

A Race-Track Trapped-Ion Quantum Processor

S. A. Moses^{1,*†}, C. H. Baldwin^{1,*‡}, M. S. Allman,¹ R. Ancona,¹ L. Ascarrunz,¹ C. Barnes,¹ J. Bartolotta¹, B. Bjork,¹ P. Blanchard,¹ M. Bohn,¹ J. G. Bohnet,¹ N. C. Brown,¹ N. Q. Burdick,² W. C. Burton¹, S. L. Campbell,¹ J. P. Campora III,¹ C. Carron,³ J. Chambers,¹ J. W. Chan,¹ Y. H. Chen,¹ A. Chernoguzov,¹ E. Chertkov¹, J. Colina,¹ J. P. Curtis,¹ R. Daniel,¹ M. DeCross¹, D. Deen³, C. Delaney,¹ J. M. Dreiling,¹ C. T. Ertsgaard,³ J. Esposito,¹ B. Estey,¹ M. Fabrikant,¹ C. Figgatt¹, C. Foltz,¹ M. Foss-Feig,¹ D. Francois,¹ J. P. Gaebler,¹ T. M. Gatterman,¹ C. N. Gilbreth,¹ J. Giles,¹ E. Glynn,¹ A. Hall,¹ A. M. Hankin,¹ A. Hansen,¹ D. Hayes,¹ B. Higashi,³ I. M. Hoffman¹, B. Horning,³ J. J. Hout,¹ R. Jacobs,¹ J. Johansen,¹ L. Jones,¹ J. Karcz,⁴ T. Klein,³ P. Lauria,¹ P. Lee,¹ D. Liefer,¹ S. T. Lu,⁴ D. Lucchetti,¹ C. Lytle,¹ A. Malm,¹ M. Matheny,¹ B. Mathewson,¹ K. Mayer,¹ D. B. Miller,¹ M. Mills,¹ B. Neyenhuis,¹ L. Nugent,¹ S. Olson,³ J. Parks,¹ G. N. Price,¹ Z. Price,¹ M. Pugh,¹ A. Ransford,¹ A. P. Reed,¹ C. Roman,¹ M. Rowe,¹ C. Ryan-Anderson,¹ S. Sanders,¹ J. Sedlacek,² P. Shevchuk,¹ P. Siegfried,¹ T. Skripka,¹ B. Spaun,¹ R. T. Sprenkle,¹ R. P. Stutz,¹ M. Swallows,¹ R. I. Tobey,¹ A. Tran,¹ T. Tran,¹ E. Vogt,⁴ C. Volin,¹ J. Walker,¹ A. M. Zolot,¹ and J. M. Pino¹

¹Quantinuum, 303 South Technology Court, Broomfield, Colorado 80021, USA

²Quantinuum, 1985 Douglas Drive North, Golden Valley, Minnesota 55422, USA

³Quantinuum, 12001 State Highway 55, Plymouth, Minnesota 55441, USA

⁴Honeywell Aerospace, 12001 State Highway 55, Plymouth, Minnesota 55441, USA

 (Received 17 May 2023; revised 15 August 2023; accepted 28 September 2023; published 18 December 2023)

We describe and benchmark a new quantum charge-coupled device (QCCD) trapped-ion quantum computer based on a linear trap with periodic boundary conditions, which resembles a race track. The new system successfully incorporates several technologies crucial to future scalability—including electrode broadcasting, multilayer rf routing, and magneto-optical trap (MOT) loading—while maintaining, and in some cases exceeding, the gate fidelities of previous QCCD systems. The system is initially operated with 32 qubits, but future upgrades will allow for more. We benchmark the performance of primitive operations, including an average state preparation and measurement error of $1.6(1) \times 10^{-3}$, an average single-qubit gate infidelity of $2.5(3) \times 10^{-5}$, and an average two-qubit gate infidelity of $1.84(5) \times 10^{-3}$. The system-level performance of the quantum processor is assessed with mirror benchmarking, linear cross-entropy benchmarking, a quantum volume measurement of $QV = 2^{16}$, and the creation of 32-qubit entanglement in a GHZ state. We also tested application benchmarks, including Hamiltonian simulation, QAOA, error correction on a repetition code, and dynamics simulations using qubit reuse. We also discuss future upgrades to the new system aimed at adding more qubits and capabilities.

DOI: [10.1103/PhysRevX.13.041052](https://doi.org/10.1103/PhysRevX.13.041052)

Subject Areas: Atomic and Molecular Physics,
Quantum Physics, Quantum Information

I. INTRODUCTION

Several technology platforms are viable candidates for large-scale quantum computation, including trapped ions [1], neutral atoms [2], and superconducting circuits [3]. However, existing demonstrations face scaling challenges to achieve

the qubit numbers and fidelities necessary for fault-tolerant quantum computing. In addition, all platforms need refinement in reliability, power consumption, form factor, and cost. This concept, known as Rent's rule, has been discussed rigorously in terms of classical computing technologies and recently generalized to include quantum processors [4].

In this work, we characterize a trapped-ion quantum computer with a new trap design based on the QCCD architecture. The new machine, Quantinuum System Model H2, significantly increases the qubit number and decreases the physical resources per qubit, all while matching—and in some instances surpassing—the high circuit fidelity of our previous generation system [5].

The QCCD architecture was proposed as a scalable method for trapped-ion quantum computation [6,7].

*These authors contributed equally to this work.

†steven.moses@quantinuum.com

‡charles.baldwin@quantinuum.com

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

Trapped-ion systems with a single trapping zone are limited in qubit number due to challenges in individually addressing single qubits within a large ion crystal, as well as motional mode crowding, which complicates achieving high-fidelity operations in a large crystal [8,9]. QCCD trapped-ion systems instead have multiple trapping zones allowing operations to always be performed with a small number of ions, thereby facilitating low-crosstalk addressing and maintaining high fidelity [10]. Two-qubit gates between arbitrary pairs of qubits are enabled by ion transport during a quantum circuit, which brings pairs to be gated into the same trapping zone. Such dynamic rearrangement enables the execution of circuits with arbitrary connectivity without the overhead of logical SWAP gates typically incurred for platforms with fixed and limited connectivity [11]. This transport requires traps with a large number of programmable electrodes, which can be achieved using microfabricated surface traps [12–14].

Our first generation hardware, H0 and H1 (based on the same linear trap design) [5,15], demonstrated many key components of the QCCD architecture and achieved high-fidelity gates with arbitrary two-qubit couplings. Since the initial operation of the linear trap, the qubit number N increased fivefold, from 4 in its initial mode of operation [5] to 20 in its latest [15], while two-qubit gate errors decreased by roughly a factor of 5. By increasing both the transport speeds and the number of gate zones, the average time required to execute a layer of N single-qubit (1Q) gates and $N/2$ two-qubit (2Q) gates on a random pairing of all N qubits was kept roughly constant as N increased.

This progress notwithstanding, linear geometries pose severe scaling challenges. The time to rearrange ions for arbitrary circuit connectivity scales poorly for the linear trap design as the number of qubits increases (proportional to N [16]). The future of QCCD systems is likely in 2D traps that offer better scaling of rearrangement times (proportional to \sqrt{N} [17]) and are also well suited to many error-correcting codes [18,19]. However, 2D traps present new engineering challenges that are still under development, such as junction transport [20,21] and signal routing under the trap’s top metal layer. Many other aspects of QCCD scaling still in development include coupling multiple surface trap die [22,23], control of a sufficient number of electrodes, laser-light generation and delivery [24,25], and detection [26]. Not all of these challenges will be met simultaneously, but rather, advances will be inserted as they are available.

This paper marks the first major trap design advancement in the H-series QCCD quantum computers. Specifically, the new trap (shown in Fig. 1) introduces the following: (1) rf tunnels, so that rf voltage electrodes do not need to be connected on the top surface that defines the trapping potential (Sec. II A), (2) voltage broadcasting to multiple control electrodes, thereby reducing the number of independent voltage sources needed to control the device (Sec. II A), and (3) MOT loading of the trap to increase the

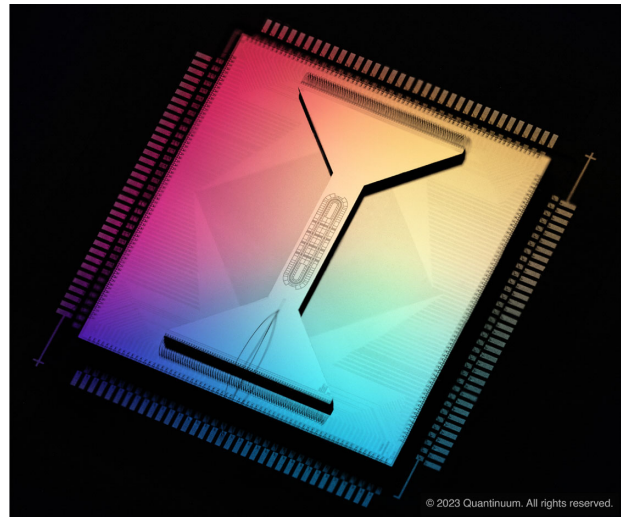


FIG. 1. Picture of the H2 surface ion trap microchip. The image has been modified to enhance visibility of the trap features. The trap sits in the isthmus in the center of the trap die. The long axis of the trap is 6.58 mm (from the edge of the dc electrodes on either side), and the isthmus width is 2.02 mm.

ion loading rate, thus decreasing the initialization time [27] (Sec. II B). In addition to the upgraded system design, we also report on upgraded operations, including higher performance and more efficient gating primitives. We present detailed benchmarking of the system performance with component benchmarking in Sec. III, system-level benchmarking in Sec. IV, and algorithmic benchmarking in Sec. V. Similar to the H1 series, the configuration described in this report is only the first of the H2 series, and we expect to make significant qubit count and gate zone operation upgrades in the near future.

II. OVERVIEW OF THE HARDWARE

A. Trap design

As shown in Fig. 2, H2 has a race track geometry similar to traps fabricated by other groups [28–30]. Two concentric rf electrodes circumscribe the center region and are driven at about 200 V and 42 MHz, creating an rf-null 70 μm from the surface where ions are trapped. The rf tunnels [29] are required for the concentric rf electrodes and allow for dc electrodes to tile the full trap perimeter shown in Fig. 2(c) and 2(d). The trap has two rows of gate zones colored in blue in Fig. 2, four on the top (UG01–UG04) and four on the bottom (DG01–DG04). In this work, we use both rows for ion rearrangement (physical swaps), but only the DG zones are used for quantum operations (gating, state preparation, and measurement). We plan to extend quantum operations to both rows in future work.

The “conveyor belt” region of the trap is colored green in Fig. 2(d). In this region, voltage “broadcasting” is used to minimize dc control signals by tying multiple dc electrodes within the trap die to the same external signal. As shown in

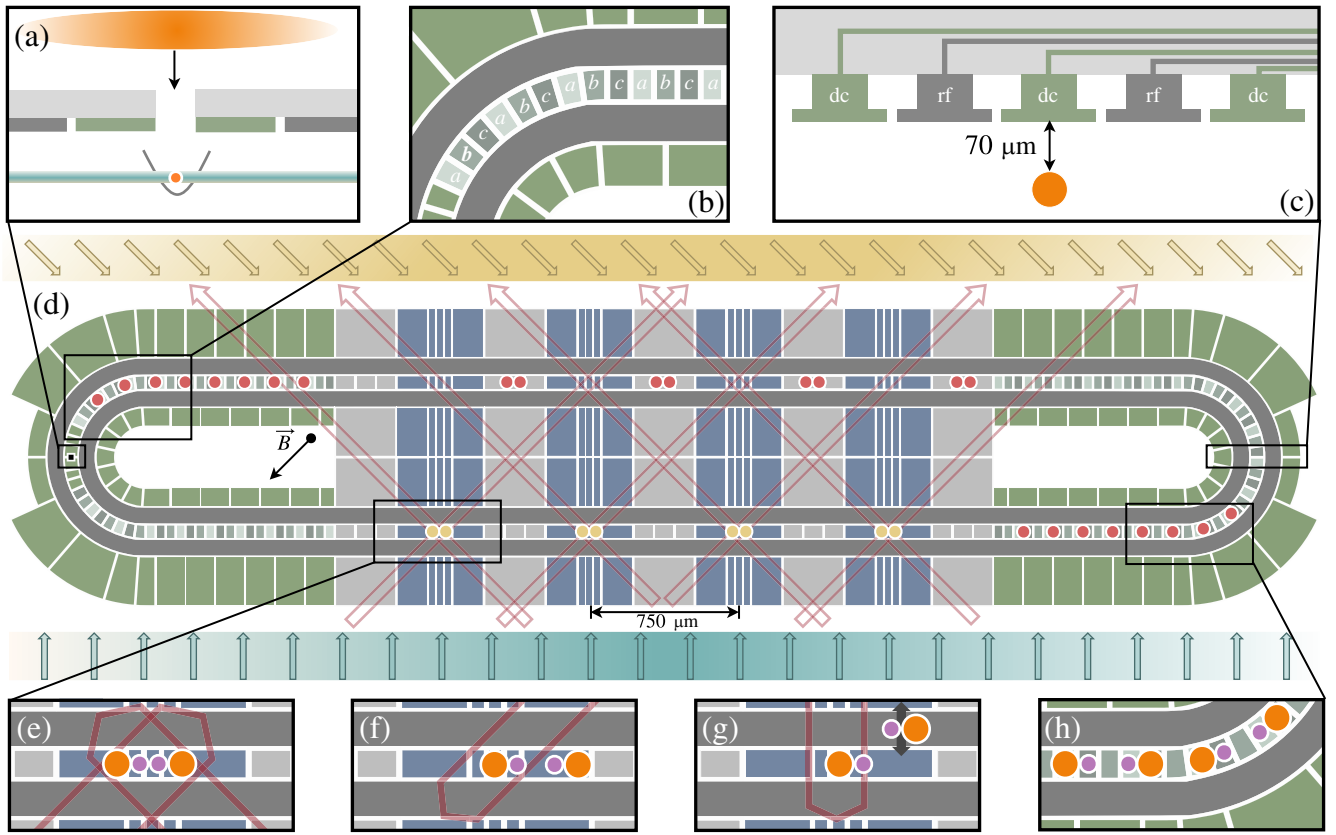


FIG. 2. Overview of the H2 trap including upgrades in trap design and gating operations. (a) 2D MOT producing a collimated beam of atoms, allowing for higher neutral atom density and faster loading than an effusive oven. (b) The abc tiling of electrodes for conveyor belt transport. (c) rf tunnels to implement inner and outer rf electrodes. Ions are trapped $70\ \mu\text{m}$ away from the trap surface. (d) Colored top metal layer of the H2 trap. Green curved zones are conveyor belt regions for ion storage. The bottom blue zones are DG01–DG04 (from left to right), which are used for quantum operations. The top blue zones are UG01–UG04 gate zones (from right to left), which are used for sorting but not quantum operations. Darker gray loops are rf electrodes. Yellow circles represent qubits that are gated while red circles represent qubits sitting in storage during gates (note that $^{138}\text{Ba}^+$ ions are omitted for simplicity). Yellow arrows indicate the Doppler sheet beam direction while blue arrows indicate the Doppler repump sheet beam direction. (e) Ion configuration and beam direction for 2Q gates. Large orange circles represent $^{171}\text{Yb}^+$ while smaller purple circles represent $^{138}\text{Ba}^+$. (f) Ion configuration and beam directions for 1Q gates on the left $^{171}\text{Yb}^+$. (g) Ion configuration and beam directions for state preparation and measurement (SPAM) operations on the left $^{171}\text{Yb}^+$ with micromotion hiding on the right $^{171}\text{Yb}^+$ [31]. (h) Storage ion configuration in the conveyor belt region.

Fig. 2(b), each conveyor belt region contains equally spaced and sized electrodes tied together in a repeating fashion ($\{a, b, c, a, b, c, \dots\}$). This requires only three total voltage signals for each side, which can support 20 wells (one for every three electrodes). Additional electrodes, called shim electrodes, are located outside of the rf electrodes and used to compensate micromotion and rotate the trap principal axes. The load hole, visible in the middle of the left-side conveyor belt region in Fig. 2(d), is surrounded by six electrodes with independent signals.

The gate zone electrode configuration is similar to that in Ref. [5], and the spacing between gate zones remains the same ($750\ \mu\text{m}$). An additional improvement to the signal count was realized by reducing the number of electrodes in the auxiliary regions around the gating zones [light gray in Fig. 2(d)]. As expected, the linear transport through the auxiliary zones is not degraded compared with H1.

In total, the trap has 376 gold-coated electrodes connected to 268 independent voltage sources and 1 rf drive. Similar to H1, H2 uses a 280-pin ceramic pin grid array to connect the trap electrodes to the dc control signals. This is a reduction in the number of electrical feedthroughs per qubit in the system, which is an important metric as the number of qubits grows. Eventually, electrode control from within the vacuum chamber will be necessary, as the number of physical connections into the vacuum chamber will become unmanageable [17].

B. Ion loading and state preparation

H2 uses a 2D MOT as a source for neutral atoms instead of an effusive atomic oven [32]. The MOT is connected to the main vacuum chamber via a differential pumping tube. The MOT cools both ^{171}Yb and ^{138}Ba neutral atoms, which

are directed toward the backside load hole in the main vacuum chamber. A fraction of the neutral atoms that pass through the load hole are ionized by the photo-ionization beams on the front side and subsequently cooled after loading into the trap [see Fig. 2(a)]. In the best case, we load one $^{171}\text{Yb}^+$ in about 1.2 ms and one $^{138}\text{Ba}^+$ in about 40 ms; however, algorithmic latency and validation procedures limit the time to load the full trap (32 $^{171}\text{Yb}^+ - ^{138}\text{Ba}^+$ [YB] pairs in a deterministic orientation) to about 3–4 minutes. Under normal operating conditions, we observe no impact on the behavior of the quantum processor with the MOT beam on. Once the trap is fully loaded, we detect individual loss events and replace the affected ion pairs, requiring only 10–15 seconds per lost pair.

The qubit subspace occupies the hyperfine approximate clock states of $^{171}\text{Yb}^+$ in the $^2S_{1/2}$ state, $|0\rangle \equiv |F=0, m_f=0\rangle$ and $|1\rangle \equiv |F=1, m_f=0\rangle$. The quantization axis is set by an externally applied magnetic field in the plane of the trap at 45° with respect to the long axis. After loading, qubits are prepared in the $|0\rangle$ state via optical pumping, similar to previous work [5,33]. State preparation is currently only possible in the DG zones, so we prepare eight qubits at a time and prepare all 32 qubits in four rounds.

C. Quantum gates

Quantum gates are implemented in the DG gate zones by stimulated Raman processes described in Ref. [5] with a laser geometry shown in Figs. 2(d)–2(f). 1Q gates use copropagating beams [Fig. 2(f)], and 2Q gates use pairs of beams with $\Delta\vec{k}$ coupling to the axial modes of motion [Fig. 2(e)]. The 2Q gates are performed on the second-lowest-frequency axial mode of a Yb-Ba-Ba-Yb crystal and are implemented with a phase-sensitive Mølmer-Sørensen (MS) gate sandwiched between 1Q wrapper pulses (using the same laser beams as the MS interaction) to generate the parametrized gate $U_{ZZ}(\theta) = \exp(-i\theta ZZ/2)$ [34–39]. The value of θ is controlled by varying the detuning, duration, and Rabi rate of the MS interaction. Specifically, down to an angle of 0.075π , we increase the detuning and decrease the MS pulse duration while holding the laser powers constant. For smaller angles, we keep the detuning and gate duration fixed at their values for the 0.075π rotation angle and attenuate one of the laser beam intensities while holding the other laser beam intensity constant. Modeling supports an average gate infidelity that decreases roughly linearly with θ down to a finite offset of approximately 5×10^{-4} as $\theta \rightarrow 0$ as shown in Fig. 3. The finite offset at zero angle exists because the wrapper pulses still occur with a delay between them, and some fraction of the 2Q laser light remains on at zero angle, leading to residual errors predominantly from laser phase noise and spontaneous emission.

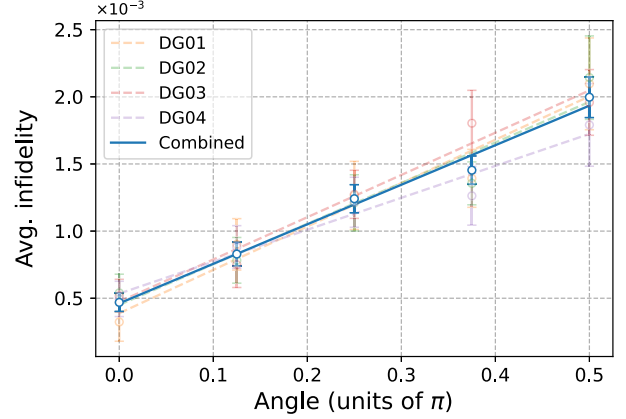


FIG. 3. Average infidelity as a function of angle for the parametrized 2Q gate $U_{ZZ}(\theta)$. Each data point is obtained by fitting the decay curves shown in Fig. 21 to an exponential decay function. The infidelity at $\theta = 0$ is due to both the wrapper pulses and memory error incurred during the cooling pulses, which are still applied in the absence of an MS gate. The linear best fit to the zone-averaged data is given by $\epsilon(\theta) = (2.9(2)\theta/\pi + 0.46(6)) \times 10^{-3}$.

The 2Q beams have the strictest requirements and consume a large portion of the total laser power budget (currently 5–10 mW are required per zone per arm for 2Q light), with the current configuration using four 2Q laser beam pairs to operate four gate zones. We note that adding one more pair of beams would enable the operation of four more gate zones on the other side of the trap, an upgrade we plan to explore in future work. To address both sides of the trap, the focus of the gate beams would shift by about 0.5 mm, which is less than the Rayleigh range (current spot size is about 15 μm), in order to achieve the same spot size for both sides of the trap. We expect that we would be able to address all eight gate zones with about 35% more power than the current setup.

D. Measurement

Measurement operations are performed in the DG zones with resonant beams traveling perpendicular to the long axis of the trap using state-dependent resonance fluorescence shown in Fig. 2(g). A photomultiplier tube array allows independent detection in all eight gate zones simultaneously, though we only implement measurement operations in the DG zones.

Similar to previous work [5,15], qubit measurement and reset may be performed in the middle of a quantum circuit while quantum information is preserved on other qubits. Midcircuit measurement and reset (MCMR) causes a small crosstalk error that acts on neighboring qubits due to stray light from the measurement and reset beams (see Sec. III and Table II). For unmeasured ions in the gate zones, this error is mitigated by the micromotion hiding technique described in Ref. [31] and depicted in Fig. 2(g). Ions in the

TABLE I. Example circuit resource estimates. Circuit time budgets are estimated from compiler information and broken into quantum operations (e.g., 1Q and 2Q gates and SPAM), transport (e.g., physical swaps and shifts), and cooling (sideband cooling before 2Q gates). Shot time does not include overheads such as postchecks of ion crystal configurations, but it does include the overhead for state preparation and initial cooling of about 17 ms/shot for a 32-qubit circuit. Number of 2Q gate rounds indicates the number of parallel 2Q gate operations with up to four 2Q gates per round.

Circuit name	Number of qubits	Circuits budget (%)		Shot time (s)	Number of 2Q gates	Number of 2Q gate rounds	Number of measurements
		quantum operations/	transport/cooling)				
2Q RB, $\ell = 128^a$	16	2/30/68		1.74	813	287	16
Transport 1Q RB, $\ell = 64$	32	1/73/26		4.01	0	0	32
MB, $\ell = 10$	32	1/67/32		1.05	320	80	32
QV	16	1/52/47		1.23	310	128	16
RCS	32	1/67/32		0.72	172	56	32
GHZ	32	2/64/34		0.18	31	14	32
TFIM, $Jt = 7$	32	1/48/51		0.59	288	72	32
QAOA, $p = 2$	32	1/69/30		0.97	96	66	32
QAOA w/qubit reuse, $p = 1$	32	1/74/25		2.83	195	166	130
Phase flip rep. code, $SE = 9$	32	2/57/41		5.4	540	540	301
HoloQUADS $t = 24$	32	1/67/32		20.6	2130	1629	129

^aThe 2Q RB is performed on eight qubits, and an additional eight qubits are required for the leakage gadget.

conveyor belt regions suffer from a similar level of cross-talk errors as ions in the gate zones, although we do not attempt to apply the micromotion hiding technique to them.

E. Ion transport

Arbitrary qubit connectivity is achieved via physical ion transport. During 32-qubit operation, the ions can be grouped into four “batches” of eight, with the four batches occupying the DG zones, UG zones, and each of the two storage regions, as shown in Fig. 2(d). The fundamental transport operations are similar to those in Ref. [5] and include split, combine, linear shifts, and physical swaps. A special type of linear shift for H2 is the batch shift, which shuttles batches of ions collectively to different regions of the trap. This operation is comparatively slow and dominates the circuit time. The fraction of total circuit time taken up by transport varies from circuit to circuit but is 60% on average (see Table I).

During ion transport, we cool all $^{138}\text{Ba}^+$ ions with Doppler cooling “sheet beams,” illustrated on the top and bottom of Fig. 2(d), which resemble sheets of laser light that cover the entire trap. These sheet beams have about a 25% variation in intensity between the center of the trap and the edges, which does not present any performance limitations.

A compiler generates a schedule of quantum gates and transport operations with the goal of minimizing the total transport time required to execute the circuit. The circuit is first decomposed into layers, which are built iteratively by looking ahead through the circuit and grouping together (into one layer) the largest possible set of 2Q gates subject to two constraints: (1) No ions participate in more than one gate in each layer, and (2) the time ordering of 2Q gates that

share one or more qubits (or any time-ordering enforced by an explicitly requested barrier) is respected. The circuit is then converted into a layered directed acyclic graph, and a modified Sugiyama algorithm [40] is applied to iteratively sort the qubits in each layer of the graph in an effort to minimize the overall transport time required to execute all layers. The periodic boundary conditions of the device are explicitly taken into account, and the resulting transport operations are computed using a parallel bubble sort routine that allows qubits to move in both directions around the device.

After compilation, the layers of the circuit are then executed sequentially, with transport primitives used to arrange the ions so that qubits scheduled to be gated in a given layer are positioned next to each other. Once arranged, we perform gates on each batch of ions, starting with the qubits already in the DG zones. Ions are transported to the center of the gate zones for quantum operations. The 1Q gates are performed after moving a single YB pair into the center of the gate zone with shift operations [Fig. 2(f)], while 2Q gates are performed with two YB pairs combined into a single four-ion crystal Yb-Ba-Ba-Yb [Fig. 2(e)]. Before 2Q gating operations, we apply resolved sideband cooling to the $^{138}\text{Ba}^+$ ions in the DG zones [5,41,42]. Ions in the UG zones are transported to the nearby auxiliary zones so that they are not addressed by the gating laser beams [see red circles in Fig. 2(d)]. After the gates are applied, we perform batch shifts to move new batches of ions into the DG zones and repeat the gating procedure until the full layer is completed.

A spatial phase tracking routine accounts for inhomogeneities in the magnetic field and spatially dependent ac Zeeman shifts from the rf current [43], which lead to spatially varying qubit frequencies. The routine calculates

extraneous phase shifts that each qubit accumulates throughout the quantum circuit and compensates for them by adjusting the phase of 1Q operations appropriately. Imperfections in the spatial phase tracking—due to temporal instabilities in the magnetic field environment and imperfections in the calibration routines that set the 1Q optical phases—lead to memory errors during the transport operations and sideband cooling time. Additional sources of memory error are the finite T1 time of several minutes, transport failures, or background gas collisions leading to an unintentional qubit reorder (the last two are difficult to distinguish experimentally).

F. Classical programming and CPU-QPU interactions

Quantum algorithm developers can write programs for H2 in different frameworks and languages so long as their programs compile to either OpenQASM 2.0 or QIR [44,45]. Both representations contain real-time support for classical operations in the middle of the circuit, conditional expressions that rely on these classical calculations that are performed in real time, and elementary feed-forward operations conditioned on measurement results.

Many quantum computing applications call for interactions between classical and quantum processing units. Perhaps the most notable example is quantum error correction schemes in which syndrome measurement results are sent to a classical computer where a decoding algorithm is used to determine recovery operations and update quantum circuits in real time. As discussed in our previous work [15,19], we have demonstrated this capability using two different frameworks: (1) OpenQASM 2.0++, which allows for real-time decision making, and (2) a more capable classical compute environment, utilizing Web Assembly (Wasm) [46], that can execute complex calculations. Option (2) has significantly enhanced capabilities aimed at the development of hybrid quantum-classical algorithms and is crucial for applications like quantum error correction.

III. COMPONENT OPERATIONS AND BENCHMARKS

As our first level of benchmarking, we measure the errors from various component operations in the system. Quantum operations (e.g., gates and SPAM) dominate the error budget but are only performed in the DG zones, and therefore, we measure the performance with a subset of eight qubits (two per DG zone). Other errors that occur during a circuit, such as memory errors, are measured with an interleaved randomized benchmarking (RB) experiment performed simultaneously on all 32 qubits. Details of each component benchmarking experiment are given below:

- (i) *SPAM experiment*: Prepare each qubit in the DG zones in $|0\rangle$ and measure the probability of finding $|1\rangle$. Repeat for preparation in $|1\rangle$ and measure the probability of finding $|0\rangle$. The average is the SPAM

error per qubit. This procedure cannot differentiate between state preparation and measurement errors; however, detailed modeling predicts that the SPAM error is dominated by measurement error for $^{171}\text{Yb}^+$ [47,48].

- (ii) *1Q gate randomized benchmarking (1Q RB)*: We use the standard Clifford-twirl randomized benchmarking for measuring the error of 1Q gates [49], with a random final Pauli to fix the asymptote [50]. We report the average infidelity per 1Q Clifford.
- (iii) *2Q gate randomized benchmarking (2Q RB)*: Similar to 1Q RB, we use the Clifford-twirl technique [49] for measuring the error of 2Q gates. Each 2Q Clifford is constructed with zero to three $U_{ZZ}(\pi/2)$ gates, and each sequence includes a random final Pauli to fix the asymptote [50]. We scale the 2Q Clifford average infidelity by the average number of $U_{ZZ}(\pi/2)$ gates per Clifford, which is 1.5, and report that as the average infidelity per 2Q gate. An example decay plot is shown in Fig. 4(a).
- (iv) *2Q SU(4) gate randomized benchmarking [2Q SU(4) RB]*: We use the same general technique as 2Q RB, but instead of 2Q Cliffords, we use unitaries randomly sampled from the Haar measure over SU(4) constructed with three parametrized $U_{ZZ}(\theta)$ gates, and for each sequence, we include a random final Pauli to fix the asymptote [50]. We report the average infidelity per SU(4) operation.
- (v) *2Q parametrized gate randomized benchmarking*: We use a direct randomized benchmarking procedure [51] to measure the average infidelity of the parametrized 2Q gate $U_{ZZ}(\theta)$ as a function of angle θ . The details of the protocol are in Appendix B 3. A plot of the average infidelity versus angle is shown in Fig. 3.
- (vi) *Measurement and reset crosstalk depumping*: Measurement and reset crosstalk errors are estimated with bright-state depumping experiments [31], where a subset of qubits are prepared in $|1\rangle$ and other qubits are measured or reset repeatedly. The qubits in $|1\rangle$ decay due to crosstalk errors from the repeated process, and the decay rate scales with the average infidelity.
- (vii) *Interleaved transport randomized benchmarking (Transport 1Q RB)*: During a circuit, qubits incur errors in between successive 2Q gates due to idling during transport and cooling (memory errors) and the application of 1Q gates (errors from any mid-circuit measurements or resets are considered above). The contribution of these errors to representative circuits is measured with an interleaved 1Q RB experiment on all 32 qubits: 1Q Clifford gates are interleaved with “dummy” 2Q gates on 16 random pairings (our choice for a representative transport sequence). The dummy 2Q gates force ion

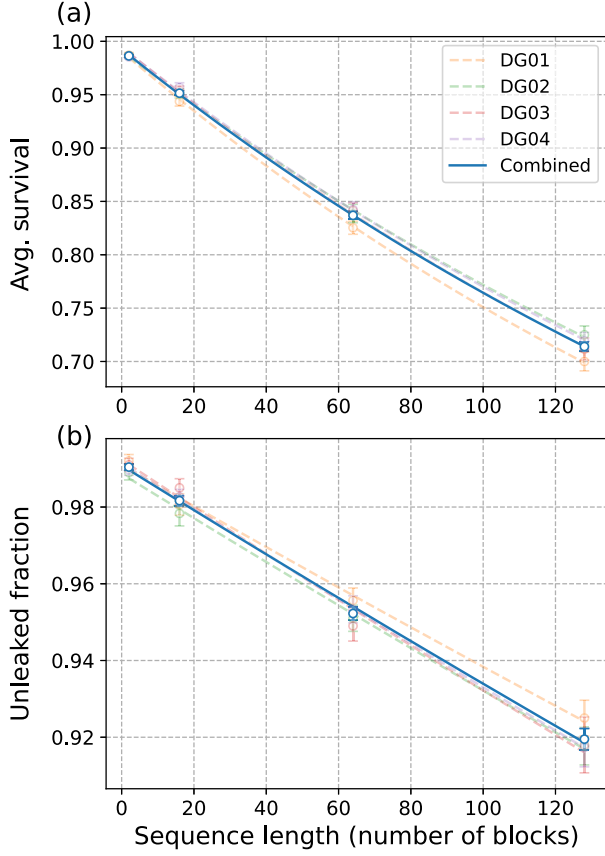


FIG. 4. The 2Q randomized benchmarking decay curves for each zone and for the combined average across all zones. (a) Standard RB decay curve. The average infidelity per 2Q gate is $1.83(5) \times 10^{-3}$ across all four gate zones. (b) Decay of fraction of shots without leakage on either qubit as identified by the leakage detection gadget, which gives a measured leakage rate per 2Q gate of $3.9(2) \times 10^{-4}$ across all four gate zones.

rearrangement, transport for 2Q gating, and side-band cooling but do not apply any 2Q lasers, avoiding 2Q gate errors and leaving the 32-qubit state fully separable. The resulting RB decay rate depends on both the average 1Q gate infidelity and memory errors, but we expect that memory errors are the dominant contribution.

Additional experimental details and data can be found in Appendix B. Results from these experiments are reported in both Table II (averaged over zones) and Table VI. An example breakdown of circuit timing for 2Q RB and transport 1Q RB is shown in Table I.

For 1Q and 2Q RB, we also measured the rate of leakage errors per gate by applying a “leakage detection gadget” at the end of each circuit, as illustrated in Fig. 5. The leakage detection gadget uses an ancilla qubit to flag shots that had a leakage error, i.e., an error that moved population outside of the computational subspace. In our system, leakage is most likely due to the unavoidable spontaneous emission that occurs in gates driven by a stimulated Raman process.

TABLE II. Average component benchmarking results for the tests outlined in Sec. III. All values are in terms of average infidelity and $\times 10^{-4}$. The values reported here are averaged over all four zones along with the one-sigma uncertainty from semi-parametric bootstrap resampling [52]. Data from individual zones are detailed in Table VI.

Test	Average infidelity ($\times 10^{-4}$)
1Q RB	0.25(3)
1Q leakage	0.04(2)
2Q RB	18.3(5)
2Q leakage	3.9(2)
2Q SU(4) RB	41(1)
2Q parametrized RB	See Fig. 3
Transport 1Q RB	2.2(3)
Measurement crosstalk	0.045(6)
Reset crosstalk	0.038(6)
SPAM	16(1)

Specifically, one half of the spontaneous Raman scattering errors are expected to result in leakage [53,54]. The leakage rate per gate r_L is defined as the rate that population leaves the computational subspace (whether 1Q or 2Q). We can estimate r_L by repeating the gate ℓ times, applying the gadget to each gated qubit, and fitting the leakage detection rate as shown in Fig. 4(b). Further details are given in Appendix B 2.

IV. SYSTEM-LEVEL BENCHMARKS

Benchmarks of component operations are a crucial fine-grained tool for estimating the contribution of various errors to quantum circuits. However, there are many potential ways in which they can mischaracterize device performance, for example, when crosstalk or non-Markovian errors are present. Therefore, it is important to also benchmark performance on a variety of more complex, multiqubit circuits, and to assess to what extent that performance can be understood from the measured performance of the constituent operations. Here, we present results from four system-level benchmarks: (A) mirror

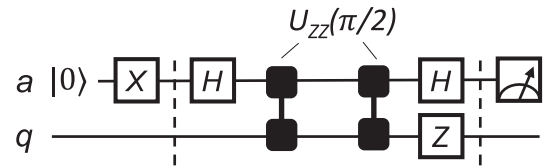


FIG. 5. Leakage detection gadget, adapted from Ref. [55]. The gadget uses an ancilla qubit “a” to detect whether qubit “q” has leaked. The ancilla is initially prepared in $|1\rangle$. If q has leaked, the 2Q gates have no logical effect, and a is measured as $|1\rangle$. If instead a has leaked, then a will still be measured as $|1\rangle$ since the leakage results in the ion being in a bright state. If neither q nor a has leaked, then the gadget (within the barriers in the circuit diagram) acts as $X_a I_q$, and a is measured as $|0\rangle$.

benchmarking [56,57], (B) quantum volume (QV) [58,59], (C) linear cross-entropy measurements for random 2D circuits [60], and (D) creation and certification of N -partite entanglement in GHZ states.

In benchmarks (A)–(C), the random structure of the circuits justifies simple heuristic arguments relating the overall circuit performance to the component operation fidelities. This allows us to assess how accurate our component-level error model translates to actual circuit implementations and identify potential gaps in our understanding of error sources. In each case, we assume that all non-SPAM errors can be attributed to the 2Q gates themselves and come in the form of a depolarizing channel (with uniform fidelity) attached to each 2Q gate. This approach accumulates all errors that happen to qubits between 2Q gates, primarily due to 1Q gate errors and memory errors, and lumps them in with the 2Q gate to form an effective “per 2Q gate” error rate, which we denote by $\epsilon_{\text{eff}}^{2\text{Q}}$. To obtain a simple but reasonable estimate for $\epsilon_{\text{eff}}^{2\text{Q}}$ based on the component benchmarks, we first determine the average angle $\bar{\theta}$ of 2Q gates used in the system-level benchmark. The data and linear fit reported in Fig. 3 then allow us to estimate the 2Q gate contribution. We then add this 2Q gate contribution together with twice the error from Transport 1Q RB, giving a predicted effective error

$$\epsilon_{\text{eff}}^{2\text{Q}} = 10^{-3}(2.9(2)\bar{\theta}/\pi + 0.9(1)). \quad (1)$$

Using analyses described in the appendixes, we also extract an inferred $\epsilon_{\text{eff}}^{2\text{Q}}$ from the system-level benchmarking data presented below and report the comparisons to the predicted values in Table III. The agreement is not perfect, nor is it expected to be, given that memory errors can be highly circuit dependent. For example, in QV circuits, multiple 2Q gates happen with very little delay in between, whereas the memory error per 2Q gate inferred from Transport 1Q RB assumes a single random reconfiguration of ions between every repeated gate on a given qubit, which likely contributes to the overestimate of $\epsilon_{\text{eff}}^{2\text{Q}}$ reported in Table III. Nevertheless, the overall reasonable agreement between predicted and inferred values suggests that the results of large-scale circuits are generally well aligned with expectations based on the individual component benchmarks.

TABLE III. Effective error per 2Q gate $\epsilon_{\text{eff}}^{2\text{Q}}$ inferred from system-level benchmarks, compared to the average per 2Q gate error estimated from combining the component benchmarks as described in the text [see Eq. (1)].

Source	$\bar{\theta}$	$\epsilon_{\text{eff}}^{2\text{Q}} (\times 10^{-3})$ (inferred)	$\epsilon_{\text{eff}}^{2\text{Q}} (\times 10^{-3})$ (predicted)
Mirror benchmarking	0.5π	2.6(2)	2.4(1)
Quantum volume	0.35π	1.7(1)	1.9(1)
Random circuit sampling	0.42π	1.9(2)	2.1(1)

A. Mirror benchmarking

Circuit mirroring was introduced as a scalable way to benchmark arbitrary quantum circuits [56,61]. We perform a randomized circuit mirroring experiment that we refer to as mirror benchmarking (MB). As described in Ref. [57], MB circuits consist of layers of 1Q gates on all qubits and 2Q gates between random pairings of the qubits with full connectivity. The 1Q gates are Clifford gates sampled uniformly at random, and each 2Q gate is the native $U_{ZZ}(\pi/2)$ gate. The circuits are “mirrored,” meaning that the inverse circuit is applied in the second half. A final random N -qubit Pauli is applied to randomize the ideal outcome for each circuit. The circuits also employ Pauli randomization on the 2Q gates so that the error channel per layer can be treated as a stochastic Pauli [62]. The circuit-averaged probability of observing the ideal outcome as a function of the number of circuit layers will then decay exponentially. If the 2Q gate error channel is depolarizing, then the decay parameter as a function of the 2Q gate average fidelity is given by an analytic formula [Eq. (C4) in Ref. [57]]. In practice, the 2Q errors in H2 are not depolarizing, but this equation still provides a useful heuristic. Fitting experimentally measured decay curves to exponentials and inverting this formula provides an effective 2Q infidelity for the system that includes 1Q gates, 2Q gates, and the memory error for random permutations.

We performed MB experiments on H2 with $N = 20, 26,$ and 32 qubits. The decay plots are shown in Fig. 6, and the results are listed in Table VII. For $N = 32$, we find $\epsilon_{\text{eff}}^{2\text{Q}} = 2.6(2) \times 10^{-3}$. Importantly, we find that $\epsilon_{\text{eff}}^{2\text{Q}}$ does not increase with qubit number.

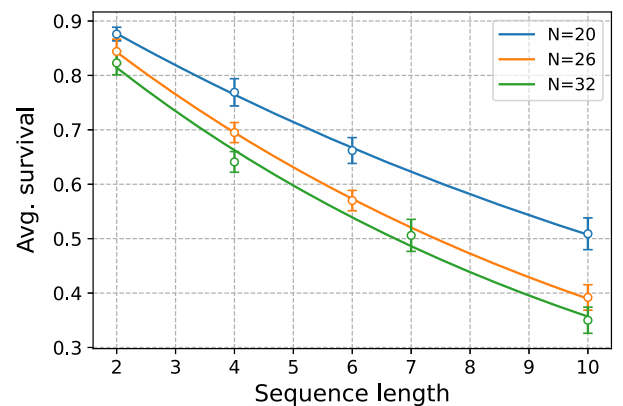


FIG. 6. Mirror benchmarking experiments on H2. The sequence lengths correspond to half the circuit depths, so circuits of length ℓ contain $2\ell N$ 2Q gates with full connectivity. Ten random circuits were run at each sequence length with 100 shots per circuit. The average survival probabilities are fit to the model $p(\ell) = Au^{\ell-1}$. The parameter u is used to obtain an effective 2Q gate average fidelity for a constant 2Q depolarizing error [57].

B. Quantum volume

Quantum volume is a system-level test designed to be comparable across gate-based quantum computers. The QV test is run with a collection of random circuits acting on N qubits. Each random circuit is generated by randomly pairing all qubits, applying random $SU(4)$ unitaries to each pair, and repeating for N rounds. The performance is assessed with a heavy-output test that requires classical simulation of the quantum circuits. The test is passed when the probability of generating heavy outputs is greater than $2/3$ with two-sigma confidence, which yields a measured value of $QV = 2^N$ [58]. A totally decohered circuit returns heavy outputs half the time, so the QV test's threshold of $2/3$ requires that the errors are small enough to be strongly distinguishable from a random distribution. Therefore, a QV of 2^N implies high performance on many circuits with more than N qubits and/or depth greater than N , as evidenced by several example algorithms run on all 32 qubits in Sec. V. QV has been measured on a variety of different systems [63], with the largest previously reported value of $QV = 2^{15}$ from H1 [64].

We performed several QV measurements, with the highest measured value being $QV = 2^{16}$. The QV = 2^{16} test data are shown in Fig. 7; we used 200 randomly generated circuits in each run with 100 shots and an average of 296 parametrized 2Q gates. The measured heavy-output probability is 68.2%, with a two-sigma lower bound of 67.3% calculated with the semiparametric bootstrap method outlined in Ref. [59].

C. Random circuit sampling (RCS)

A system-level benchmark of recent interest is the computational task of sampling the output distributions of random quantum circuits. Like QV, RCS is not a scalable

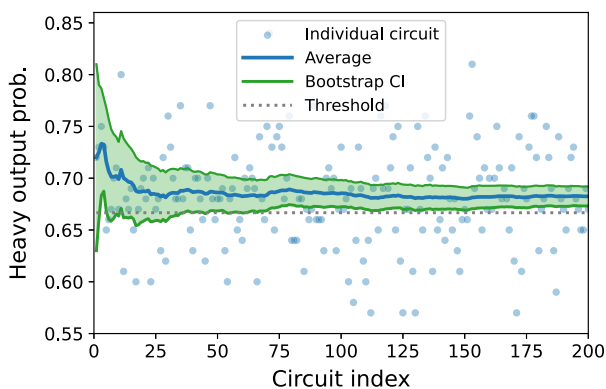


FIG. 7. Quantum volume $QV = 2^{16}$ measurement on H2. The average and two-sigma confidence intervals of the heavy-output probability are plotted as a function of the circuit index. Passing occurs when the green shaded region (two-sigma confidence interval from semiparametric bootstrap method) is above the dashed gray line at $2/3$, which we satisfy in a data set with 200 randomly generated circuits.

benchmark as it requires classical computation time exponential in N . It was recently proven [65] that at a fixed gate error, RCS is not a scalable route to quantum supremacy at large N ; however, it still tests the quantum computer's ability to faithfully execute circuits for which classical simulation methods are, at least in practice, extremely difficult given high enough gate fidelities. It has also been run on a variety of quantum computers in the context of quantum advantage demonstrations [60,66,67], making it useful from the standpoint of cross-platform comparisons.

We structure our circuits as if the qubits involved tile a two-dimensional grid with nearest-neighbor interactions [60], although we emphasize that this constraint is only imposed for fair comparison with prior art and is not a hardware constraint of H2. Future work may study whether random circuits built from randomly gating pairs of qubits with arbitrary connectivity achieve a greater degree of classical simulation difficulty at a reduced circuit depth. At each N , the grid dimensions are chosen to be as close to square as possible. In each layer from the circuit, a 1Q gate chosen randomly from $\{\sqrt{X}, \sqrt{Y}, \sqrt{W}\}$, where $W = 1/\sqrt{2}(X + Y)$, is applied to each qubit. The 1Q gate applied to a given qubit in one layer is omitted from the set of possible 1Q gates applied to the same qubit in the next layer. Subsequently, a 2Q gate is applied to pairs of qubits following a particular tiling pattern on the grid (see Ref. [60], Fig. 3). A final round of 1Q gates is applied to all qubits before measurement. We implement the exact same 2Q gate as in Ref. [60],

$$\text{fSim}\left(\frac{\pi}{2}, \frac{\pi}{6}\right) = e^{-i\pi(X\otimes X + Y\otimes Y)/4} e^{-i\pi(I-Z)\otimes(I-Z)/24}. \quad (2)$$

Up to 1Q gates,

$$\begin{aligned} \text{fSim}\left(\frac{\pi}{2}, \frac{\pi}{6}\right) &\simeq \text{iSWAP} \cdot U_{ZZ}\left(\frac{\pi}{12}\right) \\ &\simeq \text{SWAP} \cdot U_{ZZ}\left(-\frac{5\pi}{12}\right). \end{aligned} \quad (3)$$

The SWAP gate is handled in software by relabeling and transporting qubits, so the $\text{fSim}[(\pi/2), (\pi/6)]$ gate is implemented on H2 with exactly one 2Q gate. In practice, we generate the circuits using the Sycamore gate defined in the `pytket` library [68]; `pytket`'s compilation to Quantinuum hardware automatically rebases the circuits to use a $U_{ZZ}(5\pi/12)$ gate with the opposite sign from the above identity, by using a different choice of 1Q wrapper gates.

Since H2 currently supports 32 qubits, well within the classically feasible regime, we focus on the ‘‘classically verifiable’’ repeating EFGHEFGH gate tiling pattern from Ref. [60]. As developed in Ref. [69], we use linear cross-entropy benchmarking to quantify the success of

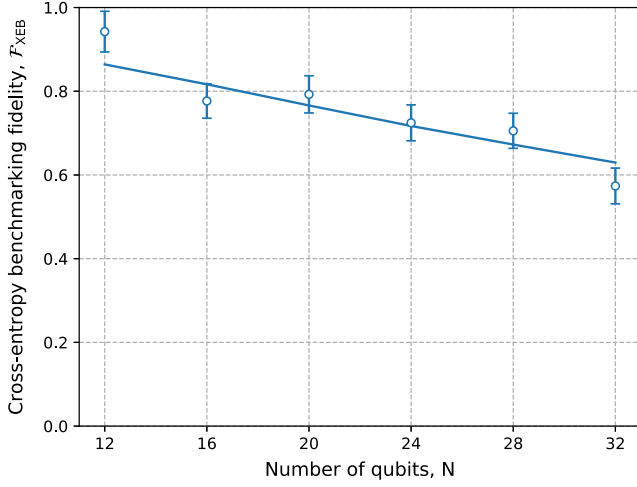


FIG. 8. Linear cross-entropy benchmarking fidelity as measured on H2 for classically verifiable random circuits. Each data point displays the combined results from ten circuits, each executed with 100 shots. The details of the best-fit curve are described in Appendix C 3.

the quantum computer in sampling from the true output distribution of each random circuit. This procedure computes a quantity called the linear cross-entropy benchmarking fidelity, \mathcal{F}_{XEB} . To match the parameters in Ref. [60], we explore random circuits of depth 14 and average the resulting \mathcal{F}_{XEB} over ten circuits at each fixed N , combining the uncertainties on each measurement of \mathcal{F}_{XEB} by inverse-variance weighting. The measured results on H2 are displayed in Fig. 8. With future improvements to the number of qubits in H2, assuming comparable $\epsilon_{\text{eff}}^{2Q}$, we expect the cross-entropy benchmark results will pose serious challenges to classical simulations.

D. N -partite entanglement certification in GHZ states

The N -qubit GHZ state [70] is defined as

$$|\text{GHZ}_N\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes N} + |1\rangle^{\otimes N}). \quad (4)$$

Producing GHZ states is a demanding test of qubit coherence, as they are maximally sensitive probes of global dephasing. Moreover, GHZ state fidelities have been widely measured and reported across a variety of quantum hardware [71–75], making this test helpful for assessing the performance of the H2 device in a broader context.

We prepare GHZ states of $N = 20, 26$, and 32 using the log-depth circuit construction given in Ref. [76] and an $N = 32$ GHZ state using a constant-depth adaptive circuit construction [77,78]. The adaptive circuit was submitted via OpenQASM 2.0++, and it exemplifies how midcircuit measurement and feed forward can be used to create

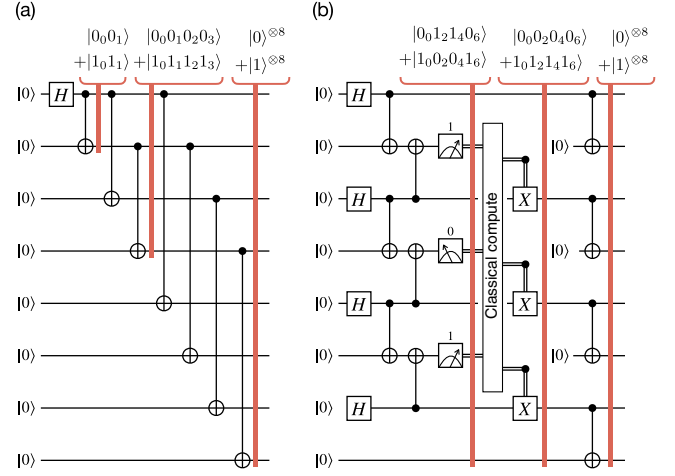


FIG. 9. GHZ state preparation circuits for (a) log-depth unitary and (b) constant-depth adaptive preparation, here shown for $N = 8$, for simplicity.

long-range-entangled states in constant depth [79–81]. Both circuit constructions are shown in Fig. 9.

We estimate the fidelity of the GHZ states using the method of Ref. [82]. The fidelity of a density matrix ρ with respect to the GHZ state is

$$F(\rho, |\text{GHZ}_N\rangle) = \frac{1}{2} \text{Tr}(\rho|0\rangle\langle 0|^{\otimes N}) + \frac{1}{2} \text{Tr}(\rho|1\rangle\langle 1|^{\otimes N}) + \frac{1}{2} \text{Tr}(\rho(|0\rangle\langle 1|^{\otimes N} + |1\rangle\langle 0|^{\otimes N})). \quad (5)$$

The first two terms are the populations in the all-zero and all-one states and are estimated by measuring all qubits in the computational basis. The third term is estimated using the fact that

$$|0\rangle\langle 1|^{\otimes N} + |1\rangle\langle 0|^{\otimes N} = \frac{1}{N} \sum_{k=1}^N (-1)^k M_k, \quad (6)$$

where the operators

$$M_k = (\cos(k\pi/N)X + \sin(k\pi/N)Y)^{\otimes N}, \quad (7)$$

for $k \in \{1, \dots, N\}$, correspond to the global parity of spin along the axis $\theta_k = k\pi/N$ on the equator of the Bloch sphere and can be measured with only 1Q rotations. The complete fidelity estimation protocol requires $N + 1$ measurement bases.

We ran one circuit with 50 shots for each of the N measurements of M_k , and N circuits with 50 shots for the population measurements. All the log-depth unitary preparation circuits across the various N were run in a random order. The results of the population and parity measurements are shown in Fig. 10, and the estimated state fidelities are listed in Table IV. For $N = 32$, we obtain

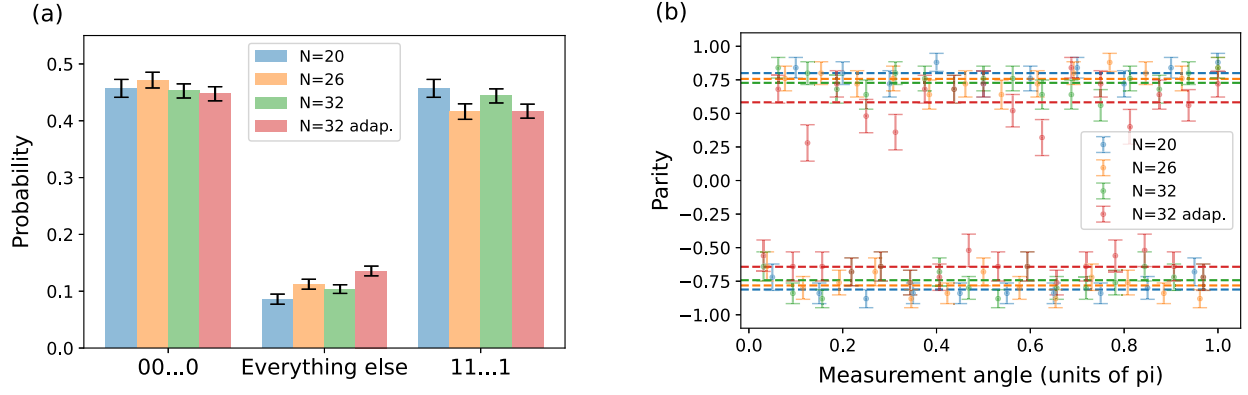


FIG. 10. Populations and parities of $N = 20$ -, 26 -, and 32 -qubit GHZ states constructed with a log-depth unitary protocol, and also of a 32 -qubit GHZ state produced with a constant-depth adaptive circuit. (a) Populations of $|0\rangle^{\otimes N}$ and $|1\rangle^{\otimes N}$. The ideal GHZ state has populations of 0.5 in these two states and zero in all other states. (b) Expectation values of the operator M_k defined in Eq. (7), plotted versus angle $\theta_k = k\pi/N$. The ideal GHZ state has values of 1 and -1 for even and odd k , respectively. The dashed lines denote the averages.

fidelities of $0.82(1)$ and $0.74(1)$ (without correcting for SPAM errors) for the unitary and adaptive state preparation circuits, respectively. By comparison, a GHZ fidelity greater than 0.5 is sufficient to witness genuine multipartite entanglement [83]. The adaptive circuit contains more $2Q$ gates (46) and measurements (48), and therefore produces a lower fidelity than the unitary circuit, which contains 31 $2Q$ gates and 32 measurements. Especially for systems with limited connectivity and appreciable memory errors, the constant-depth adaptive circuit should outperform the unitary preparation circuit at large enough N .

V. APPLICATION BENCHMARKS

The system-level benchmarks of the previous section serve to verify quantum computer performance on a well-defined set of volumetric circuits. The measured effective $2Q$ error rates for these circuits, as given in Table III, show that the component-level benchmark performance is translated to larger circuits. However, problems of practical interest tend to involve structured circuits with very specific demands on gate set and connectivity. A comprehensive survey of all such problems is beyond the scope of this work (and difficult to define), but a sampling of such applications is still helpful for evaluating the machine’s capabilities with respect to plausible near-term use cases and the demands they impose. In this section, we present the results of four application benchmarks: (A) Hamiltonian simulation, (B) QAOA, (C) large-distance

repetition codes, and (D) holographic quantum dynamics simulation. We chose these benchmarks in a complementary way, as each places a different emphasis on particular error sources. For example, Hamiltonian simulation is highly dependent on $2Q$ gate error, QAOA performance depends strongly on qubit connectivity, and repetition codes and the holographic quantum dynamics simulation require high-fidelity MCMR.

A. Hamiltonian simulation

Simulating the continuous time evolution of many-body quantum systems is an important and classically challenging problem for which quantum computers are well suited [84–86]. To benchmark the performance of the H2 quantum computer on this task, we simulate the dynamics of an $L = 32$ -site transverse-field Ising model (TFIM) in one spatial dimension, with Hamiltonian

$$H = -J \sum_{j=1}^L Z_j Z_{j+1} - h \sum_{j=1}^L X_j. \quad (8)$$

Here and elsewhere in this section, site subscripts are taken $\text{mod}(L)$ to yield periodic boundary conditions. We simulate a quantum quench where the initial state is prepared in the ground state at $h/J = \infty$, that is, $|\Psi(0)\rangle = |+\rangle_j^{\otimes L}$. The Hamiltonian is suddenly quenched to $h/J = 0.2$, and the state is then evolved up to $Jt = 7$ under the new Hamiltonian. We evaluate the dynamics of the expectation value of the Pauli X operator averaged over all qubits, i.e., $\langle X \rangle \equiv (1/L) \sum_{j=1}^L \langle X_j \rangle$. We digitally simulate the dynamics using first-order Trotterization of the time-evolution operator [87],

$$U(t) \approx \left(\prod_j \exp \left[iZ_j Z_{j+1} \frac{Jt}{r} \right] \prod_j \exp \left[iX_j \frac{ht}{r} \right] \right)^r, \quad (9)$$

TABLE IV. GHZ state fidelities.

GHZ prep	State fidelity
$N = 20$	$0.86(1)$
$N = 26$	$0.83(1)$
$N = 32$	$0.82(1)$
$N = 32$ adaptive	$0.74(1)$

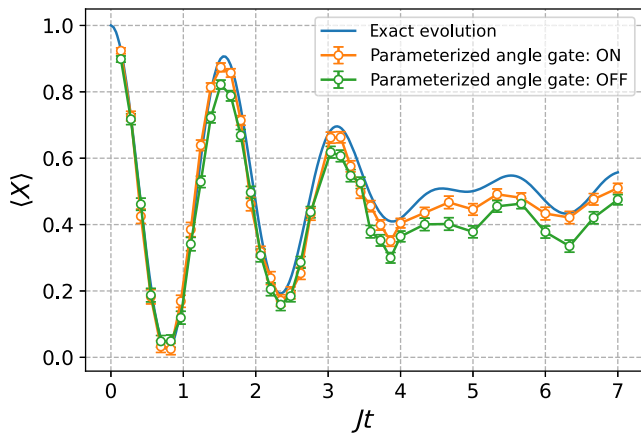


FIG. 11. Dynamics of $\langle X \rangle$ for a 32-qubit TFIM Hamiltonian simulation vs evolution time. The orange data are obtained by directly implementing each ZZ rotation in every Trotter step using our native parametrized angle $U_{ZZ}(\theta)$ gate with $\theta = 2Jt/r$. The green data are obtained by decomposing each ZZ rotation into two $U_{ZZ}(\pi/2)$ (Clifford) gates with some 1Q rotations. Each data point is obtained as the average of 100 shots of the associated Trotterized circuit for that time.

which approaches the true evolution as $r \rightarrow \infty$. The 1Q and 2Q gates in this decomposition are X rotations and $U_{ZZ}(\theta)$, which are native on H2. The number of Trotter steps r is chosen such that the errors on $\langle X \rangle$ due to Trotterization are below 0.01, as determined by explicit calculations of the noiseless Trotterized dynamics [88] and by comparison to exact results for the continuous-time evolutions [89]. This threshold ensures that Trotter errors are at or below the scale of the expected, around 1%, statistical fluctuation in the experiment (more details are given in Appendix D 1).

The results of our experiment, plotted in Fig. 11, show reasonably good agreement between our quantum simulation and the exact solution up to time $Jt = 7$, suggesting the quantum computer has small enough errors to coherently simulate quantum dynamics up to a nontrivial time (note that a completely depolarized state has $\langle X \rangle = 0$). The data have not been postprocessed or error mitigated in any way. Figure 11 also compares the circuit implementations with and without the parametrized angle $U_{ZZ}(\theta)$ gate. Without parametrized angle gates, every such gate has to be decomposed into two $U_{ZZ}(\pi/2)$ gates with additional 1Q rotations, resulting in a doubling of the number of 2Q gates and the possibility of more than doubling the error per Trotterization step (see Fig. 21). The improvements to the simulation results when using parametrized-angle 2Q gates highlights their benefit for near-term applications of quantum computers to simulating many-body physics.

B. QAOA

The quantum approximate optimization algorithm (QAOA) [90] is a near-term heuristic algorithm for solving combinatorial optimization problems of general interest

in many industries. As in previous benchmarking studies [38,91,92], we focus on solving the MaxCut problem restricted to the class of unweighted 3-regular graphs $G = (V, E)$. The standard QAOA circuit consists of alternating applications of a mixing unitary $U_B(\beta_n) = e^{-i\beta_n H_B}$ and a phase-splitting cost unitary $U_C(\gamma_n) = e^{-i\gamma_n H_C}$,

$$U(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{n=1}^p U_B(\beta_n) U_C(\gamma_n). \quad (10)$$

The $2p$ parameters β_n and γ_n are found variationally, by using a classical optimization algorithm to search for the choice of parameters that extremizes the cost of the QAOA final state,

$$\langle H_C \rangle = \langle \psi_0 | U(\boldsymbol{\beta}, \boldsymbol{\gamma})^\dagger H_C U(\boldsymbol{\beta}, \boldsymbol{\gamma}) | \psi_0 \rangle. \quad (11)$$

The initial state is taken to be $|\psi_0\rangle = |+\rangle^{\otimes N}$, the ground state of $H_B = \sum_i X_i$, while for the unweighted MaxCut problem, the cost Hamiltonian is

$$H_C = \frac{1}{2} \sum_{(i,j) \in E} (1 - Z_i Z_j). \quad (12)$$

Therefore, each term in the cost Hamiltonian comprising the cost unitary $U_C(\gamma_n)$ can be implemented with a single $U_{ZZ}(\theta)$ gate.

For the classical optimization procedure, we use the derivative-free BOBYQA optimizer [93] as implemented in the PY-BOBYQA package [94]. The BOBYQA optimizer builds a local quadratic model to the objective function within a trust region of size that decreases with iterations of the optimizer. We set the optimizer convergence conditions to be met when the precision of the variational parameters reaches the same order as the measured 2Q gate errors, 1×10^{-3} .

We study two separate experiments in this work. The first implements a larger-scale MaxCut QAOA problem ($N = 130$, $p = 1$) on 32 physical qubits using qubit-reuse compilation [38] and 100 shots per circuit. The second experiment solves an $N = 32$ MaxCut QAOA problem at $p = 2$ with 200 shots per circuit to demonstrate an improvement in solution quality compared to $p = 1$ with the more expressive and deeper ansatz. The results for these experiments are stated in terms of the average and minimum dimensionless energies measured over the course of the optimization, where the dimensionless energy corresponds to the cut value computed by the cost Hamiltonian in Eq. (12). For plotting purposes, the energy was rescaled by a sign so that, in all cases, the optimum corresponds to the solution of minimum energy.

In Fig. 12, we display the results from the $N = 130$, $p = 1$ experiment. The optimizer shows convergence within the first ten circuits. Using the tensor network methods

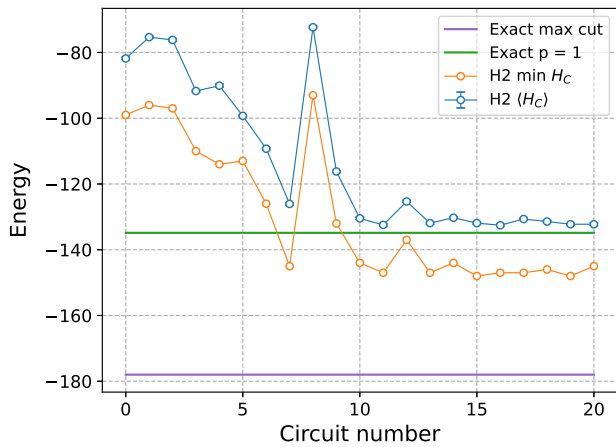


FIG. 12. Optimization trajectory of $N = 130$, $p = 1$ QAOA computed via qubit reuse on H2. The expectation value of the energy as measured experimentally at $p = 1$ (blue data) converges well to the best-possible exact value (green line). Uncertainties on the measured value of $\langle H_C \rangle$ are plotted but are smaller than the displayed point size (see Appendix D 2 for details). The best sample taken at each iteration (orange data) is also displayed relative to the true max cut (purple line).

available in the PYTHON library `qaimb` [95] in conjunction with the global Bayesian optimizer in `scikit-optimize` [96], we also exactly evaluated the best average energy possible for any $p = 1$ circuit. The convergence of the blue data to the green line in Fig. 12 demonstrates that the optimization procedure succeeded in locating the optimal parameters and that H2 evaluated the circuits with sufficiently low noise to nearly saturate the best-possible result. In Appendix D 2, we also display the optimization trace on the energy landscape, further confirming that the optimizer succeeded in locating the optimal parameters. To evaluate the performance of the algorithm in solving the combinatorial problem, we also compare the minimum value of the energy sampled in any given shot to the exact value of the max cut computed in `gurobi` [97]. As expected, since the circuit depth is only $p = 1$, the best cut value found on H2, 148, is substantially less than the exact value of 178. Nevertheless, this experiment represents substantial progress towards solving industry-scale combinatorial problems with QAOA on small quantum computers.

In Fig. 13, we demonstrate the results of the $N = 32$, $p = 2$ optimization procedure. Comparing to the best average energies possible for any $p = 1$ or $p = 2$ circuits, the experimental data for $p = 2$ consistently perform better than the best-possible $p = 1$ circuit and are close to saturating the ground-state energy for $p = 2$ circuits. Furthermore, H2 succeeded in locating solutions with the best-possible max cut of 42 for this graph.

C. Error correction: Repetition code

Large quantum computations are widely thought to only be possible through quantum error correction (QEC).

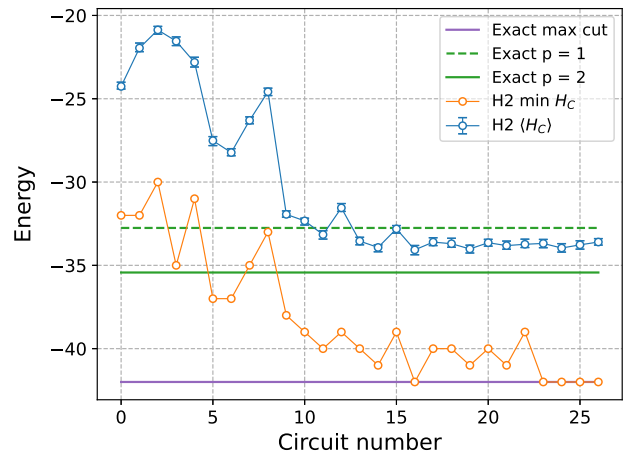


FIG. 13. Optimization trajectory of $N = 32$, $p = 2$ QAOA on H2. The expectation value of the energy as measured experimentally at $p = 2$ (blue data) surpasses the best-possible exact value for a $p = 1$ circuit (green dashed line). The best sample taken at each iteration (orange data) is also displayed relative to the true max cut (purple line).

Therefore, in the context of fault-tolerant quantum computers, perhaps the most important quantum algorithm is not a particular targeted calculation but rather the QEC algorithm being run in the background. Additionally, given the large resource overheads of QEC, the design requirements for large-scale quantum computers will likely be driven by the optimization of these codes' power and efficiency, highlighting the importance of closely related benchmarks.

Many QEC schemes are based on stabilizer codes that encode logical information into the joint subspace of many physical qubits, known as data qubits. Additional physical qubits, known as ancilla qubits, are used to make nondestructive syndrome measurements [98], which discretize errors into a manageable set of bit and phase errors, allowing for general QEC. Repetition codes are examples of stabilizer codes but can only correct a single type of error, typically either bit or phase flip errors, and thus do not constitute a full quantum error correction code. However, they make good benchmark algorithms since they possess all the components needed to implement a full quantum stabilizer code. Specifically, a distance d repetition code can reliably correct up to $(d - 1)/2$ errors. Corrections are determined by repeatedly measuring stabilizers of the code using MCMR, syndromes are decoded using algorithms similar to those used in quantum codes, and logical fidelities are calculated in the usual way. Using all 32 qubits, we implement a $d = 31$ repetition code with 31 data qubits and one ancilla, maximizing the code distance that can be tested. This low overhead implementation of the code is made possible by H2's qubit reuse capabilities and by performing 30 unique stabilizer measurements serially.

The syndrome measurements are processed in real time using Wasm calls to the classical compute environment

during the quantum circuit. At the end of the circuit, the data qubits are also measured and are used to construct a final syndrome measurement. We use this last syndrome in addition to all previously recorded syndromes to decode the logical output state and calculate the logical fidelities. The decoding uses a minimum-weight perfect-matching algorithm [99,100], which is performed online at the end of the circuit as part of the control-system software execution of each shot (i.e., while the hybrid quantum-classical program is still being executed on the actual device). Since this is performed after the logical qubit has been measured, the operation does not perform midcircuit real-time decoding, making these experiments insensitive to memory error associated with the computation time of a correction. Real-time decoding operations are possible with Wasm and the advanced classical compute environment infrastructure, but they are unnecessary for repetition-code memory experiments.

Experiments on both the $d = 31$ bit flip code and phase flip code were performed while varying the number of rounds of syndrome extraction and recording all syndrome measurements, allowing us to process subsets of the code after the program is completed. The subsets allow us to reconstruct logical fidelities for all odd distance codes less than $d = 31$. These measurements are similar to Refs. [101,102], which use a fixed architecture and parallel syndrome measurements, allowing for a direct comparison of different distances. In contrast, our architecture offers a less-direct comparison between different code distances, as syndrome measurements are performed serially, but allows for larger distance codes with lower qubit overheads. We note that the Wasm decoder was only used to calculate the $d = 31$ fidelities. All other code distance fidelities were calculated by the same minimum-weight perfect-matching algorithm offline.

The experimental results in Fig. 14 show that the larger distance codes achieve higher logical fidelities, as expected, with the bit flip code producing a higher logical fidelity compared to the phase flip code for a given distance, consistent with memory errors that are dominated by magnetic field inhomogeneities. These results demonstrate many of the necessary components for implementing scalable, real-time QEC, and show how the capabilities of the H2 system can help realize large-distance stabilizer QEC codes, all of which will be the subject of future studies.

D. Holographic quantum dynamics simulation (HoloQUADS)

High-fidelity MCMR is crucial for quantum error correction, and it can also help expand the reach of many near-term algorithms [38]. In particular, such techniques have been shown to enable the simulation of quantum dynamics from initially correlated states directly in the thermodynamic limit, with qubit number requirements set by the evolving entanglement entropy of the state rather than its

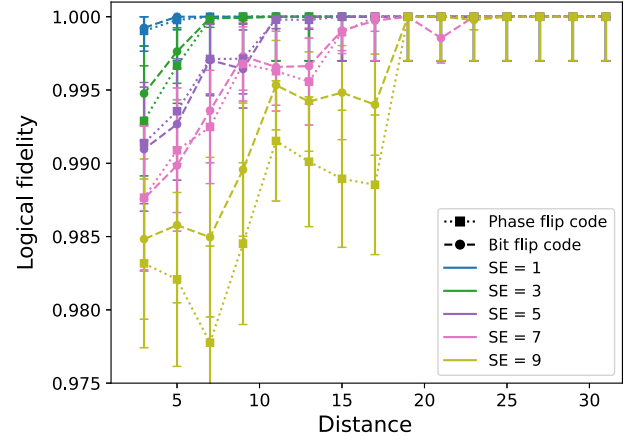


FIG. 14. Logical fidelities of the phase (square, dotted) and bit (circle, dashed) flip repetition codes as a function of distance. As the number of syndrome extraction (SE) rounds increases, more noise is injected into the system, degrading the logical fidelity. For a given number of rounds of SE, as the code distance increases, so does the logical fidelity. All error bars are calculated using jackknife resampling [103], except for those where the sampling number is too low to calculate an error (i.e., 100% fidelities), in which case the statistical rule of three was used [104].

physical size [105]. Based on the work in Refs. [106,107], Ref. [108] recently proposed and demonstrated a benchmark for such methods by simulating exactly solvable dual-unitary circuit models applied to initial matrix-product states on H1-1. Here, we use the additional resources of H2 to extend those results to longer evolution times, where the system contains more entanglement.

Following Ref. [108], we simulate time evolution under dual-unitary circuits [106,107], which are one-dimensional brick-work circuits [Fig. 15(a)] having generic properties of typical circuits (e.g., exhibiting quantum chaos and ballistic growth of entanglement [106]) and certain nongeneric properties (e.g., their correlations spread at the maximal possible velocity [109] and are confined to the light-cone boundary rather than its interior), which allow quantities such as entanglement entropy and correlation functions to be analytically determined [106,107,110]. An initial matrix-product state is prepared by applying gates between the physical qubits and an ancilla “bond” qubit [blue gates in Figs. 15(a) and 15(b)] and then time-evolving this state using the self-dual kicked Ising (SDKI) model [106,111] [green gates in Figs. 15(a) and 15(b)]. After t layers of SDKI gates are applied to $|\psi_0\rangle$, we measure the smoothed correlation functions

$$C^{xx}(r, t) = \frac{1}{2L} \sum_{j=1}^L \sum_{\delta=0,1} \langle \psi_t | X_j X_{j+r+\delta} | \psi_t \rangle, \quad (13)$$

where $|\psi_t\rangle$ is the time-evolved state and L is the system size; in Fig. 15(c), we compare the results to exact

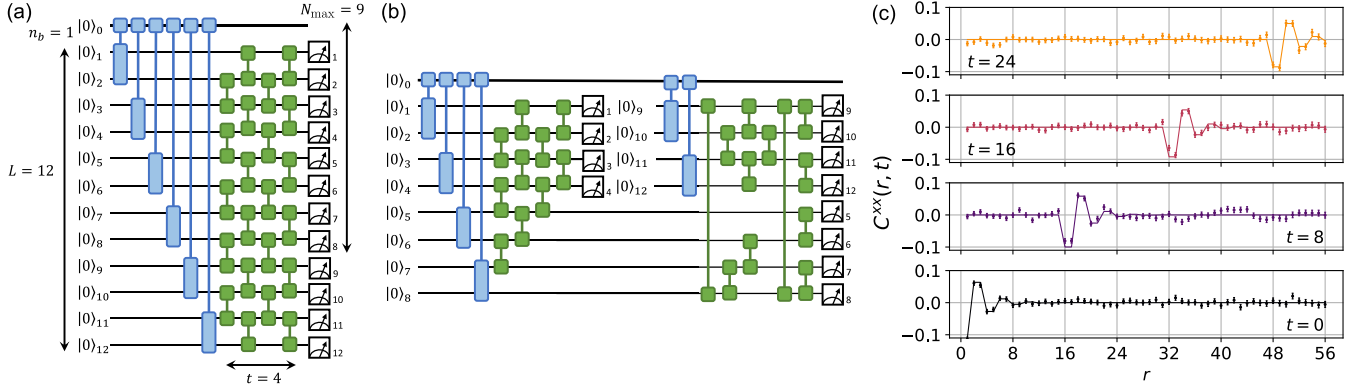


FIG. 15. (a) One-dimensional brickwork circuit of length $L = 12$ with $t = 4$ layers of gates applied to a quantum matrix-product state of bond dimension $\chi = 2^{n_b} = 2$. (b) HoloQUADS reusing qubits through MCMR to execute the same circuit with a minimal number of qubits. Here, we use $N_{\max} = 9$ qubits, but N_{\max} can be adjusted between $n_b + t + 2 = 7$ (maximally serial) and $n_b + L = 13$ (maximally parallel). (c) Experimentally measured (dots) correlation function $C^{xx}(r, t)$ for a dual-unitary circuit applied to a length $L = 128 + t$ solvable $\chi = 2$ quantum matrix-product state compared to the exact thermodynamic limit results (solid lines), using $N_{\max} = 32$ qubits up to time $t = 24$. Error bars are standard deviations of the mean from four 100-shot experiments.

theoretical calculations from Ref. [107]. We use H2’s 32 qubits to simulate $t = 0, 8, 16, 24$ layers of time evolution applied to a length $L = 128 + t = 128, 136, 144, 152$ matrix-product state.

The experimental data show close agreement with ideal noiseless results, suggesting that H2’s midcircuit measurement, midcircuit reset, crosstalk, and memory errors are low enough for sizable quantum dynamics simulations using HoloQUADS. We note that the effects of errors can be highly circuit dependent. For the particular dual-unitary circuit studied in this benchmark, their maximal velocity behavior [109] causes a number of Pauli errors proportional to t along the edges of the causal cones of qubits i and j to affect the $\langle \psi_i | X_i X_j | \psi_j \rangle$ correlation function. For a generic circuit, we would expect a number of Pauli errors proportional to t^2 within the causal cones to affect correlation functions, meaning dual-unitary circuits are less sensitive to errors than typical circuits.

VI. SUMMARY OF THE RESULTS AND OUR OUTLOOK

The H2 quantum computer is a significant upgrade from our previous H1 system, maintaining or exceeding many previous fidelity metrics while operating on more qubits. The clearest manifestation of the robust scalability of our QCCD architecture is that the system-level benchmarks are consistent with the errors measured by the component benchmarks. We also benchmarked H2’s performance on a variety of applications that are widely considered to be well suited for near-term quantum computers, with the goal of assessing the feasibility of such algorithms given current hardware performance metrics.

The current device has three error mechanisms that significantly contribute to the overall error budget: 2Q gate infidelity, memory error, and SPAM error. The relative

impact of these is circuit dependent, and we believe that all three can be significantly improved with future-generation hardware upgrades. The single largest source of error in the 2Q gate is spontaneous emission (it accounts for about 40% of the total 2Q gate error), which can, in principle, be suppressed by simultaneously increasing the laser detuning and power. Memory errors can be reduced by decreasing circuit execution time (currently dominated by transport and sympathetic cooling), reducing magnetic field noise, or reducing the qubit’s sensitivity to magnetic fields. Transport speeds are currently limited by the filtering that is necessary to achieve low heating rates at the ions’ motional-mode frequencies; the design and filtering of transport waveforms could be better matched to allow for faster transport while simultaneously reducing the required amount of sympathetic cooling. Magnetic field fluctuations could be mitigated at the apparatus design level with passive magnetic field shielding such as mu-metal or active feedback to measured magnetic fields. Magnetic field sensitivity of the qubit could be reduced by working at a lower bias field or by working with an ion species possessing a finite-field clock state. Additionally, memory errors could be mitigated at the software level via dynamical decoupling. SPAM error, which is somewhat deemphasized in most of our system-level and application benchmarks but important for fault-tolerant operations, can be significantly improved through quantum logic detection protocols [112] or by switching to a different ion species [113]. All of these schemes are under active investigation and are expected to lead to reduced circuit execution times and higher fidelity operation.

Even after these errors are substantially reduced, building ever larger systems with increased qubit number remains a significant engineering challenge. Two-dimensional trapping structures will likely be required for more efficient ion-transport operations [21] to maintain

the long-range connectivity without increasing memory errors, as well as more scalable optics delivery [24,114] to increase the number of quantum gates that can be executed in parallel. However, the H2 system demonstrates a number of key technological milestones on the path to that scaling, including ion transport controlled via broadcast electrode signals, rf signals routed under the surface of the trap, and fast MOT-based loading. These improvements are achieved in a system initially configured to operate with 32 qubits but designed to accommodate more, and collectively bolster the case for the viability of the QCCD architecture as a route to large-scale trapped-ion quantum computing.

All data presented are available in Ref. [115]. Most of the component benchmarking data are available in Ref. [116] and will be updated as H2 improvements are introduced. Quantum volume data are available in Ref. [117] and will be updated as new tests are run.

ACKNOWLEDGMENTS

We thank the entire Quantinuum team for numerous contributions that enabled this work, and we thank Matt Marcus, Philip Makotyn, and Tom Loftus for helpful and inspiring conversations. We thank Jack Ross for creating Fig. 1 and Eric Hudson for helpful comments on an earlier version of the manuscript.

APPENDIX A: ADDITIONAL EXPERIMENTAL DETAILS

The trap fabrication process is very similar to that described in Ref. [5], with the exception of the addition of the rf tunnels. The trap is cooled to 15.8 K by a cryostat setup very similar to that in Ref. [5]. The typical heating rate is tens of q/s at 1 MHz.

On average, background gas collisions will cause one of the zones in the trap to be lost, either partially or fully, roughly every 30 minutes. In addition to ion loss, a process that is also detrimental to a circuit but occurs on a much faster timescale than loss is crystal reorders since we require the crystal order to always be Yb-Ba-Ba-Yb when performing 2Q gates. The mean time for any zone to reorder is several minutes.

For a desired number of total repetitions of a given circuit, we run a smaller number of repetitions in a “chunk.” After each chunk, we check the crystal order, number of ions, and magnetic field drift over the course of running the chunk. If these checks are not satisfactory, we disregard the data, reorder or reload as necessary, and repeat the chunk. The number of repetitions in a chunk is determined such that the probability of failing these checks is low ($\lesssim 25\%$). Empirically, we find that a chunk time of 30–60 seconds is optimal. When the system is operational, about 1/3 of the time is spent running circuits, with the remainder of the time spent on compiling circuits to the set of primitive operations to run the circuit, calibrations, pre- and postchunk checks, and reloading lost ions. We have a calibration tree that keeps the system calibrated and has recoveries for various calibrations failing. With this framework, we have greatly reduced the number of instances a human needs to intervene to resolve an issue.

APPENDIX B: DETAILS OF COMPONENT BENCHMARKS

1. Randomized benchmarking parameters and data

All component benchmarks (except transport 1Q RB) were repeated for each gate zone (DG01–DG04). Transport 1Q RB used all available 32 qubits with random rearrangements, so it is not zone specific. For each RB experiment, sequences were randomly and independently generated for each qubit (or pair of qubits for all tests with 2Q gates). The sequence lengths, repetitions, and shots used for the component benchmarks are shown in Table V.

For each RB experiment, the decay is fit to the standard first-order RB function [49] with a fixed asymptote,

$$p(\ell) = Ar^\ell + 1/2^N, \quad (\text{B1})$$

where $p(\ell)$ is the observed survival probability at length ℓ , A is the SPAM fit parameter, r is the depolarizing rate and N is the number of qubits. The reported error is the average infidelity, which is given by

TABLE V. Parameters used for component benchmarking testing.

Test	Lengths	Repetitions	Shots/Circuit
1Q RB	[2, 32, 128, 512]	30	100
2Q RB	[2, 16, 64, 128]	30	100
2Q SU(4) RB	[2, 8, 32, 64]	15	100
2Q parametrized RB	[4, 50, 100]	10	100
Transport 1Q RB	[2, 16, 32, 64]	64 ^a	200
Measurement crosstalk	[0, 100, 200, 300, 400, 500]	1	1000
Reset crosstalk	[0, 100, 200, 300, 400, 500]	1	1000
SPAM	...	2	5000

^aThe transport 1Q RB test is done with 32 qubits in parallel and repeated twice.

$$\epsilon = \frac{2^N - 1}{2^N} (1 - r). \quad (\text{B2})$$

This infidelity captures only the errors in the computational subspace and not leakage errors, which are measured with the leakage detection gadget.

The reset and measurement crosstalk decay functions are fit to functions derived from error models of their respective operations in Ref. [31] with the following equations:

$$p_M(\ell) = \frac{1}{3} (2 - A_M + (4A_M - 2)e^{-3r_M\ell}), \quad (\text{B3})$$

$$p_R(\ell) = 1 - A_R + \frac{1}{3} e^{-5r_R\ell} (2 + e^{3r_R\ell}) (2A_R - 1), \quad (\text{B4})$$

where $A_{M/R}$ are the SPAM fit parameters for each method and $r_{M/R}$ is the rate of measurement or reset crosstalk scattering. Each scattering rate is then converted to the average infidelity,

$$\epsilon_M = 5r_M/6, \quad (\text{B5})$$

$$\epsilon_R = 5r_R/3. \quad (\text{B6})$$

For all component measurements, a final combined estimate is obtained by performing the RB (or crosstalk) analysis on a combined data set between all measured qubits, which is reported in Table II. For example, in 1Q RB, the combined data set is obtained by treating each qubit measurement as a single sequence randomization and performing the RB fitting averaged over every qubit's random sequences. This process leads to an RB experiment with 8×40 random sequences for each length. For the crosstalk and SPAM measurements, the combined data set is obtained by adding all circuit output counts together. Zone-specific data for each component testing experiment are shown in Table VI. Decay plots for each component benchmark are shown in Figs. 4, 16–20. We observe that most zones have similar error measurements. In cases with

notable variance between zones, the difference is likely due to fit irregularities with small observed signals (such as 1Q leakage) or variations in beam alignments and trap conditions. However, these cases all occur with very small errors, and measurements are still below prespecified thresholds for operations.

2. Leakage detection gadget

The leakage rate r_L is defined as the rate that population leaves the computational subspace due to a process Λ based on Ref. [118],

$$r_L = \frac{1}{d_C} \text{Tr}(\mathbb{1}_L \Lambda[\mathbb{1}_C]), \quad (\text{B7})$$

where $\mathbb{1}_{L/C}$ is the identity operator on the leakage/computational subspace. The number of leakage detection events is fit to the model

$$p(\ell) = A(1 - r_L)^\ell, \quad (\text{B8})$$

as shown in Figs. 4(b) and 16(b). Gate errors in the leakage detection gadget can cause false-positive or false-negative detection events, but these only contribute to the parameter A , as they are independent of ℓ , similar to the SPAM parameter in RB.

3. The 2Q parametrized randomized benchmarking

To measure the average infidelity of $U_{ZZ}(\theta)$ as a function of θ , we use direct RB. In standard RB, the unitaries comprising the RB sequence are sampled from a unitary 2-design, such as the Clifford group or $SU(2^N)$. In contrast, direct RB samples unitaries from a set of native gates that generate the group [51, 119]. Under such circumstances, the survival probability will still approximate an exponential decay, with the decay parameter linearly related to the average fidelity [120].

Our direct RB circuits are constructed by repeatedly applying $U_{ZZ}(\theta)$ (for a fixed value of θ) interleaved with

TABLE VI. Component benchmarking results for the tests outlined above. All values are in terms of average infidelity and $\times 10^{-4}$. For 1Q RB, 1Q leakage rate, measurement and reset crosstalk, and SPAM, the brackets show the average infidelity for each side of the gate zone.

Test	DG01	DG02	DG03	DG04	Combined
1Q RB	[0.30(9), 0.21(7)]	[0.12(5), 0.21(7)]	[0.27(8), 0.32(9)]	[0.27(9), 0.30(9)]	0.25(3)
1Q leakage rate	[0.08(5), 0.04(4)]	[0.00(4), 0.002(4)]	[0.04(5), 0.03(4)]	[0.05(5), 0.02(4)]	0.04(2)
2Q RB	19(1)	17.5(9)	18(1)	17.9(9)	18.3(5)
2Q leakage rate	3.6(4)	3.9(4)	4.2(5)	4.0(5)	3.9(2)
2Q SU(4) RB	39(3)	45(3)	42(3)	38(2)	41(1)
Transport 1Q RB	2.2(3)
Measurement crosstalk	[0.02(1), 0.02(2)]	[0.24(3), 0.05(2)]	[0.01(1), 0.01(1)]	[0.001(5), 0.005(9)]	0.045(6)
Reset crosstalk	[0.002(7), 0.02(1)]	[0.12(3), 0.007(2)]	[0.04(1), 0.00(2)]	[0.02(1), 0.02(1)]	0.038(6)
SPAM	[15(4), 16(4)]	[19(4), 20(4)]	[19(4), 16(3)]	[8(3), 12(3)]	16(1)

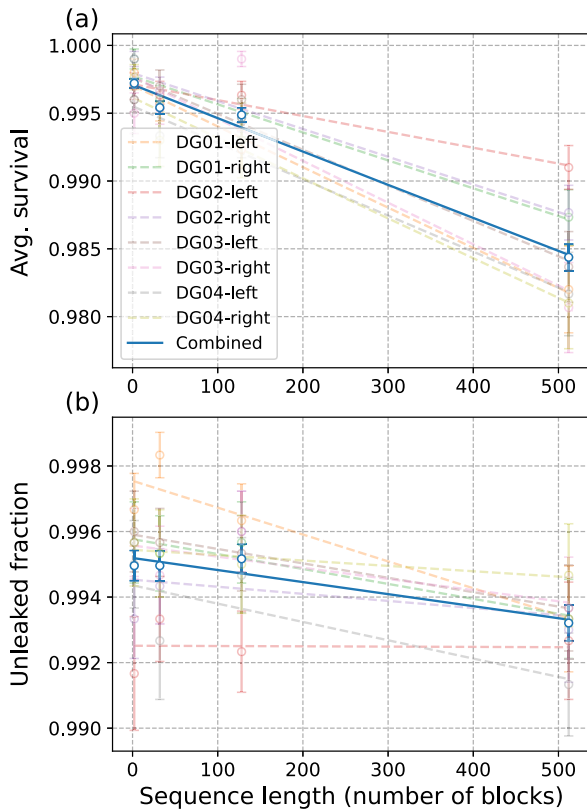


FIG. 16. The 1Q RB data with parameters given in Table V. (a) Decay of survival probability. (b) Decay of leaked fraction of shots.

Haar random $SU(2)$ gates on each qubit. For $\theta > 0$, this gate set generates $SU(4)$. The inversion unitary is applied by decomposing the resulting $SU(4)$ element into three $U_{ZZ}(\pi/2)$ 2Q gates using a standard decomposition [121]. A final random Pauli is applied to randomize the survival state. The decay curves are shown in Fig. 21, and the average fidelity is obtained by fitting to Eq. (B1).

In addition to positive values of $\theta \in \{(\pi/8), (\pi/4), (3\pi/8), (\pi/2)\}$, we also run a direct RB experiment with

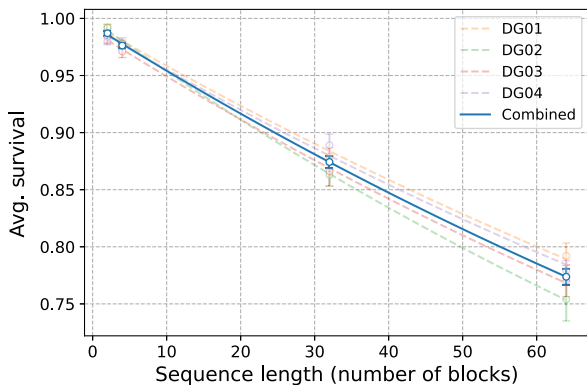


FIG. 17. The 2Q $SU(4)$ RB data with parameters given in Table V.

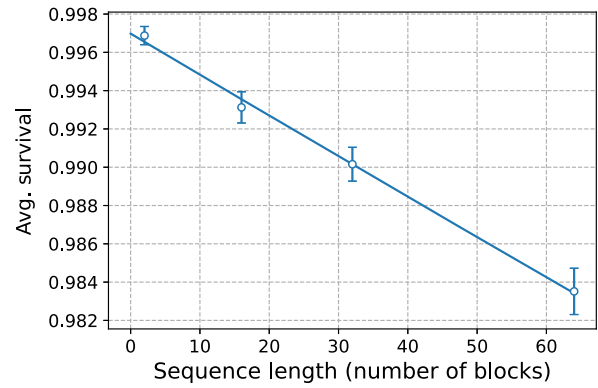


FIG. 18. Transport 1Q RB with parameters given in Table V.

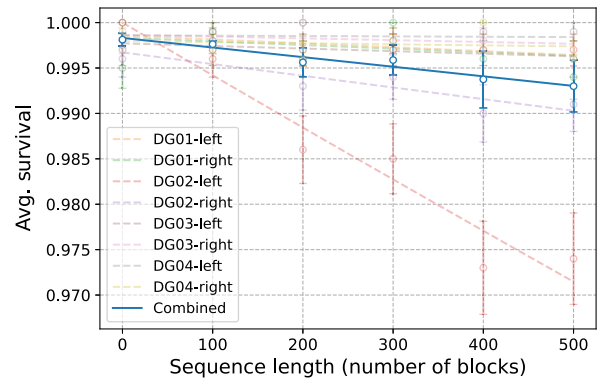


FIG. 19. Measurement crosstalk data with parameters given in Table V.

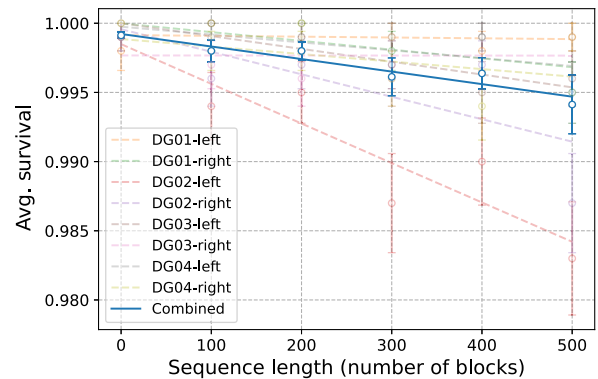


FIG. 20. Reset crosstalk data with parameters given in Table V.

θ very close to 0 (specifically 2×10^{-4} in order to force the 2Q gate to be performed), to measure the baseline error due to the MS wrapper pulses and memory error accumulated during the cooling pulses. However, for $\theta = 0$, the direct gate set reduces to $SU(2) \otimes SU(2)$, which no longer generates a unitary 2-design, and the RB theory leading to a single exponential decay no longer applies. To estimate the fidelity in this case, we use the fact that the action of $SU(2) \otimes SU(2)$ decomposes as a direct sum of four irreducible representations (irreps). (A good introduction to

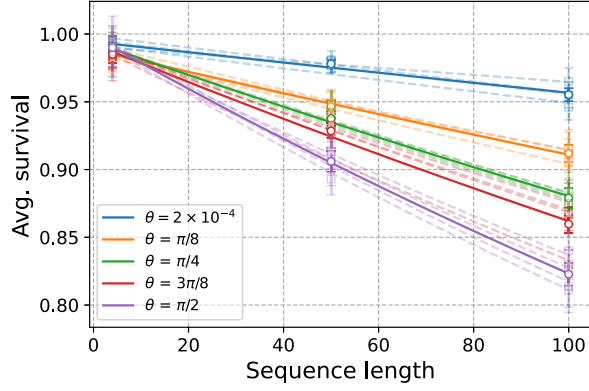


FIG. 21. Decay curves for direct RB of parametrized 2Q gates. The different sets of curves show data for θ ranging from 0 to $\pi/2$ in increments of $\pi/8$. Each experiment used sequence lengths $\ell = 4, 50$, and 100, with ten random circuits per sequence length. The circuits were run in parallel across the four gate zones, and all circuits were run in a random order. The dashed curves are the average over all zones, while the solid curve is the average over all zones. The decay curves for $\theta \in \{(\pi/8), (\pi/4), (3\pi/8), (\pi/2)\}$ are fit to the model $p(\ell) = Ar^\ell + 1/4$, and the average infidelity is computed using Eq. (B2). For $\theta = 0$, the average infidelity is computed using the procedure described in Appendix B 1. The average infidelity versus θ is shown in Fig. 3.

representation theory as it applies to RB is in Ref. [122].) The irreps are the span of the identity II , the spans of weight-1 Pauli operators on each qubit $\{IX, IY, IZ\}$ and $\{XI, YI, ZI\}$, and the span of the weight-2 Pauli operators. We let $\lambda \in \{II, IZ, ZI, ZZ\}$ label these irreps. If \mathcal{E} is the error channel for $U_{ZZ}(\theta \approx 0)$, then the twirl of \mathcal{E} over $SU(2) \otimes SU(2)$ is a linear combination of projectors onto these four irreps:

$$\mathcal{E}_T := \int_{g \in SU(2) \otimes SU(2)} d\mu(g) \phi(g) \mathcal{E} \phi(g)^{-1} = \sum_{\lambda} r_{\lambda} \Pi_{\lambda}, \quad (\text{B9})$$

where ϕ is the superoperator representation of $SU(2) \otimes SU(2)$, and Π_{λ} is the projector onto the irrep λ . The survival probability at sequence length ℓ is then given by

$$p(\ell) = \sum_{\lambda} A_{\lambda} r_{\lambda}^{\ell}. \quad (\text{B10})$$

We use the fact that $r_{II} = 1$ for trace-preserving maps, and the randomization in the survival state to fix $A_{II} = 1/4$. To reduce the number of exponential decay curves needed for the best fit, we assume qubit symmetry in the error channel, that is, $r_{IZ} = r_{ZI} = r_1$. Relabeling the SPAM parameters and defining $r_2 = r_{ZZ}$, the decay model is then given by

$$p(\ell) = A_1 r_1^{\ell} + A_2 r_2^{\ell} + \frac{1}{4}. \quad (\text{B11})$$

The entanglement (or process) fidelity is given by

$$\begin{aligned} F &= \frac{1}{16} \sum_{\lambda} \dim(\lambda) r_{\lambda} \\ &= \frac{1}{16} (1 + 6r_1 + 9r_2). \end{aligned} \quad (\text{B12})$$

The average infidelity is related to the entanglement fidelity,

$$\epsilon = \frac{d}{d+1} (1 - F), \quad (\text{B13})$$

for any d -dimensional trace-preserving error [123].

APPENDIX C: DETAILS OF SYSTEM-LEVEL BENCHMARKS

1. Mirror benchmarking

Table VII lists the survival probabilities, decay parameter, and effective 2Q average infidelity $\epsilon_{\text{eff}}^{2\text{Q}}$ for the MB experiment. The average survival probability as a function of sequence length ℓ is fit to the model,

$$p(\ell) = Au^{\ell-1}. \quad (\text{C1})$$

Let \mathcal{E} be an N -qubit error channel. Let $\{P_i\}_i$ be the N -qubit Pauli operators with $P_0 = \mathbb{I}$. The i th Pauli fidelity of \mathcal{E} is defined as

$$f_i = \frac{1}{2^N} \text{Tr}(P_i \mathcal{E}(P_i)). \quad (\text{C2})$$

By applying Pauli randomization to the TQ gates in the MB circuits, the error channel for each circuit layer is assumed to be a stochastic Pauli channel [62]. Assuming a constant stochastic Pauli error channel \mathcal{E} per circuit layer, it was shown in Ref. [57] that the decay parameter u is equal to the mean square of the nonidentity Pauli fidelities:

$$u = \frac{1}{2^N - 1} \sum_{i>0} f_i^2. \quad (\text{C3})$$

For a constant depolarizing error channel on each 2Q gate, u is given by an analytic formula [Eq. (C4) in Ref. [57]].

TABLE VII. MB survival probabilities, fit parameter (u), and effective 2Q gate average infidelity ($\epsilon_{\text{eff}}^{2\text{Q}}$).

Sequence length	$N = 20$	$N = 26$	$N = 32$
$\ell = 2$	0.88(1)	0.84(2)	0.82(2)
$\ell = 4$	0.77(3)	0.70(2)	0.64(2)
$\ell = 6$	0.66(2)	0.57(2)	
$\ell = 7$			0.51(3)
$\ell = 10$	0.51(3)	0.39(2)	0.35(3)
u	0.934(6)	0.908(7)	0.902(7)
$\epsilon_{\text{eff}}^{2\text{Q}}$	0.0027(3)	0.0030(2)	0.0026(2)

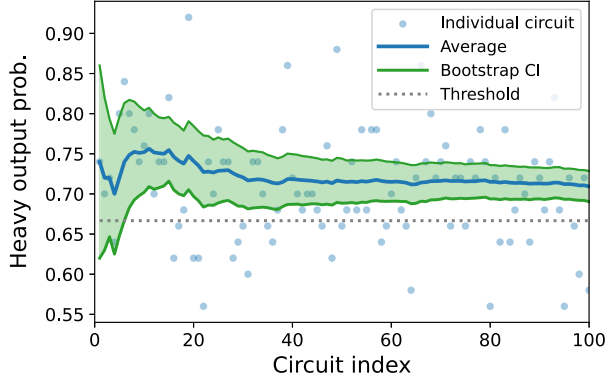


FIG. 22. Quantum volume $QV = 2^{15}$ measurement on H2. The average and two-sigma confidence interval of the heavy-output frequency are plotted as a function of the circuit index. The green-shaded region shows the two-sigma confidence interval from the semiparametric bootstrap method.

After best fitting the experimental decay curves to obtain u , this formula is used to extract $\epsilon_{\text{eff}}^{2Q}$.

2. Quantum volume

In addition to the $QV = 2^{16}$ data set presented in the main text, we also ran several smaller QV tests. In Fig. 22, we show the next largest $QV = 2^{15}$ test. This test was run with 100 random circuits, each with 50 shots and containing an average of 243 parametrized 2Q gates. The measured heavy output probability was 70.9%, with a two-sigma lower bound of 69.1% calculated by the semiparametric bootstrap resampling method.

To infer an effective 2Q error from QV data, first we convert the measured heavy-output probability to a circuit fidelity based on Eq. (13) in Ref. [59]. We then scale this based on the SPAM error and average number of 2Q gates as shown in Eq. (C6).

3. Random circuit sampling

The definition of the linear cross-entropy benchmarking fidelity is

$$\mathcal{F}_{\text{XEB}} = 2^N \langle P(x_i) \rangle - 1, \quad (\text{C4})$$

where $P(x_i)$ is the probability of measuring the output bitstring x_i in the ideal output distribution, and the expectation value is taken over the empirically measured bitstrings. The linear cross-entropy fidelity is a measure of the correlation between the empirical output distribution and the ideal output distribution. Consequently, this requires exact classical simulation of the random circuits, which is a major obstacle to scalability of the benchmark. The uncertainty on the linear cross-entropy fidelity for each circuit can be obtained from Eq. (C4) by combining the variance estimator for $P(x_i)$ with the standard uncertainty-on-the-mean formula, namely,

$$\text{var}(\mathcal{F}_{\text{XEB}}) = \frac{2^{2N} \text{var}(P_i)}{N_{\text{shots}}}. \quad (\text{C5})$$

In Fig. 8, we report a fit for the linear cross-entropy benchmarking fidelity on H2 as a function of N . This fit was obtained by the following procedure. At each fixed N , a representative random circuit was generated and compiled with `pytket` to obtain an expected number of 2Q operations. We note that the final number of 2Q U_{ZZ} operations in each circuit is equal to the number of `fSim`[($\pi/2$), ($\pi/6$)] gates in the original uncompiled circuit. The overall model for the linear cross-entropy fidelity is then

$$\mathcal{F}_{\text{XEB}} = (F_{2Q})^{\#2Q} \times (1 - \epsilon_{\text{SPAM}})^N. \quad (\text{C6})$$

Here, F_{2Q} represents the effective entanglement (or process) fidelity of two-qubit operations, while ϵ_{SPAM} is the SPAM error as measured by component benchmarking. The conversion between entanglement fidelities and average infidelities as obtained via component benchmarking in Table VI is given in Eq. (B13).

Holding fixed the average SPAM error of $1.6(1) \times 10^{-3}$ from Table VI, the model (C6) was fit to the H2 data, obtaining a best-fit value of $1 - F_{2Q} = 2.4(2) \times 10^{-3}$. In terms of average infidelity, this corresponds to $\epsilon_{2Q} = 1.9(2) \times 10^{-3}$.

APPENDIX D: DETAILS OF APPLICATION BENCHMARKS

1. Trotter steps of Hamiltonian simulation experiment

Here, we provide details on the Trotter steps used in the experiment. The Trotter steps r are determined by relative convergence with tolerance 0.0025; i.e., we choose a cutoff r such that for $r' \geq r$, neighboring steps are within the threshold $|\langle X \rangle_{r'+1} - \langle X \rangle_{r'}| \leq 0.0025$, where $\langle X \rangle_r$ is the X expectation value after r steps of propagation in a noiseless circuit. For this purpose, we compute each $\langle X \rangle_r$ exactly via a discrete-time Jordan-Wigner transformation in the Heisenberg picture [88]. We check that this 0.0025 relative error tolerance provides an absolute Trotter error tolerance of around 1% in $|\langle X \rangle_r - \langle X \rangle|$ for the times we simulate, which is at the scale of the expected statistical fluctuation of about 1% in the experiment. We choose these values because further improvements from lowering the Trotter error would not be reliably observable even if the circuit were completely noiseless, though we did not choose them in a noise-aware fashion (further lowering of the number of Trotter steps used may well give further improvements given the presence of gate errors). The steps and the corresponding Trotter errors are shown in Figs. 23(a) and 23(b). The difference between the experimental data and the exact value is shown in Fig. 23(c).

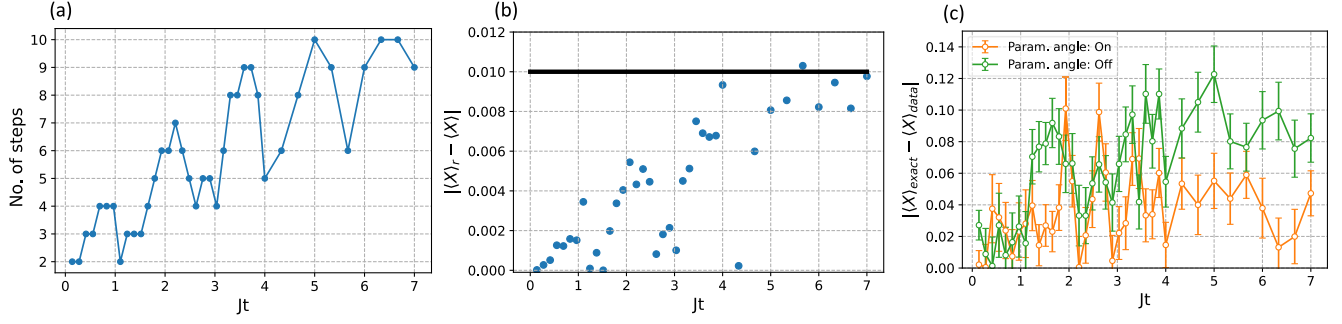


FIG. 23. (a) Number of Trotter steps used at each simulation time. (b) Absolute Trotter error $|\langle X \rangle_r - \langle X \rangle|$ at each time, where $\langle X \rangle_r$ is the expectation value using r Trotter steps under a noiseless circuit and $\langle X \rangle$ is the exact value at that time. (c) Data value relative to the exact value $|\langle X \rangle_{\text{exact}} - \langle X \rangle_{\text{data}}|$ at each simulation time. The data error bars are included to reflect the signal-to-noise ratio.

2. QAOA optimization landscape

In Figs. 12 and 13 in the main text, uncertainties were computed on the expectation value of the energy $\langle H_C \rangle$ as evaluated on H2 (blue points). These uncertainties were computed by bootstrap resampling via the reverse-percentile method [124], and they quantify the uncertainty due to shot noise but not physical noise sources on the machine. We emphasize that the different data points in Figs. 12 and 13 are evaluated at different values of the parameters β and γ .

In Fig. 24, we display the full optimization trace on the energy landscape for the $N = 130$, $p = 1$ MaxCut QAOA instance described in the main text, further justifying that the classical optimizer successfully converges to the minimum value of the energy.

3. Details of HoloQUADS experiment

Here, we provide additional details on the holographic quantum dynamics experiments performed on H2. We consider an initial matrix-product state of the form

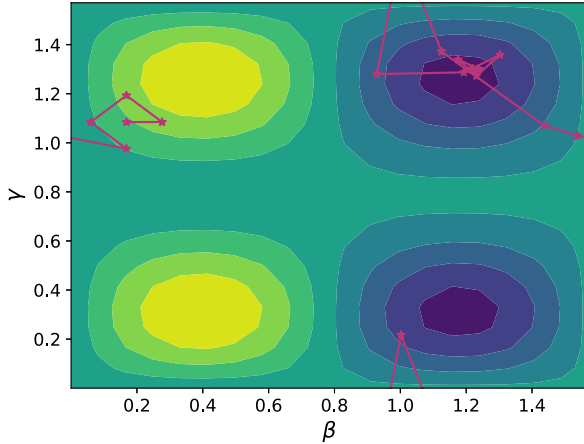


FIG. 24. Energy landscape of an $N = 130$, $p = 1$ MaxCut QAOA problem. The optimization trace (pink) starts near the maximum in the top left and eventually converges into the potential well in the top right. Individual points at which circuits were evaluated are marked with stars. Note that the landscape is periodic, so the trajectory wraps around the sides of this plot.

$$|\psi_0\rangle = \sum_{\sigma_1, \sigma_2, \dots \in \{\uparrow, \downarrow\}} e^{T \mathcal{N}(\sigma_1, \sigma_2) \mathcal{N}(\sigma_3, \sigma_4) \dots} |\sigma_1 \sigma_2 \sigma_3 \sigma_4 \dots\rangle, \quad (\text{D1})$$

with the tensor $\mathcal{N}_{i,j}^{(\sigma, \sigma')} = \langle j | \otimes \langle \sigma' | W | i \rangle \otimes | \sigma \rangle$ specified by the unitary $W = \exp[-i(K_x XX + K_y YY + K_z ZZ)]$ with $(K_x, K_y, K_z) = (0.3, 0.5, 1.25)$; this bond-dimension $\chi = 2$ matrix-product state, previously studied in Refs. [107,108], is prepared by applying gates between the physical qubits and $n_b = \log_2 \chi = 1$ ancilla “bond” qubits. We time-evolve this state using the SDKI model [106,111], which can be formulated as a dual-unitary circuit using 2Q gates,

$$U = (u_+ \otimes u_-) e^{-i\frac{\pi}{4}(XX+YY)} (v_- \otimes v_+). \quad (\text{D2})$$

Here, the 1Q gates are given by $u_+ = e^{-ihZ} e^{i\frac{\pi}{4}X} e^{-i\frac{\pi}{4}Y}$, $u_- = e^{i\frac{\pi}{4}X} e^{-i\frac{\pi}{4}Y}$, $v_- = e^{i\frac{\pi}{4}Y} e^{-ihZ}$, and $v_+ = e^{i\frac{\pi}{4}Y}$. We use $h = 0.05$, which is close to but not exactly at the integrable $h = 0$ point where the SDKI model displays no decay of correlation functions.

Qubit-reuse techniques involving MCMR [38,108] are used to construct the circuit in a way that maximizes the number of parallel 2Q gates. The leakage detection gadget (Fig. 5) is used to discard results where the bond qubit is measured to have leaked (measuring 2%, 2%, 5%, 7% bond qubit leakage for $t = 0, 8, 16, 24$). We also use a circuit identity to construct each gate U with a single parametrized $U_{ZZ}(\theta)$ gate and one physical SWAP [108].

- [1] C. D. Bruzewicz, J. Chiaverini, R. McConnell, and J. M. Sage, *Trapped-ion quantum computing: Progress and challenges*, *Appl. Phys. Rev.* **6**, 021314 (2019).
- [2] L. Henriët, L. Beguin, A. Signoles, T. Lahaye, A. Browaeys, G.-O. Reymond, and C. Jurczak, *Quantum computing with neutral atoms*, *Quantum* **4**, 327 (2020).
- [3] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, *Superconducting qubits: Current state of play*, *Annu. Rev. Condens. Matter Phys.* **11**, 369 (2020).

- [4] D. Franke, J. Clarke, L. Vandersypen, and M. Veldhorst, *Rent's rule and extensibility in quantum computing*, *Microprocess Microsyst.* **67**, 1 (2019).
- [5] 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 (London)* **592**, 209 (2021).
- [6] D. J. Wineland, C. Monroe, W. M. Itano, D. Leibfried, B. E. King, and D. M. Meekhof, *Experimental issues in coherent quantum-state manipulation of trapped atomic ions*, *J. Res. Natl. Inst. Stand. Technol.* **103**, 259 (1998).
- [7] C. Monroe and J. Kim, *Scaling the ion trap quantum processor*, *Science* **339**, 1164 (2013).
- [8] J. I. Cirac and P. Zoller, *Quantum computations with cold trapped ions*, *Phys. Rev. Lett.* **74**, 4091 (1995).
- [9] 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).
- [10] J. Home, D. Hanneke, J. Jost, J. Amini, D. Leibfried, and D. Wineland, *Complete methods set for scalable ion trap quantum information processing*, *Science* **325**, 1227 (2009).
- [11] V. Kaushal, B. Lekitsch, A. Stahl, J. Hilder, D. Pijn, C. Schmiegelow, A. Bermudez, M. Müller, F. Schmidt-Kaler, and U. Poschinger, *Shuttling-based trapped-ion quantum information processing*, *AVS Quantum Sci.* **2**, 014101 (2020).
- [12] S. Seidelin, J. Chiaverini, R. Reichle, J. J. Bollinger, D. Leibfried, J. Britton, J. H. Wesenberg, R. B. Blakestad, R. J. Epstein, D. B. Hume, W. M. Itano, J. D. Jost, C. Langer, R. Ozeri, N. Shiga, and D. J. Wineland, *Microfabricated surface-electrode ion trap for scalable quantum information processing*, *Phys. Rev. Lett.* **96**, 253003 (2006).
- [13] P. Maunz and L. Wilhelm, *High optical access trap 2.0*, SAND2016-0796R [10.2172/1237003](https://arxiv.org/abs/10.2172/1237003) (2016).
- [14] 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**, 1 (2021).
- [15] C. Ryan-Anderson *et al.*, *Implementing fault-tolerant entangling gates on the five-qubit code and the color code*, [arXiv:2208.01863](https://arxiv.org/abs/2208.01863).
- [16] A. N. Habermann, *Parallel neighbor-sort (or the glory of the induction principle)* (2000) [10.1184/R1/6608258.v1](https://arxiv.org/abs/10.1184/R1/6608258.v1).
- [17] M. Malinowski, D. T. C. Allcock, and C. J. Ballance, *How to wire a 1000-qubit trapped ion quantum computer*, [arXiv:2305.12773](https://arxiv.org/abs/2305.12773).
- [18] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, *Surface codes: Towards practical large-scale quantum computation*, *Phys. Rev. A* **86**, 032324 (2012).
- [19] C. Ryan-Anderson *et al.*, *Realization of real-time fault-tolerant quantum error correction*, *Phys. Rev. X* **11**, 041058 (2021).
- [20] W. K. Hensinger, S. Olmschenk, D. Stick, D. Hucul, M. Yeo, M. Acton, L. Deslauriers, C. Monroe, and J. Rabchuk, *T-junction ion trap array for two-dimensional ion shuttling, storage, and manipulation*, *Appl. Phys. Lett.* **88**, 034101 (2006).
- [21] W. C. Burton, B. Estey, I. M. Hoffman, A. R. Perry, C. Volin, and G. Price, *Transport of multispecies ion crystals through a junction in a radio-frequency Paul trap*, *Phys. Rev. Lett.* **130**, 173202 (2023).
- [22] K. R. Brown, J. Kim, and C. Monroe, *Co-designing a scalable quantum computer with trapped atomic ions*, *npj Quantum Inf.* **2**, 16034 (2016).
- [23] M. Akhtar, F. Bonus, F. R. Lebrun-Gallagher, N. I. Johnson, M. Siegle-Brown, S. Hong, S. J. Hile, S. A. Kulmiya, S. Weidt, and W. K. Hensinger, *A high-fidelity quantum matter-link between ion-trap microchip modules*, *Nat. Commun.* **14**, 531 (2023).
- [24] R. J. Niffenegger, J. Stuart, C. Sorace-Agaskar, D. Kharas, S. Bramhavar, C. D. Bruzewicz, W. Loh, R. T. Maxson, R. McConnell, D. Reens, G. N. West, J. M. Sage, and J. Chiaverini, *Integrated multi-wavelength control of an ion qubit*, *Nature (London)* **586**, 538 (2020).
- [25] A. J. Menssen, A. Hermans, I. Christen, T. Propson, C. Li, A. J. Leenheer, M. Zimmermann, M. Dong, H. Larocque, H. Raniwala, G. Gilbert, M. Eichenfield, and D. R. Englund, *Scalable photonic integrated circuits for programmable control of atomic systems*, [arXiv:2210.03100](https://arxiv.org/abs/2210.03100).
- [26] S. L. Todaro, V. B. Verma, K. C. McCormick, D. T. C. Allcock, R. P. Mirin, D. J. Wineland, S. W. Nam, A. C. Wilson, D. Leibfried, and D. H. Slichter, *State readout of a trapped ion qubit using a trap-integrated superconducting photon detector*, *Phys. Rev. Lett.* **126**, 010501 (2021).
- [27] C. D. Bruzewicz, R. McConnell, J. Chiaverini, and J. M. Sage, *Scalable loading of a two-dimensional trapped-ion array*, *Nat. Commun.* **7**, 13005 (2016).
- [28] J. M. Amini, H. Uys, J. H. Wesenberg, S. Seidelin, J. Britton, J. J. Bollinger, D. Leibfried, C. Ospelkaus, A. P. VanDevender, and D. J. Wineland, *Toward scalable ion traps for quantum information processing*, *New J. Phys.* **12**, 033031 (2010).
- [29] B. Tabakov, F. Benito, M. Blain, C. R. Clark, S. Clark, R. A. Haltli, P. Maunz, J. D. Sterk, C. Tigges, and D. Stick, *Assembling a ring-shaped crystal in a microfabricated surface ion trap*, *Phys. Rev. Appl.* **4**, 031001(R) (2015).
- [30] H.-K. Li, E. Urban, C. Noel, A. Chuang, Y. Xia, A. Ransford, B. Hemmerling, Y. Wang, T. Li, H. Häffner, and X. Zhang, *Realization of translational symmetry in trapped cold ion rings*, *Phys. Rev. Lett.* **118**, 053001 (2017).
- [31] J. P. Gaebler, C. H. Baldwin, S. A. Moses, J. M. Dreiling, C. Figgatt, M. Foss-Feig, D. Hayes, and J. M. Pino, *Suppression of midcircuit measurement crosstalk errors with micromotion*, *Phys. Rev. A* **104**, 062440 (2021).
- [32] J. Johansen, B. Estey, M. Rowe, and A. Ransford, *Fast loading of a trapped ion quantum computer using a 2d magneto-optical trap*, in *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)* (2022), pp. 299–303, [10.1109/QCE53715.2022.00050](https://arxiv.org/abs/10.1109/QCE53715.2022.00050).
- [33] S. Olmschenk, K. C. Younge, D. L. Moehring, D. N. Matsukevich, P. Maunz, and C. Monroe, *Manipulation and detection of a trapped Yb⁺ hyperfine qubit*, *Phys. Rev. A* **76**, 052314 (2007).

- [34] A. Sørensen and K. Mølmer, *Entanglement and quantum computation with ions in thermal motion*, *Phys. Rev. A* **62**, 022311 (2000).
- [35] 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* **7**, S371 (2005).
- [36] C.N. Self, M. Benedetti, and D. Amaro, *Protecting expressive circuits with a quantum error detection code*, [arXiv:2211.06703](https://arxiv.org/abs/2211.06703).
- [37] E. Chertkov, Z. Cheng, A. C. Potter, S. Gopalakrishnan, T. M. Gatterman, J. A. Gerber, K. Gilmore, D. Gresh, A. Hall, A. Hankin, M. Matheny, T. Mengle, D. Hayes, B. Neyenhuis, R. Stutz, and M. Foss-Feig, *Characterizing a non-equilibrium phase transition on a quantum computer*, [arXiv:2209.12889](https://arxiv.org/abs/2209.12889).
- [38] M. DeCross, E. Chertkov, M. Kohagen, and M. Foss-Feig, *Qubit-reuse compilation with mid-circuit measurement and reset*, [arXiv:2210.08039](https://arxiv.org/abs/2210.08039).
- [39] J. P. Gaebler, A. M. Meier, T. R. Tan, R. Bowler, Y. Lin, D. Hanneke, J. D. Jost, J. P. Home, E. Knill, D. Leibfried, and D. J. Wineland, *Randomized benchmarking of multiqubit gates*, *Phys. Rev. Lett.* **108**, 260503 (2012).
- [40] K. Sugiyama, S. Tagawa, and M. Toda, *Methods for visual understanding of hierarchical system structures*, *IEEE Trans. Syst. Man Cybernetics* **11**, 109 (1981).
- [41] C. Monroe, D. M. Meekhof, B. E. King, S. R. Jefferts, W. M. Itano, D. J. Wineland, and P. Gould, *Resolved-sideband Raman cooling of a bound atom to the 3D zero-point energy*, *Phys. Rev. Lett.* **75**, 4011 (1995).
- [42] M. D. Barrett, B. DeMarco, T. Schaetz, V. Meyer, D. Leibfried, J. Britton, J. Chiaverini, W. M. Itano, B. Jelenković, J. D. Jost, C. Langer, T. Rosenband, and D. J. Wineland, *Sympathetic cooling of $^9\text{Be}^+$ and $^{24}\text{Mg}^+$ for quantum logic*, *Phys. Rev. A* **68**, 042302 (2003).
- [43] T. P. Harty, *High-fidelity Microwave-Driven Quantum Logic in Intermediate-Field $^{43}\text{Ca}^+$* , Ph.D. thesis, University of Oxford (2013), <https://www.semanticscholar.org/paper/High-fidelity-microwave-driven-quantum-logic-in-Harty/b3b9be5e816994382936382e3afa754e2ef536ae>.
- [44] A. W. Cross, L. S. Bishop, J. A. Smolin, and J. M. Gambetta, *Open quantum assembly language*, [arXiv:1707.03429](https://arxiv.org/abs/1707.03429).
- [45] QIR Alliance, QIR Specification, see also <https://qir-alliance.org>.
- [46] A. Haas, A. Rossberg, D. L. Schuff, B. L. Titzer, M. Holman, D. Gohman, L. Wagner, A. Zakai, and J. Bastien, *Bringing the web up to speed with webassembly*, in *Proceedings of the 38th ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI 2017* (Association for Computing Machinery, New York, NY, 2017), p. 185–200, 10.1145/3062341.3062363.
- [47] M. Acton, K.-A. Brickman, P. C. Haljan, P. J. Lee, L. Deslauriers, and C. Monroe, *Near-perfect simultaneous measurement of a qubit register*, *Quantum Inf. Comput.* **6**, 465 (2006).
- [48] R. Noek, G. Vrijsen, D. Gaultney, E. Mount, T. Kim, P. Maunz, and J. Kim, *High speed, high fidelity detection of an atomic hyperfine qubit*, *Opt. Lett.* **38**, 4735 (2013).
- [49] E. Magesan, J. M. Gambetta, and J. Emerson, *Scalable and robust randomized benchmarking of quantum processes*, *Phys. Rev. Lett.* **106**, 180504 (2011).
- [50] R. Harper, I. Hincks, C. Ferrie, S. T. Flammia, and J. J. Wallman, *Statistical analysis of randomized benchmarking*, *Phys. Rev. A* **99**, 052350 (2019).
- [51] T. J. Proctor, A. Carignan-Dugas, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young, *Direct randomized benchmarking for multiqubit devices*, *Phys. Rev. Lett.* **123**, 030503 (2019).
- [52] A. M. Meier, *Randomized Benchmarking of Clifford Operators*, Ph.D. thesis, University of Colorado (2006).
- [53] R. Ozeri, W. M. Itano, R. B. Blakestad, J. Britton, J. Chiaverini, J. D. Jost, C. Langer, D. Leibfried, R. Reichle, S. Seidelin, J. H. Wesenberg, and D. J. Wineland, *Errors in trapped-ion quantum gates due to spontaneous photon scattering*, *Phys. Rev. A* **75**, 042329 (2007).
- [54] M. J. Boguslawski, Z. J. Wall, S. R. Vizvary, I. D. Moore, M. Bareian, D. T. C. Allcock, D. J. Wineland, E. R. Hudson, and W. C. Campbell, *Raman scattering errors in stimulated-Raman-induced logic gates in $^{133}\text{Ba}^+$* , *Phys. Rev. Lett.* **131**, 063001 (2023).
- [55] R. Stricker, D. Vodola, A. Erhard, L. Postler, M. Meth, M. Ringbauer, P. Schindler, T. Monz, M. Müller, and R. Blatt, *Experimental deterministic correction of qubit loss*, *Nature (London)* **585**, 207 (2020).
- [56] T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, *Measuring the capabilities of quantum computers*, *Nat. Phys.* **18**, 75 (2022).
- [57] K. Mayer, A. Hall, T. Gatterman, S. K. Halit, K. Lee, J. Bohnet, D. Gresh, A. Hankin, K. Gilmore, and J. Gaebler, *Theory of mirror benchmarking and demonstration on a quantum computer*, [arXiv:2108.10431](https://arxiv.org/abs/2108.10431).
- [58] 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).
- [59] 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).
- [60] F. Arute *et al.*, *Quantum supremacy using a programmable superconducting processor*, *Nature (London)* **574**, 505 (2019).
- [61] T. Proctor, S. Seritan, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young, *Scalable randomized benchmarking of quantum computers using mirror circuits*, *Phys. Rev. Lett.* **129**, 150502 (2022).
- [62] J. J. Wallman and J. Emerson, *Noise tailoring for scalable quantum computation via randomized compiling*, *Phys. Rev. A* **94**, 052325 (2016).
- [63] E. Pelofske, A. Bärttschi, and S. Eidenbenz, *Quantum volume in practice: What users can expect from NISQ devices*, *IEEE Trans. Quantum Eng.* **3**, 1 (2022).
- [64] Quantum volume reaches 5 digits for the first time: 5 perspectives on what it means for quantum computing, <https://www.quantinuum.com/news/quantum-volume-reaches-5-digits-for-the-first-time-5-perspectives-on-what-it-means-for-quantum-computing>.

- [65] D. Aharonov, X. Gao, Z. Landau, Y. Liu, and U. Vazirani, *A polynomial-time classical algorithm for noisy random circuit sampling*, in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC 2023)*, 10.1145/3564246.3585234.
- [66] Y. Wu *et al.*, *Strong quantum computational advantage using a superconducting quantum processor*, *Phys. Rev. Lett.* **127**, 180501 (2021).
- [67] Q. Zhu *et al.*, *Quantum computational advantage via 60-qubit 24-cycle random circuit sampling*, *Sci. Bull.* **67**, 240 (2022).
- [68] 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).
- [69] S. Aaronson and L. Chen, *Complexity-theoretic foundations of quantum supremacy experiments*, [arXiv:1612.05903](https://arxiv.org/abs/1612.05903).
- [70] D. M. Greenberger, M. A. Horne, and A. Zeilinger, *Going beyond Bell's theorem*, in *Bell's Theorem, Quantum Theory and Conceptions of the Universe*, edited by M. Kafatos (Springer Netherlands, Dordrecht, 1989), pp. 69–72, 10.1007/978-94-017-0849-4_10.
- [71] I. Pogorelov *et al.*, *Compact ion-trap quantum computing demonstrator*, *PRX Quantum* **2**, 020343 (2021).
- [72] A. Omran *et al.*, *Generation and manipulation of Schrödinger cat states in Rydberg atom arrays*, *Science* **365**, 570 (2019).
- [73] C. Song *et al.*, *Generation of multicomponent atomic Schrödinger cat states of up to 20 qubits*, *Science* **365**, 574 (2019).
- [74] P. Thomas, L. Ruscio, O. Morin, and G. Rempe, *Efficient generation of entangled multiphoton graph states from a single atom*, *Nature (London)* **608**, 677 (2022).
- [75] G. J. Mooney, G. A. L. White, C. D. Hill, and L. C. L. Hollenberg, *Generation and verification of 27-qubit Greenberger-Horne-Zeilinger states in a superconducting quantum computer*, *J. Phys. Commun.* **5**, 095004 (2021).
- [76] D. Cruz, R. Fournier, F. Gremion, A. Jeannerot, K. Komagata, T. Tosić, J. Thiesbrummel, C. L. Chan, N. Macris, M.-A. Dupertuis, and C. Javerzac-Galy, *Efficient quantum algorithms for GHZ and W states, and implementation on the IBM quantum computer*, *Adv. Quantum Tech.* **2**, 1900015 (2019).
- [77] H. J. Briegel and R. Raussendorf, *Persistent entanglement in arrays of interacting particles*, *Phys. Rev. Lett.* **86**, 910 (2001).
- [78] We note that while the latter circuit is constant depth, physically executing it in a constant amount of time requires a number of gate zones that scales with the number of qubits. With our current configuration of four gate zones and 32 qubits, at least three rounds of parallel gates are required.
- [79] T.-C. Lu, L. A. Lessa, I. H. Kim, and T. H. Hsieh, *Measurement as a shortcut to long-range entangled quantum matter*, *PRX Quantum* **3**, 040337 (2022).
- [80] M. Foss-Feig *et al.*, *Experimental demonstration of the advantage of adaptive quantum circuits*, [arXiv:2302.03029](https://arxiv.org/abs/2302.03029).
- [81] M. Iqbal *et al.*, *Topological order from measurements and feed-forward on a trapped ion quantum computer*, [arXiv:2302.01917](https://arxiv.org/abs/2302.01917).
- [82] O. Gühne, C.-Y. Lu, W.-B. Gao, and J.-W. Pan, *Toolbox for entanglement detection and fidelity estimation*, *Phys. Rev. A* **76**, 030305(R) (2007).
- [83] O. Gühne and M. Seevinck, *Separability criteria for genuine multipartite entanglement*, *New J. Phys.* **12**, 053002 (2010).
- [84] R. P. Feynman, *Simulating physics with computers*, *Int. J. Theor. Phys.* **21**, 467 (1982).
- [85] D. Wecker, M. B. Hastings, N. Wiebe, B. K. Clark, C. Nayak, and M. Troyer, *Solving strongly correlated electron models on a quantum computer*, *Phys. Rev. A* **92**, 062318 (2015).
- [86] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, *Toward the first quantum simulation with quantum speedup*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9456 (2018).
- [87] S. Lloyd, *Universal quantum simulators*, *Science* **273**, 1073 (1996).
- [88] B. M. Terhal and D. P. DiVincenzo, *Classical simulation of noninteracting-fermion quantum circuits*, *Phys. Rev. A* **65**, 032325 (2002).
- [89] P. Calabrese, F. H. L. Essler, and M. Fagotti, *Quantum quench in the transverse field Ising chain: I. Time evolution of order parameter correlators*, *J. Stat. Mech.* (2012) P07016.
- [90] E. Farhi, J. Goldstone, and S. Gutmann, *A quantum approximate optimization algorithm*, [arXiv:1411.4028](https://arxiv.org/abs/1411.4028).
- [91] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, *Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices*, *Phys. Rev. X* **10**, 021067 (2020).
- [92] M. P. Harrigan *et al.*, *Quantum approximate optimization of non-planar graph problems on a planar superconducting processor*, *Nat. Phys.* **17**, 332 (2021).
- [93] M. J. D. Powell, *The BOBYQA algorithm for bound constrained optimization without derivatives*, Technical Report, Cambridge University, Department of Applied Mathematics and Theoretical Physics (2009).
- [94] C. Cartis, J. Fiala, B. Marteau, and L. Roberts, *Improving the flexibility and robustness of model-based derivative-free optimization solvers*, *ACM Trans. Math. Softw.* **45**, 32:1 (2019).
- [95] J. Gray, *quimb: A PYTHON library for quantum information and many-body calculations*, *J. Open Source Software* **3**, 819 (2018).
- [96] T. Head *et al.*, *scikit-optimize/scikit-optimize: v0.5.2* (2018), 10.5281/zenodo.1157319.
- [97] *Gurobi Optimization, LLC, Gurobi Optimizer Reference Manual* (2022).
- [98] D. Gottesman, *Stabilizer codes and quantum error correction*, Ph.D. thesis, California Institute of Technology (1997).
- [99] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, *Topological quantum memory*, *J. Math. Phys. (N.Y.)* **43**, 4452 (2002).
- [100] O. Higgott, *PyMatching: A Python package for decoding quantum codes with minimum-weight perfect matching*, *ACM Trans. Quantum Comput.* **3**, 1 (2022).
- [101] Z. Chen *et al.*, *Exponential suppression of bit or phase errors with cyclic error correction*, *Nature (London)* **595**, 383 (2021).

- [102] R. Acharya *et al.*, *Suppressing quantum errors by scaling a surface code logical qubit*, *Nature (London)* **614**, 676 (2023).
- [103] B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans* (SIAM, Philadelphia, 1982), 10.1137/1.9781611970319.
- [104] E. Eypasch, R. Lefering, C. Kum, and H. Troidl, *Probability of adverse events that have not yet occurred: A statistical reminder*, *Br. Med. J.* **311**, 619 (1995).
- [105] M. Foss-Feig, D. Hayes, J. M. Dreiling, C. Figgatt, J. P. Gaebler, S. A. Moses, J. M. Pino, and A. C. Potter, *Holographic quantum algorithms for simulating correlated spin systems*, *Phys. Rev. Res.* **3**, 033002 (2021).
- [106] B. Bertini, P. Kos, and T. Prosen, *Exact spectral form factor in a minimal model of many-body quantum chaos*, *Phys. Rev. Lett.* **121**, 264101 (2018).
- [107] B. Bertini, P. Kos, and T. Prosen, *Exact correlation functions for dual-unitary lattice models in 1 + 1 dimensions*, *Phys. Rev. Lett.* **123**, 210601 (2019).
- [108] E. Chertkov, J. Bohnet, D. Francois, J. Gaebler, D. Gresh, A. Hankin, K. Lee, D. Hayes, B. Neyenhuis, R. Stutz, A. C. Potter, and M. Foss-Feig, *Holographic dynamics simulations with a trapped-ion quantum computer*, *Nat. Phys.* **18**, 1074 (2022).
- [109] P. W. Claeys and A. Lamacraft, *Maximum velocity quantum circuits*, *Phys. Rev. Res.* **2**, 033032 (2020).
- [110] S. Gopalakrishnan and A. Lamacraft, *Unitary circuits of finite depth and infinite width from quantum channels*, *Phys. Rev. B* **100**, 064309 (2019).
- [111] M. Akila, B. Gutkin, and T. Guhr, *Particle-time duality in the kicked Ising spin chain*, *J. Phys. A* **49**, 375101 (2016).
- [112] P. O. Schmidt, T. Rosenband, C. Langer, W. M. Itano, J. C. Bergquist, and D. J. Wineland, *Spectroscopy using quantum logic*, *Science* **309**, 749 (2005).
- [113] F. A. An, A. Ransford, A. Schaffer, L. R. Sletten, J. Gaebler, J. Hostetter, and G. Vittorini, *High fidelity state preparation and measurement of ion hyperfine qubits with $I > \frac{1}{2}$* , *Phys. Rev. Lett.* **129**, 130501 (2022).
- [114] K. K. Mehta, C. Zhang, M. Malinowski, T.-L. Nguyen, M. Stadler, and J. P. Home, *Integrated optical multi-ion quantum logic*, *Nature (London)* **586**, 533 (2020).
- [115] *Quantinuum hardware H2 benchmark* (2023), <https://github.com/CQCL/quantinuum-hardware-h2-benchmark>.
- [116] *Quantinuum hardware specifications* (2023), <https://github.com/CQCL/quantinuum-hardware-specifications>.
- [117] *Quantinuum hardware quantum volume data* (2023), <https://github.com/CQCL/quantinuum-hardware-quantum-volume>.
- [118] C. J. Wood and J. M. Gambetta, *Quantification and characterization of leakage errors*, *Phys. Rev. A* **97**, 032306 (2018).
- [119] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, *Randomized benchmarking of quantum gates*, *Phys. Rev. A* **77**, 012307 (2008).
- [120] A. M. Polloreno, A. Carignan-Dugas, J. Hines, R. Blume-Kohout, K. Young, and T. Proctor, *A theory of direct randomized benchmarking*, [arXiv:2302.13853](https://arxiv.org/abs/2302.13853).
- [121] D. Hanneke, J. P. Home, J. D. Jost, J. M. Amini, D. Leibfried, and D. J. Wineland, *Realization of a programmable two-qubit quantum processor*, *Nat. Phys.* **6**, 13 (2009).
- [122] J. Helsen, X. Xue, L. M. K. Vandersypen, and S. Wehner, *A new class of efficient randomized benchmarking protocols*, *npj Quantum Inf.* **5**, 71 (2019).
- [123] M. A. Nielsen, *A simple formula for the average gate fidelity of a quantum dynamical operation*, *Phys. Lett. A* **303**, 249 (2002).
- [124] A. Davison and D. Hinkley, *Bootstrap Methods and Their Application*, Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, Cambridge, England, 1997), 10.1017/CBO9780511802843.