Noise-aware circuit compilations for a continuously parameterized two-qubit gateset

Christopher G. Yale,[1,*,†] Rich Rines,[2,†,‡] Victory Omole,[2,†,‡] Bharath Thotakura,[2,†,‡] Ashlyn D. Burch,[1] Matthew N.H. Chow,[1,3,4] Megan Ivory,[1] Daniel Lobser,[1] Brian K. McFarland,[1] Melissa C. Revelle,[1] Susan M. Clark,[1] and Pranav Gokhale[2]

[1] *Sandia National Laboratories, Albuquerque, New Mexico 87123, USA*

[2] *Infleqtion, Chicago, Illinois 60604, USA*

[3] *Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131, USA*

[4] *Center for Quantum Information and Control, University of New Mexico, Albuquerque, New Mexico 87131, USA*

(Received 21 November 2024; revised 2 July 2025; accepted 8 July 2025; published 22 August 2025)

State-of-the-art noisy-intermediate-scale quantum processors are currently implemented across a variety of hardware platforms, each with their own distinct gatesets. As such, circuit compilation should not only be aware of but also deeply connect to the native gateset and noise properties of each. Trapped-ion processors are one such platform that provides a gateset that can be continuously parameterized across both one- and two-qubit gates. Here we use the Quantum Scientific Computing Open User Testbed to study noise-aware compilations focused on continuously parameterized two-qubit *ZZ* gates (based on the Mølmer-Sørensen interaction) using SUPERSTAQ, a quantum software platform for hardware-aware circuit compiler optimizations. We discuss the realization of *ZZ* gates with arbitrary angle on the all-to-all connected trapped-ion system. Then we discuss a variety of different compiler optimizations that innately target these *ZZ* gates and their noise properties. These optimizations include moving from a restricted maximally entangling gateset to a continuously parameterized one, swap mirroring to further reduce the total entangling angle of the operations, focusing the heaviest *ZZ* angle participation on the best-performing gate pairs, and circuit approximation to remove the least impactful *ZZ* gates. We demonstrate these compilation approaches on the hardware with randomized quantum volume circuits, observing the potential to realize a larger quantum volume as a result of these optimizations. Using differing yet complementary analysis techniques, we observe the distinct improvements in system performance provided by these noise-aware compilations and study the role of stochastic and coherent error channels for each compilation choice.

DOI: 10.1103/3cmg-5rk7

I. INTRODUCTION

The utility of today's quantum computers is fundamentally limited by noise. In order to execute an abstract quantum application on a real quantum computer, it must first be compiled to the discrete set of physical operations available to the hardware. The details of this compilation process can have a tremendous impact on the quality of experimental results on a limited and noisy device. Well-optimized quantum compilation has the potential to vastly expand the complexity of quantum applications that can be executed meaningfully on today's hardware. To this end, many general-purpose strategies [1–6] and frameworks [7–13] for optimizing compilation have been developed, targeting various levels of abstraction from large-scale fault-tolerant quantum applications to hardware-level gate implementations [14]. In addition to these generalized approaches, still more has been done to establish low-level, heuristic compilation rules applicable to specific quantum computing modalities—often driven by the demands of a given experimental effort [15–19]. While less universal, these targeted low-level optimizations can be particularly valuable in that they are informed by the particular capabilities, limitations, and error mechanisms of the hardware itself [4], and can serve as the building blocks for higher-level primitives and optimization strategies.

*Contact author: cgyale@sandia.gov

†These authors contributed equally to this work.

‡Contact author: superstaq@infleqtion.com

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Currently viable qubit technologies (e.g., superconducting transmons, trapped ions, neutral atoms, and silicon quantum dots) have their own unique set of control mechanisms, topological constraints, timescales, and physical capabilities and tradeoffs. Computationally, these determine the primitive gateset made available by the device and the corresponding connectivity between qubits on the device. Superconducting qubits, for example, support very fast quantum operations via rf signals, but are typically limited to static two-dimensional connectivity, whereas ions and cold atoms support both movement and longer-range interactions via optical pulses.

The physics behind each qubit modality also imbues it with a unique set of dominant error mechanisms and noise sources. To this end, optimized compilation is most successful when it is carefully tuned to the hardware, taking into account the specific capabilities and constraints of the device with a deep awareness of the possible error mechanisms. This awareness is doubly important when we account for compilation efficiency since generic optimal compilation itself is computationally hard [20], and thus the benefits of quantum circuit optimizations must be weighed against the classical complexity of the compilation steps themselves. Given that errors will drift over time, we must also account for the level of detail necessary in assessing error mechanisms to inform compilation as well as the corresponding complexity of characterizing the hardware with sufficient regularity and granularity to avoid performance degradation due to stale error descriptions. Ideally, hardware-informed compilation should therefore rely on an understanding of the general characteristics and expected trends of the system's performance rather than strict quantum characterization, verification, and validation assessments.

Here, we focus on assessing optimizations for a trapped-ion architecture in order to inform which optimizations we prioritize and provide useful heuristics to reduce the complexity of these optimizations. Trapped-ion approaches are a significant player in the quantum computing landscape providing high-fidelity operations [21,22], long-lived coherence [23], high degree of connectivity [24], and paths to fault tolerance [21,25–27].

At current register sizes, ion-trap quantum computers either comprise a single ion chain [16,28,29] with complete, or "all-to-all," connectivity mediated through shared motional modes [30], or employ a quantum charge-coupled device architecture [19] in which ions are shuttled between a fixed number of smaller interaction sites to achieve full connectivity [22,31]. The former approach offers a number of advantages in terms of compilation efficiency: no additional gates or operations are needed to bridge any missing connectivity, and qubits can be relabeled to take advantage of better-performing gate pairs and/or through the use of virtual swaps in protocols such as swap mirroring [4,15]. However, the maximum size of a single ion chain is fundamentally limited due to a variety of factors, including low axial confinement, reduced ion-ion interaction strengths, and motional mode crowding [32,33]. Furthermore, for a chain of n ions, the number of possible qubit interconnections (i.e., the number possible two-qubit gates) grows as $\frac{1}{2}(n^2 - n)$. Future developments may therefore rely on hybrid approaches consisting of subregisters of tens of ions connected via shuttling routines [34], and eventually photonic interconnects [24].

This work focuses specifically on optimized compilation within a single ion chain, which we see not only as a tool for maximizing the capability of current single-chain architectures, but also as a key component of compilation for future full-scale hybrid devices. We further limit our consideration to low-level gate and circuit optimizations, rather than higher-level or application-specific circuit transformations.

The ability to realize a continuously parameterized two-qubit gateset has also become a key feature of trapped-ion processors in recent years [16,19,35,36]. In many cases, maximally entangling gates are neither required nor desired to realize quantum circuits [17,37–39]. As smaller-angle entanglers tend to incur less error than maximally entangling gates [16,17,19,36,40,41], this continuous range of entangling angles helps to reduce the overall error of a circuit.

Here, we develop a suite of compilation optimizations in SUPERSTAQ [4,42] harnessing the flexibility that is gained by these two features, all-to-all connectivity and continuously parameterized two-qubit gates. We then demonstrate and examine their role in improving performance on a trapped-ion system.

II. QUANTUM SCIENTIFIC COMPUTING OPEN USER TESTBED

The experiments here are performed on a trapped-ion quantum processor, the Quantum Scientific Computing Open User Testbed (QSCOUT) at Sandia National Laboratories [29]. In this work, the qubit register consists of a linear chain of either four or five ions with all-to-all connectivity. Physical gates are executed using Raman transitions on a hyperfine "clock" transition of $^{171}\mathrm{Yb}^{+}$ qubits with a pulsed 355-nm laser. The native gateset consists of a continuously parameterized single-qubit rotation gate, $R_\phi(\theta) = e^{-i\phi Z/2}e^{i\theta X/2}e^{i\phi Z/2}$ (with Pauli operators X and Z, and rotation angle θ about an equatorial axis defined by azimuthal phase ϕ), and a continuously parameterized two-qubit gate, $ZZ(\theta) = e^{i\theta Z\otimes Z/2}$. Single-qubit phase gates can be realized via virtual $Z(\theta)$ gates, which shift the phase of all subsequent waveforms.

For single-qubit gates, $R_\phi(\theta)$, the rotation angle θ is realized by the pulse duration, and the phase ϕ is the programmed relative phase between both legs of the Raman transition. These gates are constructed by

putting both tones of the Raman transition on the qubit's respective individual addressing (IA) beam, a copropagating and motional-insensitive configuration. Additionally, the single-qubit gates are run in parallel and mitigate first-order crosstalk through unique single-photon detunings [43].

The two-qubit gate, based on a Mølmer-Sørensen (MS) interaction, is enabled by the shared motional modes of the ion chain, and we can implement a gate between any pair by on-demand driving of these modes to generate the desired entanglement. In particular, we use the IA beams, along with a global addressing beam (propagating counter to the IA beams), to perform the necessary motionally sensitive red- and blue-detuned Raman transitions to generate each $ZZ(\theta)$ gate pair [28,29].

While the core two-qubit interaction in the system is about an equatorial rotation axis (XX-like) [44], we convert it to $ZZ(\theta)$ via a phase-agnostic formulation [4,45–47] that limits phase-dependent optical crosstalk [43]. This conversion is achieved through the use of single-qubit counterpropagating "wrapper" gates $R^{(\mathrm{wr})}_{\pm\pi/2}(\pi/2)$ which are also needed to combat the phase instabilities which arise from path length differences from mixing counterpropagating two-qubit gates and the standard copropagating single-qubit gates [45]. To realize the continuously parameterized entangling angle, θ, the key degree of freedom exploited in this work, we adjust a Raman beam amplitude [36] while keeping all $ZZ(\theta)$ gates equivalent in duration (250 μs). The gate has a Gaussian pulse shape and utilizes a constructive combination of radial motional modes (mode balancing) for frequency robustness when possible [40]. In particular, in a four-qubit register all $ZZ(\theta)$ take advantage of this mode balancing. However, in a five-qubit register, the center qubit does not participate in several motional modes [48], and therefore $ZZ(\theta)$ involving the center qubit are unable to utilize mode balancing and instead have only one motional mode contributing significantly.

Typical single-qubit gate fidelities, $R_\phi(\pi/2)$, are around 0.993 while typical two-qubit gate fidelities, $ZZ(\pi/2)$ are about 0.96–0.99 depending on the pair. The virtual single-qubit phase gate, $Z(\theta)$, is assumed perfect. Estimated fidelities for each experimental run are provided in Sec. V and Appendix A.

Empirically, we observe that smaller entangling angle leads to improved two-qubit gate performance on the QSCOUT system. A detailed discussion of the performance and realization of these gates can be found in Ref. [36], while a discussion of theoretical error scalings of the gate can be found in Ref. [40]. Likewise, on other trapped-ion processors, similar error scaling behavior has been reported [16,19]. Because of improved gate performance at smaller angles, here we aim to both develop and test compilation optimizations using SUPERSTAQ to leverage this continuously parameterized two-qubit gateset within an all-to-all connected trapped-ion register.

III. COMPILATION OPTIMIZATIONS

Here we describe the basics of our circuit decomposition and optimization procedures for compiling arbitrary circuits for QSCOUT. Because we expect two-qubit gates to be the dominant source of error on QSCOUT, and the said error to scale roughly with the absolute rotation angle $|\theta| \leq \pi/2$ of each $ZZ(\theta)$ gate, the primary goals of our compiler optimizations are to minimize both the number and total angle of ZZ gates in the compiled circuit.

Our workflow for compiling circuits for QSCOUT comprises the following steps:

1. Merge adjacent two-qubit operations.
2. Decompose the merged two-qubit unitaries into $ZZ(\theta)$ operations and arbitrary single-qubit gates.
3. Merge adjacent single-qubit operations.
4. Decompose the merged single-qubit unitaries into $R_\phi(\theta)$ operations and virtual phase gates.

Due to the register's all-to-all connectivity, no additional gates or operations are needed to bridge missing qubit connectivity between two-qubit operations.

Single-qubit unitaries can always be implemented with a single physical pulse via their Euler decomposition,

$$U = e^{i\gamma Z/2} e^{i\beta X/2} e^{i\alpha Z/2} = Z(\gamma + \alpha) R_\alpha(\beta), \qquad (1)$$

where the leading term is implemented virtually.

In the following, we describe our baseline two-qubit gate implementation, and then four optimizations intended to improve performance over the baseline circuit construction. The first two reduce the overall total ZZ rotation angle. The next approach utilizes the all-to-all connectivity to reassign qubit roles based on gate performance, while the final optimization aims to reduce the overall number of two-qubit gates through circuit approximation.

A. Baseline: Fixed two-qubit gate

As a baseline for comparisons, we restrict our gateset and decompose two-qubit unitaries to static $ZZ(\pi/2)$ entangling gates—ignoring (for now) the hardware's ability to implement arbitrary rotation angles. A well-known result, Cartan's KAK decomposition, [49] guarantees that any operation in $SU(4)$ can be decomposed into at most three such entangling pulses, interleaved with and surrounded by single-qubit operations on each qubit. As a result, the total ZZ rotation angle, Θ_U, for each Haar-random $SU(4)$ operation composing the quantum volume (QV) circuits considered, is $\Theta_U = 3\pi/2$.

B. Continuously parameterized two-qubit gate

We can decompose arbitrary two-qubit unitaries to QSCOUT's continuously parameterized $ZZ(\theta)$ operations by observing their KAK decompositions [49]. Given an arbitrary unitary operation $U \in SU(4)$, we can always compute

$$U = (C \otimes D)e^{ixXX/2}e^{iyYY/2}e^{izZZ/2}(A \otimes B), \qquad (2)$$

where A, B, C, D are single-qubit unitaries and $\pi/2 \geq x \geq y \geq |z| \geq 0$ are the locally invariant Weyl coordinates describing the entangling properties of the operation. The XX and YY entangling terms can be converted to ZZ operations via conjugation by $R_{\pi/2}(\pi/2)^{\otimes 2}$ and $R_0(\pi/2)^{\otimes 2}$, respectively. The total ZZ rotation angle $\Theta_U = x + y + |z| \leq 3\pi/2$ is then upper-bounded by that of the baseline decomposition, with a mean (for Haar-random unitaries) of $3\pi/4$. Using this decomposition in step 2 of the above compilation procedure, we therefore reduce the average rotation angle $\mathbb{E}[|\theta|]$ of $ZZ(\theta)$ operations in QV circuits by a factor of 2 relative to the baseline's static gateset. The single-qubit rotations are then merged with the other single-qubit unitaries and decomposed into QSCOUT's gateset as before.

C. Swap mirroring

As described in [4], we can further reduce the total ZZ cost by selectively inserting pairs of SWAP gates (or "swaps") alongside existing two-qubit gates in the circuit. By performing one of the inserted swaps virtually (i.e., by relabeling qubits) and absorbing the second into the gate U to be decomposed, we are left with a new gate U' with Weyl coordinates $(\pi/4 - x, \pi/4 - y, z - \pi/4)$. During the decomposition of merged two-qubit unitaries (step 2 of the compilation sequence outlined above), we therefore choose to insert swaps whenever U' would require either fewer total ZZ gates, or the same number of ZZ gates but a smaller total ZZ angle than the original unitary U (i.e., $\Theta_{U'} < \Theta_U$). This choice further reduces the maximum necessary Θ_U per selected unitary (U or U') to $3\pi/4$ and average to about 2.0 radians (determined empirically by sampling 10^4 Haar-random two-qubit unitaries)—a mean reduction of just over 15%. The net permutation incurred by the inserted virtual swaps is tracked in order to correctly map measurement outcomes on physical qubits to their original (circuit) positions. QSCOUT's all-to-all connectivity ensures that the propagation of virtual swaps does not add any complexity in terms of routing to the compiled circuit.

D. Gate ranking

In practice, interactions between different pairs of ions are not always equal. The fidelity of ZZ gates can vary between different qubit pairs due to the calibration of the drive lasers and geometry and interaction strength of the shared motional modes. QSCOUT's all-to-all connectivity means we are free to arbitrarily assign circuit qubits to physical qubits (i.e., ions) on the processor. We take further advantage of this connectivity for noise-aware qubit allocation [50], in order to best utilize the highest-performing qubit pairs. After inserting mirror swaps, we sum the total rotation angle of ZZ gates on each pair of circuit qubits in the decomposed circuit. We then choose a physical qubit assignment by minimizing the cost function $C = \sum_{ij} \epsilon_{ij} |\theta_{ij}|$, where θ_{ij} is the ZZ rotation angle that would occur on physical qubits (i, j) in the mapped circuit, and ϵ_{ij} is the expected rate of error for ZZ gates on those qubits. For the small circuits considered here we employ a simple brute-force approach to optimize qubit assignment; we discuss how this approach could be generalized to support larger ion chains in Sec. III F.

While the procedure described here is applicable to computations run within a single ion chain, we envision a similar procedure being applicable for a multichain architecture in which applications are first mapped to make the most use of the best-performing ion chains, and then qubits are allocated within each chain according to the performance of individual ion pairs.

E. Circuit approximation

The compilation processes above sometimes result in gates with very small rotation angles. Though we expect that gate error to increases with rotation angle, we also expect a finite amount of error due to implementing the gate at all (due both to the implementation itself and to the extended runtime due to the constant duration of ZZ operations). For a gate that is sufficiently well approximated by an identity operation, these errors can outweigh those which would be incurred from eliding the gate altogether. To this end, the compilation stack includes a variable tolerance parameter fixing the minimum rotation angle θ_{min} below which gates will be dropped in order to minimize the total number of ZZ pulses in the circuit.

Approximate decomposition provides further opportunity for gate elimination when used in conjunction with mirror swapping. In this case, if a gate $ZZ(\theta)$ is sufficiently well approximated by a $\pi/2$ rotation (i.e., $|\theta - \pi/2| \leq \theta_{\text{min}}$), the insertion of a mirror swap will mean we can drop that gate entirely. We therefore make the mirror swapping insertion pass aware of the approximation tolerance parameter, and always insert (do not insert) a mirror swap if the resulting decomposition will require fewer (more) entangling operations. A similar technique was employed in [15], in which mirror swaps were used to reduce the number of controlled-NOT gates required for approximating two-qubit unitaries in QV circuits.

F. Scalability

The baseline compilation steps, optimized gate decompositions, swap mirroring, and circuit approximation steps described here can all be implemented gate by gate, and therefore have complexity which is roughly linear in circuit size. However, the generic gate-ranking procedure described above does not efficiently scale to larger arrays. The complexity of the brute-force solver used for optimal qubit assignment grows factorially with the number of qubits being assigned. We find empirically that this approach can handle up to nine or ten qubits efficiently (i.e., with runtime comparable to the rest of the compilation toolchain). While this is sufficient for the four- and five-qubit circuits demonstrated in this work, it is likely prohibitive when compiling for a general architecture comprising subarrays with tens of qubits each. Commonly used strategies for noise-aware layout based on iterating over (sub)graph isomorphisms [51,52] offer little advantage in the case of all-to-all connectivity due to the breadth of the search space. Similar results can be achieved by simply cutting off the brute-force solver after sampling a limited subset of possible assignments. Alternative search methods for approximate qubit allocation algorithms [53–55] are likely to yield better results, but are not explored in this work.

However, we note that optimal gate ranking becomes tractable for a number of physically motivated cases (including those studied below). In particular, if two-qubit gate error is dominated by a particular set of similarly poorly-performing qubit pairs, gate ranking amounts to finding the qubit assignment which minimizes the sum total ZZ rotation angle on these pairs. While still computationally difficult in general, this assignment can be solved efficiently in some common cases. For example:

(1) If the poorly-performing "bad pairs" are disjoint (i.e., no two pairs share a qubit), we can treat qubit assignment as an instance of (fixed-cardinality) minimum-weight matching. This can be solved efficiently using variants of Edmond's Blossom algorithm [56,57].

(2) If two-qubit gate error is instead dominated by all pairs involving a small number (n_b) of "bad qubits," we can modify our brute-force approach to consider only the assignment of those qubits. The complexity is then reduced to $\binom{n}{n_b}$, easily manageable for a subarray with tens of qubits if a small fraction are known to perform poorly.

(3) An analogous approach would be applicable to a system with just a few "bad pairs" in a larger array, with complexity corresponding to the number of ways these pairs of qubits can be assigned to pairs of qubits in the input circuit.

IV. EVALUATION APPROACHES FOR COMPILATIONS WITH RANDOM CIRCUITS

A. Circuit choice

To examine the impact of these various compilation optimizations harnessing the continuously parameterized gateset and full connectivity, we generate randomized circuits across all qubits in the register. These types of circuits are used in the prevalent QV benchmarking approach [15,58,59]. In particular, these randomized circuits are a square measure consisting of n ions at a depth of $d = n$ layers. Each layer within the circuit consists of Haar-random SU(4) operations applied to randomly selected pairs of qubits. For each layer, a permutation is performed such that new pairings are available for subsequent layers. An important point is that in some cases (especially for the qubit register size demonstrated here, $n = 4$ or 5), the permutation might select the same pairing(s) from the previous layer, in which case the two SU(4) operations can be merged. As described in Sec. III, this approach is used for all circuits.

B. Analysis approaches

Here we introduce three analysis approaches to quantify improvements realized by the various compilation techniques. The first two are variants on the typical heavy output analysis, one we call the "unaware heavy output probability," h_U, as well as the standard approach, which we term the "aware heavy output probability," h_A. The final analysis approach is the Hellinger infidelity, $\mathcal{I}_H$, used to measure the closeness of the empirical probability distribution to the ideal distribution. In addition to quantifying performance improvements, these techniques also help to elucidate error channels in our system. We now describe each in detail below.

The first metric, the unaware heavy output, h_U, is the sum of all output probabilities greater than the median of the resulting probability distribution. This rudimentary version is susceptible only to incoherent/stochastic errors (dephasing, depolarizing, etc.), and not to coherent errors because it is ignorant of the ideal outputs and only examines the empirical outputs, M, of a given circuit to assess its heaviness. For all 2^n possible observable qubit states (bit strings), $q \in \{0,1\}^n$, we find the states

$$q_M = \{q \in \{0,1\}^n \text{ such that } M(q) > \tilde{M}\}, \tag{3}$$

where $\tilde{M}$ is the median of the set of (measured) output probabilities, M. The unaware heavy output, h_U is thus

$$h_U = \sum_{q \in q_M} M(q). \tag{4}$$

Because the median $\tilde{M}$ is computed from the empirical output distribution, measurement of h_U requires the number of measurements per circuit to be large relative to 2^n.

The second metric is the standard approach to QV benchmarking. The aware heavy output, h_A, is the sum of those empirical outputs whose corresponding ideal outputs (simulated) lie above the ideal median [15,58], thus penalizing both coherent errors and stochastic errors. With this measure, we find the states q_S from the ideal simulated output distribution, S, such that

$$q_S = \{q \in \{0,1\}^n \text{ such that } S(q) > \tilde{S}\}, \tag{5}$$

where $\tilde{S}$ is the median of the S ideal set. We take only those qubit states whose *ideal* output is classified as heavy, and then sum the corresponding empirical outputs:

$$h_A = \sum_{q \in q_S} M(q). \tag{6}$$

Because the median $\tilde{S}$ is computed from the ideal distribution rather than the empirical output, measurements of h_A do not depend on having a high number of shots per circuit (provided the number of random circuits is sufficiently high)—projection noise instead depends on the number of measurements taken across all random circuits. This can make this measurement significantly less costly experimentally, especially for larger numbers of qubits. However, classically computing the ideal distribution to find q_S itself scales exponentially with n, and thus there exists a tradeoff between quantum and classical costs for h_A and h_U. At our register sizes, neither of these factors is significant, and thus we find that computing h_U alongside h_A can provide us with further insights into the dominant errors impacting the system.

In the limit of large n, we expect both heavy output probabilities to approach $\mathbb{E}[h_U] = \mathbb{E}[h_A] = (1 + \ln 2)/2 \approx 0.847$ for Haar-random pure quantum states. For smaller n this value needs to be reduced slightly [58]—for example, we find empirically that $\mathbb{E}[h_U] = \mathbb{E}[h_A] \approx 0.839$ for Haar-random states at $n = 5$, and $\mathbb{E}[h_U] = \mathbb{E}[h_A] \approx 0.831$ at $n = 4$. Approaching a fully mixed state, we expect the outputs to approach a uniform distribution and therefore the probability of measuring a heavy output to approach $1/2$ (independent of n).

For h_U this floor increases slightly when the number of samples approaches 2^n, as the expected median $\mathbb{E}[\tilde{M}]$ is biased below the mean $\mathbb{E}[M] = 2^{-n}$ due to projection noise. For example, we find $\mathbb{E}[h_U] \approx 0.554$ for 200 samples of a uniform distribution with $n = 4$. Because h_A uses the ideal distribution to compute $\tilde{S}$, it is not impacted by sample size in the same way.

We also note that the h_U and h_A measures are typically understood in terms of their mean corresponding to a large number of random circuits. From this, a quantum volume of 2^n is certified when both the h_A mean and its 2σ confidence interval (CI) lower bound lie above the 2/3 threshold for n qubits at a depth of $d = n$ layers. This 2/3 threshold is inspired by theoretical studies of quantum advantage and speedup [60]. Traditionally, the CI corresponds to a Wald interval [15], but more recently, bootstrapping methods have been able to yield tighter CIs, resulting in fewer circuits needed to certify the quantum volume [58]. In this work, however, we will use a more conservative measure than the Wald interval, the Wilson score interval, to determine 2σ intervals. The Wilson score interval is asymmetric and avoids the potential overshoot issues associated with the Wald interval especially for smaller sample sizes [61]. At significantly large numbers of circuits sampled, and with means sufficiently far from the limits, the Wald and Wilson score intervals converge.

The final metric we use is Hellinger infidelity. Hellinger fidelity is a measure similar to classical fidelity but is derived from the Hellinger distance, d_H, a measure of the distance between two probability distributions [62,63]. Here the two distributions are the ideal simulated distribution, S, and the experimentally observed distribution, M. Thus, d_H is defined as

$$d_H(S,M) = \frac{1}{\sqrt{2}} \sqrt{\sum_{q \in \{0,1\}^n} (\sqrt{S(q)} - \sqrt{M(q)})^2}. \tag{7}$$

The Hellinger fidelity, $\mathcal{F}_H$, is then

$$\mathcal{F}_H = (1 - d_H^2)^2, \tag{8}$$

and the Hellinger infidelity, $\mathcal{I}_H$, is simply

$$\mathcal{I}_H = 1 - \mathcal{F}_H. \tag{9}$$

Like the unaware heavy output probability, measurements of Hellinger infidelity are impacted by shot noise. For example, with 200 shots we expect shot noise to increase infidelities by about 0.02 for four-qubit QV circuits and 0.05 for five-qubit QV circuits. This is due to an increase in the total number of observable states within a probability distribution; that is, as n increases, the number of observable states grows as 2^n, leading to a significant scaling problem for Hellinger infidelity [63]. Therefore, in addition to a sufficient number of distinct circuits, a sufficient number of shots per circuit should be used.

All three measures help to give a picture of the underlying error sources of the continuously parameterized two-qubit gate. For each compilation optimization, we generally find improvements in all measures, but some demonstrate a stronger improvement than other measures. It is important to note each measure has its limitations due to sampling or convolution of differing effects, and utilizing all three provides a more holistic picture of the

comparison. They also diverge in terms of experimental overhead and scalability, with h_A requiring significantly fewer shots than either h_U or $\mathcal{I}_H$.

C. Distinguishing errors

The unaware heavy output probability h_U is totally independent of the ideal circuit outputs. Instead, it characterizes the degree to which measurement outcomes align with the expectations from random circuit sampling, that is, how well they converge upon the expected Porter-Thomas distribution of output probabilities. In most cases a circuit with coherent (unitary) errors still resembles a random circuit—but not the circuit we are attempting. Because the calculation of h_U is agnostic toward the ideal circuits, it is largely insensitive to these coherent errors so long as the collection of faulty circuits is still approximately Haar random. While there are cases for which Haar randomness is not preserved (take, for example, the limiting case of systematic 100% underrotation of all gates, that is, no gates are performed at all), we find that in practice this divergence is negligible for the types and magnitudes of coherent errors that could feasibly be present in our experiments. Stochastic errors on the other hand are not information-preserving, and so are readily observed as changes in h_U.

On the other hand, the aware heavy output probability h_A *does* take into account the ideal output distribution of the circuit, and therefore is expected to reflect both stochastic and coherent errors. The comparison between these values provides us with a simple heuristic for distinguishing between certain classes of errors: with no errors, we expect the two metrics to be identical; for stochastic noise we expect a shift in heavy outputs from both metrics; and for coherent errors we expect a drop in h_A but not h_U.

We validate this behavior empirically by simulating 400 compiled four-qubit QV circuits subject to a variety of coherent error sources and stochastic noise models, described in Table I. The resulting unaware and aware heavy outputs are presented as a function of error rate in Fig. 1. As expected, we see a clear dependence of h_U on error rate for all of the stochastic error channels (dashed lines). This dependence disappears for the coherent channels (solid lines), which show almost no variation except at the highest rates (which are well beyond what we could reasonably expect to see on the hardware). For the same set of circuits, we see that h_A has a clear dependence on error rate for both the stochastic and coherent channels. In Sec. VI we will use this analysis to estimate the amount

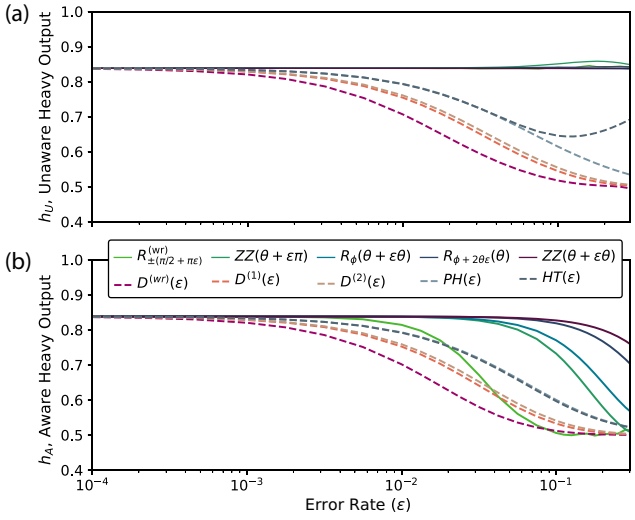


FIG. 1. Simulated heavy output probabilities h_U (a) and h_A (b) for 400 quantum volume four-qubit circuits as a function of error rate for a variety of coherent (solid lines of varying color) and stochastic (dashed lines of varying color) error models, demonstrating the negligible impact of coherent errors on unaware heavy outputs. See Table I for descriptions of the individual models. We also see that h_U can also be slightly more sensitive to some stochastic errors than h_A; this is especially noticeable for $HT(\epsilon)$ at very high error rates, where we see a slight uptick in h_U due to the resulting bias toward measuring qubits in their excited states.

TABLE I. Description of noise models used in error simulations.

Coherent error channels:	
$ZZ(\theta+\epsilon\theta)$	Relative overrotation of $ZZ(\theta)$ gates
$ZZ(\theta+\epsilon\pi)$	Fixed overrotation of $ZZ(\theta)$ gates
$R_\phi(\theta+\theta\epsilon)$	Relative overrotation of $R_\phi(\theta)$ gates
$R_{\phi+2\theta\epsilon}(\theta)$	Angle-dependent phase error in single-qubit $R_\phi(\theta)$ gates
$R^{(\mathrm{wr})}_{\pm(\pi/2+\pi\epsilon)}$	Phase errors in counterpropagating "wrapper" $R^{(\mathrm{wr})}_{\pm\pi/2}(\pi/2)$ gates
Stochastic error channels:	
$D^{(1)}(\epsilon)$	Single-qubit depolarization: X, Y, or Z error with probability $\epsilon/3$ after each nonwrapper $R_\phi(\theta)$ gate
$D^{(\mathrm{wr})}(\epsilon)$	Single-qubit depolarization of wrapper gates only
$D^{(2)}(\epsilon)$	Two-qubit depolarization: X, Y, or Z error with probability $\epsilon/3$ on each qubit after a $ZZ(\theta)$ gate
$PH(\epsilon)$	Dephasing: time-dependent phase error between gates with characteristic decay rate ϵ/T_{ZZ}, where $T_{ZZ}=250\ \mu$s is the duration of a single ZZ gate
$HT(\epsilon)$	Heating: time-dependent projection into the $\|1\rangle$ state with characteristic heating rate ϵ/T_{ZZ}

of stochastic noise in our experiments, and observe how it is affected by the various compiler optimizations described in Sec. III.

V. EMPIRICAL COMPARISONS OF COMPILATION APPROACHES

With these various evaluations, we now turn our attention to comparing these compilation optimizations on hardware, the trapped-ion QSCOUT system described in Sec. II.

A. Continuously parameterized two-qubit gate

We begin by comparing a set of 200 randomly selected QV circuits, generated via QISKIT's quantum volume library [7], on a four-qubit register with a depth of four layers. These 200 circuits are first compiled via SUPERSTAQ with only fixed entangling gates, $ZZ(\pi/2)$, as described in Sec. III to act as our baseline. These same circuits are then compiled using the continuously parameterized entangling gate, $ZZ(\theta)$. During the experimental run, the circuits from the different sets (along with the swap mirrored circuits described in Sec. V B) are interleaved such that they are all susceptible to the same noise and drift during the course of the experimental run. In particular, each version of the circuit is run 200 times before moving on to the next variation of that circuit. After all variations of a particular circuit are run, the next distinct circuit and its different compiler variations are run. The resulting qubit state output probabilities are then analyzed with the methods described in Sec. IV. For these experiments and all subsequent ones, the total number of circuits and shots performed is chosen to maximize the effectiveness of each analysis metric (h_U, h_A, $\mathcal{I}_H$) while still keeping the experimental runtime within a single calibration window of the experiment (below about 30 min).

In Fig. 2(a), we see an increase in h_U for the circuits compiled with $ZZ(\theta)$ relative to those compiled with $ZZ(\pi/2)$. Likewise, in Fig. 2(b), while we see the overall heaviness is reduced in both cases due to additional coherent error, the relative increase in h_A of the $ZZ(\theta)$ circuits over the $ZZ(\pi/2)$ circuits is similar. There are a variety of ways to further quantify the improvement in performance as measured by h_U and h_A. When considering heavy output analyses, it might be prudent to measure the empirically determined mean heavy output probabilities relative to the ideal limits of the measurement (see Sec. IV B) rather than simply comparing the numeric increase. Here we consider

$$\Gamma_{h_\alpha} = 1 - \frac{\mathbb{E}[h_\alpha] - \bar{h}_{\alpha,c'}}{\mathbb{E}[h_\alpha] - \bar{h}_{\alpha,c}}, \tag{10}$$

in which c, c' refer to the two different compilations being considered, α refers to the aware A or unaware U case, and the limits $\mathbb{E}[h_U] = \mathbb{E}[h_A] \approx 0.831$ at $n = 4$. This acts as a quantifier of how much the distance to the ideal case has been reduced. Thus, we find the mean values of h_U and h_A of the $ZZ(\theta)$ compilation are about 32% and 29% closer to the ideal heavy output respectively than those of the $ZZ(\pi/2)$ compilation.

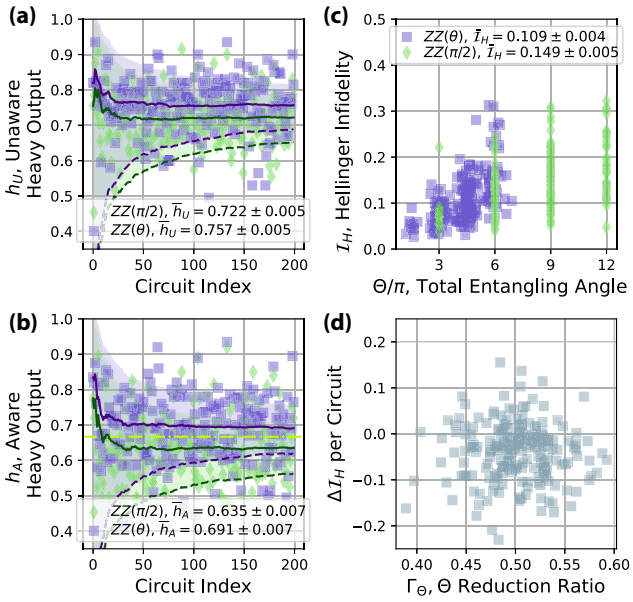


FIG. 2. Comparison of 200 QV circuits on a four-qubit register compiled with either fixed entangling gates, $ZZ(\pi/2)$ (green diamonds), or continuously parameterized entangling gates, $ZZ(\theta)$ (purple squares). Analysis with (a) unaware heavy outputs versus runtime circuit ordering, (b) aware heavy outputs versus runtime circuit ordering, (c) Hellinger infidelity versus total entangling angle, Θ, and (d) per-circuit change in Hellinger infidelity versus overall Θ reduction ratio, Γ_Θ. In (a),(b), each running mean is presented via a solid line, while 2σ Wilson score intervals are presented with shading in between the bounds with the lower bound presented via a dashed line. In (b), the 2/3 QV threshold is presented as a yellow dash-dotted line.

This compilation technique also transforms a series of circuits whose h_A mean probability lies below the QV threshold of 2/3 to a series of circuits whose h_A mean probability surpasses the threshold (albeit without a sufficient number of circuits for the 2σ CIs to surpass the threshold).

Because the ideal heavy outputs vary significantly between individual random circuits, we instead use $\mathcal{I}_H$ to elucidate potential trends in the data. In Fig. 2(c), we plot $\mathcal{I}_H$ versus the overall entangling angle (summed over all gates and qubits in the circuit) $\Theta = \sum_k |\theta_k|$, where θ_k is the entangling angle for each $ZZ(\theta)$ gate k within the circuit. Here, we see a clear trend of increasing infidelity with increasing Θ. There is also distinct clustering in the data which is reflective of the role that SU(4) merges play on the circuit length. There are four sets of clusters along Θ in each dataset (note that the clusters of the $ZZ(\theta)$ compilation are more apparent in Fig. 3). Each cluster within one

compilation corresponds to a cluster in the other, for example the $\Theta = 12\pi$ cluster of the $ZZ(\pi/2)$ compilation corresponds to the same set of circuits in the $\Theta \sim 6\pi$ cluster of the $ZZ(\theta)$ compilation (for completeness, the $ZZ(\theta)$ clusters centered at $\Theta \sim [6\pi, 4.5\pi, 3\pi, 1.5\pi]$ correspond to the $ZZ(\pi/2)$ clusters at exactly $[12\pi, 9\pi, 6\pi, 3\pi]$, respectively). In general, we see each $ZZ(\theta)$ cluster occupies a lower range of Hellinger infidelities than its corresponding $ZZ(\pi/2)$ cluster.

In the aggregate, we can also assess the performance improvement in a similar manner to Eq. (10); however, since $\mathcal{I}_H$ is already measured relative to its ideal limit (i.e., $\mathcal{I}_H \to 0$), the metric is simply

$$\Gamma_{\mathcal{I}_H} = 1 - \frac{\mathcal{I}_{H,c'}}{\mathcal{I}_{H,c}} \tag{11}$$

for the two different compilations c, c'. We find $\Gamma_{\mathcal{I}_H} \sim 27\%$.

Next, we examine $\mathcal{I}_H$ on a per-circuit basis in Fig. 2(d). Here, each circuit is sorted by how much the total Θ has been reduced via a reduction ratio, Γ_Θ, defined as

$$\Gamma_\Theta = \frac{\Theta_c - \Theta_{c'}}{\Theta_c} \tag{12}$$

for the two compilation optimizations c, c'. In this example, they refer to the baseline $ZZ(\pi/2)$ versus $ZZ(\theta)$ compilations respectively. We see the entangling angle is reduced by $\Gamma_\Theta \sim 0.5$, and in general, we see a reduction in the per-circuit $\mathcal{I}_H$ as a result.

B. Swap mirroring

With those same 200 circuits described above, the next variation we test continues to compile the circuits with $ZZ(\theta)$ but with the added feature of swap mirroring described in Sec. III. This experiment was interleaved (as described above) with those in Fig. 2. For simplicity, we present only the comparison of the $ZZ(\theta)$ compilations with and without swap mirroring in Fig. 3.

The improvements for this approach are more subtle than those seen in switching from a fixed entangling gate to the continuously parameterized entangler. Both the h_U and h_A measures show distinct but smaller improvements in performance in Figs. 3(a) and 3(b) as compared to Figs. 2(a) and 2(b). The performance improvements, as assessed by Eq. (10), are thus $\Gamma_{h_U} \sim 22\%$ and $\Gamma_{h_A} \sim 11\%$.

These improvements are also seen in the $\mathcal{I}_H$ measures presented in Fig. 3(c) where $\Gamma_{\mathcal{I}_H} \sim 6\%$. We once again see the four distinct clusters from the differing degrees of SU(4) merges, and the swap-mirrored circuits all inhabit clusters at lower entangling angles (centered at $\Theta \sim [5.0\pi, 3.8\pi, 2.6\pi, 1.3\pi]$) than the corresponding clusters containing the same circuits compiled without swap mirroring (centered at $\Theta \sim [6.0\pi, 4.5\pi, 3.0\pi, 1.5\pi]$). However, the general trend of improvements in $\mathcal{I}_H$ between

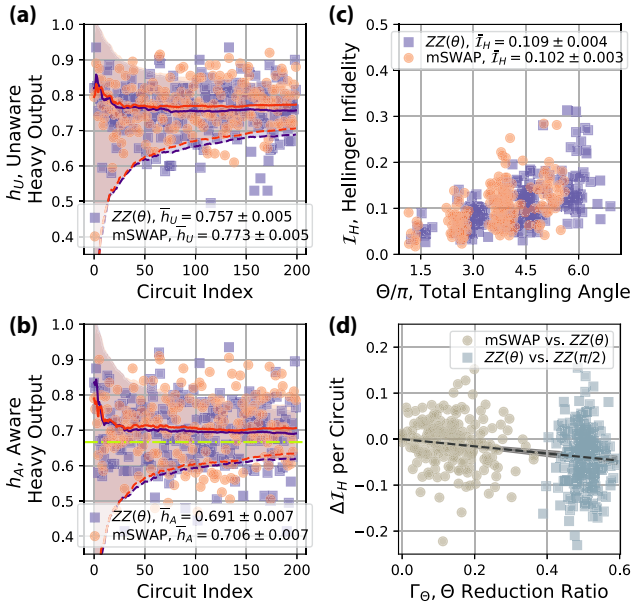


FIG. 3. Comparison of 200 quantum volume circuits on a four-qubit register compiled with continuously parameterized entangling gates, $ZZ(\theta)$, with swap mirroring (orange circles, denoted "mSWAP" in this and subsequent figures for conciseness) and without (purple squares). Analysis with (a) unaware heavy outputs versus runtime circuit ordering, (b) aware heavy outputs versus runtime circuit ordering, (c) Hellinger infidelity versus total entangling angle, Θ, and (d) per-circuit change in Hellinger infidelity versus overall Θ reduction ratio, Γ_Θ. Gold circles represent $\Delta\mathcal{I}_H$ for $ZZ(\theta)$ with swap mirroring versus $ZZ(\theta)$ without. Gray squares represent $\Delta\mathcal{I}_H$ for $ZZ(\theta)$ without versus $ZZ(\pi/2)$ (reproduced from Fig. 2). In (a) and (b), each running mean is presented via a solid line, while 2σ Wilson score intervals are presented with shading in between the bounds with the lower bound presented via a dashed line. In (b), the 2/3 quantum volume threshold is presented as a yellow dash-dotted line. In (d), a trend line is fitted to both sets (circles and squares) of $\Delta\mathcal{I}_H$ data with a 2σ error envelope. Note that Figs. 2 and 3 are the same 200 circuits, in which all three compilation choices were taken interleaved with one another.

corresponding clusters is, as expected, more subtle than in Fig. 2(c). Likewise, when the per-circuit Θ reduction, Γ_Θ, is plotted in Fig. 3(d) [together with the Γ_Θ data from Fig. 2(d)], we see a reduction in the $\mathcal{I}_H$ that follows a trend for decreasing infidelity with increasing Γ_Θ.

These more marginal improvements in both heavy output measures and $\mathcal{I}_H$ are to be expected as the overall reduction in total entangling angle is $\Gamma_\Theta \sim 0.15$ on average across all of the circuits. However, there are classes of circuits in which this effect is more pronounced. One such example is a previous demonstration examining a series of parameterized FSim gates, defined in [64]. For a subset of these gates, when compiled into the native gateset of QSCOUT, swap mirroring always reduced the entangling angle [4]. Not only did the overall entangling angle, Θ,

decrease when swap mirroring was enabled, but the entangling angle, θ, on *each* constituent ZZ gate also decreased leading to an average $\Gamma_\Theta \sim 0.61$ across 50 circuits tested. This yielded more drastic improvements in performance on the QSCOUT system [4], comparable to the average infidelity changes seen in Fig. 2.

C. Gate ranking

The next approach we investigate relies on knowledge of the machine's performance on a gate-by-gate basis. In particular, the compiler will attempt to reduce the total entangling angle on the lesser-performing two-qubit gates. For these demonstrations, we intentionally depress the performance of certain gates in order to highlight this compilation approach. Here, we will compile all circuits with $ZZ(\theta)$ entanglers and swap mirroring enabled.

1. "Bad pairs"

The first demonstration uses a four-qubit register, where we find the pairs between next-nearest neighbors are typically lower-fidelity operations (see Fig. 4). We refer to these pairs as "orthogonal" as they do not share any qubits with one another. These operations incur more optical crosstalk on the skipped qubit (i.e., the skipped qubit is susceptible to nearest-neighbor spillover light from *both* qubits around it, unlike all other pair configurations), and in particular, the amount of crosstalk will vary based on the alignment and beam profile of the individual addressing beams on any given day. As such, they are ideal for compiling around. To better emphasize the technique and to focus on the general behavioral trends of the system, we disentangle the day-to-day inconsistency of the hardware by intentionally depressing the fidelities of both of the next-nearest-neighbor gates to a similar level. We do so by mismatching the red- and blue-detuned Raman amplitudes—specifically, decreasing the blue-detuned Raman amplitude to 80% of the optimal performance. This incurs additional light shift (due to the Autler-Townes or ac Stark effect) on the participating qubits as the detuned carrier light shifts from red- and blue-detuned drives no longer cancel [65]. This additional physical light shift is canceled by a virtual *in-situ* Z gate, but the resulting additional dephasing noise from the stronger light shift remains [36].

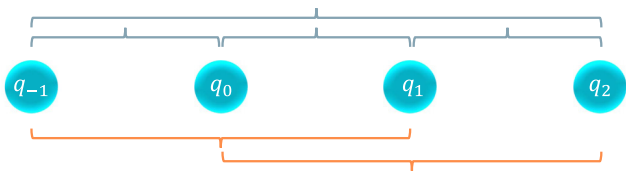


FIG. 4. Four-ion chain with all-to-all connectivity. Nearest-neighbor interactions and outer interaction (in gray brackets) are approximately 0.99 estimated fidelity. Next-nearest-neighbor interactions (in orange brackets) intentionally depressed to approximately 0.95 estimated fidelity.

TABLE II. Fidelity estimates of two-qubit gate pairs in a four-ion chain. Qubit name assignments are in Fig. 4. Performance of gate pairs $\{-1, 1\}$ and $\{0, 2\}$, highlighted in orange, has been intentionally depressed through a mismatch of amplitudes of the red- and blue-detuned Raman transitions during the $ZZ(\theta)$ interaction. To *estimate* the fidelity of the fully entangling interaction efficiently prior to runtime, we repurpose our light shift cancelation calibration consisting of two successive bare $MS(\pi/2)$ gate interactions (no "wrapper" gates to create the $ZZ(\pi/2)$ interaction) while varying the *in-situ* Z gate to cancel the light shift. We determine the greatest probability of $|11\rangle$ through a maximum likelihood estimation fit and extract its 2σ confidence intervals. To then estimate the fidelity of a maximally entangled state generated by a single gate, the square root of the $|11\rangle$ probability is used, while the uncertainties are derived from its 2σ confidence intervals.

Gate pair	Connectivity	Est. state fidelity
$\{-1, 0\}$	nearest neighbor	$0.987^{+0.004}_{-0.005}$
$\{0, 1\}$	nearest neighbor	$0.987^{+0.004}_{-0.005}$
$\{1, 2\}$	nearest neighbor	$0.989^{+0.004}_{-0.005}$
$\{-1, 1\}$	next-nearest neighbor	$0.954^{+0.007}_{-0.007}$
$\{0, 2\}$	next-nearest neighbor	$0.952^{+0.007}_{-0.007}$
{-1, 2}	outer	$0.988^{+0.004}_{-0.005}$

In Table II, we list the measured state fidelity estimates determined by measuring the $|11\rangle$ state after two successive $MS(\pi/2)$ gates (i.e., $ZZ(\pi/2)$ without the "wrapper" single-qubit gates to transform its basis). As this is a rough measure and used for relative comparisons as opposed to absolutes, no correction for state readout errors is implemented. These values act as an upper bound on the true fidelity of the operation, and we see, after our intentional miscalibration, the next-nearest-neighbor pairs are limited to an upper bound of about 0.953 for their fidelity.

We compile 400 randomly selected QV circuits with knowledge of the more poorly performing gates, referred to as "ranked" circuits, as well as without knowledge, referred to as "unranked." We find of the 400 circuits, approximately 2/3 of the circuits are able to take advantage of this compilation approach, and within that subset, more than half are able to remove participation of those gate pairs completely (see Appendix B).

The circuits are once again performed interleaved with one another at 200 shots, and we find that there is improvement in all three analysis measures in Fig. 5. Interestingly, we see a smaller gap between the h_U measures than in the h_A measures, suggesting we introduced substantial coherent error on our $ZZ(\theta)$ gates on the next-nearest-neighbor pairs [Figs. 5(a) and 5(b)]. When assessing the closeness to the ideal limit, we find reductions in distance of $\Gamma_{h_U} \sim 19\%$ and $\Gamma_{h_A} \sim 28\%$. These error channels are explored further in Sec. VI. In terms of quantum volume, when

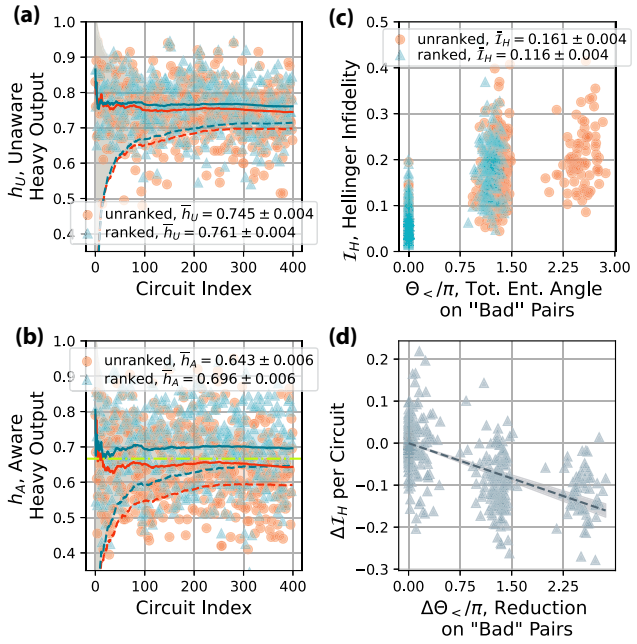


FIG. 5. Comparison of 400 quantum volume circuits on a four-qubit register (compiled with continuously parameterized entangling gates, $ZZ(\theta)$, with swap mirroring by default). with compilation that aims to put the largest entangling angle on the best-performing gates (teal triangles) versus no knowledge of gate performance (orange circles). Here, the two orthogonal (and next-nearest-neighbor) gate pairs were intentionally made worse. Analysis with (a) unaware heavy outputs versus runtime circuit ordering, (b) aware heavy outputs versus runtime circuit ordering, (c) Hellinger infidelity versus total entangling angle on the more poorly performing gates, $\Theta_<$, and (d) per-circuit change in Hellinger infidelity versus overall $\Theta_<$ reduction on the more poorly performing gates. In (a),(b), each running mean is presented via a solid line, while 2σ Wilson score intervals are presented with shading in between the bounds with the lower bound presented via a dashed line. In (b), the 2/3 quantum volume threshold is presented as a yellow dash-dotted line. In (d), a trend line is presented with a 2σ error envelope.

using the "unranked" circuits, which are agnostic to gate performance, the h_A mean remains below the QV threshold of 2/3; however, when compiling with knowledge of the "bad pairs," the h_A mean now lies above that threshold (once again, without a sufficient number of circuits for the 2σ CIs to cross the threshold which would have certified quantum volume).

In Fig. 5(c), the $\mathcal{I}_H$ measure shows a noisy but distinct worsening trend for increasing entangling angle on the lesser-performing gate pairs, $\Theta_< = \sum_i \theta_{i,p}$ for all pairs p that are lesser-performing. When examined on a per-circuit basis, and sorting the circuits based on how much total entangling angle has been removed from the "bad" pairs, $\Delta\Theta_<$, and reassigned to better-performing pairs (i.e., "offloaded"), we see a clear reduction trend in the $\mathcal{I}_H$ for increased offloading from those next-nearest-neighbor pairs [Fig. 5(d)]. Here, however, the clustering is less reflective of SU(4) merges and instead reflects different scenarios for how much entangling angle is offloaded from the more poorly performing gates and is described further in Appendix B. Additionally, when assessed in aggregate, we find $\Gamma_{\mathcal{I}_H} \sim 28\%$.

2. "Bad qubit"

Another scenario we evaluate is the case where one qubit's ZZ gate pairs all perform worse than the other pairs available in the chain. Here, we increase the qubit register to $n = 5$ (and similarly, the circuit depth, $d = 5$) to better demonstrate this scenario. In particular, QV circuits differ substantially between even- and odd-numbered registers. As the circuit construction relies upon pairing qubits within an SU(4) operation, odd registers will always have one qubit that does not participate within any given layer [15,58]. Therefore, for any given circuit, this gate-ranking scenario essentially recasts the least participatory qubit to be the worst-performing one. At larger odd-register sizes (and thus depths), this asymmetry in participation will be greatly diluted; however, this scenario would still be advantageous in other classes of circuits that display strong participation asymmetry, such as those in which an auxiliary qubit participates only marginally with the rest of the register.

Here, the leftmost qubit in the chain, q_{-2}, is deemed to be the "bad qubit" (Fig. 6). Here, we intentionally inject error into all of its corresponding gate pairs to ensure all its pairs' fidelity estimates are below every other ZZ gate pairs via the same mismatch of blue- and red-detuned amplitudes (at a ratio of 80%) that was done for the "bad pairs" scenario. Fidelity estimates are provided in Table III.

To examine this scenario, we generate 400 randomly selected circuits, once again performed at 200 shots and interleaved. Here we find small yet distinct improvements in the h_U and h_A measures [Figs. 7(a) and 7(b)]. This is similarly reflected in the metric from Eq. (10), as we only find reductions of $\Gamma_{h_U} \sim 5\%$ and $\Gamma_{h_A} \sim 4\%$. We note these improvements are more subtle than the case of the "bad pairs" scenario, likely due to less overall average offloading of entangling angle for this scenario. We also see

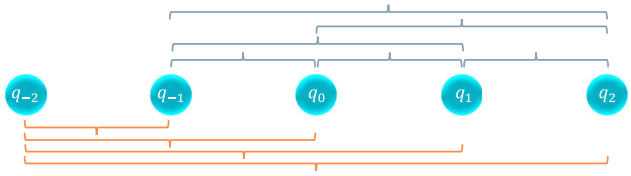


FIG. 6. Five-ion chain with all-to-all connectivity. All interactions with one of the outer qubits, q_{-2}, are intentionally depressed in estimated fidelity to about 0.89–0.92 (in orange brackets), while all other interactions have fidelity estimates of about 0.95–0.97 (in gray brackets).

TABLE III. Fidelity estimates of two-qubit gate pairs in a five-ion chain. Qubit name assignments are found in Fig. 6. The performance of all gate pairs involving q_{-2}, highlighted in orange, have been intentionally depressed through a mismatch of amplitudes of the red- and blue-detuned Raman amplitudes during the ZZ interaction. As described in the caption to Table II, fidelity estimates are determined via measurements taken just prior to runtime.

Gate pair	Connectivity	Est. state fidelity
$\{-2,-1\}$	nearest neighbor	$0.897^{+0.010}_{-0.010}$
$\{-1,0\}$	nearest neighbor	$0.965^{+0.007}_{-0.008}$
$\{0,1\}$	nearest neighbor	$0.961^{+0.007}_{-0.008}$
$\{1,2\}$	nearest neighbor	$0.968^{+0.007}_{-0.008}$
$\{-2,0\}$	next-nearest neighbor	$0.916^{+0.009}_{-0.010}$
$\{-1,1\}$	next-nearest neighbor	$0.961^{+0.007}_{-0.008}$
$\{0,2\}$	next-nearest neighbor	$0.959^{+0.008}_{-0.008}$
$\{-2,1\}$	third-nearest neighbor	$0.922^{+0.009}_{-0.009}$
$\{-1,2\}$	third-nearest neighbor	$0.947^{+0.008}_{-0.009}$
$\{-2,2\}$	outer	$0.890^{+0.010}_{-0.010}$

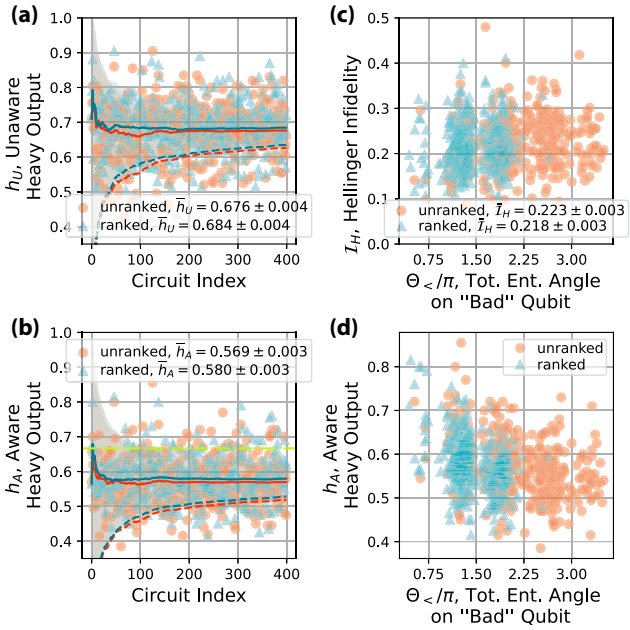


FIG. 7. Comparison of 400 quantum volume circuits on a five-qubit register (compiled with continuously parameterized entangling gates, $ZZ(\theta)$ with swap mirroring by default) with compilation that aims to put the largest entangling angle on the best-performing gates (teal triangles) versus no knowledge of gate performance (orange circles). Here, pairs involving q_{-2} were intentionally made worse. Analysis with (a) unaware heavy outputs versus runtime circuit ordering, (b) aware heavy outputs versus runtime circuit ordering, (c) Hellinger infidelity versus total entangling angle on the more poorly performing gates, $\Theta_<$, and (d) aware heavy outputs versus overall $\Theta_<$ reduction. In (a),(b), each running mean is presented via a solid line, while 2σ Wilson score intervals are presented with shading in between the bounds with the lower bound presented via a dashed line. In (b), the 2/3 quantum volume threshold is presented as a yellow dash-dotted line.

only a minor improvement in the $\mathcal{I}_H$ measure [Fig. 7(c)] with $\Gamma_{\mathcal{I}_H} \sim 5\%$. This minor change in the $\mathcal{I}_H$ is partially due to sample size susceptibility of this analysis approach (described in Sec. IV B) as a five-qubit chain has twice as many qubit bitstrings as a four-qubit chain, leading the measure to be subsumed by more shot noise (for the same number of shots) and thus an increased floor. To better quantify this difference, more experimental shots would be required. Instead, to examine the effects on a per-circuit basis, we look at the h_A measure as a function of reduced entangling angle on the "bad qubit," and see distinct improvements for smaller amounts of entangling angle in Fig. 7(d). In particular, the less distinct clustering seen here reveals the different participatory factors of q_{-2} convolved with various SU(4) merges.

D. Circuit approximation

Our final approach investigates the role that circuit approximation plays in improving empirical circuit performance at the expense of implementing the desired circuit exactly. Here, we use SUPERSTAQ to eliminate all entangling $ZZ(\theta)$ gates with an entangling angle, $|\theta|$, below a threshold of 0.10 radians along with the merging of the leftover single-qubit gates remaining. Like the gate-ranking investigations, all these experiments include continuously parameterized $ZZ(\theta)$ and swap mirroring as default options within SUPERSTAQ. A threshold of 0.10 was chosen as it eliminated a substantial number of $ZZ(\theta)$ gates present in the circuit (about 15%), while only adding an average infidelity of $\sim$0.006 (with a maximum infidelity of $\sim$0.025) to the approximate circuit when compared to the intended circuit (see Appendix B). Further study of the role that the threshold plays on the infidelity of the circuit (both in ideal simulation and in empirical practice) is warranted but beyond the scope of this survey of compilation optimization techniques.

Experimentally, we test 400 randomly selected circuits performed at 200 shots, interleaving the exact circuit and its approximation. There are distinct yet subtle improvements in the mean h_U and h_A measures presented in Figs. 8(a) and 8(b) represented by $\Gamma_{h_U} \sim 13\%$ and $\Gamma_{h_A} \sim 7\%$. Additionally, we see both the exact and approximate circuits certify a quantum volume of 2^4 as the h_A mean and the corresponding 2σ CIs are above the 2/3 threshold. Another measure to quantify the improvement from a given optimization is the number of circuits required to certify quantum volume (i.e., at what point the CI crosses the threshold). Here, the approximate circuits are able certify the QV measure at fewer total circuits (297 circuits) than the exact circuits require (397 circuits). Put simply, approximately 3/4 of the number of circuits are required

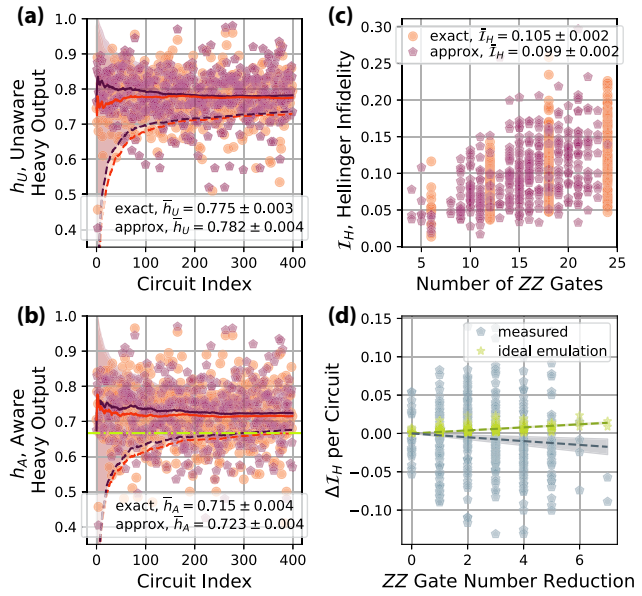


FIG. 8. Comparison of 400 quantum volume circuits on a four-qubit register compiled with continuously parameterized entangling gates, $ZZ(\theta)$, with swap mirroring. Compilation using the exact circuit required (orange circles) versus approximate circuits in which entangling gates with $|\theta| < 0.10$ were removed (magenta pentagons). Analysis with (a) unaware heavy outputs versus runtime circuit ordering, (b) aware heavy outputs versus runtime circuit ordering, (c) Hellinger infidelity versus total number of MS gates, and (d) per-circuit change in experimentally measured Hellinger infidelity between the approximate and exact measured circuits (gray pentagons) alongside the expected (i.e., emulated) Hellinger infidelity increase for approximating the circuits (yellow stars) plotted against the overall reduction in the number of MS gates. In (a),(b), each running mean is presented via a solid line, while 2σ Wilson score intervals are presented with shading in between the bounds with the lower bound presented via a dashed line. In (b), the 2/3 quantum volume threshold is presented as a yellow dash-dotted line. In (d), trend lines are presented with a 2σ error envelope.

to certify the QV benchmark when using approximate circuits.

With the $\mathcal{I}_H$ measure, we also see a small drop when approximating the circuits in Fig. 8(c) corresponding to a $\Gamma_{\mathcal{I}_H} \sim 6\%$. This is particularly notable as the approximated circuits are able to overcome their inherent additional infidelity incurred solely by approximation. This is more clearly seen in Fig. 8(d), where the experimentally determined per-circuit $\Delta\mathcal{I}_H$ is presented alongside the expected per-circuit $\Delta\mathcal{I}_H$ from pure simulation, both relative to the number of $ZZ(\theta)$ gates by which the circuit has been reduced. (The reduction in gates is a more relevant measure than Γ_Θ which remains virtually identical per circuit on average.) There is an upward trend in the simulated results as is to be expected, while there is a noisy yet still downward trend in the experimental results, indicating fidelity improvements gained by reducing the total number of entangling gates on average outweigh the loss in fidelity incurred by approximating the circuit.

VI. EVALUATING STOCHASTIC AND COHERENT ERROR CONTRIBUTIONS

In this section, we demonstrate how the analysis approaches outlined Sec. IV B can be applied to the empirical results in Sec. V in order to further illuminate details about the dominant errors affecting the system, and the impact of the compiler optimizations thereon. In particular, the use of both the aware and unaware heavy outputs in tandem provides a qualitative window into the relative roles of stochastic and coherent error sources.

We first use our experimental measurements of h_U to estimate the relative contributions of stochastic error. For each compilation strategy, we simulate the complete set of QV circuits with various simple noise models, and tune the probability of error such that the simulated h_U matches the experimentally measured value. This error probability is taken as an estimate of the stochastic error rate of the hardware, assuming that particular error channel was dominant. By comparing these error rates across experiments we gain some insight into the impacts of each of our compiler optimizations.

We demonstrate this procedure using four different stochastic error models from Table I: gate-driven depolarization after one-qubit gates, two-qubit gates, or the "wrapper" gates ($D^{(1)}$, $D^{(2)}$, $D^{(\mathrm{wr})}$) and time-based dephasing (PH). Note that our analysis is incapable of differentiating between these (or other) stochastic channels; the two channels are simply intended serve as representative examples indicating trends in stochastic error generally.

The predicted error probabilities assuming each noise channel are shown in Fig. 9(a). With each model we see a clear drop in stochastic error as a result of both parameterized $ZZ(\theta)$ decomposition and mirror swapping, reaffirming the θ dependence of two-qubit gate noise. We also see a reduction in stochastic error as a result of gate-ranking in both the "bad qubit" and "bad pair" compiled circuits, demonstrating the additional stochastic error resulting from our intentional gate miscalibrations and indicating that it is successfully being circumvented by the corresponding compiler pass. Because the approximate compilation pass removes small-angle gates but in general does not modify the remaining operations, we expect any per-gate reduction in error to be reflected in both the noisy simulations and the experimental results. In this case we do not find an appreciable difference between the stochastic error probability predicted from our experimental results with and without circuit approximation, which is consistent with a reduction in overall circuit error due to stochastic noise.

We then repeat each simulation with the addition of each coherent error channel described in Table I (with

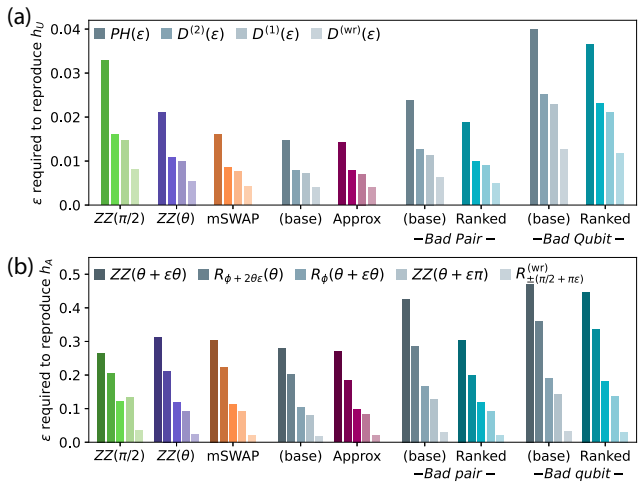


FIG. 9. (a) Simulated stochastic error rates needed to realize the empirically determined h_U for each compilation, assuming a single source of stochastic error. Each cluster of bars represents a single compilation method with four different error sources: *PH* (darker bars on left), $D^{(2)}$, $D^{(1)}$, and $D^{(\mathrm{wr})}$ (lighter bars on right). "Base" (gray) refers to the circuits without the studied optimization turned on, but with $ZZ(\theta)$ and swap mirroring. (b) Estimated coherent error rate ϵ needed to realize the measured h_A for each compilation method, assuming the dominant source is (from left to right within each cluster) $ZZ(\theta+\epsilon\theta)$, $R_{\phi+2\theta\epsilon}(\theta)$, $R_\phi(\theta+\epsilon\theta)$, $ZZ(\theta+\epsilon\theta)$, or $R^{(\mathrm{wr})}_{\pm(\pi/2+\pi\epsilon)}$. Note that in both (a),(b), due to differences in parameterization the relative strength each error channel is not directly comparable; rather each serves to indicate trends between the different compilation methods assuming one or the other source is dominant. See Table I for descriptions of each error model.

the stochastic error rates fixed to that predicted from our experimental h_U). We compare these results to the experimentally measured value of h_A to predict the coherent error rate, now assuming both a single dominant coherent and single stochastic channel. Notably, we find that these coherent error estimates are largely agnostic to the chosen stochastic channel, predicting the same coherent error rate (within $\pm 1\%$) regardless of the chosen stochastic noise model. We therefore average these predictions to arrive at estimates of the coherent error rate assuming each coherent channel.

The resulting error rate predictions are shown for a variety of coherent channels in Fig. 9(b). While the true error model is likely a combination of all of these sources (and others), isolating each one can be instructive in further elucidating the dominant coherent error mechanisms in the system. For example, we would require a 20–30% overrotation to account for our experimental results if that was the only source of coherent error, which is well outside the range of errors we can reasonably expect to be present on the system. By contrast, a much smaller phase error (induced by light shifts that are currently uncompensated) on just the single-qubit "wrapper" gates implemented as part of $ZZ(\theta)$ would account for the same h_A, making this a good candidate for additional benchmarking and calibration.

We can now examine the role the compilation optimizations have on the relative impact of coherent and stochastic error sources on the circuits. For instance, as we switch from the $ZZ(\pi/2)$ compilation to the $ZZ(\theta)$ compilation, there is a significant decrease in the stochastic error rate [Fig. 9(a)], but at the expense of a more marginal increase in the coherent error rate [Fig. 9(b)]. This is expected as there should be a small increase in rotation errors in a continuously parameterized gateset (which relies on an interpolated relationship between θ and laser amplitude) over entangling gates fixed at $\theta = \pi/2$. Likewise, in the "bad pairs" scenario, there is a larger decrease in the coherent error rate than in the stochastic error rate, revealing that the types of errors we are mitigating by this compilation are largely coherent. Thus, we have a method to see how much each compilation affects coherent and stochastic sources. It should be noted, however, that within a "family" of error sources (e.g., stochastic errors or coherent errors), this technique is unable to distinguish how the compilations might affect different channels with respect to one another.

In this way, our use of both the unaware and aware heavy output analyses in tandem provides us with insight into the relative roles of stochastic and coherent errors in our system. Unfortunately, neither analysis is scalable: while h_U circumvents the classical simulations required to determine ideal output distribution when computing h_A, it incurs experimental overhead as the number of required shots grows exponentially in system size. The utility of these methods is therefore likely limited to diagnostic probing of smaller subsets of a larger system in order to inform more targeted characterization experiments. We contrast this with other common methods for device characterization: randomized benchmarking and its derivatives [66–70], which can scale much more efficiently to larger systems but which tend to wash out coherent errors; and full gateset tomography [71], which precisely characterizes both coherent and stochastic errors (provided they are Markovian) but which incurs even higher experimental overhead making it impractical for more than two or three qubits.

VII. CONCLUSIONS

In summary, we have developed and demonstrated a series of circuit compiler optimizations within SUPERSTAQ to improve the performance of a trapped-ion processor. These compilations are tuned to the specific noise characteristics and performance of the two-qubit gates on the QSCOUT platform, leveraging the continuously parameterized two-qubit gateset as well as its inherent all-to-all connectivity. We first demonstrate a substantial

improvement in circuit performance just by utilizing this continuously parameterized two-qubit gateset over a more restrictive set of maximal entanglers only. We then realize further improvements through swap mirroring which further reduces the total entangling angle per circuit. We explore reassigning qubit roles to best utilize the highest-performing gates under two artificially constructed scenarios, one in which two orthogonal pairs perform worse than others and one in which all gate pairs associated with a single qubit perform worse. In both scenarios, we find that the circuits with knowledge of the gate fidelities outperform those without. We then demonstrate that circuit approximation to remove the least impactful two-qubit gates also can lead to marginal gains. Finally, we examine how different analysis approaches can elucidate the impact of these optimizations as well as estimate the role of stochastic versus coherent error channels.

Perhaps the most conspicuous result is that these optimizations can lead to the realization of a higher quantum volume for a processor (2^4 in this demonstration) when using specifically tuned and optimized circuits relative to using circuits with more restrictive gatesets or those unaware of gate fidelities. Additionally, better understanding the stochastic and coherent error channels of a particular platform provides pathways to investigate and mitigate those error sources.

This suite of circuit optimizations shows promise in the context not only of benchmarking like quantum volume, but also of other noisy-intermediate-scale quantum era applications. One such application is a quantum cavity electrodynamics simulation that demonstrated performance improvements when using these compilation techniques [72]. Additionally, while several optimizations presented here yielded marginal gains on these Haar-random circuits, more structured circuit designs could benefit greatly from these optimizations. One example is a class of FSim gates for which swap mirroring is always successful in which on average $\Gamma_\Theta \sim 0.61$, yielding an average improvement of approximately 40% in $\mathcal{I}_H$ [4]. Likewise, many circuits do not require all-to-all connectivity, and thus the gate-ranking methodology could potentially eliminate the need to even invoke the more poorly performing gates.

While these particular optimizations are tuned to the performance of the QSCOUT system, they are broadly applicable to other fully connected trapped-ion processors that take advantage of continuously parameterized two-qubit gates. Similarly, this study demonstrates a general methodology that tailors circuits directly to a hardware platform and in doing so yields performance improvements, increasing the hardware's utility.

ACKNOWLEDGMENTS

We thank Joshua Goldberg and Antonio Russo for QSCOUT software support and development. This material is supported by the US Department of Energy, Office of Science, Office of Advanced Scientific Computing Research under Award Number DE-SC0021526 and under its Quantum Testbed Program. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the US Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the US Department of Energy or the United States Government.

DATA AVAILABILITY

The data that support the findings of this paper are openly available [73].

APPENDIX A: EXPERIMENTAL FIDELITY ESTIMATES

In this appendix, for completeness, we present the fidelity estimates for the *ZZ* gates for the remaining datasets (whose estimates are not detailed in the main text). These fidelity estimates are determined across each pair in the chain prior to the each measurement presented in the main text. As these are measured prior to runtime, a full fidelity estimate per pair is too time-intensive to be practical. A discussion on how actual fidelity estimates compare to these upper bound estimates on the QSCOUT system can be found in Ref. [74]. Table IV presents the estimates determined prior to the experiments described in Sec. V A, which compare compilations using fixed entanglers versus those using continuously parameterized entanglers, as well as those described in Sec. V B, which compare the same compilation using continuously parameterized entanglers

TABLE IV. Fidelity estimates of two-qubit gate pairs in a four-ion chain taken prior to the measurements presented in Figs. 2 and 3, comparing fixed entanglers versus continuously parameterized entanglers versus continuously parameterized entanglers with swap mirroring enabled. Fidelity estimates are determined in the same fashion as described in Tables II and III. Qubit name assignments can be seen in Fig. 4.

Gate pair	Connectivity	Est. state fidelity
$\{-1, 0\}$	nearest neighbor	$0.988^{+0.004}_{-0.005}$
$\{0, 1\}$	nearest neighbor	$0.984^{+0.005}_{-0.005}$
$\{1, 2\}$	nearest neighbor	$0.982^{+0.005}_{-0.006}$
$\{-1, 1\}$	next-nearest neighbor	$0.977^{+0.005}_{-0.006}$
$\{0, 2\}$	next-nearest neighbor	$0.956^{+0.007}_{-0.007}$
$\{-1, 2\}$	outer	$0.988^{+0.004}_{-0.005}$

TABLE V. Fidelity estimates of two-qubit gate pairs in a four-ion chain taken prior to the measurements presented in Fig. 8, comparing exact circuit formulations versus approximate circuit formulations. Fidelity estimates are determined in the same fashion as described in Tables II and III. Qubit name assignments can be seen in Fig. 4.

Gate pair	Connectivity	Est. state fidelity
$\{-1,0\}$	nearest neighbor	$0.978^{+0.005}_{-0.006}$
$\{0,1\}$	nearest neighbor	$0.984^{+0.005}_{-0.005}$
$\{1,2\}$	nearest neighbor	$0.979^{+0.005}_{-0.006}$
$\{-1,1\}$	next-nearest neighbor	$0.967^{+0.006}_{-0.006}$
$\{0,2\}$	next-nearest neighbor	$0.958^{+0.007}_{-0.007}$
$\{-1,2\}$	outer	$0.987^{+0.004}_{-0.005}$

versus a compilation which also adds the swap mirroring optimization.

Table V presents the fidelity estimates determined prior to the experiment described in Sec. V D, which compares a compilation using the exact circuit formulation versus an approximation by removal of the smallest $ZZ(\theta)$ gates with $\theta < 0.10$ radians.

APPENDIX B: COMPILATION OPTIMIZATION STATISTICS

In this appendix, we present further analysis of the circuits being performed for some of the compilation optimizations studied. Additionally, we present further statistical analyses on all of the data collected from the experimental apparatus (corresponding to all compilations studied). We begin by analyzing the circuits themselves to better understand when and how a particular compilation affects each particular randomly generated circuit.

For the "bad pairs" scenario in which two orthogonal pairs in a four-qubit register have diminished performance (as described in Sec. V C 1), we find four distinct possibilities for how the compilation affects any given circuit [Fig. 10(a)] across all 400 circuits used. First, there are 71 circuits (17.75%) that have no participation from those pairs to begin with and are unaffected. Another 61 circuits (15.25%) are unaffected by this compilation as all pairs are involved in the circuit and those two orthogonal pairs are already participating with the least entangling angle possible. A total of 113 circuits (28.25%) partially offload participation of the orthogonal pairs, while the remaining 155 circuits (38.75%) offload participation of those two pairs completely. Thus, at this register size of four qubits, we find this compilation was effective for approximately 2/3 of the circuits. We define an offloading ratio, Γ_{ol} [similar to the reduction ratio Γ_Θ defined in Eq. (12)], which is the reduction in $\Theta_<$ between the unranked and ranked circuits as compared to the total $\Theta_<$ in the unranked circuit,

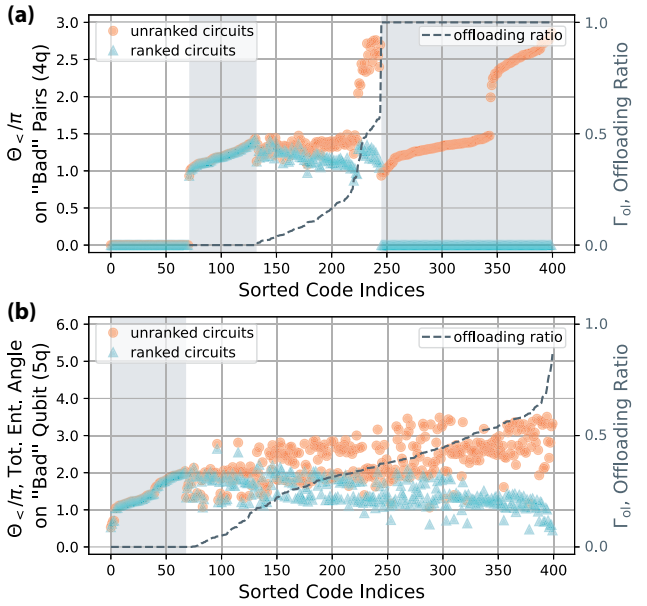


FIG. 10. Per-circuit analysis of total entangling angle, Θ, on the (a) orthogonal "bad pairs" in a four-qubit register and (b) the "bad qubit" in a five-qubit register. The original circuits (orange circles) and ranked circuits (teal triangles) are sorted to reveal different groups of the effectiveness of the ranking compilation and shaded accordingly. In (a), the first 71 circuits have no participation of the "bad pairs" regardless. The next 61 are circuits in which participation of the "bad pairs" already has the lowest possible Θ. The following 113 reveal circuits where partial offloading is achieved, and the remaining 155 indicate circuits in which complete offloading of Θ is realized. In (b), the first 67 circuits are those in which the participation of the "bad qubit" is already minimized. The remaining 333 are circuits where partial offloading occurs. The dashed gray line indicates the offloading ratio per circuit and corresponds to the right axis in both plots.

or:

$$\Gamma_{\mathrm{ol}} = \frac{\Theta_{<,u} - \Theta_{<,r}}{\Theta_{<,u}} \tag{B1}$$

in which u, r refer to the unranked and ranked circuits, respectively. This is plotted in Fig. 10 with a dashed line referenced to the right axis.

With the "bad qubit" scenario in the five-qubit register (Sec. V C 2), the optimizer works to relabel the qubit to reduce its entangling angle largely based on how many layers it participates in. In the 400 circuits tested, this leads to a partial offloading on 333 circuits (83.25%), and the remainder are identical to the original circuit. In Fig. 10(b), the overall entangling angle on the "bad qubit" is presented for both the ranked and unranked 400 circuits. In particular, there are no cases of complete offloading, so all circuits are still affected by the more poorly performing gates to some degree.

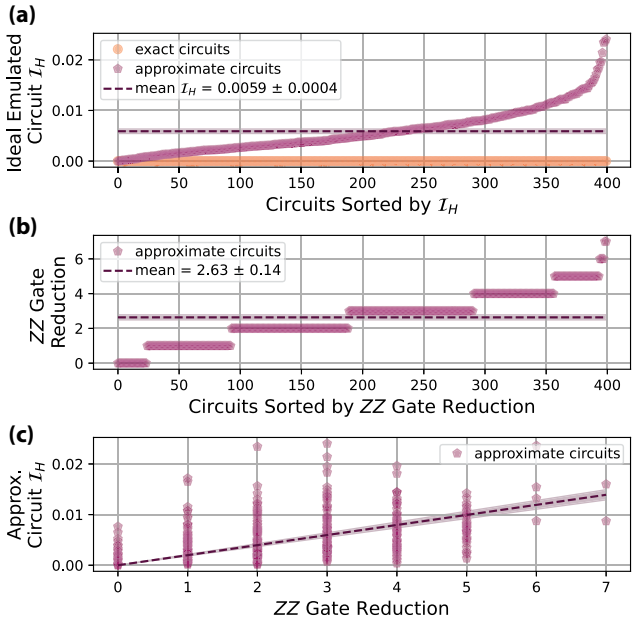


FIG. 11. Per-circuit analysis of the approximate circuits. (a) Plot of error-free simulation of the natural infidelity of the exact circuits (orange circles) and the approximate circuits (magenta pentagons) in which the circuits are sorted by increasing infidelity, along with a mean line (and 2σ standard error on the mean) indicating the average infidelity per approximate circuit. (b) The number of ZZ gates by which each approximate circuit is reduced are plotted with the circuits sorted by increasing ZZ gate reduction. A mean line and 2σ standard error on the mean are also plotted. (c) Plot of the natural infidelity of the approximate circuits versus the ZZ gate reduction. A trend line is presented with 2σ error.

In the case of the circuit approximation, we examine the natural circuit infidelities that arise from this particular approximation: removing all $ZZ(\theta)$ gates with $|\theta| < 0.10$ for a four-qubit QV circuit. Simulating the circuits, we find the average $\mathcal{I}_H$ is approximately 0.006 with a maximal $\mathcal{I}_H$ of approximately 0.025 as shown in Fig. 11(a). Examining the circuits more closely, we find that on average about 2.6 ZZ gates are removed per circuit, with a maximal removal of seven gates in total for two of the 400 circuits, in Fig. 11(b). As is to be expected, there is also rough trend for *decreasing* performance (i.e., increasing $\mathcal{I}_H$) as a function of increasing ZZ gate reduction in Fig. 11(c). However, it is important to note that the empirical results in Fig. 8(d) instead show a rough trend for a *gain* in performance (i.e., decreasing infidelity), suggesting that a reduction in ZZ gates via this approximation in general is able to outperform the loss of fidelity that is incurred by virtue of the approximation.

Next, in Table VI, we process the collected empirical data to compute a t-statistic and p-value for all comparisons presented in the main text across all analysis approaches: h_U, h_A, and $\mathcal{I}_H$. While there are certainly more complex (and more accurate) ways to determine the likelihood the analyzed empirical distributions are drawn from the same distribution, this is a rudimentary and common method. It should be noted that while the absolute values of some t-statistics (and p-values) *do not* fall above (below) typical thresholds of 2 (0.05), for each comparison we make, there is at least one analysis approach which yields a t-statistic (and p-value) which *does* lie within the generally accepted bounds for determining distinctness, suggesting all compilation optimizations provide a distinct change to the performance of the system.

Finally, in Table VII, we examine the certification of quantum volume based on the distributions measured in the main text. While there are some slight differences in how QV certification is defined across the literature [15,58,59], it in general occurs when the lower bound 2σ CI crosses the 2/3 threshold. In our case, we use one of the more conservative approaches to defining CIs, with the Wilson score interval [61] with continuity correction defined as

$$w^{-} = \max\left\{0, \frac{2N\hat{p} + z^2 - z\sqrt{z^2 - 1/N + 4N\hat{p}(1-\hat{p}) + (4\hat{p} - 2)} - 1}{2(N + z^2)}\right\}, \tag{B2}$$

TABLE VI. Further statistics to compare compilation optimizations. To determine the likelihood the measured distributions are drawn from the same distribution, the t-statistic and p-value for each compilation optimization and analysis method are presented. Highlighted in teal are values which pass the typical threshold to determine the statistical significance that two distributions are distinct from one another. For statistical significance, we require the $|t\text{-statistic}| > 2$ and the $p\text{-value} < 0.05$.

Comparison	h_U		h_A		$\mathcal{I}_H$	
	t-statistic	p-value	t-statistic	p-value	t-statistic	p-value
$ZZ(\pi/2)$versus $ZZ(\theta)$	−4.679	0.000	−5.810	0.000	6.740	0.000
$ZZ(\theta)$ versus +mSWAP	−2.284	0.023	−1.609	0.108	1.407	0.160
unranked versus ranked "bad pairs"	−2.934	0.003	−6.647	0.000	8.419	0.000
unranked versus ranked "bad qubit"	−1.574	0.116	−2.232	0.026	1.145	0.253
exact versus approx	−1.330	0.184	−1.354	0.176	2.003	0.046

TABLE VII. Quantum volume certification. Number of circuits required to certify quantum volume of 2^4 (or 2^5 in the case of Fig. 7) for all compilations presented in the main text. For distributions with means below 2/3, no quantum volume certification can be achieved. For distributions with means above 2/3 but confidence interval below 2/3, the extrapolated number of circuits required to certify is presented in italics. For distributions with means and confidence interval above 2/3, the measured number of circuits is presented.

Compilation	# of circuits	Compilation	# of circuits	Figure
$ZZ(\pi/2)$	...	$ZZ(\theta)$	*1536*	2
$ZZ(\theta)$	*1536*	+mSWAP	*592*	3
unranked	...	"bad pairs"	*1057*	5
unranked	...	"bad qubit"	...	7
exact	397	approximate	297	8

where N is the number of random circuits evaluated, $z = 2$ defines the 2σ error level, and $\hat{p}$ is the mean of the distribution.

We then evaluate the experimental results for the different compilation optimizations presented in the main text. For results in which the mean lies below the 2/3 threshold, no certification is possible. In several cases, the mean lies above the 2/3 threshold, but for the limited number of circuits tested (200 or 400), its empirically determined CI still lies below the threshold. As such, using the experimentally determined mean, we extrapolate the number of circuits required to have a CI that lies above the threshold based on Eq. (B2). In particular, when comparing the $ZZ(\theta)$ compilations with and without swap mirroring (Fig. 2), we find that the mean of the circuits with swap mirroring would only require 592 circuits, while the same circuits without would require 1536 circuits to be evaluated. Our final demonstration looking at exact and approximate circuits showed the CI did cross the 2/3 threshold for both measurements given the number of circuits evaluated. Here, the approximate circuits reach certification at 297 circuits, while the exact circuits require 397 circuits.

We note that a 2σ Wald interval [15] would certify these optimizations at fewer circuits. For example, the approximate circuits are certified at 234 circuits with the Wald interval rather than 297 circuits with the its corresponding Wilson score interval. Likewise, a bootstrapping method which also takes into account *the number of shots* rather than just the number of distinct circuits in each measurement, such as the one presented in [58], would likely certify at even fewer circuits.

[1] R. Wille, S. Hillmich, and L. Burgholzer, in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)* (IEEE, New York, NY, USA, 2020), p. 1.

[2] R. S. Smith, E. C. Peterson, M. G. Skilbeck, and E. J. Davis, An open-source, industrial-strength optimizing compiler for quantum programs, Quantum Sci. Technol. **5**, 044001 (2020).

[3] L. Moro, M. G. A. Paris, M. Restelli, and E. Prati, Quantum compiling by deep reinforcement learning, Commun. Phys. **4**, 178 (2021).

[4] C. Campbell, *et al.*, in *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*, Vol. 1 (IEEE, New York NY, USA, 2023), p. 1020.

[5] J. Liu, A. Gonzales, B. Huang, Z. H. Saleem, and P. Hovland, in *2025 IEEE International Symposium on High Performance Computer Architecture (HPCA)* (IEEE, New York NY, USA, 2025), p. 158.

[6] F. J. R. Ruiz, T. Laakkonen, J. Bausch, M. Balog, M. Barekatain, F. J. H. Heras, A. Novikov, N. Fitzpatrick, B. Romera-Paredes, J. van de Wetering, A. Fawzi, K. Meichanetzidis, and P. Kohli, Quantum circuit optimization with AlphaTensor, Nat. Mach. Intell. **7**, 374 (2025).

[7] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta, Quantum computing with Qiskit, arXiv:2405.08810.

[8] A. Javadi-Abhari, S. Patil, D. Kudrow, J. Heckey, A. Lvov, F. T. Chong, and M. Martonosi, ScaffCC: Scalable compilation and analysis of quantum programs, Parallel Comput. **45**, 2 (2015).

[9] Cirq Developers, Cirq, Zenodo, 2024.

[10] S. Sivarajah, S. Dilkes, A. Cowtan, W. Simmons, A. Edgington, and R. Duncan, t|ket⟩: A retargetable compiler for NISQ devices, Quantum Sci. Technol. **6**, 014003 (2020).

[11] The CUDA-Q development team, CUDA-Q, GitHub, 2025. https://github.com/NVIDIA/cuda-quantum

[12] A. Kissinger and J. van de Wetering, in *Proceedings 16th International Conference on Quantum Physics and Logic, Chapman University, Orange, CA, USA, 10–14 June 2019*, Electronic Proceedings in Theoretical Computer Science, edited by B. Coecke and M. Leifer (Open Publishing Association, Waterloo, NSW, Australia, 2020), Vol. 318, p. 229.

[13] E. Younis, C. Iancu, W. Lavrijsen, M. Davis, and E. Smith, Berkeley Quantum Synthesis Toolkit (BQSKit) v1 (2021).

[14] M. Maronese, L. Moro, L. Rocutto, and E. Prati, in *Quantum Computing Environments* (Springer International Publishing, Cham, Switzerland, 2022), p. 39.

[15] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta, Validating quantum computers using randomized model circuits, Phys. Rev. A **100**, 032328 (2019).

[16] Y. Nam, *et al.*, Ground-state energy estimation of the water molecule on a trapped-ion quantum computer, npj Quantum Inf. **6**, 33 (2020).
[17] A. Hashim, R. Rines, V. Omole, R. K. Naik, J. M. Kreikebaum, D. I. Santiago, F. T. Chong, I. Siddiqi, and P. Gokhale, Optimized SWAP networks with equivalent circuit averaging for QAOA, Phys. Rev. Res. **4**, 033028 (2022).
[18] M. Bowman, P. Gokhale, J. Larson, J. Liu, and M. Suchara, Hardware-conscious optimization of the quantum Toffoli gate, ACM Tran. Quantum Comput. **4**, 26 (2023).
[19] S. A. Moses, *et al.*, A race-track trapped-ion quantum processor, Phys. Rev. X **13**, 041052 (2023).
[20] A. Botea, A. Kishimoto, and R. Marinescu, On the complexity of quantum circuit compilation, Proc. Int. Symp. Comb. Search **9**, 138 (2021).
[21] L. Egan, D. M. Debroy, C. Noel, A. Risinger, D. Zhu, D. Biswas, M. Newman, M. Li, K. R. Brown, M. Cetina, and C. Monroe, Fault-tolerant control of an error-corrected qubit, Nature **598**, 281 (2021).
[22] J. M. Pino, J. M. Dreiling, C. Figgatt, J. P. Gaebler, S. A. Moses, M. S. Allman, C. H. Baldwin, M. Foss-Feig, D. Hayes, K. Mayer, C. Ryan-Anderson, and B. Neyenhuis, Demonstration of the trapped-ion quantum ccd computer architecture, Nature **592**, 209 (2021).
[23] P. Wang, C.-Y. Luan, M. Qiao, M. Um, J. Zhang, Y. Wang, X. Yuan, M. Gu, J. Zhang, and K. Kim, Single ion qubit with estimated coherence time exceeding one hour, Nat. Commun. **12**, 233 (2021).
[24] C. Monroe and J. Kim, Scaling the ion trap quantum processor, Science **339**, 1164 (2013).
[25] A. Erhard, H. Poulsen Nautrup, M. Meth, L. Postler, R. Stricker, M. Stadler, V. Negnevitsky, M. Ringbauer, P. Schindler, H. J. Briegel, R. Blatt, N. Friis, and T. Monz, Entangling logical qubits with lattice surgery, Nature **589**, 220 (2021).
[26] C. Ryan-Anderson, *et al.*, High-fidelity teleportation of a logical qubit using transversal gates and lattice surgery, Science **385**, 1327 (2024).
[27] Y. Hong, E. Durso-Sabina, D. Hayes, and A. Lucas, Entangling four logical qubits beyond break-even in a nonlocal code, Phys. Rev. Lett. **133**, 180601 (2024).
[28] S. Debnath, N. M. Linke, C. Figgatt, K. A. Landsman, K. Wright, and C. Monroe, Demonstration of a small programmable quantum computer with atomic qubits, Nature **536**, 63 (2016).
[29] S. M. Clark, D. Lobser, M. C. Revelle, C. G. Yale, D. Bossert, A. D. Burch, M. N. Chow, C. W. Hogle, M. Ivory, J. Pehr, B. Salzbrenner, D. Stick, W. Sweatt, J. M. Wilson, E. Winrow, and P. Maunz, Engineering the Quantum Scientific Computing Open User Testbed, IEEE Trans. Quantum Eng. **2**, 3102832 (2021).
[30] K. A. Landsman, Y. Wu, P. H. Leung, D. Zhu, N. M. Linke, K. R. Brown, L. Duan, and C. Monroe, Two-qubit entangling gates within arbitrarily long chains of trapped ions, Phys. Rev. A **100**, 022332 (2019).
[31] D. Kielpinski, C. Monroe, and D. J. Wineland, Architecture for a large-scale ion-trap quantum computer, Nature **417**, 709 (2002).
[32] P. H. Leung and K. R. Brown, Entangling an arbitrary pair of qubits in a long ion crystal, Phys. Rev. A **98**, 032318 (2018).
[33] M. Cetina, L. Egan, C. Noel, M. Goldman, D. Biswas, A. Risinger, D. Zhu, and C. Monroe, Control of transverse motion for quantum gates on individually addressed atomic qubits, PRX Quantum **3**, 010334 (2022).
[34] P. Murali, D. M. Debroy, K. R. Brown, and M. Martonosi, Toward systematic architectural design of near-term trapped ion quantum computers, Commun. ACM **65**, 101 (2022).
[35] R. Shaffer, H. Ren, E. Dyrenkova, C. G. Yale, D. S. Lobser, A. D. Burch, M. N. H. Chow, M. C. Revelle, S. M. Clark, and H. Häffner, Sample-efficient verification of continuously-parameterized quantum gates for small quantum processors, Quantum **7**, 997 (2023).
[36] C. G. Yale, A. D. Burch, M. N. H. Chow, B. P. Ruzic, D. S. Lobser, B. K. McFarland, M. C. Revelle, and S. M. Clark, Realization and calibration of continuously parameterized two-qubit gates on a trapped-ion quantum processor, arXiv:2504.06259.
[37] B. P. Lanyon, C. Hempel, D. Nigg, M. Müller, R. Gerritsma, F. Zähringer, P. Schindler, J. T. Barreiro, M. Rambach, G. Kirchmair, M. Hennrich, P. Zoller, R. Blatt, and C. F. Roos, Universal digital quantum simulation with trapped ions, Science **334**, 57 (2011).
[38] K. Blekos, D. Brand, A. Ceschini, C.-H. Chou, R.-H. Li, K. Pandya, and A. Summer, A review on quantum approximate optimization algorithm and its variants, Phys. Rep. **1068**, 1 (2024).
[39] T. G. Draper, Addition on a quantum computer, arXiv: quant-ph/0008033v1.
[40] B. P. Ruzic, M. N. Chow, A. D. Burch, D. S. Lobser, M. C. Revelle, J. M. Wilson, C. G. Yale, and S. M. Clark, Leveraging motional-mode balancing and simply parametrized waveforms to perform frequency-robust entangling gates, Phys. Rev. Appl. **22**, 014007 (2024).
[41] D. R. Pérez, P. Varosy, Z. Li, T. Roy, E. Kapit, and D. Schuster, Error-divisible two-qubit gates, Phys. Rev. Appl. **19**, 024043 (2023).
[42] Infleqtion, Superstaq documentation, 2024. https://superstaq.readthedocs.io
[43] M. N. H. Chow, C. G. Yale, A. D. Burch, M. Ivory, D. S. Lobser, M. C. Revelle, and S. M. Clark, First-order crosstalk mitigation in parallel quantum gates driven with multi-photon transitions, Appl. Phys. Lett. **124**, 044002 (2024).
[44] A. Sørensen and K. Mølmer, Entanglement and quantum computation with ions in thermal motion, Phys. Rev. A **62**, 022311 (2000).
[45] P. J. Lee, K. A. Brickman, L. Deslauriers, P. C. Haljan, L. M. Duan, and C. Monroe, Phase control of trapped ion quantum gates, J. Opt. B: Quantum Semiclassical Opt. **7**, S371 (2005).
[46] T. R. Tan, J. P. Gaebler, Y. Lin, Y. Wan, R. Bowler, D. Leibfried, and D. J. Wineland, Multi-element logic gates for trapped-ion qubits, Nature **528**, 380 (2015).
[47] C. H. Baldwin, B. J. Bjork, J. P. Gaebler, D. Hayes, and D. Stack, Subspace benchmarking high-fidelity entangling

operations with trapped ions, Phys. Rev. Res. **2**, 013317 (2020).
[48] D. F. V. James, Quantum dynamics of cold trapped ions with application to quantum computation, Appl. Phys. B **66**, 181 (1998).
[49] R. R. Tucci, An introduction to Cartan's KAK decomposition for QC programmers, arXiv:quant-ph/0507171.
[50] S. S. Tannu and M. K. Qureshi, in *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, ASPLOS '19 (Association for Computing Machinery, New York, 2019), p. 987.
[51] A. Ash-Saki, M. Alam, and S. Ghosh, in *DAC '19: Proceedings of the 56th Annual Design Automation Conference* (Association for Computer Machinery, New York, NY, USA, 2019).
[52] P. D. Nation and M. Treinish, Suppressing quantum circuit errors due to system variability, PRX Quantum **4**, 010327 (2023).
[53] W. Finigan, M. Cubeddu, T. Lively, J. Flick, and P. Narang, Qubit allocation for noisy intermediate-scale quantum computers, arXiv:1810.08291.
[54] S. Park, D. Kim, M. Kweon, J.-Y. Sim, and S. Kang, in *DAC '22: Proceedings of the 59th ACM/IEEE Design Automation Conference* (Association for Computing Machinery, San Francisco, CA, USA, 2022), p. 13.
[55] T. LeCompte, F. Qi, X. Yuan, N.-F. Tzeng, M. H. Najafi, and L. Peng, Machine-learning-based qubit allocation for error reduction in quantum circuits, IEEE Trans. Quantum Eng. **4**, 1 (2023).
[56] J. Edmonds, Maximum matching and a polyhedron with 0,1-vertices, J. Res. Natl. Bureau Stand. Section B: Math. Math. Phys. **69**, 125 (1965).
[57] V. Kaibel and M. Walter, A note on matchings constructed during Edmonds' weighted perfect matching algorithm, arXiv:1703.09505.
[58] C. H. Baldwin, K. Mayer, N. C. Brown, C. Ryan-Anderson, and D. Hayes, Re-examining the quantum volume test: Ideal distributions, compiler optimizations, confidence intervals, and scalable resource estimations, Quantum **6**, 707 (2022).
[59] E. Pelofske, A. Bärtschi, and S. Eidenbenz, Quantum volume in practice: What users can expect from NISQ devices, IEEE Trans. Quantum Eng. **3**, 3102119 (2022).
[60] S. Aaronson and L. Chen, in *CCC '17: Proceedings of the 32nd Computational Complexity Conference* (Schloss Dagstuhl—Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2017).
[61] E. B. Wilson, Probable inference, the law of succession, and statistical inference, J. Am. Stat. Assoc. **22**, 209 (1927).
[62] E. Hellinger, Neue Begründung der Theorie quadratischer Formen von unendlich vielen Veränderlichen, J. Reine Angew. Math. **136**, 210 (1909).
[63] S. Dasgupta, Stability of quantum computers, Ph.D. thesis, University of Tennessee, Knoxville, 2024, https://trace.tennessee.edu/utk_graddiss/10111.
[64] B. Foxen, *et al.*, Demonstrating a continuous set of two-qubit gates for near-term quantum algorithms, Phys. Rev. Lett. **125**, 120504 (2020).
[65] A. C. Lee, J. Smith, P. Richerme, B. Neyenhuis, P. W. Hess, J. Zhang, and C. Monroe, Engineering large Stark shifts for control of individual clock state qubits, Phys. Rev. A **94**, 042308 (2016).
[66] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, J. Opt. B: Quantum Semiclassical Opt. **7**, S347 (2005).
[67] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, Phys. Rev. A **80**, 012304 (2009).
[68] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, Phys. Rev. Lett. **106**, 180504 (2011).
[69] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, Efficient measurement of quantum gate error by interleaved randomized benchmarking, Phys. Rev. Lett. **109**, 080505 (2012).
[70] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt, Characterizing large-scale quantum computers via cycle benchmarking, Nat. Commun. **10**, 5347 (2019).
[71] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, Quantum **5**, 557 (2021).
[72] A. H. Rubin, B. Marinelli, V. A. Norman, Z. Rizvi, A. D. Burch, R. K. Naik, J. M. Kreikebaum, M. N. H. Chow, D. S. Lobser, M. C. Revelle, C. G. Yale, M. Ivory, D. I. Santiago, C. Spitzer, M. Krstic-Marinkovic, S. M. Clark, I. Siddiqi, and M. Radulaski, Quantum digital simulation of cavity quantum electrodynamics: Insights from superconducting and trapped ion quantum testbeds, arXiv:2404.03861.
[73] R. Rines and C. Yale, Data for "Noise-aware circuit compilations for a continuously parameterized two-qubit gateset", Zenodo (2025), https://doi.org/10.5281/zenodo.16283542.
[74] M. N. H. Chow, Scalable methods for performant control of hyperfine qubits in atoms and ions, Ph.D. thesis, University of New Mexico, 2024, https://digitalrepository.unm.edu/phyc_etds/335.