End-to-End Optimization of Partial Fault Tolerance in Neutral-Atom Arrays

Stefano Veroni,* Aydin Deger, and Andrew J. Daley
Department of Physics, Clarendon Laboratory, University of Oxford, Oxford OX1 3PU, United Kingdom
(Dated: October 19, 2026)

Fault-tolerant (FT) quantum computation offers a pathway to quantum advantage, but its realization is hindered by its significant overhead. The partially FT "STAR" architecture reduces the cost of universal computation at the price of relaxing FT requirements for non-Clifford operations, increasing exposure to hardware noise. Therefore, maximizing its performance and applicability demands careful co-design with the underlying hardware. In this work we investigate the co-design of the STAR architecture with neutral-atom platforms, connecting the underlying atomic physics to logical error modeling and application-level performance. At the physical level, we identify a favorable noise bias in ^{171}Yb Rydberg gates, and show how it can be enhanced through error-selective pulse optimization. At the logical level, we identify an intrinsic error suppression mechanism, yielding vanishing logical coherent noise in leading order. Detailed circuit-level simulations of surface code implementations show more than an order of magnitude reduction in logical error relative to previous estimates. As a concrete target application, we focus on the Fermi-Hubbard model with periodic boundaries, for which we design a simulation circuit particularly suited to early implementations. We quantify the resource tradeoff of partially and fully FT approaches to gate synthesis, with the former showing approximately an order of magnitude reduction in state-injection depth, and an even larger reduction in Clifford resources. Comparing with classical and fully FT approaches, we identify regimes where partial fault tolerance provides both lower overheads and lower error rates. In our estimates, this is accessible with moderate error rates $p \sim 10^{-3}$ and a few thousand qubits, underlining that for these types of problems, partial FT approaches provide a compelling pathway to early quantum advantage.

I. INTRODUCTION

Large-scale quantum advantage is expected to require quantum error correction (QEC) to overcome the inherently noisy nature of quantum devices. QEC encodes logical information across multiple physical qubits, and repeatedly detects physical faults by means of syndrome measurements [1]. Remarkable progress in QEC has been demonstrated across different architectures, such as superconducting qubits [2–6], trapped ions [7–10], nitrogen vacancies in diamonds [11], and neutral-atom arrays [12–18]. Critical to the success of large-scale quantum computation is the concept of fault tolerance. A protocol is fault-tolerant (FT) if it constrains the propagation of errors, thus suppressing noise, provided that the hardware error rate is lower than a threshold value. Fault tolerance is essential not only to error correction, but also to the execution of logical operations.

Transversal logical gates, which do not couple physical qubits within the same code block, prevent uncontrolled error propagation and are naturally FT. However, no QEC code admits a gate set that is both transversal and universal [19]. Several methods have therefore been proposed to enable the FT realization of a universal gate set [20–23], but they come with significant space-time overhead. In particular, arbitrary operations are synthesized by tens of non-Clifford gates [24], each considerably more demanding than transversal operations. To reduce this overhead, *partially* fault-tolerant computation has recently been proposed [25–30]. In the so-called "Space-Time efficient Analog Rotation" (STAR) architecture, the execution of Clifford gates remains FT, whereas arbitrary-angle rotations are natively implemented rather than synthesized via tens of FT gates. This overhead reduction comes at the expense of reduced resilience against specific protocol-dependent faults, making accurate modeling and co-design with the underlying hardware indispensable for its success. Among the most promising hardware approaches, neutral-atom architectures have recently demonstrated a rich collection of desirable features, including mid-circuit measurements and reconfigurable non-local connectivity [31]. These enable fast transversal logical gates, hence a significant reduction in QEC overhead [32–35], and have recently been incorporated within the framework of partial fault tolerance, in the so-called transversal STAR architecture [28].

In our work, we study the co-design of partially FT computation with neutral-atom architectures [31], spanning gate-pulse optimization, analysis of logical operations, and application-layer design, taking the Fermi-Hubbard model as representative case-study. Altogether, our work opens a near-term regime in which partially FT computation surpasses the reach of classical and NISQ simulations, as well as that of fully FT implementations at comparable hardware size. More precisely, we show that it improves over standard FT approaches not only by yielding shorter circuit depth, but also by achieving lower logical noise, thereby enabling simulations of larger system sizes. The STAR architecture thus emerges not only as a lower-overhead alternative, but as an all-round compelling route for early quantum advantage. Concretely, we first uncover intrinsic error-suppression mechanisms

* stefano.veroni@physics.ox.ac.uk

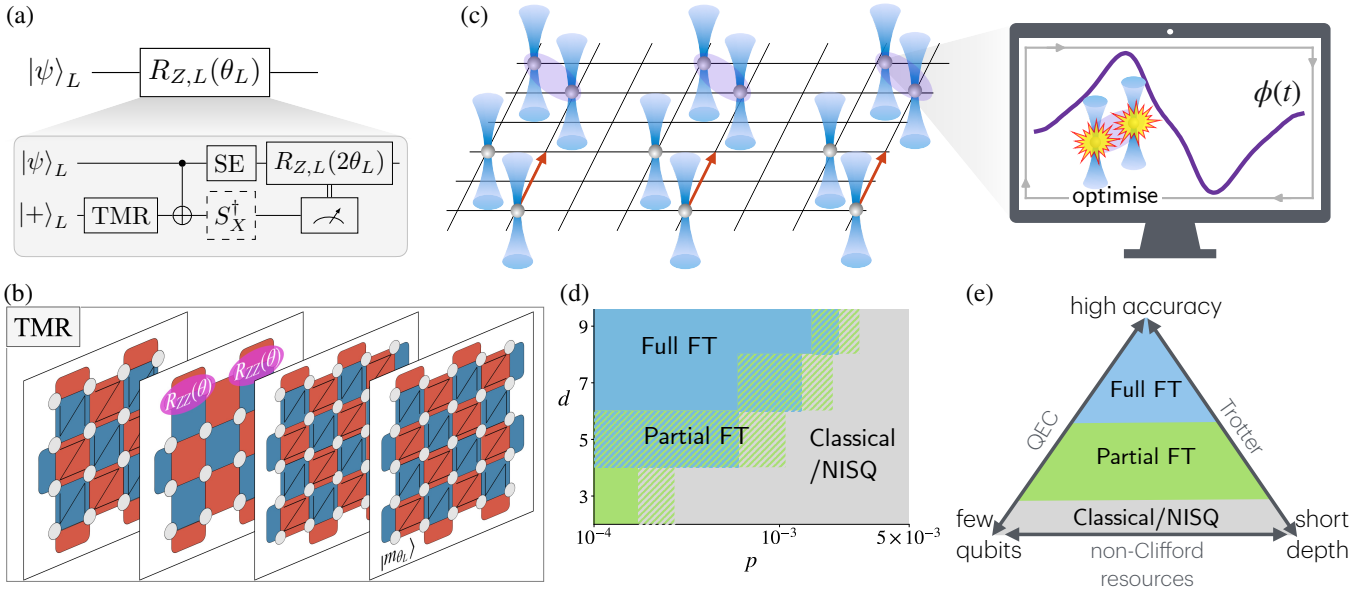


Figure 1. Overview. (a) Partially fault-tolerant realization of arbitrary-angle $R_{Z,L}(\theta_L)$ rotations. An angle-dependent resource state $|m_{\theta_L}\rangle_L$ is encoded via the Transversal Multi-Rotation (TMR) protocol, then gate teleportation is performed and eventually repeated until the target measurement is obtained. The Clifford $S_X^\dagger$ is required if an even number of transversal rotations is applied in the TMR stage. (b) TMR encoding of the $|m_{\theta_L}\rangle_L$ resource state in the rotated surface code. First, a logical patch is encoded with code distance d_s; then r multi-qubit transversal rotations by a suitable angle are applied; finally, other rounds of syndrome extraction are performed, growing the code distance if needed. Post-selection is used upon nontrivial syndromes. In this example, $d_s = 4$, $r = 2$ and the final code distance is $d = 5$. (c) Operation and optimization of a neutral atom device. Shuttling enables non-local connectivity and transversal gates. Pulse-level optimization of gates under realistic noise models is performed to increase the logical performance of the scheme. (d) Regimes of advantage of different computational paradigms, in terms of code distance d and physical error rate p, taking time evolution in the Fermi-Hubbard model as concrete benchmark. Green and blue denote regions where partial and full fault tolerance, respectively, yield lower error, and are capable of simulating larger systems than classical/NISQ alternatives with a contained error mitigation overhead. The striped region highlights the extended near-term regime, opened by the improvements detailed in the manuscript, in which the transversal STAR architecture is not only a lower-overhead alternative to full fault-tolerance, but also yields the lowest errors. The background color behind the green stripes indicates the alternative method that would have yielded lower errors under previous estimates [28]. (e) Trilemma of quantum computation for Trotterised evolution, where high accuracy, short depth, and low qubit counts are mutually exclusive. A balance is struck by tuning QEC approach, Trotterisation, and availability of non-Clifford resources.

in partially FT computation, elucidating further noise reduction approaches. Second, we tailor partially FT computation to neutral-atom architectures by identifying a favorable noise bias in the relevant gates, and increasing it through pulse-level optimization. Third, focusing on time-evolution in the Fermi-Hubbard model with periodic boundaries, we design a simulation circuit particularly suited to early implementations. Then we assess the performance of partially FT computation, both by investigating its resilience to logical noise, and by quantifying its tradeoff between physical resources and circuit depth.

The manuscript is organized as follows. Sec. II provides a concise outline of our approach and highlights our main findings. In Sec. III, we introduce partially fault-tolerant computation, discuss its dominant error mechanisms, and develop a detailed model of error propagation in the protocol. In Sec. IV, we specialize the analysis to neutral-atom architectures, characterize the noise properties of Rydberg gates, and improves their performances with pulse-level optimization. In Sec. V, we present simulations under circuit-level noise of partially FT computation with the rotated surface code, and quantify the resulting logical performance. In Sec. VI, we compare the robustness and resource requirements of the partially and fully FT approaches, and examine the consequences for near-term quantum simulation of the Hubbard model.

II. OVERVIEW AND KEY RESULTS

This section provides a summary of our approach and results in the co-design of partial fault-tolerance with neutral-atom architectures, spanning the physical, logical and application layers.

First, we develop an accurate model of error propagation in partially FT operations, independent of the underlying hardware. In the STAR architecture, logical rotations $R_{Z,L}$ are natively implemented via injection of a suitably encoded resource state. We provide a detailed enumeration of leading noise sources and their impact on the injection and state preparation, shown in Figs. 1(a) and 1(b), revealing that the protocol is naturally resilient to physical coherent errors—often the dominant contri-

butions in non-Clifford operations. Moreover, we find that the resulting logical noise has a vanishing coherent component in leading order. These observations are critical to a correct characterization of noise in partially FT operations, and the development of suitable mitigation approaches.

Secondly, we analyze and improve the robustness of the protocol in the concrete setting of neutral-atom arrays, schematically depicted in Fig. 1(c), with a particular focus on ^{171}Yb implementations. We show that Rydberg gates—to which partially FT computation is particularly sensitive in this setting—exhibit noise scaling and bias that are particularly favorable for this approach. Then, as sketched in Fig. 1(c), we optimize these gates at the pulse level to reduce the rate of undetectable errors, yielding a further two-fold improvement in logical performance with no additional overhead.

Third, we perform full circuit-level simulations of the logical protocol in the rotated surface code. For this purpose, we develop a tailored simulation method that captures the effects of non-Clifford operations and noise in the protocol, and propagates them through multiple iterations of the scheme. These simulations demonstrate more than one order of magnitude reduction in logical error relative to previous estimates [27, 28].

Finally, we assess the implications for near-term quantum simulation. As a concrete setting, we focus on time evolution in the Hubbard model on a square lattice with periodic boundaries, with evolution time growing linearly with lattice length. We evaluate implementations with the rotated surface code in neutral-atom architectures, accounting for the cost of non-Clifford resources, as well as of transversal Clifford operations. The latter are found to have non-negligible impact at moderate distances, hence we design a new circuit implementation tailored to early FT approaches by lowering the number of Clifford gates, exploiting a previously proposed construction [36, 37]. First, our analysis quantifies the accumulation of logical errors. Quantum simulations are assumed to be supported by error mitigation, hence are considered practically feasible when the sum of gates' error rates is below unity. The reach of classical methods for simulating the Hubbard model is understood to be somewhat below a system size of a 10×10 lattice for dynamics that spread information over the whole system [37, 38].

We identify regimes of advantage of partially FT, fully FT and classical computation, as summarized in Fig. 1(d). For comparison purposes, assuming periodic boundaries and evolution time comparable to the length of the lattice, we place the limit of current classical methods at 6×6 sites, as indicated by Refs. [38, 39]. Remarkably, the improvements obtained in our framework open up a near-term regime (striped green) characterized by moderate code distance and error rate $p \sim 10^{-3}$, in which the transversal STAR architecture yields lower errors than all alternatives considered, in contrast to the conclusion following from previous modeling [28]. Notably, we estimate partial fault-tolerance to enable simulations beyond the reach of both current classical methods and NISQ approaches already on devices with only $5,000$ physical qubits. As the code distance is increased, partially FT computation eventually becomes limited by its fundamental constraints, and fully FT approaches achieve lower error rates. Second, we extend the analysis to the broader tradeoffs of quantum computation—accuracy, qubit count, and depth—as sketched in Fig. 1(e). We show that partially FT computation exhibits a much more favorable spacetime tradeoff relative to fully FT alternatives. As a result, it achieves approximately an order of magnitude reduction in non-Clifford depth, and an even greater reduction in Clifford resources for gate synthesis, up to three orders of magnitude depending on ancillary availability. Altogether, our work establishes partially FT computation as a compelling route toward near-term demonstrations of practical quantum advantage.

III. PARTIAL FAULT-TOLERANCE

Two-dimensional stabilizer codes only admit transversal gates belonging to the Clifford group [40]. Thus, several methods have been proposed to enable the FT realization of specific gates, e.g. T or CCZ gates, to be combined with Clifford operations to form a universal digital gate set. These include magic state distillation [20, 21, 41–44] or cultivation [22, 45, 46], code switching [23, 47–49], and code deformation [50, 51]. However, these approaches are considerably more demanding than transversal operations, and arbitrary operations must additionally be synthesized by tens of these non-Clifford gates [24].

Partial fault tolerance reduces the overhead of universal logical computation, by avoiding both the cost of realizing fully FT non-Clifford gates and the need to approximate arbitrary operations with long gate sequences. In the transversal STAR architecture [28], the logical gate set contains Clifford operations together with arbitrary-angle single qubit rotations $R_L(\theta_L)$. Clifford gates are applied transversally—with qubit permutations when needed—and therefore have low overhead and typically higher fidelity. By contrast, arbitrary-angle rotations are implemented in a *partially* FT fashion, in which the significant reduction in overhead comes at the price of an increased exposure to physical faults.

The STAR architecture, originally proposed in Ref. [25], and further improved in subsequent works [26–28], relies on the preparation and injection of a suitable resource state

$$|m_{\theta_L}\rangle_L = R_{Z,L}(\theta_L)\,|+\rangle_L = \cos\frac{\theta_L}{2}\,|+\rangle_L - i\sin\frac{\theta_L}{2}\,|-\rangle_L\,, \tag{1}$$

which encodes the arbitrary-angle logical rotation

$$R_{Z,L}(\theta_L) = e^{-i\frac{\theta_L}{2}Z_L}. \tag{2}$$

The resource state is then consumed to inject the rotation onto a target state $|\psi\rangle_L$ via gate teleportation, as

shown in Fig. 1(a). Gate teleportation is probabilistic: with probability 1/2, upon a +1 measurement outcome, the output is the correctly rotated state $R_{Z,L}(\theta_L)\,|\psi\rangle_L$, while the opposite outcome produces the inversely rotated state $R_{Z,L}(-\theta_L)\,|\psi\rangle_L$. In the latter case, one repeats the resource-state preparation and injection with a doubled rotation angle, until a successful +1 outcome is observed. This *repeat-until-success* (RUS) procedure is successful on average in two trials.

In the remainder of this section, we review the state preparation and injection protocol (Sec. III A), discuss the error mechanisms that limit its performance (Sec. III B), and finally quantify error propagation within the broader RUS procedure (Sec. III C). Remarkably, we show that the RUS nature of the scheme leads to the beneficial cancellation of coherent error sources, as well as leading-order suppression of coherent logical noise.

A. Transversal Multi-rotation Protocol

The resource state preparation is referred to as the Transversal Multi-Rotation (TMR) protocol [27], for reasons that will soon become apparent, and is illustrated in Fig. 1(b). Let us denote by $P_q^{\otimes w}$ the tensor product of the operator P over w consecutive qubits, starting at index wq. Then, let us define a set of r transversal $Z^{\otimes w}$-rotations

$$\bigotimes_{q=0}^{r-1} R_{Z_q^{\otimes w}}(\theta) = \bigotimes_{q=0}^{r-1} \left[\cos\frac{\theta}{2}\, I_q^{\otimes w} - i\sin\frac{\theta}{2}\, Z_q^{\otimes w}\right], \quad (3)$$

where $R_{Z_q^{\otimes w}}(\theta) = e^{-i\theta Z_q^{\otimes w}/2}$. Next, consider a logical qubit in the state $|+\rangle_L$, with code distance d and logical operator $Z_L = Z^{\otimes d}$. Assuming $d = wr$, applying the transversal rotations on the support of Z_L yields

$$\bigotimes_{q=0}^{r-1} R_{Z_q^{\otimes w}}(\theta)\,|+\rangle_L = \cos^r\frac{\theta}{2}\,|+\rangle_L + (-i)^r \sin^r\frac{\theta}{2}\,|-\rangle_L + (\text{Z-error terms}), \quad (4)$$

where the Z-error terms encompass all terms that project the state outside the logical subspace. Defining a logical rotation angle θ_L satisfying

$$\tan\frac{\theta_L}{2} = \tan^r\frac{\theta}{2}, \quad (5)$$

with $0 \le \theta, \theta_L \le \pi/2$, the above expression is written in terms of the desired resource state in Eq. (1) as

$$\bigotimes_{q=0}^{r-1} R_{Z_q^{\otimes w}}(\theta)\,|+\rangle_L = \begin{cases} \sqrt{p_{\theta_L}}\cdot |m_{(-1)^{\lfloor r/2\rfloor}\theta_L}\rangle_L + (Z\text{-error terms}) & (r \text{ odd}), \\ \sqrt{p_{\theta_L}}\cdot e^{i\frac{\pi}{4}}\, S_X\, |m_{(-1)^{\lfloor r/2\rfloor}\theta_L}\rangle_L + (Z\text{-error terms}) & (r \text{ even}), \end{cases} \quad (6)$$

where

$$p_{\theta_L}(\theta_L, r) = \cos^{2r}\frac{\theta}{2} + \sin^{2r}\frac{\theta}{2}, \quad (7)$$

is the probability that, upon a following syndrome measurement, the state collapses onto the desired resource state. $S_X = R_{X,L}(\pi/2)$ can be removed by applying $S_X^\dagger$ before the teleportation measurement, as in Fig. 1(a). We highlight that in the rotated surface code this can be implemented with only transversal gates and qubit permutations [52]. Finally, we note that no additional correction needs to be applied to reverse the sign of the target angle θ_L when this is negative, as when $\lfloor r/2\rfloor$ is odd, due to the random nature of gate teleportation.

In summary, the TMR protocol encodes the desired resource state $|m_{\theta_L}\rangle_L$ via the following steps: initialize $|+\rangle_L$, apply the transversal w-qubits physical rotations $R_{Z^{\otimes w}}$, and then perform syndrome measurements with suitable post-selection.

Although error detection and post-selection can strongly suppress faults, several error mechanisms can still prevent successful implementation of the protocol; these must therefore be quantified. In what follows, p denotes the physical circuit-level error rate, whereas the noise $\mathcal{E}$ acting on each transversal rotation is best modeled by its chi-matrix $\chi_{P,P'}$, such that

$$\mathcal{E}(\rho) = \sum_{P,P'\in\mathcal{P}^{\otimes w}} \chi_{P,P'} P\rho P', \quad (8)$$

where $\mathcal{P}^{\otimes w}$ denotes the set of w-qubit Pauli operators.

B. Error sources of the TMR protocol

All error-corrected schemes of logical computation suffer from noise in syndrome measurements and transversal logical gates. However, two main error mechanisms limit the performance of the TMR protocol: (i) undetected correlated phase-flips, modifying the logical rotation angle as depicted in Fig. 2(a); (ii) undetected repeating bit-flips, inverting the sign of the logical rotation angle as depicted in Fig. 2(b). In this subsection, we provide a concise overview of the former, while deferring detailed modeling to App. A 2 and A 3. The latter error source, despite its potential to ultimately limit the performance of the protocol, has a negligible contribution in the setting we consider, hence its treatment is deferred to App. A 4.

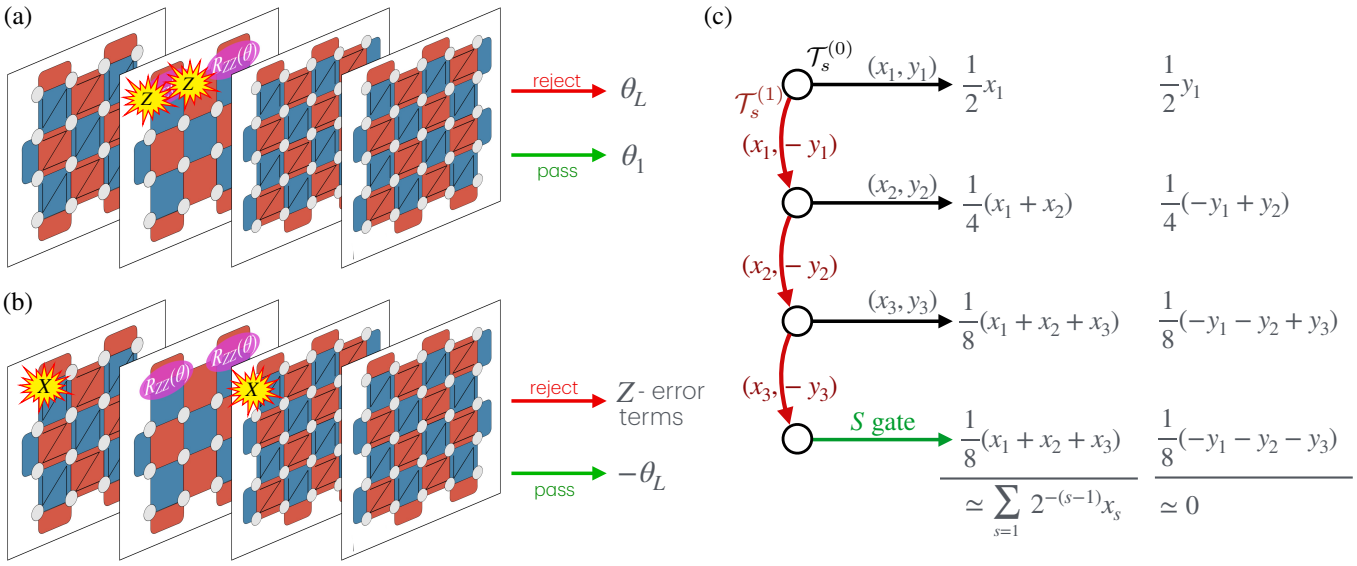


Figure 2. Overview of noise sources and propagation. (a) Correlated phase-flips, occurring shortly before or after the physical rotations, cause the rejection of the correctly rotated state, while passing a wrongful rotation by angle θ_1 (see Eq. (10)). (b) Two repeating bit-flips, occurring shortly before and after the physical rotations respectively, lead to the inverse of the target θ_L rotation. Z-error terms are correctly rejected upon post-selection. (c) Schematic depiction of leading-order error propagation in the Clifford-terminated RUS (ctRUS) procedure. Subscripts denote the step within the RUS process, in this example from 1 to 3. Black (red) arrows represent gate-teleportation channels $\mathcal{T}_s^{(\mu)}$ with ideal (opposite) measurement outcome $\mu = 0(1)$. Each has probability 1/2 in leading order. The green arrow represents the action of the terminating Clifford S, here noiseless for simplicity. Each arrow points to the leading-order estimate of the incoherent x and coherent y noise on the resulting state, weighted by the probability such state is produced. Incoherent noise x_s adds up, whereas coherent noise y_s vanishes in leading order, owing to the opposite sign this noise takes upon opposite measurement outcomes.

The first error mechanism originates from $Z^{\otimes w}$ faults occurring shortly before or after the physical rotations Eq. (3), as illustrated in Fig. 2(a). To understand its origin and effect, it should first be observed that the error terms in Eq. (6) that are closest to the logical subspace—hence dominating the error budget—take the form

$$-i \sum_{q=0}^{r-1} \sqrt{p_{\theta_1}}\, Z_q^{\otimes w}\ |m_{(-1)^{\lfloor r/2 \rfloor}\theta_1}\rangle_L\,, \tag{9}$$

where their effective rotation angle $\theta_1(\theta_L, r)$ satisfies

$$\tan\frac{\theta_1}{2} = -\tan^{(1-\frac{2}{r})}\frac{\theta_L}{2} = -\tan^{r-2}\frac{\theta}{2}, \tag{10}$$

and the probability $p_{\theta_1}(\theta_L, r)$ of collapsing onto each one of these terms is

$$p_{\theta_1} = \cos^{2(r-1)}\frac{\theta}{2}\sin^2\frac{\theta}{2} + \sin^{2(r-1)}\frac{\theta}{2}\cos^2\frac{\theta}{2}. \tag{11}$$

These terms lie outside the logical subspace, thus noiseless syndrome extraction and error detection successfully rejects them. However, a $Z_q^{\otimes w}$ fault exchanges the syndrome subspace of the desired term and the corresponding error component. Ideal post-selection therefore rejects the desired term affected by this fault while accepting the component proportional to p_{θ_1}, producing the wrong logical angle θ_1. The probability of occurrence of one of these faults on any of the physical rotations is

$$p_{Z^{\otimes w}} \simeq r[a_Z\, p^w + \chi_{Z^{\otimes w}, Z^{\otimes w}}(\theta)], \tag{12}$$

where $\chi_{Z^{\otimes w}, Z^{\otimes w}}$ denotes the probability of its occurrence in one of the transversal rotations, and $a_Z > 0$ is a combinatorial prefactor.

C. Repeat Until Success

The previous subsection established the main ingredients required to accurately model noise in the TMR protocol. In this subsection, we present their combined effect in the whole RUS scheme. In particular, we derive a remarkable property of the procedure, neglected in previous works : an intrinsic suppression of logical coherent errors.

The RUS scheme consists of a repetition of gate-teleportation attempts, each succeeding with half probability. In principle, each s^{th} step, with $s = 1, 2, \ldots$ targets a rotation angle twice as large as the previous, i.e. $\theta_{L,s} = 2\,\theta_{L,s-1}$, with $\theta_{L,1} = \theta_L$. However, as will be shown shortly, smaller angles yield smaller error rates. Then, as $\pi/2$ rotations are FT Clifford gates, the preferred logical rotation angle at every step is the *least absolute remainder* with respect to $\pi/2$, i.e. the one that has the smallest absolute value up to multiples of $\pi/2$,

$$\theta_{L,s} = 2^{s-1}\theta_L - \frac{\pi}{2}\underset{k\in\mathbb{Z}}{\operatorname{argmin}}\left|2^{s-1}\theta_L - \frac{k\pi}{2}\right|. \tag{13}$$

In general, we denote with s subscript the quantities pertaining to each s^{th} step, conveniently summarized

Table I. Summary of the key parameters characterizing the Transversal Multi-Rotation (TMR) protocol at the s^{th} step of the Repeat-Until-Success (RUS) scheme, ultimately realizing a rotation by the angle θ_L. The expressions detail their analytical dependence on the number r of physical rotations and the rotation angles.

Symbol	Description	Definition
$\theta_{L,s}$	Target logical rotation angle	$2^{s-1}\theta_L - \frac{\pi}{2}\underset{k\in\mathbb{Z}}{\operatorname{argmin}}\left\lvert 2^{s-1}\theta_L - \frac{k\pi}{2}\right\rvert$
θ_s	Physical transversal rotation angle	$2\arctan\left(\tan^{1/r}\frac{\theta_{L,s}}{2}\right)$
$p_{\theta_{L,s}}$	Probability of collapse onto the desired resource state	$\cos^{2r}\frac{\theta_s}{2} + \sin^{2r}\frac{\theta_s}{2}$
$\theta_{1,s}$	Logical rotation angle of leading Z-error terms	$2\arctan\left(-\tan^{r-2}\frac{\theta_s}{2}\right)$
$p_{\theta_{1,s}}$	Probability of collapsing onto one of the leading Z-error terms	$\cos^{2(r-1)}\frac{\theta_s}{2}\sin^2\frac{\theta_s}{2} + \sin^{2(r-1)}\frac{\theta_s}{2}\cos^2\frac{\theta_s}{2}$

in Table I. Moreover, we assume that the RUS scheme can be terminated at the C^{th} step with full certainty, by application of the Clifford S gate. As the target angle is doubled at every step of the scheme, this is the case for any target angle of the form

$$\theta_L = \frac{\pi}{2}\frac{2k+1}{2^C}, \tag{14}$$

for any integer k. It should be noted that this form provides a dense set of angles, which can approximate any other with arbitrary precision.

To model noise propagation, we first observe that the error sources specific to the TMR protocol amount to incoherent and coherent noise along the Z_L axis. Each teleportation attempt within the RUS procedure is well described by the ansatz

$$\begin{aligned}\mathcal{T}_s(\Psi,\theta_{L,s};\,x_s,y_s) = {}& (1-x_s)\Psi_{\theta_{L,s}} + x_s\, Z_L\Psi_{\theta_{L,s}}Z_L \\ & - iy_s\, Z_L\Psi_{\theta_{L,s}} + iy_s^*\Psi_{\theta_{L,s}}Z_L,\end{aligned} \tag{15}$$

where Ψ is the input state; $\Psi_\bullet$ denotes $R_Z(\bullet)\Psi\, R_Z^\dagger(\bullet)$ such that $\Psi_{\theta_{L,s}}$ is the ideal output; x_s and y_s represent, respectively, the incoherent and coherent error contributions, both dependent on the rotation angle θ_s and the number r of physical rotations. Importantly, upon opposite teleportation outcomes, the signs of both the encoded angle $\theta_{L,s}$ and the coherent error y_s are flipped, as explained in App. A 5.

A step-by-step derivation of error accumulation in the RUS realization of the protocol, including in the absence of Clifford termination, is deferred to App. A 6, while here we proceed to state the result. Assuming first that the terminating Clifford is noiseless, the result of the Clifford-terminated repeat-until-success (ctRUS) is described by the channel

$$\begin{aligned}\rho_{\text{ctRUS}}(\Psi,\theta_L) = {}& (1-p_{\text{ctRUS}})\Psi_{\theta_L} + p_{\text{ctRUS}}\, Z_L\Psi_{\theta_L}Z_L \\ & + O(x_s^2 + x_s y_s + y_s^3),\end{aligned} \tag{16}$$

yielding the target output Ψ_{θ_L} under stochastic Z-noise of strength $p_{\text{ctRUS}} = \sum_{s=1}^{C} 2^{-(s-1)}x_s$. Remarkably, owing to the difference in sign of the coherent noise upon opposite teleportation outcomes, the leading order contribution to the overall logical coherent noise *vanishes*. Further intuition on the RUS error accumulation or cancellation is provided by the diagram in Fig. 2(c), which illustrates the propagation of leading order errors in the ctRUS scheme.

This result does not change upon inclusion of the noise from the S gate, which has negligible coherent noise owing to the applicability of Pauli twirling to Clifford operations [53, 54]. Moreover, the noise from the S gate is suppressed exponentially with code distance and is further reduced by the probability $\sim 2^{-C}$ of its application, thus is generally negligible. The dominant noise terms x_s are determined by the error mechanisms presented in the previous subsection. Only considering the effect of undetected correlated phase-flips—the dominant error mechanism—yields

$$x_s \simeq \frac{p_{\theta_{L,s}}}{n_{\theta_{L,s}}}\, p_{\theta_{1,s}}\, p_{Z^{\otimes w}}\, \sin^2\frac{\theta_{1,s}-\theta_{L,s}}{2} \tag{17}$$

where $p_{Z^{\otimes w}}$ generally depends on $\theta_{L,s}$ and r as per Eq. (12), and with normalization factor $n_{\theta_{L,s}}$ defined in Eq. (A16) and close to one in the small angle scenario. A more complete expression taking into account also the effect of undetected bit-flips is provided in App. A 5.

The results outlined in this section highlight some important points. First, the RUS realization of logical rotations benefits from intrinsic cancellation of coherent noise in leading order, avoiding the need for the additional mitigation scheme devised in [27], and making partially FT computation more amenable to typical quantum error mitigation techniques [55]. Second, noise is expected to decrease with the rotation angle θ_L, which is desirable given the expected applications of the protocol. Third, by increasing the size w of the transversal rotations, the contribution from physical noise can in principle be reduced, namely that coming from correlated phase flips in Eq. (12). In practice, however, gates beyond two qubits have been rarely implemented, and with lower fidelity than two-qubit

gates [56]. For this reason, we assume $w = 2$ throughout the rest of this paper. Finally, it is evident that a reliable estimate of the protocol performance requires an accurate characterization of the noise in the transversal $R_{Z^{\otimes w}}$ rotations.

Accordingly, in the next section we perform a detailed analysis of R_{ZZ} gates in realistic neutral-atom settings. We show that the undetectable error $\chi_{zz,zz}$ exhibits a favourable dependence on the rotation angle, and is comparable in magnitude to p^2, despite not scaling as such. Both features contrast with assumptions made in previous works [27, 28]. We also optimize the gate at the pulse level to further improve the performance of the protocol. Then, in Sec. V we simulate and discuss the resulting behavior of the logical protocol.

IV. RYDBERG GATE CHARACTERIZATION AND OPTIMIZATION

As explained in the previous section, partial fault tolerance significantly reduces the overhead of universal logical computation, at the price of exposure to physical faults. Specifically, the protocol is particularly sensitive to ZZ-type errors in the application of physical $R_{ZZ}(\theta)$ gates, as they modify the logical angle of rotation while going undetected, as explained in Sec. III B. Therefore, it is critical to accurately investigate the magnitude of these undetectable faults in the realistic scenario of interest, and to reduce it if possible. In this regard, we note that Ref. [29] has recently proposed an error reduction technique exploiting the structure of dissipative noise, particularly suitable to superconducting qubits. Our work considers a different platform and approach.

We address the exposure to critical faults of partial fault-tolerance in the context of neutral-atom architectures, one of the most promising platforms for digital quantum computation, with recent demonstrations showcasing a broad range of capabilities across different implementations [31]. These include high-fidelity gates [56–61], mid-circuit measurements [13, 17, 61–66], scalability beyond a thousand atoms [67–69], continuous reloading of the array [69–72], and reconfigurable non-local connectivity via atomic shuttling [12–14, 73].

In this section, we focus on the modeling and optimization of $R_{ZZ}(\theta)$ Rydberg gates implemented with ^{171}Yb atoms [59, 60, 62], noting that our considerations readily extend to other alkali-earth atoms, such as strontium [57]. Unlike the construction proposed in Ref. [29], which to our knowledge has not yet been tested in neutral-atom platforms, our study builds directly on experimentally demonstrated gate implementations. We first introduce Rydberg gates and their dominant noise sources, then optimize the pulse specifically against faults that remain undetected by the TMR protocol, and finally quantify the error channel of the gates.

A. Rydberg Gates

Multi-qubit Z-type rotations can be natively implemented in neutral atoms. Typically, these are two-qubit gates relying on the repulsive van der Waals V interaction, which arises when the atoms are in the same Rydberg state $|r\rangle$, a highly excited atomic state. Specifically, Rydberg gates are implemented by state-selectively coupling the atomic qubit of each atom, conventionally from its $|1\rangle$ state, to the Rydberg level via a laser, as illustrated for two atoms in Fig. 3(a). The laser amplitude Ω, phase ϕ, and detuning Δ can be modulated in time to yield the desired gate, ranging from the maximally-entangling CZ gate [74–76], to arbitrary angle $R_{ZZ}(\theta)$ rotations [56, 77]. In practice, only one between laser phase and detuning is typically modulated, as they are related up to a change of reference frame, as $\dot{\phi} = \Delta$ [77].

General N_q-qubits Rydberg gates are realized by the Hamiltonian

$$H(t) = H_{\rm drv}(t) + H_{\rm int}(t), \tag{18}$$

with

$$\begin{aligned} H_{\rm drv}(t)/\hbar &= \sum_{q}^{N_q} \left(\frac{\tilde{\Omega}(t)}{2} |r\rangle \langle 1|_q + \text{h.c.} \right) - \Delta(t) |r\rangle \langle r|_q , \\ H_{\rm int}(t)/\hbar &= \sum_{q<q'}^{N_q} V_{qq'} |rr\rangle \langle rr|_{qq'} , \end{aligned} \tag{19}$$

where $\tilde{\Omega}(t) = \Omega(t) e^{-i\phi(t)}$ and $V_{qq'} = C_6/R^6_{qq'}$, where C_6 is a coefficient of the chosen Rydberg state, and $R_{qq'}$ is the distance between atoms q and q'.

State-of-the-art realizations have achieved CZ gates fidelity exceeding 99.5%, and native three-qubit CCZ gates have also been demonstrated, with fidelity $\sim 98\%$ [56].

B. Noise on Rydberg Gates

In this subsection, the noise model considered for the Rydberg gates is outlined. Specifically, we consider (i) atomic motion due to finite temperature and photon recoil, (ii) driving imperfections, including shot-to-shot fluctuations and finite amplitude rise- and fall- times, and (iii) Rydberg decay, both into and out of the computational subspace. These noise sources are schematically depicted in Fig. 3, together with the relevant parameters adopted in our simulations. The laser phase is modulated in time and the laser detuning set to zero.

For modeling of atomic motion, we assume that the atoms sit in harmonic traps with radial frequency $\omega_{\rm tr}$, inverse temperature β, and initial thermal state $\rho \propto \exp\left(-\beta\hbar\omega_{\rm tr} a^\dagger a\right)$, where a ($a^\dagger$) is the annihilation (creation) operator of the relevant motional modes, described by the usual Hamiltonian $H_{\rm mot} = \hbar\omega_{\rm tr} a^\dagger a$. Then, the driving term in the Hamiltonian is modified to include

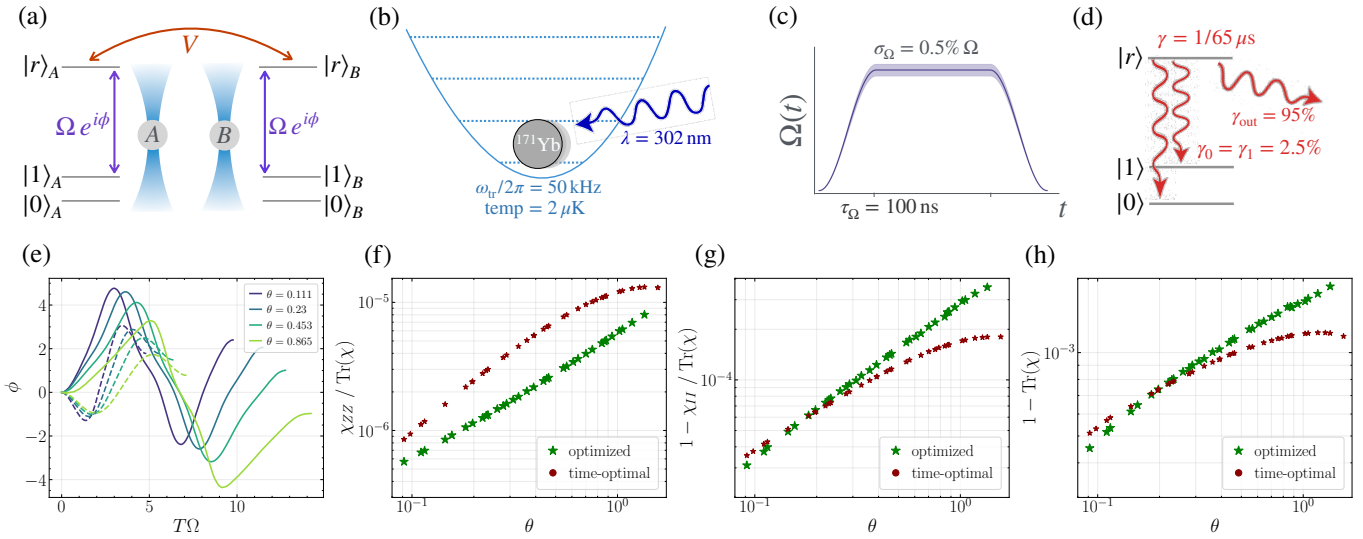


Figure 3. Rydberg gate and associated noise, with relevant parameters. (a) The two-qubit Rydberg gate exploits the blockade interaction $V\,|rr\rangle\langle rr|$, arising when the $|1\rangle$ state of both atoms is coupled to the highly-excited Rydberg state $|r\rangle$. (b) We consider ^{171}Yb atoms trapped at low yet finite temperature, thus undergoing atomic motion coupled with recoil from the driving photons. (c) The amplitude of the driving laser has finite ramping time, modeled by a cosine-tapered envelope, as well as gaussian fluctuations in its value. (d) The Rydberg state has finite lifetime $1/\gamma$, thus experiences decay both outside the computational subspace, as well as back to the computational states. (e) Examples of time-optimal (dashed) and noise-optimized (solid) pulses for different physical angles of rotation. (f,g,h) Comparison of relevant error metrics between time-optimal and noise-optimized pulses at different angles of rotation: (f) probability of undetectable error post-selected upon leakage; (g) error rate post-selected upon leakage; (h) leakage probability.

the momentum transfer of the laser field, with wavelength λ,

$$H_{\rm drv}(t)/\hbar = \frac{1}{2}\sum_{q}^{N_q}\left(\tilde{\Omega}(t)e^{i\eta(a_q+a_q^\dagger)}\,|r\rangle\langle 1|_q + \text{h.c.}\right), \quad (20)$$

with Lamb-Dicke parameter $\eta = 2\pi x_{\rm osc}/\lambda$, where $x_{\rm osc} = \sqrt{\hbar/(2m\omega_{\rm tr})}$. Moreover, the van der Waals potentials are expanded to first order in the interatomic distance fluctuations $\delta R_{qq'} = x_q - x_{q'}$, where $x_q = x_{\rm osc}(a_q^\dagger + a_q)$, such that

$$H_{\rm int}/\hbar \simeq \sum_{q<q'}^{N_q} V_{qq'}\left(1 - \frac{6(x_q - x_{q'})}{R_{qq'}}\right)|rr\rangle\langle rr|_{qq'}\,. \quad (21)$$

In App. B 1 we argue that axial fluctuations, ignored in our computation, indeed only have a small impact in the considered set-up.

The driving amplitude $\Omega = |\tilde{\Omega}|$, rather than being constant throughout the protocol, is assumed to have a finite ramping time τ_Ω, with cosine-tapered shape [60]. That is, the amplitude is initially ramped-up with profile $\Omega(t < \tau_\Omega) = \Omega_{\max}\sin^2[\pi t/(2\tau_\Omega)]$, and similarly ramped-down in the final stage. The laser phase is accordingly adapted. Moreover, the amplitude has shot-to-shot gaussian fluctuations $\delta\Omega$ with standard deviation σ_Ω. Due to the off-resonant coupling of the laser to other states in the Rydberg manifold, these deviations lead to self-light shifts, i.e detuning fluctuations correlated with those on the driving amplitude, as $\delta\Delta_q = 2\kappa\,\Omega\delta\Omega$. With regards to alkali-earth implementations, $2\pi\kappa \approx 10\,\text{kHz/MHz}^2$ was observed in strontium [78], and a twice-as-large upper bound was determined in ytterbium [73]. In our simulations, we set $\kappa = 2\,\text{kHz/MHz}^2$.

Finally, Rydberg gates suffer from decay from the Rydberg state, often one of the major sources of infidelity [56–61]. Rydberg decay outside of the computational subspace can be modeled by the addition of a non-Hermitian effective Hamiltonian

$$H_{\rm out}/\hbar = -i\frac{\gamma_{\rm out}}{2}\sum_{q=1}^{N_q}|r\rangle\langle r|_q\,, \quad (22)$$

whereas decay to the computational state $|0\rangle$ and $|1\rangle$ is modeled by jump operators $L_q^{(i)} = \sqrt{\gamma_i}\,|i\rangle\langle r|_q$, where $\gamma_{0/1}$ is the decay rate to $|0/1\rangle$. It has been observed that in the metastable qubit of ^{171}Yb here considered, $\gamma_{\rm out}$ accounts for 95% of the total Rydberg decay rate $\gamma \approx 1/(65\,\mu\text{s})$ [62]. This results in leakage to levels whose population can be continuously monitored without affecting the qubit, thus detectable *erasure* errors [79]. In the context of the transversal rotations on the TMR's resource state, the state can be discarded upon detection of erasures, thus leakage does not contribute to the logical error rate. Finally, we note that by using alternative leakage-detection units, most leakage events can be detected in absence of built-in erasure capabilities, at the price of additional overhead [80].

C. Gate Optimization

State-of-the-art implementations adopt the so-called time-optimal gate proposed in Ref. [76], or an approximation thereof [56–61]. This is the laser pulse realizing the desired gate in the shortest time in a noiseless scenario, up to additional single-qubit Z-rotations. However, gates can also be optimized to increase their robustness to noise, as previously demonstrated for Rydberg gates in different scenarios [81–85].

Given the partially FT nature of the TMR protocol, specifically its sensitivity to ZZ-type errors in the application of physical $R_{ZZ}(\theta)$ gates, we optimize the latter at the pulse level, to further enhance the logical performance of the protocol. Therefore, we perform *error-selective* pulse optimization, not to suppress the overall infidelity, as commonly done, but specifically the contribution of those undetectable ZZ-type errors. Such error-selective approaches have only recently attracted attention in neutral atoms [85, 86] and superconducting architectures [87], albeit within the context of quantum memories rather than logical computation.

In the remainder of the section, we focus on two-qubit Rydberg gates. The laser field is assumed to be global, in line with most experimental implementations [17, 57, 58], with constant amplitude Ω, null detuning Δ, and time-modulated phase $\phi(t)$. We adopt the cost function

$$\text{cost} = \chi_{zz,zz} + (1-F)\,\eta_F, \tag{23}$$

where $\chi_{zz,zz}$ denotes the probability of a ZZ-error following the gate, whereas $(1-F)$ is the gate's infidelity, entering the cost function with weight η_F. We choose $\eta_F = 10^{-3}$, to make the ZZ-error the focus of the optimization process, while ensuring the gate maintains sufficiently high fidelity. The cost function is computed under the effect of the noise sources mentioned in the previous subsection, although appropriately approximated to be efficiently estimated, as detailed in App. B 2, while the computation of $\chi_{zz,zz}$ is described in App. B 3.

Numerical optimization is carried out via gradient-ascent pulse engineering (GRAPE) [88–91], in which the time-dependent control function $\tilde{\Omega}(t)$ is discretised as a piecewise function over a time-interval T. The optimization is then described by $N \gg 1$ parameters $\phi_1, ..., \phi_N$, such that $\tilde{\Omega}(t_i) \equiv \Omega e^{-i\phi_i}$, where $t_i = (i+1/2)dt$, and $dt = T/N$. We choose $N = 600$. Two other parameters are optimized over. The first is a residual single-qubit phase, as usual in the context of Rydberg gates [76, 83, 84]. The second is the total gate time T, optimized together with the other parameters to reduce the number of iterations. This is uncommon, as usually the optimization is repeated for increasing gate times, until a time T is found for which the cost function drops significantly. We observe that the piecewise approximation of the control function, and the combined optimization of parameters and gate time, can lead to numerical instabilities, suppressed by appropriately designed penalties in the cost function, as further commented on in App. B 4.

D. Gates Performance

We computed the performance of the time-optimal and noise-optimised gates, by simulating the Lindbladian dynamics described by the Hamiltonian and jump operators in Sec. IV B. The performance was evaluated at experimentally realistic amplitude $\Omega = 7.5 \times 2\pi\,\text{MHz}$, atomic distance $R = 2.5\ \mu$m, and van der Waals potential $V = 63.7\Omega$. The simulation was discretised in 3000 time-steps; 8 vibrational levels were used per atom; shot-to-shot noise in the Rabi frequency was sampled according to the Gauss-Hermite quadrature over 15 realizations. It was verified that these parameters were sufficient to obtain converged results. Residual population in the Rydberg state at the end of the pulse was mapped to computational states and leakage according to the branching ratios in Fig. 3(d).

In Fig. 3, we compare the time-optimal and optimized pulses for arbitrary angle rotations $R_{ZZ}(\theta)$. We note that, with the adopted parameters, the error rate of the CZ gate amounts to 0.14%, thus only slightly better than best reported experimental values. Figure 3(e) compares their shape, where we observe that the optimized pulses exhibit a smooth profile, and are implemented in approximately twice the time of the time-optimal counterparts. Figure 3(f) compares the probability of undetectable errors under realistic noise, with the optimized pulses consistently providing a two- to three-fold reduction. Finally, Figures 3(g) and 3(h) show that the optimized gates' infidelity—post-selected upon leakage events—and probability of leakage are only mildly increased with respect to the time-optimal pulse. As it will be shown in the next section, these results directly translate to a twofold reduction in the logical error rate of the protocol, with negligible increase in post-selection.

However, two more remarks are in order. First, we note that previous work has shown that no pulse can suppress leading-order sensitivity of Rydberg gates to atomic motion, unless combined with fine-tuned experimental tricks [83]. These are in turn exposed to other error sources and experimental challenges, and are not assumed in our work. This vulnerability could limit the suppression of the undetectable error in our optimized pulses. Secondly, we note that the reduction of the cost function during the optimization process, as opposed to the value associated with time-optimal pulses, was by nearly an order of magnitude, hence more pronounced than observed in Fig. 3(f). This discrepancy is likely due the difference between the approximate noise model adopted in the optimization and the detailed noise model used in the final benchmarking. In a real experimental scenario, the method's improvements may be maximized by spending additional resources in a more detailed optimization process. We defer this investigation to future work.

V. LOGICAL PERFORMANCE

In this section, we first outline the underlying assumptions in the simulation of the logical TMR protocol, specifically the adopted noise model and decoding. Then we show and comment on the numerical performance of its repeat-until-success implementation.

A. Assumptions

The protocol is assumed to exploit the all-to-all connectivity enabled by atomic shuttling [13, 73], in turn enabling the realization of (fold-)transversal logical gates. Owing to their deterministic nature and the use of correlated decoding, these require only ~ 1 round of syndrome extraction [32], as opposed to traditional lattice surgery logic. Despite increasing the total volume of classical decoding, efficient decoding has been shown to remain viable [28, 33, 34]. As such, our simulations incorporate a single noisy round of stabilizer extraction on the logical qubit following the encoding of the resource state and gate teleportation, as shown in Fig. 1(a). Classical decoding is performed with the recently developed Minimum Weight Parity Factor (MWPF) decoder [92], with cluster node limit set to 300. More details on the decoder, and its comparison against popular alternatives, can be found in Ref. [28].

The protocol is simulated against circuit-level noise of varying strength, modeled by the error rate $p_2 \equiv p$ of CZ gates—the only natively available Clifford two-qubit gates. Single qubit gates have depolarizing rate $p_1 = p_2/5$, while measurements and resets—only assumed to be available in the computational basis—have bit-flip rate $p_m = p_2/2$ and $p_i = p_2/5$, respectively. These values roughly reproduce typical state-of-the-art devices using alkali-earth atoms [58, 59, 62, 93]. The error channel $\chi(\theta)$ of the non-Clifford transversal rotations, numerically computed as per section Sec. IV with parameters corresponding to $p_2 = 0.14\%$, was linearly rescaled to match the circuit-level error rate.

A key component of the model is the biased erasure noise of the CZ gates, the native two-qubit gates in atomic platforms, exhibited by alkali-earth atoms [62, 79]. As mentioned in Sec. IV B, Rydberg decay is one of the main sources of infidelity, and mainly results in *state-dependent* leakage from the $|1\rangle$ state. As such, re-initialization in $|1\rangle$, or replacement of the atom by a qubit in $|1\rangle$, does not flip its computational state, but only affects the phase, resulting in a Z-biased noise channel. While the details of the CZ noise modeling, including the adopted Kraus channel, are referred to App. C, we note that previous work has shown that this form of bias can significantly increase the QEC threshold [62, 79, 94, 95]. Moreover, due to the sensitivity of the TMR protocol to phase-flips, erasures in CZ gates acting on the first two rows of the logical patch—thus in proximity of the transversal physical rotations—are post-selected upon.

B. Simulation and Results

We numerically simulated the logical performance of the RUS procedure for the realization of arbitrary-angle logical rotations $R_{Z,L}(\theta_L)$ in the rotated surface code with distance $d = 5, 7,$ or 9. The resource state is initially encoded by one round of syndrome extraction at distance $d_s = d-1$ respectively, then $r = d_s/2$ transversal physical rotations are applied, and finally the state is grown to the target distance d. The resource state is injected upon measurement of trivial syndromes, and discarded otherwise.

The key figure of merit used to assess the protocol's robustness is the diamond-norm error $\varepsilon_\diamond$, also known as the worst-case error rate. This is more sensitive than infidelity to coherent errors [96, 97], determines the maximum probability of distinguishing the noisy and ideal outputs [98, 99], and plays a central role in rigorous proofs of fault tolerance thresholds [100, 101]. In this scenario, due to the strong suppression of coherent noise in the RUS protocol, $\varepsilon_\diamond \simeq p_L$. More details on the diamond-norm error can be found in App. A 1. Another important metric to assess the scheme's viability is the average number $A(\theta_L)$ of attempts required to successfully encode a resource state. Although omitted for compactness, this also depends on the physical error model and details of the TMR implementation.

We reconstruct these metrics by combining Clifford-circuit simulations, performed with `stim` [102], with the physical chi matrix of the transversal rotations. Our method is thoroughly described in App. D, and we stress that it enables the evaluation of the logical channel without further approximation in the treatment of the non-Clifford operations. Details on the number of shots simulated at different distances and error rates, and on the estimation of uncertainty, can be found in App. D 5. The results are shown in Fig. 4, for two-qubit error rates $p \equiv p_2 = 10^{-3}$ and 5×10^{-4}. The simulations compare the use in the transversal rotations of time-optimal Rydberg gates against the optimized pulses found in Sec. IV.

The logical error rate is displayed in Figs. 4(a) and 4(b), and is well approximated by the following ansatz

$$p_L(\theta_L, p, d, r) \simeq \alpha_{\mathrm{RUS}}(p, r)\, p\, |\theta_L| + 2p_L^{(\mathrm{inj})}(p, d), \quad (24)$$

as highlighted in the insets. The first term is specific to the TMR protocol, depends linearly on $|\theta_L|$, and dominates for larger $|\theta_L|$ and relatively small code distance and error rate. As expected, the use of optimized pulses yields approximately a twofold reduction in α_{RUS}, notably without additional overhead. We estimate $\alpha_{\mathrm{RUS}}(r, p) \simeq r(0.01+4p)$ upon use of the optimized pulses. This is between one and two orders of magnitude lower than previous estimates [27, 28], which assumed an angle-independent noise model for the transversal rotations. Such a reduction stems from the use of post-selection upon leakage, the favorable noise bias in Rydberg gates observed in Sec. IV, and the pulse-level optimization.

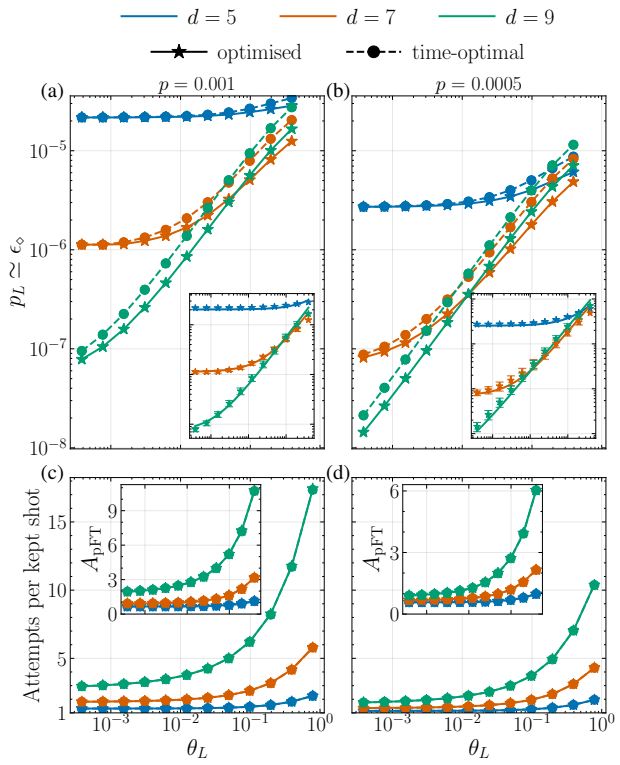


Figure 4. Logical performance of the RUS execution of the TMR protocol. The rows show the logical error rate, and the average number of encoding attempts per resource state. Columns correspond to two-qubit error rates $p = 0.1\%$ and 0.05% respectively, based on which other error rates are rescaled. Code distances $d = 5, 7, 9$ are considered, with corresponding distance $d_s = d - 1$ at which the physical transversal rotations are applied. The use of time-optimal (dashed) and noise-optimized (solid) Rydberg gates is compared. (a-b) Logical error rate of an arbitrary-angle $R_Z(\theta_L)$ rotation. Markers are presents for both p_L and $\epsilon_\diamond$, indistinguishable due to their closeness. The inset compares numerical simulations (markers) against the analytic fit in Eq. (24) (lines) upon use of the optimized pulses, showing excellent agreement. Errorbars represent the standard error on the numerical estimates, and are shown in the inset only for visual clarity. (c-d) Average number of attempts per encoded resource state, i.e. inverse of the probability the encoding procedure passes post-selection. Numerical estimates (markers) are compared against the analytical ansatz in Eq. (26) (lines) showing excellent agreement. The inset shows the behavior of the $A_{\mathrm{pFT}}(\theta_L)$ metric, which can be interpreted as an *effective* number of attempts per encoded state in the RUS process, over a batch of parallel rotations (see App. F 1).

This reduction significantly improves the application-level resource estimates considered in the next section. The second error term comprises errors in the Clifford operations involved in state injection, and is suppressed exponentially in the code distance as

$$p_L^{(\mathrm{inj})}(p, d) = \frac{a_{\mathrm{inj}}}{d-1}\left(\frac{p}{p_{\mathrm{th}}}\right)^{\lceil d/2 \rceil}, \tag{25}$$

with fitted $a_{\mathrm{inj}} \simeq 0.070$ and $p_{\mathrm{th}} \simeq 0.012$ under the considered noise model. Its contribution is twice as much because the RUS protocol requires approximately two repetitions on average. At near term code distances and error rates, it represents the main limitation to the performance of the TMR protocol at small $|\theta_L| \lesssim 10^{-3}$. This suggests that the recently proposed combination of the TMR protocol and magic state cultivation [103], whose advantages manifest at smaller angles, will only be attained in neutral-atom devices at lower error rates.

The average number $A(\theta_L)$ of encoding attempts required per kept resource state is displayed in Figs. 4(c) and 4(d). This is inverse of the probability of accepting the encoded state, thus scaling as

$$A^{-1}(\theta_L) \approx p_{\theta_L} \cdot (1 - p_{\mathrm{erasure}}) \cdot p_{\mathrm{pass}}. \tag{26}$$

All factors depend on the details of the TMR implementation and error model, and are, respectively, the probability Eq. (7) of collapsing on the correct state, Eq. (C2) of observing no erasure error, and that of observing no other faults. With the empirical fit $p_{\mathrm{pass}} \approx (1 - 4.5p)^{[2(d^2-d)+(d_s^2-d_s)]}$, excellent agreement is achieved with the numerical simulations.

As further motivated in the next section and in App. F 1, $A(\theta_L)$ does not by itself capture the overhead of parallel executions of multiple RUS rotations. For application-level scheduling, the overhead is better described by the metric $A_{\mathrm{pFT}}(\theta_L)$. This can be interpreted, in analogy with standard gate synthesis by Clifford+T decomposition, as the *effective* number of attempts per injection over the whole RUS protocol. Specifically, for the RUS realization of a rotation by θ_L, with each s^{th} step targeting $\theta_{L,s}$, it is defined as

$$\begin{aligned} A_{\mathrm{pFT}}(\theta_L) &= \max_{s \geq 1} \; q_s(\theta_L) - 1, \\ q_s(\theta_L) &= 2^{-s} \max\left(A(\theta_{L,s}, \frac{A(\theta_{L,s+1})}{2} + 1 \right). \end{aligned} \tag{27}$$

Its dependence on θ_L, d and p is shown in the insets of Figs. 4(c) and 4(d). Remarkably, in the regime of main interest, $A_{\mathrm{pFT}} \simeq 1$ at moderate code distances, hence smaller than the equivalent metric in magic state cultivation [22, 45, 46]. The relevance of the metric to quantifying the overhead of partial fault tolerance in realistic applications is further discussed in Sec. VI C.

VI. RESOURCE ANALYSIS FOR NEAR-TERM APPLICATIONS

This section quantifies the resource requirements of the transversal STAR architecture and compares this with two fully FT approaches: standard Clifford+T synthesis [24] and Hamming Weight Phasing (HWP) [104, 105]. Throughout we quantify the limitations and trade-offs between low qubit counts, high accuracy, and short circuit depth, broken down as follows.

Subsection VI A addresses the limitations from hardware size on the availability of logical qubits. Given an implemented QEC code and code distance, independent of the approach, hardware size poses stringent requirements on the availability of logical qubits. See also Fig. 5(a).

Subsection VI B addresses the limitations from logical error rates on the achievable simulation size. Focusing on Trotterised dynamics, gate counts are defined by its digital implementation and the target Trotter error; code distance and physical noise define a logical error model. Together, these determine the feasibility of quantum simulations at different system sizes—assuming the combination of QEC and quantum error mitigation. See also Fig. 5(b).

Subsection VI C addresses the limitations from circuit depth. In particular it quantifies the trade-off between depth and the availability of non-Clifford resources, depending on the computation approach. See also Fig. 5(c).

We consider implementations with the rotated surface code. Altogether, we argue that at moderate code distances $d = 5$ or 7, partial fault tolerance can support the simulation of many-body physics at system sizes beyond the reach of its fully FT counterpart, as well as of classical methods, thus providing a pathway to early demonstrations of quantum advantage. In particular, simulations at distance 5 beyond the reach of classical methods are expected to require less than 5,000 qubits. Moreover, we show that partial fault tolerance reduces "feed-forward" and Clifford depths by roughly one and two orders of magnitude, respectively, relative to fully FT implementations. The reduction in depth is accentuated when only a small fraction of the hardware is allocated to non-Clifford resource-state production, as expected in early FT devices.

In quantifying these trade-offs, we focus on the specific case of time evolution in the (Fermi–)Hubbard model, a paradigmatic model of strongly correlated fermions. The model describes spin-1/2 fermions hopping between neighboring lattice sites and interacting through an on-site interaction, with Hamiltonian

$$H_{FH}/\hbar = -J \sum_{\langle i,j\rangle,\sigma} \left(c^\dagger_{i\sigma} c_{j\sigma} + \text{h.c.} \right) + U \sum_{i=1} n_{i\uparrow} n_{i\downarrow}, \quad (28)$$

where $c^\dagger_{j\sigma}(c_{j\sigma})$ is the fermionic creation(annihilation) operator, $\langle i,j\rangle$ denotes nearest-neighbors on the lattice, and $n_{i\sigma} = c^\dagger_{i\sigma} c_{i\sigma}$. Despite its simple structure, it exhibits rich many-body physics such as antiferromagnetism, Mott insulating behavior, and a simplified picture of high-temperature superconductivity [106]. Its relative simplicity compared with full electronic-structure Hamiltonians, together with its display of strongly correlated dynamics, has made the Hubbard model a long-standing benchmark for analogue and digital quantum simulation [27, 37, 105–113]. For its digital implementation, the fermionic Hamiltonian is mapped to Pauli operators using the Jordan–Wigner (JW) transformation, requiring no additional ancillary qubits. Our protocol is tailored to early FT architectures, where Clifford operations still have a non-negligible cost, by making frequent use of the Clifford-count reduction technique introduced in Ref. [36], and adopted in Ref. [37], of which we provide an extension to periodic boundary conditions. Details of the Hamiltonian, the adopted Trotter decomposition, and our circuit implementation are given in App. E. Results provided in the remainder of the section assume $U = 4J$ and half-filling, as this represents a challenging parameter regime for classical simulation [105–107].

A. Limitations from hardware size

The physical-qubit budget imposes the first bound on the simulated lattice size. The spatial footprint of the error-corrected simulation is determined by the following factors.

(i) The size of each logical qubit, defined by the encoding and code distance. We consider encoding in the rotated surface code, requiring $2d^2 - 1$ physical qubits per logical patch.

(ii) The number Q_{alg} of *algorithmic* logical qubits, i.e. those required for the algorithm implementation, neglecting the cost of non-Clifford operations. This is determined by the selected application and related system size. For the simulation of the Hubbard model on a square $L_x \times L_y$ lattice with periodic boundary conditions, our implementation requires $Q_{\text{alg}} = 2L_xL_y + L_y$ algorithmic qubits (see App. E for more details).

(iii) The number Q_{syn} of ancillary patches dedicated to the synthesis of non-Clifford operations.

Altogether, the number of physical qubits required amounts to

$$Q_{\text{phys}} \simeq (2d^2 - 1)(Q_{\text{alg}} + Q_{\text{syn}}). \quad (29)$$

Figure 5(a) quantifies the maximum lattice size in the Hubbard model that can be simulated at different hardware sizes, depending on the adopted code distance and the fraction of resources dedicated to non-Clifford synthesis. It is not surprising that larger system sizes are achieved at smaller code distance and ancillary footprint. However, these result in higher error rates and larger circuit depth, respectively, as discussed in the next subsections.

B. Limitations from logical errors

Early FT quantum computation is likely to be combined with quantum error mitigation (QEM) to extend its applicability. QEM aims to return the noiseless expectation value $\langle O\rangle$ of a chosen observable, at the price of an increased variance in its estimator, hence sampling

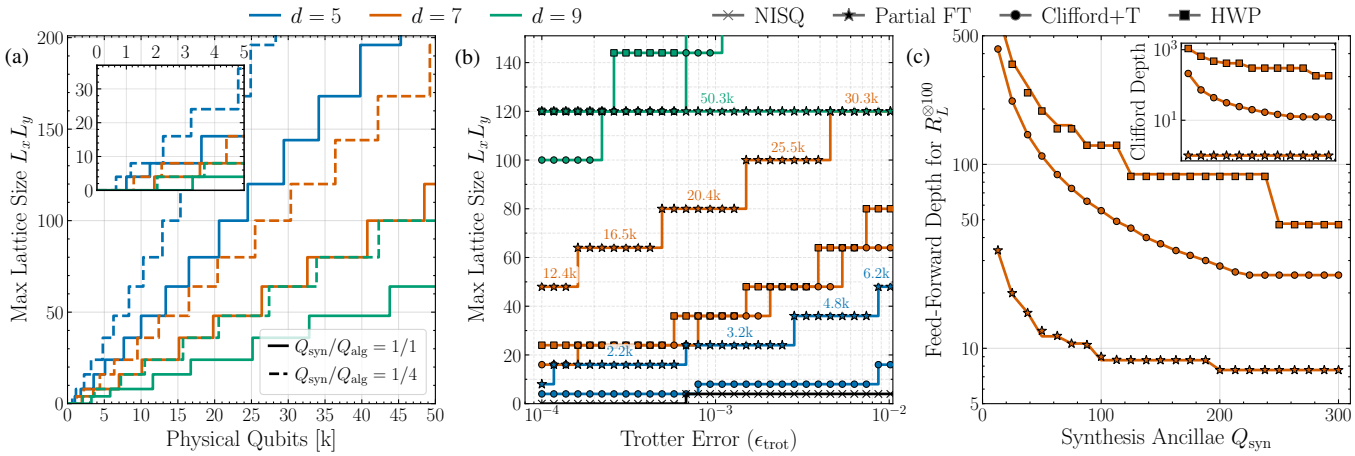


Figure 5. Overview of resource requirements for partially and fully FT simulations. For concreteness, the problem considered is the time evolution of the Fermi-Hubbard model on a square lattice with periodic boundaries. We consider rotated surface encoding at moderate code distances $d = 5, 7, 9$, and physical error rate $p = 10^{-3}$. The partially FT approach (stars) is compared against standard Clifford+T synthesis (circles) and Hamming weight phasing HWP (squares), as well as a NISQ implementation (crosses) when applicable. (a) Maximum lattice size supported by growing hardware sizes. This is independent of the approach, and determined by code distance d and the fraction $Q_{\rm syn}/Q_{\rm alg}$ of logical patches dedicated to the synthesis of non-Clifford operations (solid and dashed lines). The inset provides a closer look to smaller hardware sizes. (b) Maximum lattice size per allowed Trotter error on the final observable, with evolution-time proportional to the lattice's side. At smaller distances, the partially FT approach enables simulation of larger systems, whereas the advantage is lost at distance $d \geq 9$. Labels over partially FT datapoints indicate the number of physical qubits required for simulations at given code distance and lattice size, with $Q_{\rm syn}/Q_{\rm alg} = 1/4$. (c) Circuit depth for the synthesis of one layer of 100 parallel $R_Z(\theta_L)$ rotations, at code distance $d = 7$, given $Q_{\rm syn}$ ancillary logical patches dedicated to the realization of non-Clifford operations. Depth is shown both in terms of the injection of resource states (main) and the associated Clifford depth (inset). Datapoints are numerical estimates, whereas lines are heuristic analytic estimates, showing excellent agreement.

overhead. For typical QEM techniques, the overhead increases exponentially with the accumulated error rate

$$\Sigma_L = \sum_{g \,\in\, \text{circuit}} p_{L,g}, \tag{30}$$

where each gate g has a logical error rate $p_{L,g}$, independently of the circuit's noise propagation. For more details, we refer the reader to the review in Ref. [55]. Therefore, $\Sigma_L \lesssim 1$ is demanded, as a general rule of thumb, to keep the sampling overhead within reasonable limits. In this subsection, we quantify the maximum system size that can be simulated while ensuring that $\Sigma_L \lesssim 1$.

The accumulated logical error depends, in the first place, on the number of simulated Trotter steps, which is fixed by the target Trotter error. This is in turn a function of the evolution time T and the order of the Trotter decomposition. The Trotter error is taken to scale as

$$\epsilon_{\rm trot}^{(\rm ord)} \propto N_{\rm trot} \left(\frac{TJ}{N_{\rm trot}} \right)^{\rm ord+1}, \tag{31}$$

where ord denotes the order of the decomposition. Rather than relying on theoretical bounds on the Trotter error, which are often loose [105, 107, 111], the proportionality constants were inferred from simulations performed with the `ffsim` package [114], as described in App. E. For typical observables—such as average density, occupancy, spin, double occupancy, density correlations, and spin correlations—representative proportionalities 10^{-2} and 3×10^{-3} were inferred for the second- and fourth-order decompositions of the Hubbard Hamiltonian considered in the manuscript. Setting the evolution time to grow with size as $TJ = (L_x + L_y)/2$ (so that information approximately propagates across the whole 2D system as we scale the size), the number of Trotter steps required to bound the Trotter error below a given target is readily estimated. Gate counts are then obtained, given the adopted implementation of the Trotter steps. The implementation we devised is detailed in App. E.

The accumulated logical error then depends on the error model of the adopted gates. We assume a single round of syndrome extraction is interleaved with each operation, as in the framework of algorithmic fault tolerance [32–35]. The available Clifford operations are H, S, S_X, CX, CZ, with error rates given in Table II as derived in App. D 4. Partially FT computation realizes arbitrary-angle $\{R_Z, R_X\}$ rotations by means of the TMR protocol, with error rate Eq. (16). Fully FT computation decomposes them into $\{T, T_X\}$ and Clifford gates. With moderate code distances, the error budget of T gates is assumed to be dominated by the injection process, such that $p_L^{(T)} \simeq p_L^{(\rm inj)}$, defined in Eq. (25). Given near-term constraints, such as limited ancillary resources, rotations in the fully-FT scenario are assumed to be realized by the most efficient *ancilla-free* synthesis protocol, requiring $N_T^{\rm Tdec}(\epsilon) \simeq 1.52 \log_2(\epsilon^{-1}) - 0.01$ T gates per rotation to

Table II. FT gates, followed by a single syndrome extraction round, have logical error rate $p_L = a_\bullet (p/p_{\text{th},\bullet})^{\lceil d/2 \rceil}/(d-1)$, with gate-dependent thresholds $p_{\text{th},\bullet}$ and prefactors $a_\bullet$ provided below, as derived in App. D 4. T gates' noise is assumed to be dominated by Clifford contributions, given the moderate code distances considered in the manuscript.

	H	S & S_X	CX & CZ	T & T_X
Threshold $p_{\text{th},\bullet}$	0.015	0.009	0.008	0.012
Prefactor $a_\bullet$	0.032	0.033	0.025	0.070

achieve precision ϵ [24]. The optimal precision is chosen by minimizing the total error, i.e. the sum of ϵ and the error rates of the T+Clifford gates involved—further details are provided in App. F 2. Finally, a digital NISQ approach is included in the comparison, with a simplified and optimistic error model: error rate p for two-qubit gates and noiseless single qubit operations.

Once the ϵ_{trot}-dependent gate counts and the (d, p)-dependent logical error rates are set, the accumulated error rate Σ_L, governing the cost of error mitigation, can be estimated at given system size. Figure 5(b) shows the maximum system size that can be achieved in a square layout before the accumulated error rate exceeds 1, for different target trotter errors. We recall that evolution time is chosen to scale proportional to the lattice's side. The figure compares partially and fully FT approaches at a physical error rate $p = 0.001$ and code distances $d = 5, 7, 9$. For each data point, we evaluate both second- and fourth-order Trotter formulas and retain the best result. Often, the second-order decomposition was preferable for the partially FT approach, whereas full fault tolerance showed better performance under the fourth-order decomposition.

The NISQ approach included in the comparison shows poor performance compared to that of error-corrected computation. At moderate distances 5 and 7, partial fault tolerance yields the lowest error rates. This is due to the relatively large contribution of errors from Clifford operations, whose usage is significantly reduced in the TMR protocol. At distances 5 and 7, the partially FT protocol targeting Trotter error $\sim 0.5\%$ accommodates lattice sizes 36 and 120 respectively, beyond the reach of current classical approaches and NISQ digital computation for equivalent evolution times [38, 39, 115–117], as well as much larger than the fully FT counterparts. This identifies the STAR architecture as a more promising candidate for near-term demonstrations of quantum advantage. At higher distance, noise from Clifford operations is strongly reduced, and fully FT approaches achieve lower error rates, enabling the simulation of larger system sizes.

This comparison, here outlined for the representative error rate $p = 0.001$, was repeated over the range $p \in [10^{-4}, 5 \times 10^{-3}]$. By comparing the maximum lattice size attainable by error-corrected approaches with target Trotter error 0.5%, regimes of advantage of partial and full fault-tolerance are identified. For the purposes of comparing against current classical methods, the limit of classical simulations of the 2D Hubbard model is assumed to be just below lattice size 36 [38, 39] in the equivalent setting, i.e. on a square lattice with periodic boundaries and with evolution times comparable to the length of the lattice. Moreover, we repeated the comparison among methods under the assumption that $\alpha_{\text{RUS}} = 1.5$ in Eq. (24), as previously estimated for the transversal STAR architecture [28]. Figure 1(d) shows that the reduction of the effective RUS prefactor α_{RUS} obtained in our framework establishes a near-term region in which the transversal STAR architecture is not only a lower-overhead alternative to full fault-tolerance, but yields lower errors than the other approaches. In the long-term, instead, full fault tolerance still appears as a more robust solution, albeit with a significantly larger overhead, as quantified in the next subsection.

C. Limitations from circuit depth

In this subsection, we analyze the circuit depth of different approaches in the realization of parallel arbitrary-angle rotations. We show that the TMR protocol not only requires lower depth, but also that it has a much more advantageous spacetime tradeoff upon reduction of physical resources, independently of the target application.

App. F provides a detailed derivation of the resources required for a maximally parallel implementation of the methods—i.e. with the shortest possible depth—and of the increase in depth following a reduction of these resources. Table III summarizes the findings, and compares the different approaches to gate synthesis. Specifically, it assumes that a batch of b parallel $R_Z(\theta_L)$ rotations is to be implemented, with θ_L a dyadic fraction of π. The table provides reliable analytic approximations of the circuit depth as the availability of ancillae is reduced by a factor $\mathfrak{r}$ with respect to the maximally parallel scenario. We note that while contributions to depth from feed-forward and Clifford operations are provided separately for generality, the total circuit depth is likely less than their direct sum, due to potential concurrent scheduling: this relies on additional practical assumptions and its quantitative characterization is left for future work. The modeling assumes comparable times for the encoding and the injection of resource states, justified in the context of neutral atom implementations, where measurements and shuttling represent the main time bottleneck.

The feed-forward depth of the TMR protocol has a small logarithmic dependence on the batch; rather remarkably, upon reduction of its ancillary reservoir it only receives an additive increase. The only Clifford contribution to its circuit depth is a potential S or $S^\dagger$ correction to the RUS process. In contrast, for Clifford+T synthesis, the depth is set by the target approximation error, with typical T-count $N_T^{\text{Tdec}}(\epsilon) > 20$, and approximately half as many S-type gates. Reducing the ancillary reservoir in-

Table III. Comparison of different synthesis protocols for the realization of a batch of b parallel $R_Z(\theta_L)$ rotations. Estimates are derived in App. F, and represent well-approximating ansatze. The rotation angle is assumed to be a dyadic fraction of π, $\theta_L = (2k+1)\pi/2^{C+1}$ with k and C integers. A_{FT} is the average number of attempts required to encode the relevant fully FT resource state; A_{pFT} is the effective analogous for the partially FT RUS protocol. These depend on code distance and physical noise: typical values for neutral-atom implementations at moderate distance and error rate $p = 0.001$ are $A_{\mathrm{pFT}} \gtrsim 1$ (see Fig. 4(c)) and $A_{\mathrm{FT}} \simeq 1.2$ [46]. $N_T^{\mathrm{Tdec}}(\epsilon)$ is the average number of T gates required to synthesize R_Z rotations with precision ϵ. The reduction factor $\mathfrak{r} \geq 1$ governs the spacetime tradeoff between the availability of ancillary logical patches and circuit depth.

	Ancillae for non-Clifford synthesis	Feed-forward depth per batch	Clifford depth per batch
Partial FT	$\left\lceil (1+A_{\mathrm{pFT}})\frac{b}{\mathfrak{r}} \right\rceil$	$\min(\log_2 b + 1.33, C) + 2\mathfrak{r}\left(1 - 2^{-\min\{C, \lceil \log_2 \mathfrak{r} \rceil\}}\right)$	$\lesssim 1$
Clifford+T	$\left\lceil (1+A_{\mathrm{FT}})\frac{b}{\mathfrak{r}} \right\rceil$	$\left\lceil \mathfrak{r} N_T^{\mathrm{Tdec}}(\epsilon) \right\rceil$	$\left\lceil \mathfrak{r} N_T^{\mathrm{Tdec}}(\epsilon)/2 \right\rceil$
Reduced HWP	$\left\lceil \frac{4}{3}\left(1+\frac{2}{3}A_{\mathrm{FT}}\right)\frac{b}{\mathfrak{r}} \right\rceil$	$\mathfrak{r}\left[2\log_{3/2}\left(\frac{b}{\mathfrak{r}}\right) + N_T^{\mathrm{Tdec}}(\epsilon)\right]$	$\mathfrak{r}\left[16\log_{3/2}\left(\frac{b}{\mathfrak{r}}\right) + \frac{N_T^{\mathrm{Tdec}}(\epsilon)}{2}\right]$

creases both depths multiplicatively. Regarding HWP, its requirements and resources are determined by its specific implementation. As motivated in App. F 3, we considered a "reduced" version of the protocol, which is particularly suited for near-term implementations with moderate code distance and an ancillary reservoir of limited size. While HWP reduces the number of rotations to be synthesized, it increases both ancillae requirements and circuit depth.

The significant difference in circuit depth is evident in Fig. 5(c). The figure shows the circuit depth required to implement one batch of $b = 100$ parallel R_Z rotations against the number Q_{syn} of logical patches available for the synthesis of non-Clifford resources. Typical values for implementations at code distance $d = 7$ and physical error rate $p = 0.001$ are considered, with effective average number of encoding attempts per resource state $A_{\mathrm{pFT}} = 1.05$, $A_{\mathrm{FT}} = 1.2$, and $N_T^{\mathrm{Tdec}} = 25$ T gates per synthesized rotation. As the size of the non-Clifford resources is comparable to or smaller than the batch size, matching near-term constraints, the feed-forward depth of the TMR protocol is roughly an order of magnitude shorter than the FT counterparts. Even once the maximally parallel regime is attained, requiring much larger ancillae reservoirs, the partially FT depth remains ~ 5 times shorter.

In terms of Clifford operations, partial fault-tolerance yields at least an order of magnitude reduction, growing up to three orders of magnitude once the availability of ancillae is reduced. However, we note that in near-term regimes with fewer ancillary patches, non-Clifford operations cannot be fully parallelized, leading where possible to concurrent scheduling of Clifford and non-Clifford gates. This implies that the total depth is not simply the sum of the separate contributions, but is subject to further assumptions—the details of the Clifford+T decomposition, hardware routing, and scheduling optimization—beyond the scope of our work.

VII. SUMMARY AND OUTLOOK

Partial fault tolerance significantly reduces the resources required for robust quantum computation. In the STAR architecture, Clifford gates remain fault-tolerant (FT), while arbitrary-angle rotations are natively implemented, avoiding long Clifford+T synthesis sequences [25–30, 103, 113]. Specifically, arbitrary-angle rotations are realised via the Transversal Multi-Rotation (TMR) protocol, which combines the use of r physical $R_{ZZ}(\theta)$ rotations with syndrome extraction and post-selection, to prepare angle-dependent resource states. These are then injected in a repeat-until-success (RUS) process, as described in Sec. III. This approach reduces the space-time cost of arbitrary rotations, at the price of exposing the protocol to a small set of hardware-dependent faults that must be carefully modeled. In neutral-atom hardware, reconfigurable connectivity enables the execution of transversal operations, further reducing space-time overhead and logical error rates [28].

In this work, we built on these advances by investigating the co-design of partially FT computation and neutral-atom hardware, ranging from the underlying atomic physics to noise propagation in the logical protocol, and the final target application. With a combination of detailed modeling and optimization, partial fault tolerance emerges not only as a lower-overhead option, but also as a more robust route for near-term demonstrations of quantum advantage.

Concretely, we first developed a refined model of noise propagation from the physical operations to the logical protocol (cf. Sec. III). In the first place, we show that the RUS nature of the protocol suppresses the leading coherent contribution to the logical channel, better informing future error mitigation approaches. Secondly, while it is known that dominant undetectable faults arise in the aforementioned r physical $R_{ZZ}(\theta)$ rotations [25–27], our work revealed an intrinsic resilience to their coherent noise, guiding subsequent optimization of its per-

formance. Therefore we tailored the execution of the $R_{ZZ}(\theta)$ Rydberg gates in neutral-atom hardware to the logical protocol, exploiting accurate knowledge of the underlying atomic physics (see Sec. IV). In the first place, our modeling reveals a much more favorable scaling in these rotations than assumed in previous analysis [27, 28], as well as a beneficial noise bias amplified by the possibility of post-selecting on leakage events. Moreover, we further suppressed undetectable physical faults via pulse-level error-selective gate optimization, yielding a two- to three-fold reduction of the critical faults, without additional overhead. We specifically focused on implementations in ^{171}Yb, a popular candidate for neutral-atom quantum computation [58–60], but our findings can be extended to other atomic species. Our refined analysis and optimized implementation predict logical error rates approximately an order of magnitude below previous STAR estimates [27, 28] (cf. Sec. V). Specifically, the error rate of θ_L rotations in the rotated surface code, at distance d and physical error rate p, is well modeled by

$$p_L(\theta_L, p, d, r) \simeq \alpha_{\mathrm{RUS}}(p, r)\, p\, |\theta_L| + 2p_L^{(\mathrm{inj})}(p, d), \quad (32)$$

with $\alpha_{\mathrm{RUS}}(p, r) \simeq r(0.01 + 4p)$ for the optimized Rydberg pulses and circuit-level noise model considered here. The first term captures the angle-dependent contribution specific to the TMR encoding—depending on the number r of physical transversal $R_{ZZ}(\theta)$ rotations—while the second term captures the Clifford and injection errors accumulated over the RUS attempts, two on average.

Then, we benchmarked the applicability and scalability of our framework against fully FT alternatives, namely standard Clifford+T synthesis and Hamming Weight Phasing (see Sec. VI). As a representative case-study, we considered time evolution in the Hubbard model with periodic boundaries, for which we devised a simulation circuit tailored to early FT scenarios. The aforementioned improvements make partial fault tolerance most compelling at moderate code distances $d \leq 7$, where the asymptotic advantages of full fault tolerance are not yet dominant. At physical error rate $p \sim 10^{-3}$, partially FT computation, in combination with error mitigation, is expected to surpass the reach of classical methods and NISQ simulations, as well as that of fully FT alternatives at comparable physical resources. Notably, we predict simulations of 36 sites at $d = 5$ and 100 sites at $d = 7$ to be achievable, with physical footprints of approximately 5 and 25 thousand physical qubits respectively.

Finally, we analyzed the circuit depth and spacetime tradeoffs of the techniques considered (cf Sec. VI C). Even in regimes where fully FT synthesis is expected to achieve lower error rates, the STAR architecture retains a substantially better spacetime tradeoff, with order-of-magnitude reductions in circuit depth. This leads to order-of-magnitude reductions in circuit depth, particularly pronounced when the number of ancillary logical patches is limited. For example, for a batch of $b = 100$ parallel rotations at $d = 7$ and $p = 10^{-3}$, repeat-until-success TMR injection depth with realistic ancillary reservoirs is approximately one order of magnitude shorter than fully FT alternatives, and remains five times shorter even with extensive availability of resource states. The Clifford depth of STAR rotations is lower than the FT counterparts by one to three orders of magnitude, depending on the number of ancillary logical patches available. Therefore, while partial fault tolerance asymptotically yields higher error rates, its depth and hardware efficiency make it a compelling alternative even in the longer term.

Several directions follow from these results. In the first place, partially FT protocols can be extended to alternative error-correcting codes, hardware and applications. The rotated surface code provides a versatile baseline for neutral-atom implementations, but other code families may offer better tradeoffs. In particular, high-rate QLDPC codes are expected to reduce the physical overhead, although their applicability is likely to depend on the target application, as explored in recent works [28, 30]. Due to generally lower thresholds, their advantage is likely to become more pronounced in longer-term implementations. Secondly, the different noise structure of partially FT operations calls for tailored quantum error mitigation techniques, as well as suggesting potential connections to quantum metrology [118, 119]. Our end-to-end analysis and improvements provide a natural starting point for these investigations.

Altogether, our results open new prospects for the resilience and scalability of partially FT computation, and establish a near-term regime in which it is expected to surpass the reach of both classical methods and fully FT approaches. This makes partial fault-tolerance a compelling route to early demonstrations of practical quantum advantage.

VIII. ACKNOWLEDGMENTS

We thank Giacomo Giudice, Martin Kiffner, Sebastian Wüster, and Jonathan Pritchard for stimulating discussions. The authors acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work (http://dx.doi.org/10.5281/zenodo.22558). This work was supported by the EPSRC through the QQQS programme grant (EP/Y01510X/1) and by the QCi3 hub (EP/Z53318X/1). Work by SV was supported by PlanQC GmbH.

AI tools were used in the following form in this work: ChatGPT 5.6 Sol was used to suggest improvements to syntax and readability during the writing process. Different versions, from ChatGPT 5.0 to 5.6 Sol, assisted with debugging and with code generation, proposing adjustments to the simulation code, producing Mathematica and SymPy scripts to verify mathematical derivations, and implementing requested changes in the layout of figures (which were produced separately). AI suggestions were in each case reviewed and edited by the authors and further verified against independent calculations and numerical tests.

Appendix A: Detailed Derivation of the Error Budget of the TMR protocol

In this appendix, analytic estimates for the error budget of the TMR protocol are derived. As it will be quantified, three main error mechanisms limit the performance of the TMR protocol: noise from stabilizer extraction and transversal operations, undetected correlated phase-flips, and undetected repeating bit-flips. The effect of other faults is included in numerical simulations, but not dealt with analytically in this appendix. This appendix is organized as follows:

(i) App. A 1 provides a few remarks on the adopted error metrics, i.e. average infidelity and diamond-norm error.

(ii) App. A 2 briefly comments on the error contributions from noisy Clifford operations.

(iii) App. A 3 and A 4 introduce the two sources of noise specific to the TMR protocol—which *partial* fault tolerance owes its name to—and models their effects on the resource state.

(iv) App. A 5 quantifies error propagation onto the target logical qubits upon gate-teleportation.

(v) App. A 6 quantifies error propagation in the whole repeat-until-success scheme, showing error accumulation of incoherent noise, but also a remarkable cancellation of leading coherent noise.

For the purposes of clarifying the notation and procedure, it should be recalled from the main text that the TMR protocol aims to encode the resource state $|m_{\theta_L}\rangle = R_{Z,L}(\theta_L)\,|+\rangle$, where $R_{Z,L}(\theta_L) = \exp(-i\theta_L Z_L/2)$.

It does so by applying r transversal rotations, each of weight w, followed by stabilizer measurements and post-selection. In this appendix, $Z_q^{\otimes w}$ denotes the tensor product of the Z operator over w consecutive qubits, starting at index wq. Then, the transversal rotations have form $\bigotimes_{q=0}^{r-1} R_{Z_q^{\otimes w}}(\theta)$, for some angle θ. The noise $\mathcal{E}$ acting on each transversal rotation is modeled by the chi-matrix $\chi_{P,P'}$, such that

$$\mathcal{E}(\rho) = \sum_{P,P'\in\mathcal{P}^{\otimes w}} \chi_{P,P'} P\rho P', \tag{A1}$$

with $\mathcal{P}^{\otimes w}$ the set of w-qubit Pauli operators.

1. Error Metrics

As mentioned in the main text, the adopted figures of merit are the average infidelity $1-F$, and the diamond-norm error $\varepsilon_\diamond$. The latter, also known as the worst-case error rate, is defined as the diamond-norm distance between the noise channel $\mathcal{E}$ that we wish to characterize and the identity channel $\mathcal{I}$,

$$\varepsilon_\diamond(\mathcal{E}) = \frac{1}{2}\|\mathcal{E}-\mathcal{I}\|_\diamond = \frac{1}{2}\sup_\rho \|\left(\mathcal{E}\otimes\mathcal{I}-\mathcal{I}^{\otimes 2}\right)(\rho)\|_1, \tag{A2}$$

where $\|A\|_1 = \mathrm{Tr}\sqrt{A^\dagger A}$, and the supremum is over all density matrices of dimension D^2, with D the Hilbert-space dimension of the system acted upon by $\mathcal{E}$. The diamond-norm error is intuitively understood as the maximum probability of distinguishing the outputs of the noisy and ideal operation [98, 99]. It is sub-additive, and it plays a central role in rigorous proofs of fault tolerance thresholds [98–101]. Notably, the diamond-norm error is more sensitive than the average infidelity to coherent errors [96, 97].

As an example—which will be of great interest in the remainder of the appendix—given a general completely-positive trace-preserving Z-error channel of the form

$$\mathcal{N}(\rho;p,q) = (1-p)\rho + pZ\rho Z - (iqZ\rho + \text{h.c.}), \tag{A3}$$

where p, q are real, then $1-F = 2p/3$, whereas $\varepsilon_\diamond = \sqrt{p^2+q^2} = p\sqrt{1+(q/p)^2}$ [97]. This implies that the average infidelity and diamond-norm error between rotations by a target angle θ_0 and an incorrect angle θ_1 are

$$\begin{aligned} 1-F(\theta_0,\theta_1) &= \frac{2}{3}\sin^2\left(\frac{\theta_0-\theta_1}{2}\right), \\ \varepsilon_\diamond(\theta_0,\theta_1) &= \left|\sin\left(\frac{\theta_0-\theta_1}{2}\right)\right|, \end{aligned} \tag{A4}$$

such that potentially $\varepsilon_\diamond \gg 1-F$. Finally, we note that, allowing for a non-CPTP channel of the form in Eq. (A3) with complex q, as it is possible upon post-selection, the diamond distance generalizes to $\varepsilon_\diamond = \sqrt{p^2+|q|^2}$, as tested numerically.

2. Noise from QEC

The first contribution to logical faults is given by noise in the syndrome extraction and state injection. This is common to FT quantum computation in general, and it is exponentially suppressed with the code distance d, provided that the physical error rate p is below threshold p_{th}. It is often well captured by an effective Pauli-noise model, with

$$p_L^{(\text{inj})} \simeq a_{\text{inj}}\left(\frac{p}{p_{\text{th}}}\right)^{\lceil\frac{d+1}{2}\rceil}, \tag{A5}$$

where a_{inj} and p_{inj} depend on the physical noise model and details of the specific implementation. For stochastic Pauli noise, $1-F^{(\text{inj})} = 2p_L^{\text{inj}}/3$ and $\varepsilon_\diamond^{(\text{inj})} = p_L^{(\text{inj})}$.

3. Noisy resource state under correlated phase-flips

The first error mechanism specific to the TMR protocol originates from undetected correlated phase flips that occur shortly before or shortly after the transversal $R_{Z^{\otimes w}}$ rotations, and is typically the dominant error. To understand how these flips affect the protocol, recall Eq. (6): the transversal rotations amount to the sum of the desired θ_L rotation, plus other Z-error terms that map the logical state outside the logical subspace. Assuming that, for even r, the $S_X^\dagger$ correction is applied as specified in the previous section, the error terms that are closest to the logical subspace—hence dominating the error budget—are

$$-i\sum_{q=0}^{r-1}\sqrt{p_{\theta_1}}\,Z_q^{\otimes w}\;|m_{(-1)^{\lfloor r/2\rfloor}\theta_1}\rangle_L\,, \tag{A6}$$

where the effective rotation angle θ_1 satisfies

$$\tan\frac{\theta_1}{2}=-\tan^{r-2}\frac{\theta}{2}=-\tan^{(1-\frac{2}{r})}\frac{\theta_L}{2}, \tag{A7}$$

and the probability of collapsing onto each one of these terms is

$$p_{\theta_1}(\theta,r)=\cos^{2(r-1)}\frac{\theta}{2}\sin^2\frac{\theta}{2}+\sin^{2(r-1)}\frac{\theta}{2}\cos^2\frac{\theta}{2}. \tag{A8}$$

Therefore, if the logical state collapses onto any of these error terms, and, in addition, the corresponding $Z_q^{\otimes w}$ error occurs shortly before or after [120], then the net effect is the production of the logical state $|m_{\theta_1}\rangle_L$: the state is rotated incorrectly, yet undetectably so.

Denoting by p the circuit-level error rate, and by χ the chi-matrix of the error channel acting on each of the transversal rotations, the incoherent probability of correlated-phase flips on any of the physical rotations is given by

$$p_{Z^{\otimes w}}\simeq r\,[a_Z\,p^w+\chi_{Z^{\otimes w},Z^{\otimes w}}(\theta)]. \tag{A9}$$

Moreover, in contrast to previous works, we provide a complete modeling that does not neglect the contribution of coherent (i.e. off-diagonal) errors in the physical rotations, which amounts to

$$q_{Z^{\otimes w}}\simeq -i\,r\,\chi_{Z^{\otimes w},I}(\theta). \tag{A10}$$

Then, after the transversal rotations and stabilizers measurements, the resource state is in leading order

$$\begin{aligned}m_{\rm un}^{(Z^{\otimes w})}=&(1-p_{Z^{\otimes w}})\,p_{\theta_L}\;|m_{\theta_L}\rangle\langle m_{\theta_L}|_L\\&+\;p_{Z^{\otimes w}}\,p_{\theta_1}\;|m_{\theta_1}\rangle\langle m_{\theta_1}|_L\\&+\qquad q_{\rm off}\;|m_{\theta_1}\rangle\langle m_{\theta_L}|_L+\text{h.c.},\end{aligned} \tag{A11}$$

where $q_{\rm off}=q_{Z^{\otimes w}}\sqrt{p_{\theta_L}p_{\theta_1}}$. Note, the resource state is currently unnormalised due to post-selection. Denoting $\Delta=\theta_1-\theta_L$, this can then be re-written as

$$\begin{aligned}m_{\rm un}^{(Z^{\otimes w})}(\theta_L)=&(1-\,p_{Z^{\otimes w}})\,p_{\theta_L}\;|m_{\theta_L}\rangle\langle m_{\theta_L}|_L\\&+p_{Z^{\otimes w}}\,p_{\theta_1}\,R_{Z,L}(\Delta)\,|m_{\theta_L}\rangle\langle m_{\theta_L}|_L\,R_{Z,L}^\dagger(\Delta)\\&+q_{\rm off}\,R_{Z,L}(\Delta)\,|m_{\theta_L}\rangle\langle m_{\theta_L}|_L+\text{h.c.}.\end{aligned} \tag{A12}$$

4. Noisy resource state under repeating flips

The second error contribution specific to the TMR protocol originates from undetected flips that anticommute with the Z_L operator. Practically, such faults require two events:

1. an X- or Y-flip shortly before the rotations, which effectively reverses the rotation direction, as $R_{Z_1^{\otimes w}}X_1=X_1R_{Z_1^{\otimes w}}^\dagger$, and occurs with probability $O(p)$;

2. a second flip shortly after, either during the rotation itself with probability $\chi_{X,X}$ or $\chi_{Y,Y}$, or in one of the subsequent gates with probability $O(p)$.

In summary, the first flip inverts the sign of rotation, while the second makes the fault go undetected. The probability of this combination of errors scales linearly with the number of qubits in the support of Z_L, which is the smaller code distance d_s at the time of the transversal rotations.

Therefore, the probability $p_{\rm flip}$ for this error reads

$$p_{\rm flip}=d_s[a_X\,p^2+a_{X,r}\,p\,(\chi_{X,X}+\chi_{Y,Y})(\theta)], \tag{A13}$$

where $a_X,a_{X,r}>0$ re circuit- and noise-dependent prefactors arising from the combinatorics of the contributing error processes; $(\chi_{X,X}+\chi_{Y,Y})$ depend on the physical angle of rotation θ. In this scenario, off-diagonal errors in the transversal rotations, e.g. $\chi_{X,Y}$, are projected out by stabilizer measurements. As this error anticommutes with Z_L, it realizes a flipped rotation $R_{Z,L}(-\theta_L)$. Then, to leading order, the noisy resource state is

$$\begin{aligned}m_{\rm un}^{(\rm flip)}(\theta_L)\simeq&\;(1-p_{\rm flip})p_{\theta_L}\,|m_{\theta_L}\rangle\langle m_{\theta_L}|_L\\&+p_{\rm flip}\,p_{\theta_L}\;|m_{-\theta_L}\rangle\langle m_{-\theta_L}|_L\,.\end{aligned} \tag{A14}$$

Note, the state is currently unnormalised due to post-selection.

In previous studies [27, 28], this contribution was neglected because their noise models assumed $\chi_{Z^{\otimes w},Z^{\otimes w}}\sim O(p)$, making the correlated phase-flips (see App. A 3) overwhelmingly dominant. As it is shown in later sections, this is not necessarily the case. Moreover, due to its quadratic dependence on the error rate, this noise source can contribute to the ultimate limits of the TMR protocol, independently of the code distance and of the implementation of the transversal rotations.

5. Gate-teleportation Noise

In this subsection, we first derive the noisy state $\mathfrak{M}(\theta_L)$ of the resource qubit. Then, we quantify noise propagation onto the input logical qubit Ψ through the teleportation channel $\mathcal{T}^{(\mu)}(\Psi;\theta_L)$, conditioned on the teleportation measurement $\mu=0$ or 1.

In the previous subsections, we defined $m_{\rm un}^{(Z^{\otimes w})}(\theta_L)$ and $m_{\rm un}^{(\rm flip)}(\theta_L)$ as the unnormalised resource states under the most relevant error sources: correlated phase-flips and repeating flips, respectively. The resulting state under their combined action is

$$\begin{aligned}\mathfrak{M}(\theta_L) = &\ (1-x_{\theta_L})\,|m_{\theta_L}\rangle\langle m_{\theta_L}|_L \\ &+ x_{\theta_L}\,Z_L\,|m_{\theta_L}\rangle\langle m_{\theta_L}|_L\,Z_L \\ &- iy_{\theta_L}Z_L\,|m_{\theta_L}\rangle\langle m_{\theta_L}|_L + \text{h.c.},\end{aligned} \tag{A15}$$

where, in leading order, the noise coefficients are

$$\begin{aligned}x_{\theta_L} &\simeq \frac{p_{\theta_L}}{n_{\theta_L}}\left[p_{Z^{\otimes w}}\,p_{\theta_1}\sin^2\frac{\Delta}{2} + p_{\rm flip}\,p_{\theta_L}\sin^2\theta_L\right], \\ y_{\theta_L} &\simeq \frac{p_{\theta_L}}{n_{\theta_L}}\left[\left(p_{Z^{\otimes w}}\,p_{\theta_1}\cos\frac{\Delta}{2} + q_{\rm off}\right)\sin\frac{\Delta}{2}\right. \\ &\qquad\left. - p_{\rm flip}\,p_{\theta_L}\sin\theta_L\right],\end{aligned} \tag{A16}$$

with normalisation factor n_{θ_L}—close to one in the small angle scenario—defined by

$$\frac{n_{\theta_L}}{p_{\theta_L}} = \left[(1-p_{Z^{\otimes w}})p_{\theta_L} + p_{Z^{\otimes w}}\,p_{\theta_1} + 2\,\mathrm{Re}(q_{\rm off})\cos\frac{\Delta}{2}\right], \tag{A17}$$

where terms of order $O(p_{\rm flip}\,p_{Z^{\otimes w}} + p_{\rm flip}\,q_{\rm off})$ have been neglected. To model gate teleportation, note that the resource state has matrix elements

$$\begin{aligned}\mathfrak{M}_{0,0}(\theta_L) &= [1+2\,\mathrm{Im}(y_{\theta_L})]/2 \\ \mathfrak{M}_{0,1}(\theta_L) &= e^{-i\theta_L}\,[1-2x_{\theta_L}-2i\,\mathrm{Re}(y_{\theta_L})]/2 \\ \mathfrak{M}_{1,0}(\theta_L) &= e^{+i\theta_L}\,[1-2x_{\theta_L}+2i\,\mathrm{Re}(y_{\theta_L})]/2 \\ \mathfrak{M}_{1,1}(\theta_L) &= [1-2\,\mathrm{Im}(y_{\theta_L})]/2.\end{aligned} \tag{A18}$$

Upon gate teleportation, the Z_L-type errors on the noisy resource state $\mathfrak{M}(\theta_L)$ propagate on the target logical qubit, originally in state Ψ. To perform teleportation, a $\mathrm{CX} = |0\rangle\langle 0|\otimes\hat{\mathbb{I}} + |1\rangle\langle 1|\otimes\hat{X}$ gate controlled on Ψ targeting $\mathfrak{M}$ is applied, yielding

$$\begin{aligned}\mathrm{CX}\,(\Psi\otimes\mathfrak{M})\,\mathrm{CX}^\dagger = &\ \Psi_{00}\,|0\rangle\langle 0|\otimes\mathfrak{M} \\ &+ \Psi_{01}\,|0\rangle\langle 1|\otimes\mathfrak{M}X \\ &+ \Psi_{10}\,|1\rangle\langle 0|\otimes X\mathfrak{M} \\ &+ \Psi_{11}\,|1\rangle\langle 1|\otimes X\mathfrak{M}X.\end{aligned} \tag{A19}$$

This is followed by the measurement of the resource state with outcome μ, yielding the unnormalised output

$$\begin{pmatrix}\Psi_{00}\,\mathfrak{M}(\theta_L)_{0\oplus\mu,0\oplus\mu} & \Psi_{01}\,\mathfrak{M}(\theta_L)_{0\oplus\mu,1\oplus\mu} \\ \Psi_{10}\,\mathfrak{M}(\theta_L)_{1\oplus\mu,0\oplus\mu} & \Psi_{11}\,\mathfrak{M}(\theta_L)_{1\oplus\mu,1\oplus\mu}\end{pmatrix}. \tag{A20}$$

We denote the teleportation outcome $\mathcal{T}^{(\mu)}(\Psi;\theta_L)$ as twice this post-measurement state, as normalized for real y_{θ_L}. This can be re-written as

$$\begin{aligned}\mathcal{T}^{(\mu)}(\Psi;\theta_L) = &\ (1-x_{\theta_L})\,\Psi_{(-1)^\mu\theta_L} \\ &+ x_{\theta_L}\,Z_L\Psi_{(-1)^\mu\theta_L}Z_L \\ &- iy^{(\mu)}_{\theta_L}Z_L\Psi_{(-1)^\mu\theta_L} + \text{h.c.},\end{aligned} \tag{A21}$$

where $\Psi_\bullet = R_{Z,L}(\bullet)\,\Psi\,R^\dagger_{Z,L}(\bullet)$, such that $\Psi_{(-1)^\mu\theta_L}$ is the noiseless output upon the recorded measurement, whereas $y^{(0)}_{\theta_L} = -y^{(1)}_{\theta_L} = y_{\theta_L}$. In fact, upon opposite measurements, both the rotation angle θ_L and the coherence y_{θ_L} are flipped, as $\mathfrak{M}_{i,j}$ transforms into $\mathfrak{M}_{i\oplus 1,j\oplus 1}$ under sign-flip of both θ_L and y_{θ_L}.

The normalized output is

$$\tilde{\mathcal{T}}^{(\mu)}(\Psi;\theta_L) = \frac{\mathcal{T}^{(\mu)}(\Psi;\theta_L)}{2P^{(\mu)}(\Psi;\theta_L)}, \tag{A22}$$

where

$$P^{(\mu)}(\Psi;\theta_L) = \frac{1}{2} + (-1)^\mu\mathrm{Im}(y_{\theta_L})\mathrm{Tr}(Z_L\Psi) \tag{A23}$$

is the probability the μ measurement is observed, conditioned on the input Ψ.

We note that $\tilde{\mathcal{T}}(\Psi;\theta_L)$ is also protocol- and noise-dependent, as most quantities depend on the number r of transversal rotations and the noise strength p. The dependence is kept implicit for compactness. The resulting infidelity and diamond-norm error are $2x_{\theta_L}/3$ and $\sqrt{x^2_{\theta_L}+|y_{\theta_L}|^2}$ respectively. However, these values only describe an individual instance of gate teleportation. As shown in the next subsection, complex error accumulation and cancellation occurs within the whole repeat-until-success protocol.

6. Noise propagation in the RUS process

Gate teleportation is probabilistic, as the gate is correctly injected with probability $\sim 1/2$ upon successful measurement, whereas the inverse gate is injected upon opposite measurement. A flipped gate can be corrected simply by implementing gate-teleportation again, with double the original angle. Therefore, a complete gate injection protocol is a repeat-until-success (RUS) scheme, in which gate teleportation is repeated until the desired measurement is observed. At every $s^{\rm th}$ step, with $s = 1, 2, \ldots$, a different angle $\theta_{L,s}$ is injected (see main text).

The previous subsection modeled a single teleportation attempt. In this subsection, error propagation in the whole RUS protocol is examined. It will be useful to define the following ansatz describing the action of a $R_{Z,L}(\theta)$ rotation on an input Ψ, followed by noise, whose incoherent and coherent error contributions are x, y :

$$\begin{aligned}\mathcal{N}(\Psi;\,\theta,x,y) = &\ (1-x)\,\Psi_\theta + x\,Z_L\,\Psi_\theta\,Z_L \\ &- iy\,Z_L\,\Psi_\theta\ + \text{h.c.},\end{aligned} \tag{A24}$$

recalling that Ψ_θ denotes $R_{Z,L}(\theta)\ \Psi\ R^\dagger_{Z,L}(\theta)$. Coefficients x, y generally depend on the angle of rotation θ.

As it will become clear at the end of the subsection, it is now convenient to work with generally unnormalised states. Extending the notation adopted in previous subsections, the action of gate-teleportation at the s^{th} step of the RUS process on an input Ψ can be modeled as

$$\mathcal{T}_s^{(\mu)}(\Psi;\theta_L) = \mathcal{N}\left(\Psi;\ (-1)^\mu\theta_{L,s}, x_s, y_s^{(\mu)}\right) \tag{A25}$$

where μ is the difference between the observed and target measurement. The x_s, y_s error weights denote the error with respect to the state associated to the observed measurement μ, and, owing to their angle-dependence, they are defined by s. As commented in the previous subsection, and critical to the error modeling of the RUS process, the diagonal error x is independent of the measurement outcome, whereas the coherence y is inverted upon inverted measurement, i.e. $y_s^{(1)} = -y_s^{(0)}$, due to the one-sided action of the Z operator propagating onto the measurement outcome.

While $\mathcal{T}_s^{(\mu)}$ describes the action of each individual instance of gate-teleportation, let $\rho_s^{(\mu)}(\rho;\theta_L)$ be the cumulative result after the s^{th} step. Thus, it is defined as

$$\rho_s^{(\mu)}(\Psi;\theta_L) = \mathcal{T}_s^{(\mu)} \circ \mathcal{T}_{s-1}^{(1)} \circ ... \circ \mathcal{T}_1^{(1)}(\Psi;\theta_L), \tag{A26}$$

and can be modeled as

$$\rho_s^{(\mu)}(\Psi;\theta_L) = \mathcal{N}(\Psi, \Theta_s^{(\mu)};\ p_s^{(\mu)}, q_s^{(\mu)}) + \delta_s^{(\mu)}\Psi_{\Theta_s^{(\mu)}}, \tag{A27}$$

where $\Theta_s^{(\mu)} = [(-1)^\mu 2^{s-1} - (2^{s-1}-1)]\theta_L$, such that $\Theta_s^{(0)} = \theta_L$, whereas $\Theta_s^{(1)}$ is the accumulation of wrongful rotations; $\delta_s^{(\mu)}$, defined shortly, tracks a variation in the trace stemming from the use of unnormalised states and complex $y_s^{(\mu)}$. These relations allow to study the propagation of errors in between attempts of gate-teleportation. As $\rho_s^{(\mu)}(\Psi;\theta_L) = \mathcal{T}_s^{(\mu)}(\rho_{s-1}^{(1)};\theta_L)$, we observe the recursive relation

$$\begin{aligned}
p_s^{(\mu)} &= p_{s-1}^{(1)} + x_s\ - 2x_s\, p_{s-1}^{(1)} + 2\text{Re}(q_{s-1}^{(1)*}\, y_s^{(\mu)}) \\
&\quad +\ x_s\delta_{s-1}^{(1)}, \\
q_s^{(\mu)} &= q_{s-1}^{(1)} + y_s^{(\mu)} - 2x_s\text{Re}(q_{s-1}^{(1)}) - 2p_{s-1}^{(1)}\text{Re}(y_s^{(\mu)}) \\
&\quad +\ y_s^{(\mu)}\delta_{s-1}^{(1)}, \\
\delta_s^{(\mu)} &= \delta_{s-1}^{(1)} + 4\,\text{Im}(y_s^{(\mu)})\text{Im}(q_{s-1}^{(1)}).
\end{aligned} \tag{A28}$$

It can be noted that the form of the update of $\delta_s^{(\mu)}$ recalls that of the non-hermitian term in Eq. (A23). To leading order, unravelling the recursive relations yields

$$\begin{aligned}
p_s^{(\mu)} &\simeq \sum_{t=1}^{s} x_t + 2\text{Re}\left(y_s^{(\mu)}\sum_{t=1}^{s-1} y_t^{(1)*} + \sum_{t=2}^{s-1}\sum_{u=1}^{t-1} y_t^{(1)*} y_u^{(1)}\right), \\
q_s^{(\mu)} &\simeq y_s^{(\mu)} + \sum_{t=1}^{s-1} y_t^{(1)}, \\
\delta_s^{(\mu)} &\simeq 4\,\text{Im}\left(y_s^{(\mu)}\right)\sum_{t=1}^{s-1}\text{Im}\left(y_t^{(1)}\right) \\
&\quad + 4\sum_{t=2}^{s-1}\text{Im}\left(y_t^{(1)}\right)\sum_{u=1}^{t-1}\text{Im}\left(y_u^{(1)}\right),
\end{aligned} \tag{A29}$$

up to, respectively, $O(x_s^2 + x_s y_s^2 + y_s^4)$, $O(x_s y_s + y_s^3)$, and $O(x_s y_s^2 + y_s^4)$.

The RUS process terminates once a measurement $\mu = 0$ is obtained. Thus, its final result is the average of the $\mu = 0$ normalized cumulative outputs $\tilde{\rho}_s^{(0)}(\Psi,\theta_L)$, weighted by the probability of their occurrence, which we denote $\Pi_s^{(0)}(\Psi;\theta_L)$. The normalized states are described by

$$\begin{aligned}
\tilde{\rho}_s^{(0)}(\Psi;\theta_L) &= \tilde{\mathcal{T}}_s^{(0)} \circ \tilde{\mathcal{T}}_{s-1}^{(1)} \circ ... \circ \tilde{\mathcal{T}}_1^{(1)}(\Psi;\theta_L) \\
&= \frac{1}{2^s\Pi_s^{(0)}}\ \mathcal{T}_s^{(0)} \circ \mathcal{T}_{s-1}^{(1)} \circ ... \circ \mathcal{T}_1^{(1)}(\Psi;\theta_L) \\
&= \frac{1}{2^s\Pi_s^{(0)}}\ \rho_s^{(0)}(\Psi;\theta_L).
\end{aligned} \tag{A30}$$

Therefore, the final result of the RUS process is best described by the unnormalised outputs of the protocol, as

$$\begin{aligned}
\rho_{\text{RUS}}(\Psi;\theta_L) &= \sum_{s=1}^{\infty}\Pi_s^{(0)}(\Psi;\theta_L)\ \tilde{\rho}_s^{(0)}(\Psi;\theta_L) \\
&= \sum_{s=1}^{\infty}\frac{1}{2^s}\rho_s^{(0)}(\Psi;\theta_L) \\
&= \mathcal{N}(\Psi;\ \theta_L, p_{\text{RUS}}, q_{\text{RUS}}) + \delta_{\text{RUS}}(\theta_L)\Psi_{\theta_L},
\end{aligned} \tag{A31}$$

recalling that $\Psi_{\theta_L} = R_{Z,L}(\theta_L)\ \Psi\ R^\dagger_{Z,L}(\theta_L)$ denotes the correctly rotated state, and we where have

$$\begin{aligned}
p_{\text{RUS}}(\theta_L) &= \sum_{s=1}^{\infty}\frac{1}{2^s}p_s^{(0)}(\theta_L), \\
q_{\text{RUS}}(\theta_L) &= \sum_{s=1}^{\infty}\frac{1}{2^s}q_s^{(0)}(\theta_L), \\
\delta_{\text{RUS}}(\theta_L) &= \sum_{s=1}^{\infty}\frac{1}{2^s}\delta_s^{(0)}(\theta_L).
\end{aligned} \tag{A32}$$

As it will be argued in the following subsection, $\delta_{\text{RUS}}(\theta_L) \simeq 0$.

a. Noise Cancellation in the Clifford-terminated RUS

Previous work [27, 28] has observed that, by choosing $\theta_L = \pi/2^{C+1}$, the RUS process is terminated deterministically after C unsuccessful steps by the fully-FT application of an S gate, as the $(C+1)^{\text{th}}$ step requires the injection of $\theta_{L,C+1} = \pi/2$. In the first place, we highlight that this Clifford termination is applicable to any dyadic fraction of $\pi/2$, i.e.

$$\theta_L = \frac{\pi}{2}\frac{2k+1}{2^C}, \tag{A33}$$

for any integer k. The form of Eq. (A33) drastically expands the set of convenient choices for θ_L for algorithmic purposes, as it provides a dense set of angles, capable of approximating any other up to arbitrary precision.

Remarkably, we observe that the Clifford-terminated RUS (ctRUS) protocol has *near-vanishing* coherent error. For the moment, we assume that the terminating Clifford S correction is noiseless. The output of the Clifford-terminated RUS protocol is

$$\rho_{\text{ctRUS}}(\Psi;\theta_L,C) = \sum_{s=1}^{C}\frac{1}{2^s}\rho_s^{(0)}(\Psi;\theta_L) + \frac{1}{2^C} S\,\rho_C^{(1)}(\Psi;\theta_L)\,S^\dagger. \tag{A34}$$

The perfect Clifford correction maps the final failed branch back to the target logical rotation. Therefore,

$$S\rho_C^{(1)}(\Psi;\theta_L)S^\dagger = \mathcal{N}\left(\Psi;\,\theta_L, p_C^{(1)}, q_C^{(1)}\right) + \delta_C^{(1)}\Psi_{\theta_L}. \tag{A35}$$

It follows that the full Clifford-terminated protocol is described by a single effective noise channel,

$$\rho_{\text{ctRUS}}(\Psi;\theta_L,C) = \mathcal{N}\left(\rho;\,\theta_L, p_{\text{ctRUS}}, q_{\text{ctRUS}}\right) + \delta_{\text{ctRUS}}(\theta_L,C)\Psi_{\theta_L}, \tag{A36}$$

with

$$\begin{aligned} p_{\text{ctRUS}}(\theta_L,C) &= p_{\text{RUS}}^{(C)} + \frac{1}{2^C}p_C^{(1)},\\ q_{\text{ctRUS}}(\theta_L,C) &= q_{\text{RUS}}^{(C)} + \frac{1}{2^C}q_C^{(1)},\\ \delta_{\text{ctRUS}}(\theta_L,C) &= \delta_{\text{RUS}}^{(C)} + \frac{1}{2^C}\delta_C^{(1)}, \end{aligned} \tag{A37}$$

as $p_{\text{RUS}}^{(C)}$, $q_{\text{RUS}}^{(C)}$, and $\delta_{\text{RUS}}^{(C)}$ denote the weighted contribution of the successful RUS branches appearing before Clifford termination. These are, to leading order,

$$\begin{aligned} p_{\text{RUS}}^{(C)} &\simeq \sum_{s=1}^{C}\left[\left(\frac{1}{2^{s-1}}-\frac{1}{2^C}\right)x_s - \frac{2}{2^C}\text{Re}\left(y_s^{(0)}\sum_{t=1}^{s-1}y_t^{(0)*}\right)\right],\\ q_{\text{RUS}}^{(C)} &\simeq \frac{1}{2^C}\sum_{s=1}^{C}y_s^{(0)},\\ \delta_{\text{RUS}}^{(C)} &\simeq -\frac{4}{2^C}\sum_{s=2}^{C}\text{Im}\left(y_s^{(0)}\right)\sum_{t=1}^{s-1}\text{Im}\left(y_t^{(0)}\right). \end{aligned} \tag{A38}$$

Before proceeding further, we can note that even in absence of cut-off, for $C\to\infty$, both δ_{RUS} and q_{RUS} approach zero. The final failed branch has, to leading order,

$$\begin{aligned} p_C^{(1)} &\simeq \sum_{s=1}^{C}x_s + 2\text{Re}\left[y_C^{(0)}\sum_{s=1}^{C-1}y_s^{(0)*} + \sum_{s=1}^{C-1}y_s^{(0)}\sum_{t=1}^{s-1}y_t^{(0)*}\right],\\ q_C^{(1)} &\simeq -\sum_{s=1}^{C}y_s^{(0)},\\ \delta_C^{(1)} &\simeq 4\sum_{s=2}^{C}\text{Im}\left(y_s^{(0)}\right)\sum_{t=1}^{s-1}\text{Im}\left(y_t^{(0)}\right). \end{aligned} \tag{A39}$$

Substituting Eqs. (A38) and (A39) into Eq. (A37), we observe that the leading coherent contribution cancels. The final state in Eq. (A36) has noise coefficients

$$\begin{aligned} p_{\text{ctRUS}}(\theta_L,C) &= \sum_{s=1}^{C}\frac{1}{2^{s-1}}x_s + O(x_s^2 + x_s y_s^2 + y_s^4),\\ q_{\text{ctRUS}}(\theta_L,C) &= O(x_s y_s + y_s^3),\\ \delta_{\text{ctRUS}}(\theta_L,C) &= O(x_s y_s^2 + y_s^4). \end{aligned} \tag{A40}$$

In practice, the terminating Clifford gate is generally noisy. If its stochastic error probability is denoted by $p_L^{(S)}$, then the above expressions neglect contributions of order $2^{-C}p_L^{(S)}$, together with higher-order products between the noise of the S gate and the RUS steps. The Clifford error rate $p_L^{(S)}$ is suppressed exponentially in code distance, and these terms are further reduced by the probability 2^{-C} of reaching the terminating branch, hence are typically negligible. Moreover, since the terminating correction is a Clifford operation, its noise can be Pauli-twirled [53, 54], hence it does not contribute a coherent-noise term to q_{ctRUS}.

Altogether, the termination cancels the leading contribution to the coherent-noise coefficient q_{ctRUS}, as well as the $O(y_s^2)$ coherent-noise contribution to p_{ctRUS}. The corresponding branch-normalisation contribution δ_{ctRUS} cancels as well. This provides an intrinsic error-suppression mechanism for dyadic choices of θ_L, in addition to giving a natural deterministic cutoff for the RUS process. Moreover, this result extends to other Clifford-terminated probabilistic processes, such as the fully FT injection of T gates.

Appendix B: Details on Rydberg Gates Optimization

1. On the negligible impact of axial motion

Rydberg gates in alkali atoms (Rb, Cs) are significantly impacted not only by radial fluctuations, but axial too. In this subsection, we estimate the contribution of axial position fluctuations to the variance of the Rydberg interaction for the ytterbium parameters used in this work. Specifically, we show that, within a semiclassical picture

of atomic motion, their impact is an increase in interaction fluctuations by just a few percentage points.

Under the assumption of harmonic traps and thermal equilibrium, atoms experience shot-to-shot fluctuations in position and velocity with variance

$$\sigma_x^2 \simeq \frac{\hbar}{2m\omega_x}\coth\left(\frac{\beta\hbar\omega_x}{2}\right),$$
$$\sigma_{v_x}^2 \simeq \frac{\hbar\omega_x}{2m}\coth\left(\frac{\beta\hbar\omega_x}{2}\right), \tag{B1}$$

where ω_x is the trap frequency along the axis considered.

In our setup, we take the interatomic separation to be $\vec{R} = (R_0 + \Delta x,\, \Delta y,\, \Delta z)$, with the radial trap frequency inducing fluctuations in the x, y plane, whereas z is identified as the axial direction. The Rydberg potential is $V = V_0(R/R_0)^{-6}$, thus

$$V = V_0\left[\left(1+\frac{\Delta x}{R_0}\right)^2 + \frac{\Delta y^2}{R_0^2} + \frac{\Delta z^2}{R_0^2}\right]^{-3}. \tag{B2}$$

The standard deviations of the relevant terms are

$$\begin{aligned}\sigma_{\Delta x} = \sigma_{\Delta y} = \sqrt{2}\sigma_x &\simeq 0.019R_0,\\ \sigma_{\Delta x^2} = \sigma_{\Delta y^2} = 2\sqrt{2}\sigma_x^2 &\simeq 0.0005R_0^2,\\ \sigma_{\Delta z^2} = 2\sqrt{2}\sigma_z^2 &\simeq 0.011R_0^2,\end{aligned} \tag{B3}$$

where numerical values are obtained under the assumption of interatomic separation $R_0 = 2.5\mu$m, temperature $2\,\mu$K, radial trap frequency $\omega_x/2\pi = 50\,$kHz, and axial frequency $\omega_z/2\pi = 10\,$kHz, roughly reproducing state-of-the-art values [59, 66]. These values justify the expansion

$$V \approx V_0\left[1 - \frac{6\Delta x}{R_0} - \frac{3\,\Delta z^2}{R_0^2}\right], \tag{B4}$$

where second-order fluctuations in x and y were neglected due to their much smaller deviation. The isolated contributions from x and z displacements to fluctuations in interaction strength are $\sigma_{V,x} = 6\sqrt{2}V_0\,\sigma_x/R_0$ and $\sigma_{V,z} = 6\sqrt{2}V_0\,\sigma_z^2/R_0^2$, respectively. As they add in quadrature, their combination yields

$$\sigma_V = \sigma_{V,x}\sqrt{1+\frac{\sigma_{V,z}^2}{\sigma_{V,x}^2}} \simeq 1.04\,\sigma_{V,x}. \tag{B5}$$

That is, including axial fluctuations leads to only a $\sim 4\%$ increase in the deviation of the interaction strength. Given the computational cost of its inclusion in simulations, they were neglected.

2. Noise Approximation for Gate Optimization

In the optimization process, a semiclassical approximation of the noise was considered, reducing the computational cost of the procedure. Specifically, the Hamiltonian in Eq. (18) is assumed, with shot-to-shot gaussian fluctuations on its parameters, determined by the semiclassical variance of position and velocity in Eq. (B1):

(i) Fluctuations $\delta\Delta_q$ in detuning—different for each qubit—with standard deviation $\sigma_\Delta = 2\pi\sigma_v/\lambda$. We considered $\sigma_\Delta = 0.005\,\Omega$.

(ii) Fluctuations δV in the interaction strength, with standard deviation $\sigma_V = 6\sqrt{2}V\,\sigma_x/R$. We considered $\sigma_V = 7\,\Omega$.

(iii) Fluctuations $\delta\Omega$ in amplitude, which also give rise to additional shifts in detuning, due to self-light shifts. We considered $\sigma_\Omega = 5\times10^{-3}\Omega$, and $\delta\Delta = 0.2\,\delta\Omega$.

The driving amplitude was assumed to be constant throughout the protocol, ignoring finite ramping times.

Despite the possibility of detecting and post-selecting upon most decay errors (see Sec. IV B), inclusion of Rydberg decay in the optimization process was found to be beneficial. Details of inclusion in the computation of the cost function are provided in the next subsection.

3. Cost Function Computation

We wish to implement the target gate U_0 on N_q qubits. In practice, noisy versions U_n of the gate are implemented, yielding the channel $\mathcal{U}(\rho) = \overline{U_n\rho U_n^\dagger}$, where the overline denotes the average over noisy realizations. To make the role of noise explicit, note that $\mathcal{U}(\rho)$ is equivalent to an "afterwards" noise channel $\mathcal{E}(\rho)$ following the application of the target gate, such that

$$\begin{aligned}\mathcal{U}(\rho) &= \mathcal{E}(U_0\rho U_0^\dagger),\\ \mathcal{E}(\rho) &= \mathcal{U}(U_0^\dagger\rho U_0).\end{aligned} \tag{B6}$$

Equivalently, this expands in terms of its chi-matrix as

$$\mathcal{E}(\rho) = \sum_{Q,Q'}\chi_{Q,Q'}Q\rho Q', \tag{B7}$$

Its diagonal entries are computed as

$$\chi_{Q,Q} = \frac{1}{2^{3N_q}}\sum_{P\in\mathcal{P}^{\otimes N_q}}(-1)^{c_{P,Q}}\,\mathrm{tr}\left\{P\,\mathcal{E}(P)\right\}, \tag{B8}$$

where $\mathcal{P}^{\otimes N_q}$ denotes the set of N_q-qubits Pauli operators, and $c_{P,Q} = 0\,(1)$ if P and Q (anti)commute. This expression can then be used to compute the relevant chi-matrix element in numerical simulations.

Combining the previous expressions, again denoting with the overline average over noisy realizations,

$$\begin{aligned}\chi_{Q,Q} &= \frac{1}{2^{3N_q}}\sum_{P\in\mathcal{P}^{\otimes N_q}}(-1)^{c_{Q,P}}\,\mathrm{Tr}\left\{P\,\overline{U_nU_0^\dagger PU_0U_n^\dagger}\right\}\\ &\simeq \frac{1}{2^{3N_q}}\sum_{P\in\mathcal{P}^{\otimes N_q}}(-1)^{c_{Q,P}}\sum_{i\in\mathcal{C}}\langle i|\overline{PU_nU_0^\dagger PU_0U_n^\dagger}|i\rangle,\end{aligned} \tag{B9}$$

where $\mathcal{C}$ denotes a set of orthonormal basis states in the computational subspace of N_q qubits. Then, with $P = \sum_{s,t\in\mathcal{C}} P_{st}|s\rangle\langle t|$, where $P_{st} = \langle s|P|t\rangle$, it reads

$$\chi_{Q,Q} = \frac{1}{2^{3N_q}} \sum_{P\in\mathcal{P}^{\otimes N_q}} (-1)^{c_{Q,P}} \sum_{s,t,p,q\in\mathcal{C}} P_{st}P_{pq}\overline{\langle t|U_n U_0^\dagger\,|p\rangle\,\langle q|\,U_0 U_n^\dagger|s\rangle}. \tag{B10}$$

Conveniently rearranging, this yields

$$\chi_{Q,Q} = \frac{1}{2^{3N_q}} \sum_{s,t,p,q\in\mathcal{C}} \sigma_{st,pq}(Q)\ \overline{\langle t|U_n U_0^\dagger\,|p\rangle\,\langle q|\,U_0 U_n^\dagger|s\rangle}, \tag{B11}$$

where

$$\sigma_{st,pq}(Q) = \sum_{P\in\mathcal{P}^{\otimes N_q}} (-1)^{c_{Q,P}} P_{st}P_{pq}, \tag{B12}$$

are gate-independent coefficients.

For simplicity, the average over noises was taken to be the sum of separate contributions from fluctuations in Hamiltonian parameters and Rydberg decay, each with different form of the noisy evolution operator U_n and weighting of the average. The noisy evolution depends on the optimized control parameters and total duration. i.e. $U_n \equiv U_n(\vec{x}, T)$: x_0 denotes a residual single-qubit phase after the application of the Rydberg Hamiltonian, whereas each $x_{k>0}$ parameter represents the laser phase at the k^{th} time-step. Computation of the time evolution operator occurs in the limit of a large enough number of time-steps, thus is well approximated by a simple first-order decomposition.

Shot-to-shot fluctuations were incorporated in the cost function by averaging over multiple realizations

$$U_n^{(H)}(\vec{x}, T) = R_Z^{\otimes N_q}(x_0) \prod_{k=N}^{1} e^{-i\frac{H_n(x_k)}{\hbar}\frac{T}{N}}, \tag{B13}$$

by means of Gauss-Hermite quadrature. That is, rather than using probabilistic sampling, the noisy realizations were selected at fixed values, and averaged with fixed weights over their mesh. We averaged over the 2,000 most probable noisy realizations in the mesh, with the mesh given by seven points per noisy variable, i.e. 7^4 realizations for two qubits.

Rydberg decay is incorporated by replacing the time-evolution operator at time step k' with a jump operator $L_q^{(i)} = \sqrt{\gamma_i}\,|i\rangle\,\langle r|$, representing Rydberg decay of the q^{th} qubit to the computational state $|i\rangle$. Thus, the average was performed, with equal weight dt, on all possible operators

$$U_{q,i,k'}^{(\gamma)}(\vec{x}, T) = R_Z^{\otimes N_q}(x_0) \left(\prod_{k=N}^{k'+1} e^{-i\frac{H_0(x_k)}{\hbar}\frac{T}{N}}\right) L_q^{(i)} \left(\prod_{k=k'-1}^{1} e^{-i\frac{H_0(x_k)}{\hbar}\frac{T}{N}}\right), \tag{B14}$$

where H_0 is the noiseless Hamiltonian, similar to the approach used in [85].

Finally, we point out that the gradient can be analytically computed using the product rule, both for the gate parameters $\vec{x}$ and gate duration T.

4. Numerical Instabilities

In this subsection, we comment on some undesired behavior in the numerical optimization of the pulse shape $\phi(t)$, and on the adopted solutions. Concretely, we added penalty-terms to the cost-function in Eq. (23), to (i) ensure a smooth pulse profile, hence its experimental feasibility, and (ii) to avoid erroneous local solutions stemming from the combined optimization of the pulse shape and its duration T.

The pulse was smoothened to match experimental constraints, by the addition of the regularizer

$$\text{cost}_{\text{sm}} = \eta_{\text{sm}} N \sum_{i=1}^{N-1} (\phi_{i+1} - \phi_i)^2, \tag{B15}$$

already used in the optimization of Rydberg interactions [84, 121]. Our results were found by setting $\eta_{\text{sm}} = 3 \times 10^{-9}$.

The optimization of the gate duration T if unconstrained, was observed to occasionally lead to undesired behavior such as null gate time, or, oppositely, overly long gate durations. These were prevented by the regularizer

$$\text{cost}_T = \frac{1 - \tanh\left[50\Omega(T - T_{\text{tmop}})\right]}{2} + \eta_T \Omega(T - T_{\text{tmop}}), \tag{B16}$$

conditioned on the duration T_{tmop} of the corresponding time-optimal pulse, where the first term imposes a duration longer than that of the time-optimal solution, as necessarily the case, and the second term provides a small penalty against longer pulses. We set $\eta_T = 10^{-9}$, to ensure that the latter penalty only affects the optimization process once a robust solution is found, driving the procedure to shorter equally-robust solutions.

Appendix C: Circuit-level Noise

As mentioned in Sec. V A, the logical protocol is simulated against circuit-level noise of varying strength, modeled by the error rate $p \equiv p_2$ of two-qubit gates. This

section provides the exact noise acting on CZ gates in our simulations, accounting for both Rydberg decay errors and other noise sources. In particular, we assume that a representative fraction $f_d = 0.5$ of the error rate comes from Rydberg decay, allowing the impact of different noise sources to be captured by the model. Then, we assume that

1. a fraction $R_e = 0.98$ of the errors due to decay results in biased erasures, in line with the estimates from [62, 79];
2. the remaining noise due to decay leads to depolarizing noise, but with no contribution from XX, XY, YX, and YY errors, which would require unlikely double-decay events, as excitations to the Rydberg state are only possible from $|1\rangle$;
3. the remaining fraction $1 - f_d = 0.5$ of the error rate comes from sources other than decay; this contributes to Z-biased noise, owing to the fact that, in Rydberg gates, $|0\rangle$ and $|1\rangle$ are uncoupled in the absence of decay.

Therefore, for the majority of CZ gates, we assume the noise channel to be Pauli noise where the probabilities for each noise term

$$\begin{pmatrix} & p_{IX} & p_{IY} & p_{IZ} \\ p_{XI} & p_{XX} & p_{XY} & p_{XZ} \\ p_{YI} & p_{YX} & p_{YY} & p_{YZ} \\ p_{ZI} & p_{ZX} & p_{ZY} & p_{ZZ} \end{pmatrix}$$

are the following fraction of p_2

$$\left(\frac{1-f_d}{3} + f_d\frac{R_e}{4}\right)\begin{pmatrix} 0&0&0&1\\0&0&0&0\\0&0&0&0\\1&0&0&1\end{pmatrix} + f_d\frac{1-R_e}{11}\begin{pmatrix} 1&1&1&1\\1&0&0&1\\1&0&0&1\\1&1&1&1\end{pmatrix} \tag{C1}$$

and $p_{II} = 1 - \sum_{P\neq II} p_P$. Note that the weight $f_d R_e/4$ follows from the fact that one of the Pauli operators originating from biased erasures is the identity.

Since post-selection is used in the encoding of the resource state, erasure in the CZ gates acting on the first two rows of the logical patch was assumed to be post-selected upon, instead of replaced by biased errors. These locations were chosen due to their proximity to the support of the transversal rotations, such that phase errors on these CZ gates can more easily increase the failure rate of the TMR protocol. Therefore, their error model does not include the $f_d R_e/4$ term in Eq. (C1). Moreover, the rate of acceptance of the resource state was reduced by the probability of observing no erasure,

$$1 - p_{\text{erasure}} = (1 - p_2 f_d R_e)^{n_{CZ}^{(2)}} \tag{C2}$$

where $n_{CZ}^{(2)}$ is the distance-dependent number of relevant CZ gates.

For completeness, we also recall that single qubit gates have depolarizing rate $p_2/5$, while measurements and resets, only available in the computational basis, have bit-flip rate $p_m = p_2/2$ and $p_i = p_2/5$, respectively.

Appendix D: Numerical Methods for Simulation of Logical Gadgets

In this appendix, we outline the numerical modeling we implemented for the simulation of logical gadgets used in the manuscript : the TMR protocol and other Clifford operations. This appendix is organised as follows

(i) In App. D 1, we derive an explicit modeling of the noisy TMR channel, accounting for both the noise in the non-Clifford transversal rotations and standard circuit-level noise.

(ii) In App. D 2, we outline how Clifford simulators, such as `stim` [102], can be used to accurately estimate the noisy TMR channel. We note that the exactness of the method will be obtained at the price of a loss in efficiency.

(iii) In App. D 3 we show how to compute the performance of the TMR protocol within a deeper circuit, to better reproduce its performance in a realistic scenario. For the purpose, we also compute the logical flip rate of quantum memory experiments and state injection.

(iv) In App. D 4 we benchmark the performance of transversal Clifford operations under the considered noise.

(v) In App. D 5 we comment on the sampling statistics and uncertainty estimation.

1. Full working modeling of the TMR protocol

This subsection derives a closed form describing the full TMR channel under noise. This includes accounting for noise in (i) the initial error detection QED_a, (ii) the transversal noisy rotations $\mathcal{R}^{\otimes r}$, (iii) the following error detection QED_b, and finally (iv) gate teleportation $\mathcal{T}$. We stress that our method allows for the noisy non-Clifford $\mathcal{R}^{\otimes r}(\rho) = \sum_{\alpha\beta}\mathcal{R}_{\alpha\beta}P_\alpha\rho P_\beta$ rotations to be accounted for without approximations, and to address all noise sources outlined in Sec. III B and III C, and thoroughly described in App. A. The modeling exploits the Clifford nature of the syndrome extraction circuit, and assumes Pauli noise on all gates other than the transversal rotations, as customary in QEC and legitimated by the practical implementation of Pauli Twirling for Clifford operations.

The resource state is, before any encoding, $|0\rangle\langle 0|^{\otimes n}$, over n physical qubits. The logical resource state is initialized by the suitable syndrome extraction circuit QED_a. As this is a Clifford circuit, any Pauli error can be propagated to its end, up to a global phase. Therefore, the resulting state after QED_a can be written as

$$m_L^{(1)} = \text{QED}_1(|0\rangle\langle 0|^{\otimes n}) = \sum_{\mathbf{a}} p_{\mathbf{a}} P_{\mathbf{a}} |+\rangle\langle +|_L P_{\mathbf{a}}, \tag{D1}$$

where $|+\rangle\langle+|_L$ is the result of the ideal circuit, $P_{\mathbf{a}}$ denotes a physical Pauli operator, and any global phase has been canceled.

After the transversal rotations of the TMR, here encoded by the chi-matrix $\mathcal{R}_{\alpha\beta}$, the state is in

$$\begin{aligned} m_L^{(TMR)} &= \mathcal{R}^{\otimes r}(m_L^{(1)}) \\ &= \sum_{\alpha,\beta} \mathcal{R}_{\alpha\beta} P_\alpha m_L^{(1)} P_\beta \\ &= \sum_{\mathbf{a},\alpha,\beta} \mathcal{R}_{\alpha\beta}\, p_{\mathbf{a}}\, P_\alpha P_{\mathbf{a}} |+\rangle\langle+|_L P_{\mathbf{a}} P_\beta. \end{aligned} \tag{D2}$$

Note that $\mathcal{R}_{\alpha\beta}$ represents the overall action of the noisy gates, not an error channel following their ideal implementation.

The resource state then goes through more rounds of error syndrome extraction, denoted $\mathrm{QED_b}$. Its final state, which we denote $\mathfrak{M}_L$ as in App. A 5, reads $\mathfrak{M}_L = \mathrm{QED}_2(m_L^{(TMR)})$. Breaking down the action of the second error detection cycle, it can be written as

$$\mathfrak{M}_L = \mathrm{Tr}_{\mathbf{z}} \left(\mathcal{E} \left(\sum_{\mathbf{a},\alpha,\beta} \mathcal{R}_{\alpha\beta}\, p_{\mathbf{a}}\, P_\alpha P_{\mathbf{a}} |+\rangle\langle+|_L P_{\mathbf{a}} P_\beta \otimes |\mathbf{z}_{\mathbf{a},\alpha}\rangle\langle\mathbf{z}_{\mathbf{a},\beta}| \right) \right), \tag{D3}$$

where $\mathcal{E}$ represents noise occurring during the error detection process, $\mathrm{Tr}_{\mathbf{z}}$ represents the ancillae measurement, and $\mathbf{z}_{\mathbf{a},l}$ is the ancillae bit-string measured after the stabilizers' extraction. Note that Pauli noise after the transversal rotation can be considered to be applied at the end of a noiseless extraction owing to the Clifford nature of the circuits, as previously explained for $\mathrm{QED_a}$. Upon the ancillae measurement, only terms with identical left and right syndromes $\mathbf{z}_{\mathbf{a},l} = \mathbf{z}_{\mathbf{a},r}$ are preserved, in turn requiring $P_\alpha \equiv_L P_\beta$, i.e. $P_\alpha = P_\beta$ up to logical operators and stabilizers. Concretely, as the TMR is applied on the support of Z_L, the relation $P_\alpha \equiv_L P_\beta$ imposes

$$P_\beta = i^{v_p(\alpha,\beta)} Z_L^{v_0(\alpha,\beta)} \prod_k S_k^{v_k(\alpha,\beta)} P_\alpha \tag{D4}$$

where S_k denotes the stabilizers with full support on the qubits acted upon by the transversal rotations, $v_p \in \{0,1,2,3\}$, and $v_0, v_k \in \{0,1\}$, . Therefore, after measurement and post-selection, we can write

$$\begin{aligned} \mathfrak{M}_L &= \mathcal{E}' \left(\sum_{(\alpha,\beta)_L} \mathcal{R}_{\alpha\beta} P_\alpha \rho_L^{(1)} P_\beta \right) \\ &= \sum_{(\alpha,\beta)_L} \mathcal{R}_{\alpha\beta} \sum_{\mathbf{a},\mathbf{b}} p_{\mathbf{a},\mathbf{b}}\, P_{\mathbf{b}} P_\alpha P_{\mathbf{a}} |+\rangle\langle+|_L P_{\mathbf{a}} P_\beta P_{\mathbf{b}}\,, \end{aligned} \tag{D5}$$

where $(\alpha,\beta)_L$ denotes all pairs of Pauli operators satisfying Eq. (D4), and $\mathbf{b}$ identifies the average over noisy trajectories in QED_b.

The same errors have the same logical correction. Therefore, we wish to replace the right-applied operator P_β with P_α. Using Eq. (D4) and rearranging yields

$$\begin{aligned} \mathfrak{M}_L = \sum_{(\alpha,\beta)_L} & \mathcal{R}_{\alpha\beta} \\ & \sum_{\mathbf{a},\mathbf{b}} p_{\mathbf{a},\mathbf{b}}\, s_{\mathbf{a},\alpha,\beta}\, P_{\mathbf{b}} P_\alpha P_{\mathbf{a}} |+\rangle\langle+|_L Z_L^{v_0(\alpha,\beta)} P_{\mathbf{a}} P_\alpha P_{\mathbf{b}}, \end{aligned} \tag{D6}$$

where $s_{\mathbf{a},\alpha,\beta} = \pm i^{v_p(\alpha,\beta)}$, with the sign determined by the commutation relation between $P_{\mathbf{a}}$ and $Z_L^{v_0(\alpha,\beta)} \prod_k S_k^{v_k(\alpha,\beta)}$.

The resource state $\mathfrak{M}_L$ is then injected via gate-teleportation onto a target logical qubit, initially in Ψ. The outcome of the teleportation measurement is denoted μ. Under a Pauli noise model, the action of the protocol is again modeled as the ideal scheme followed by residual noise $Q_{\mathbf{c}}$, while previously accumulated Pauli errors $P_{\mathbf{b}} P_\alpha P_{\mathbf{a}}$ propagate through the protocol to residual errors $Q_{\mathbf{b}} Q_l Q_{\mathbf{a}}$. Denoting $\mathbf{t}$ as the combination of each $\mathbf{a}, \mathbf{b}, \mathbf{c}$ trajectories, and $Q_{\mathbf{t},l} = Q_{\mathbf{c}} Q_{\mathbf{b}} Q_l Q_{\mathbf{a}}$, the result reads

$$\begin{aligned} \rho_L(\Psi) = \sum_{(\alpha,\beta)_L} & \mathcal{R}_{\alpha\beta} (-1)^{\mu \cdot v_0(\alpha,\beta)} \\ & \sum_{\mathbf{t}} p_{\mathbf{t}}\, s_{\mathbf{t},\alpha,\beta}\, Q_{\mathbf{t},l}\, \Psi Z_L^{v_0(\alpha,\beta)} Q_{\mathbf{t},l}^\dagger. \end{aligned} \tag{D7}$$

where the phase $(-1)^{\mu \cdot v_0(\alpha,\beta)}$, arises from the one-sided action of $Z_L^{v_0(\alpha,\beta)}$ on the teleportation measurement outcome μ. It is determined by observing that

1. if $v_0(\alpha,\beta) = 0$, then $Z_L^{v_0(\alpha,\beta)} \equiv I$, and no additional phase is ever acquired;

2. if $v_0(\alpha,\beta) = 1$, then the adopted resource state is effectively $|+\rangle\langle-|_L$, such that the phase between measurements in $\mu = 0$ and 1 differs by $(-1)^\mu$.

The result μ of the measurement can be determined and used in post-processing. With r physical rotations, the target outcome is $\mu = 0$ when $\lfloor r/2 \rfloor$ is even and $\mu = 1$ when it is odd.

Upon decoding, the state is mapped back onto the logical computational state, yielding the final state

$$\rho_L^{(\mu)}(\Psi) = \sum_{(\alpha,\beta)_L} \mathcal{R}_{\alpha\beta}(-1)^{\mu\cdot v_0(\alpha,\beta)} \sum_{\mathbf{t}} p_{\mathbf{t}}\, s_{\mathbf{t},\alpha,\beta}\, Q_{L,(\mathbf{t},l)} \Psi Z_L^{v_0(\alpha,\beta)} Q_{L,(\mathbf{t},l)}. \quad \text{(D8)}$$

While in most QEC experiments it suffices to determine whether the logical state of interest was flipped, here, due to the one-sided action of the $Z_L^{v_0(\alpha,\beta)}$ operator, the exact resulting logical operator Q_L must be determined. Fortunately, this can be done by simply applying each trajectory to both initial states $|+\rangle_L, |0\rangle_L$. Any other global phase in $Q_{L,(\mathbf{t},i)}$ cancels out.

Finally, it will be useful for its numerical evaluation, to express the final state in terms of probabilities for different outcomes, rather than a sum over trajectories. First, we introduce $p_\eta(\alpha,\beta)$ as the fraction of accepted trajectories with commutation parity η over all attempted trajectories. More precisely, $p_\eta(\alpha,\beta)$ for $\eta = 0\,(1)$ is the unnormalised probability that a trajectory passed post-selection given the error P_a before the transversal rotations, such that P_a (anti)commutes with the difference between P_β and P_α, i.e. $Z_L^{v_0(\alpha,\beta)} \prod_k S_k^{v_k(\alpha,\beta)}$. As such, $p_\eta(\alpha,\beta)$ encodes the probability distribution over trajectories of the ± 1 sign of $s_{\mathbf{t},\alpha,\beta}$. These probabilities are unnormalised to track the loss of norm due to post-selection. Second, we define $p_{Q,\alpha,\eta}$ as the probability of having a logical flip Q_L conditioned on P_α and η. This yields

$$\rho_L^{(\mu)}(\Psi) = \sum_{(\alpha,\beta)_L} \mathcal{R}_{\alpha\beta} i^{v_p(\alpha,\beta)} (-1)^{\mu\cdot v_0(\alpha,\beta)} \sum_{\eta\in 0,1} (-1)^\eta\, p_\eta(\alpha,\beta) \sum_{Q} p_{Q,\alpha,\eta}\, Q_L \Psi Z_L^{v_0(\alpha,\beta)} Q_L. \quad \text{(D9)}$$

Finally, the range of the outermost sum can be further reduced by considering a subset of Pauli operators $\mathcal{G}$, such that all P_α can be generated from them, the logical Z_L operator and S_k stabilizers. Then, following similar steps as before,

$$\rho_L^{(\mu)}(\Psi) = \sum_{g\in\mathcal{G}} \sum_{\alpha,\beta \equiv_L g} \mathcal{R}_{\alpha\beta} i^{v_p(\alpha,\beta)} (-1)^{\mu\cdot v_0(\alpha,\beta)} \sum_{\eta\in 0,1} (-1)^\eta\, p_\eta(\alpha,\beta) \sum_{Q} p_{Q,g,\eta}\, Q_L Z_L^{v_0(\alpha,g)} \Psi Z_L^{v_0(\beta,g)} Q_L. \quad \text{(D10)}$$

By computing this expressions for a suitable selection of input states, one obtains not only the logical infidelity, but an accurate description of the whole logical channel. By repeated application of the TMR channel, with, at each step, suitable measurement outcome μ and angle-dependent matrix $\mathcal{R}_{\alpha\beta}(\theta_L)$, the channel for the whole RUS procedure can be estimated.

a. *Clifford Correction for even r*

As described in Sec. III A and depicted in Fig. 1(a), if the TMR involves an even number of transversal physical rotations, then the gate $S_{X_L}^\dagger = \exp(i\pi X_L/4)$ is to be applied on the resource state before teleportation measurement. In this brief subsection, we highlight how to incorporate its action on the final state, as well as to propagate its noise.

To incorporate its action on the resource state $\mathfrak{M}_L$ described in Eq. (D6), note that it imparts a phase $e^{i\pi/4}$ on the left side, while on the right side its action depends on $v_0(\alpha,\beta)$, as X_L and Z_L anti-commute: it imparts a (α,β)-dependent phase $e^{-i\pi/4}(e^{+i\pi/4})$ for $v_0(\alpha,\beta) = 0\,(1)$. This phase then propagates onto the final logical state.

To incorporate its infidelity, we first note that it is Clifford and has a transversal implementation up to a qubit permutation [52], hence its infidelity falls exponentially with the code distance. Nevertheless, errors can occur and propagate : specifically, X_L and Y_L flips modify the teleportation-measurement outcome. Denoting their probability $p_{\mathrm{flip}_L}^{(S_X)}$, the final result $\rho_L^{(\mu)}(|\sigma_L\rangle)$ of a gate teleportation attempt with measurement μ on the target logical qubit Ψ, effectively is

$$\rho_L^{(\mu)}\left(\Psi; p_{\mathrm{flip}_L}^{(S_X)}\right) = \left(1 - p_{\mathrm{flip}_L}^{(S_X)}\right) \rho_L^{(\mu)}(\Psi;0) + p_{\mathrm{flip}_L}^{(S_X)}\, \rho_L^{(\mu\oplus 1)}(\Psi;0). \quad \text{(D11)}$$

We numerically computed under the assumed noise model $p_{\mathrm{flip}_L}^{(S_X)} \simeq 0.026(p/0.009)^{\lceil d/2\rceil}/(d-1)$, as part of the Clifford simulations described in App. D 4.

2. Numerical estimation of TMR performance

The previous subsections derived the quantities required for numerical reconstruction. In this subsection, we explicitly explain how the relevant error parameters are accessed by means of Clifford simulations performed in `stim` [102]. As evident from Eq. (D10), an accurate estimate of the logical errors in the TMR protocol requires more information than usually extracted from Clifford simulations. This can be obtained by running the same noise realisation for multiple initial states, in turn made possible by manually inserting errors in the circuit according to their probability distribution—at the price of slowing down significantly the simulations' speed.

The outermost sum is dealt with by simulating multiple noisy trajectories for each generating Pauli $g \in \mathcal{G}$, in place

of the transversal rotation. To compute $p_\eta(\alpha,\beta)$, the initialization of the resource state is simulated with the same errors for both $|0\rangle_L$ and $|+\rangle_L$—differing by a layer of H gates—allowing the measurement of Z_L and X-type stabilizers respectively. Their flip indicates anti-commutation of $P_{\mathbf{a}}$ and said operators, allowing to compute $s_{\mathbf{t},\alpha,\beta}$ for each trajectory, hence $p_\eta(\alpha,\beta)$. Then, the same trajectory is extended with the following operations, and possible additional errors, to compute the probability distribution $p_{Q,g,\eta}$ of logical flips, conditioned on P_g and η. To exactly individuate what logical error occurred in each trajectory, this was simulated for the target logical qubit in both $|0\rangle_L$ and $|+\rangle_L$.

Finally, we would like to point out how each noise source discussed in App. A is reflected in Eq. (D10): (i) logical flips from syndrome extraction are represented by $p_{Q,g,\eta}$; (ii) noise from repeating flips is encoded in $p_\eta(\alpha,\beta)$, and (iii) $\mathcal{R}_{\alpha\beta}$ encodes noise in the transversal rotations themselves, with dominant error contributions being variation in the correlated phase-flips terms with respect to the noiseless scenario.

3. TMR logical performance in deep circuit

Evaluating the protocol's performance, specifically the $p_{Q,g,\eta}$ rates, in practical scenario requires embedding it within multiple QEC rounds. First, the logical TMR gadget (comprising gate teleportation and a single syndrome-extraction round) is simulated in a deeper circuit, as preceded and followed by $\lfloor d/2 \rfloor$ noisy QEC rounds. To isolate the noise contribution from the TMR gadget alone, simulations of the QEC memory are also performed.

To accurately parameterize Eq. (D10), we must extract the isolated TMR error probabilities, $p_{Q,g,\eta}$, from the total logical probabilities obtained in the deep circuit simulations, which we denote as $p^{(\mathrm{deep})}_{Q,g,\eta}$. As the total deep circuit noise is the composition of the TMR gadget channel and the QEC memory channel, their contributions can be separated working with the Pauli Transfer Matrix (PTM) formalism. For any stochastic Pauli channel with error distribution p_Q, the PTM eigenvalues λ_P are

$$\lambda_P = \sum_{Q\in\mathcal{P}} (-1)^{c_{P,Q}}\, p_Q, \tag{D12}$$

where $\mathcal{P} = \{I,X,Y,Z\}$ and $c_{P,Q} = 0\,(1)$ if P,Q (anti)commute. Their composition in the deep circuit corresponds to the multiplication of their respective eigenvalues $\lambda^{(\mathrm{deep})}_{P,l,\eta} = \lambda^{(\mathrm{TMR})}_{P,l,\eta}\lambda^{(\mathrm{QEC})}_{P}$. We can therefore deconvolve [122] the channels and isolate the eigenvalues of the TMR protocol simply as

$$\lambda^{(\mathrm{TMR})}_{P,l,\eta} = \frac{\sum_{R\in\mathcal{P}}(-1)^{c_{P,R}}\, p^{(\mathrm{deep})}_{R,l,\eta}}{\sum_{R\in\mathcal{P}}(-1)^{c_{P,R}}\, p^{(\mathrm{QEC})}_{R}}. \tag{D13}$$

Finally, the isolated TMR logical error probabilities $p_{Q,\alpha,\eta}$ to be inserted into Eq. (D10) are recovered by applying the inverse transform to the isolated eigenvalues

$$\begin{aligned} p_{Q,\alpha,\eta} &= \frac{1}{4}\sum_{P\in\mathcal{P}} (-1)^{c_{P,Q}}\, \lambda^{(\mathrm{TMR})}_{P,l,\eta} \\ &= \frac{1}{4}\sum_{P\in\mathcal{P}} (-1)^{c_{P,Q}}\, \frac{\sum_{R\in\mathcal{P}}(-1)^{c_{P,R}}\, p^{(\mathrm{deep})}_{R,l,\eta}}{\sum_{R\in\mathcal{P}}(-1)^{c_{P,R}}\, p^{(\mathrm{QEC})}_{R}}. \end{aligned} \tag{D14}$$

Due to statistical deviations, the procedure occasionally results in negative, albeit very small, error rates. We therefore project the reconstructed Pauli distribution onto the probability simplex, i.e. assumed the closest valid Pauli distribution in Euclidean distance. minimizing its euclidean distance from the unconstrained estimate using the sampled covariance matrix.

Quantum memory error rates were determined by simulations of logical flips on states $|0\rangle_L$ and $|+\rangle_L$, as customary in studies of quantum memories and Clifford gates. This it motivated by the assumption of a negligible Y_L error rate—as is usually the case for the surface code, and further motivated by our Q_L-resolving TMR simulations. Results are displayed in Fig. 6(a). The memory error rate over $d-1$ rounds fits the standard ansatz

$$p^{(\mathrm{QEC})}_L(p,d) = a_{\mathrm{QEC}}\left(\frac{p}{p_{\mathrm{th}}}\right)^{\lceil d/2\rceil}, \tag{D15}$$

with $a_{\mathrm{QEC}} \simeq 0.088$ and $p_{\mathrm{th}} \simeq 0.02$. In the regime considered, the error rate per extraction round at code distance d can be computed as simply $p^{(\mathrm{QEC})}_L(p,d)/(d-1)$.

Moreover, additional simulations were performed to probe logical error rates in the TMR protocol at small rotation angle $|\theta_L|$. In fact, the effect of the noise sources specific to the TMR protocol decreases with the angle θ_L of rotation—as outlined in Sec. III B and derived in App. A 3 and A 4. Then, the main error contribution is simply noise from syndrome extraction and transversal gates. To accurately probe their behavior at low error rates and high code distance, many more trajectories are needed than can be simulated with the resource-intensive procedure outlined in App. D 2. Therefore, additional simulations were performed to simply compute the probabilities of logical flips on states $|0\rangle_L$ and $|+\rangle_L$, for the deep circuit comprised of the full TMR gadget within rounds of syndrome extraction. Results are displayed in Fig. 6(b). Combining simulations of the deep circuit and the memory experiment, the contribution of the TMR gadget was isolated. This is shown in Fig. 6(c), and fitted by

$$p^{(\mathrm{inj})}_L(p,d) = \frac{a_{\mathrm{inj}}}{d-1}\left(\frac{p}{p_{\mathrm{th}}}\right)^{\lceil d/2\rceil}, \tag{D16}$$

with $a_{\mathrm{inj}} \simeq 0.070$ and $p_{\mathrm{th}} \simeq 0.012$. We note that, considering the chi-matrix decomposition of the transversal rotations in the TMR protocol, $\mathcal{R}^{\otimes r}(\rho) = \sum_{\alpha,\beta}\mathcal{R}_{\alpha\beta}P_\alpha\rho P_\beta$, those simulations were performed only for the term $P_\alpha = I^{\otimes 2r}$, which is overwhelmingly dominating in the

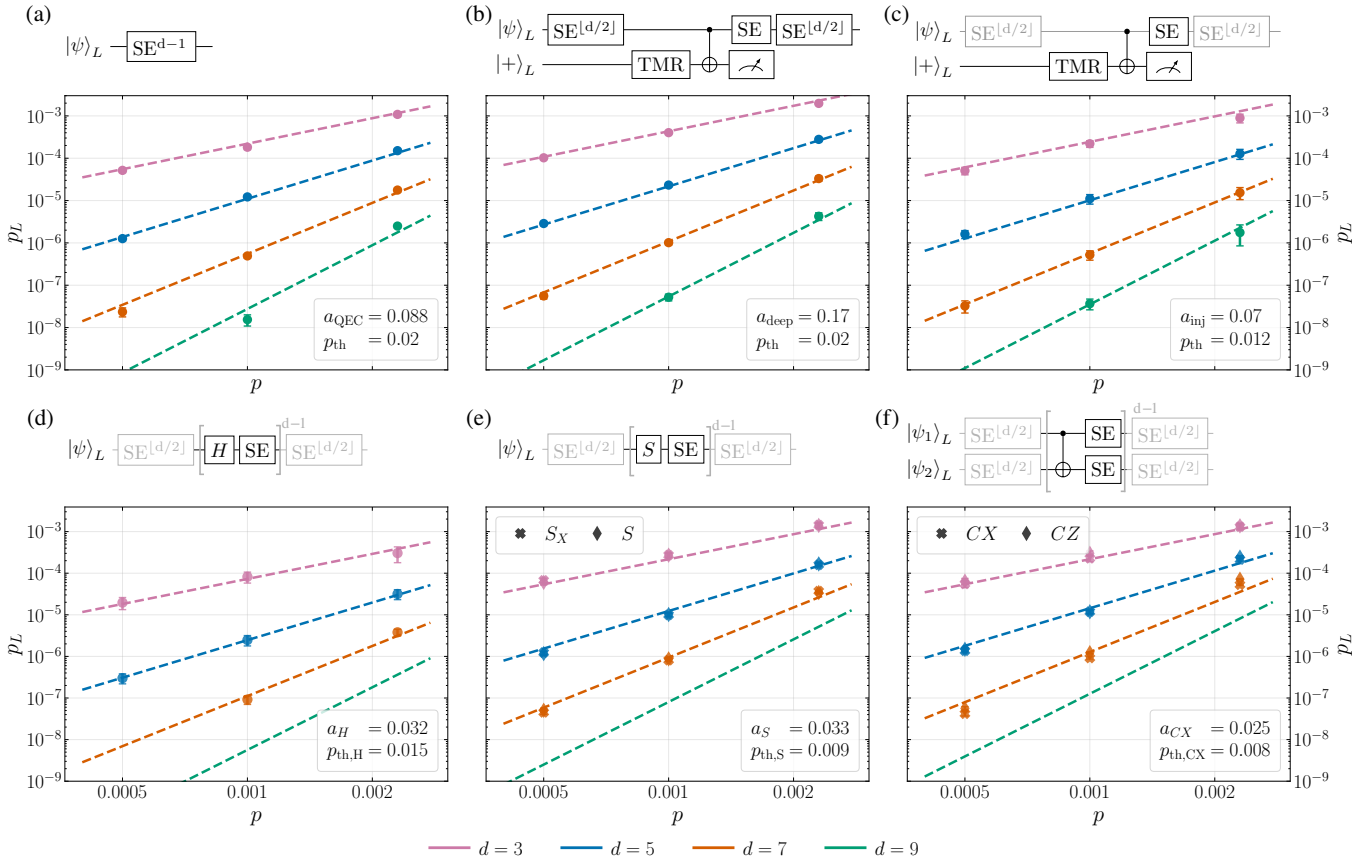


Figure 6. Logical error rate of Clifford operations against physical error rate p, for code distances $3, 5, 7$ and 9. In each panel, the diagram above the plot displays the simulated circuit, including both the components whose error rate is measured (black), and those whose contribution is deconvolved away (light grey). Fitted thresholds and noise prefactors are shown, with the logical error rate of isolated gadgets (c-f) further normalised by a factor $d-1$. Errorbars denote sampling standard error. (a) $d-1$ rounds of syndrome extraction. (b) TMR gadget in between $d-1$ rounds of syndrome extraction. (c) Isolated TMR gadget, i.e. TMR encoding, teleportation, and one round of extraction. (d) Isolated logical H gadget. (e) Isolated logical S or S_X gadget. (f) Isolated logical CX or CZ gadget.

context of small rotation angle and error rates. Moreover, owing to the use of extensive post-selection in the resource state encoding, the failure rate is dominated by the transversal CNOT and following syndrome extraction, thus providing a good estimate of the error rate of a general resource state injection.

4. Logical Clifford performance in deep circuit

We simulated the logical Clifford gates required by the application benchmarked in this manuscript, i.e. the simulation of the Hubbard model (see Sec. VI B). These are $\{H, S, S_X, CX, CZ\}$. Working in the framework of algorithmic fault tolerance [32–35], we identify a Clifford gadget as the corresponding gate followed by a single round of syndrome extraction. We note that the transversal S gate proposed in Ref. [52] and hereby adopted, includes a round of syndrome extraction in its minimal execution, thus bringing the total number of extraction rounds per gadget to two.

As explained for the TMR protocol in the previous subsection, we simulated logical gadgets preceded and followed by multiple rounds of syndrome extraction, and then deconvolved the relevant channel. Each Clifford gadget was repeated $d-1$ times to better capture its performance within a deep circuit. Simulation ran for initial states $|0\rangle_L$ and $|+\rangle_L$ (and product thereof for controlled gates), again under the assumption of a negligible Y_L error rate.

Results are shown in Figs. 6(d) to 6(f) for, respectively, the H gate, the S-like gates, and the controlled gates. The S-like gates had equivalent performance, and so did the controlled gates. Performance of the individual gadget was fit against the ansatz $p_L^{(\bullet)} = a_\bullet (p/p_{\mathrm{th},\bullet})^{\lceil d/2 \rceil}/(d-1)$. Corresponding thresholds and prefactors are provided in the figure and Table II.

5. Details of numerical simulations

For the estimation of the TMR channel in Eq. (D10), the number of shots was increased with the code distance to maintain good accuracy as logical errors become more rare. Specifically, for each generating operator P_g, simulations included 4, 16, 64 million trajectories at code distances 5,7,9 respectively, or half as many passing post-selection, whichever was reached first. We also observe that the (noisy) channel of the non-Clifford transversal rotations $\mathcal{R}_{\alpha\beta}$ enters the numerical computation Eq. (D10) by means of post-processing, rather than by sampling its noise. This reduces the required sampling overhead, and allows to use the same simulations for different rotation angles θ_L. Finally, we note that owing to the structure of $\mathcal{R}_{\alpha\beta}$ and the extensive use of post-selection, multiple P_g yield negligible contributions: their simulations thus had a reduced number of shots, with the statistical uncertainty reflected in the errorbars in Fig. 4.

For the estimation of logical flips in the memory, deep TMR circuits, and Clifford gates, the number of shots was increased with both the code distance and the inverse of the physical error rate. Memory simulations used $\min(2.5 \cdot 10^3 \cdot (0.02/p)^{\lceil d/2 \rceil}, 2 \times 10^9)$ shots per logical basis. Deep TMR simulations proceeded until that many shots were simulated or half as many accepted after post-selection. As these simulations have much larger statistics than the ones performed to estimate the full TMR channel, their results were used to provide a more accurate estimate of $p_{Q,I^{\otimes 2r},0}$, i.e. of the logical flip rates upon trivial commutation on the statistically-leading term. Clifford gates' simulations had half as many shots per logical basis as in the memory experiments, owing to their longer runtime, and to their generally higher error rate which demanded fewer statistics.

Sampling standard errors were estimated using a Dirichlet smoothing model, followed by linear uncertainty propagation. For each simulated Pauli term, acceptance is modeled as binomial with Jeffreys Beta$(1/2, 1/2)$ prior. Conditioned on acceptance, the joint distribution of commutation outcomes and logical flips was modeled as multinomial. Probabilities are then evaluated at the mean of a Dirichlet posterior, with prior concentration 0.8 uniformly distributed over the observed categories, following the recommendation from Ref. [123, 124]. Finally, the multinomial covariance evaluated at those estimates determines the uncertainties. For the TMR protocol, these uncertainties, affecting the $p_\eta(\alpha, \beta)$ and $p_{Q,\alpha,\eta}$ coefficients entering the logical channel, were propagated through normalization, RUS procedure, and calculation of the logical error rate, using linear propagation implemented by the Python `uncertainties` package [125]. The resulting error bars are reported in Fig. 4.

Appendix E: Details on the simulation of the Fermi-Hubbard model

The Hubbard model is one of the central paradigms in the study of strongly correlated quantum many-body systems, and has long served as a benchmark target for quantum simulations [27, 37, 105–113]. Its Hamiltonian $H = H_J + H_U$ is the sum of a hopping (kinetic) term H_J and an on-site potential term H_U, defined as

$$\begin{aligned} H_J/\hbar &= -J \sum_{\langle i,j\rangle,\sigma} \left(c^\dagger_{i\sigma} c_{j\sigma} + \text{h.c.} \right), \\ H_U/\hbar &= U \sum_{i=1} n_{i\uparrow} n_{i\downarrow}, \end{aligned} \tag{E1}$$

Here $\langle i, j\rangle$ denotes nearest-neighbour lattice sites and $\sigma \in \{\uparrow, \downarrow\}$ labels the spin orbital. The system has M lattice sites and therefore $2M$ fermionic modes.

For digital simulation, the fermionic Hamiltonian is mapped to Pauli operators using the Jordan–Wigner (JW) transformation, chosen because it requires no additional ancillary qubits, oppositely to other mappings. The on-site potential term is simple. For fixed-particle number simulations, after a convenient shift of the chemical potential, which can be accounted for in classical post-processing [107, 111], the on-site interaction becomes

$$H_U/\hbar = \frac{U}{4} \sum_i Z_{i\uparrow} Z_{i\downarrow}. \tag{E2}$$

By contrast, the hopping terms are non-local. A hopping term between modes i and j maps to an $XX+YY$ rotation dressed by the JW parity string

$$Z^{(i,j)}_{\text{JW}} = \prod_{i<q<j} Z_q, \tag{E3}$$

acting on all qubits lying between i and j in the chosen JW ordering. Explicitly, the kinetic term becomes

$$H_J/\hbar = -\frac{J}{2} \sum_{\langle i,j\rangle,\sigma} (X_{i\sigma} X_{j\sigma} + Y_{i\sigma} Y_{j\sigma}) Z^{(i,j)}_{\text{JW}}. \tag{E4}$$

In the following, we provide an explicit scheme for the Trotterised simulation of time evolution in the Fermi-Hubbard model, specific to a two-dimensional $L_x \times L_y$ lattice with periodic boundary conditions (BCs). The main difficulty is the implementation of the hopping Hamiltonian H_J, as the JW strings make geometrically local hopping terms generally non-local in qubit space. To reduce gate overhead, following Refs. [36, 37], we use the fact that each JW string only depends on the parity of the qubits lying between the two hopping endpoints. These parities can be computed in a nested fashion with a small number of ancillae. Specifically, the parity of the $(i, i+1)$ qubits is first stored on an ancilla; then re-used when calculating that of $(i-1, i+2)$; this is in turn stored and then re-used when calculating that of $(i-2, i+3)$,

	$R_{X/Z}$		S_X		CX/CZ		Algorithmic ancillae
	Count	Depth	Count	Depth	Count	Depth	
$H_{v,\text{odd}}$	$2L_xL_y$	1	$4L_xL_y$	2	$8L_xL_y - 6L_y$	$6L_x - 4$	L_y
H_U	L_xL_y	1	0	0	$2L_xL_y$	2	0
$H_h^{(\text{pBCs})}$	$4L_xL_y$	2	$8L_xL_y$	4	$6L_xL_y - 6L_y$	$L_x/2 + 3$	0
$H_{v,\text{even}}^{(\text{pBCs})}$	$2L_xL_y$	1	$4L_xL_y$	2	$8L_xL_y - 8$	$12L_x + L_y - 6$	L_y
$\text{Total}^{(\text{pBCs})}$	$(14N{+}2)L_xL_y$	$8N{+}1$	$(24N{+}4)L_xL_y$	$12N{+}2$	$(32N{+}8)L_xL_y$ $-(18N{+}6)L_y - 8N$	$(19L_x{+}L_y)N$ $+6L_x - 4$	L_y
$H_h^{(\text{oBCs})}$	$4(L_xL_y - L_y)$	2	$8(L_xL_y - L_y)$	4	$4(L_xL_y - L_y)$	4	0
$H_{v,\text{even}}^{(\text{oBCs})}$	$2(L_xL_y - 2L_x)$	1	$4(L_xL_y - 2L_x)$	2	$8L_xL_y - 16L_x - 6L_y + 12$	$6L_x - 4$	L_y
$\text{Total}^{(\text{oBCs})}$	$(14N{+}2)L_xL_y$ $-4N(L_x + 2L_y)$	$8N{+}1$	$(24N{+}4)L_xL_y$ $-8N(L_x + 2L_y)$	$12N{+}2$	$(28N{+}8)L_xL_y - 16NL_x$ $-(20N{+}6)L_y + 12N$	$(12L_x{+}4)N$ $+6L_x - 4$	L_y

Table IV. Gate counts and circuit depths for single sweeps of the Hubbard Hamiltonian terms on an $L_x \times L_y$ lattice, with $L_x, L_y \geq 4$, and their total. $H_{v,\text{odd}}$ and H_U are independent of boundary conditions, whereas the other estimates are separated between periodic (pBCs) and open (oBCs). Hopping counts are summed over both spin species, which execute in parallel. Gates are grouped into single-qubit rotations $R_{X/Z}$, S_X gates, and two-qubit Clifford gates CX/CZ. The last column gives the ancillae required for the optimized computation of JW strings, reused throughout the circuit. The total rows give the resources for a second-order Trotterised evolution of N steps with the ordering $H_{v,\text{odd}}/2 \to H_U/2 \to H_h/2 \to H_{v,\text{even}} \to H_h/2 \to H_U/2 \to H_{v,\text{odd}}/2$, with consecutive $H_{v,\text{odd}}/2$ terms merged in between steps.

and so on. Compared with evaluating every JW string independently, this nested computation substantially reduces the gate overhead. Moreover, hopping between two endpoints preserves the parity of the qubits in the intermediate string, so the hopping rotations can be commuted through parity-computation layers and executed in parallel.

Using this trick, we extend the scheme proposed in Ref. [37] to periodic BCs, assuming that both L_x and L_y are even. We decompose the Hamiltonian into five non-commuting parts: the on-site interaction H_U and four hopping Hamiltonians,

$$H_J = H_{h,\text{even}} + H_{h,\text{odd}} + H_{v,\text{even}} + H_{v,\text{odd}}. \tag{E5}$$

These four hopping terms are depicted in Fig. 7(a) as, respectively, green, yellow, blue and red bonds. The even terms include the hopping terms that wrap around the periodic boundaries.

We now assume a second-order Trotter decomposition. Half the second-order Trotter step—essentially the originating first-order step—is depicted in Fig. 7(b). Explicitly, the full second-order step is

$$V_2(t) = e^{-iH_{v,\text{odd}}\frac{t}{2}} e^{-iH_U\frac{t}{2}} e^{-iH_{h,\text{odd}}\frac{t}{2}} e^{-iH_{h,\text{even}}\frac{t}{2}} \; e^{-iH_{v,\text{even}}t} \; e^{-iH_{h,\text{even}}\frac{t}{2}} e^{-iH_{h,\text{odd}}\frac{t}{2}} \; e^{-iH_U\frac{t}{2}} e^{-iH_{v,\text{odd}}\frac{t}{2}}. \tag{E6}$$

Over multiple Trotter steps, consecutive half-steps of the $H_{v,\text{odd}}$ term are merged. This ordering was chosen as it allows for the simulation of the two most costly terms, $H_{v,\text{even}}$ and $H_{v,\text{odd}}$ to occur half as frequently as the others.

The potential term is readily implemented, independently of the ordering, as

$$e^{-iH_U t} = \bigotimes_{j=1}^{M} e^{-i\frac{Ut}{4} Z_{j\uparrow} Z_{j\downarrow}}, \tag{E7}$$

with the implementation of each j^{th} term shown in Fig. 7(c).

The simulation of the hopping terms requires two other building blocks: fermionic SWAPs and tunneling $XX + YY$ rotations, respectively depicted in Figs. 7(d) and 7(e). The starting JW ordering follows the pattern shown in Fig. 7(f), obtained from the typical snake ordering by interchanging the two rightmost sites in each row. For each of the two spin orbitals, the simulation of the hopping Hamiltonian proceeds as follows, in one-to-one correspondence with Fig. 7(f):

(f.1) The vertical term $H_{v,\text{odd}}$ requires the computation of JW strings' parities, as per the trick used in Refs. [36, 37]. These are computed onto $L_y/2$ ancillae—one per pairs of rows—in a nested fashion (CNOTs in the figure). Interleaved with the computation of the parities, each $Z_{\text{JW}}^{(i,j)}$ operator is applied on one of the hopping qubits (CZ in the figure).

(f.2) As the JW strings have been applied, the corresponding tunneling rotations are simulated.

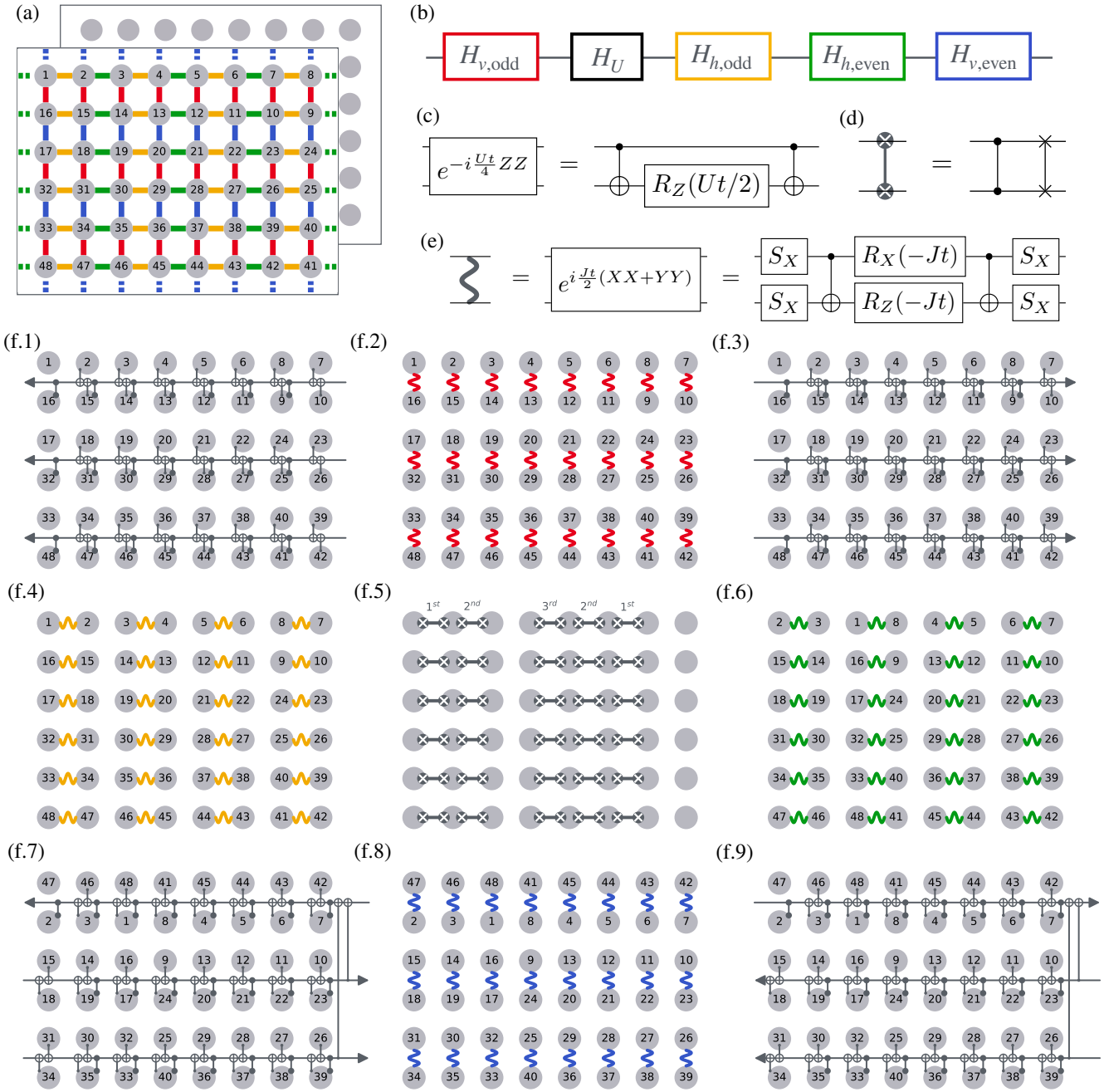


Figure 7. Overview of the digital simulation of the 2D Fermi-Hubbard model with periodic BCs. (a) Schematic depiction of the model on a 8×6 lattice: the two planes identify the two spin orbitals; the sites' numbering follows the traditional "snake ordering"; the colors identify sets of mutually commuting hopping terms. (b) First-order Trotter decomposition. The same ordering of the terms—with its inverse—is used to construct the higher order decompositions we adopt. (c) ZZ rotation. (d) Fermionic SWAP. (e) Tunneling operator. (f) Simulation of the hopping terms. Site labels are those of panel (a). For visualization purposes, sites' positions are displayed such that the JW ordering matches an underlying snake-path. Initially, this only differs from the sites' indexing by interchanging the two rightmost columns. (f.1) Efficient computation of the JW strings for $H_{v,\text{odd}}$. (f.2) Simulation of $H_{v,\text{odd}}$ hopping. (f.3) Uncomputation of the JW strings used for $H_{v,\text{odd}}$. (f.4) Simulation of $H_{h,\text{odd}}$, trivially implemented due to a convenient initial choice of JW ordering. (f.5) Application of fermionic SWAPs, in the order specified by the labels, mutating the JW ordering. (f.6) Simulation of $H_{h,\text{even}}$, including hoppings wrapping across the boundaries. (f.7) Efficient computation of the JW strings for $H_{v,\text{even}}$. The last row is displayed at the top for visualisation purposes, without implying a novel change in the JW ordering. Bulk JW parities are computed first. Then these are mapped on another ancilla, which is used to further store and update the parity of qubits between the first and last row, to account for the periodic BCs. (f.8) Simulation of $H_{v,\text{even}}$ hopping. (f.9) Uncomputation of the JW strings for $H_{v,\text{even}}$.

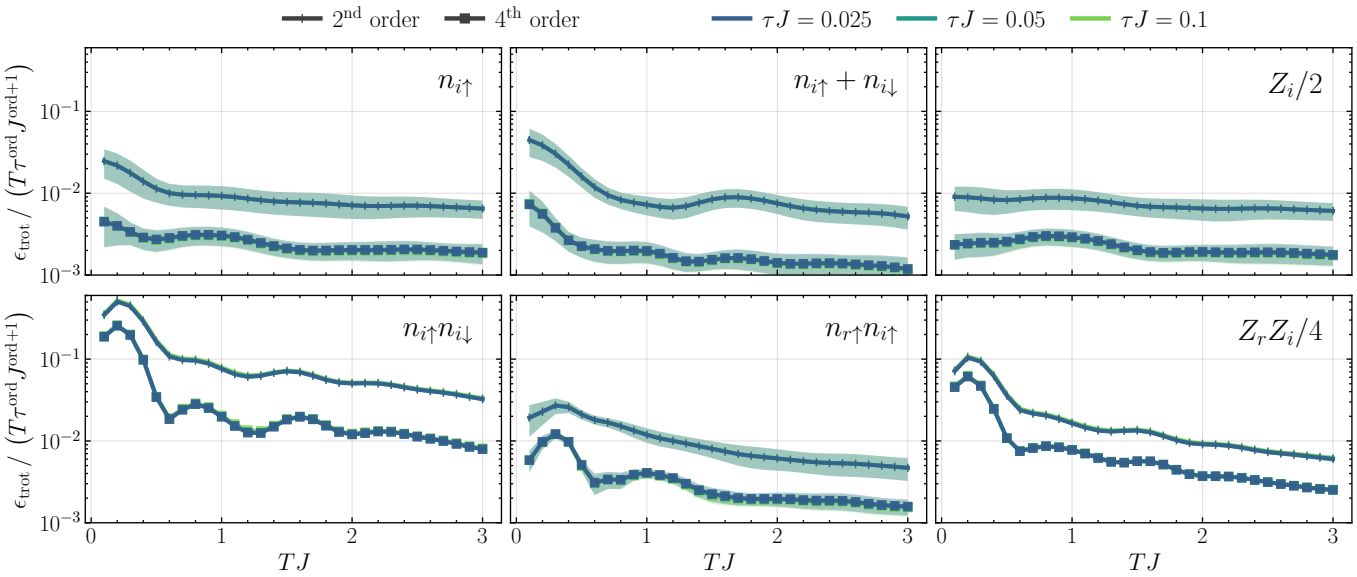


Figure 8. Estimates of the Trotter error in the adopted 2nd and 4th order decompositions, at half-filling and $U = 4J$, simulated in a 4×4 lattice. Each panel shows for the annotated observable, at different evolution times T, the estimated proportionality between the Trotter error ϵ_{trot} and $T\tau^{\text{ord}}J^{\text{ord}+1}$, with τ denoting the step size, and the error defined as per Eq. (E9). The curves' collapse confirms the expected scaling $\epsilon_{\text{trot}} \propto (\tau J)^{\text{ord}}$ at fixed T. The increasingly flat profile legitimates the scaling $\epsilon_{\text{trot}} \propto TJ$ at fixed τ. The shaded regions highlight 1σ standard deviation over 40 initial states, uniformly sampled from the degenerate ground subspace of the hopping Hamiltonian.

(f.3) The JW strings are then uncomputed by reversing the parity-computation circuit of (f.1). This completes the implementation of $H_{v,\text{odd}}$.

(f.4) $H_{h,\text{odd}}$ currently contains only nearest-neighbor terms. It is therefore straightforwardly simulated by applying the local tunneling rotations.

(f.5) The JW ordering is then modified by applying fermionic swaps across each row.

(f.6) In the updated ordering, $H_{h,\text{even}}$ is now local and readily simulated with tunneling rotations.

(f.7) The nested parity-computation procedure is used for $H_{v,\text{even}}$. Because of the periodic BCs, the JW strings connecting the first and last rows include the bulk string, and are therefore computed after the shorter bulk parities have been obtained.

(f.8) After these JW strings have been computed, the corresponding tunneling rotations are applied.

(f.9) Finally, the JW strings are uncomputed by reversing the circuit of (f.7), completing the implementation of $H_{v,\text{even}}$.

We observe that this scheme yields a lower 2-qubit gate count than FSWAP networks, as well as the most recent implementation with JW mapping compiled to an equivalent digital gate set [117]. However, it requires L_y ancillae and larger circuit depth.

The resulting per-step gate counts and circuit depths are reported in Table IV for each term in the Hamiltonian as well as the combination over the second-order Trotter decomposition considered. The counts and depths are further separated into single-qubit rotations $R_{X/Z}$, S_X gates, and two-qubit Clifford gates CX/CZ. As this scheme adapts straightforwardly to open BCs, estimates are provided for this case as well. We note that in this scenario the requirement that L_x and L_y be even is relaxed, since it was only needed for an efficient treatment of the hopping terms wrapping around the periodic boundaries.

Finally, we note that the construction extends naturally to higher-order product formulas. We consider a fourth-order Suzuki decomposition obtained as a symmetric product of five second-order steps,

$$V_4(t) = V_2(vt)\, V_2(vt)\, V_2\big([1-4v]t\big)\, V_2(vt)\, V_2(vt), \quad \text{(E8)}$$

with $v = 1/(4 - 4^{1/3})$. As in the second-order case, the half-steps of adjacent second-order steps merge at every interior boundary. The difference in time step affects only the rotation angles, leaving the gate counts unchanged. Consequently, the per-step counts and depths are exactly five times those of the second-order scheme reported in Table IV, for both periodic and open BCs.

Finally, we estimate the Trotter error $\epsilon_{\text{trot}}^{(\text{ord})}$ of the considered 2nd and 4th order decompositions. Specifically, we estimate the proportionality constant for the ansatz $\epsilon_{\text{trot}}^{(\text{ord})} \propto N\,(\tau J)^{\text{ord}+1}$, where N is the number of Trotter steps and τ is the step size. Note, $N\,\tau^{\text{ord}+1} = T\,\tau^{\text{ord}}$, where T is the total evolution time. We use the `ffsim` package [114] to simulate trotterised and exact time-

evolution, and compute the root-mean-square difference

$$\epsilon_{\rm trot} = \sqrt{\frac{1}{L_x L_y}\sum_{i=1}^{L_x L_y} |\langle O_i\rangle_{\rm trot} - \langle O_i\rangle_{\rm exact}|^2} \qquad \text{(E9)}$$

on various observables O of typical interest, averaged over the lattice. These are density $n_{i,\uparrow}$, occupancy $n_{i,\uparrow}+n_{i,\downarrow}$, double occupancy $n_{i,\uparrow}n_{i,\downarrow}$, spin $Z_i/2$, density correlation $n_{r,\uparrow}n_{i,\uparrow}$, and spin correlation $Z_rZ_i/4$ [108–110]. In the latter two, r denotes a reference site in the periodic lattice. The considered setting is the classically challenging regime characterized by half-filling and $U = 4J$. Simulations were performed at lattice size 4×4. We intend our analysis to be representative of low-temperature behavior, thus we considered initial states uniformly sampled in the span of the degenerate ground subspace of the hopping Hamiltonian H_J [108–110].

Figure 8 displays, separately for each observable, the estimated value of the proportionality constant at different steps of the evolution. Simulations were performed over evolution times up to $3J^{-1}$, for three different step sizes $\tau J = 0.025$, 0.05, 0.1, and averaging results over 40 initial states. On the one hand, the curves' collapse confirms the expected scaling $\epsilon_{\rm trot}^{(\rm ord)} \propto (\tau J)^{\rm ord}$ at fixed T. On the other hand, curves are not perfectly flat, implying a more complex dependence on evolution time than simply linear. However, an increasingly flat profile is observed with larger times, legitimating the adopted ansatz. Different observables inevitably provide quantitatively different estimates. To provide a versatile heuristic description, we set the proportionality constants to 10^{-2} and 3×10^{-3} for $2^{\rm nd}$ and $4^{\rm th}$ order decompositions respectively.

Appendix F: Resources for Execution of Arbitrary Rotations

In this appendix, we compare three ways of executing a batch of b logical single-qubit rotations by an arbitrary-angle θ_L: partial fault-tolerance, Clifford+T decomposition with magic state cultivation, and Hamming Weight Phasing (HWP). Throughout this subsection, resources are quoted for one batch; in an algorithm containing N_R such rotations, the corresponding counts should be multiplied by N_R/b, up to integer roundings. The total number of ancillary patches available is denoted $Q_{\rm syn}$, as in Sec. VI A. We focus on the cost of both Clifford and non-Clifford operations, as both have a significant impact in near term quantum computation.

We further assume θ_L to be a dyadic fraction of $\pi/2$ of the form $\theta_L = (2k+1)\pi/2^{C+1}$, where k and C are integers. This is a convenient ansatz, as it provides a few practical advantages (as explained later in this section) and is capable of approximating any angle to arbitrary precision.

In the remainder of the section, for each method, the following are determined

1. the average number of operations (both Clifford and non);
2. the required number of ancillae for a maximally parallel implementation, i.e. with the smallest possible circuit depth.
3. the circuit depth in the maximally parallel implementation, i.e. upon availability of a large enough ancilla reservoir to fully parallelize the encoding of the resource states;
4. the effect of spacetime tradeoffs, in which the number of ancillae is reduced at the expense of a larger depth.

With regard to circuit depth, analytic estimates will often be only approximate. Nevertheless, they show excellent afreement with numerical simulations, and provide further insight on the depth's scaling in the regime of interest and on the relevant spacetime tradeoffs. Results presented in the main text use estimates derived from numerical simulations rather than simplified analytic ansätze. To convert these counts into a common depth unit, we assume comparable times for the encoding and the injection of resource states. This assumption is justified in the context of neutral atom implementations, in which measurements and shuttling represent the main time bottleneck.

1. Resource Overhead of Partial Fault tolerance

In the partially FT implementation, each logical rotation is executed by preparing a resource state and injecting it through a RUS procedure. No fully fault-tolerant T gate is required in this model. At each $s^{\rm th}$ step of the RUS scheme, with $s = 1, 2, \ldots$ the injected angle is doubled, such that, after C unsuccessful steps, the angle to be injected is an odd multiple of $\pi/2$. This corresponds to a Clifford gate, hence it can be implemented deterministically, thus ending the RUS injection.

To determine the gate count, it should be noted that a qubit reaches stage s with probability $2^{-(s-1)}$. Therefore, the expected number of injections in a batch of size b at the $s^{\rm th}$ step is $b_s = 2^{-(s-1)}b$, and the total over the whole RUS process is

$$N_{\rm inj}^{\rm pFT}(b, C) = \sum_{s=1}^{C} b_s = 2b\left(1 - 2^{-C}\right). \qquad \text{(F1)}$$

To determine depth and ancillae requirements, first we focus on the maximally parallel implementation of the scheme. We denote $A_s \equiv A(\theta_{L,s})$ the average number of attempts required to successfully encode a resource state at the required angle. We note that this further depends on the code distance and physical error rate, and is described in Sec. V B. For a maximally parallel implementation, enough ancillae are required to be able to carry out the $s^{\rm th}$ step's injection while simultaneously

preparing resource states for the next step. The injection requires b_s resource states; the preparation step targets the encoding of b_{s+1}, and can re-use the $b_s(A_s-1)$ resource states that failed to be encoded in the previous step. The total number of ancillae required at step s grows with batch size, so we denote it as $b\,q_s$, where from the previous arguments it follows that

$$b\,q_s = b_s + \max[\,b_{s+1}A_{s+1},\, b_s(A_s-1)\,] \tag{F2}$$

such that

$$q_s = 2^{-(s-1)}\left[1+\max\left(\frac{A_{s+1}}{2}, A_s-1\right)\right]. \tag{F3}$$

Then, to be maximally parallel we need an ancilla reservoir of size

$$Q_{\rm syn}^{\rm pFT,prl}(b) = \lceil b(A_{\rm pFT}+1)\rceil, \tag{F4}$$

where $A_{\rm pFT} = \max_s q_s - 1$ is a batch-independent metric, and can be interpreted as an *effective* number of attempts per injection in the RUS protocol. In fact, its definition makes it directly comparable to the average number of attempts of T states in magic state cultivation (see next subsection).

The average circuit depth is the expectation value of the step $s_{\max}$ at which the RUS process over the batch is terminated.. This can be computed using the tail-sum formula for non-negative integer random variables,

$$\begin{aligned} D^{\rm pFT,prl}(b,C) =& \mathbb{E}[s_{\max}] \\ =& \sum_{s=1}^{\infty} 1-[1-2^{-(s-1)}]^b, \end{aligned} \tag{F5}$$

where the sum runs over the probabilities that the RUS process over the whole batch extends to or beyond step s. This expectation value, up to negligible corrections, evaluates to $\log_2 b + 1/2 + \gamma_{em}/\ln 2$ where $\gamma_{em}\approx 0.577$ is the Euler-Mascheroni constant [126]. Once the cutoff C is taken into account, the average depth is approximately

$$D^{\rm pFT,prl}(b,C) \simeq \min(\log_2 b + 1.33, C). \tag{F6}$$

Finally, we consider the effects of a spacetime tradeoff in which the size of the ancillae reservoir is reduced by a factor $\mathfrak{r}$, i.e.

$$Q_{\rm syn}^{\rm pFT}(b;\mathfrak{r}) = \lceil b(A_{\rm pFT}+1)/\mathfrak{r}\rceil\,. \tag{F7}$$

In this scenario, each step of the RUS protocol contributes on average depth $\lceil bq_s/Q_{\rm syn}^{\rm pFT}(b;\mathfrak{r})\rceil \approx \lceil \mathfrak{r}\,q_s/(A_{\rm pFT}+1)\rceil$ instead of unity. The change in depth then is, on average,

$$\Delta D^{\rm pFT}(b,C;\mathfrak{r}) \approx \sum_{s=1}^{C}\left[\left\lceil\frac{\mathfrak{r}\,q_s}{A_{\rm pFT}+1}\right\rceil - 1\right]. \tag{F8}$$

Under the assumption, satisfied in practice in Sec. V B, that the A_s coefficients vary slowly enough to approximate $\lceil q_s/(A_{\rm pFT}+1)\rceil \sim \lceil 2^{-(s-1)}\rceil$, only the first $s_{\mathfrak{r}} = \lceil\log_2\mathfrak{r}\rceil$ steps are impacted by the reduction in ancillae, and we further simplify the terms in the above sum to $\sim \mathfrak{r}\,2^{-(s-1)}$. Taking into account the presence of the cutoff, we denote $m_{\mathfrak{r}C} = \min(s_{\mathfrak{r}}, C)$, and estimate

$$\begin{aligned} \Delta D^{\rm pFT}(b,C;\mathfrak{r}) &\sim \sum_{s=1}^{m_{\mathfrak{r}C}} 2^{-(s-1)}\mathfrak{r} \\ &\sim 2\mathfrak{r}(1-2^{-m_{\mathfrak{r}C}}) \end{aligned} \tag{F9}$$

Therefore, the total depth is approximately

$$\begin{aligned} D^{\rm pFT}(b,C;\mathfrak{r}) \simeq\ & \min(\log_2 b + 1.33, C) \\ &+ 2\mathfrak{r}(1-2^{-m_{\mathfrak{r}C}}), \end{aligned} \tag{F10}$$

in good agreement with exact numerical simulations, as shown in Fig. 9(a). Remarkably, the depth is only logarithmic in batch size, and, upon spacetime tradeoffs, it only has a small additive increase.

Finally, we also note that the only Clifford operation that must be accounted for at the logical level is the possible S correction to be applied after the $C^{\rm th}$ step of the RUS process, which can always be applied as one parallel Clifford layer. Therefore, partial fault tolerance has Clifford depth per parallel batch $\lesssim 1$.

2. Resource Overhead of Clifford+T Decomposition

The fully FT baseline synthesizes each arbitrary rotation directly into a Clifford+T sequence. Near-term quantum hardware is likely to have limited size, thus we adopt the most efficient ancilla-free synthesis protocol, estimated to require

$$N_T^{\rm Tdec}(\epsilon) \simeq 1.52\log_2(\epsilon^{-1})\ -0.01 \tag{F11}$$

T gates. This is the mixed protocol provided in Ref. [24], in turn improving the original proposals in Refs. [127, 128].

The corresponding Clifford overhead is determined by the Matsumoto–Amano normal form of the Clifford+T decomposition [129, 130], which is independent of the synthesis technique adopted. In the normal form, up to at most single boundary operators, each T gate appears in one of two elementary patterns, both containing H gates, and only one of which contains an S gate. Therefore, it is expected that, on average,

$$N_H^{\rm Tdec} \simeq N_T^{\rm Tdec}, \qquad N_S^{\rm Tdec} \simeq N_T^{\rm Tdec}/2. \tag{F12}$$

This was numerically verified for Ross–Selinger decompositions generated with `pygridsynth` [131, 132] over the precision range relevant here. The ratio between S and T count remains the same once the probabilistic S-gate correction to T-gate teleportation is taken into account and merged with surrounding Clifford operators, as shown in Fig. 9(b). The number of H gates is reduced to $N_T^{\rm Tdec} \bmod 2 = 0$ or 1 if T_X and S_X gates are available. Finally, we note that all Clifford gates can be moved to

the end of the synthesis and effectively reduced to a single Clifford gate if T_Y gates are also natively available [133], but this scenario extends beyond the assumptions of the manuscript.

Let A_{FT} be the mean number of attempts required to successfully prepare a T resource state. Recent proposals optimized for a neutral-atom noise mode [46] estimate $A_{\mathrm{FT}} \simeq 1.3$ for relevant error rates and code distances. In the maximally parallel scenario, the ancilla reservoir provides b resource states to be injected, while using $A_{\mathrm{FT}} b$ logical patches to encode the next layer's resource states, thus has size

$$Q_{\mathrm{syn}}^{\mathrm{Tdec,prl}}(b) = \lceil b(A_{\mathrm{FT}} + 1) \rceil. \tag{F13}$$

With this footprint, all b rotations are synthesized in parallel, yielding non-Clifford depth

$$D^{\mathrm{Tdec,prl}}(\epsilon) = N_T^{\mathrm{Tdec}}(\epsilon). \tag{F14}$$

As all qubits in the batch require the same amount of injection, reducing the side-patch budget by a factor $\mathfrak{r} \geq 1$ yields a multiplicative increase in depth. Thus, the spacetime tradeoff is

$$\begin{aligned} Q_{\mathrm{syn}}^{\mathrm{Tdec}}(b;\mathfrak{r}) &= \left\lceil \frac{b(A_{\mathrm{FT}}+1)}{\mathfrak{r}} \right\rceil, \\ D^{\mathrm{Tdec}}(\epsilon;\mathfrak{r}) &= \left\lceil \mathfrak{r} N_T^{\mathrm{Tdec}}(\epsilon) \right\rceil. \end{aligned} \tag{F15}$$

The Clifford depth is proportional to the T-depth as specified by Eq. (F12).

3. Resource Overhead of Hamming Weight Phasing

Hamming Weight Phasing (HWP) is a technique introduced to reduce the number of synthesized arbitrary rotations, applicable when equiangular rotations are applied in parallel [104, 105]. Instead of synthesizing b independent rotations, one computes the Hamming weight of the batch, applies rotations to the Hamming-weight register of size m_{Cb}, and then uncomputes the register. For rotation angles that are dyadic fractions of π, only the first $m_{Cb} = \min(C-1, \lfloor \log_2 b \rfloor + 1)$ digits of the Hamming weight are to be computed, as no advantage is obtained from computing the next ones. Moreover, it will be useful to define $\sigma_2(b, C)$ as the Hamming weight of the lowest m_{Cb} bits of b in binary form.

The reduction in synthesis cost comes at the expense of additional Clifford and T gates required to compute and uncompute the Hamming weight. The building blocks are the so called full- and half-adders, in turn relying on the use of logical-AND and inverse-AND gadgets [104]. These are depicted in Fig. 9(e) as from Ref. [112]. Essentially, each qubit is assigned a weight, and can be "processed" or "unprocessed". Initially, all qubits have weight 1 and are unprocessed, i.e., they have not yet contributed to the computation of the Hamming weight. A full(half)-adder introduces an ancillary qubit, consumes four T states, and takes three(two) unprocessed qubits of the same weight; then outputs two(one) processed qubits of same weight, one unprocessed qubit of same weight, and one unprocessed qubit of weight twice as large. The computation of the Hamming-weight bits terminates when exactly one unprocessed qubit is left for each weight $w_\ell = 2^\ell$, with $\ell = 0, \ldots, m_{Cb} - 1$, constituting the Hamming-weight register. Full adders are mostly used, with half adders only required at most once per weight, if two unprocessed qubits of same weight are left.

To estimate non-Clifford counts, the number of full and half adders need to be determined. Denote b_ℓ, f_ℓ, h_ℓ, respectively, the number of qubits, full-adders and half-adders of weight w_ℓ to be processed. Moreover, note that, as each adder produces one higher-weight qubit, $b_{\ell+1} = f_\ell + h_\ell$. As a full adder effectively removes two qubits of the given weight to account for, and is applied as long as more than two weight w_ℓ qubits are present: $b_{\ell+1} = \lfloor b_\ell/2 \rfloor = \lfloor b/2^{\ell+1} \rfloor$, and a half-adder is only required if b_ℓ is even. Since $\lfloor \cdot/2^\ell \rfloor$ is a binary shift operator—such that the binary representation of b_ℓ is that of b shifted by ℓ places—it is convenient to shift to a binary description where $b = \sum_{\ell=0}^{\lfloor \log_2 b \rfloor} \beta_\ell 2^\ell$. It immediately follows that $h_\ell = 1 \oplus \beta_\ell$, hence

$$\begin{aligned} h_{\mathrm{tot}}(b, C) &= \sum_{\ell=0}^{m_{Cb}-1} h_\ell = m_{Cb} - \sigma_2(b, C), \\ f_{\mathrm{tot}}(b, C) &= \sum_{\ell=0}^{m_{Cb}-1} b_{\ell+1} - h_\ell = b - m_{Cb} - \lfloor b 2^{-m_{Cb}} \rfloor, \end{aligned} \tag{F16}$$

where evaluation of the latter used the 2-adic valuation of Legendre's formula [134]. Then, the total number of adders is

$$N_{\mathrm{add}}^{\mathrm{HWP}}(b, C) = b - \sigma_2(b, C) - \left\lfloor b 2^{-m_{Cb}} \right\rfloor, \tag{F17}$$

exactly matching previous estimates in the absence of the cutoff C [107].

To determine the gate count, we note that each full- or half-adder consumes four T states, while the measurement-based inverse adders used during uncomputation are T-free. The m_{Cb} rotations applied to the Hamming-weight register must also be synthesized, while the $\lfloor b 2^{-m_{Cb}} \rfloor$ residual carry qubits at the cutoff weight are handled by individual T gates. The total T count for one batch is therefore

$$N_T^{\mathrm{HWP}}(b, C, \epsilon) = 4 N_{\mathrm{add}}^{\mathrm{HWP}}(b, C) + \left\lfloor \frac{b}{2^{m_{Cb}}} \right\rfloor + m_{Cb} N_T^{\mathrm{Tdec}}(\epsilon). \tag{F18}$$

The Clifford count also follows from the total number of adders used in the Hamming weight computation, and their inverse in the subsequent uncomputation. We do not write this count explicitly, since it is directly obtained from f_{tot}, h_{tot}, and the adder gadgets. The synthesized rotations on the Hamming-weight register add the same average Clifford overhead discussed in App. F 2.

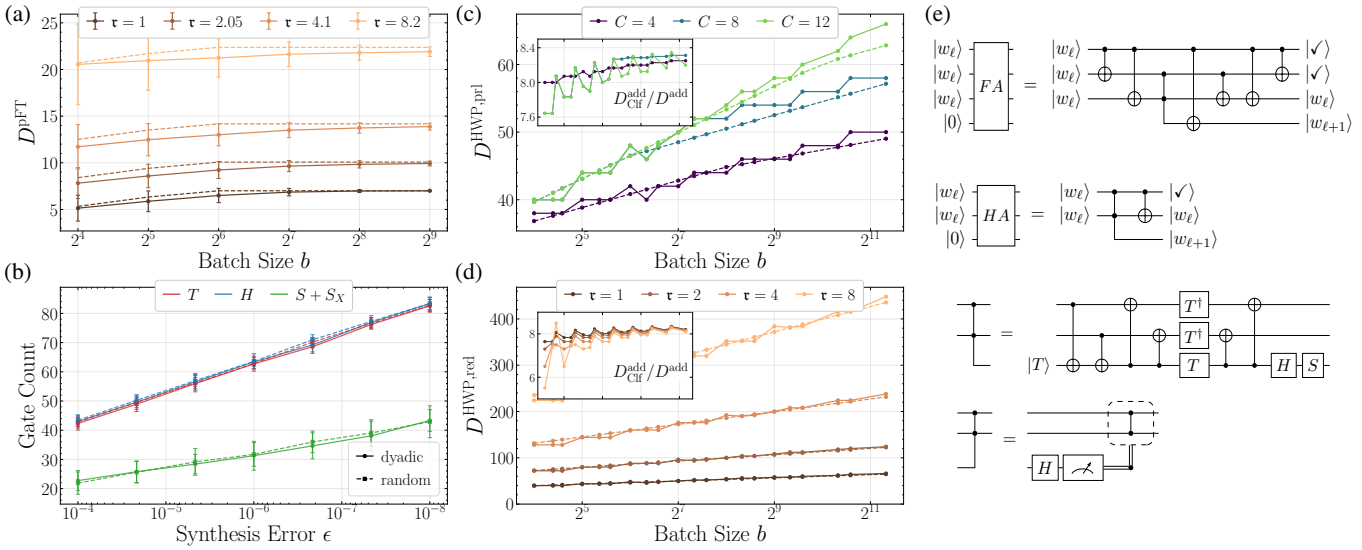


Figure 9. Overview of resource estimation for the synthesis of arbitrary-angle rotations. The angles considered are modeled as dyadic fractions of π, i.e. $\theta_L = (2k+1)\pi/2^{C+1}$ with integer k, unless specified otherwise. (a) Mean feed-forward depth of the partially FT approach for different batch sizes, at $C = 7$. It shows good agreement between exact scheduling (solid) and the analytic approximation in Eq. (F9) (dashed). $\mathfrak{r}$ denotes the reduction in the size of the ancilla-reservoir (see Eq. (F7)), with values chosen to minimize rounding effects given $A_{\mathrm{pFT}} \approx 1.05$. (b) Mean gate counts for Clifford+T synthesis with the Ross-Selinger decomposition. Similar estimates are returned whether the target angle is a dyadic fraction of π or random. Benchmarked for angles in $[0.003, 0.2]$. (c) Mean feed-forward depth of maximally-parallel HWP for different batch sizes and cutoffs C, assuming T-depth 26 for the synthesis of the required rotations. It shows good agreement between exact numerical simulations and the heuristic ansatz in Eq. (F25). The inset shows the ratio of the Clifford depth with respect to the feed-forward depth, over the same x-axis range. (d) Mean feed-forward depth of reduced HWP for different batch sizes and no cutoff $C \to \infty$, at different ancillae-reduction factor $\mathfrak{r}$. It shows good agreement between exact numerical simulations (solid) and the analytic ansatz in Eq. (F31) (dashed). The inset shows the ratio of the Clifford depth with respect to the feed-forward depth, over the same x-axis range. (e) Building blocks of HWP. $FA(HA)$ denotes a full(half)-adder, taking three(two) unprocessed input qubits of weight w_ℓ and returning two(one) processed qubits, one unprocessed qubit of weight w_ℓ and one of weight $w_{\ell+1}$. The final CNOT of the full-adder can be neglected. Below, the logical-AND and the inverse-AND gadgets are depicted, to be used in the adders and in their inverse, respectively.

To determine the spacetime cost, we first consider the maximally parallel implementation. In the Hamming-weight computation, each adder requires one ancilla, which cannot be re-used once the adder has been implemented, and the simultaneous production and injection of T states, in turn requiring an injection reservoir. Thus, the overall ancillae requirement is the sum of separate terms

$$Q_{\mathrm{syn}}^{\mathrm{HWP,prl}}(b, C, A_{\mathrm{FT}}) = Q_{\mathrm{add}}^{\mathrm{HWP}}(b, C) + Q_{\mathrm{inj}}^{\mathrm{HWP,prl}}(b, C, A_{\mathrm{FT}}). \tag{F19}$$

with

$$\begin{aligned} Q_{\mathrm{add}}^{\mathrm{HWP}}(b, C) &= N_{\mathrm{add}}^{\mathrm{HWP}}(b, C), \\ Q_{\mathrm{inj}}^{\mathrm{HWP,prl}}(b, C, A_{\mathrm{FT}}) &= \left\lceil b\, q_{\mathrm{inj}}^{\mathrm{HWP}}(b, C, A_{\mathrm{FT}}) \right\rceil, \end{aligned} \tag{F20}$$

where $q_{\mathrm{inj}}^{\mathrm{HWP}}(b, C, A_{\mathrm{FT}})$ is the number of injection side patches per input qubit needed to keep the adder tree supplied with T states. This depends on the number of attempts per kept T states A_{FT} introduced in the previous subsection; moreover, oppositely to the equivalent quantities for previous methods, it has a (mild) dependence on batch size. It will now be determined. As each full-adder takes three unprocessed input qubits and returns two, we assume that the number of adders per parallel step is

$$N_{\mathrm{add,s}}^{\mathrm{HWP}}(b) \approx \frac{b}{3}\left(\frac{2}{3}\right)^{s}. \tag{F21}$$

It should be noted that each parallel step s does not coincide with a weight layer ℓ: a single step can involve adders processing different weights, and, conversely, a weight might require more than a single step to be processed. Moreover, the adder's ancillae can be used to produce T gates before their use in the adders. Therefore, the ancillae available at step s for T states are—omitting their dependences for compactness,

$$Q_{\mathrm{inj,s}}^{\mathrm{HWP,prl}} \approx Q_{\mathrm{inj}}^{\mathrm{HWP,prl}} + Q_{\mathrm{add}}^{\mathrm{HWP}} - \sum_{t=0}^{s-1} N_{\mathrm{add,t}}^{\mathrm{HWP}} \tag{F22}$$

At all steps, the injection-ancillae reservoir must be large enough to supply T states for the injection as well as encode states for the next step—with the possibility of re-using the $4(A_{\mathrm{FT}} - 1)N_{\mathrm{add,s}}^{\mathrm{HWP}}$ logical patches that failed

to encode T states. This requires

$$Q_{\mathrm{inj,s}}^{\mathrm{HWP,prl}} \geq 4N_{\mathrm{add,s}}^{\mathrm{HWP}} \left[1 + \max\left(A_{\mathrm{FT}} - 1, \frac{2}{3}A_{\mathrm{FT}}\right)\right] \tag{F23}$$

at all steps. The minimum overhead required thus is

$$q_{\mathrm{inj}}^{\mathrm{HWP}}(b, C, A_{\mathrm{FT}}) \approx \frac{1}{3}\begin{cases} 4A_{\mathrm{FT}} - 3 & \text{if } A_{\mathrm{FT}} \geq 3 \\ \frac{8}{3}A_{\mathrm{FT}} + 1, & \text{otherwise} \end{cases} + 2^{-m_{C\flat}} + \sigma_2(b, C)/b. \tag{F24}$$

The circuit depth is well approximated by analytic closed forms obtained by simple heuristic arguments, yet empirically justified—see Fig. 9(c). In the absence of cutoff, the protocol terminates once a number of qubits of order unity is left, and, since every full-adder effectively reduces by 2/3 the number of unprocessed qubits, the depth is roughly $\log_{3/2} b \approx 1.7 \log_2 b$. In the presence of a cutoff much smaller than $\log_2 b$, only the first relatively-few layers matter, and, since they are reduced by 1/3 at each step, the depth scales as $\log_3 b$. In an intermediate stage, the depth is well approximated by $\log_2 b$ scaling. Therefore, the cumulative adders' depth is empirically approximated by the following piecewise ansatz

$$D^{\mathrm{add,prl}}(b, C) \sim \begin{cases} \log_{3/2} b & \text{if } \lfloor \log_2 b \rfloor \lesssim C - 2 \\ \log_2 b + \text{const.} & \text{else if } \lfloor \log_2 b \rfloor \lesssim 2C \\ \log_3 b + \text{const.} & \text{otherwise} \end{cases} \tag{F25}$$

where the constants' values depend on the cutoff and are fit to match the boundaries. The uncomputation of the Hamming weight uses measurements and feedforward, but no T gates, hence has the maximally parallel adder depth, independently of the availability of non-Clifford resources. The phasing stage consists of rotation synthesis by means of Clifford+T decomposition, with depth determined as per Eq. (F14). Therefore, the maximally parallel HWP depth is

$$D^{\mathrm{HWP,prl}}(b, C, \epsilon) = 2D^{\mathrm{add,prl}}(b, C) + D^{\mathrm{Tdec,prl}}(\epsilon). \tag{F26}$$

The Clifford depth in the computation and uncomputation processes follows the same adder-tree structure, with the forward and inverse adders contributing a constant number of Clifford layers per step. This is roughly 8 times the feed-forward depth D^{add}, as shown in the inset of Fig. 9(c), and as expected from the structure of the adders and their inverse, depicted in Fig. 9(e). The synthesized Hamming-weight rotations contribute the same Clifford depth discussed for T-gate synthesis.

Finally, we consider spacetime tradeoffs. In the near term, limited hardware size is likely to limit the size of the ancillae reservoir to a significantly reduced size $Q_{\mathrm{red}} < b$. Since the number of ancillae needed for the adders approximately matches the batch size, only a "reduced" version of HWP is possible, where the batch b is accordingly split into balanced chunks b_j,

$$b = \sum_{j=1}^{\mathfrak{r}} b_j, \qquad b_j \in \left\{ \left\lfloor \frac{b}{\mathfrak{r}} \right\rfloor, \left\lceil \frac{b}{\mathfrak{r}} \right\rceil \right\}, \tag{F27}$$

with $\mathfrak{r}$ chosen as the smallest integer such that

$$\max_j Q_{\mathrm{syn}}^{\mathrm{HWP,prl}}(b_j, C, A_{\mathrm{FT}}) \leq Q_{\mathrm{red}}. \tag{F28}$$

These are executed sequentially, and the peak ancilla count is then the maximum chunk footprint. The novel gate count and depth are obtained by summing the requirements of the chunks,

$$N_T^{\mathrm{HWP,red}}(b, C, \epsilon, \mathfrak{r}) = \sum_{j=1}^{\mathfrak{r}} N_T^{\mathrm{HWP}}(b_j, C, \epsilon). \tag{F29}$$

$$D^{\mathrm{HWP,red}}(b, C, \epsilon, \mathfrak{r}) = \sum_{j=1}^{\mathfrak{r}} D^{\mathrm{HWP,prl}}(b_j, C, \epsilon). \tag{F30}$$

To get an intuitive understanding of the spacetime tradeoff, we note that in practical implementations in the near-term, the depth can be generally approximated as

$$D^{\mathrm{HWP,red}}(b, \epsilon, \mathfrak{r}) \sim \mathfrak{r}\left[2\log_{3/2}(b/\mathfrak{r}) + D^{\mathrm{Tdec,prl}}(\epsilon)\right]. \tag{F31}$$

In fact, $b_j \sim b/\mathfrak{r} < 2^{C-1}$ is expected, due to the combination of relatively small system sizes and high $\mathfrak{r}$. The resulting depth for different batch sizes and $\mathfrak{r}$ values is portrayed in Fig. 9(d), showing excellent agreement between exact numerical simulations and the analytic ansatz above. The same chunking logic applying to the Clifford depth. Thus, the Clifford depth in the adders and their inverse remains roughly 8 times the feed-forward depth D^{add}, as shown in the inset of Fig. 9(d), and the total Clifford depth is approximately

$$D_{\mathrm{Clf}}^{\mathrm{HWP,red}}(b, \epsilon, \mathfrak{r}) \sim \mathfrak{r}\left[16\log_{3/2}(b/\mathfrak{r}) + D^{\mathrm{Tdec}}(\epsilon)\right]. \tag{F32}$$

A "grouped" version of HWP also exists [105], where the Hamming weight of the whole batch is computed one sub-batch at a time. Unlike the reduced HWP, the rotation synthesis is still applied only once, to the Hamming weight of the full batch b. However, this approach significantly increases circuit depth and Clifford usage, as it requires repeating weight computation and uncomputation twice for each sub-batch—more details can be found in Ref. [105]. Therefore, it is unlikely to be a viable solution in the near-term, when shorter depth is preferable, and the combination of higher error rate and smaller code distance make noise from Clifford gates relevant.

[1] B. M. Terhal, Quantum error correction for quantum memories, Rev. Mod. Phys. **87**, 307 (2015).

[2] Google Quantum AI, Suppressing quantum errors by scaling a surface code logical qubit, Nature **614**, 676 (2023).

[3] R. S. Gupta, N. Sundaresan, T. Alexander, C. J. Wood, S. T. Merkel, M. B. Healy, M. Hillenbrand, T. Jochym-O'Connor, J. R. Wootton, T. J. Yoder, *et al.*, Encoding a magic state with beyond break-even fidelity, Nature **625**, 259 (2024).

[4] L. Caune, L. Skoric, N. S. Blunt, A. Ruban, J. McDaniel, J. A. Valery, A. D. Patterson, A. V. Gramolin, J. Majaniemi, K. M. Barnes, *et al.*, Demonstrating real-time and low-latency quantum error correction with superconducting qubits, Nature Communications **17**, 7383 (2026).

[5] Google Quantum AI and Collaborators, Quantum error correction below the surface code threshold, Nature 10.1038/s41586-024-08449-y (2024).

[6] N. Lacroix, A. Bourassa, F. J. H. Heras, L. M. Zhang, J. Bausch, A. W. Senior, T. Edlich, N. Shutty, V. Sivak, A. Bengtsson, *et al.*, Scaling and logic in the colour code on a superconducting quantum processor, Nature **645**, 614 (2025).

[7] I. Pogorelov, F. Butt, L. Postler, C. D. Marciniak, P. Schindler, M. Müller, and T. Monz, Experimental fault-tolerant code switching, Nature Physics **21**, 298 (2025).

[8] C. Ryan-Anderson, N. C. Brown, C. H. Baldwin, J. M. Dreiling, C. Foltz, J. P. Gaebler, T. M. Gatterman, N. Hewitt, C. Holliman, C. V. Horst, *et al.*, High-fidelity teleportation of a logical qubit using transversal gates and lattice surgery, Science **385**, 1327 (2024).

[9] A. Paetznick, M. P. da Silva, C. Ryan-Anderson, J. M. Bello-Rivas, J. P. C. III, A. Chernoguzov, J. M. Dreiling, C. Foltz, F. Frachon, J. P. Gaebler, T. M. Gatterman, L. Grans-Samuelsson, D. Gresh, D. Hayes, N. Hewitt, C. Holliman, C. V. Horst, J. Johansen, D. Lucchetti, Y. Matsuoka, *et al.*, Demonstration of logical qubits and repeated error correction with better-than-physical error rates (2024), arXiv:2404.02280.

[10] S. Dasu, M. DeCross, A. Y. Guo, A. Lavasani, J. Behrends, A. Benhemou, Y.-H. Chen, K. Mayer, C. N. Self, S. Simsek, B. Srivastava, M. S. Allman, J. Arkinstall, J. G. Bohnet, N. Q. Burdick, J. P. C. III, A. Chernoguzov, S. F. Cooper, R. D. Delaney, J. M. Dreiling, *et al.*, Computing with many encoded logical qubits beyond break-even (2026), arXiv:2602.22211.

[11] M. H. Abobeih, Y. Wang, J. Randall, S. J. H. Loenen, C. E. Bradley, M. Markham, D. J. Twitchen, B. M. Terhal, and T. H. Taminiau, Fault-tolerant operation of a logical qubit in a diamond quantum processor, Nature **606**, 884 (2022).

[12] D. Bluvstein, H. Levine, G. Semeghini, T. T. Wang, S. Ebadi, M. Kalinowski, A. Keesling, N. Maskara, H. Pichler, M. Greiner, *et al.*, A quantum processor based on coherent transport of entangled atom arrays, Nature **604**, 451 (2022).

[13] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, *et al.*, Logical quantum processor based on reconfigurable atom arrays, Nature **626**, 58 (2024).

[14] B. W. Reichardt, A. Paetznick, D. Aasen, I. Basov, J. M. Bello-Rivas, P. Bonderson, R. Chao, W. van Dam, M. B. Hastings, R. V. Mishmash, A. Paz, M. P. da Silva, A. Sundaram, K. M. Svore, A. Vaschillo, Z. Wang, M. Zanner, W. B. Cairncross, C.-A. Chen, D. Crow, *et al.*, Fault-tolerant quantum computation with a neutral atom processor (2025), arXiv:2411.11822.

[15] W. C. Chung, D. C. Cole, P. Gokhale, E. B. Jones, K. W. Kuper, D. Mason, V. Omole, A. G. Radnaev, R. Rines, M. H. Teo, *et al.*, Fault-tolerant operation and materials science with neutral atom logical qubits, npj Quantum Information **11**, 193 (2025).

[16] P. S. Rodriguez, J. M. Robinson, P. N. Jepsen, Z. He, C. Duckering, C. Zhao, K.-H. Wu, J. Campo, K. Bagnall, M. Kwon, *et al.*, Experimental demonstration of logical magic state distillation, Nature **645**, 620 (2025).

[17] D. Bluvstein, A. A. Geim, S. H. Li, S. J. Evered, J. P. B. Ataides, G. Baranes, A. Gu, T. Manovitz, M. Xu, M. Kalinowski, *et al.*, A fault-tolerant neutral-atom architecture for universal quantum computation, Nature **649**, 39 (2026).

[18] R. Rines, B. Hall, M. H. Teo, J. Viszlai, D. C. Cole, D. Mason, C. Barker, M. J. Bedalov, M. Blakely, T. Bothwell, C. Carnahan, F. T. Chong, S. Y. Eubanks, B. Fields, M. Gillette, P. Goiporia, P. Gokhale, G. T. Hickman, M. Iliev, E. B. Jones, *et al.*, Demonstration of a logical architecture uniting motion and in-place entanglement (2026), arXiv:2509.13247.

[19] B. Eastin and E. Knill, Restrictions on transversal encoded quantum gate sets, Phys. Rev. Lett. **102**, 110502 (2009).

[20] S. Bravyi and A. Kitaev, Universal quantum computation with ideal clifford gates and noisy ancillas, Phys. Rev. A **71**, 022316 (2005).

[21] D. Litinski, Magic State Distillation: Not as Costly as You Think, Quantum **3**, 205 (2019).

[22] C. Gidney, N. Shutty, and C. Jones, Magic state cultivation: growing t states as cheap as cnot gates (2024), arXiv:2409.17595.

[23] J. T. Anderson, G. Duclos-Cianci, and D. Poulin, Fault-tolerant conversion between the Steane and reed-muller quantum codes, Phys. Rev. Lett. **113**, 080501 (2014).

[24] V. Kliuchnikov, K. Lauter, R. Minko, A. Paetznick, and C. Petit, Shorter quantum circuits via single-qubit gate approximation, Quantum **7**, 1208 (2023).

[25] 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**, 10.1103/PRXQuantum.5.010337 (2024).

[26] H. Choi, F. T. Chong, D. Englund, and Y. Ding, Fault tolerant non-clifford state preparation for arbitrary rotations (2023), arXiv:2303.17380.

[27] R. Toshio, Y. Akahoshi, J. Fujisaki, H. Oshima, S. Sato, and K. Fujii, Practical quantum advantage on partially fault-tolerant quantum computer, Physical Review X **15**, 10.1103/PhysRevX.15.021057 (2025).

[28] R. Ismail, I.-C. Chen, C. Zhao, R. Weiss, F. Liu, H. Zhou, S.-T. Wang, A. Sornborger, and M. Kornjača, Transver-

sal architecture for megaquop-scale quantum simulation with neutral atoms, PRX Quantum **7**, 020343 (2026).
[29] P. Zeng, G. Zheng, Q. Xu, and L. Jiang, Error-structure-tailored early fault-tolerant quantum computing (2025), arXiv:2511.19983.
[30] R. Ismail, M. Kornjača, H.-Y. Hu, N. Maskara, S.-T. Wang, H. Zhou, and C. Zhao, Fast and parallel high-rate star architecture for megaquop quantum simulation (2026), 2606.25011.
[31] A. J. Menssen, T. Wang, M. Gullans, T. Manovitz, J. M. Taylor, J. Cong, J. Sinclair, Z. Aqua, D. J. Blumenthal, J. P. B. Ataides, J. Borregaard, A. Browaeys, P. Cappellaro, S. Choi, A. Cooper, R. Côté, J. P. Covey, A. Dauphin, I. Dimitrova, M. Eichenfield, *et al.*, Strategic plan for neutral atom quantum computation (2026), arXiv:2607.21554.
[32] H. Zhou, C. Zhao, M. Cain, D. Bluvstein, N. Maskara, C. Duckering, H.-Y. Hu, S.-T. Wang, A. Kubica, and M. D. Lukin, Low-overhead transversal fault tolerance for universal quantum computation, Nature **646**, 303 (2025).
[33] M. Cain, C. Zhao, H. Zhou, N. Meister, J. P. B. Ataides, A. Jaffe, D. Bluvstein, and M. D. Lukin, Correlated decoding of logical algorithms with transversal gates, Phys. Rev. Lett. **133**, 240602 (2024).
[34] M. Cain, D. Bluvstein, C. Zhao, S. Gu, N. Maskara, M. Kalinowski, A. A. Geim, A. Kubica, M. D. Lukin, and H. Zhou, Fast correlated decoding of transversal logical algorithms (2025), arXiv:2505.13587.
[35] M. Serra-Peralta, M. H. Shaw, and B. M. Terhal, Decoding across transversal clifford gates in the surface code, PRX Quantum **7**, 010335 (2026).
[36] M. B. Hastings, D. Wecker, B. Bauer, and M. Troyer, Improving quantum algorithms for quantum chemistry, Quantum Information & Computation **15**, 1 (2015).
[37] S. Flannigan, N. Pearson, G. H. Low, A. Buyskikh, I. Bloch, P. Zoller, M. Troyer, and A. J. Daley, Propagation of errors and quantitative quantum simulation with quantum advantage, Quantum Science and Technology **7**, 045025 (2022).
[38] M. Papič, M. G. Algaba, E. Godinez-Ramirez, I. de Vega, A. Auer, F. I. Šimkovic, and A. Calzona, Near-term fermionic simulation with subspace noise tailored quantum error mitigation, npj Quantum Information **12**, 10.1038/s41534-026-01248-5 (2026).
[39] F. Alam, J. L. Bosse, I. Čepaitė, A. Chapman, L. Clinton, M. Crichigno, E. Crosson, T. Cubitt, C. Derby, O. Dowinton, N. Eassa, P. K. Faehrmann, S. Flammia, B. Flynn, F. M. Gambetta, R. García-Patrón, M. Hunter-Gordon, G. Jones, A. Khedkar, J. Klassen, *et al.*, Fermionic dynamics on a trapped-ion quantum computer beyond exact classical simulation (2025), arXiv:2510.26300.
[40] S. Bravyi and R. König, Classification of topologically protected gates for local stabilizer codes, Phys. Rev. Lett. **110**, 170503 (2013).
[41] E. T. Campbell, H. Anwar, and D. E. Browne, Magic-state distillation in all prime dimensions using quantum reed-muller codes, Phys. Rev. X **2**, 041021 (2012).
[42] E. T. Campbell and M. Howard, Unified framework for magic state distillation and multiqubit gate synthesis with reduced resource cost, Phys. Rev. A **95**, 022316 (2017).
[43] J. O'Gorman and E. T. Campbell, Quantum computation with realistic magic-state factories, Phys. Rev. A **95**, 032338 (2017).
[44] A. Wills, M.-H. Hsieh, and H. Yamasaki, Constant-overhead magic state distillation, Nature Physics **21**, 1842 (2025).
[45] K. Sahay, P.-K. Tsai, K. K. Chang, Q. Su, T. B. Smith, S. Singh, and S. Puri, Fold-transversal surface code cultivation, PRX Quantum **7**, 033006 (2026).
[46] Y. Vaknin, S. Jacoby, A. Grimsmo, and A. Retzker, High rate magic state cultivation on the surface code, PRX Quantum **7**, 010353 (2026).
[47] M. E. Beverland, A. Kubica, and K. M. Svore, Cost of universality: A comparative study of the overhead of state distillation and code switching with color codes, PRX Quantum **2**, 020341 (2021).
[48] F. Butt, S. Heußen, M. Rispler, and M. Müller, Fault-tolerant code-switching protocols for near-term quantum processors, PRX Quantum **5**, 020345 (2024).
[49] F. Butt, D. F. Locher, K. Brechtelsbauer, H. P. Büchler, and M. Müller, Measurement-free, scalable, and fault-tolerant universal quantum computing, Science Advances **11**, eadv2590 (2025).
[50] T. J. Yoder, Universal fault-tolerant quantum computation with Bacon-Shor codes (2017), arXiv:1705.01686.
[51] S. Veroni, A. Paler, and G. Giudice, Universal quantum computation via scalable measurement-free error correction, PRX Quantum **6**, 040337 (2025).
[52] Z.-H. Chen, M.-C. Chen, C.-Y. Lu, and J.-W. Pan, Transversal logical clifford gates on the rotated surface code with reconfigurable neutral atom arrays, Phys. Rev. Lett. , (2026).
[53] W. Dür, M. Hein, J. I. Cirac, and H.-J. Briegel, Standard forms of noisy quantum operations via depolarization, Phys. Rev. A **72**, 052326 (2005).
[54] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh, D. G. Cory, and R. Laflamme, Symmetrized characterization of noisy quantum processes, Science **317**, 1893 (2007).
[55] 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, Reviews of Modern Physics **95**, 045005 (2023).
[56] S. J. Evered, D. Bluvstein, M. Kalinowski, S. Ebadi, T. Manovitz, H. Zhou, S. H. Li, A. A. Geim, T. T. Wang, N. Maskara, *et al.*, High-fidelity parallel entangling gates on a neutral-atom quantum computer, Nature **622**, 268 (2023).
[57] R. B.-S. Tsai, X. Sun, A. L. Shaw, R. Finkelstein, and M. Endres, Benchmarking and fidelity response theory of high-fidelity rydberg entangling gates, PRX Quantum **6**, 010331 (2025).
[58] J. A. Muniz, M. Stone, D. T. Stack, M. Jaffe, J. M. Kindem, L. Wadleigh, E. Zalys-Geller, X. Zhang, C.-A. Chen, M. A. Norcia, *et al.*, High-fidelity universal gates in the ^{171}Yb ground-state nuclear-spin qubit, PRX Quantum **6**, 020334 (2025).
[59] M. Peper, Y. Li, D. Y. Knapp, M. Bileska, S. Ma, G. Liu, P. Peng, B. Zhang, S. P. Horvath, A. P. Burgers, *et al.*, Spectroscopy and modeling of ^{171}Yb rydberg states for high-fidelity two-qubit gates, Phys. Rev. X **15**, 011009 (2025).
[60] A. Senoo, A. Baumgärtner, J. W. Lis, G. M. Vaidya, Z. Zeng, G. Giudici, H. Pichler, and A. M. Kaufman, High-fidelity entanglement and coherent multi-qubit mapping in an atom array, Nature Physics **22**, 903

(2026).
[61] A. Radnaev, W. Chung, D. Cole, D. Mason, T. Ballance, M. Bedalov, D. Belknap, M. Berman, M. Blakely, I. Bloomfield, *et al.*, Universal neutral-atom quantum computer with individual optical addressing and nondestructive readout, PRX Quantum **6**, 030334 (2025).
[62] S. Ma, G. Liu, P. Peng, B. Zhang, S. Jandura, J. Claes, A. P. Burgers, G. Pupillo, S. Puri, and J. D. Thompson, High-fidelity gates and mid-circuit erasure conversion in an atomic qubit, Nature **622**, 279 (2023).
[63] K. Singh, C. E. Bradley, S. Anand, V. Ramesh, R. White, and H. Bernien, Mid-circuit correction of correlated phase errors using an array of spectator qubits, Science **380**, 1265 (2023).
[64] P. Scholl, A. L. Shaw, R. B.-S. Tsai, R. Finkelstein, J. Choi, and M. Endres, Erasure conversion in a high-fidelity rydberg quantum simulator, Nature **622**, 273 (2023).
[65] M. A. Norcia, W. B. Cairncross, K. Barnes, P. Battaglino, A. Brown, M. O. Brown, K. Cassella, C.-A. Chen, R. Coxe, D. Crow, *et al.*, Midcircuit qubit measurement and rearrangement in a ^{171}Yb atomic array, Phys. Rev. X **13**, 041034 (2023).
[66] J. A. Muniz, D. Crow, H. Kim, J. M. Kindem, W. B. Cairncross, A. Ryou, T. C. Bohdanowicz, C.-A. Chen, Y. Ji, A. M. W. Jones, *et al.*, Repeated ancilla reuse for logical computation on a neutral atom quantum computer, Phys. Rev. X **15**, 041040 (2025).
[67] L. Pause, L. Sturm, M. Mittenbühler, S. Amann, T. Preuschoff, D. Schäffner, M. Schlosser, and G. Birkl, Supercharged two-dimensional tweezer array with more than 1000 atomic qubits, Optica **11**, 222 (2024).
[68] H. J. Manetsch, G. Nomura, E. Bataille, K. H. Leung, X. Lv, and M. Endres, A tweezer array with 6,100 highly coherent atomic qubits, Nature **647**, 60 (2025).
[69] N.-C. Chiu, E. C. Trapp, J. Guo, M. H. Abobeih, L. M. Stewart, S. Hollerith, P. L. Stroganov, M. Kalinowski, A. A. Geim, S. J. Evered, *et al.*, Continuous operation of a coherent 3,000-qubit system, Nature **646**, 1075 (2025).
[70] K. Singh, S. Anand, A. Pocklington, J. T. Kemp, and H. Bernien, Dual-element, two-dimensional atom array with continuous-mode operation, Phys. Rev. X **12**, 011040 (2022).
[71] M. A. Norcia, H. Kim, W. B. Cairncross, M. Stone, A. Ryou, M. Jaffe, M. O. Brown, K. Barnes, P. Battaglino, T. C. Bohdanowicz, *et al.*, Iterative assembly of ^{171}Yb atom arrays with cavity-enhanced optical lattices, PRX Quantum **5**, 030316 (2024).
[72] F. Gyger, M. Ammenwerth, R. Tao, H. Timme, S. Snigirev, I. Bloch, and J. Zeiher, Continuous operation of large-scale atom arrays in optical lattices, Phys. Rev. Res. **6**, 033104 (2024).
[73] B. Zhang, G. Liu, G. Bornet, S. P. Horvath, P. Peng, S. Ma, S. Huang, S. Puri, and J. D. Thompson, Logical qubits with erasure conversion using metastable neutral atoms, Nature Physics **22**, 910 (2026).
[74] D. Jaksch, J. I. Cirac, P. Zoller, S. L. Rolston, R. Côt'e, and M. D. Lukin, Fast quantum gates for neutral atoms, Phys. Rev. Lett. **85**, 2208 (2000).
[75] H. Levine, A. Keesling, G. Semeghini, A. Omran, T. T. Wang, S. Ebadi, H. Bernien, M. Greiner, V. Vuleti'c, H. Pichler, *et al.*, Parallel implementation of high-fidelity multiqubit gates with neutral atoms, Phys. Rev. Lett. **123**, 170503 (2019).
[76] S. Jandura and G. Pupillo, Time-optimal two- and three-qubit gates for rydberg atoms, Quantum **6**, 712 (2022).
[77] C. Fromonteil, R. Tricarico, F. Cesa, and H. Pichler, Hamilton-jacobi-bellman equations for rydberg-blockade processes, Phys. Rev. Res. **6**, 033333 (2024).
[78] I. S. Madjarov, J. P. Covey, A. L. Shaw, J. Choi, A. Kale, A. Cooper, H. Pichler, and M. Endres, High-fidelity entanglement and detection of alkaline-earth rydberg atoms, Nature Physics **16**, 857 (2020).
[79] K. Sahay, J. Jin, J. Claes, J. D. Thompson, and S. Puri, High-threshold codes for neutral-atom qubits with biased erasure errors, Phys. Rev. X **13**, 041013 (2023).
[80] I. Cong, H. Levine, A. Keesling, D. Bluvstein, S.-T. Wang, and M. D. Lukin, Hardware-efficient, fault-tolerant quantum computation with Rydberg atoms, Phys. Rev. X **12**, 021049 (2022).
[81] G. Pelegrí, A. J. Daley, and J. D. Pritchard, High-fidelity multiqubit rydberg gates via two-photon adiabatic rapid passage, Quantum Science and Technology **7**, 045020 (2022).
[82] C. Fromonteil, D. Bluvstein, and H. Pichler, Protocols for rydberg entangling gates featuring robustness against quasistatic errors, PRX Quantum **4**, 020335 (2023).
[83] S. Jandura, J. D. Thompson, and G. Pupillo, Optimizing rydberg gates for logical-qubit performance, PRX Quantum **4**, 020336 (2023).
[84] G. Giudici, S. Veroni, G. Giudice, H. Pichler, and J. Zeiher, Fast entangling gates for rydberg atoms via resonant dipole-dipole interaction, PRX Quantum **6**, 030308 (2025).
[85] S. Jandura, L. Pecorari, and G. Pupillo, Surface code stabilizer measurements for Rydberg atoms, Physical Review Research **8**, L022059 (2026).
[86] L. Pecorari, S. Jandura, and G. Pupillo, Low-depth quantum error correction via three-qubit gates in rydberg atom arrays, Phys. Rev. Lett. **135**, 240602 (2025).
[87] S. Tasler, J. Old, L. Heunisch, V. Feulner, T. Eckstein, M. Müller, and M. J. Hartmann, Optimizing superconducting three-qubit gates for surface-code error correction (2025), arXiv:2506.09028.
[88] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Optimal control of coupled spin dynamics: design of nmr pulse sequences by gradient ascent algorithms, Journal of Magnetic Resonance **172**, 296 (2005).
[89] A. Garon, S. J. Glaser, and D. Sugny, Time-optimal control of SU(2) quantum operations, Phys. Rev. A **88**, 043422 (2013).
[90] B. E. Anderson, H. Sosa-Martinez, C. A. Riofrío, I. H. Deutsch, and P. S. Jessen, Accurate and Robust Unitary Transformations of a High-Dimensional Quantum System, Phys. Rev. Lett. **114**, 240401 (2015).
[91] B. Riaz, C. Shuang, and S. Qamar, Optimal control methods for quantum gate preparation: a comparative study, Quantum Information Processing **18**, 100 (2019).
[92] Y. Wu, B. Li, K. Chang, S. Puri, and L. Zhong, Minimum-weight parity factor decoder for quantum error correction (2025), arXiv:2508.04969.
[93] R. Finkelstein, R. B.-S. Tsai, X. Sun, P. Scholl, S. Direkci, T. Gefen, J. Choi, A. L. Shaw, and M. Endres, Universal quantum operations and ancilla-based read-out for tweezer clocks, Nature **634**, 321 (2024).
[94] Y. Wu, S. Kolkowitz, S. Puri, and J. D. Thompson, Erasure conversion for fault-tolerant quantum computing

in alkaline earth rydberg atom arrays, Nat. Commun. **13**, 4657 (2022).
[95] S. Veroni, M. Müller, and G. Giudice, Optimized measurement-free and fault-tolerant quantum error correction for neutral atoms, Phys. Rev. Res. **6**, 043253 (2024).
[96] Y. R. Sanders, J. J. Wallman, and B. C. Sanders, Bounding quantum gate error rate based on reported average fidelity, New Journal of Physics **18**, 012002 (2016).
[97] E. Huang, A. C. Doherty, and S. Flammia, Performance of quantum error correction with coherent errors, Phys. Rev. A **99**, 022313 (2019).
[98] R. Kueng, D. M. Long, A. C. Doherty, and S. T. Flammia, Comparing experiments to the fault-tolerance threshold, Phys. Rev. Lett. **117**, 170502 (2016).
[99] M. Kliesch and I. Roth, Theory of quantum system certification, PRX Quantum **2**, 010201 (2021).
[100] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error rate, SIAM Journal on Computing **38**, 1207 (2008).
[101] P. Aliferis, D. Gottesman, and J. Preskill, Quantum accuracy threshold for concatenated distance-3 codes (2005), arXiv:quant-ph/0504218.
[102] C. Gidney, Stim: a fast stabilizer circuit simulator, Quantum **5**, 497 (2021).
[103] R. Toshio, S. Kanasugi, J. Fujisaki, H. Oshima, S. Sato, and K. Fujii, Star-magic mutation: Even more efficient analog rotation gates for early fault-tolerant quantum computer (2026), arXiv:2603.22891.
[104] C. Gidney, Halving the cost of quantum addition, Quantum **2**, 74 (2018).
[105] I. D. Kivlichan, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, W. Sun, Z. Jiang, N. Rubin, A. Fowler, A. Aspuru-Guzik, *et al.*, Improved fault-tolerant quantum simulation of condensed-phase correlated electrons via Trotterization, Quantum **4**, 296 (2020).
[106] D. P. Arovas, E. Berg, S. A. Kivelson, and S. Raghu, The hubbard model, Annual Review of Condensed Matter Physics **13**, 239 (2022).
[107] E. T. Campbell, Early fault-tolerant simulations of the Hubbard model, Quantum Science and Technology **7**, 015007 (2022).
[108] 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).
[109] Z. Jiang, K. J. Sung, K. Kechedzhi, V. N. Smelyanskiy, and S. Boixo, Quantum algorithms to simulate many-body physics of correlated fermions, Phys. Rev. Appl. **9**, 044036 (2018).
[110] S. Stanisic, J. L. Bosse, F. M. Gambetta, R. A. Santos, W. Mruczkiewicz, T. E. O'Brien, E. Ostby, and A. Montanaro, Observing ground-state properties of the Fermi-Hubbard model using a scalable algorithm on a quantum computer, Nature Communications **13**, 5743 (2022).
[111] A. J. Bay-Smidt, F. R. Klausen, C. Sünderhauf, R. Izsák, G. C. Solomon, and N. S. Blunt, Fault-tolerant quantum simulation of generalized Hubbard models, PRX Quantum **6**, 030318 (2025).
[112] A. Kan and B. C. B. Symons, Resource-optimized fault-tolerant simulation of the Fermi-Hubbard model and high-temperature superconductor models, npj Quantum Information **11**, 138 (2025).
[113] Y. Akahoshi, R. Toshio, J. Fujisaki, H. Oshima, S. Sato, and K. Fujii, Compilation of trotter-based time evolution for partially fault-tolerant quantum computing architecture, PRX Quantum **6**, 040319 (2025).
[114] K. J. Sung, I. Choi, M. Amico, B. Andrews, E. Ayantuna, Y. Kawashima, W.-H. Lin, D. Omanovic, S. Piccinelli, J. R. Moreno, A. A. Saki, J. Shee, S. Shin, M. C. Tran, K. Ueda, H. Zhang, and M. Motta, ffsim: Faster simulation of fermionic quantum circuits (2026), arXiv:2605.03123.
[115] M. D'Anna, J. Nys, and J. Carrasquilla, Majorana string simulation of nonequilibrium dynamics in two-dimensional lattice fermion systems, Phys. Rev. B **114**, 034305 (2026).
[116] A. P. Thompson, A. Soeteman, C. Cade, and I. Niesen, Non-zero noise extrapolation: accurately simulating noisy quantum circuits with tensor networks (2025), arXiv:2501.13237.
[117] F. Alam, J. L. Bosse, I. Čepaitė, A. Chapman, L. Clinton, M. Crichigno, E. Crosson, T. Cubitt, C. Derby, O. Dowinton, P. K. Faehrmann, S. Flammia, B. Flynn, F. M. Gambetta, R. García-Patrón, M. Hunter-Gordon, G. Jones, A. Khedkar, J. Klassen, M. Kreshchuk, *et al.*, Programmable digital quantum simulation of 2d fermi-hubbard dynamics using 72 superconducting qubits (2026), arXiv:2510.26845.
[118] R. Ott, T. V. Zache, S. Choi, A. M. Kaufman, and H. Pichler, Rare-event quantum sensing using logical qubits, Phys. Rev. Lett. (2026).
[119] E. Huang, P.-G. Rozon, A. Dua, S. Gopalakrishnan, and M. J. Gullans, A robust phase of continuous transversal gates in quantum stabilizer codes (2026), arXiv:2510.01319.
[120] A $Z_q^{\otimes w}$ fault occurring sufficiently far before or after the transversal rotations would be detected by the intervening stabilizer measurements.
[121] Z. Zeng, G. Giudici, A. Senoo, A. Baumgärtner, A. M. Kaufman, and H. Pichler, Adiabatic echo protocols for robust quantum many-body state preparation, Phys. Rev. Lett. **136**, 120404 (2026).
[122] S. Mangini, L. Maccone, and C. Macchiavello, Qubit noise deconvolution, EPJ Quantum Technology **9**, 29 (2022).
[123] J. O. Berger, J. M. Bernardo, and D. Sun, Overall objective priors, Bayesian Analysis **10**, 189 (2015).
[124] D. Alvares, C. Armero, and A. Forte, What does objective mean in a dirichlet-multinomial process?, International Statistical Review **86**, 106 (2018).
[125] E. O. Lebigot, Uncertainties: A python package for calculations with uncertainties, `http://pythonhosted.org/uncertainties/`, accessed: 2026-06-17.
[126] B. Eisenberg, On the expectation of the maximum of IID geometric random variables, Statistics & Probability Letters **78**, 135 (2008).
[127] E. Campbell, Shorter gate sequences for quantum computing by mixing unitaries, Physical Review A **95**, 042306 (2017).
[128] M. B. Hastings, Turning gate synthesis errors into incoherent errors, Quantum Information and Computation **17**, 488 (2017).
[129] K. Matsumoto and K. Amano, Representation of quantum circuits with clifford and $\pi/8$ gates (2008), arXiv:0806.3834.
[130] B. Giles and P. Selinger, Remarks on matsumoto and

amano's normal form for single-qubit clifford+T operators (2013), arXiv:1312.6584.

[131] N. J. Ross and P. Selinger, Optimal ancilla-free Clifford+T approximation of z-rotations, Quantum Information and Computation **16**, 901 (2016).

[132] S. Yamamoto and N. Yoshioka, `pygridsynth`: a fast numerical tool for ancilla-free clifford+T synthesis (2026), arXiv:2604.21333.

[133] D. Litinski, A Game of Surface Codes: Large-Scale Quantum Computing with Lattice Surgery, Quantum **3**, 128 (2019).

[134] E. Beyerstedt and V. H. Moll, The p-adic valuation of the ASM numbers, Journal of Integer Sequences **14** (2011).