




Measuring qubit stability in a gate-based NISQ hardware processor

Kübra Yeter-Aydeniz^{1,7} · Zachary Parks² · Aadithya Nair Thekkiniyadath² · Erik Gustafson^{3,8} · Alexander F. Kemper⁴ · Raphael C. Pooser⁵ · Yannick Meurice³ · Patrick Dreher⁶ 

Received: 15 February 2022 / Accepted: 27 December 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Some of the most problematic issues that limit the implementation of applications on noisy intermediate-scale quantum 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.

Keywords Analysis of qubit stability · Cycle Benchmarking protocol · Quantum Capacity · Transverse-field Ising model

This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC0500OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for the United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan.

✉ Patrick Dreher
padreher@ncsu.edu

Extended author information available on the last page of the article

1 Introduction

Researchers and application developers today 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.

There have been advances in quantum computing hardware development over the past few years, and the number of qubits on a processor is projected to grow to several thousand within the next few years (for example, the IBM Hardware Roadmap [2]). Although the number of qubits will increase, these processors will still be NISQ type devices. Many of the gates today built from this NISQ hardware do not operate sufficiently below the error correction threshold. As a result, the number of levels of code concatenation that are required for a given circuit depth will determine the number of qubits required to construct “logical qubits” that can compensate and correct for noise in these systems.

Nevertheless, there are error mitigation protocols that can be employed to reduce the errors and improve the coherence times for computations implemented on these platforms. Understanding the different types of errors, such as coherent and incoherent errors, and developing methods to mitigate them are a critical area of research essential for advancing quantum information science.

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 extrapolate the signal output back to a zero-noise limit. This has been extensively studied over the past few years [3–5]. Coherent errors arise from multiple sources. These may include cross-talk, global external fields and unwanted qubit-qubit interactions. These errors have been quantified in several recent publications [6–8].

Developing a full characterization of these errors on quantum hardware platforms is a challenge because doing so requires quantum process tomography [9–13] or gate set tomography [14–16], both of which require resources that grow exponentially with the number of qubits. Alternatives to such an exhaustive option are process fidelity measurements that can be obtained from randomized benchmarking (RB) [17–19] and cycle benchmarking (CB) [20]. The combination of using cycle benchmarking with randomized compiling (RC) [6] is an especially powerful method to address the issue of coherent errors [1, 21–32]. The randomized compiling procedure converts coherent errors in a circuit into stochastic noise. This reduces the magnitude of the unpredictable coherent errors and enables estimates of algorithm performance to be measured using the cycle benchmarking protocol.

Using the IBM quantum `ibmq_boeblingen` processor, we did an initial examination of the processor's re-calibration data and backend properties for both single and two qubits. 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 1 in Appendix C. 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, 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 in several sections. In Sect. 2, we describe the transverse-field Ising model used to study the error characterization in an IBM superconducting transmon platform. Section 3 discusses the methodology and conditions that were implemented using this model and the types of data that were collected. Section 4 summarizes the results of the computations using these 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, Sect. 5 summarizes our observations and discusses next steps.

2 Physics model

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

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

as an example application for studying qubit stability in an IBM quantum computing hardware platform (`ibmq_boeblingen`). 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 [3, 4, 12, 32–52]

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 [53–55].

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

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

Following Refs. [54–56], the Trotter approximation is applied to the evolution operator with the explicit form:

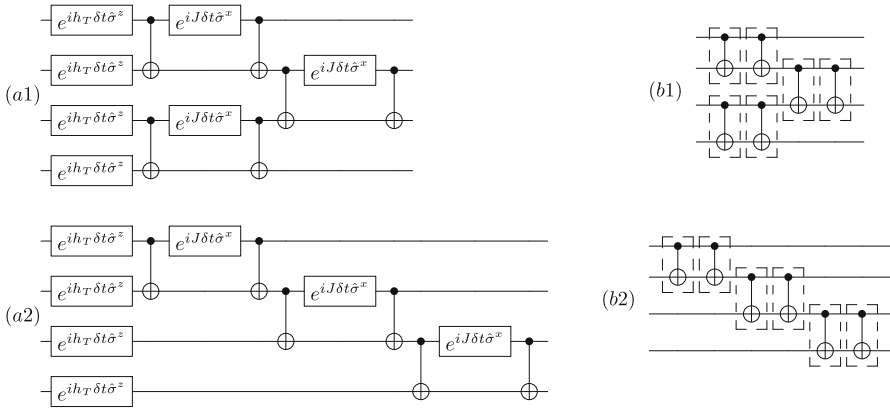


Fig. 1 Left side **a1**, **a2**: Quantum circuits for one Trotter step of the time evolution with the open boundary condition Ising model Hamiltonian. We define the quantum circuit in **a1** and **a2** as Circuit 1 and Circuit 2, respectively. Right side **b1**, **b2** CNOT hard cycles used at each Trotter step for Circuit 1 (**b1**) and Circuit 2 (**b2**). The quantum circuits were drawn using the Q-circuit package [60]

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

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 [54, 57] 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 [58, 59].

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 (labeled as Circuit 1 and Circuit 2). These two quantum circuits are equivalent combinations because the gates representing 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 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) [20] and the Quantum Capacity (QCAP).

We chose CB because it is a fully scalable protocol that can assess the performance of a specified combination of single-qubit, two-qubit or multi-qubit subset of gates within an n -qubit circuit when a clock cycle of operations is applied to a specific quantum register. By selecting a gate set of interest and combining that gate set with random Paulis, one can construct several sequence lengths. One can then measure the expectation values of the signal in these circuits as a function of these sequence lengths. From this information, one can determine the process infidelities for various Pauli decay terms and the overall process infidelity for this chosen gate set.

The QCAP measurement provides a bound on the performance of the full set of gates in a circuit under the process of randomized compiling. The QCAP gives a measure of the fidelity of the actual circuit in comparison with its idealized version. This measurement differs from CB because the QCAP measures the performance of the entire subset of gates under consideration within the overall n -qubit circuit.

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 [61]. 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 is implemented within the True-Q software in more detail.

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 of the combination of two-qubit cycles [(0, 1) and (2, 3), (0, 1), (1, 2) and (2, 3)].

We ran a series of computations at three different preset time periods each day. Figure 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

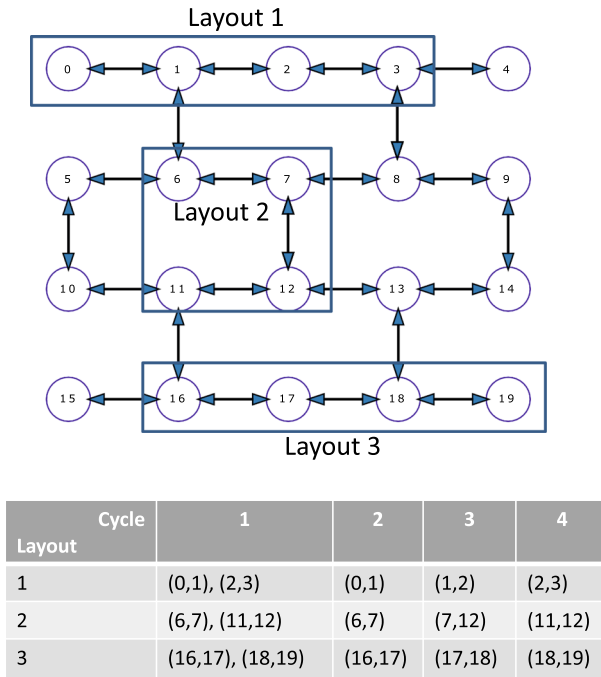


Fig. 2 Top: Qubit layout for the `ibmq_boeblingen` quantum computer. Three different groups of qubits were selected to run the set of cycle benchmarking and TFIM computations described in this paper. Layout 1 refers to qubits [0, 1, 2, 3], Layout 2 refers to qubits [6, 7, 12, 11] and Layout 3 refers 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`

(January 24th through January 31st) 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 dedicated access to the IBM hardware platform. This mode of operation allowed the qubit calibrations, computations and measurements to run uninterrupted and undisturbed by other users accessing these qubits during this time. This assured that the results would represent the response and performance of the system without external levels of interference.

Each day there was a block of dedicated reserved time on `ibmq_boeblingen` from 4 am to 10 am and again from 3 pm to 11 pm. Each morning at 4 am, IBM did a full re-calibration of all qubits and a two-qubit re-calibration at 6 pm. After the 4 am and 6 pm IBM re-calibrations were completed, we recorded `ibmq_boeblingen`'s back-end properties. Appendix A summarizes the daily IBM re-calibration schedule and procedures.

At 4 am 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

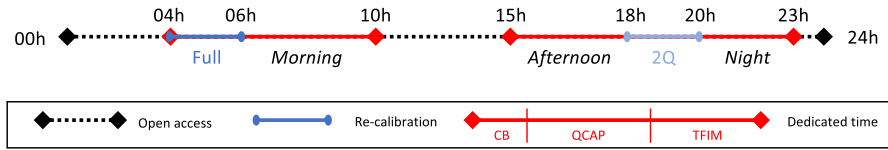


Fig. 3 Daily timeline for calibrations, computations and measurements. The red lines denote the dedicated reservation time periods. The blue lines denote the time periods when IBM performed a full 1- and 2-qubit re-calibration each morning and a two-qubit (2Q) only re-calibration each night

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 to 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 measurements were done using cycle benchmarking (CB) and randomized benchmarking (RB) to study the errors present in the circuits from Fig. 1. Cycle benchmarking (CB) was performed on Layouts 1, 2, and 3 using Circuits 1 and 2. We computed the process infidelity and the Quantum Capacity bound (QCAP). QCAP measurements were done on the three layouts only with Circuit 1. After the CB computations were finished, the Trotterization of the TFIM Hamiltonian was run on each of the three different physical qubit layouts using only the Circuit 1 gate design so that a direct comparison could be made both to the previously published results in [55] and the CB and QCAP measurements.

4 Results

This section reports on the measurements obtained from the cycle benchmarking and TFIM computations on the IBM `ibmq_boeblingen` processor during the eight-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 summarizes 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 (Sects. 4.1 and 4.2), dependencies on qubit layouts for

concurrent calculations (Sect. 4.3) and impacts from different circuit structure choices (Sect. 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 eight-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

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

where $d = 2^n$ (n is the number of qubits) [62]. 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

$$e_F = r \frac{d+1}{d} \quad (7)$$

from [63].

The overall process infidelity from the CB calculations and the RB measured process infidelities (taken from the backend properties that IBM published each day after re-calibration) 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 (Fig. 5). For Cycles 2, 3 and 4, the process infidelity computations using the CB protocol captured additional sources of error, such as spatially correlated errors and cross-talk that the RB procedure could not measure. As a result, the CB process infidelity measurement was sometimes 2 to 4 times greater than the corresponding RB measurement for the same 2 qubits (Table 2 in Appendix C). For Cycle 1, there is no analog to the CB two 2-qubit gate set measurement provided through the IBM backend properties and so only the CB results are plotted for Cycle 1. However, an effective RB process infidelity can be computed by just using the individual [6,7] and [12,11] two-qubit gates error measurements, converting each to a process fidelity, multiplying them together and then finally calculating an effective process infidelity. Upon doing so, it is found that the Cycle 1 effective RB process infidelity on January 24th was 0.0191 and on January 29th it was 0.0438.

The QCAP bound as a function of number Trotter steps was then calculated using CB, and both the QCAP_{CB} and QCAP_{RB} results are plotted in Fig. 5. All of the

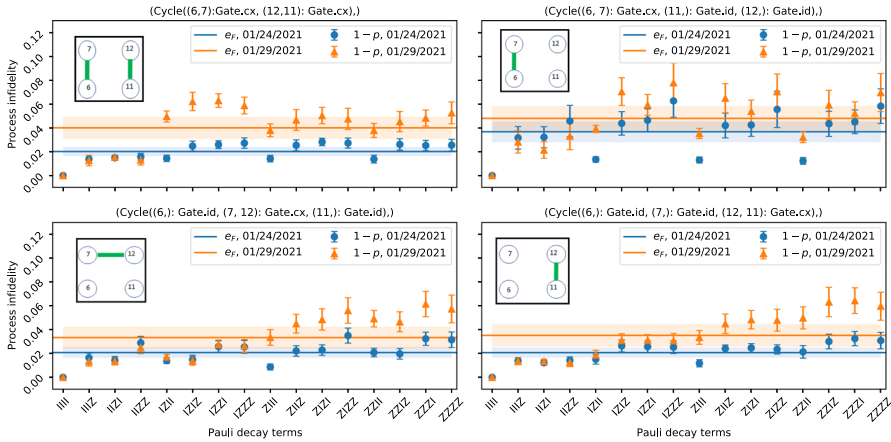


Fig. 4 Pauli infidelities for each hard cycle calculated for each Pauli decay term from morning run on qubits [6, 7, 12, 11] (Layout 2) on days 01/24/2021 (blue lines and data points) and 01/29/2021 (orange lines and data points). The upper left graph is Cycle 1, the upper right is Cycle 2, lower left 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 (Color figure online)

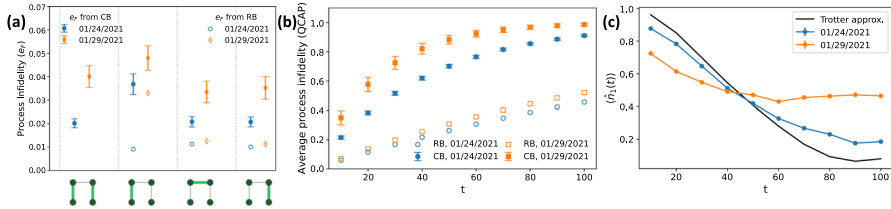


Fig. 5 Inter-day drift in error as characterized by RB and CB. *Left* The process infidelities calculated using RB and CB such that the two-qubit CNOT cycles 1, 2, 3 and 4 are identified along the horizontal axis from left to right, *Center* the QCAP bound as a function of evolution time calculated using RB ($QCAP_{RB}$) and CB ($QCAP_{CB}$). *Right* 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 (Color figure online)

individual CNOT gates were included in the QCAP circuit. The RB computation multiplied the RB backend 2-qubit gate value for each distinct individual CNOT gate by the total number of CNOT gates of that type in the circuit and added the results to obtain the total $QCAP_{RB}$

After this QCAP calculation was completed, the TFIM Trotterization on Circuit 1 was run while the `ibmq_boeblingen` processor while the processor continued to be in dedicated mode. Figure 5c 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.

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.

For each of these measurement time periods, the RB average process infidelity was recorded based on the published backend properties 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 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. The figure shows four individual histograms, one for each of the distinct 2-qubit CNOT cycles. Each of the Pauli decay terms is listed along the x-axis underlying the histograms. The y-axis of each histogram displays the values of the process infidelity. Each Pauli decay term displays 4 individual bars, each representing the magnitude of the process infidelity for that specific Pauli decay term measured at that specific date and time. The black line at the top of each bar represents the error bar range that was calculated for that specific intra-day measurement. There is a color-coded legend at the top of the figure that indicates the specific date and time of each measurement. At the extreme right-hand side of each histogram is the average process infidelity (e_f) for each 2-qubit CNOT cycle measured at each specific date and time.

Figure 7a shows the CB versus RB process infidelity data and analysis, and Fig. 7b shows the graph of the $QCAP_{CB}$ and the $QCAP_{RB}$ values versus evolution time for the January 27th data. Similarly, Fig. 7c shows the CB versus RB process infidelity results and Fig. 7d shows the graph of the $QCAP_{CB}$ 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:

- 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

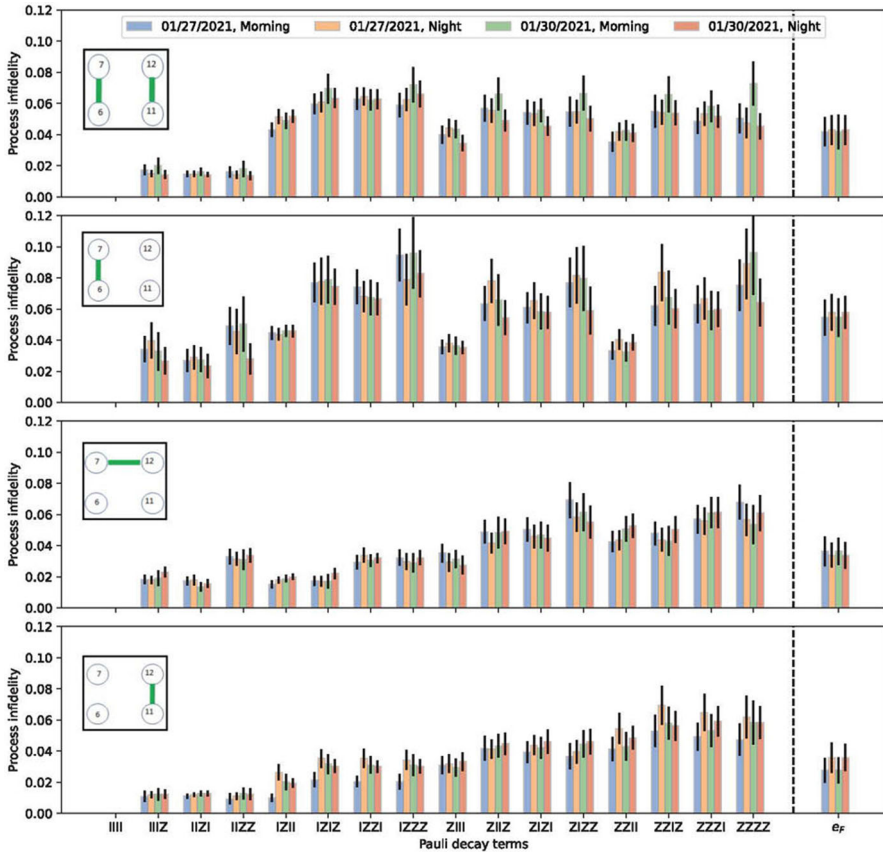


Fig. 6 Histograms of intra-day Pauli process infidelities and error bars for each Pauli decay term for the four distinct 2-qubit cycles measured during the morning and night runs on January 27th and January 30th 2021 on Layout 2 (qubits [6, 7, 12, 11]). At the extreme right side of each histogram is the bar measurement of the process infidelity (e_f) averaged from all the Pauli decay term contributions for that date and time measurement. The black line at the top of each computed average process infidelity is the computed error range for that day’s overall process infidelity measurement (Color figure online)

who try to simply aggregate their results from different runs and perform statistical fits on their aggregated data.

- The daily IBM one-qubit and two-qubit published re-calibrations utilize randomized benchmarking. The average error rates measured by randomized benchmarking and related protocols are not sensitive to the full impact of coherent errors and therefore do not reliably predict the global performance of quantum algorithms. In contrast, the cycle benchmarking protocol employed here can also estimate the global and local error mechanisms when a clock cycle of operations is applied to a quantum register. As a result, the cycle benchmarking process infidelities provide a more inclusive estimate on the circuit coherence times and can more reliably predict the global performance of quantum algorithms.

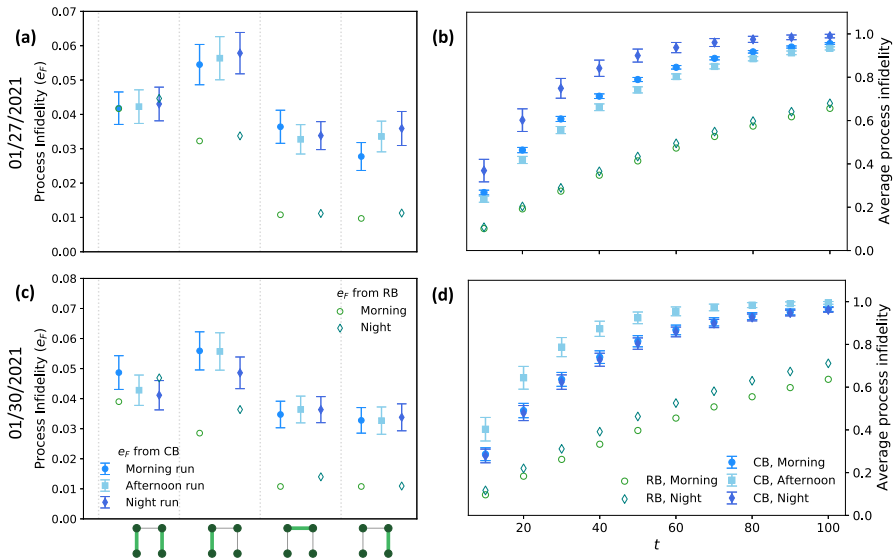


Fig. 7 Error characterization via CB vs. RB. The process infidelities for cycles 1, 2, 3 and 4 calculated using RB and cycle CB from morning, afternoon and night runs of Layout 2 [qubits 6, 7, 12, 11] on January 27, 2021 (a), and January 30, 2021 (c). (See also Table 3). The QCAP bound as a function 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 January 27, 2021 (b) and January 30, 2021 (d) (Color figure online)

- 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. Figures 5 and 7 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 measurements 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 eight-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]). Figure 8 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. The computed QCAP bound uses the set of two qubits illustrated in Fig. 1—diagram b1 for the QCAP computation. The resulting QCAP plot illustrates the deviation of the measured circuit from what would be its ideal equivalent. The point of Fig. 8 is as follows:

- (i) Using only the direct two-qubit randomized benchmark measured error rates from the IBM backend properties as a measure of signal quality will underreport the total error in the computation. This gives the mistaken assurance that it is possible to include data from a larger number Trotter steps than can actually be justified.
- (ii) The practice of “replicating” a quantum computing circuit in a similar scatter-gather procedure used in digital high-performance computing hardware platforms to speed up / increase the statistics for the computation is suspect when applied to algorithms run on current NISQ hardware platforms. Although the hardware on today’s HPC platforms is sufficiently stable so that the results from a scatter-gather operation when combined will fall within each computation’s overall error bars, this same assumption cannot be justified on a quantum processor. Our results show that the output from the scatter-gather operation on Layout 1 when compared to the results from Layout 2 is not even within overlapping error bars of each individual computation so that aggregating these results together can be problematic.
- (iii) 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.
- (iv) The $QCAP_{RB}$ measurements are not sensitive to the impacts of coherent errors and cross-talk affecting the results from Circuit 1 running on 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

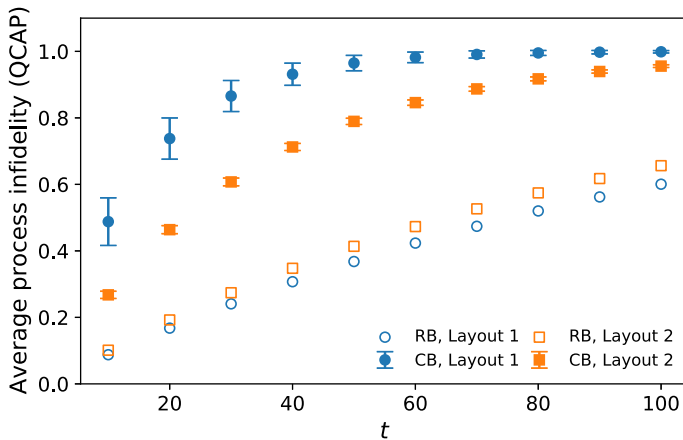


Fig. 8 QCAP bound as a function of evolution time calculated using randomized benchmarking (QCAP_{RB}) and cycle benchmarking (QCAP_{CB}) from the morning run of Circuit 1 on both Layout 1 and Layout 2 on January 27, 2021 (Color figure online)

aggregated together because their results cannot be duplicated within the QCAP bound error bars for the computations from each circuit.

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. 9.

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. 9.

- (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

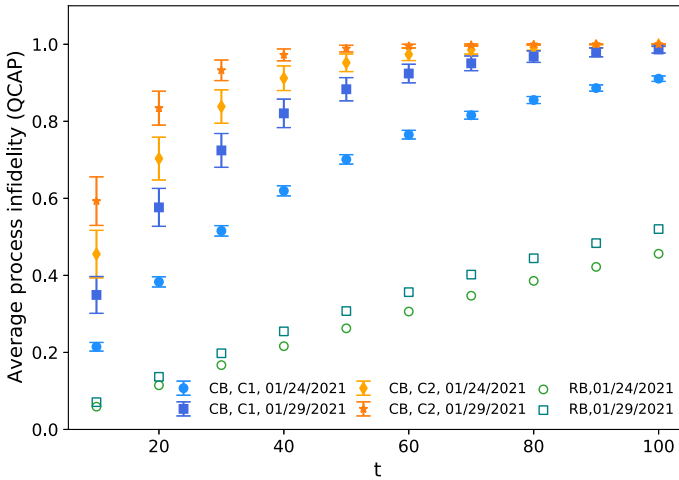


Fig. 9 Comparing two Trotter decomposition that gives the same physics. The QCAP bound as a function of evolution time calculated using RB ($QCAP_{RB}$) and CB ($QCAP_{CB}$) from morning run of Layout 2 (qubits [6, 7, 12, 11]) on days 01/24/2021 and 01/29/2021 for Circuit 1 and Circuit 2 with only CNOT gates used as hard cycles. The plotted error bars only show the statistical error (Color figure online)

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 impacts 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 main goal of our project was quantitatively measure two-qubit process infidelities over different time intervals. Using a TFIM quantum circuit, a series of identical calculations were performed on the same quantum computing hardware platform under as similar conditions as possible at multiple different time periods. The measurements showed time dependent deterioration in the fidelity of the computations that are likely to be attributable to multiple sources.

Decoherent errors associated with the actual hardware environment housing the processor can definitely contribute to this deterioration. There may be also be additional error contributions attributable to environmental factors such as the room temperature hardware and related equipment connected to the superconducting transmon quantum computing device. Decoherent errors grow linearly as a function of system size, and the average error rate can be measured using randomized benchmarking such as T1, T2 and certain environmental properties for the processor.

Coherent errors also contribute to deterioration of the coherence of the qubits in the circuit. Coherent errors are known to have quadratic error growth with gate infidelities that may differ up to an order of magnitude from the incoherent average error rate. Furthermore, because the coherent errors may add and subtract unpredictably, the global impact of these coherent errors is difficult to systematically calibrate in structured circuits.

Because two-qubit gates are most likely to be the source of the largest error rates, this project focused on analyzing these two-qubit gate measurements. The complexity in characterizing these qubit errors indicates that these hardware error rates are a complex mix of the choice of qubits, the choice of gates, the order or sequence that the gates are applied in the circuit, the local direct gate combinations, and the impacts of surrounding spectator qubits.

We used Cycle Benchmarking and Quantum Capacity protocols to measure the two-qubit process infidelities, including the computation of the error bars. The project observed that in many cases the comparison between measurements run on the same subset of qubits at several different time periods showed strikingly different results for both the two-qubit process infidelity measurement and the quantum capacity measurements. These instabilities appear as both inter-day and intra-day qubit drift. Depending on which subset of qubits were used for the computation, variations in the process infidelities among different layouts were also measured.

This implies that comparisons of the two-qubit process infidelities and associated error bars when run at different time periods and/or on different subsets of qubits under as identical as possible hardware conditions may not yield consistent results. What is significant is that in many instances these measurements may not even overlap each other within the range of their error bars.

These results raise questions as to the reproducibility of computations over time scales of days, weeks or months and the potential instability and drift of the qubits and/or the overall quantum computing hardware platform itself. These should be a serious concern to the community.

At the present time, the number of qubits used for quantum computing computations is quite modest. If there are concerns about potential discrepancies between computational results measured at two different time periods, the accuracy of the quantum computing results can always be checked with digital computers. However, as the number of qubits used in quantum computations grows over the next several years, at some point direct verification of the quantum computer's output will be beyond the reach of any digital computer. In addition, the quantum computing hardware will still not have a sufficient number of qubits to support fully fault tolerant error correction mechanisms for these computations.

If the user community is to have confidence in the published results from running quantum circuits using a very large number of qubits, there must be demonstrated mitigation strategies to control the noise and errors on these systems. These mitigation strategies must demonstrate that the measurements from identical circuits run at different time periods of days, weeks or even longer under identical conditions on the same processor will deliver process instability results that are within error bars of each other. Without such supporting evidence or the ability to verify results using digital computers, these computations will be overshadowed by the possibility that the data may carry uncertainties that may cause users to view the results as unreliable or lacking a sufficiently high confidence level to be trusted.

Over the past few years, great efforts have been directed toward error mitigation and improvements in coherence time for individual computations [3, 64–69]. In addition, a recently introduced IBM Qiskit Runtime Program [70] helps to streamline computations and minimize intra-day drift. Other groups are also investigating additional methods for improving the reliability of quantum processing computations [71–75]. Furthermore, recent improvements extending the coherence time of a superconducting transmon qubit device to 0.5 milliseconds [76] have allowed these error mitigation and noise reductions protocols and techniques to improve overall coherence times on these hardware platforms.

Although these improvements do extend the stability of an individual computation, it is not clear they also improve computations impacted by inter-day and intra-day qubit drift and hardware instabilities. It would be useful if an equivalent level of effort and set of in-depth studies could be directed to check the stability and reproducibility of the results from an application's quantum circuits implemented on a processor's qubits over time periods of days or weeks.

Although `ibmq_boeblingen` is an older generation quantum processor, the error mitigation protocols and methodologies summarized here are still applicable for checking the stability of the newer hardware platforms. This group has also begun several other projects measuring qubit properties through additional robust characterization and stability analysis protocols. These new methods are now investigating how the recent hardware and software improvements may have improved qubit stability of the new superconducting transmon hardware platforms. The results from these new projects will be reported in future publications.

Acknowledgements PD was supported in part by the U.S. Department of Energy (DoE) under award DE-AC05-00OR22725. KYA and RCP were supported by the Quantum Information Science Enabled Discovery (QuantISED) for High Energy Physics program at ORNL under FWP number ERKAP61 and used resources of Oak Ridge Leadership Computing Facility located at ORNL, which is supported by the Office of Science of the Department of Energy under contract No. DE-AC05-00OR22725. YM and EG are supported by a Department of Energy QuantISED grant DE-SC0019139. ZP gratefully acknowledges funding support from the NSF with an NSF pre-doctoral fellowship. We acknowledge the use of IBM Quantum services for this work, especially discussions with Nathan Earnest-Noble, Matthew Stypulkoski, Azia Ngoueya and Patrick Mensac from IBM. We thank North Carolina State University (NCSU) for access to the IBM Quantum Network quantum computing hardware platforms through the NCSU IBM Quantum Hub and thank IBM Research for the extended dedicated mode reservations and availability of the `ibmq_boeblingen` processor on which the computations were performed. The views expressed are those of the authors and do not reflect the official policy or position of IBM or the IBM Quantum team. The project team acknowledges the use of True-Q software from Keysight Technologies and useful discussions with Ian Hincks, Dar Dahlen and Arnaud Carignan-Dugas from Keysight.

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 original TFIM code and the physics model circuits. ANT built the tables in the paper. AFK, RP, YM and PD edited and reviewed the document.

Data availability Data are available from the corresponding author upon reasonable request.

Declarations

Conflict of interests The authors declare no competing interests.

Appendix A: Daily `ibmq_boeblingen` qubit re-calibration schedule

For this specific project, IBM agreed to supply our team with approximately 140 h 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 h 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 [77].

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 [20] and is schematically represented in Fig. 10. 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{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{G} represents the noisy implementation of the gate(s) being measured in the circuit. The blocks $\tilde{R}_{i,j}$ are random Paulis represented by the j th tensor factor of the i th gate inserted into the cycle to create an effective Pauli channel for the gate \tilde{G} being measured. The blocks \tilde{B} and \tilde{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{G} is calculated and is represented by the solid line and shaded band on the graph.

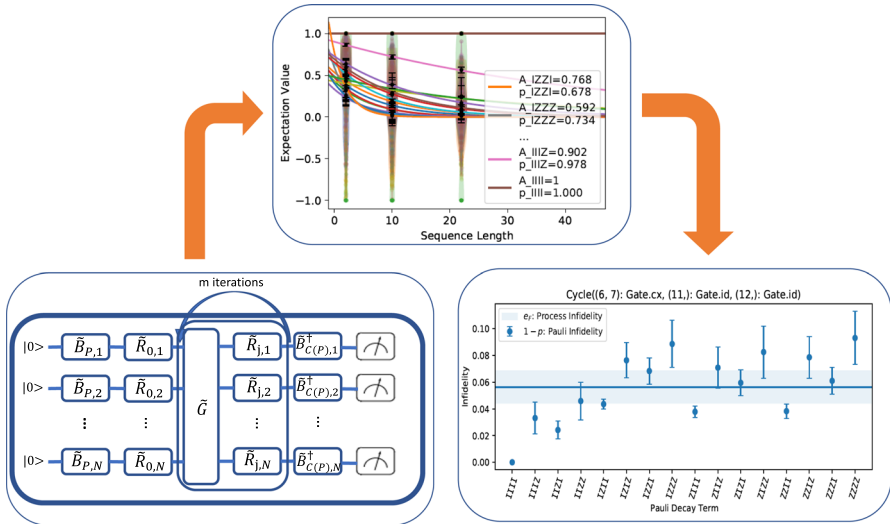


Fig. 10 Block diagram of the CB protocol implementation

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 need to be chosen carefully such that 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., the 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 of the 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 measurements 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 overall process infidelity are shown in Fig. 4 for inter-day and Fig. 6 for intra-day computations.

Appendix B.2: Quantum Capacity

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 was run $N_{shots} = 128$.

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 [54]. 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.

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

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

See Tables 1, 2 and 3.

Table 1 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

<i>Inter-day, Layout 2,</i>				IBM backend readout		
single-qubit errors	<i>Qubits</i>	T_1 (μs)	T_2 (μs)	error ($\times 10^{-2}$)	U_2 ($\times 10^{-4}$)	U_3 ($\times 10^{-4}$)
01/24/2021 Morning run	6	67.1	99.9	2.54	2.87	5.74
	7	94.8	86.8	2.30	3.05	6.71
	12	97.5	88.5	3.13	2.91	5.82
	11	95.1	71.6	3.55	4.44	8.88
01/29/2021 Morning run	6	24.1	4.97	9.19	25.0	50.0
	7	78.8	103.4	2.84	3.97	7.94
	12	80.8	114.6	3.47	3.49	6.98
	11	48.4	87.8	3.22	4.64	9.29

Table 2 Cycle 2, 3, and 4 values for both the CB process infidelity computed for the two-qubit pairs [6, 7], [7,12], [12, 11] and the RB process infidelity computed from the recorded IBM back-end property error rates 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

<i>Inter-day, Layout 2, two-qubit process infidelities</i>				
	Cycle	Qubits	Cycle benchmark ($\times 10^{-2}$)	IBM backend ($\times 10^{-2}$)
01/24/2021 Morning run	2	[6,7]	3.67	0.908
	3	[7,12]	2.08	1.128
	4	[12,11]	2.07	1.010
01/29/2021 Morning run	2	[6,7]	4.79	3.30
	3	[7,12]	3.34	1.25
	4	[12,11]	3.52	1.124

Table 3 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

<i>Layout 2 intra-day two-qubit process infidelities</i>				
	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

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

References

1. Preskill, J.: Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018). <https://doi.org/10.22331/q-2018-08-06-79>
2. IBM: Our New 2022 Development Roadmap. <https://www.ibm.com/quantum/roadmap>. Accessed 23 June 2022
3. Temme, K., Bravyi, S., Gambetta, J.M.: Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.* **119**(18), 180509 (2017). <https://doi.org/10.1103/PhysRevLett.119.180509>

4. Li, Y., Benjamin, S.C.: Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X* **7**(2), 021050 (2017). <https://doi.org/10.1103/PhysRevX.7.021050>
5. Giurgica-Tiron, T., Hindy, Y., LaRose, R., Mari, A., Zeng, W.J.: 2020 IEEE International Conference on Quantum Computing and Engineering (QCE) (2020). <https://doi.org/10.1109/QCE49297.2020.00045>
6. Wallman, J.J., Emerson, J.: Noise tailoring for scalable quantum computation via randomized compiling. *Phys. Rev. A* **94**(5), 052325 (2016). <https://doi.org/10.1103/PhysRevA.94.052325>
7. 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., Siddiqi, I.: Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor (2021). [arXiv:2010.00215](https://arxiv.org/abs/2010.00215)
8. Greenbaum, D., Dutton, Z.: Modeling coherent errors in quantum error correction. *Quantum Sci. Technol.* **3**, 015007 (2017). <https://doi.org/10.1088/2058-9565/aa9a06>
9. Mohseni, M., Rezaekhani, A.T., Lidar, D.A.: Quantum-process tomography: resource analysis of different strategies. *Phys. Rev. A* **77**, 032322 (2008). <https://doi.org/10.1103/PhysRevA.77.032322>
10. Flammia, S.T., Gross, D., Liu, Y.K., Eisert, J.: Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New J. Phys.* **14**, 095022 (2012). <https://doi.org/10.1088/1367-2630/14/9/095022>
11. Chuang, I.L., Nielsen, M.A.: Prescription for experimental determination of the dynamics of a quantum black box. *J. Mod. Opt.* **44**, 2455–2467 (1997). <https://doi.org/10.1080/09500349708231894>
12. Mostame, S., Schützhold, R.: Quantum simulator for the Ising model with electrons floating on a helium film. *Phys. Rev. Lett.* **101**(22), 220501 (2008). <https://doi.org/10.1103/PhysRevLett.101.220501>
13. Branderhorst, M.P.A., Nunn, J., Walmsley, I.A., Kosut, R.L.: Simplified quantum process tomography. *New J. Phys.* **11**, 115010 (2009). <https://doi.org/10.1088/1367-2630/11/11/115010>
14. Merkel, S.T., Gambetta, J.M., Smolin, J.A., Poletto, S., Córcoles, A.D., Johnson, B.R., Ryan, C.A., Steffen, M.: Self-consistent quantum process tomography. *Phys. Rev. A* **87**, 062119 (2013). <https://doi.org/10.1103/PhysRevA.87.062119>
15. Blume-Kohout, R., Gamble, J.K., Nielsen, E., Mizrahi, J., Sterk, J.D., Maunz, P.: Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit (2013). [arXiv:1310.4492](https://arxiv.org/abs/1310.4492)
16. Scholten, T.L., Blume-Kohout, R.: Behavior of the maximum likelihood in quantum state tomography. *New J. Phys.* **20**, 023050 (2018). <https://doi.org/10.1088/1367-2630/aaa7e2>
17. Emerson, J., Alicki, R., Życzkowski, K.: Scalable noise estimation with random unitary operators. *J. Opt. B Quantum Semiclassical Opt.* **7**, S347–S352 (2005). <https://doi.org/10.1088/1464-4266/7/10/021>
18. Dankert, C., Cleve, R., Emerson, J., Livine, E.: Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A* **80**, 012304 (2009). <https://doi.org/10.1103/PhysRevA.80.012304>
19. Magesan, E., Blume-Kohout, R., Emerson, J.: Gate fidelity fluctuations and quantum process invariants. *Phys. Rev. A* **84**, 012309 (2011). <https://doi.org/10.1103/PhysRevA.84.012309>
20. Erhard, A., Wallman, J.J., Postler, L., Meth, M., Stricker, R., Martinez, E.A., Schindler, P., Monz, T., Emerson, J., Blatt, R.: Characterizing large-scale quantum computers via cycle benchmarking. *Nat. Commun.* **10**, 1–7 (2019). <https://doi.org/10.1038/s41467-019-13068-7>
21. 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., Pan, J.W.: Genuine 12-qubit entanglement on a superconducting quantum processor. *Phys. Rev. Lett.* **122**(11), 110501 (2019). <https://doi.org/10.1103/PhysRevLett.122.110501>
22. McKay, D.C., Sheldon, S., Smolin, J.A., Chow, J.M., Gambetta, J.M.: Three-qubit randomized benchmarking. *Phys. Rev. Lett.* **122**(20), 200502 (2019). <https://doi.org/10.1103/PhysRevLett.122.200502>
23. Sundaresan, N., Lauer, I., Pritchett, E., Magesan, E., Jurcevic, P., Gambetta, J.M.: Reducing unitary and spectator errors in cross resonance with optimized rotary echoes. *PRX Quantum* **1**, 020318 (2020). <https://doi.org/10.1103/PRXQuantum.1.020318>
24. Piltz, C., Sriarunothai, T., Varón, A., Wunderlich, C.: A trapped-ion-based quantum byte with 10–5 next-neighbour cross-talk. *Nat. Commun.* **5**, 1–10 (2014). <https://doi.org/10.1038/ncomms5679>
25. Reagar, M., Osborn, C.B., Tezak, N., Staley, A., Prawiroatmodjo, G., Scheer, M., Alidoust, N., Sete, E.A., Didier, N., da Silva, M.P., et al.: Demonstration of universal parametric entangling gates on a multi-qubit lattice. *Sci. Adv.* **4**, eaao3603 (2018). <https://doi.org/10.1126/sciadv.aao3603>

26. Sarovar, M., Proctor, T., Rudinger, K., Young, K., Nielsen, E., Blume-Kohout, R.: Detecting crosstalk errors in quantum information processors. *Quantum* **4**, 321 (2020). <https://doi.org/10.22331/q-2020-09-11-321>
27. 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., et al.: Simple all-microwave entangling gate for fixed-frequency superconducting qubits. *Phys. Rev. Lett.* **107**, 080502 (2011). <https://doi.org/10.1103/PhysRevLett.107.080502>
28. Tripathi, V., Chen, H., Khezri, M., Yip, K.W., Levenson-Falk, E.M., Lidar, D.A.: Suppression of crosstalk in superconducting qubits using dynamical decoupling (2021). [arXiv:2108.04530](https://arxiv.org/abs/2108.04530)
29. Bialczak, R.C., Ansmann, M., Hofheinz, M., Lucero, E., Neeley, M., O'Connell, A.D., Sank, D., Wang, H., Wenner, J., Steffen, M., et al.: Quantum process tomography of a universal entangling gate implemented with Josephson phase qubits. *Nat. Phys.* **6**, 409–413 (2010). <https://doi.org/10.1038/nphys1639>
30. Zhao, P., Lan, D., Xu, P., Xue, G., Blank, M., Tan, X., Yu, H., Yu, Y.: Suppression of static ZZ interaction in an all-transmon quantum processor. *Phys. Rev. Appl.* **16**, 024037 (2021). <https://doi.org/10.1103/PhysRevApplied.16.024037>
31. 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., Pappas, D.P.: A universal quantum gate set for transmon qubits with strong zz interactions (2021). [arXiv:2103.12305](https://arxiv.org/abs/2103.12305)
32. Clark, C.R., Metodi, T.S., Gasster, S.D., Brown, K.R.: Resource requirements for fault-tolerant quantum simulation: the ground state of the transverse Ising model. *Phys. Rev. A* **79**(6), 062314 (2009). <https://doi.org/10.1103/PhysRevA.79.062314>
33. Kandala, A., Temme, K., Córcoles, A.D., Mezzacapo, A., Chow, J.M., Gambetta, J.M.: Error mitigation extends the computational reach of a noisy quantum processor. *Nature* **567**, 491–495 (2019). <https://doi.org/10.1038/s41586-019-1040-7>
34. Yeter-Aydeniz, K., Siopsis, G., Pooser, R.C.: Scattering in the Ising model with the quantum Lanczos algorithm *. *New J. Phys.* **23**, 043033 (2021). <https://doi.org/10.1088/1367-2630/abe63d>
35. Lamm, H., Lawrence, S.: Simulation of nonequilibrium dynamics on a quantum computer. *Phys. Rev. Lett.* **121**, 170501 (2018)
36. Cervera-Lierta, A.: Exact Ising model simulation on a quantum computer. *Quantum* **2**, 114 (2018). <https://doi.org/10.22331/q-2018-12-21-114>
37. Somma, R.D.: Quantum simulations of one dimensional quantum systems (2016). [arXiv:1503.06319](https://arxiv.org/abs/1503.06319)
38. Johanning, M., Varón, A.F., Wunderlich, C.: Quantum simulations with cold trapped ions. *J. Phys. B At. Mol. Opt. Phys.* **42**, 154009 (2009). <https://doi.org/10.1088/0953-4075/42/15/154009>
39. Smith, A., Kim, M.S., Pollmann, F., Knolle, J.: Simulating quantum many-body dynamics on a current digital quantum computer. *npj Quantum Inf.* **5**, 1–13 (2019). <https://doi.org/10.1038/s41534-019-0217-0>
40. Vovrosh, J., Knolle, J.: Confinement and entanglement dynamics on a digital quantum computer (2020). [arXiv:2001.03044](https://arxiv.org/abs/2001.03044)
41. Salathé, Y., Mondal, M., Oppliger, M., Heinsoo, J., Kurpiers, P., Potočník, A., Mezzacapo, A., Las Heras, U., Lamata, L., Solano, E., et al.: Digital quantum simulation of spin models with circuit quantum electrodynamics. *Phys. Rev. X* **5**, 021027 (2015). <https://doi.org/10.1103/PhysRevX.5.021027>
42. Bernien, H., Schwartz, S., Keesling, A., Levine, H., Omran, A., Pichler, H., Choi, S., Zibrov, A.S., Endres, M., Greiner, M., Vuletić, V., Lukin, M.D.: Probing many-body dynamics on a 51-atom quantum simulator. *Nature* **551**, 579–584 (2017)
43. Zhang, J., Pagano, G., Hess, P.W., Kyprianidis, A., Becker, P., Kaplan, H., Gorshkov, A.V., Gong, Z.X., Monroe, C.: Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator. *Nature* **551**, 601–604 (2017). <https://doi.org/10.1038/nature24654>
44. Labuhn, H., Barredo, D., Ravets, S., de Léséleuc, S., Macrì, T., Lahaye, T., Browaeys, A.: Tunable two-dimensional arrays of single Rydberg atoms for realizing quantum Ising models. *Nature* **534**, 667–670 (2016). <https://doi.org/10.1038/nature18274>
45. 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., Foss-Feig, M.: Holographic dynamics simulations with a trapped ion quantum computer (2021). [arXiv:2105.09324](https://arxiv.org/abs/2105.09324)
46. Kandala, A., Mezzacapo, A., Temme, K., Takita, M., Brink, M., Chow, J.M., Gambetta, J.M.: Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242–246 (2017). <https://doi.org/10.1038/nature23879>


47. You, H., Geller, M.R., Stancil, P.C.: Simulating the transverse Ising model on a quantum computer: error correction with the surface code. *Phys. Rev. A* **87**(3), 032341 (2013). <https://doi.org/10.1103/PhysRevA.87.032341>
48. Hebenstreit, M., Alsina, D., Latorre, J.I., Kraus, B.: Compressed quantum computation using a remote five-qubit quantum computer. *Phys. Rev. A* **95**(5), 052339 (2017). <https://doi.org/10.1103/PhysRevA.95.052339>
49. Schauss, P.: Quantum simulation of transverse Ising models with Rydberg atoms. *Quantum Sci. Technol.* **3**, 023001 (2018). <https://doi.org/10.1088/2058-9565/aa9c59>
50. Flammia, S.T., Wallman, J.J.: Efficient estimation of Pauli channels. *ACM Trans. Quantum Comput.* **1**, 1–32 (2020). <https://doi.org/10.1145/3408039>
51. Keesling, A., et al.: Quantum Kibble–Zurek mechanism and critical dynamics on a programmable Rydberg simulator. *Nature* **568**, 207–211 (2019)
52. Kim, Y., Wood, C.J., Yoder, T.J., Merkel, S.T., Gambetta, J.M., Temme, K., Kandala, A.: Scalable error mitigation for noisy quantum circuits produces competitive expectation values (2021). [arXiv:2108.09197](https://arxiv.org/abs/2108.09197)
53. Gustafson, E., Zhu, Y., Dreher, P., Linke, N.M., Meurice, Y.: Real-time quantum calculations of phase shifts using wave packet time delays. *Phys. Rev. D* **104**, 054507 (2021)
54. Gustafson, E., Dreher, P., Hang, Z., Meurice, Y.: Benchmarking quantum computers for real-time evolution of a $(1 + 1)$ field theory with error mitigation. *Quantum Sci. Technol.* **6**, 045020 (2021). <https://doi.org/10.1088/2058-9565/ac1dff>
55. Gustafson, E., Meurice, Y., Unmuth-Yockey, J.: Quantum simulation of scattering in the quantum Ising model. *Phys. Rev. D* **99**, 094503 (2019)
56. Lloyd, S.: Universal quantum simulators. *Science* **273**, 1073–1078 (1996). <https://doi.org/10.1126/science.273.5278.1073>
57. Meurice, Y.: *Quantum Field Theory: A Quantum Computation Approach*. Institute of Physics Publishing, Bristol (2021)
58. Childs, A.M., Su, Y., Tran, M.C., Wiebe, N., Zhu, S.: Theory of trotter error with commutator scaling. *Phys. Rev. X* **11**(1), 011020 (2021). <https://doi.org/10.1103/PhysRevX.11.011020>
59. Layden, D.: First-order trotter error from a second-order perspective (2021). [arXiv:2107.08032](https://arxiv.org/abs/2107.08032)
60. Eastin, B., Flammia, S.T.: Q-circuit tutorial (2004). [arXiv:quant-ph/0406003v2](https://arxiv.org/abs/quant-ph/0406003v2)
61. Beale, S.J., Carignan-Dugas, A., Dahlen, D., Emerson, J., Hincks, I., Iyer, P., Jain, A., Hufnagel, D., Ospadov, E., Saunders, J., Stasiuk, A., Wallman, J.J., Winick, A.: True-q (2020). <https://doi.org/10.5281/zenodo.3945250>
62. Qi, J., Ng, H.K.: Comparing the randomized benchmarking figure with the average infidelity of a quantum gate-set. *Int. J. Quantum Inf.* **17**, 1950031 (2019). <https://doi.org/10.1142/S021974991950031X>
63. Nielsen, M.A.: A simple formula for the average gate fidelity of a quantum dynamical operation. *Phys. Lett. A* **303**, 249–252 (2002). [https://doi.org/10.1016/S0375-9601\(02\)01272-0](https://doi.org/10.1016/S0375-9601(02)01272-0)
64. McKay, D.C., Cross, A.W., Wood, C.J., Gambetta, J.M.: Correlated randomized benchmarking (2020). [arXiv:2003.02354](https://arxiv.org/abs/2003.02354)
65. Bravyi, S., Sheldon, S., Kandala, A., McKay, D.C., Gambetta, J.M.: Mitigating measurement errors in multiqubit experiments. *Phys. Rev. A* **103**, 042605 (2021). <https://doi.org/10.1103/PhysRevA.103.042605>
66. Ware, M., Johnson, B.R., Gambetta, J.M., Ohki, T.A., Chow, J.M., Plourde, B.L.T.: Cross-resonance interactions between superconducting qubits with variable detuning (2019). [arXiv:1905.11480](https://arxiv.org/abs/1905.11480)
67. Nation, P.D., Kang, H., Sundaresan, N., Gambetta, J.M.: Scalable mitigation of measurement errors on quantum computers. *PRX Quantum* **2**, 040326 (2021). <https://doi.org/10.1103/PRXQuantum.2.040326>
68. Kim, Y., Wood, C.J., Yoder, T.J., Merkel, S.T., Gambetta, J.M., Temme, K., Kandala, A.: Scalable error mitigation for noisy quantum circuits produces competitive expectation values (2021). [arXiv:2108.09197](https://arxiv.org/abs/2108.09197)
69. Van den Berg, E., Mineev, Z.K., Kandala, A., Temme, K.: Probabilistic error cancellation with sparse pauli-lindblad models on noisy quantum processors (2022). [arXiv:2201.09866](https://arxiv.org/abs/2201.09866)
70. IBM Qiskit Runtime. <https://www.ibm.com/quantum/qiskit-runtime>
71. Lin, J., Wallman, J.J., Hincks, I., Laflamme, R.: Independent state and measurement characterization for quantum computers. *Phys. Rev. Res.* **3**(3), 033285 (2021). <https://doi.org/10.1103/PhysRevResearch.3.033285>

72. Lin, J., Wallman, J.J., Hincks, I., Laflamme, R.: Independent state and measurement characterization for quantum computers. *Phys. Rev. Res.* **3**, 033285 (2021). <https://doi.org/10.1103/PhysRevResearch.3.033285>
73. He, A., Nachman, B., de Jong, W.A., Bauer, C.W.: Zero-noise extrapolation for quantum-gate error mitigation with identity insertions. *Phys. Rev. A* **102**, 012426 (2020). <https://doi.org/10.1103/PhysRevA.102.012426>
74. Pasuzzi, V.R., He, A., Bauer, C.W., de Jong, W.A., Nachman, B.: Computationally efficient zero noise extrapolation for quantum gate error mitigation (2021). [arXiv:2110.13338](https://arxiv.org/abs/2110.13338)
75. Urbaneck, M., Nachman, B., Pasuzzi, V.R., He, A., Bauer, C.W., de Jong, W.A.: Mitigating depolarizing noise on quantum computers with noise-estimation circuits. *Phys. Rev. Lett.* **127**, 270502 (2021). <https://doi.org/10.1103/PhysRevLett.127.270502>
76. Guo, J., Lou, H., Li, R., Fang, W., Liu, J., Long, P., Ying, S., Ying, M.: isq: towards a practical software stack for quantum programming (2022). [arXiv:2205.03866](https://arxiv.org/abs/2205.03866)
77. 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., Siddiqi, I.: Leveraging randomized compiling for the QITE algorithm (2021). [arXiv:2104.08785](https://arxiv.org/abs/2104.08785)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Kübra Yeter-Aydeniz^{1,7} · Zachary Parks² · Aadithya Nair Thekkiniyadath² · Erik Gustafson^{3,8} · Alexander F. Kemper⁴ · Raphael C. Pooser⁵ · Yannick Meurice³ · Patrick Dreher⁶ 

¹ Physics Division and Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

² Department of Computer Science, North Carolina State University, Raleigh, NC 27695, USA

³ Department of Physics and Astronomy, The University of Iowa, Iowa City, IA 52242, USA

⁴ Department of Physics, North Carolina State University, Raleigh, NC 27695, USA

⁵ Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

⁶ Electrical and Computer Engineering and Department of Physics, North Carolina State University, Raleigh, NC 27695, USA

⁷ Present Address: Emerging Engineering and Physical Sciences Department., The MITRE Corporation, 7515 Colshire Drive, McLean, VA 22102-7539, USA

⁸ Present Address: Theoretical Physics Division, Fermi National Accelerator Lab, Batavia, IL 60510, USA