu Y, Guo C, Sun L, Castellano A D, Wang H, Peng C, Lu C Y, Zhu X and Pan J W 2019 *Phys. Rev. Lett.* **122**(11) 110501 URL https://link.aps.org/doi/10.1103/PhysRevLett.122.110501

[20] McKay D C, Sheldon S, Smolin J A, Chow J M and Gambetta J M 2019 *Phys. Rev. Lett.* **122**(20) 200502 URL https://link.aps.org/doi/10.1103/PhysRevLett.122.200502

[21] Sundaresan N, Lauer I, Pritchett E, Magesan E, Jurcevic P and Gambetta J M 2020 *PRX Quantum* **1** ISSN 2691-3399 URL http://dx.doi.org/10.1103/PRXQuantum.1.020318

[22] Piltz C, Sriarunothai T, Varón A and Wunderlich C 2014 *Nature Communications* **5** ISSN 2041-1723 URL http://dx.doi.org/10.1038/ncomms5679

[23] Reagor M, Osborn C B, Tezak N, Staley A, Prawiroatmodjo G, Scheer M, Alidoust N, Sete E A,

Didier N, da Silva M P and et al 2018 *Science Advances* **4** eaao3603 ISSN 2375-2548 URL http://dx.doi.org/10.1126/sciadv.aao3603

[24] Sarovar M, Proctor T, Rudinger K, Young K, Nielsen E and Blume-Kohout R 2020 *Quantum* **4** 321 ISSN 2521-327X URL http://dx.doi.org/10.22331/q-2020-09-11-321

[25] Chow J M, Córcoles A D, Gambetta J M, Rigetti C, Johnson B R, Smolin J A, Rozen J R, Keefe G A, Rothwell M B, Ketchen M B and et al 2011 *Physical Review Letters* **107** ISSN 1079-7114 URL http://dx.doi.org/10.1103/PhysRevLett.107.080502

[26] Tripathi V, Chen H, Khezri M, Yip K W, Levenson-Falk E M and Lidar D A 2021 (*Preprint* 2108.04530)

[27] Bialczak R C, Ansmann M, Hofheinz M, Lucero E, Neeley M, O'Connell A D, Sank D, Wang H, Wenner J, Steffen M and et al 2010 *Nature Physics* **6** 409–413 ISSN 1745-2481 URL http://dx.doi.org/10.1038/nphys1639

[28] Zhao P, Lan D, Xu P, Xue G, Blank M, Tan X, Yu H and Yu Y 2021 *Physical Review Applied* **16** ISSN 2331-7019 URL http://dx.doi.org/10.1103/PhysRevApplied.16.024037

[29] Long J, Zhao T, Bal M, Zhao R, Barron G S, sheng Ku H, Howard J A, Wu X, McRae C R H, Deng X H, Ribeill G J, Singh M, Ohki T A, Barnes E, Economou S E and Pappas D P 2021 A universal quantum gate set for transmon qubits with strong zz interactions (*Preprint* 2103.12305)

[30] Clark C R, Metodi T S, Gasster S D and Brown K R 2009 *Phys. Rev. A* **79**(6) 062314 URL https://link.aps.org/doi/10.1103/PhysRevA.79.062314

[31] Yeter-Aydeniz K, Siopsis G and Pooser R C 2021 *New Journal of Physics* **23** 043033 URL https://doi.org/10.1088/1367-2630/abe63d

[32] Lamm H and Lawrence S 2018 *Phys. Rev. Lett.* **121** 170501 (*Preprint* 1806.06649)

[33] Cervera-Lierta A 2018 *Quantum* **2** 114 ISSN 2521-327X URL https://doi.org/10.22331/q-2018-12-21-114

[34] Kandala A, Temme K, Córcoles A D, Mezzacapo A, Chow J M and Gambetta J M 2019 *Nature* **567** 491–495 ISSN 1476-4687 URL http://dx.doi.org/10.1038/s41586-019-1040-7

[35] Somma R D 2016 (*Preprint* 1503.06319)

[36] Johanning M, Varón A F and Wunderlich C 2009 *Journal of Physics B: Atomic, Molecular and Optical Physics* **42** 154009 ISSN 1361-6455 URL http://dx.doi.org/10.1088/0953-4075/42/15/154009

[37] Smith A, Kim M S, Pollmann F and Knolle J 2019 *npj Quantum Information* **5** ISSN 2056-6387 URL http://dx.doi.org/10.1038/s41534-019-0217-0

[38] Vovrosh J and Knolle J 2020 (*Preprint* 2001.03044)

[39] Salathé Y, Mondal M, Oppliger M, Heinsoo J, Kurpiers P, Potočnik A, Mezzacapo A, Las Heras U, Lamata L, Solano E and et al 2015 *Physical Review X* **5** ISSN 2160-3308 URL http://dx.doi.org/10.1103/PhysRevX.5.021027

[40] Bernien H, Schwartz S, Keesling A, Levine H, Omran A, Pichler H, Choi S, Zibrov A S, Endres M, Greiner M, Vuletić V and Lukin M D 2017 *Nature* **551** 579–584

[41] Zhang J, Pagano G, Hess P W, Kyprianidis A, Becker P, Kaplan H, Gorshkov A V, Gong Z X and Monroe C 2017 *Nature* **551** 601–604 (*Preprint* 1708.01044)

[42] Labuhn H, Barredo D, Ravets S, de Léséleuc S, Macrì T, Lahaye T and Browaeys A 2016 *Nature* **534** 667–670 ISSN 1476-4687 URL http://dx.doi.org/10.1038/nature18274

[43] Chertkov E, Bohnet J, Francois D, Gaebler J, Gresh D, Hankin A, Lee K, Tobey R, Hayes D, Neyenhuis B, Stutz R, Potter A C and Foss-Feig M 2021 Holographic dynamics simulations with a trapped ion quantum computer (*Preprint* 2105.09324)

[44] Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow J M and Gambetta J M 2017 *Nature* **549** 242 EP – URL https://doi.org/10.1038/nature23879

[45] You H, Geller M R and Stancil P C 2013 *Phys. Rev. A* **87**(3) 032341 URL https://link.aps.org/doi/10.1103/PhysRevA.87.032341

[46] Hebenstreit M, Alsina D, Latorre J I and Kraus B 2017 *Phys. Rev. A* **95**(5) 052339 URL https://link.aps.org/doi/10.1103/PhysRevA.95.052339

[47] Schauss P 2018 *Quantum Science and Technology* **3** 023001 URL https://doi.org/10.1088% 2F2058-9565%2Faa9c59

[48] Flammia S T and Wallman J J 2020 *ACM Transactions on Quantum Computing* **1** ISSN 2643-6809 URL https://doi.org/10.1145/3408039

[49] Keesling A *et al.* 2019 *Nature* **568** 207–211 (*Preprint* 1809.05540)

[50] Kim Y, Wood C J, Yoder T J, Merkel S T, Gambetta J M, Temme K and Kandala A 2021 *arXiv e-prints* arXiv:2108.09197 (*Preprint* 2108.09197)

[51] Gustafson E, Zhu Y, Dreher P, Linke N M and Meurice Y 2021 *Phys. Rev. D* **104** 054507 (*Preprint* 2103.06848)

[52] Gustafson E, Dreher P, Hang Z and Meurice Y 2021 *Quantum Science and Technology* **6** 045020 URL https://doi.org/10.1088/2058-9565/ac1dff

[53] Gustafson E, Meurice Y and Unmuth-Yockey J 2019 *Physical Review D* **99**

[54] Lloyd S 1996 *Science* **273** 1073–1078 ISSN 0036-8075 URL http://science.sciencemag.org/content/273/5278/1073

[55] Meurice Y 2021 *Quantum field theory: a quantum computation approach* (Bristol: Institute of Physics Publishing)

[56] Childs A M, Su Y, Tran M C, Wiebe N and Zhu S 2021 *Phys. Rev. X* **11**(1) 011020 URL https://link.aps.org/doi/10.1103/PhysRevX.11.011020

[57] Layden D 2021 *arXiv e-prints* arXiv:2107.08032 (*Preprint* 2107.08032)

[58] Eastin B and Flammia S T 2004 (*Preprint* arXiv:quant-ph/0406003v2) URL https://arxiv.org/abs/quant-ph/0406003

[59] Qi J and Ng H K 2019 *International Journal of Quantum Information* **17** 1950031 ISSN 1793-6918 URL http://dx.doi.org/10.1142/S021974991950031X

[60] Nielsen M A 2002 *Physics Letters A* **303** 249–252 ISSN 0375-9601 URL https://www.sciencedirect.com/science/article/pii/S0375960102012720

[61] McKay D C, Cross A W, Wood C J and Gambetta J M 2020 Correlated randomized benchmarking (*Preprint* 2003.02354)

[62] Bravyi S, Sheldon S, Kandala A, Mckay D C and Gambetta J M 2021 *Physical Review A* **103** ISSN 2469-9934 URL http://dx.doi.org/10.1103/PhysRevA.103.042605

[63] Ware M, Johnson B R, Gambetta J M, Ohki T A, Chow J M and Plourde B L T 2019 Cross-resonance interactions between superconducting qubits with variable detuning (*Preprint* 1905.11480)

[64] Nation P D, Kang H, Sundaresan N and Gambetta J M 2021 *PRX Quantum* **2** ISSN 2691-3399 URL http://dx.doi.org/10.1103/PRXQuantum.2.040326

[65] Kim Y, Wood C J, Yoder T J, Merkel S T, Gambetta J M, Temme K and Kandala A 2021 Scalable error mitigation for noisy quantum circuits produces competitive expectation values (*Preprint* 2108.09197)

[66] Lin J, Wallman J J, Hincks I and Laflamme R 2021 *Phys. Rev. Research* **3**(3) 033285 URL https://link.aps.org/doi/10.1103/PhysRevResearch.3.033285

[67] Lin J, Wallman J J, Hincks I and Laflamme R 2021 *Physical Review Research* **3** ISSN 2643-1564 URL http://dx.doi.org/10.1103/PhysRevResearch.3.033285

[68] He A, Nachman B, de Jong W A and Bauer C W 2020 *Physical Review A* **102** ISSN 2469-9934 URL http://dx.doi.org/10.1103/PhysRevA.102.012426

[69] Pascuzzi V R, He A, Bauer C W, de Jong W A and Nachman B 2021 Computationally efficient zero noise extrapolation for quantum gate error mitigation (*Preprint* 2110.13338)

[70] Urbanek M, Nachman B, Pascuzzi V R, He A, Bauer C W and de Jong W A 2021 *Physical Review Letters* **127** ISSN 1079-7114 URL http://dx.doi.org/10.1103/PhysRevLett.127.270502

[71] Ville J L, Morvan A, Hashim A, Naik R K, Mitchell B, Kreikebaum J M, O'Brien K P, Wallman J J, Hincks I, Emerson J, Smith E, Younis E, Iancu C, Santiago D I and Siddiqi I 2021 (*Preprint* 2104.08785)

9. Appendix

Appendix A. Daily `ibmq_boeblingen` Qubit Re-Calibration Schedule

For this specific project IBM agreed to supply our team with approximately 140 hours of dedicated reservation time and to follow an agreed upon customized calibration schedule.

The customized schedule for dedicated time included a period in the morning from 4 am until 10 am and again in the afternoon from 3 pm until 11 pm. The complete `ibmq_boeblingen` re-calibration for both single and two-qubit gate gates was scheduled at 4:00 am ET, the beginning of the morning dedicated reservation time. A second re-calibration for only two-qubit gates ran at 6:00 pm ET, approximately 3 hours into the afternoon dedicated reservation time. The calibration jobs took approximately an hour and a half to complete. Our team executed no external jobs on the device during the calibration process, allowing the calibration jobs to run without interference.

The single-qubit calibration process consisted of Ramsey and Rabi experiments to measure the frequency and amplitude of each qubit along with calibration of the optimal scaling factor of the Derivative Removal by Adiabatic Gate (DRAG) pulse used in single-qubit gates on superconducting hardware. The T1/T2 coherence times and measurement error rates of each qubit were also measured and recorded. Randomized benchmarking of the single-qubit gates was then performed in batches of non-adjacent qubits. The two-qubit calibration process was done in a similar manner. Calibration of the amplitude and phase of each pulse was completed before performing randomized benchmarking in batches of well-separated gates of similar length in order to measure the average gate fidelities. Each time the `ibmq_boeblingen` quantum computing hardware platform was re-calibrated and benchmarked, IBM published and made these backend properties available through Qiskit, the open-source quantum software development kit.

Appendix B. Cycle Benchmarking and Quantum Capacity (QCAP) Protocols

This appendix summarizes both cycle benchmarking and quantum capacity protocols and their True-Q software implementation and parameter settings used for the computations reported in this paper.

Appendix B.1. Cycle Benchmarking

Cycle benchmarking (CB) is a scalable noise characterization protocol that was selected to identify local and global errors across multi-qubit quantum processors. The CB protocol can measure errors such as process infidelity containing any combination of single gates, two-qubit gates, and idle qubits, across an entire quantum device. CB helps keep track of each twirling gate and makes the process scalable with the number of qubits [71].

This protocol has the feature that the number of measurements required to estimate the process fidelity to a fixed precision is approximately independent of the number of qubits and is also insensitive to State Preparation and Measurement (SPAM) errors. Robustness to SPAM is very important characteristic because these type of errors can dominate the gate error measurement.

The CB protocol is presented in detail in reference [18] and is schematically represented in Figure B1. In CB, a gate cycle is an arbitrary set of native operations that act on a quantum register within a single clock cycle of time. Furthermore, within the CB protocol, there is a distinction between operations that can be physically implemented with relatively small and large amounts of noise, respectively called 'easy' and 'hard' gate cycles.

The box on the left hand side of the figure shows the CB protocol "dressing" a primitive gate cycle of interest (represented by $\tilde{\mathcal{G}}$) by composing the cycle with independent, random n-qubit Pauli operators in such a way that the effective logical circuit remains unchanged. In the figure the block $\tilde{\mathcal{G}}$ represents the noisy implementation of the gate(s) being measured in the circuit. The blocks $\tilde{\mathcal{R}}_{i,j}$ are random Paulis represented by the jth tensor factor of the ith gate inserted into the cycle to create an effective Pauli channel for the gate $\tilde{\mathcal{G}}$ being measured. The blocks $\tilde{\mathcal{B}}$ and $\tilde{\mathcal{B}}^{\dagger}$ represent basis changing operations connected with controlling SPAM errors.

CB decouples state preparation and measurement errors from the process fidelity estimation of a particular gate cycle by applying the noisy, dressed cycle to the system m number of times, (called the sequence length), and extracting the process fidelity from the average decay rate as a function of this sequence length. This Pauli twirling of gate cycles map coherent errors into stochastic Pauli errors, which are then measured in the prepared eigenstates of the Pauli basis set.

This is represented by the top box in the center of the figure showing all measured Pauli decay expectation values plotted as a function of the gate sequence length. In practice, this process is computed using at least three distinct gate sequence lengths. Each measurement sequence produces an exponential decay of the expectation value versus the sequence length. Taken together, the set of exponential decays of the form Ap^m can be fit to the cycle of interest as a function of the circuit depth for each basis preparation state.

Using the fitted exponential decay, the individual process infidelity for each Pauli Decay term e_F can be measured as shown in the box on the right hand side of the figure. An average process infidelity and error for the particular cycle $\tilde{\mathcal{G}}$ is calculated and is represented by the solid line and shaded band on the graph.

For our project in order to measure the error characterization associated with the two qubit gates in the TFIM circuits, the cycle benchmarking (CB) protocol was implemented using the True-Q software package. This package included a function *make_cb* that can produce quantitative measurements showing the effect of global and local error mechanisms affecting different primitive cycle operations of interest using CB.

The *make_cb* in True-Q uses a set of input parameters for the calculation. The first parameter is the cycle of interest. The second parameter sets how many times to apply the dressed cycle to observe the decay of the expectation values. Here, dressed cycle is the term that is used for denoting the target cycle preceded by a cycle of random elements of the twirling group. The number of random cycles needs to be chosen carefully such that it leads to exponential decay is evident and the fidelity can be accurately estimated.

The third parameter in the function is the number of circuits for each circuit length determined in the second parameter, i.e. number of random cycles. The last parameter in the function is the number randomly chosen Pauli decay strings. One can also specify the twirling group to be used that will be used in the process to automatically instantiate a twirl based on the labels in the given cycles. The supported twirling groups in True-Q software are Pauli, Clifford, unitary and identity. The software also offers initializing a twirl with single-qubit Cliffords. After the circuits are generated using this function the expectation values of the Pauli operators are calculated which then gives the process infidelity for the cycle of interest by using an exponential fit to the decay of the expectation values.

The Clifford (C1) gates for the hard gate twirling were selected to minimize the computation time so that they would complete within the morning and night dedicated time windows available on `ibmq_boeblingen`. The C1 twirling used random single qubit Cliffords which had the effect of symmeterizing the X, Y and Z noise. This ultimately allowed for an analysis of the depolarization error, which is one of the simplest of the systematic errors to measure and study.

To calculate the contribution of each of the Pauli decay terms to the average process infidelity e_F, C1 twirling was done using gate sequence circuit lengths of 2, 10, and 22 random Clifford gates were applied to each of the different pair combinations of CNOTs. Here, the sequence length refers to the number of times the cycle of interest appears apart from state inversion. We used 48 random circuits in each sequence length, and 128 shots. The combination of the CNOT gate being measured and the sequence of random Cliffords defines a dressed cycle of the CNOT gates being measured. For each three different circuit lengths the expectation values were calculated for all 16 of the Pauli decay terms. From these expectation values, fits to the exponential decay Ap^m (SPAM parameter A and the decay parameter p) are calculated for each Pauli decay term.

Individual process infidelity measurement were recorded for every CNOT pair for each of the three different qubit layouts on `ibmq_boeblingen` device as shown in Fig. 2. For example, on Layout 1 measurements included all of the combination of two-qubit cycles ([0, 1 and 2, 3], [0, 1], [1, 2] and [2, 3]). Similar measurements were taken on the CNOT cycles for Layouts 2 and 3. Hence, there are four cycles studied for each qubit layout

The average process infidelity of the dressed cycle for that CNOT pair was computed based on the calculated values of each of the Pauli decay terms. Both the individual process infidelity and average process infidelity measurements were computed and used

in the stability analysis of the qubits on the `ibmq_boeblingen` processor. The individual process infidelities for each CNOT pair and the overal process infidelity are shown in Fig 4, Fig 6, Fig 7, Fig 8 and Fig 9.
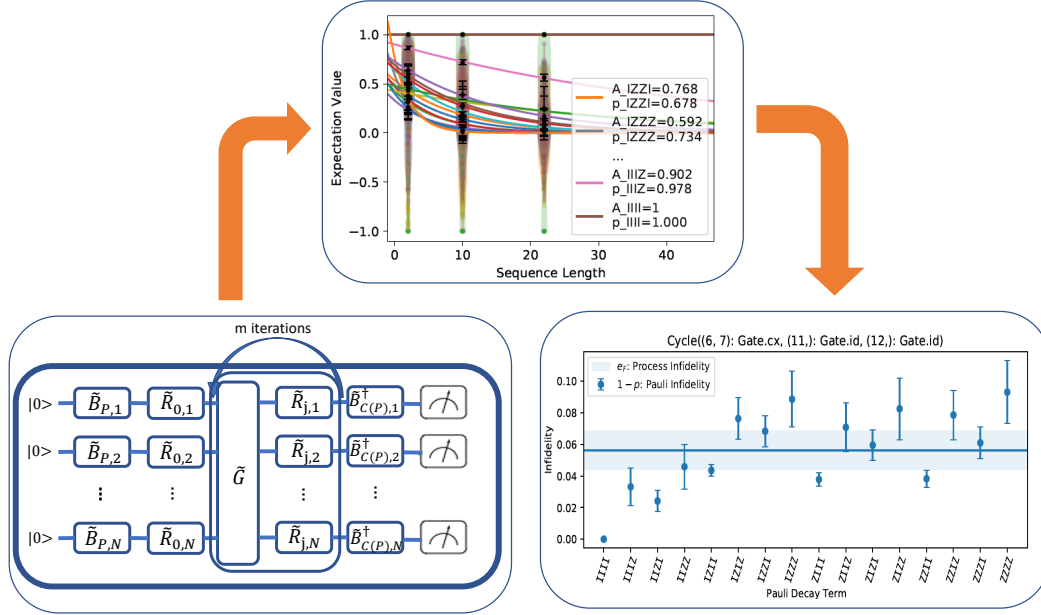

Figure B1: Block diagram of the CB protocol implementation.

Appendix B.2. Quantum Capacity Bound

The QCAP protocol was used for comparing the measured performance of a circuit that is loaded onto a quantum computing hardware processor to the measurement of an equivalent idealized version of that same circuit. A value of "0" for a QCAP result means that the circuit being tested is identical to its idealized equivalent whereas a QCAP value of "1" implies that the circuit being measured has no equivalent performance characteristics to its idealized equivalent. An increase in the QCAP bound as a function of evolution time is a measure as to how many time evolution steps can be included in a result before the signal being measured is overcome by noise in the circuit.

For the actual QCAP measurements the *make_qcap* and *qcap_bound* functions in the True-Q software were used to obtain an equivalent bound on the performance of a circuit as if it were computed using randomized compiling for calculating the process infidelity of the entire circuit of interest. The parameters to generate the collection of circuits for the quantum capacity bound *make_qcap* function use similar parameters as *make_cb* function, i.e. the circuit of interest, a list for the number of random cycles, number of circuits for each random cycle, and total number of randomly chosen Pauli decay strings. After generation of quantum circuits, these circuits are embedded into *qcap_bound* as

well as the circuit of interest to return a bound on the circuit performance. In this particular project, due to limited access to the dedicated mode on `ibmq_boeblingen` device we used sequence lengths of 4 and 16. The number of random circuits in this case is 30 and each of these circuits were run $N_{\text{shots}} = 128$.

We selected three separate groups of qubits on the `ibmq_boeblingen` hardware platform as shown in Fig. 2 to study the error characterization of TFIM Trotterization circuits using CB. The quantum circuit for evolution under the TFIM Hamiltonian has three pairs of two-qubit CNOT gates (c.f. Fig. 1). For the QCAP computation we selected circuit 1 shown in Fig 1 in order to compare to previous TFIM measurements [52]. We computed an estimate to the QCAP bound of the circuit 1 CNOT cycles in the TFIM Trotterization quantum circuits as a function of the number of Trotter steps. We also calculated the QCAP bound from the CNOT error rates reported by IBM using RB. To this end, we used the expression for the relationship between the average process fidelity and the average gate fidelity as seen in Eq. (6). For a quantum circuit with N CNOT gates (Eq. B.1) the QCAP bound is calculated using CNOT error rates provided by IBM.

$$\text{QCAP}_{\text{RB}} = 1 - \prod_{i=1}^{N} \left(1 - \frac{d+1}{d} r_i \right) \ . \tag{B.1}$$

The QCAP bound versus step size was then plotted as the average process infidelity (QCAP bound) as a function of number of Trotter steps (as a function of time). From this graph the performance of the circuit implemented on the set of specific qubits on that specific hardware platform can be measured over time.

Appendix C. Tables

				IBM Backend	U_2	U_3
	Qubits	$\mathbf{T_1}(\mu s)$	$\mathbf{T_2}(\mu s)$	**Readout Error** $(\times 10^{-2})$	$(\times 10^{-4})$	$(\times 10^{-4})$
	6	67.1	99.9	2.54	2.87	5.74
01/24/2021	7	94.8	86.8	2.30	3.05	6.71
Morning Run	12	97.5	88.5	3.13	2.91	5.82
	11	95.1	71.6	3.55	4.44	8.88
	6	24.1	4.97	9.19	25.0	50.0
01/29/2021	7	78.8	103.4	2.84	3.97	7.94
Morning Run	12	80.8	114.6	3.47	3.49	6.98
	11	48.4	87.8	3.22	4.64	9.29

Table spanning header: Inter-Day, Layout 2, Single-Qubit Errors

Table C1: T_1, T_2 values, readout errors and single-qubit errors for basis gates U_2 and U_3 for Layout 2 (qubits [6,7,12,11]) extracted from the recorded IBM back-end properties immediately after IBM completed a full re-calibration of the `ibmq_boeblingen` quantum chip on the morning of January 24, 2021 and January 29, 2021.

	Cycle	**Qubits**	**Cycle Benchmark** $(\times 10^{-2})$	**IBM Backend** $(\times 10^{-2})$
01/24/2021	2	[6,7]	3.67	0.908
Morning Run	3	[7,12]	2.08	1.128
	4	[12,11]	2.07	1.010
01/29/2021	2	[6,7]	4.79	3.30
Morning Run	3	[7,12]	3.34	1.25
	4	[12,11]	3.52	1.124

Table spanning header: Inter-Day, Layout 2, Two-Qubit Process Infidelities

Table C2: Values for two-qubit process infidelities for qubit pairs [6, 7], [7,12], [12, 11] extracted from the recorded IBM back-end properties immediately after IBM completed a full re-calibration of the `ibmq_boeblingen` quantum chip on the morning of January 24, 2021 and January 29, 2021 and the process infidelity values for the two-qubit pairs from cycle 2, 3, and 4 from the cycle benchmarking computations.

Layout 2 Intra-Day Two-Qubit Process Infidelities				
	Cycle	**Qubits**	**Cycle Benchmark** $(\times 10^{-2})$	**IBM Backend** $(\times 10^{-2})$
01/27/2021 Morning Run	2	[6,7]	5.45	3.23
	3	[7,12]	3.64	1.08
	4	[12,11]	2.77	0.98
01/27/2021 Night Run	2	[6,7]	5.78	3.38
	3	[7,12]	3.38	1.11
	4	[12,11]	3.59	1.12
01/30/2021 Morning Run	2	[6,7]	5.59	2.85
	3	[7,12]	3.47	1.08
	4	[12,11]	3.52	1.06
01/30/2021 Night Run	2	[6,7]	4.86	3.64
	3	[7,12]	3.63	1.40
	4	[12,11]	3.28	1.09

Table C3: Values for two-qubit process infidelities for qubit pairs [6, 7], [7,12], [12, 11] extracted from the recorded IBM back-end properties immediately after IBM completed full re-calibrations of the `ibmq_boeblingen` quantum chip on the morning of January 27th and 30th 2021 and after the two-qubit re-calibrations at night on January 27th and 30th 2021. The table also lists the process infidelities for the two-qubit pairs from cycle 2, 3, and 4 obtained from the cycle benchmarking computations for those same time periods.