

Space-time trade-off in networked virtual distillation

Tenzan Araki^{1,2,*}, Joseph F. Goodwin¹, and Bálint Koczor²

¹*Department of Physics, Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom*

²*Mathematical Institute, University of Oxford, Woodstock Road, Oxford OX2 6GG, United Kingdom*



(Received 20 May 2025; accepted 29 September 2025; published 27 October 2025)

In contrast to monolithic devices, modular, networked quantum architectures are based on interconnecting smaller quantum hardware nodes using quantum communication links and offer a promising approach to scalability. Virtual distillation (VD) is a technique that can, under ideal conditions, suppress errors exponentially as the number of quantum state copies increases. However, additional gate operations required for VD introduce further errors, which may limit its practical effectiveness. In this work, we analyze three practical implementations of VD that correspond to edge cases that maximize space-time trade-offs. Specifically, we consider an implementation that minimizes the number of qubits but introduces significantly deeper quantum circuits and contrast it with implementations that parallelize the preparation of copies using additional qubits, including a constant-depth implementation. We rigorously characterize their circuit depth and gate count requirements and develop explicit architectures for implementing them in networked quantum systems—while also detailing implementations in early fault-tolerant quantum architectures. We numerically compare the performance of the three implementations under realistic noise characteristics of networked ion trap systems and conclude the following. First, VD effectively suppresses errors even for very noisy states. Second, the constant-depth implementation consistently outperforms the implementation that minimizes the number of qubits. Finally, the approach is highly robust to errors in remote entangling operations, with noise in local gates being the main limiting factor to its performance.

DOI: [10.1103/t828-wvgl](https://doi.org/10.1103/t828-wvgl)

I. INTRODUCTION

There has recently been remarkable progress in the development of quantum hardware, but achieving full fault-tolerance remains challenging. Algorithms that are relatively robust to errors and therefore amenable to near-term quantum hardware have been proposed [1,2], but noise levels on current devices are still limiting their practical efficacy [3]. Solving practical problems beyond system sizes that are classically tractable before achieving full fault tolerance requires mitigating errors either in physical gates or logical gates in early fault-tolerant architectures [4–6]. Since one is often concerned with a specific observable property of a quantum system in typical applications of quantum computers, a broad range of quantum error mitigation (QEM) [7] techniques have been proposed that suppress errors in observable expectation values by postprocessing results from an ensemble of noisy quantum circuits—as opposed to fully correcting all errors in the quantum states.

Suppose we want to obtain the expectation value of an observable O with respect to some quantum state $|\psi_{\text{id}}\rangle$ (i.e.,

$\langle\psi_{\text{id}}|O|\psi_{\text{id}}\rangle$ or equivalently $\text{Tr}[O|\psi_{\text{id}}\rangle\langle\psi_{\text{id}}|]$) that would be prepared by an ideal noise-free quantum circuit. In practice, due to noise in the quantum hardware, the output of the circuit is a mixed state ρ , and an expected value measurement therefore yields $\text{Tr}[O\rho] \neq \langle\psi_{\text{id}}|O|\psi_{\text{id}}\rangle$. Virtual distillation (VD) [8,9] applies a shallow entangling operation (known as a controlled “derangement”) to n copies of the noisy state ρ as illustrated in Fig. 1(a), where each copy is prepared independently by applying the same noisy state preparation circuit [10]. Then, measuring an ancilla qubit gives rise to an estimator that approaches $\langle\psi|O|\psi\rangle$ exponentially with respect to n . Here $|\psi\rangle$ is the dominant eigenvector of ρ , which is subject to a slight deviation from the ideal state, i.e., $|\langle\psi||\psi_{\text{id}}\rangle| \neq 1$, known as a coherent mismatch [11]. In typical practical scenarios, this is negligible compared to the incoherent loss of fidelity, where the latter is suppressed exponentially effectively by VD.

While coherent errors are not suppressed by VD, they can be addressed either by Pauli twirling [12–15] or by combining it with other QEM techniques [16–19]. A more serious drawback of VD is the considerable overhead in the total number of qubits due to the preparation of multiple copies, a significant concern given that scaling qubit number remains the foremost challenge for almost all proposed quantum computing platforms. The problem can be alleviated by combining VD with dual-state purification [20,21], in which case the number of qubits can be halved. A qubit-efficient implementation of VD was also proposed [22], such that only two quantum registers are used to store n copies by repeatedly preparing new copies

*Contact author: tenzan.araki@physics.ox.ac.uk

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

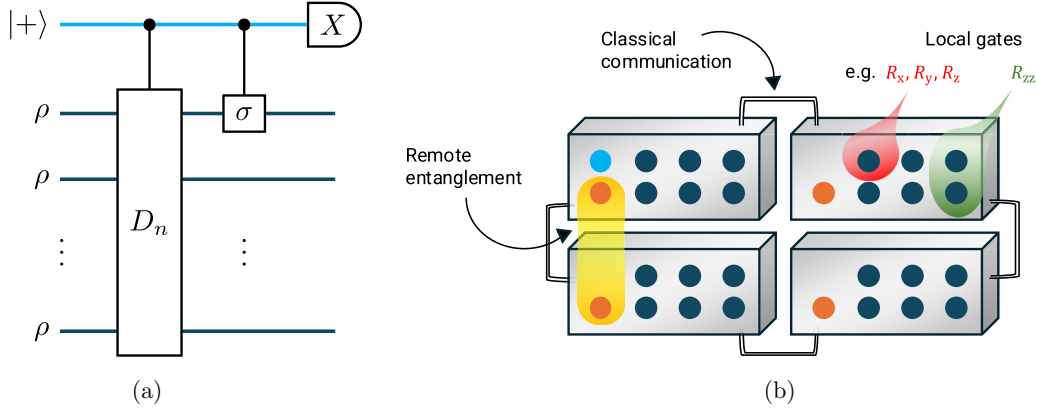


FIG. 1. (a) Quantum circuit implementing VD using n copies of the noisy quantum state ρ (dark blue) and a single ancilla qubit (light blue) initialized in $|+\rangle = (1/\sqrt{2})(|0\rangle + |1\rangle)$. A derangement operation is applied to all n copies (denoted by D_n), conditioned on the state of the ancilla qubit, and is followed by a set of controlled Pauli gates (denoted by σ) associated with each term in the Pauli basis expansion of O . A measurement of the ancilla qubit in the X -basis reveals the nonlinear functional $\text{Tr}[\sigma \rho^n]$, which is central to the scheme. (b) Conceptual illustration of applying VD with four copies of six data qubits across a four-node quantum network, where each copy (dark blue) is placed in a distinct node. The ancilla qubit (light blue) is placed in one of the nodes, and each node has an additional network qubit (orange) used to establish remote entanglement between nodes. With Bell state generation, local operations, and classical communication enabled, VD can be applied across the quantum network.

and resetting those that are no longer needed. However, while reducing the number of qubits required, such implementations lead to deep circuits which suffer from increased idling errors. In the present work, we rigorously analyze circuit depth and gate count requirements of a variety of such implementations, including a variant that achieves a constant circuit depth, albeit requiring an increased number of qubits.

With VD protocols requiring at least double the number of qubits versus unmitigated circuits, consideration of practical routes to scaling and their corresponding impact on the implementation of the VD routine are essential in predicting real-world performance. In widely used trapped-ion quantum architectures, e.g., scaling via producing a long chain of ions results in slower entangling gate speeds and lower fidelities, due to increasingly crowded motional mode frequencies; meanwhile, instead using many spatially separated small chains increases ion shuttling overheads and control hardware complexity. Beyond the barriers to scaling qubit number, monolithic quantum architectures pose challenges for implementing VD and other large-scale protocols due to, e.g., connectivity constraints. Therefore, in this work we consider applying VD across a modular network of quantum processing units (QPUs) [23–26] as illustrated in Fig. 1(b).

Coherent quantum operations across distinct nodes currently represent a bottleneck in modular quantum architectures due to their relatively slow and noisy operations. However, typical practical applications, e.g., simulating time evolution via Trotterization, require state preparation circuits that are significantly—potentially orders of magnitude—deeper than the remote entangling operations required for VD which are applied across the entire network—VD is therefore well suited to modular architectures [27]. While in this work we illustrate applications of VD in noisy devices, we will detail that our implementations are immediately compatible with early fault-tolerant quantum devices, whereby applying QEM techniques for mitigating residual, non-negligible logical errors is expected to be a major enabler—and VD is

expected to be one of the most promising candidates for the early fault-tolerant regime [3].

In Sec. II, we describe three general implementations of VD that represent edge cases in achieving space-time trade-offs. We then introduce quantum networks and detail architectures for implementing VD in Sec. III. We also briefly discuss implementation details of VD in modular early fault-tolerant quantum architectures in this section. In Sec. IV, we demonstrate simulation results for comparing the performance of different VD implementations in a prototypical example of a quantum network of trapped-ion QPUs, considering errors that accumulate during initialization, idling time, detection, and local and remote gate operations. We finally conclude in Sec. V.

II. SPACE-TIME TRADE-OFFS IN VD IMPLEMENTATIONS

Given n copies of a quantum state ρ and some Pauli string σ , VD can be performed via a Hadamard test as illustrated in Fig. 1(a), where sampling the ancilla qubit allows one to estimate $\text{Tr}[\sigma \rho^n]$. Doing so for the individual terms σ_i in the Pauli decomposition of an observable $O = \sum_i c_i \sigma_i$ leads to an approximation of the ideal expectation value as

$$\sum_i c_i \frac{\text{Tr}[\sigma_i \rho^n]}{\text{Tr}[\rho^n]} = \langle \psi | O | \psi \rangle + \mathcal{E}, \quad (1)$$

where $\mathcal{E}$ is an error that decays exponentially with respect to n and $|\psi\rangle$ is the dominant eigenvector of ρ . Please refer to Refs. [8] and [9] for proofs. In typical practical scenarios, $|\psi\rangle$ is a good approximation of the ideal quantum state [11], such that $\langle \psi_{\text{id}} | O | \psi_{\text{id}} \rangle \approx \langle \psi | O | \psi \rangle$. The derangement operation, denoted as D_n in Fig. 1(a), is an n -cycle permutation of the n copies of ρ , which can be decomposed nonuniquely into a minimum of $n-1$ pairwise swaps [28]. Therefore, the circuit requires $n-1$ controlled SWAP (C-SWAP) operations applied transversally across each pair of N -qubit registers, and there is significant flexibility in choosing a specific permutation

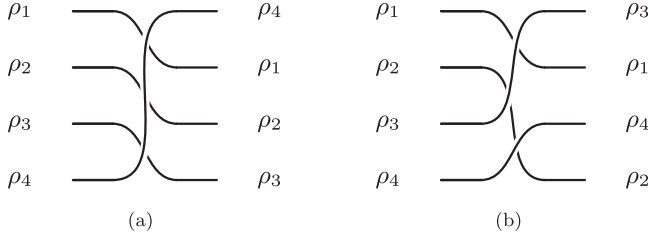


FIG. 2. n -cycles applied as (a) cyclic rotation and (b) brickwork, shown for the example where $n = 4$.

pattern. C-SWAP gates can also be implemented in parallel by introducing additional ancilla qubits, which was studied in Ref. [29] as a space-time trade-off with respect to both N and n . In this work, however, we restrict our attention to the space-time trade-off in n by applying a sequential implementation controlled by a single ancilla qubit, as this is the most relevant case to the networked architecture that we consider (Sec. III).

In the following, we present three practical implementations of VD considering general hardware architectures (i.e., either monolithic or modular), where the minimal number of $n-1$ transversal C-SWAP gates are applied. The implementations are each associated with one of the two derangements shown in Fig. 2, which are called *cyclic rotation* and *brickwork* in this work. We also note that the number of distinct n -cycles grows rapidly with the number of copies as $(n-1)!$ [9], but here we focus on three particular choices to explore edge cases in space-time trade-offs, summarized in Table I. We anticipate that the space-time trade-off in the intermediate regime varies monotonically between these extremes.

In order to minimize circuit depth, we apply each implementation destructively by replacing the final layer of transversal C-SWAP gates with a layer of transversal Bell state measurements (BSMs), i.e., pairs of qubits are measured in the basis spanned by Bell states, as detailed in Appendix A. Since C-SWAP gates can rarely be implemented natively in existing hardware platforms, trading them with BSMs can lead to significantly shallower circuits after compilation. As we will show, this trick is particularly effective when applied to the below BW, where half of the C-SWAP gates are in the final layer, and can therefore be removed, along with a slight yet experimentally crucial reduction in the size of the multipartite entangled state that the implementation utilizes. While further, potentially radical reduction in quantum resources is possible by exploiting the locality of entanglement

in certain many-body systems [30], we do not make use of those schemes as our aim is to assess the performance of VD in a general setting.

A. Cyclic rotation

The most commonly applied derangement for VD is shown in Fig. 2(a), where a set of noisy copies initially prepared in the order $\{\rho_1, \rho_2, \rho_3, \dots, \rho_n\}$ undergoes a permutation to $\{\rho_n, \rho_1, \rho_2, \dots, \rho_{n-1}\}$. While various arrangements of pairwise swaps can give rise to such cyclic rotation (CR), here we consider sequentially swapping between registers 1 and k , where k increments from 2 to n . Due to the sequential nature, only a single set of transversal C-SWAP gates can be applied destructively: the final set. The quantum circuit with its final C-SWAP gates applied destructively is shown in Fig. 3(a). Cyclic rotation requires in total $nN + 1$ qubits, i.e., $n-1$ copies of an N -qubit register and a register with $N+1$ qubits to account for the single ancilla qubit. This implementation then requires $(n-1)N$ C-SWAP gates and N BSMs.

B. Qubit-efficient cyclic rotation

By noticing that most registers do not participate in subsequent operations after being swapped in CR, a qubit-efficient version of CR has also been proposed [22], where the same derangement can be applied with a qubit number that is constant in the number of copies, as shown in Fig. 3(b). While CR prepares the copies in parallel, the latter does so by resetting the data register that is no longer in use and repreparing new copies in it. Noting that the preparation of new copies is typically the deepest part of the quantum circuit, the drawback of qubit-efficient cyclic rotation (QEQR) is the additional circuit depth and the resulting accumulation of idling errors. Only $2N + 1$ qubits are required in QEQR, consisting of N and $(N+1)$ -qubit registers. Due to the same derangement applied, the number of C-SWAP gates and BSMs is the same as in CR.

C. Brickwork

While both implementations described above lead to circuit depths that scale linearly in n , VD can also be applied in a constant circuit depth [29,31] by considering the derangement shown in Fig. 2(b). In this third implementation, two layers of transversal C-SWAP gates are applied in a brickwork (BW) structure for any n . The resulting quantum circuit, shown

TABLE I. Quantum resources required for qubit-efficient cyclic rotation (QEQR), CR, and BW for general architectures, showing a space-time trade-off. C-SWAP and Bell state measurements (BSM) counts are given by the number of individual operations rather than the number of transversal set of operations. The bottom row shows the total circuit depth in terms of the circuit depth of preparing each copy (d_ρ) and applying transversal C-SWAP gates (d_S), transversal BSMs (d_B), and the controlled Pauli gates (d_σ). We assume that d_ρ is much greater than d_σ and the circuit depth required to prepare a GHZ state.

	QEQR	CR	BW
Register count	2	n	n
Qubit count	$2N + 1$	$nN + 1$	$nN + \lfloor \frac{n-1}{2} \rfloor$
C-SWAP count	$(n-1)N$	$(n-1)N$	$\lfloor \frac{n-1}{2} \rfloor N$
BSM count	N	N	$\lfloor \frac{n}{2} \rfloor N$
Circuit depth	$(n-1)d_\rho + (n-2)d_S + d_B$	$d_\rho + (n-2)d_S + d_B + d_\sigma$	$d_\rho + d_S + d_B + d_\sigma$

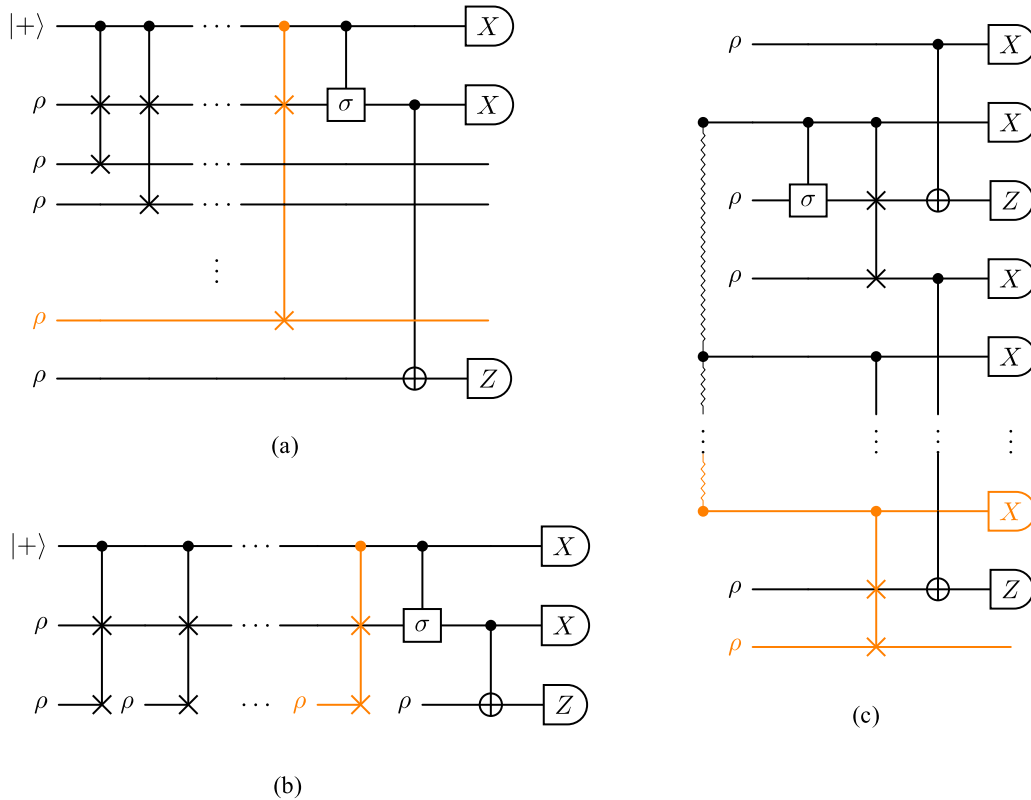


FIG. 3. Three implementations of VD after replacing the final layer of C-SWAP gates with BSMs: (a) Cyclic rotation (CR), (b) qubit-efficient cyclic rotation (QECR), and (c) brickwork (BW). The components highlighted in orange are present (absent) when the number of copies is odd (even). In every case, the controlled derangement in Fig. 1(a) is decomposed into C-SWAP gates and BSMs, where the latter consists of a CNOT gate followed by X - and Z -basis measurements. The wavy line across multiple ancilla qubits represents a GHZ state. The amount of quantum resources used in each implementation is summarized in Table I.

in Fig. 3(c), involves $\lfloor (n-1)/2 \rfloor$ ancilla qubits that share a Greenberger-Horne-Zeilinger (GHZ) state in order to parallelize the transversal C-SWAP gates. In Appendix B, we show that the circuit construction can be viewed as fanning out [32] the ancilla qubit and fanning it back in. While also benefiting from the parallelism in the preparation of new copies seen in CR, the BW implementation could furthermore lead to much shallower circuit depths in practice. This is not only due to the constant-depth nature of the construction but also because half of the transversal C-SWAP gates (specifically, all of those in the second layer) can be replaced by BSMs. As a result, this implementation involves a total of $nN + \lfloor (n-1)/2 \rfloor$ qubits, $\lfloor (n-1)/2 \rfloor N$ C-SWAP gates, and $\lfloor n/2 \rfloor N$ BSMs. Each qubit of the GHZ state is located in every other register, such that only N or $N+1$ qubits is required in each of the n registers.

III. DISTRIBUTED IMPLEMENTATIONS

In distributed, networked quantum computing architectures, multiple QPUs of moderate scale are interconnected by quantum (and classical) communication links. Each QPU, equipped with the usual initialization, local gate, and measurement operations, is a *module* or *node* of the quantum network. Crucially, quantum gates can be applied remotely across pairs of nodes by consuming Bell states prepared between them [33], enabling universal quantum computation across the quantum network. One or more qubits per node are

typically dedicated to storing a half of a Bell pair; these are called *network qubits*. In this work, we refer to the qubits on which the main computation is performed (i.e., those storing the copies ρ in VD) as *data qubits*. The controlled derangement operation is applied on these data qubits, conditioned on the *ancilla qubit*.

Although promising experimental progress has been made in the remote preparation of Bell states [34–36], they are and will likely remain a bottleneck for such architectures with regards to achievable rate and fidelity versus local bipartite entangling gates. In this view, the structures of CR, QECR, and BW implementations of VD are well suited for implementation in a distributed architecture: by placing each copy in a different node, the deepest part of the computation (the preparation of copies) can be performed independently within each node, and only a small number of remote operations bridge them together in the end. Therefore, in this work, we study the performance of VD across a quantum network when placing one copy per node at any given time and at most one ancilla qubit per node.

A. Connectivity

The connectivity requirement of the quantum network depends on the implementation considered, and centers around the nodes that contain the ancilla qubits. For clarity, we refer to these nodes that contain ancilla qubits as *control nodes* and

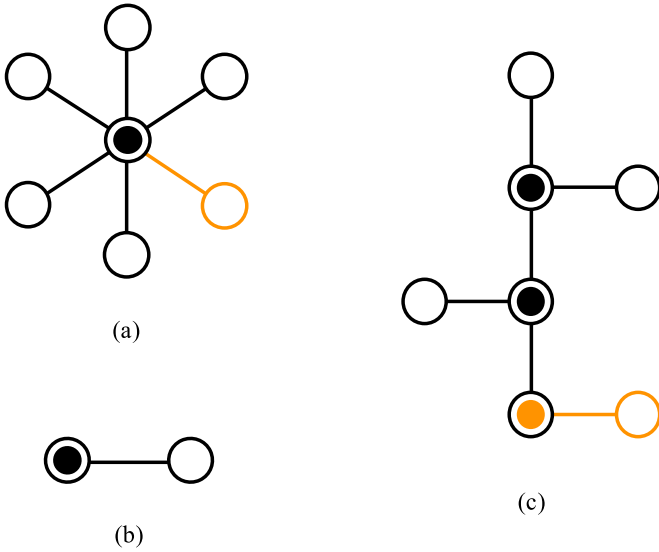


FIG. 4. Connectivities between nodes required to apply VD with six or six copies across a quantum network, where one copy is stored in each node at every moment, using (a) CR, (b) QECR, and (c) BW implementations. Each white circle represents a node containing a single copy, and those with a black solid circle also includes an ancilla qubit. Bell states can be prepared between each pair of nodes that are connected by a line. Components highlighted in orange are present (absent) when considering seven (six) copies.

the others as *target nodes*. If CR is applied across a quantum network, then remote operations will be applied between a single control node and all the target nodes, leading to a star graph connectivity [Fig. 4(a)]. If QECR is chosen instead, only two nodes are sufficient for any number of copies, consisting of one control node and one target node [Fig. 4(b)]. For BW, the $\lfloor (n-1)/2 \rfloor$ ancilla qubits that collectively share

a GHZ state are all placed in different control nodes, requiring at least a linear connectivity between them. If n is odd, then one of the two outermost control nodes is connected to two target nodes, and each of the other control nodes is connected to only one target node. It is similar in the case where n is even, but the outermost control node on the other end would also be connected to two target nodes [Fig. 4(c)].

B. Remote operations

The following three remote operations, illustrated in Fig. 5, are required when applying VD across a quantum network using the implementations described in Sec. II.

1. Remote C-SWAP gate

One of the copies to be swapped is placed in a control node and the other in a target node. While it is possible to decompose the C-SWAP gate into a series of single- and two-qubit gates and apply the two-qubit gates remotely, such an attempt would require a suboptimal number of Bell state preparations that in general cannot be parallelized. A simpler method considered in this work is to use quantum teleportation, which transfers the state of the qubit in the target node to the control node [see Fig. 5(a)], such that the C-SWAP gate can be applied locally within the control node. Note that in all implementations, we only teleport a copy to the control node but never teleport it back. For CR and QECR, this is because the copy that undergoes teleportation does not participate in the rest of the computation after the C-SWAP gate. While most of the copies involved in the C-SWAP layer of BW are still needed in the operations that follow, the subsequent operations can be performed from within the control node. Thus, each remote C-SWAP gate can be applied by consuming a single Bell pair prepared across one network qubit in each of the two nodes.

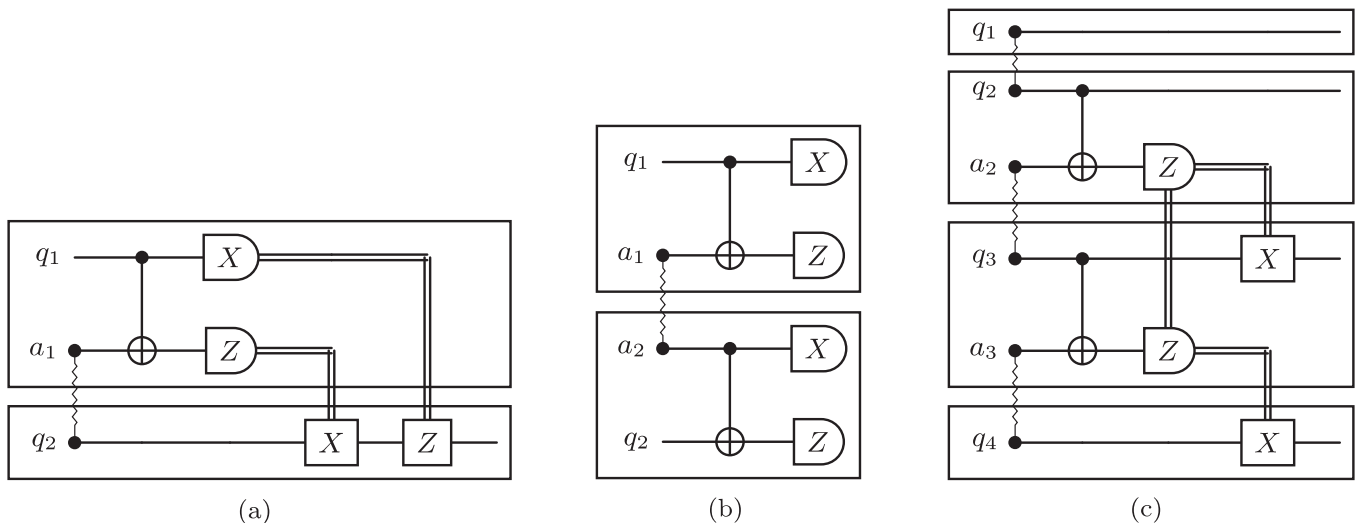


FIG. 5. Remote operations used to apply VD across a quantum network. (a) Teleportation of the state of qubit q_1 to a qubit q_2 in a different node, given a Bell pair shared between qubits a_1 and q_2 . (b) BSM between qubits q_1 and q_2 in different nodes. A remote BSM outcome of both 1 corresponds to when both pair of qubits (q_1, a_2) and (a_1, q_2) produce odd parity outcomes. (c) Preparation of an n -qubit GHZ state in constant time—example for $n = 4$ where the qubits q_1, q_2, q_3 , and q_4 are located in distinct nodes and ancilla qubits a_2 and a_3 are used. The X gates are corrections applied based on the collective outcome of the Z -basis measurements. The scheme scales straightforwardly to $n > 4$ qubit GHZ states.

2. Remote BSM

This is applied in the second layer of BW. We consider the implementation shown in Fig. 5(b), which corresponds to applying local BSMs after sharing a single Bell pair between the two participating nodes. The qubits on which remote BSM is applied do not need to be teleported to a different node after the C-SWAP layer of BW.

3. Remote GHZ state preparation

This is required to achieve the highly parallel structure of BW and can be realized in various ways (see, e.g., Refs. [37,38]). In particular, a GHZ state of N_{GHZ} qubits can be prepared in a constant circuit depth with the help of at most one network qubit per node using $N_{\text{GHZ}} - 1$ Bell states, along with local gates and adaptive operations, as shown in Fig. 5(c). The network qubits used in this step can be erased and reused to apply VD, such that no additional network qubits are required.

C. Fault-tolerant architectures

While our numerical simulations will assume implementations in physical qubits using physical gates, all space-time trade-off considerations apply equally to logical qubits and logical operations. It has recently been argued that VD is one of the most promising candidates for error mitigation in the early fault-tolerant era, whereby logical qubits are in limited supply and have non-negligible logical error rates [3]. Applying VD to mitigate the effect of residual logical errors in expected value measurements may allow—in certain error regimes—for a more efficient noise reduction than committing the increased number of physical qubits to increase the code distance. Furthermore, exploiting modular quantum architectures for VD remains equally relevant in the early fault-tolerant context, such that VD can be combined with error-corrected qubits similarly as concatenating QEC codes in distributed schemes.

When QEC is applied across a quantum network, the error correction cycle will be enacted remotely across multiple nodes, each holding data qubits to be protected. These data qubits could be individual physical qubits but may even be logical qubits that are concatenated by the code on the higher layer. Syndromes can be extracted using single ancillary registers in the standard approach, but a more architecture-native method involves measuring parities through GHZ states [39,40]. Entanglement distillation [25,41–44] can be applied either physically or logically to improve the fidelity of remotely established Bell pairs.

Consider the same distributed implementation of VD as discussed in the rest of this section but applied to logical qubits instead of physical ones. Then each copy ρ is encoded logically and may deviate from $|\psi_{\text{id}}\rangle$ due to, e.g., residual logical errors or compilation errors in randomized compiling [18,45–47]. The controlled derangement operation can be implemented fault tolerantly and remotely across the quantum network with the use of network qubits—the required transversal C-SWAP gates can similarly be implemented by teleporting single logical qubits from the target node to the control node and by applying local logical C-SWAP gates,

each of which can be decomposed into a sequence of a small number of Clifford+T gates [48,49].

IV. NUMERICAL SIMULATIONS

In the following, we numerically compare and analyze the performance of the three practical implementations of VD described in Sec. II when applied across a quantum network. In particular, we consider a noisy quantum network of trapped-ion QPUs as a concrete example. Moreover, we consider a practically motivated benchmark problem whereby we estimate an observable expectation value after time-evolving a simple initial state under a random-field Heisenberg model [50,51] using Trotterization [52,53] (see Appendix D5 for details). Thus, the simulations involve the practical implementations shown in Fig. 3 compiled to the trapped-ion quantum network, with input states ρ corresponding to noisy realizations of Trotter circuits.

A. Details of noise models

We decompose all quantum gates involved in the Trotter and VD parts of the circuits into primitives native to trapped-ion platforms (see Appendix D3 for details). We model noise in native single (two) qubit gates by applying ideal gates followed by single (two) qubit depolarizing error channels. Coherent errors are expected to be negligible in these systems due to precise calibration of control pulses [54]. Furthermore, we model the error associated with remote Bell state generation via an incoherent mixture of all four Bell states, which is formally equivalent to a single-qubit depolarizing channel applied only to one qubit of an ideal Bell pair. We use experimentally realistic values as reference error probabilities, assuming performance typical of leading experiments, specifically $\{p_{1Q}^*, p_{2Q}^*, p_{\text{Bell}}^*\} = \{10^{-4}, 10^{-3}, 10^{-2}\}$ for single-qubit gates, two-qubit gates, and remote entanglement generation, respectively. We then conduct a range of simulations considering a range of errors around these nominal values.

Detection errors occur at midcircuit or final measurements and correspond to unbiased bit-flips in the measurement outcomes. In our simulations, we fix this to an error probability of 10^{-3} . Since there are no time constraints in the initialization of the data qubits in the beginning of the circuit run, we assume this can be performed with negligible error [55]. Moreover, any midcircuit state preparation that occur during remote entanglement generation lead only to a marginal reduction in the entanglement *rate* with negligible impact on the fidelity. We therefore do not explicitly simulate these state preparation errors. However, the remaining midcircuit state preparation required to reprepare new copies in QECC are subject to stricter time constraints, and we model the resulting errors as unbiased bit-flips of probability 10^{-3} . Assuming that idling errors occur with probability 10^{-5} per $1\mu\text{s}$, they are simulated by applying single-qubit dephasing channels with error probabilities obtained from the accumulated idling times between consecutive operations.

B. Comparing different implementations

We perform direct density matrix simulations to probe expectation values free of shot noise for QECC and CR up

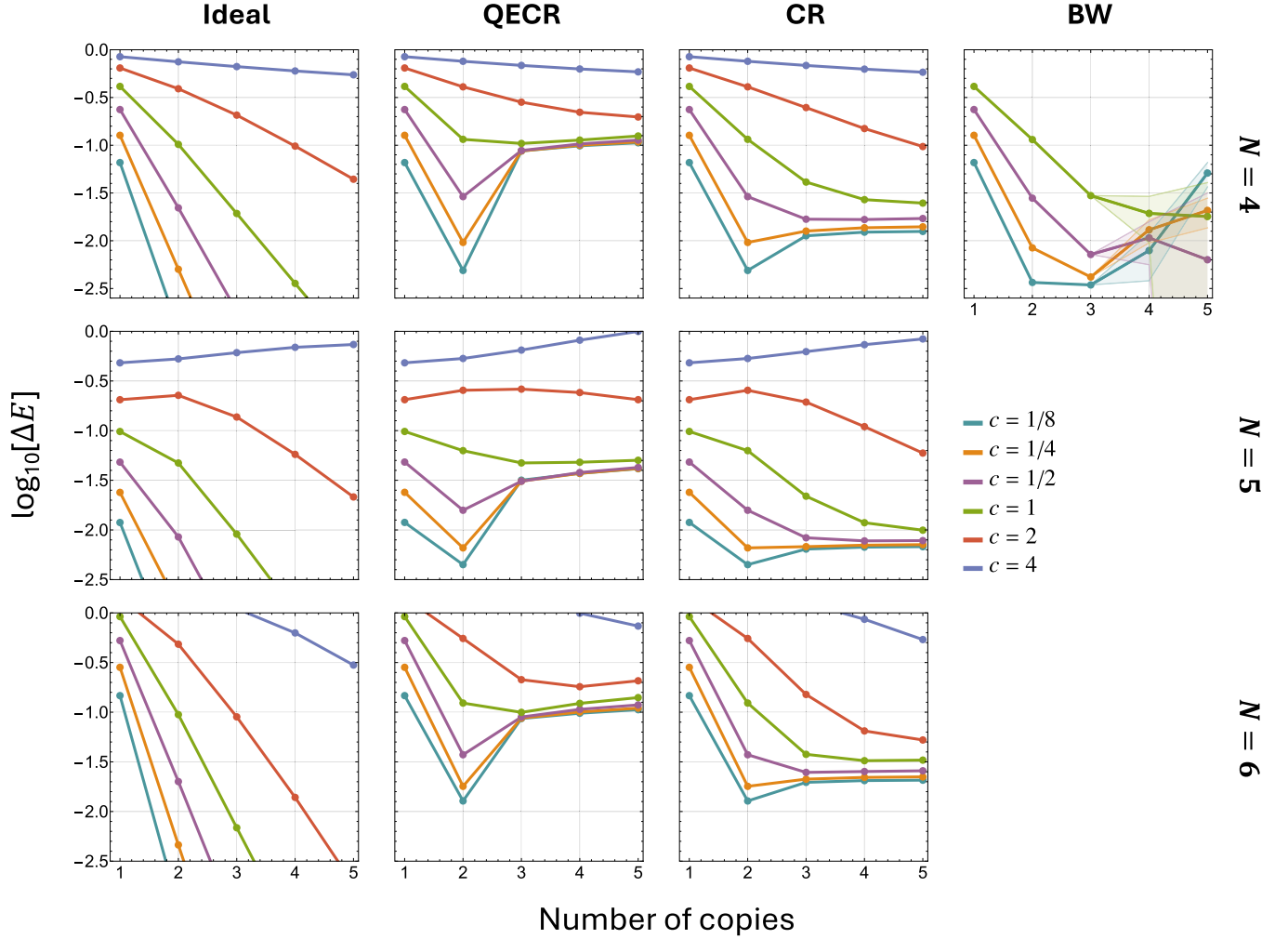


FIG. 6. Comparison between an ideal VD and QECR, CR, and BW (columns) for system sizes $N \in \{4, 5, 6\}$ (rows) and error probabilities $\{p_{1Q}, p_{2Q}, p_{\text{Bell}}\} = \{cp_{1Q}^*, cp_{2Q}^*, cp_{\text{Bell}}^*\}$ controlled by the parameter $c \in \{1/8, 1/4, 1/2, 1, 2, 4\}$ (line colors). In each case, the base-10 logarithm of the absolute error $\Delta E := |\langle O \rangle_{\text{VD}}^{(n)} - \langle \psi | O | \psi \rangle|$ is plotted against the number of copies. BW is only simulated for $N = 4$ and $c \in \{1/8, 1/4, 1/2, 1\}$ due to the demanding computation, and estimated standard errors for its Monte Carlo simulation are represented by shaded areas. By comparing QECR, CR, and BW with the ideal implementation, we observe that errors in the system limits the suppression power of VD, but it is still able to suppress errors for increased number of copies and high c . QECR is less effective than CR and BW due to the significant idling errors associated with the repeated preparation of copies, highlighting the importance of preparing n copies in parallel in n nodes.

to six qubits. For BW, we only simulate the case with four data qubits and use pure-state Monte Carlo simulations for four and five copies to circumvent high memory requirements of density matrix simulations. The accuracy of the latter approximate simulation depends on the number of Monte Carlo samples used, as detailed in Appendix D2. Therefore, the largest simulation we perform uses $5 \times 4 + 2 = 22$ qubits, which includes two ancilla qubits that are initialized in a Bell pair (a bipartite GHZ state). By denoting the error-mitigated expectation value from an ideal or noisy implementation of VD with n copies as $\langle O \rangle_{\text{VD}}^{(n)}$, in Fig. 6, we plot the absolute error from the expectation value obtained from the dominant eigenvector, i.e., $\Delta E := |\langle O \rangle_{\text{VD}}^{(n)} - \langle \psi | O | \psi \rangle|$, for an increasing number of copies $n \in \{1, 2, 3, 4, 5\}$ in each case. In these simulations, we scale all reference error probabilities by a factor of $c \in 2^{\{-3, \dots, 2\}}$, i.e., $\{p_{1Q}, p_{2Q}, p_{\text{Bell}}\} = c\{p_{1Q}^*, p_{2Q}^*, p_{\text{Bell}}^*\}$, but keep idling errors constant. For our BW Monte Carlo

simulations, we use shaded areas to show the estimated standard errors in the figure, and we do not consider $c \in \{2, 4\}$ due to its excessive computational cost. The regime where quantum error mitigation can be effective is when the circuit error rate is of order 1. We choose the number of Trotter steps consistently such that the circuit error rate of each copy is approximately 1 for the $c = 1$ case, and therefore the circuit error rate is approximately c for the different simulations.

While theoretical error bounds indeed decrease monotonically as we increase n , the absolute error may increase for small n even for an ideal implementation of VD—but then the absolute error decreases exponentially when the number of copies is sufficiently high [9]. This increase for small n can be observed in the case of $N = 5$ and $c = 4$ of Fig. 6 and we further plot in Appendix D6 that the absolute error indeed decreases for $n \geq 6$. The variations between $N = 4, 5, 6$ are

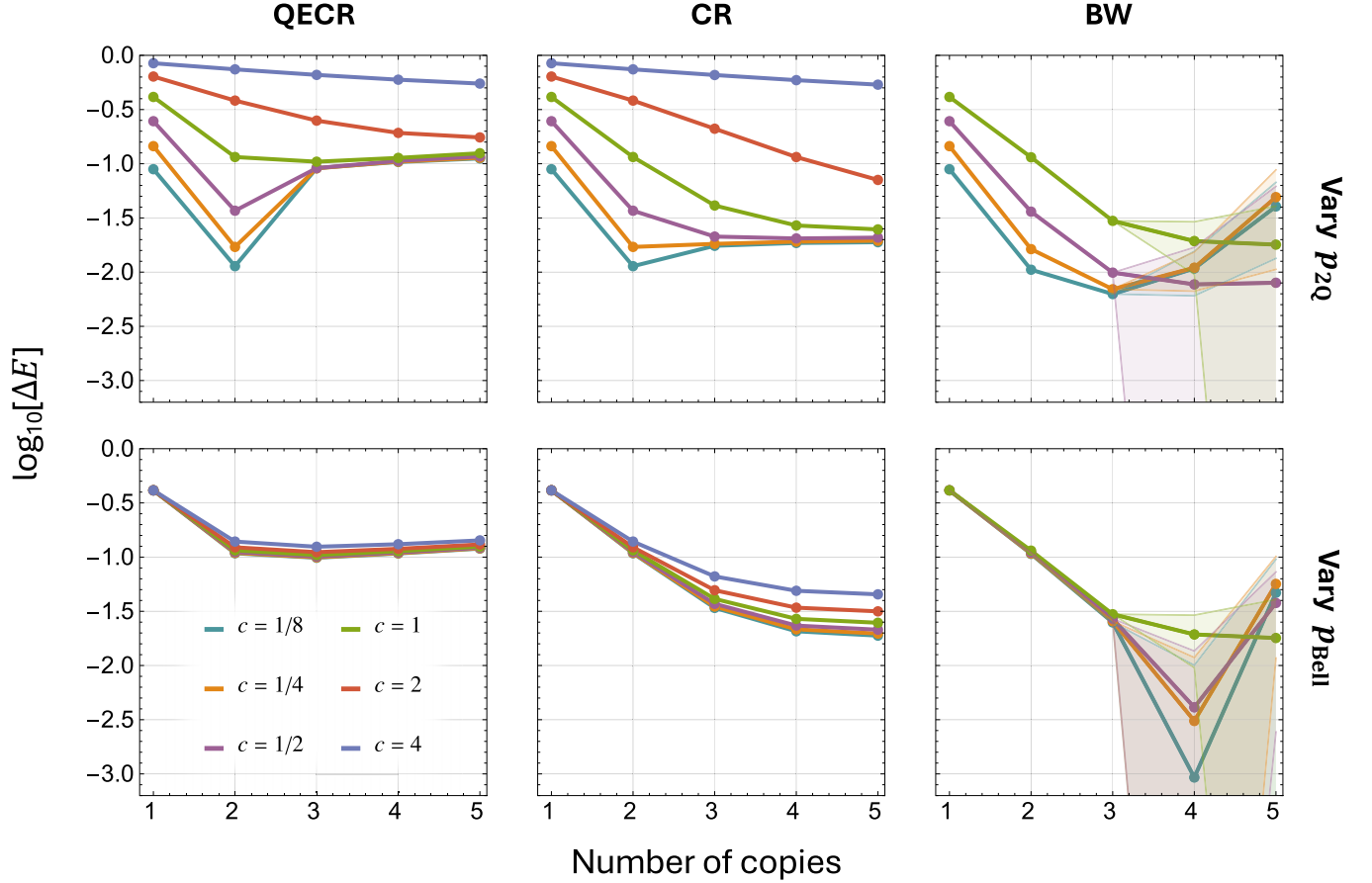


FIG. 7. Performance of QEQR, CR, and BW (columns) when varying only two-qubit gate error probability as $\{p_{1Q}, p_{2Q}, p_{Bell}\} = \{p_{1Q}^*, c p_{2Q}^*, p_{Bell}^*\}$ (first row) and when varying only entanglement error probability as $\{p_{1Q}, p_{2Q}, p_{Bell}\} = \{p_{1Q}^*, p_{2Q}^*, c p_{Bell}^*\}$ (second row) controlled by the parameter $c \in \{1/8, 1/4, 1/2, 1, 2, 4\}$ (line colors). In each case, the base-10 logarithm of the absolute error $\Delta E := |\langle O \rangle_{\psi_D}^{(n)} - \langle \psi | O | \psi \rangle|$ is plotted against the number of copies. BW is only simulated for $c \in \{1/8, 1/4, 1/2, 1\}$ due to the demanding computation, and estimated standard errors for its Monte Carlo simulation are represented by shaded areas. The robustness of the performance of the three implementations against entanglement errors relative to two-qubit gate errors shows the compatibility of VD with quantum networks.

expected due to the different random instances of the system Hamiltonian considered (see Appendix D 5).

The performance of the three practical implementations is affected to various degrees by the presence of errors in the derangement subcircuit. For $N = 4$, we observe that both QEQR and CR lead to improvements consistently when $n = 2$ but only when c is large enough for $n \geq 3$. In the latter case, their absolute errors plateau, as can be observed for $c \leq 1$ in both implementations. The plateau is around 10^{-1} for QEQR and around 10^{-2} for CR, indicating that the presence of significant idling errors in QEQR due to repeated resetting of quantum registers and preparing new copies has a detrimental impact in practice. Similar features can be seen for $N = 5$ and $N = 6$, highlighting the importance of preparing copies in parallel. The performance of BW for $c \leq 1$ is more similar to that of CR than QEQR due to the parallel preparation of copies. Furthermore, it improves the absolute error up to $n = 3$ and to lower values compared to CR consistently for all $c \leq 1$. This can be attributed to the reduced circuit depth and two-qubit gate counts given by its constant-depth nature and the replacement of more C-SWAP gates to BSMs in the final layer. Nevertheless, the ancillary GHZ state utilized in BW also introduces additional

error sources, and even BW does not show improvements for $n \geq 4$.

C. Analyzing robustness against different error sources

Next we study how varying the two-qubit gate and entanglement generation error probabilities each affect the performances of the three practical implementations in Fig. 7. Focusing on $N = 4$, we set $\{p_{1Q}, p_{2Q}, p_{Bell}\} = \{p_{1Q}^*, c p_{2Q}^*, p_{Bell}^*\}$ in the former and $\{p_{1Q}, p_{2Q}, p_{Bell}\} = \{p_{1Q}^*, p_{2Q}^*, c p_{Bell}^*\}$ in the latter. Since we choose the circuit depth such that the circuit error rate of each copy is close to 1 when $p_{2Q} = p_{2Q}^*$ (see Appendix D 5), we vary the circuit error rate, along with the errors in the controlled derangement, when varying p_{2Q} .

When varying only two-qubit gate error probabilities, we observe the same qualitative features as the comparison in Fig. 6 for $N = 4$. This shows that the relative performance between the three practical implementations depends strongly on two-qubit gate error probabilities. On the other hand, varying only errors in remote entanglement preparation has a minimal impact on the performance of BW when $n \leq 3$, where only at most $2N$ remote Bell pairs are generated, and for all n of QEQR and CR. In particular, the absolute error

improves monotonically for all n for CR and at least for all $n \leq 4$ for BW in all entanglement error probabilities considered. Thus, while entanglement errors do compromise the performance of VD when applied across a quantum network, the effects are much less concerning than local two-qubit gate errors for realistic error rates, confirming the compatibility between the technique and the hardware.

V. CONCLUSION

In this work, we considered practical implementations of VD and compared their performance in realistic scenarios where the qubits and operations involved in VD are subject to errors. We focused on implementation variants of VD that can be applied across a quantum network, including three specific edge cases in space-time trade-offs. In particular, we considered the standard CR, its qubit-efficient variant (QECR), and a depth-efficient variant that can be implemented in a constant circuit depth (in the number of copies) that we term BW. We comprehensively detailed implementations of these variants in general and characterized the total numbers of qubits, gate counts and circuit depths required. Furthermore, we explicitly constructed optimized architectures assuming characteristics of networked ion-trap platforms.

We then performed comprehensive numerical simulations assuming realistic noise characteristics of ion trap systems, using density matrix and Monte Carlo simulation methods—the largest systems we simulated involved 22 noisy qubits. We confirmed that the accumulation of idling errors in QECR limits its performance and underscores the advantage of CR and BW, which both prepare copies in parallel. While entangling many copies may present a challenge for monolithic quantum devices, the circuits involved are particularly well suited to networked quantum devices. The primary challenge then becomes the generation of remote entanglement between nodes for the controlled derangement operation, a process we assume will suffer from considerably lower fidelity than can be achieved via local two-qubit gates. However, we observed that, for realistic assumptions, the performance of VD remains primarily limited by the fidelity of local gates, demonstrating its robustness to errors in remote operations. We conclude that, when applied across a quantum network, VD can potentially be effective with more than three copies if the copies are prepared in parallel and reliable implementations of local operations are available.

While our numerical simulations focused on physical gate error characteristics, we highlighted that our implementations of VD can be applied in a networked architecture in an early fault-tolerant setting. In this scenario, each copy stores error-corrected qubits that are manipulated by logical quantum operations [3] and VD is used to mitigate residual logical errors or algorithmic errors due to, e.g., approximate randomized compiling [18,45–47].

Rather than storing each copy in a different node, one can also envision placing individual logical qubits in each node, which is often of interest in distributed logical architectures, where each node can only accommodate a limited number of physical qubits. In this case, the use of one logical ancilla qubit per logical data qubit could be utilized to enable more flexible choices of network connectivities. One can even

distribute single logical qubits across physical data qubits placed in distinct nodes. However, these implementations would require logical gates to be applied remotely, leading to higher susceptibility to errors. Combining VD with classical codes applied on quantum architectures with highly biased noise [56] could also serve as a resource-efficient means of suppressing errors. Integrating VD into modular early fault-tolerant quantum devices presents a promising direction for future exploration.

ACKNOWLEDGMENTS

Numerical experiments for this work utilize the Quantum Exact Simulation Toolkit (QuEST) [57] via the QuESTlink [58] frontend. Quantum circuits in the main text were illustrated using the Quantikz LaTeX package [59]. T.A. thanks Tyson Jones for software assistance, such as enabling functionalities in QuESTlink that improved simulation efficiency, and Arjun Rao and Tom Hinde for valuable comments on the manuscript. T.A. acknowledges support by the Oxford-Uehiro Graduate Scholarship Programme. T.A. also acknowledges the use of the University of Oxford Advanced Research Computing (ARC) facility [60] in carrying out this work. J.G. thanks the UK Engineering and Physical Sciences Research Council (EPSRC) for funding under the Horizon Europe Guarantee (EP/Y026438/1) and the European Research Council (ERC) for selecting and approving the proposal (ERC Starting Grant: MICRON-QC-101077098). B.K. thanks UKRI for the Future Leaders Fellowship Theory to Enable Practical Quantum Advantage (MR/Y015843/1). The authors also acknowledge funding from the EPSRC projects Robust and Reliable Quantum Computing (RoaRQ, EP/W032635/1) and Software Enabling Early Quantum Advantage (SEEQA, EP/Y004655/1). B.K. thanks the University of Oxford for a Glasstone Research Fellowship and Lady Margaret Hall, Oxford for a Research Fellowship. This research was funded in part by UKRI (MR/Y015843/1).

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

APPENDIX A: REPLACING C-SWAP GATES WITH BSMS

For each of the three implementations introduced in Sec. II, all C-SWAP gates on the final layer are replaced by BSMS. The quantum circuits for the three implementations prior to performing this replacement are shown in Fig. 8. Since the C-SWAP gate is not typically available as a native operation in today's quantum architectures, replacing it with BSMS can lead to significant reductions in the compiled circuit depth. Furthermore, this replacement is most effective for BW, since it involves $\lfloor \frac{n}{2} \rfloor$ transversal C-SWAP gates on the final layer, whereas each of the other methods only has 1. Here we use

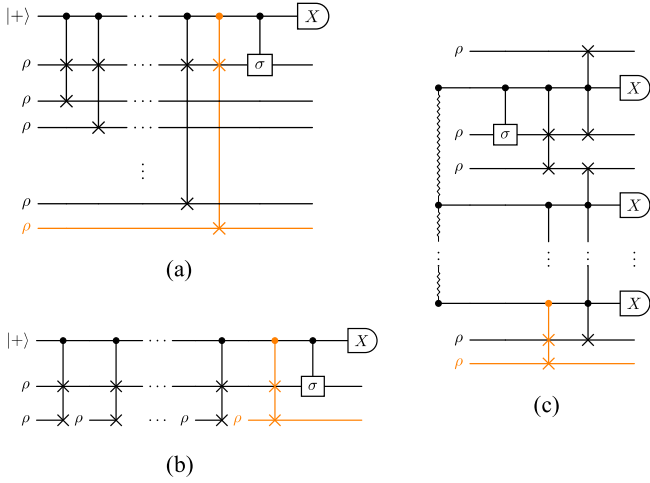


FIG. 8. Three implementations of VD before replacing the final layer of C-SWAP gates with BSMs: (a) CR, (b) QECR, and (c) BW. The components highlighted in orange are present (absent) when the number of copies is odd (even). As opposed to the circuits shown in Fig. 3, all circuits shown here involve $n - 1$ C-SWAP gates and no BSMs. The wavy line across multiple ancilla qubits represents a GHZ state.

the destructive swap test as a minimal example to describe how to arrive at this replacement for VD.

The swap test evaluates the overlap $\text{Tr}[\rho_1 \rho_2]$ between two quantum states ρ_1 and ρ_2 using the quantum circuit shown in Fig. 9(a). While this is its traditional nondestructive implementation, it is known that the swap test can also be performed destructively as in Fig. 9(f), where the ancilla qubit is removed and the C-SWAP gate is replaced by a BSM [61,62]. Their equivalence can be understood by noticing that the state of the ancilla qubit, which is the only qubit that is measured in the end, is unaffected by the application of subsequent operations on all other qubits. In particular, an additional BSM can be applied across the two quantum registers as shown in Fig. 9(b). The circuit can be simplified by decomposing the C-SWAP gate into a Toffoli gate sandwiched by a pair of CNOT gates, leading to Fig. 9(c) following a cancellation of two neighboring CNOT gates. Changing the X-basis

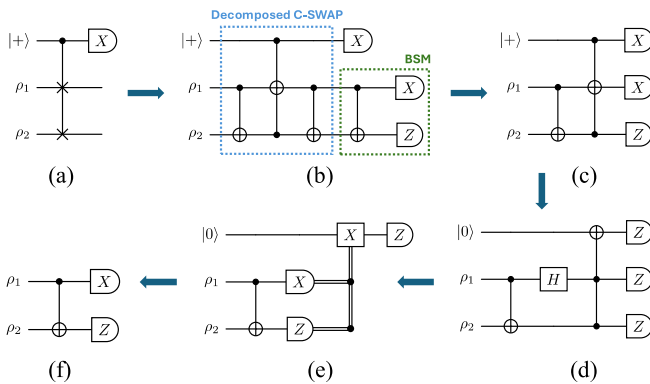


FIG. 9. Sketch of equivalence between the nondestructive and destructive swap tests. The same argument allows us to replace the final layer of C-SWAP gates in VD with BSMs.

measurements to Z-basis measurements and commuting the Hadamard gates over to the left, we arrive at Fig. 9(d), where the ancilla qubit is the target of the Toffoli gate. By the principle of deferred measurement, the Toffoli gate can be changed to an X gate conditioned on classical measurement outcomes as in Fig. 9(e). Since, at this stage, only classical information processing is performed on the ancilla qubit, this qubit can be removed from the quantum circuit, noting that the probability of measuring 1 on the ancilla qubit in the original circuit is now translated to whether or not both measurements in Fig. 9(f) produce 1s in this destructive version.

The same procedure can be applied to each of the C-SWAP gates on the final layer of any VD implementation. Note that bringing the controlled Pauli gates to before the derangement does not change the final measurement outcomes, so we do so to maximize the number of C-SWAP gates that can be replaced by BSMs. When n is even in BW, we also notice that this replacement makes one of the ancilla qubits uninvolved in the derangement, such that we can remove this qubit as highlighted in orange in Fig. 3(c). This is advantageous in practical implementations, since this implies that VD can be applied with the same number of copies but consuming a smaller GHZ state.

APPENDIX B: GATE PARALLELIZATION WITH QUANTUM FAN-OUT AND FAN-IN

The quantum fan-out and fan-in primitives have been utilized to parallelize quantum gates conditioned on ancilla qubits, and they are particularly well suited for modular quantum architectures [32]. Here we review the fan-out and fan-in primitives to describe how fanning out the ancilla qubit in CR [Fig. 8(a)] leads to BW [Fig. 8(c)].

Given an arbitrary pure single-qubit state $|\phi^{(1)}\rangle = \alpha|0\rangle + \beta|1\rangle$, a single fan-out expands it to $|\phi^{(2)}\rangle = \alpha|00\rangle + \beta|11\rangle$ for complex amplitudes α and β . This can be achieved by applying a CNOT gate controlled by the qubit storing $|\phi^{(1)}\rangle$ and targeted on another qubit initialized to $|0\rangle$. The technique can be scaled straightforwardly by applying CNOT gates from the same qubit targeted on additional $|0\rangle$ states, from which we obtain the s -qubit state $|\phi^{(s)}\rangle = \alpha|00\dots 0\rangle + \beta|11\dots 1\rangle$ for some $s \geq 2$. When a quantum circuit relies on separable gates that are controlled by a qubit in $|\phi^{(1)}\rangle$, as in the case of VD, fanning it out into $|\phi^{(s)}\rangle$ enables parallel applications of the controlled gates. After using the fanned-out state for parallel gate operations, a reverse fan-in operation brings $|\phi^{(s)}\rangle$ back to $|\phi^{(1)}\rangle$ before its measurement. While the most conceptually simple way to perform fan-in is by applying the same set of CNOT operations as in the fan-out part, the fact that the additional qubits always return to $|0\rangle$ can be utilized to arrive at a destructive fan-in circuit that does not require any two-qubit gate.

In VD and in many other quantum circuits (e.g., the Hadamard test), an ancilla qubit is initially prepared in the $|+\rangle$ state and finally measured in the X (or Y) basis after being used as a control qubit for the main computation. In this case, the ancilla qubit will be fanned out to the Bell state $|\phi^{(2)}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ for $s = 2$ and more generally to s -partite GHZ states. Thus, instead of performing sequential CNOT gates after initializing the ancilla qubit to $|+\rangle$, this

fanned out state can be prepared using existing GHZ state preparation techniques that could be more efficient, including those discussed in Sec. III. Similarly, fan-in can also be simplified in this case by noticing that a Z gate conditioned on a classical measurement outcome only induces a bit flip on the final X - or Y -basis measurement. This means that all the classical communication and the following conditional Z gates are redundant, provided one simply keeps track of the parity between the measurement outcomes. These two simplifications are reflected in the quantum circuit for BW in Figs. 3(c) and 8(c).

APPENDIX C: PHYSICAL CHARACTERISTICS OF A QUANTUM NETWORK OF TRAPPED IONS

As a concrete example of a quantum network, we consider for the purposes of this study a trapped-ion quantum architecture in which qubits are encoded in a long-lived transition within some Group-II or Group-II-like ionic species. Such systems exhibit very long coherence times, often via encoding in hyperfine “clock” transitions [63–66] with first-order-insensitivity of frequency to magnetic field. Alternatively, other metastable transitions (e.g., ground-state Zeeman qubits) can be used in combination with ultrastable magnetic bias fields; in either case decoherence is generally dominated by qubit dephasing. Qubits within each node are assumed to be confined in one or more linear Paul traps such that all-to-all connectivity is achieved either across a single, long chain of ions [67] or via shorter chains of ions, shuttled, split, and combined across numerous discrete trapping potentials [68].

Depending on the qubit type and physical details of the platform, gates may be realized via the application of radiation at optical, microwave, or radiofrequency bands. Whatever the mechanism used, this gives rise to a native single qubit gate set consisting of rotations $R_x(\theta)$, $R_y(\theta)$, $R_z(\theta)$, parameterized by angles θ about the three perpendicular axes X , Y , and Z . While R_x and R_y are noisy physical operations applied over a 1- μ s timescale [69], R_z is typically an instantaneous noiseless virtual Z gate [70]. A wide variety of two-qubit gate operations have been demonstrated in recent decades, which achieve entanglement via conditional rotations of a two-qubit state around either equatorial or azimuthal axes [71]. For the purposes of this work, we assume that local entangling operations are achieved via a geometric phase gate which applies a maximally entangling $R_{zz}(\theta)$ rotation in 10 μ s [69,72,73]. These single- and two-qubit gates together form a universal gate set within each node.

By enabling remote entanglement, universal quantum computation can be performed across the quantum network. This can be achieved by entangling the internal state of the participating ion in each of two nodes with a state of a photon emitted from that ion. A variety of photonic flying-qubit encodings can be used, including frequency, photon number, polarization, or time bin. The photons from the two nodes to be entangled are then directed such that they interfere on a beamsplitter, and their subsequent coincident detection on the far side of this beamsplitter projects the entanglement onto the ions, resulting in an ion-ion Bell state heralded by the coincidence pattern measured.

Remote entanglement rates are a significant bottleneck in quantum network performance, with leading experiments demonstrating Bell-pair generation rates on the order of 100 Hz [34]. While remote rates are likely to remain lower than those achievable locally, significant increases in performance are anticipated via the use of optical cavities to mediate network ion-photon coupling. In this work we assume that remote entanglement can be achieved at a rate of 10 kHz with fidelities an order of magnitude lower than local two-qubit gates. Both fidelity and rate assumptions exceed the state-of-the-art in current cavity-mediated remote entanglement experiments [35] but are consistent with realistic engineering assumptions and stand-alone demonstrations of photon collection efficiencies achievable with optical cavities [74].

Qubits can be initialized or reset during computation by applying a sequence of cooling techniques, such as Doppler cooling and sideband cooling, followed by optical pumping. The duration of the entire process is on the order of 1 μ s for the preparation fidelity required for midcircuit remote entanglement generation ($\sim 99\%$); high-fidelity state preparation ($\sim 99.99\%$) is typically at least an order of magnitude slower but is only required during initialization of data qubits ρ , before the circuit is applied. The detection of ions is performed by collecting photons from state-dependent fluorescence. The probability of detection errors depends on the duration of the illumination and collection efficiency, but high-fidelity detection is possible in $\sim 100 \mu$ s.

APPENDIX D: SIMULATION DETAILS

The results presented in Sec. IV analyses the performance of VD when applied across a quantum network of trapped ions. Here we describe details of the simulation.

1. Exact simulations

In this work, exact simulations are adopted whenever computational resources are manageable for the system sizes of interest. This includes simulations of the ideal VD and also noisy VD for the single-copy case, CR, QECR, and up to three copies of BW. All final measurements in the quantum circuits are computed as exact expectation values of observables instead of sampling random shots.

In an ideal VD, only the copies are noisy and the VD part of the circuit is entirely error free. Therefore, the simulation can be performed simply by obtaining the density matrix ρ of one noisy copy and computing $\text{Tr}[O\rho^n]$ without explicitly constructing the entire circuit. The resource required to perform this computation is not problematic for the system sizes ($N = 4, 5, 6$) considered in this work.

The single-copy limit of a noisy VD corresponds to not applying VD, so it can be performed in an ancilla-free fashion using N qubits. After preparing a single copy and storing it in memory, a noisy measurement in the σ basis is performed for all Pauli term σ in the Pauli basis expansion of the observable O . The desired quantity $\text{Tr}[O\rho]$ can then be obtained by combining the results, which can be done easily for the circuit depths considered.

For QECR, 2 registers of N qubits each store copies onto which VD is applied, with the help of a single ancilla qubit

and 2 network qubits. Therefore, the total number of qubits is 15 when $N = 6$, which is manageable. Since the simulations for l and $l + 1$ copies ($l \geq 2$) only differ by an additional preparation of the copy followed by a layer of C-SWAP gates, we store the intermediate result from the l th copy so that the simulation of the $l + 1$ th copy does not need to repeat the simulations for all copies up to the l th. Similarly, the circuit for each l is equivalent up to the final layer of controlled Pauli gates for each term in the observable, so we also store intermediate results within each l to minimize the amount of redundant computation. Since most of the gates of the whole circuit are used to prepare each copy, this significantly reduces the time-complexity of the simulation.

An experimental realization of CR would require a total of $n(N + 1) + 1$ qubits (n registers with $N + 1$ qubits each, including the network qubit, and the ancilla qubit), which is generally much more than the $2N + 3$ qubits needed in QEQR. However, in our model, CR can be simulated using the same amount of resources as QEQR. Since the difference between the two implementations only lie in the errors associated with the preparation of new copies, the simulation of CR can be done by simply switching off these errors that only QEQR suffers from.

The simulation of BW gets increasingly challenging with the number of copies for fixed N . To alleviate the memory requirement of the simulation, we leave out all network qubits that would otherwise be present in a real experiment. Since we only simulate up to five copies, the largest GHZ state that needs to be prepared across the ancilla qubits is a Bell state, which can be prepared directly on the ancilla qubits, so network qubits are not needed for this step. However, errors associated with the other remote operations, namely quantum teleportation and BSMs, must be simulated in slightly modified ways compared to QEQR and CR due to the suppression of network qubits. Since we model an entanglement error as a single-qubit depolarizing channel acting on one side of an ideal Bell state, this error can be simulated by directly applying the same depolarizing channel on the teleported qubit without explicitly performing the teleportation (see Fig. 16 in Ref. [27]). Similarly, having a depolarizing entanglement error of probability p_{Bell} in the remote BSM is identical to having a detection error of probability $p_{\text{Bell}}/3$ after the control qubit and $2p_{\text{Bell}}/3$ after the target qubit of the CNOT gate. While idling errors would have accumulated also on the network qubits, those are ignored in our simulations since they are small when network qubits are initialized after the copies are prepared in parallel and are repeatedly reset and reinitialized for remote operations. After removing network qubits, the quantum circuit involves $nN + \lfloor \frac{n-1}{2} \rfloor$ qubits total, such that the largest quantum circuit for BW that we simulate exactly ($N = 4$ and $n = 3$) involves 13 qubits.

2. Approximate simulations

When exact simulations are not feasible, approximate results can be obtained via Monte Carlo sampling. In this work, we do so for four and five copies of BW. Instead of applying generally nonunitary operations on density matrices as in exact simulations, in this case we only store state vectors, thus quadratically reducing the memory requirement.

Each nonunitary channel involved in the simulation, excluding measurements, can be defined as a linear combination of tensor products of Pauli operators, so the Monte Carlo simulation can be performed by sampling its terms with probabilities corresponding to its coefficients. For example, a single-qubit depolarizing channel of probability λ is defined as

$$\Delta_\lambda = (1 - \lambda)\mathcal{I} + \frac{\lambda}{3}\mathcal{X} + \frac{\lambda}{3}\mathcal{Y} + \frac{\lambda}{3}\mathcal{Z}, \quad (\text{D1})$$

where $\mathcal{I}$, $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ are quantum channels that apply the identity, Pauli X , Y , and Z operators, respectively. In a Monte Carlo simulation, this channel is replaced by an identity channel with probability $1 - \lambda$ and either one of the Pauli X , Y , or Z gates with probability $\lambda/3$.

Given that each of those choices are unitary operations, in a Monte Carlo sampling one needs to only simulate pure states that undergo these random unitary errors. We denote the expected value of the Pauli string σ that we obtain from each random circuit variant as the random variable $A_{\rho,n,\sigma}$ whose mean value is the noise expected value as $\mathbb{E}[A_{\rho,n,\sigma}] = \text{Tr}[\sigma \rho^n]$. Ultimately, we want to estimate the ratio that approaches the expectation value with respect to the dominant eigenvector shown in Eq. (1), as $\frac{\mathbb{E}[A_{\rho,n,O}]}{\mathbb{E}[A_{\rho,n,\mathbb{I}}]}$, where we defined $A_{\rho,n,O} = \sum_i c_i A_{\rho,n,\sigma_i}$.

The inaccuracy of the simulation, quantified by standard errors, can only be suppressed by taking enough samples. In this work, and in particular for error bars in Sec. IV B, the standard error of the biased estimator is estimated by the following procedure: (1) Obtain M samples of $A_{\rho,n,O}$ and $A_{\rho,n,\mathbb{I}}$, where M is chosen as a multiple of 100. (2) Split the samples for $A_{\rho,n,O}$ and $A_{\rho,n,\mathbb{I}}$ into 100 batches each, labeled by $j \in \{1, 2, \dots, 100\}$, and compute the mean values $\mu_O^{(j)}$ and $\mu_{\mathbb{I}}^{(j)}$ within each batch. (3) Take the standard deviation between the 100 ratios $\mu_O^{(j)}/\mu_{\mathbb{I}}^{(j)}$.

To analyze the sampled data, consider the example where $N = 4$, $n = 2$, and $\{p_{1Q}, p_{2Q}, p_{\text{Bell}}\} = \{p_{1Q}^*, p_{2Q}^*, p_{\text{Bell}}^*\}$. The ratios of batches are distributed as shown in the histogram in Fig. 10(a) for $M = 50\,000$. To better visualize the distribution, here we resample from the original dataset to obtain 100 000 batches that each contain $M/100 = 500$ random samples, and obtain ratios of batch means. The dashed vertical red line shows the average value obtained from the sampled data and is situated close to the exact density matrix simulation result, which is the solid vertical line. An agreement with the standard Monte Carlo scaling of $1/\sqrt{M}$ is observed by fitting the standard deviation against 500 to 50 000 number of samples per 100 000 resampled batches, as shown in Fig. 10(b). The y intercept of the linear fit shows that the underlying standard deviation of a single sample is $10^{0.466204} \approx 2.92533$, and the number of samples required to obtain some target error can be extrapolated from the fit.

During the Monte Carlo simulation, many sampled quantum circuits correspond to the ideal quantum circuit for small enough error probabilities. Therefore, we save the ideal case in memory and retrieve it whenever the sampled circuit is ideal to avoid performing redundant simulations. Specifically, we store $|\psi_{\text{id}}\rangle^{\otimes n}$, which is the quantum state corresponding to preparing all the ideal copies in parallel. We also store the

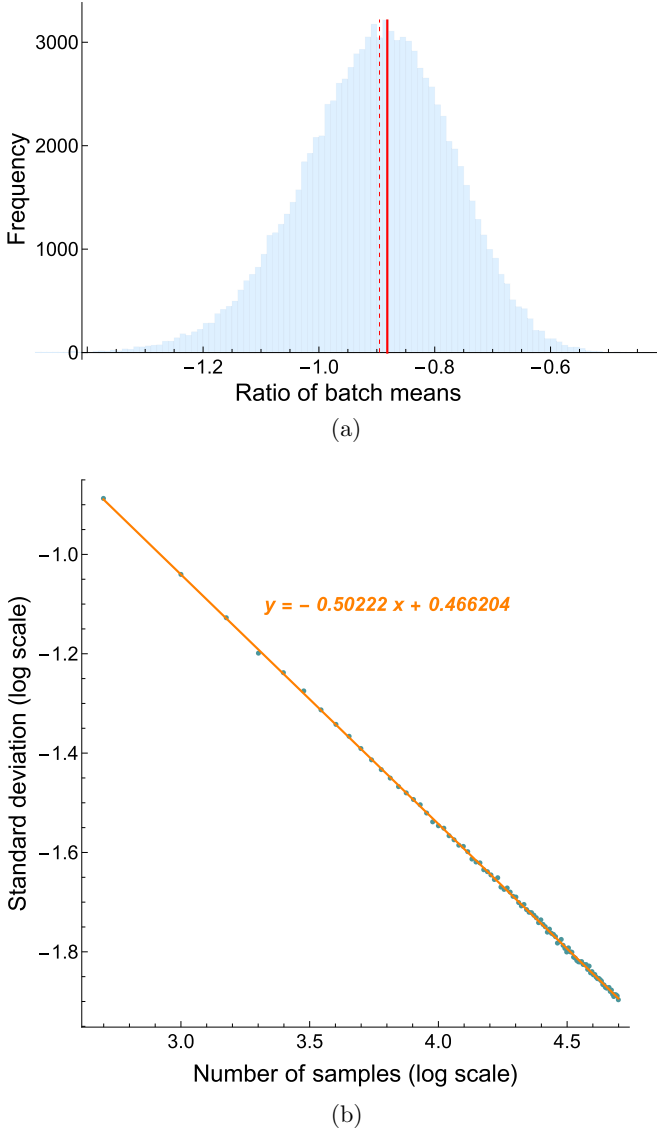


FIG. 10. Analysis of data sampled for Monte Carlo simulation with $N = 4$, $n = 2$, and $\{p_{1Q}, p_{2Q}, p_{\text{Bell}}\} = \{p_{1Q}^*, p_{2Q}^*, p_{\text{Bell}}^*\}$. (a) Distribution of 100 000 ratios of batch means, resampled from a dataset of $M = 50\,000$. The solid and dashed vertical red lines are the exact and approximate simulation results, where the latter case corresponds to the ratio between the average over M samples. (b) Standard deviations (turquoise circles) across 100 000 resampled ratios of batch means, where the number of samples per batch is varied from 500 to 50 000. A linear fit (orange line) in the log scale reveals an agreement with the standard $1/\sqrt{M}$ scaling from Monte Carlo sampling via the slope and the value of the underlying standard deviation via the y-intercept.

ideal results $A_{|\psi_{\text{id}}\rangle, n, \sigma_i}$ from single runs for all terms indexed by i of the observable. Thus, when the entire circuit is ideal, $A_{|\psi_{\text{id}}\rangle, n, \sigma_i}$ is returned without any further computation. If, instead, the copies are ideal but the VD part is noisy, then the prepared ideal copies $|\psi_{\text{id}}\rangle^{\otimes n}$ are retrieved from memory and applied with the noisy VD operations. The entire circuit is simulated otherwise. This is most effective for simulations with low error probabilities, where the probability that the sampled circuit corresponds to the ideal circuit is high.

3. Native-gate decompositions

For all noisy simulations, we decompose quantum gates into those native in the ion trap hardware described in Appendix C. Here we describe how the gates are decomposed, using equal signs to denote equalities up to global phases and order gates temporally such that those on the left are applied before those on the right. Noting that R_z is a virtual Z gate, the Hadamard gate can be applied (on some qubit q) with a single physical gate as $H_q = R_{z_q}(\frac{\pi}{2})R_{x_q}(\frac{\pi}{2})R_{z_q}(\frac{\pi}{2})$. The controlled Pauli gates that correspond to the terms of the observable are each decomposed into single Rzz gates as

$$\begin{aligned} CX_q^p &= R_{z_p}(\frac{\pi}{2})R_{z_q}(\frac{\pi}{2})R_{x_q}(\frac{\pi}{2})R_{z_q}(\pi)R_{zz_q^p}(\frac{-\pi}{2})H_q \\ CY_q^p &= R_{z_p}(\frac{\pi}{2})R_{x_q}(\frac{\pi}{2})R_{z_q}(\frac{3\pi}{2}) \\ &\quad \times R_{zz_q^p}(\frac{-\pi}{2})R_{x_q}(\frac{\pi}{2})R_{z_q}(\pi) \\ CZ_q^p &= R_{z_p}(\frac{\pi}{2})R_{z_q}(\frac{\pi}{2})R_{zz_q^p}(\frac{-\pi}{2}), \end{aligned} \quad (\text{D2})$$

where p and q are the control and target qubits of the controlled Pauli gates, respectively. A decomposition of the C-SWAP gate was obtained using a recompiler, as described in detail in Ref. [9]. The decomposition of the C-SWAP gate is shown in Fig. 11 for the native gates considered in this work and includes six Rzz gates.

4. Artificial gates

In all numerical simulations that involve explicit simulations of the quantum circuit, the final measurement outcomes are obtained by taking expectation values of observables instead of sampling random shots as would be done in real experiments. For example, considering that we have applied single-qubit gates in the end such that all measurements are in the Z basis, BW requires a measurement of the observable

$$\bigotimes_i Z_i \bigotimes_j (1 - Z_{j_1})(1 - Z_{j_2}), \quad (\text{D3})$$

where i runs over all $\lfloor \frac{n-1}{2} \rfloor$ indices of the ancilla qubits of BW and $j = (j_1, j_2)$ are $\lfloor \frac{n}{2} \rfloor N$ pairs of indices corresponding to the two qubits involved in individual BSMs. This shows that the number of terms of the observable is $4^{\lfloor \frac{n}{2} \rfloor N}$, which becomes prohibitive for the larger choices of n in this work.

To overcome this issue, we introduce instantaneous and noiseless artificial gates, which do not contribute to the main circuit but are inserted for the sole purpose of converting the measurement observable into a single Pauli string in the numerical simulations. The artificial gates are Toffoli gates (or CCX gates) applied after each conversion to the Bell basis and any one of the ancilla qubits, such that the outcome of the ancilla qubit is flipped when both outcomes of the BSM are 1. Note that this can be seen as applying the final layer of C-SWAP gates as in Fig. 9(d) instead of reducing it further down to Fig. 9(f). After applying these artificial gates, the full observable in Eq. (D3) would not be needed, and a single Pauli string $\bigotimes_i Z_i$ would be sufficient.

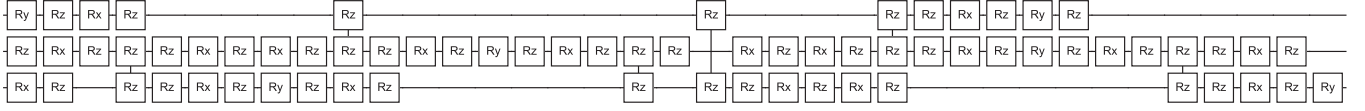


FIG. 11. C-SWAP gate decomposed into native gates in the ion trap architecture considered in this work (see Appendix C), where rotation angles are all integer multiples of $\pi/4$.

For QEQR and CR, the observable without any artificial gates is given by

$$Z_\alpha \bigotimes_j (1 - Z_{j_{q1}} Z_{j_{a2}}) (1 - Z_{j_{a1}} Z_{j_{q2}}), \quad (\text{D4})$$

where α is the index for the ancilla qubit and $j = (j_{q1}, j_{a1}, j_{q2}, j_{a2})$ correspond to all four qubits involved in the j th remote BSM according to the notation in Fig. 5(b).

To convert the factor for BSMs in Eq. (D4) into the same form as that in Eq. (D3), we store the parities between qubits q_1 and a_2 into q_1 and between a_1 and q_2 into q_2 . This is done by including artificial CNOT gates controlled by a_2 (a_1) and targeted to q_1 (q_2) for each of the remote BSMs.

While the number of terms for simulations of QEQR and CR is only 4^N due to the fact that only a single set of C-SWAP gates can be replaced by BSMs, applying the artificial Toffoli gates is still crucial. This is due to the complication that network qubits are reinitialized after being used for each remote BSM, while our measurement of Eq. (D4) requires that all qubits involved be stored in memory after being prepared in the appropriate bases. By inserting the artificial Toffoli gates, along with the artificial CNOT gates, the observable simplifies to Z_α , so that a final measurement of the ancilla qubit is sufficient.

5. Random-field Heisenberg model and Trotterization

We consider the example problem to be a one-dimensional spin chain with nearest-neighbor interactions and random magnetic fields, known as the random-field Heisenberg model [50,51]. Its Hamiltonian is generally given by

$$\mathcal{H} = J \sum_{j=1}^N \vec{S}_j \cdot \vec{S}_{j+1} + \sum_{j=1}^N h_j Z_j, \quad (\text{D5})$$

where $\vec{S}_j = (X_j, Y_j, Z_j)$ denotes a vector of Pauli operators acting on site j , J is an interaction strength assumed to be uniform across all pairs of interacting spins, and h_j are random magnetic field strengths that are extracted uniformly from the

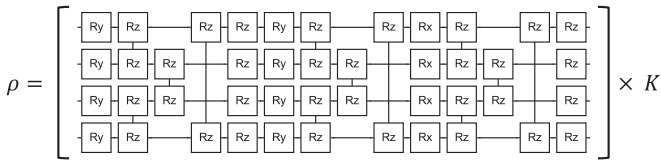


FIG. 12. Quantum circuit that prepares each copy of ρ considered in this work when applied onto $|\psi_{\text{init}}\rangle$, shown for $N = 4$. All gates, which are noisy except for the (virtual) Rz gates, are repeated K times, corresponding to K Trotter steps. The preparation of $|\psi_{\text{init}}\rangle$ can be combined with the first layer of Ry gates in the first Trotter steps, since $|\psi_{\text{init}}\rangle$ is a computational basis state.

interval $[-h, h]$. We impose periodic boundary conditions, i.e., site $N + 1$ is identified with site 1, and choose $J = h = 1$. Denoting the N -dimensional vector of random field strengths as $\vec{h}^{(N)}$, their values in the three system sizes studied in Sec. IV B are

$$\begin{aligned} \vec{h}^{(4)} &= (-0.887, -0.925, -0.72, 0.08) \\ \vec{h}^{(5)} &= (0.206, -0.649, 0.598, -0.826, 0.702) \\ \vec{h}^{(6)} &= (-0.859, 0.396, -0.354, 0.634, -0.893, 0.198) \end{aligned} \quad (\text{D6})$$

to three decimal places.

The ideal state $|\psi_{\text{id}}\rangle$ corresponds to a time evolution under this Hamiltonian, approximated with the Lie-Trotter-Suzuki formula [52,53]. In this context, an evolution of duration t is separated into K uniform time steps $\delta t = t/K$, within which the time-evolution operator generated by $\mathcal{H}$ can approximately be split into simple quantum rotation gates. Suppose we know how to write the Hamiltonian as $\mathcal{H} = \sum_l \mathcal{H}_l$, where all terms in $\mathcal{H}_l$ in the Pauli basis commute with each other, but $[\mathcal{H}_{l_1}, \mathcal{H}_{l_2}] \neq 0$ for $l_1 \neq l_2$. The ideal state would then be

$$|\psi_{\text{id}}\rangle = \left(\prod_l e^{-i\mathcal{H}_l \delta t} \right)^K |\psi_{\text{init}}\rangle \approx e^{-i\mathcal{H}t} |\psi_{\text{init}}\rangle. \quad (\text{D7})$$

The initial state $|\psi_{\text{init}}\rangle$ is chosen to be the computational basis state with only site 3 being $|1\rangle$ for all N considered, and we fix $\delta t = 0.01$ such that longer time evolutions are simulated by increasing K . The algorithmic error in Eq. (D7) depends

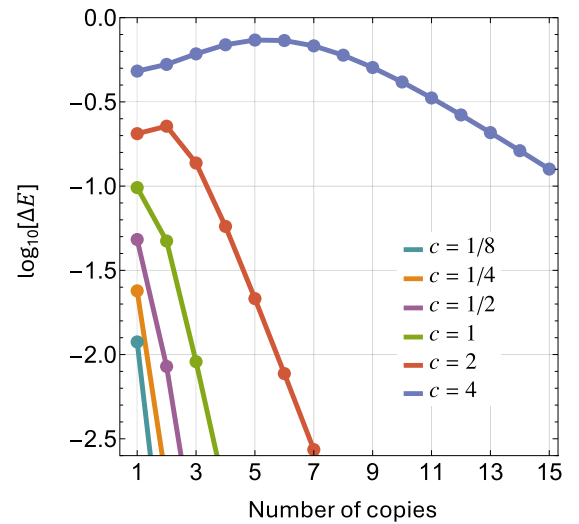


FIG. 13. Ideal implementation of VD with $N = 5$ simulated with the same parameter choices as in Sec. IV B but up to 15 copies. Even though, as in the corresponding plot in Fig. 6, the absolute error ΔE in the $c = 4$ case increases in the range $n \in \{1, 2, 3, 4, 5\}$, it starts to decrease as n is increased further.

on δt , but its choice is unimportant for this work, since we are only interested in suppressing physical errors.

The Hamiltonian in Eq. (D5) can be grouped into sums of mutually commuting Pauli terms by splitting the sites into an even and odd part. This gives rise to a Trotterized quantum circuit consisting of three layers of two-qubit gates (for three interaction directions) interleaved by single-qubit rotations, as shown in Fig. 12 for $N = 4$ after compiling for the ion trap quantum hardware described in Appendix C. Since the initial state is a computational basis state, its preparation is combined with the first layer of Ry gates in the first Trotter step. Within each Trotter step, all single-qubit gates except for the final layer of Rz gates are rotations either by $\frac{\pi}{2}$ or $-\frac{\pi}{2}$. The Rzz gates all share the angle $2J\delta t$, and the final layer of Rz gates have angles given by the random field strengths, $2h_j\delta t$ for $j \in \{1, 2, \dots, N\}$. In line with the applicability of QEM techniques, we choose $K = \lfloor 1/(3Np_{2Q}^*) \rfloor$ such that the circuit error rate of each copy is approximately 1, where $3N$ is the number of two-qubit gates in a single Trotter step.

6. Performance of VD for $N = 5$ and $c = 4$ up to 15 copies

In practice, increasing the number of copies leads to an exponential increase in the sampling cost, such that a value of $n \gtrsim 10$ may be considered prohibitive. Nevertheless, we may numerically simulate the effect of VD for higher number of copies under an infinite number of circuit runs to study the behavior of the $N = 5$ and $c = 4$ case in Fig. 6, for which the absolute error with the ideal expectation value increases as the number of copies increase in $n \in \{1, 2, 3, 4, 5\}$. With the same parameter choices as in Sec. IV B, we simulate the ideal implementation of VD up to 15 copies, and the result is shown in Fig. 13. As we can see, the absolute error begins to decrease from the 6th copy and falls below the $n = 1$ case beyond the 10th copy. This shows that even in this high error rate, the ideal state remains to be the dominant component, but it is less dominant than the rest of the values of c such that the exponential error suppression was not observed from the low number of copies considered in Fig. 6.

-
- [1] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).
 - [2] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, *Rev. Mod. Phys.* **94**, 015004 (2022).
 - [3] Z. Zimborás, B. Koczor, Z. Holmes, E.-M. Borrelli, A. Gilyén, H.-Y. Huang, Z. Cai, A. Acín, L. Aolita, L. Bianchi, F. G. S. L. Brandão, D. Cavalcanti, T. Cubitt, S. N. Filippov, G. García-Pérez, J. Goold, O. Kálmán, E. Kyoseva, M. A. C. Rossi, B. Sokolov *et al.*, Myths around quantum computation before full fault tolerance: What no-go theorems rule out and what they don't, *arXiv:2501.05694*.
 - [4] C. Piveteau, D. Sutter, S. Bravyi, J. M. Gambetta, and K. Temme, Error mitigation for universal gates on encoded qubits, *Phys. Rev. Lett.* **127**, 200505 (2021).
 - [5] Y. Suzuki, S. Endo, K. Fujii, and Y. Tokunaga, Quantum error mitigation as a universal error reduction technique: Applications from the NISQ to the fault-tolerant quantum computing eras, *PRX Quantum* **3**, 010345 (2022).
 - [6] Y. Akahoshi, K. Maruyama, H. Oshima, S. Sato, and K. Fujii, Partially fault-tolerant quantum computing architecture with error-corrected Clifford gates and space-time efficient analog rotations, *PRX Quantum* **5**, 010337 (2024).
 - [7] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, *Rev. Mod. Phys.* **95**, 045005 (2023).
 - [8] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, Virtual distillation for quantum error mitigation, *Phys. Rev. X* **11**, 041036 (2021).
 - [9] B. Koczor, Exponential error suppression for near-term quantum devices, *Phys. Rev. X* **11**, 031057 (2021).
 - [10] In practice, even if identical gates are applied, the copies may not be identically prepared due to variations in the noise channels. VD remains effective as long as $|\psi\rangle$ is the dominant eigenvector across all copies, and its performance is limited by the lowest dominant eigenvalue amongst the copies as shown in Ref. [9].
 - [11] B. Koczor, The dominant eigenvector of a noisy quantum state, *New J. Phys.* **23**, 123047 (2021).
 - [12] M. Silva, E. Magesan, D. W. Kribs, and J. Emerson, Scalable protocol for identification of correctable codes, *Phys. Rev. A* **78**, 012347 (2008).
 - [13] E. Magesan, J. M. Gambetta, and J. Emerson, Characterizing quantum gates via randomized benchmarking, *Phys. Rev. A* **85**, 042311 (2012).
 - [14] Z. Cai and S. C. Benjamin, Constructing smaller Pauli twirling sets for arbitrary error channels, *Sci. Rep.* **9**, 11281 (2019).
 - [15] Z. Cai, X. Xu, and S. C. Benjamin, Mitigating coherent noise using Pauli conjugation, *npj Quantum Inf.* **6**, 17 (2020).
 - [16] N. Yoshioka, H. Hakoshima, Y. Matsuzaki, Y. Tokunaga, Y. Suzuki, and S. Endo, Generalized quantum subspace expansion, *Phys. Rev. Lett.* **129**, 020502 (2022).
 - [17] B. Yang, N. Yoshioka, H. Harada, S. Hakkaku, Y. Tokunaga, H. Hakoshima, K. Yamamoto, and S. Endo, Resource-efficient generalized quantum subspace expansion, *Phys. Rev. Appl.* **23**, 054021 (2025).
 - [18] B. Koczor, Sparse probabilistic synthesis of quantum operations, *PRX Quantum* **5**, 040352 (2024).
 - [19] B. Koczor, J. J. L. Morton, and S. C. Benjamin, Probabilistic interpolation of quantum rotation angles, *Phys. Rev. Lett.* **132**, 130602 (2024).
 - [20] M. Huo and Y. Li, Dual-state purification for practical quantum error mitigation, *Phys. Rev. A* **105**, 022427 (2022).
 - [21] Z. Cai, Resource-efficient purification-based quantum error mitigation, *arXiv:2107.07279*.
 - [22] P. Czarnik, A. Arrasmith, L. Cincio, and P. J. Coles, Qubit-efficient exponential suppression of errors, *arXiv:2102.06056*.
 - [23] D. Gottesman and I. L. Chuang, Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations, *Nature (London)* **402**, 390 (1999).
 - [24] C. Monroe, R. Raussendorf, A. Ruthven, K. R. Brown, P. Maunz, L.-M. Duan, and J. Kim, Large-scale modular quantum-

- computer architecture with atomic memory and photonic interconnects, *Phys. Rev. A* **89**, 022317 (2014).
- [25] L. Jiang, J. M. Taylor, A. S. Sørensen, and M. D. Lukin, Distributed quantum computation based on small quantum registers, *Phys. Rev. A* **76**, 062323 (2007).
- [26] M. Caleffi, M. Amoretti, D. Ferrari, J. Illiano, A. Manzalini, and A. S. Cacciapuoti, Distributed quantum computing: A survey, *Comput. Netw.* **254**, 110672 (2024).
- [27] H. Jnane, B. Undseth, Z. Cai, S. C. Benjamin, and B. Koczor, Multicore quantum computing, *Phys. Rev. Appl.* **18**, 044064 (2022).
- [28] J. Dénes, The representation of a permutation as the product of a minimal number of transpositions and its connection with the theory of graphs, *Publ. Math. Inst. Hung. Acad. Sci.* **4**, 63 (1959).
- [29] J.-M. Liang, Q.-Q. Lv, Z.-X. Wang, and S.-M. Fei, Unified multivariate trace estimation and quantum error mitigation, *Phys. Rev. A* **107**, 012606 (2023).
- [30] H. Hakoshima, S. Endo, K. Yamamoto, Y. Matsuzaki, and N. Yoshioka, Localized virtual purification, *Phys. Rev. Lett.* **133**, 080601 (2024).
- [31] Y. Quek, E. Kaur, and M. M. Wilde, Multivariate trace estimation in constant quantum depth, *Quantum* **8**, 1220 (2024).
- [32] T. Häner, D. S. Steiger, T. Hoeffler, and M. Troyer, Distributed quantum computing with QMPI, in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (Association for Computing Machinery, New York, 2021).
- [33] D. Main, P. Drmota, D. P. Nadlinger, E. M. Ainley, A. Agrawal, B. C. Nichol, R. Srinivas, G. Araneda, and D. M. Lucas, Distributed quantum computing across an optical network link, *Nature (London)* **638**, 383 (2025).
- [34] L. J. Stephenson, D. P. Nadlinger, B. C. Nichol, S. An, P. Drmota, T. G. Ballance, K. Thirumalai, J. F. Goodwin, D. M. Lucas, and C. J. Ballance, High-rate, high-fidelity entanglement of qubits across an elementary quantum network, *Phys. Rev. Lett.* **124**, 110501 (2020).
- [35] V. Krutyanskiy, M. Galli, V. Krcmarsky, S. Baier, D. A. Fioretto, Y. Pu, A. Mazloom, P. Sekatski, M. Canteri, M. Teller, J. Schupp, J. Bate, M. Meraner, N. Sangouard, B. P. Lanyon, and T. E. Northup, Entanglement of trapped-ion qubits separated by 230 meters, *Phys. Rev. Lett.* **130**, 050803 (2023).
- [36] S. Saha, M. Shalae, J. O'Reilly, I. Goetting, G. Toh, A. Kalakuntla, Y. Yu, and C. Monroe, High-fidelity remote entanglement of trapped atoms mediated by time-bin photons, *Nat. Commun.* **16**, 2533 (2025).
- [37] S. de Bone, R. Ouyang, K. Goodenough, and D. Elkouss, Protocols for creating and distilling multipartite GHZ states with Bell pairs, *IEEE Trans. Quant. Eng.* **1**, 1 (2020).
- [38] E. M. Ainley, A. Agrawal, D. Main, P. Drmota, D. P. Nadlinger, B. C. Nichol, R. Srinivas, and G. Araneda, Multipartite entanglement for multi-node quantum networks, [arXiv:2408.00149](https://arxiv.org/abs/2408.00149).
- [39] N. H. Nickerson, Y. Li, and S. C. Benjamin, Topological quantum computing with a very noisy network and local error rates approaching one percent, *Nat. Commun.* **4**, 1756 (2013).
- [40] N. H. Nickerson, J. F. Fitzsimons, and S. C. Benjamin, Freely scalable quantum technologies using cells of 5-to-50 qubits with very lossy and noisy photonic links, *Phys. Rev. X* **4**, 041041 (2014).
- [41] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, Purification of noisy entanglement and faithful teleportation via noisy channels, *Phys. Rev. Lett.* **76**, 722 (1996).
- [42] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Mixed-state entanglement and quantum error correction, *Phys. Rev. A* **54**, 3824 (1996).
- [43] W. Dür and H.-J. Briegel, Entanglement purification for quantum computation, *Phys. Rev. Lett.* **90**, 067901 (2003).
- [44] K. Yamamoto, Y. Matsuzaki, Y. Suzuki, Y. Tokunaga, and S. Endo, Virtual entanglement purification via noisy entanglement, [arXiv:2411.10024](https://arxiv.org/abs/2411.10024).
- [45] E. Campbell, Random compiler for fast Hamiltonian simulation, *Phys. Rev. Lett.* **123**, 070503 (2019).
- [46] C. Kiumi and B. Koczor, TE-PAI: Exact time evolution by sampling random circuits, *Quantum Sci. Technol.* **10**, 045071 (2025).
- [47] V. Kliuchnikov, K. Lauter, R. Minko, A. Paetznick, and C. Petit, Shorter quantum circuits via single-qubit gate approximation, *Quantum* **7**, 1208 (2023).
- [48] P. Selinger, Quantum circuits of T -depth one, *Phys. Rev. A* **87**, 042302 (2013).
- [49] M. Amy, D. Maslov, M. Mosca, and M. Roetteler, A meet-in-the-middle algorithm for fast synthesis of depth-optimal quantum circuits, *IEEE Trans. Comput.-Aided Des. Integr. Circ. Syst.* **32**, 818 (2013).
- [50] D. J. Luitz, N. Laflorencie, and F. Alet, Many-body localization edge in the random-field Heisenberg chain, *Phys. Rev. B* **91**, 081103(R) (2015).
- [51] 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. USA* **115**, 9456 (2018).
- [52] H. F. Trotter, On the product of semi-groups of operators, *Proc. Am. Math. Soc.* **10**, 545 (1959).
- [53] M. Suzuki, Generalized trotter's formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems, *Commun. Math. Phys.* **51**, 183 (1976).
- [54] Any residual coherent contribution from single-qubit gates can be converted into incoherent errors by the application of twirling techniques.
- [55] A. S. Sotirova, J. D. Leppard, A. Vazquez-Brennan, S. M. Decoppet, F. Pokorny, M. Malinowski, and C. J. Ballance, High-fidelity heralded quantum state preparation and measurement, [arXiv:2409.05805](https://arxiv.org/abs/2409.05805).
- [56] Y. Seis, B. J. Brown, A. S. Sørensen, and J. F. Goodwin, Improving trapped-ion-qubit memories via code-mediated error-channel balancing, *Phys. Rev. A* **107**, 052417 (2023).
- [57] T. Jones, A. Brown, I. Bush, and S. C. Benjamin, QuEST and high performance simulation of quantum computers, *Sci. Rep.* **9**, 10736 (2019).
- [58] T. Jones and S. Benjamin, QuESTlink—Mathematica embellished by a hardware-optimised quantum emulator, *Quantum Sci. Technol.* **5**, 034012 (2020).
- [59] A. Kay, Tutorial on the Quantikz package, [arXiv:1809.03842](https://arxiv.org/abs/1809.03842).
- [60] A. Richards, University of Oxford Advanced Research Computing, Zenodo (2015), doi:[10.5281/zenodo.22558](https://doi.org/10.5281/zenodo.22558).
- [61] J. C. Garcia-Escartin and P. Chamorro-Posada, Swap test and Hong-Ou-Mandel effect are equivalent, *Phys. Rev. A* **87**, 052330 (2013).

- [62] L. Cincio, Y. Subaşı, A. T. Sornborger, and P. J. Coles, Learning the quantum algorithm for state overlap, *New J. Phys.* **20**, 113022 (2018).
- [63] C. Langer, R. Ozeri, J. D. Jost, J. Chiaverini, B. DeMarco, A. Ben-Kish, R. B. Blakestad, J. Britton, D. B. Hume, W. M. Itano, D. Leibfried, R. Reichle, T. Rosenband, T. Schaetz, P. O. Schmidt, and D. J. Wineland, Long-lived qubit memory using atomic ions, *Phys. Rev. Lett.* **95**, 060502 (2005).
- [64] 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).
- [65] M. A. Sepiol, A. C. Hughes, J. E. Tarlton, D. P. Nadlinger, T. G. Ballance, C. J. Ballance, T. P. Harty, A. M. Steane, J. F. Goodwin, and D. M. Lucas, Probing qubit memory errors at the part-per-million level, *Phys. Rev. Lett.* **123**, 110503 (2019).
- [66] Y. Wang, M. Um, J. Zhang, S. An, M. Lyu, J.-N. Zhang, L.-M. Duan, D. Yum, and K. Kim, Single-qubit quantum memory exceeding ten-minute coherence time, *Nat. Photon.* **11**, 646 (2017).
- [67] Y. Wang, S. Crain, C. Fang, B. Zhang, S. Huang, Q. Liang, P. H. Leung, K. R. Brown, and J. Kim, High-fidelity two-qubit gates using a microelectromechanical-system-based beam steering system for individual qubit addressing, *Phys. Rev. Lett.* **125**, 150505 (2020).
- [68] D. Kielpinski, C. Monroe, and D. J. Wineland, Architecture for a large-scale ion-trap quantum computer, *Nature (London)* **417**, 709 (2002).
- [69] C. J. Ballance, T. P. Harty, N. M. Linke, M. A. Sepiol, and D. M. Lucas, High-fidelity quantum logic gates using trapped-ion hyperfine qubits, *Phys. Rev. Lett.* **117**, 060504 (2016).
- [70] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, Efficient Z gates for quantum computing, *Phys. Rev. A* **96**, 022330 (2017).
- [71] C. F. Roos, Ion trap quantum gates with amplitude-modulated laser beams, *New J. Phys.* **10**, 013002 (2008).
- [72] J. P. Gaebler, T. R. Tan, Y. Lin, Y. Wan, R. Bowler, A. C. Keith, S. Glancy, K. Coakley, E. Knill, D. Leibfried, and D. J. Wineland, High-fidelity universal gate set for $^9\text{Be}^+$ ion qubits, *Phys. Rev. Lett.* **117**, 060505 (2016).
- [73] V. M. Schäfer, C. J. Ballance, K. Thirumalai, L. J. Stephenson, T. G. Ballance, A. M. Steane, and D. M. Lucas, Fast quantum logic gates with trapped-ion qubits, *Nature (London)* **555**, 75 (2018).
- [74] J. Schupp, V. Krcmarsky, V. Krutyanskiy, M. Meraner, T. E. Northup, and B. P. Lanyon, Interface between trapped-ion qubits and traveling photons with close-to-optimal efficiency, *PRX Quantum* **2**, 020331 (2021).