State preparation with parallel-sequential circuits

Zhi-Yuan Wei (魏志远)[1,2,3] and Daniel Malz[4]

[1] Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Str. 1, 85748 Garching, Germany
[2] Munich Center for Quantum Science and Technology (MCQST), Schellingstr. 4, 80799 München, Germany
[3] Joint Quantum Institute and Joint Center for Quantum Information and Computer Science, NIST/University of Maryland, College Park, Maryland 20742, USA
[4]Department of Mathematical Sciences, University of Copenhagen, 2100 Copenhagen, Denmark

arXiv:2503.14645v2 [quant-ph] 15 Apr 2026

We introduce parallel-sequential (PS) circuits, a family of quantum circuit layouts that interpolate between brickwall and sequential circuits, which introduces control parameters governing a trade-off between the amount of entanglement and the maximum correlation range they can express. We provide numerical evidence that PS circuits can efficiently prepare many-body ground states in one dimension. On noisy devices, characterized through both idling errors and two-qubit gate errors, we show that in a wide parameter regime, PS circuits outperform brickwall, sequential, and the log-depth circuits from [Malz, Styliaris, Wei, Cirac, PRL 132, 040404 (2024)]. Additionally, we demonstrate that properly chosen noisy random PS circuits suppress error proliferation and, when employed as a variational ansatz, exhibit superior trainability.

1 Introduction

Quantum circuits are a unified framework to describe operations on quantum processors [1], but can also serve as a model of many-body dynamics. In this vein, different circuit layouts (arrangement of gates) correspond to distinct processes and can be used to define different variational families of states. For example, brickwall circuits naturally model evolution with spatial locality [2] and thus efficiently prepare states obtained in a quench from a product state. In contrast, sequential circuits prepare states with constant entanglement entropy, which in one dimension includes all matrix product states (MPS) [3] and in higher dimensions a rich family of states including topologically ordered ones [4–8]. The multi-scale entanglement renormalization ansatz (MERA) [9] can model critical states and can be interpreted as arising from holographic dynamics.

The advantage of using a variational family that is equivalent to a quantum circuit applied to a product state is that one can directly prepare states in that family on quantum devices. This approach is thus widely used to prepare quantum states represented as tensor network states by converting (or compiling) them into a circuit [10–14], but also in variational quantum algorithms that employ a quantum computer to optimize with a family defined through a circuit layout (usually brickwall circuits) [15–17].

When preparing states on noisy devices, the circuit that achieves the highest fidelity in the noiseless setting (has the lowest *approximation error*) may no longer be optimal. *Idling errors* accumulate with circuit depth and *gate errors* with the number of gates. The optimal circuit must balance these three sources of error. This depends on the architecture considered. For concreteness, we mainly consider one-dimensional devices with spatially local gates, but in section 5 we also introduce a higher-dimensional generalization. In the one-dimensional setting, sequential circuits and MERA have a depth (in terms of local gates) linear in system size N and are therefore very susceptible to idling error. Brickwall circuits are the densest possible arrangement of spatially local gates and thus achieve the lowest depth for any task, but may use too many gates. If the goal is to prepare gapped ground states, which have constant (in system size) entanglement entropy [18, 19], as well as exponentially decaying correlations [20], one might heuristically expect that the optimal circuit has logarithmic depth, but a constant number of gates crossing any particular bipartition of the chain, which is fulfilled by none of the aforementioned layouts. While the *RG circuits* in Ref. [21] can be used to prepare gapped ground states in terms of MPS with such an asymptotical scaling, they incur significant overhead when compiled into two-qubit gates [cf. section B.2].

In this work, we introduce *parallel-sequential* (PS) circuits, designed to manage this trade-off optimally. PS circuits interpolate between brickwall and sequential circuits, with tunable degree of entanglement and correlation range. In contrast to sequential circuits, PS circuits are shallow, which suppresses idling errors, and compared to brickwall circuits they use fewer gates, which mitigates gate errors. Focusing on one dimension (see section 5 for higher-dimensional gen-

Zhi-Yuan Wei (魏志远): zywei@umd.edu
Daniel Malz: daniel.malz@unibas.ch, Current affiliation: University of Basel, Klingelbergstrasse 82, CH-4056

eralizations), we find that PS circuits are sufficiently expressive, more noise robust than other circuits, and are also easier to find variationally. Specifically, we first show numerically that in the absence of noise, single-layer PS circuits can represent short-range correlated MPS of bond dimension $D = 2$ with a small infidelity ε using a circuit depth $T \sim \log(N/\varepsilon)$ that is substantially shallower than the state-of-the-art RG circuits [21]. Then we provide evidence that the PS circuits outperform the brickwall and sequential circuits for finding ground state of the XY model in noisy devices. To further support our findings, we analyse random PS circuits and show that those layouts with a constant number of layers have better trainability and less error proliferation compared to brickwall circuits of the same depth and sequential circuits.

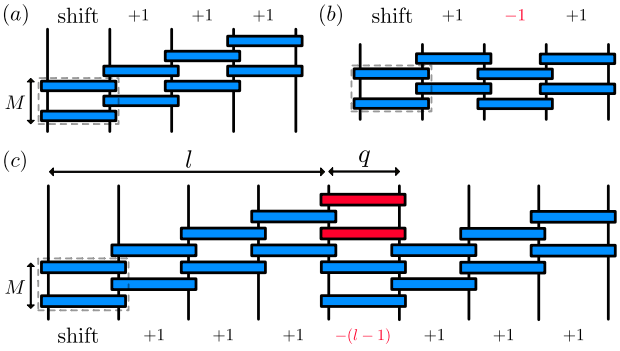


Figure 1: **(a) Sequential, (b) Brickwall, and (c) Parallel-Sequential (PS) Circuits.** The sequential and brickwall circuits both apply an identical stack of M gates (dashed boxes) to each pair of neighboring qubits. The distinction lies in the shift pattern: in sequential circuits, each successive stack is shifted upward by one position ('+1') relative to the previous one, whereas in brickwall circuits, the stacks alternate between being shifted up ('+1') and down ('−1'). In PS circuits (panel c), the stack is shifted upward l times—forming a sequential segment—followed by a single downward shift of $l-1$ steps, creating a break. To mitigate this break, we allow sequential segments to overlap, such that adjacent segments share a region of q gates (highlighted in red; $q = 2$ in the figure).

2 Definition

Before we introduce 1D parallel-sequential circuits, we note that M-layer sequential circuits [cf. Fig. 1(a)] and depth-$2M$ brickwall circuits [cf. Fig. 1(b)] can both be viewed as placing the same stack of M gates on each pair of neighbouring qubits, but with the difference that in sequential circuits, each subsequent stack is shifted up relative to its predecessor, whereas in brickwall circuits, they are alternately shifted up and down. In both cases, the entanglement entropy across any cut is bounded by $S \leq 2M$, but the correlations in a brickwall circuit extend only over $4M$ sites, whereas in a sequential circuit, they extend over the entire system.

In PS circuits [cf. Fig. 1(c)], we take a mixed approach: the stack is shifted up l times (generating a sequential chunk) and then once down by $l-1$ (a break). To smoothen the break, we allow the sequential chunks to extend, such that neighboring chunks overlap over a distance q (red gates, $q = 2$ in the figure). This defines PS circuits with $n_C = \lceil N/l \rceil$ chunks and circuit depth $T = l + q + 2M - 3$. In the rest of this article, we understand brickwall and sequential circuits as opposite limits of PS circuits, where M-layer PS circuits with $l = 2, q = 1$, coincide with depth-$2M$ brickwall circuits, while those with $l = N - 1, q = 1$ are M-layer sequential circuits.

With gate parameters $\boldsymbol{\theta}$ and layout hyperparameters N, M, l, q, PS circuits can be expressed as a unitary $U_{\mathrm{PS}}(\boldsymbol{\theta}, N, M, l, q)$, which also defines the associated variational family

$$|\Psi_{\mathrm{PS}}(\boldsymbol{\theta}, N, M, l, q)\rangle = U_{\mathrm{PS}}(\boldsymbol{\theta}, N, M, l, q)|0\rangle^{\otimes N}. \quad (1)$$

The entanglement entropy of $|\Psi_{\mathrm{PS}}\rangle$ across any non-overlapping cut is bounded by $S \leq 2M$, and the maximal correlation range can be tuned via M, l, and q [cf. section B.1.2]. By choosing appropriate scaling of parameters, PS circuits can create states with diverse entanglement and correlation scalings. For instance, by choosing constant M and $l, q \sim \log N$, PS circuits produce states with the same entanglement and correlation scaling as gapped 1D ground states.

3 PS circuits for quantum state preparation

MPS (with open boundary conditions) of a chain of N qubits and with bond dimension D are defined as

$$|\phi_{\mathrm{MPS}}\rangle = \sum_{i_1,\ldots,i_N=0}^{1} A_1^{i_1} \ldots A_N^{i_N} |i_1 \ldots i_N\rangle, \quad (2)$$

where $\{A_n^i\}$ in the bulk are matrices of dimension $D \times D$. Typical MPS exhibit short-range correlations [22–24] and correspond to the ground states of local gapped Hamiltonians [25]. Please see a more detailed review of MPS in section A.

3.1 Preparing MPS of $D = 2$ with single-layer PS circuits

It is known that an MPS with bond dimension $D = 2$ can be prepared by a single-layer ($M = 1$) sequential circuit acting on a product state of qubits [3], with a circuit depth that scales linearly with the system size N. In the following, we show that single-layer PS circuits can approximate short-range correlated MPS with $D = 2$ using exponentially smaller depth. For simplicity, we consider MPS with identical tensors $A_i = A$, which we refer to as bulk translation-invariant (bulk-TI) MPS and denote as

$|\phi^{\text{bulk-TI}}_{\text{MPS}}\rangle$ [cf. section A]. We determine the optimal approximation $|\Psi_{\text{PS}}\rangle$ within the single-layer PS circuits to the target state $|\phi^{\text{bulk-TI}}_{\text{MPS}}\rangle$ by locally optimizing the gates in the overlap region. Each overlap region introduces an error, such that the state fidelity $\mathcal{F} \equiv |\langle\Psi_{\text{PS}}|\phi^{\text{bulk-TI}}_{\text{MPS}}\rangle|^2$ decays exponentially with the number of chunks n_C as [cf. section B.1.2]

$$\mathcal{F}(n_C, q) \approx \exp\big[-\kappa(q)\cdot(n_C-1)\big]. \tag{3}$$

We numerically establish that the error density $\kappa(q)$ exhibits a generic scaling [cf. Fig. 2(a)]

$$\kappa(q) \sim \exp\big(-\gamma\cdot q/\xi\big). \tag{4}$$

We test this for a family of bulk-TI MPS with tunable correlation length ξ [26] (the constant factor $\gamma \approx 2$ for this family), as well as random bulk-TI MPS with $D = 2$ [cf. section A]. This behavior indicates that the infidelity originates from the absence of correlations in PS circuits near the location where the stack shifts down [cf. Fig. 1(c)] [cf. section B.1.2]. Notably, the exponential decay of $\kappa(q)$ quantitatively matches the scaling observed in the RG circuit [21], suggesting that the PS circuit is asymptotically optimal.

The scaling of $\mathcal{F}(n_C, q)$ and $\kappa(q)$ leads to a favorable circuit depth scaling, given by

$$T \sim \xi\cdot\log\big(N/\epsilon\big), \tag{5}$$

for preparing $|\phi^{\text{bulk-TI}}_{\text{MPS}}\rangle$ of size N and bond dimension $D = 2$ with a small infidelity ϵ using the single-layer PS circuit. From this, we estimate a lower bound on the CNOT gate depth T_{CNOT} after compiling the circuit [cf. Fig. 2(b)], and compare it with that from the RG circuits [21]. Notably, PS circuits consistently achieve a lower T_{CNOT} than RG circuits, as the latter involve gates acting on $1 + \lceil 2\log_2 D\rceil$ qubits for preparing an MPS of bond dimension D. This results in a compilation overhead for RG circuits that increases T_{CNOT}, which becomes more significant when preparing MPS with higher bond dimensions [cf. section B.2].

3.2 Preparation of states with more entanglement

By increasing the number of layers M, PS circuits (and the two limiting cases, brickwall and sequential circuits) can represent quantum states with higher entanglement. Multi-layer structures have been widely explored in the case of brickwall and sequential circuits and generally have been very effective [11, 12, 14, 17, 27–34]. In particular, the task of preparing short-range correlated states with higher entanglement can be decomposed into a sequence of preparations of MPS with bond dimension $D = 2$ [11]. Therefore, our results for $D = 2$ MPS [cf. Fig. 2] may suggest that multi-layer PS circuits are likewise effective for representing short-range correlated states

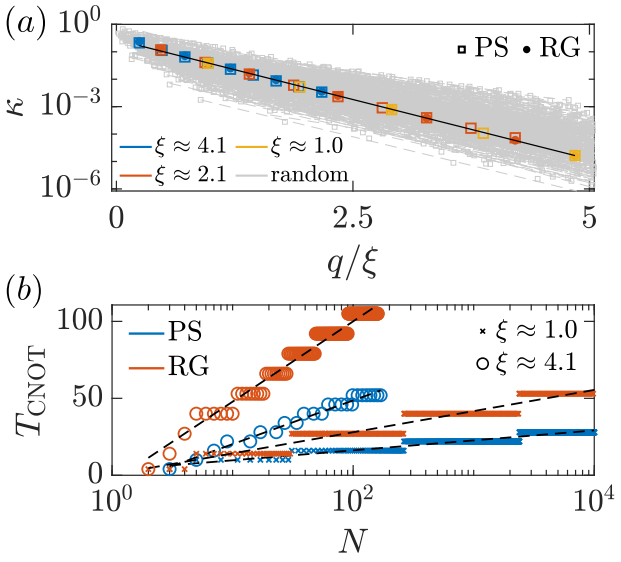


Figure 2: **Single-layer PS circuits for MPS preparation.** (a) The error density κ as a function of q/ξ, for the preparation of a family of bulk-TI MPS with various correlation lengths ξ [26], as well as for 500 random bulk-TI MPS with $D = 2$, using PS circuits (hollow squares) and the RG circuit [21] (dots). The lines are exponential fits to the corresponding data from the same MPS. (b) Lower bound of the CNOT gate depth T_{CNOT} for the PS circuit and the RG circuit for preparing the MPS family of size N with fidelity $\mathcal{F} = 0.95$. The dashed lines are log fits of the data.

with higher entanglement. To demonstrate this more concretely, we investigate the ground-state preparation of the XY model as an illustrative example in the following. Since the XY model is gapless and exhibits power-law decaying correlations, preparing its ground state is expected to be more challenging than for gapped models. This model has been used in several studies [11, 35] as a benchmark for ground-state preparation using multi-layer sequential circuits (a subclass of the PS circuits studied here).

The Hamiltonian of the XY model is

$$H_{XY} = \sum_{i=1}^{N-1} \sigma_i^x\sigma_{i+1}^x + \sigma_i^y\sigma_{i+1}^y, \tag{6}$$

where $\sigma_i^{x,y,z}$ are the Pauli matrices at site i. We quantify the state-preparation performance using the energy density above the ground state, defined as

$$\nu_{\text{XY}} \equiv (E - E_{\text{GS}})/N, \tag{7}$$

where E_{GS} is the exact ground state energy. We employ a DMRG-like circuit optimization algorithm [36] to find the optimal PS state $|\Psi_{\text{PS}}\rangle$ that minimizes the energy $E = \langle\Psi_{\text{PS}}|H_{XY}|\Psi_{\text{PS}}\rangle$.

In Fig. 3(a), we plot the energy density ν_{XY} along lines of constant $M \in \{1, 2\}$ and for $q = 1$, interpolating between brickwall and sequential circuit. As we increase the length of the sequential chunks l, the approximation error monotonically decreases, because

the circuit is better able to capture longer ranged correlations. Increasing M from 1 (blue) to 2 (purple) shifts the interpolation curve down uniformly, as the circuits can approximate more of the entanglement spectrum. This observation suggests decomposing the approximation error into two distinct contributions, expressed as

$$\nu_{\mathrm{XY}} \approx \nu_{\mathrm{ent}} + \nu_{\mathrm{cor}}, \tag{8}$$

arising respectively from the error due to lacking enough entanglement (ν_{ent}) and the correlation range (ν_{cor}), respectively. Importantly, the correlation error ν_{cor} drops off rapidly in l, q, allowing the PS circuit to achieve an energy density ν_{XY} comparable to the sequential circuit with the same number of layers M, while requiring a significantly smaller circuit depth [cf. Fig. 3(a)].

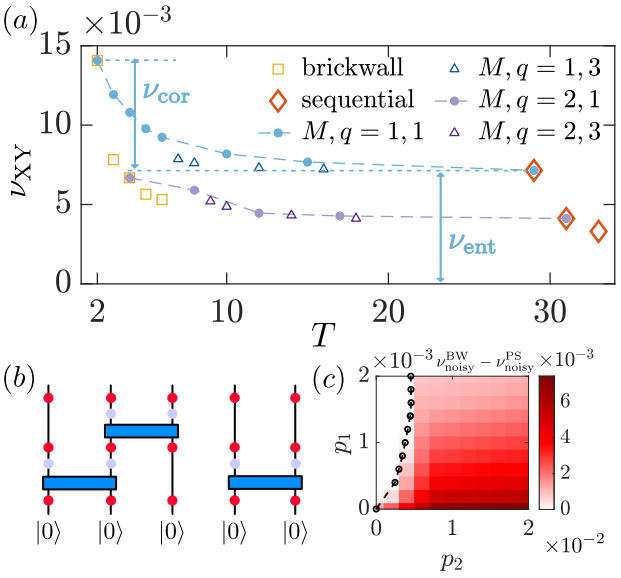


Figure 3: **Preparing the ground state of the XY model.** (a) The ground-state energy density ν_{XY} of the XY model for brickwall, sequential, and various PS circuits for $M \in \{1, 2\}$ (dots connected by the same line indicate tuning l). We visually indicate the correlation error ν_{cor} and entanglement error ν_{ent} for the $M = 1$ case. (b) Illustration of the noise model, where red (purple) dots represent idling errors (two-qubit gate errors) applied with probabilities p_1 (p_2). (c) The difference between the optimal energy densities [Eq. (A26)] obtained from noisy brickwall circuits, denoted as $\nu^{\mathrm{BW}}_{\mathrm{noisy}}$, and those obtained from PS circuits (which include brickwall circuits), denoted as $\nu^{\mathrm{PS}}_{\mathrm{noisy}}$, for various values of p_1 and p_2. The black dots represent the "phase boundary" where $\nu^{\mathrm{BW}}_{\mathrm{noisy}} = \nu^{\mathrm{PS}}_{\mathrm{noisy}}$. The system size for all plots is $N = 30$.

4 PS circuits on noisy devices

The low depth and sparsity of PS circuits combined with their competitive approximation error suggest that they offer an advantage in the presence of both idling and gate error. To investigate this, we consider an incoherent error model, illustrated in Fig. 3(b), since coherent errors can often be converted into incoherent ones using randomized compiling techniques [37]. In this model, single-qubit depolarizing errors act on the entire system at each time step with rate p_1, representing idling errors, Additionally, each two-qubit gate is followed by depolarizing errors acting simultaneously on the involved qubits at a rate p_2, representing two-qubit gate errors.

4.1 Energy error scaling

We simulate the the ground-state preparation circuits obtained in Fig. 3(a) in the presence of noise and compute the resulting energy $E_{\mathrm{noisy}} = \mathrm{Tr}(\rho_{\mathrm{PS}} H_{XY})$ and the energy density $\nu_{\mathrm{noisy}} \equiv (E_{\mathrm{noisy}} - E_{\mathrm{GS}})/N$. For dilute noise (given by $p_1 T$ and $2p_2 M$), each error event perturbs only a local patch of the state. First-order perturbation theory then implies that their contributions to any extensive observable (e.g. the energy E_{noisy}) add up linearly. Consequently, the energy density error scales as

$$\nu_{\mathrm{noisy}} \approx \nu_{\mathrm{XY}} + c_E(p_1 T + 2p_2 M), \tag{9}$$

which we confirm numerically in section C.1. Here $p_1 T$ (resp. $2p_2 M$) counts the total density of idling errors (two-qubit gate errors), with a prefactor $c_E \approx 0.7$ for the XY model. For any $p_1 > 0$, the error density in sequential circuits grows linearly in system size $\nu^{\mathrm{SEQ}}_{\mathrm{noisy}} \sim c_1 p_1 N$, which make them unsuitable for large systems. Therefore, we focus on comparing brickwall and the new layouts within PS circuits in the following analysis.

Given the noise rates p_1 and p_2, we ask whether there exists a layout within the PS circuits that outperforms brickwall circuits. In the regime in which $p_2 \lesssim p_1$, this generally is not the case, as idling is just as detrimental as performing a gate. The more interesting setting is when the gate error rate is substantially larger than the idling error rate ($p_2 \gg p_1$), which is also the experimentally relevant regime [38–40]. In this regime, slightly increasing the circuit depth to capture longer-range correlations is inexpensive, provided the total number of gates remains unchanged. This can be precisely achieved in PS circuits by using longer sequential segments (increasing l). This analysis is numerically verified in Fig. 3(c), where we evaluate ν_{noisy} for all circuits examined in Fig. 3(a). Specifically, we compare the minimal energy density obtained from all depths of brickwall circuits (denoted as $\nu^{\mathrm{BW}}_{\mathrm{noisy}}$) to that obtained from PS circuits (denoted as $\nu^{\mathrm{PS}}_{\mathrm{noisy}}$). Over a broad range of parameters $p_2 \gtrsim p_1 > 0$, optimized PS circuits consistently yield lower energy densities compared to brickwall circuits. This advantage diminishes as p_1 approaches p_2, delineating a kind of "phase boundary". Note that, since brickwall circuit layouts are contained within the PS circuit layouts, in the regime $p_2 \lesssim p_1$ [the white region in Fig. 3(c)], the optimal PS

circuit reduces to the brickwall circuit. Consequently, $\nu^{\mathrm{BW}}_{\mathrm{noisy}} - \nu^{\mathrm{PS}}_{\mathrm{noisy}} = 0$ in that region.

These results show a superior performance of PS circuits when preparing one dimensional ground states on noisy devices in a broad and relevant parameter regime. In the following, we show that their optimized layout also translates to better trainability and lower error propagation.

4.2 Gradient variance scaling

To explore the trainability of noisy PS circuits under gradient-based optimization, we parameterize the circuit gates with parameters $\boldsymbol{\theta}$, such that the output density matrix is $\rho_{\mathrm{PS}}(\boldsymbol{\theta})$ [cf. section C.2]. Choosing E_{noisy} as the cost function, we investigate the scaling of the gradient variance,

$$V_E(M,T,p_1,p_2) \equiv \langle(\partial_{\theta_j} E_{\mathrm{noisy}})^2\rangle_{\boldsymbol{\theta}} - \langle\partial_{\theta_j} E_{\mathrm{noisy}}\rangle^2_{\boldsymbol{\theta}}, \tag{10}$$

where the averaging is performed over random gate realizations and all variational parameters $\boldsymbol{\theta}$.

As H_{XY} is a sum of local observables, it has been shown that $V_E^{\mathrm{ideal}} \equiv V_E^{\mathrm{noisy}}(M,T,0,0) \sim e^{-\alpha M}$ for noiseless circuits [41, 42], and $V_E^{\mathrm{noisy}}(M,T,p_1,0) \sim e^{-\alpha M - c_V p_1 T}$ for circuits under idling depolarizing noise [43], where α and c_V are constant factors determined by the cost function. Building on these results, we expect that [cf. Eq. (11)]

$$V_E^{\mathrm{noisy}}(M,T,p_1,p_2) \sim e^{-\alpha M - c_V(p_1 T + 2p_2 M)}, \tag{11}$$

which we numerically verify for PS circuits (including brickwall and sequential), with the cost-function dependent prefactors $\alpha \approx 0.6$ and $c_V \approx 5.5$ [cf. section C.2].

One consequence of this scaling is the noise-induced barren plateaus for sequential circuits [43], where $V_E^{\mathrm{SEQ}} \sim e^{-c_V p_1 N}$ decays exponentially with N. In contrast, PS circuits (including brickwall circuits) with depth $T = O(\log N)$ exhibit at most a polynomial decay of V_E with system size N, although for a large number of layers M, the variance can still be small. PS circuits achieve a similar precision to brickwall circuits when approximating short-range correlated states but using fewer layers M [cf. examples in Fig. 2 and 3], and thus exhibit better trainability than brickwall circuits, essentially because the PS circuits are sparser. We provide a numerical example to illustrate the behavior of V_E in Fig. 4(a). Together with the result shown in Fig. 3(c), which demonstrates that there exist PS circuits achieving a ground-state energy density lower than that of brickwall circuits with arbitrary depth, this indicates that, on noisy devices, PS circuits admit layouts that simultaneously achieve higher accuracy and better trainability than brickwall and sequential circuits.

4.3 Error propagation

Gates are known to proliferate errors in the sense that an error in a single location can lead to errors on many qubits [1, 44]. To probe how error propagation depends on circuit layout, we consider the setting proposed in Ref. [44]. Given a circuit layout, the authors consider the family of random unitaries U obtained by taking each gate in the circuit to be Haar random. In the absence of errors, applying $U^\dagger U$ to $|0\rangle^{\otimes N}$ first entangles the qubits and then disentangles them again. If single-qubit depolarizing error occurs at a certain rate p_1 during this evolution, the system fails to return to its initial state and instead has some number of depolarized qubits. This process can be modeled as a Markov chain, where the system is represented as a string of zeros and ones: zeros denote noiseless qubits, and ones denote depolarized qubits [cf. section C.3]. Using Markov-chain Monte Carlo, the average number of depolarized qubits, $\langle\eta\rangle$ (equivalently, the average number of ones in the Markov chain), can be computed at the end of the circuit evolution [44], as shown in Fig. 4(b). The results in this subsection further demonstrate the noise robustness of PS circuits, albeit in a generic average case, where we can make analytical progress.

We now investigate how $\langle\eta\rangle$ depends on circuit layout in the limit of low error densities $p_1 T \ll 1$. If no gates are applied, the system idles for time $2T$, and thus the depolarizing dynamics leads to $\langle\eta\rangle/N = 2p_1 T + O((p_1 T)^2)$. Adding in gates, it can be shown that their effect when acting on a clean and a depolarized qubit amounts to depolarizing both (propagating the error) with probability 4/5 or removing the error with probability 1/5 [44]. In one-dimensional random brickwall circuits an error will thus propagate to a number of qubits that scales linearly in T (for $T < N$), and thus $\langle\eta\rangle/N \propto p_1 T^2$ [44], with the quadratic scaling $\propto T^2$ indicating strong error propagation.

Before we extend the analysis to general PS circuits, we first consider sequential circuits with M layers (the opposite limit of brickwall circuits), with $M \ll N$. A simple argument shows that a single layer propagates an error on average over 3 additional sites, and M layers produce on average a string of error of length $L(M)$, with $L(1) = 4$, and $L(2) = 6.4$ and asymptotically, $L(M) \to 2 + 9M/4$ [cf. section C.3]. Neglecting the effect of boundaries, overlapping errors, and errors occuring on a qubit between the first and last unitary applied to it, we thus obtain (note that $T = N$ here) $\langle\eta\rangle/N = p_1 T(1 + L(M))$ [cf. section C.3].

In general PS circuits, we numerically find that $\langle\eta\rangle$ interpolates between the brickwall and sequential limits and find that it is well approximated as

$$\langle\eta\rangle/N \approx p_1 T(c_{\eta,1} + c_{\eta,2} M), \tag{12}$$

with prefactors $c_{\eta,1}, c_{\eta,2}$ of order unity and exhibit a mild dependence on the circuit layout (as well as T

and M), but only very weakly on system size [cf. section C.3]. Importantly, we always find the dominant contribution (in the limit of sparse errors) to grow with $p_1 TM$ (see Fig. 4(b) and section C.3). Thus, in PS circuits with constant M, the errors are just *linear* in T, as compared to the T^2 scaling in brickwall circuits.

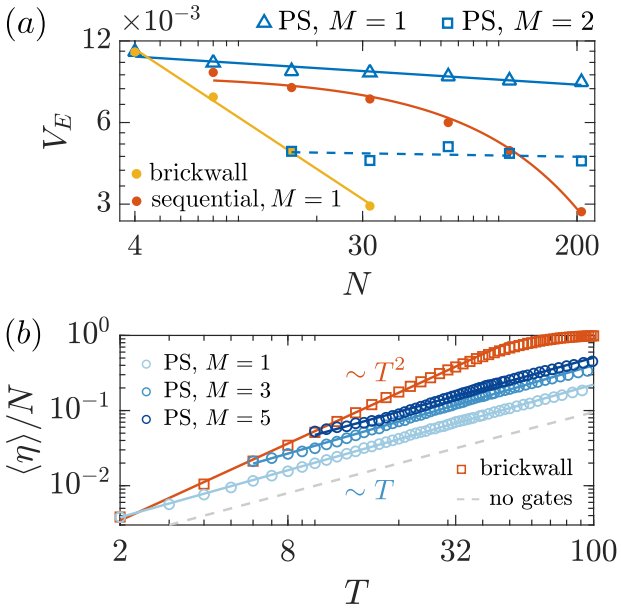


Figure 4: **Noise robustness of random PS circuits.** (a) The scaling of gradient variance V_E with system size N for sequential circuit ($M = 1$), brickwall of depth $T = \log_2 N$, and PS circuits with the same depth, but fixed $M = 1, 2, q = 1$. The noise rates are $p_1 = 10^{-3}$ and $p_2 = 10^{-2}$. The lines are fits to the data. (b) The fraction of depolarized qubits, $\langle\eta\rangle/N$, as a function of circuit depth T for brickwall circuits, PS circuits with fixed M and $q = 1$, and purely depolarizing dynamics with no gates. The system parameters are $N = 1000$, $p_1 = 5 \times 10^{-4}$, and $p_2 = 0$. The solid lines are fits to the data in the regime of $\langle\eta\rangle/N \ll 1$.

5 Parallel-sequential circuits in higher dimensions

In higher-dimensional lattices, various types of sequential and brickwall circuits have been extensively studied. These include applications such as the preparation of long-range entangled states using sequential circuits [4, 6–8] and the demonstration of quantum computational advantage with deep brickwall circuits [45–48]. Extending the one-dimensional lattice analysis, one can define PS circuits in arbitrary higher dimensions that interpolate between higher-dimensional sequential circuits and brickwall circuits. Due to the freedom of defining the sequential circuits and the brickwall circuits, such an interpolation is not unique. For concreteness, here we consider the multi-layer sequentially-generated states (SGS) [4] [1] as the representative sequential circuit and a particular ordering of applying the commuting brickwall layers, which will be detailed later. In the following, we define PS circuits on the two-dimensional square lattice, while the generalization to arbitrary-dimensional cubic lattices is straightforward.

Consider a two-dimensional square lattice of size $N = L \times L$ [c.f. Fig. 5]. We define the PS *unitary* acting on the j-th row of the 2D lattice as $U^H_{\mathrm{PS}}(j, \boldsymbol{\theta}, L, M, l, q)$, where the gate layout within the unitary follows that of the 1D PS circuit. Here $\boldsymbol{\theta}$ denote the gate parameters, and the circuit layout is characterized by the system linear size L, number of gate layers M, the chunk length $l \geq 2$, and the overlapping distance $1 \leq q \leq l$ [cf. Fig. 1 and Eq. (1)]. Similarly, we define the PS unitary acting on the k-th column of the 2D lattice as $U^V_{\mathrm{PS}}(k, \boldsymbol{\theta}, L, M, l, q)$.

Inspired by the construction of SGS [4], we construct the 2D PS circuits by first applying row unitaries, followed by column unitaries, to cover the 2D lattice (as illustrated in Fig. 5), as

$$U^{\mathrm{2D}}_{\mathrm{PS}}(\boldsymbol{\theta}_{\mathrm{2D}}, L, M, l, q) = \prod_{k=1}^{L} U^V_{\mathrm{PS}}(k, \boldsymbol{\theta}_{V,k}, L, M, l, q) \prod_{j=1}^{L} U^H_{\mathrm{PS}}(j, \boldsymbol{\theta}_{H,j}, L, M, l, q). \tag{13}$$

Here, $\boldsymbol{\theta}_{\mathrm{2D}} \equiv (\{\boldsymbol{\theta}_{V,k}\}, \{\boldsymbol{\theta}_{H,j}\})$ represents all gate parameters in the 2D PS circuits. The circuit depth is given by $T = 2(l + q + 2M - 3)$, and the variational states associated with the 2D PS circuits are expressed as

$$|\Psi^{2D}_{\mathrm{PS}}\rangle = U^{\mathrm{2D}}_{\mathrm{PS}}(\boldsymbol{\theta}_{\mathrm{2D}}, L, M, l, q)|0\rangle^{\otimes N}. \tag{14}$$

Similar to the 1D case, the number of layers M controls the amount of entanglement in the 2D PS circuits, such that the entanglement entropy S of $|\Psi^{2D}_{\mathrm{PS}}\rangle$ across any horizontal or vertical cut (not in the overlapping regions) satisfies $S \leq 2ML$. By fixing M and $q = 1$ while varying l, the 2D PS circuits interpolate between the multi-layer SGS ($l = L - 1$) and the 2D brickwall circuits ($l = 2$), which consist of multiple commuting layers of brickwall gates acting first along the horizontal direction, followed by the vertical direction. This analysis generalizes directly to arbitrary higher-dimensional cubic lattices, where the PS circuits are constructed by consecutively applying 'lines' of 1D PS circuits (as defined in the main text) along all lattice directions to entangle the entire

[1] Here the multi-layer SGS means that we replace the arbitrary-bond-dimension sequential circuits used in Ref. [4] by multiple layers of sequential circuits consisting of two-qubit gates.

lattice.

PS circuits in higher dimensions exhibit several interesting properties. For instance, $|\Psi_{\mathrm{PS}}^{2D}\rangle$, created by single-layer ($M = 1$) PS circuits, already includes cluster states in arbitrary dimensions, which serve as universal resources for measurement-based quantum computations in two- and higher-dimensional lattices [49]. By choosing $M = O(1)$ and $l, q = O(\log N)$, the PS circuits are capable of generating states with area-law entanglement and exponentially decaying correlations, making it a potential circuit model for gapped ground states in higher dimensions [50]. Furthermore, constant-layer 2D PS circuits enable the efficient computation of few-body correlators using the methods outlined in Ref. [4, 6]. Finally, following the arguments presented in the main text, we anticipate that higher-dimensional noisy PS circuits will also suppress error proliferation and demonstrate superior trainability and evaluation accuracy compared to their higher-dimensional noisy sequential circuit and noisy brickwall circuit counterparts.

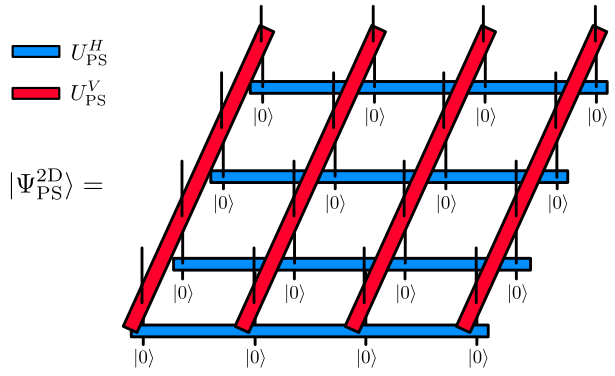


Figure 5: Two-dimensional parallel-sequential circuits, $U_{\mathrm{PS}}^{2\mathrm{D}}$, and their associated variational states, $|\Psi_{\mathrm{PS}}^{2D}\rangle$, are constructed by consecutively applying 'lines' of 1D PS circuits $U_{\mathrm{PS}}^{H}, U_{\mathrm{PS}}^{V}$ along both lattice directions to entangle the entire lattice [cf. Eq. (13)], inspired by the construction of the sequentially-generated states [4].

6 Discussion and Outlook

Parallel-sequential (PS) circuits interpolate between brickwall and sequential circuits and thus constitute a new class of circuit layouts that are simultaneously expressive and robust to noise. These layouts enable the discovery of better circuits for ground state preparation. Our work suggests a versatile, easy-to-implement strategy for enhancing the performance of quantum tasks on noisy devices by replacing sequential or brickwall circuits by the optimal layout within the PS family.

Although our discussion has focused on one-dimensional systems, we expect that many of our results extend to higher-dimensional PS circuits, which can be defined analogously [cf. section 5]. Future work should explore the properties of PS circuits, such as their information dynamics [51], simulation complexity [48], and their applications in variational quantum algorithms [15]. Additionally, PS circuits can serve as a 'skeleton' for applying adaptive circuit growth techniques [52] to identify more optimal circuit layouts for representing quantum states. It is also worth investigating the use of PS circuits to enhance numerous protocols [11, 14, 28–34] that rely on representing MPS of bond dimension $D = 2$ as a single-layer sequential circuit.

7 Acknowledgments

We thank Ignacio Cirac, Guillermo González-García, Khadijeh Sona Najafi, Adam Smith, and Georgios Styliaris for insightful discussions. The research is part of the Munich Quantum Valley, which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus. ZYW acknowledge funding from the German Federal Ministry of Education and Research (BMBF) through EQUAHUMO (Grant No. 13N16066) within the funding program Quantum Technologies—From Basic Research to Market. ZYW also acknowledges support from the U.S. Department of Energy, Office of Science, Accelerated Research in Quantum Computing, Fundamental Algorithmic Research toward Quantum Utility (FAR-Qu). DM acknowledges support from Novo Nordisk Fonden under grant numbers NNF22OC0071934 and NNF20OC0059939. The numerical calculations were performed using the ITensor.jl [53] and PastaQ.jl [54].

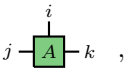
A Brief review of Matrix product state (MPS)

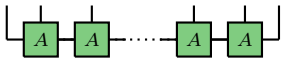
Consider a one-dimensional chain of N qudits, each of dimension d (we focus on qubit case $d = 2$ in this work), with open boundary conditions. An MPS on this chain is expressed as:

$$|\phi_{\mathrm{MPS}}\rangle = \sum_{i_1,\ldots,i_N=0}^{d-1} A_1^{i_1} \ldots A_N^{i_N} |i_1 \ldots i_N\rangle, \qquad \text{(A1)}$$

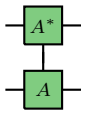
where $\{A_{[j]}^{i_j}\}$ in the bulk are matrices of dimension $D \times D$, with D denoting the bond dimension of the MPS. The boundary tensors $A_1^{i_1}$ and $(A_N^{i_N})$ are of dimensions $1 \times D$ and $D \times 1$, respectively. Every MPS can be transformed into a canonical form [55], where the tensors satisfy the condition $\sum_{i_k,\alpha} (A_k^{*i_k})_{\alpha,\beta'} (A_k^{i_k})_{\alpha,\beta} = \delta_{\beta,\beta'}$. An MPS in canonical form can be represented exactly by a single *dense* "layer" of a sequential circuit [3, 56]. For example, MPS with $D = d = 2$ can be prepared by such a "single-layer" sequential circuit composed of two-qubit gates, as illustrated in

Fig. A1(a). Multi-layer sequential circuits are circuits in which a single-layer circuit composed of two-qubit gates is applied multiple times [11, 12].

In Fig. 2, we focus on cases where the MPS tensors in the bulk are identical, resulting in states that are approximately translational invariant for sufficiently large system sizes. Such MPS are referred to as *bulk-TI* MPS. Specifically, we denote the tensors in the bulk as A, using a graphical notation $(A^i)_{jk} =$  , and the state is denoted as

$$|\phi_{\text{MPS}}^{\text{bulk-TI}}\rangle =$$  (A2)

where the boundary tensors are taken to be the identity. We also define the *transfer matrix* associated with A as

$$E_{AA} = \sum_{i=1}^{d} (A^i)^* \otimes A^i =$$  . (A3)

The correlations within MPS are characterized by the two-body correlation function:

$$C(i,j) = \langle O_i O_j \rangle - \langle O_i \rangle \langle O_j \rangle, \tag{A4}$$

where O_i and O_j are single-site operators acting on sites i and j, respectively. For bulk-TI MPS of $N \gg 1$, the correlation function follows a translationally invariant form in the bulk, $C(i,j) = C(|i-j|)$, which can be derived from the transfer matrix of the MPS [26]:

$$C(r) = \sum_{j=1}^{D^2-1} C_j \exp\left(-r/\xi_j\right), \tag{A5}$$

where $C_1, C_2, \ldots, C_{D^2-1}$ are constants, and $\xi \geq \xi_2 \geq \cdots \geq \xi_{D^2-1} > 0$ represent the length scale of each decay. The largest ξ is referred to as the correlation length of the state and characterizes the asymptotic exponential decay of correlations.

In Fig. 2, we examine a bulk-TI MPS family [26] and random bulk-TI MPS with bond dimension $D = 2$. The bulk tensors for the bulk-TI MPS family are given by [cf. Eq. (A1)]:

$$A_{[i]}^0(g) = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \quad A_{[i]}^1(g) = \begin{pmatrix} 1 & g \\ 0 & 0 \end{pmatrix}, \tag{A6}$$

where $g \in (-1, 0)$ interpolates the state from the cluster state ($g = -1$) to the GHZ state ($g = 0$). In this regime, the eigenvalues of the transfer matrix are $\{\lambda_i\} = \{1, \frac{1+g}{1-g}, 0, 0\}$. Consequently, the correlation function $C(r)$ decays as a single exponential [cf. Eq. (A4)], with the correlation length given by

$$\xi = \left(\ln \frac{1-g}{1+g}\right)^{-1}. \tag{A7}$$

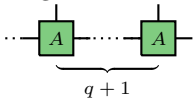
By tuning g, we can construct MPS of varying correlation lengths. Note that $g \in (-1, 0)$ encompasses all states with $g < 0$, as the tensors $\{A(g)\}$ in Eq. (A6) can be mapped to $\{A(1/g)\}$ through a gauge transformation [26, 57]. For the random bulk-TI MPS with $D = 2$, we sample the bulk tensor from the Haar random distribution [22, 23].

B PS circuits for quantum state preparation

In this section, we discuss various aspects of PS circuits for quantum state preparation, including the optimization method used to approximate MPS with $D = 2$, their performance analysis, and a comparison with the log-depth RG circuit introduced in Ref. [21].

B.1 Approximating MPS of $D = 2$

B.1.1 Local optimization algorithm

Given a single-layer sequential circuit corresponding to the target $|\phi_{\text{MPS}}\rangle$ [cf. Eq. (A1)] with bond dimension $D = 2$, we propose an efficient local optimization method to obtain the PS state $|\Psi_{\text{PS}}\rangle$ that best approximates $|\phi_{\text{MPS}}\rangle$. This approach is illustrated in Fig. A1(a), where it is shown that a sequential circuit can be parallelized by locally inverting the order of gates in a local isometry $W_{\text{SEQ}} =$  , resulting in another isometry $W_{\text{PS}}(\boldsymbol{\theta})$. This local gate-order inversion leads to a PS circuit with one additional sequential chunk. While an exact solution $\boldsymbol{\theta}_0$ satisfying $W_{\text{PS}}(\boldsymbol{\theta}_0) = W_{\text{SEQ}}$ typically does not exist, we can numerically minimize the distance between $W_{\text{PS}}(\boldsymbol{\theta})$ and $W_{\text{SEQ}}(\boldsymbol{\theta})$ using the following cost function

$$\begin{aligned} \mathcal{C}(\boldsymbol{\theta}) &= \frac{1}{2d} \|W_{\text{PS}} - W_{\text{SEQ}}\|_F^2 \\ &= 1 - \frac{1}{d} \text{Re}[\text{Tr}(W_{\text{PS}}^\dagger(\boldsymbol{\theta}) W_{\text{SEQ}})] \end{aligned} \tag{A8}$$

to find the optimal parameters $\boldsymbol{\theta}_{\text{best}}$, where $\|A\|_F = \sqrt{\text{Tr}(A^\dagger A)}$ is the Frobenius norm. Figure A1(b) graphically depicts $\mathcal{C}(\boldsymbol{\theta})$ for generic PS circuits with overlapping distance q. Since this local optimization involves only $O(q)$ gates, it can be efficiently solved using a DMRG-like sweep [36]. This optimization algorithm can be employed to prepare general MPS with bond dimension $D = 2$. To achieve this, it needs to be applied to multiple regions of the 1D chain to construct the PS circuit that approximately prepares the target MPS. In the following, we focus on bulk-TI MPS, where the same solution $W_{\text{PS}}(\boldsymbol{\theta}_{\text{best}})$ can be used to replace $W_{\text{SEQ}}(\boldsymbol{\theta})$ in different regions of the chain. For example, one can apply this replacement to create a PS circuit of chunk length l (with $n_C = \lceil N/l \rceil$

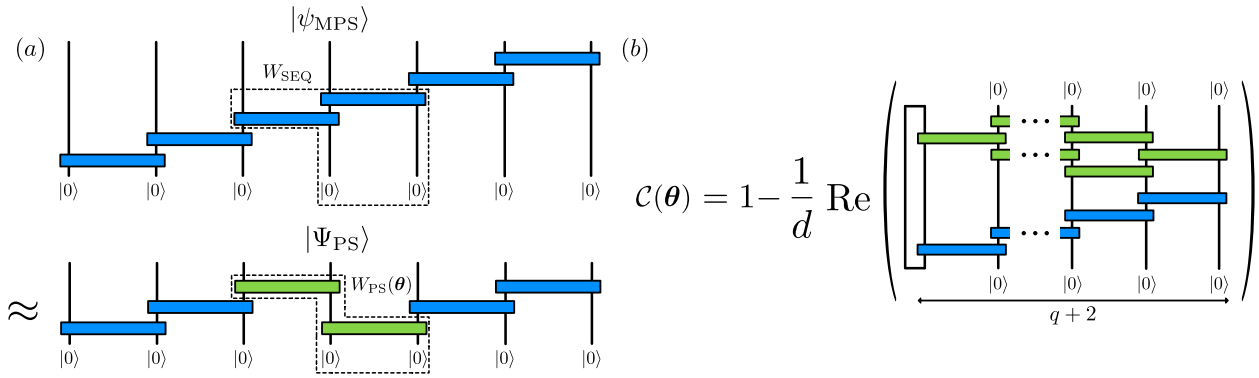


Figure A1: (a) The target MPS $|\phi_{\mathrm{MPS}}\rangle$ with bond dimension $D = d = 2$ [Eq. (A1)] can be exactly represented as a single-layer sequential circuit composed of two-qubit gates. By finding a local isometry $W_{\mathrm{PS}}(\boldsymbol{\theta})$ (denoted by the dashed box on the right side) that approximates the isometry W_{SEQ} (denoted by the dashed box on the left side) in the sequential circuit, one can invert the order of gates within W_{SEQ}, obtaining an approximation $|\Psi_{\mathrm{PS}}(\boldsymbol{\theta})\rangle \approx |\phi_{\mathrm{MPS}}\rangle$ using a PS circuit with one additional sequential chunk and reduced circuit depth. (b) Graphical representation of the cost function $\mathcal{C}(\boldsymbol{\theta})$ [Eq. (A8)] for general PS circuit with overlapping distance q.

chunks) and overlapping distance q, with its structure graphically represented as

$$|\Psi_{\mathrm{PS}}\rangle = \cdots A \cdots A\; W_{\mathrm{PS}}\; A \cdots A\; W_{\mathrm{PS}}\; A \cdots A \cdots \tag{A9}$$
$$l-q-1 \qquad q+1 \qquad l-q-1 \qquad q+1 \qquad l-q-1$$

B.1.2 Error due to missing correlations, and the scaling of the fidelity $\mathcal{F}$ for bulk-TI MPS

The gate order inversion [cf. Fig. A1] cannot be realized perfectly due to the inherent light-cone structure of the PS circuits. This is illustrated in Fig. A2, where sites i and $i+q+M+1$ ($M = 1$ here for MPS of $D = 2$) are not correlated, i.e., $C(i, i+q+M+1) = 0$ [cf. Eq. (A4)] for arbitrary local operators. As a result, approximation errors arise when attempting to use PS circuits to capture states with correlations beyond the distance $\mathcal{R}_c = q + M$, such as gapped ground states with exponentially decaying correlations. These errors can be reduced by increasing q, as shown in Fig. 2(a).

We can quantify this error due to the missed correlation by studying the state fidelity between the target bulk-TI MPS $|\phi^{\text{bulk-TI}}_{\mathrm{MPS}}\rangle$ with bond dimension $D = 2$ and its best approximation $|\Psi_{\mathrm{PS}}(\boldsymbol{\theta}_{\mathrm{best}})\rangle$ within a given PS circuit, defined as $\mathcal{F} = |\langle\Psi_{\mathrm{PS}}(\boldsymbol{\theta}_{\mathrm{best}})|\phi^{\text{bulk-TI}}_{\mathrm{MPS}}\rangle|^2$. To compute $\mathcal{F}$, we define the overlap matrix between W_{PS} and W_{SEQ} as $E_{AW} =$ 

, and the fidelity $\mathcal{F}$ can be expressed as

$$\begin{aligned}\mathcal{F} &= |\langle\Psi_{\mathrm{PS}}|\phi^{\text{bulk-TI}}_{\mathrm{MPS}}\rangle|^2 \\ &= |\langle I|(E_{AA}^{l-q-1}E_{AW})^{n_C-1}E_{AA}^{l-q-1}|I\rangle|^2,\end{aligned} \tag{A10}$$

where $\langle I|$ and $|I\rangle$ denote the left and right boundary conditions. This already implies the exponential decay of $\mathcal{F}$ with the number of sequential chunks n_C (i.e. the times that we replace W_{SEQ} with W_{PS}), as

$$\mathcal{F} \approx \exp\big[-\kappa(q)\cdot(n_C-1)\big], \tag{A11}$$

where $\kappa(q)$ represents the error rate. This scaling behavior is numerically verified for all bulk-TI MPS analyzed in this work (see Fig. A4(a) for numerical results corresponding to the MPS family discussed in Eq. (A6)). The extracted numerical scaling is utilized to determine $\kappa(q)$, as shown in Fig. 2(a).

Moreover, for $l-q-1 \gg 1$, the power of E_{AA} can be approximated as $E_{AA}^{l-q-1} \approx |\Lambda_L\rangle\langle\Lambda_R|$, where $|\Lambda_L\rangle$ and $|\Lambda_R\rangle$ denote the left and right eigenvectors associated with the eigenvalue 1. The normalization condition implies

$$\langle I|\Lambda_L\rangle\langle\Lambda_R|I\rangle \approx \langle\phi^{\text{bulk-TI}}_{\mathrm{MPS}}|\phi^{\text{bulk-TI}}_{\mathrm{MPS}}\rangle = 1. \tag{A12}$$

Consequently, in this regime, the fidelity can be expressed as

$$\begin{aligned}\mathcal{F} &\approx |\langle I|\Lambda_L\rangle\langle\Lambda_R|I\rangle(\langle\Lambda_R|E_{AW}|\Lambda_L\rangle)^{n_C-1}|^2 \\ &\approx |\langle\Lambda_R|E_{AW}|\Lambda_L\rangle|^{2(n_C-1)}\end{aligned} \tag{A13}$$

which provide another way to extract $\kappa(q) \approx -\log\big(|\langle\Lambda_R|E_{AW}|\Lambda_L\rangle|^2\big)$.

B.1.3 Details on CNOT depth scaling for $D = 2$ bulk-TI MPS [Fig. 2(b)]

The scaling of fidelity $\mathcal{F}$ [Fig. A4(a) and Eq. (A13)] and the error density $\kappa(q) = \kappa_0 \exp\big(-\gamma q/\xi\big)$

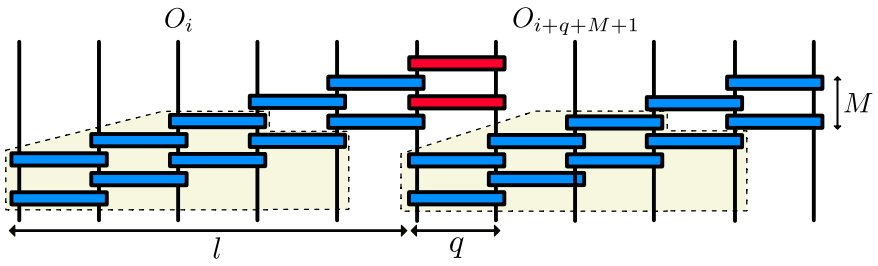


Figure A2: The correlation structure in PS circuits. Sites within each chunk (of length l) are correlated with each other, whereas correlations between sites in neighboring chunks are more limited. As an example, the regions within the dashed boxes illustrate the inverse light cones of site i and site $i+q+M+1$ near the overlap region, with $q=2$ and $M=2$ shown here. Since the inverse light cones do not overlap, no correlation exists between these two sites.

[cf. Fig. 2(a)] allow us to deduce the optimal PS circuit and the scaling of its parameters for preparing the target $D=2$ MPS of size N with a given infidelity $\epsilon = 1-\mathcal{F}$.

From Eq. (A13), for a given q and target infidelity $\epsilon \ll 1$, the maximum number of sequential chunks should be approximately $n_C \approx \epsilon/\kappa(q)+1$. Since $n_C \approx N/l$ and the circuit depth for single-layer PS circuits is $T = l+q-1$, we choose $l = q+1$ to minimize T. This choice corresponds to the minimal l that allows our optimization algorithm (which locally replaces W_{SEQ} with W_{PS}) to function effectively, as illustrated in [cf. Fig. A3]. Combining these considerations, we obtain the relation

$$\frac{N}{q+1} \approx \frac{\epsilon}{\kappa_0}\exp\big(\gamma q/\xi\big)+1, \tag{A14}$$

from which we deduce the asymptotic scaling of q (and thus l) as $q \sim \xi\cdot\log\big(\kappa_0 N/\epsilon\big)$, and the scaling of circuit depth as $T = 2q \sim \xi\cdot\log\big(\kappa_0 N/\epsilon\big)$.

To estimate the CNOT depth, we use the theoretical lower bound for implementing a generic m-to-n qubit isometry [58]:

$$T_{\mathrm{iso}}(m,n) = \left\lceil \frac{1}{4}\left(2^{n+m+1} - 2^{2m} - 2n - m - 1\right)\right\rceil. \tag{A15}$$

In some cases, this bound is tight [59]. For the first layer of gates acting on the initial qubits, we have $T_{\mathrm{iso}}(0,2)=1$, and for generic two-qubit gates, $T_{\mathrm{iso}}(2,2)=3$. Thus, the CNOT depth is $T_{\mathrm{CNOT}} = 3T \sim \xi\cdot\log\big(\kappa_0 N/\epsilon\big)$, as numerically demonstrated in Fig. 2(b).

For comparison, we also show T_{CNOT} for the sequential-RG circuit described in Ref. [21] for preparing the same set of $D=2$ MPS, in Fig. 2(b). The details of the T_{CNOT} calculation are provided in the supplementary materials of Ref. [21]. We find that PS circuits consistently yield shallower circuits for preparing the same state for a given system size N. This difference arises from the multi-qubit isometries used in RG circuits [21], such as the 2-to-3 isometries (with $T_{\mathrm{iso}}(2,3)=14$) employed in sequential-RG circuits for $D=2$ MPS. In section B.2, we show that the advantage of PS circuits over the RG circuits persists for states with higher bond dimensions.

B.1.4 Robustness against local minima

To determine the gate parameters $\boldsymbol{\theta}_{\mathrm{best}}$ of the PS circuit for preparing the bulk-TI MPS $|\phi_{\mathrm{MPS}}^{\mathrm{bulk-TI}}\rangle$ of size N, correlation length ξ, and infidelity ϵ, it suffices to find the global minimum of the cost function $\mathcal{C}(\boldsymbol{\theta})$ [cf. Fig. A1(c)] for a one-dimensional local structure characterized by the overlapping distance $q \sim \xi\cdot\log\big(N/\epsilon\big)$ [cf. section B.1.3]. This ensures computational efficiency for each run of the DMRG-like sweep [36] inside $W_{\mathrm{PS}}(\boldsymbol{\theta})$. However, optimizations on quantum circuits are known to suffer from the presence of local minima in the optimization landscape [15, 16], raising concerns about whether our algorithm can asymptotically and efficiently find $\boldsymbol{\theta}_{\mathrm{best}}$ corresponding to the globally minimal fidelity error density $\kappa(q)$.

To investigate the impact of local minima, in Fig. A4(b), we present the error density $\kappa(q)$ obtained from individual runs of the optimization algorithm with randomly initialized gates for single-layer PS circuits with various q and $l=q+1$ [cf. Fig. A3]. As q increases (corresponding to a larger optimization problem size), we observe multiple instances where the optimization halts at local minima. We estimate the success probability $p_{\mathrm{succ}}(q)$ of finding the optimal κ by calculating the ratio of optimizations that find the global minimum $\kappa(q)$ within a relative error of 1% and show the results in the inset of Fig. A4(b). Our results indicate that beyond a threshold value of q, p_{succ} decreases almost exponentially with q. This behavior reflects the NP-hardness [60] of classically optimizing variational circuits.

Remarkably, despite the near-exponential decay of $p_{\mathrm{succ}}(q)$ with q, our optimization algorithm is expected to run $\mathrm{poly}(N/\epsilon)$ times to find the global minimum, as q scales only logarithmically with N/ϵ. This suggests that the local optimization algorithm [cf. section B.1.1] can scalably find sufficiently good solutions.

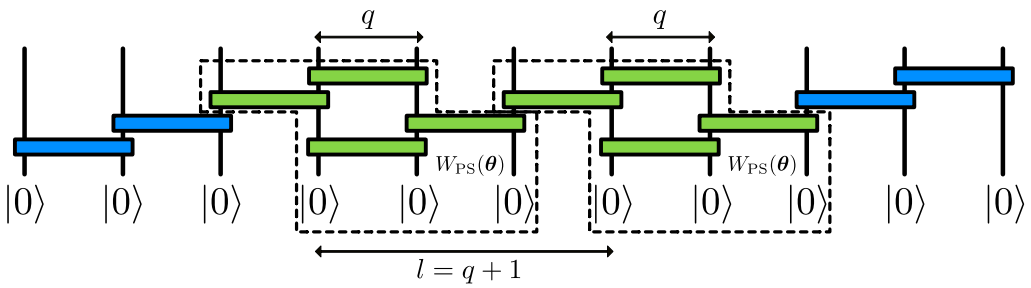


Figure A3: Illustration of the single-layer PS circuit with $l = q + 1$, representing the maximal level of parallelization (i.e., the minimal value of l) at which the local optimization algorithm remains applicable.

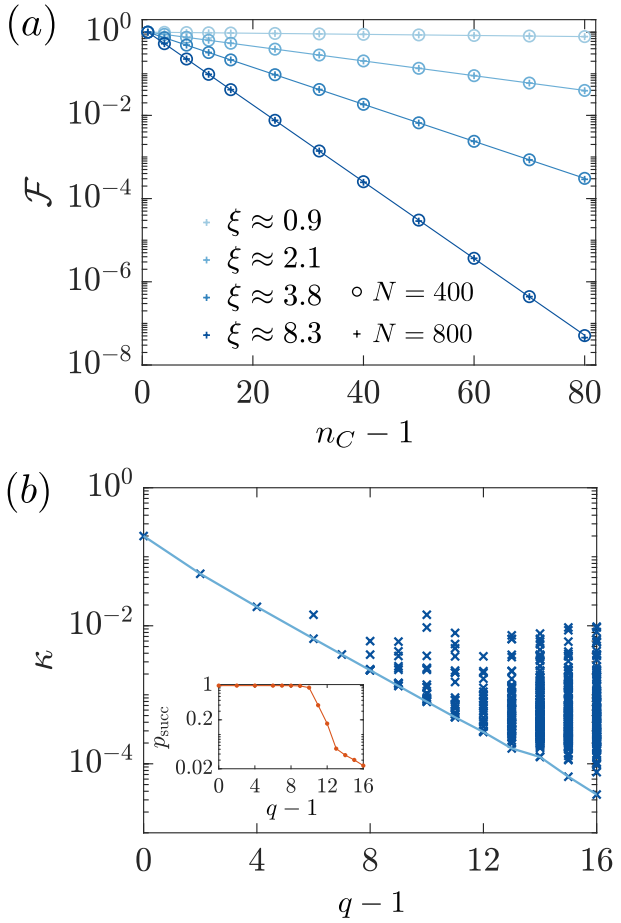


Figure A4: (a) The scaling of fidelity $\mathcal{F}$ [Eq. (A10)] with the number of chunks n_C for various states in the MPS family [cf. Eq. (A6)] and system sizes. Here we fix the overlapping distance $q = 2$, and the chunk length l is varied to get different n_C. (b) The error density $\kappa(q)$ achieved by individual runs of the optimization algorithm with randomly initialized gates for PS circuits [cf. Fig. A3] with various q, for a target MPS [cf. Eq. (A6)] with $\xi \approx 3.8$. The line connects the minimal κ values achieved for each q, representing the exponential decay of the global minima $\kappa(q)$ with q, as observed in Fig. 2(a). The inset shows the success probability $p_{\rm succ}$ of the optimization algorithm finding the global minimum $\kappa(q)$ within a relative error of 1%. We sample enough points to ensure the optimization meets the success criterion at least 10^3 times for each q.

B.2 Circuit compilation overhead of RG circuits [21] for preparing MPS of high bond dimensions

As discussed in the main text, by varying the number of layers M, PS circuits can represent quantum states with entanglement across non-overlapping cuts bounded by $S \leq 2M$. This has been explored in the case of multi-layer sequential circuits for preparing MPS with higher bond dimensions, such as the ground states of local Hamiltonians and MPS arising from algorithmic applications or quantum dynamics [11, 12, 14, 28–34]. These studies generally suggest that a favorable number of layers, $M = O(1)$, is sufficient to prepare relevant MPS with entanglement $S = O(1)$ (where the bond dimension $D \sim e^S$). While the circuit depth for sequential circuits scales linearly with the system size as $T_{\rm SEQ} \sim N$, PS circuits are expected to achieve comparable performance to sequential circuits with the same number of layers while requiring significantly shallower depths, as demonstrated for bulk-TI MPS with $D = 2$ and the ground state of the XY model [cf. Fig. 2 and 3].

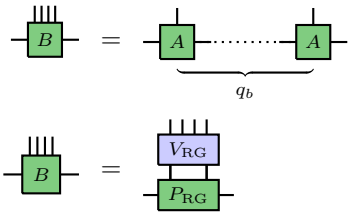
On the other hand, there exists a rigorous algorithm to convert arbitrary MPS with high bond dimensions into a single dense "layer" of a sequential circuit [3] or to convert short-range correlated MPS into log-depth RG circuits [21]. It is already established that multi-layer sequential circuits (corresponding to PS circuits with $l = N - 1$ and $q = 1$) typically offer a more compact representation of high bond-dimension MPS compared to the 'dense' sequential circuits [11–14]. To complement these results, we focus on analyzing the gate compilation overhead of RG circuits for preparing short-range correlated MPS with higher bond dimensions and qualitatively discuss why this makes PS circuits a preferable choice over RG circuits for preparing states with higher entanglement.

B.2.1 $T_{\rm CNOT}$ scaling of the RG circuits

Here, we briefly review the RG circuits introduced in Ref. [21] and derive the scaling of $T_{\rm CNOT}$ for preparing an N-qubit ($d = 2$) MPS [cf. Eq. (A1)] with bond dimension D and correlation length ξ, up to a small

infidelity ϵ. Note that the analysis below can be easily generalized to physical dimension d.

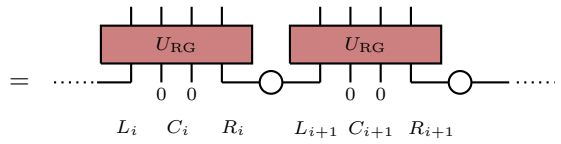
Starting from a given MPS, the RG circuits are constructed by grouping each set of q_b neighboring tensors $\{A\}$ into a single tensor B, followed by a polar decomposition $B = V_{\mathrm{RG}} P_{\mathrm{RG}}$ to extract the isometry V_{RG}, which renormalizes the MPS toward its fixed point [21].



(A16)

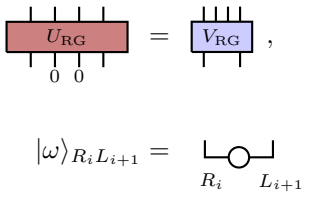
Therefore, one can construct an approximation of the target MPS as

$$|\widetilde{\phi}_N\rangle = \left(\bigotimes_{i=1}^{N/q_b} U_i\right) \bigotimes_{i=1}^{N/q_b} \left(|\omega\rangle_{R_i L_{i+1}} |0\dots 0\rangle_{C_i}\right)$$



(A17)

where the unitary is constructed such that it implements the required isometry when acting on a product state $|0\rangle^{\otimes(q_b-2)}$ over the 'central' region ($q_b = 4$ in the illustration below), and the fixed-point state $|\Omega\rangle = \bigotimes_{i=1}^{N/q} |\omega^i\rangle_{R_i L_{i+1}}$ is a tensor product of entangled pairs of qudit with qudit dimension D,



(A18)

The infidelity ϵ between the approximation $|\widetilde{\phi}_N\rangle$ and the target MPS is shown to scale as $\epsilon = O\left(\frac{N}{q} e^{-\gamma q/\xi}\right)$ [21, 61], which is numerically demonstrated for bulk-TI MPS of $D = 2$ in Fig. 2(a) [2]

Ref. [21] provides two explicit constructions for the circuits within the unitary U_{RG}, illustrated in Fig. A5. The *sequential-RG* circuit iteratively applies singular value decompositions to construct a sequential circuit on q_b sites, consisting of multi-qubit isometries with a mapping dimension of $D^2 \to 2D^2$ and a layer of $O(q)$ SWAP gates. In contrast, the *tree-RG* approach iteratively applies blocking and polar decomposition for $O(\log\log N)$ times, yielding a tree-like circuit with $O(\log\log N)$ layers of long-range isometries, each with a mapping dimension of $D^2 \to D^4$. Taking $k = \lceil \log_2 D \rceil$,

By using the lower bound from Eq. (A15) for the CNOT depth required to implement an isometry, we can derive a lower bound of the total CNOT depth, T_{CNOT}, for the RG circuit with a blocking size of $q_b > 2k$ as:

$$T_{\mathrm{CNOT}}^{\mathrm{Seq-RG}} \geq T_{\mathrm{pair}} + T_{\mathrm{SWAP}} + T_{\mathrm{up}}, \tag{A19}$$

where each component is given by:

$$\begin{aligned} T_{\mathrm{pair}} &= T_{\mathrm{iso}}(0, 2k) \\ &= \left\lceil \frac{1}{4}\left(2^{2k+1} - 4k - 2\right)\right\rceil \sim D^2 \end{aligned} \tag{A20}$$

$$T_{\mathrm{SWAP}} = 3(q_b - k - 1), \tag{A21}$$

$$\begin{aligned} T_{\mathrm{up}} &= (q_b - 2k)\, T_{\mathrm{iso}}(2k, 2k+1) \\ &= (q_b - 2k)\left\lceil \frac{1}{4}\left(3 \cdot 2^{4k} - 6k - 3\right)\right\rceil \sim D^4 \end{aligned} \tag{A22}$$

For the tree-RG circuits, the total CNOT depth is:

$$T_{\mathrm{CNOT}}^{\mathrm{tree-RG}} \geq T_{\mathrm{pair}} + T_{\mathrm{tree}} \tag{A23}$$

$$\begin{aligned} T_{\mathrm{tree}} &= (q_b - 2k)\, T_{\mathrm{iso}}(2k, 4k) \\ &= (q_b - 2k)\left\lceil \frac{1}{4}\left(2^{6k+1} - 2^{4k} - 10k - 1\right)\right\rceil \sim D^6 \end{aligned} \tag{A24}$$

Therefore, for $D \gg 1$, we asymptotically have $T_{\mathrm{CNOT}}^{\mathrm{Seq-RG}} = \tilde{\Omega}(D^4)$, and $T_{\mathrm{CNOT}}^{\mathrm{tree-RG}} = \tilde{\Omega}(D^6)$. Here the notation $\tilde{\Omega}$ suppresses the logarithmic factors.

The above scalings suggest that RG circuits incur a significant circuit compilation overhead for preparing short-range correlated MPS with higher bond dimensions. As a simple example, achieving an MPS bond dimension of $D = 64$ corresponds to an approximate 10^6-fold (sequential-RG) and 10^9-fold (tree-RG) increase in T_{CNOT} for fixed q_b, rendering RG circuits impractical for preparing MPS with high bond dimensions. In contrast, a PS circuit with $M = 3$ layers can already prepare a certain subclass of MPS with a bond dimension of $D = 2^{2M} = 64$. Although it remains to be explored how the number of parameters scales for PS circuits to prepare generic MPS with high bond dimensions, we expect that PS circuits will be a better choice than RG circuits for preparing states with higher entanglement on near-term noisy quantum devices.

We also note that it is possible to consider circuit layouts formed by multiple layers of RG circuits

[2] For a given q of the PS circuit, we plot the $q_b = q + 1$ result of the RG circuit in Fig. 2(a). This choice corresponds to comparing the RG circuit and the PS circuit with the same maximal correlation range [cf. Fig. A2].

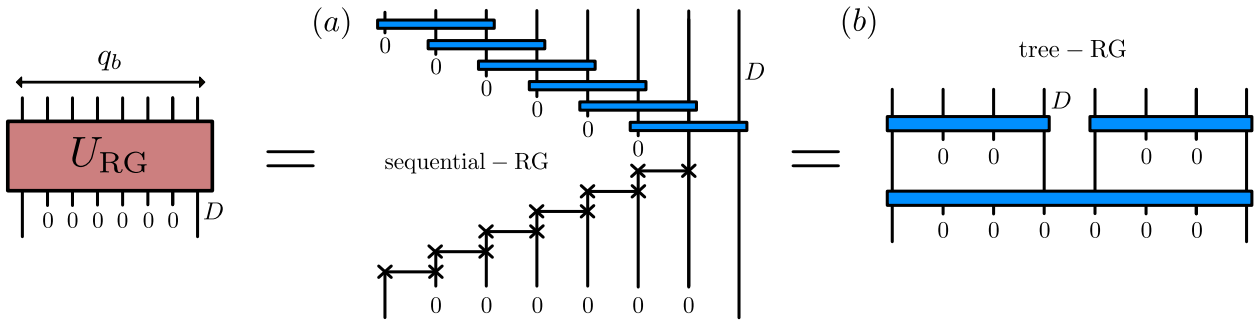


Figure A5: Decomposition of the isometry U_{RG} as (a) the sequential-RG circuit, or (b) tree-RG circuit. The '0' denotes the input product state $|0\rangle$ of the isometries. See Ref. [21] for details.

for $D = 2$ to reduce the circuit depth required for RG circuits to create entanglement. However, since the single-layer PS circuits achieve the same performance as RG circuits while having a shallower depth [cf. Fig. 2(a,b)], we expect that multi-layer PS circuits will also outperform such multi-layer RG circuits for MPS preparation.

C Noise robustness of PS circuits

In this section, we provide additional details on the noise robustness of PS circuits, including error propagation, gradient scaling, and evaluation accuracy.

C.1 The impact of noise on the energy density

Given a circuit layout, we use a DMRG-like algorithm [36] to determine the circuit parameters $\boldsymbol{\theta}_{\mathrm{best}}$ that minimize the energy

$$E = \langle \Psi_{\mathrm{PS}}(\boldsymbol{\theta}_{\mathrm{best}}) | H_{XY} | \Psi_{\mathrm{PS}}(\boldsymbol{\theta}_{\mathrm{best}}) \rangle \tag{A25}$$

for a system size of $N = 30$, ensuring that finite-size effects are negligible. We compare the ground-state energy error density, $\nu_{\mathrm{XY}} \equiv (E - E_{\mathrm{GS}})/N$, where E_{GS} is the exact ground-state energy, with the results shown in Fig. 3(a).

To evaluate the impact of noise, we introduce errors following a noise model that includes both idling errors and two-qubit gate errors [cf. Fig. 3(b)]. We then calculate the energy error density $\nu_{\mathrm{noisy}} \equiv (E_{\mathrm{noisy}} - E_{\mathrm{GS}})/N$, where $E_{\mathrm{noisy}} = \mathrm{Tr}[\rho(\boldsymbol{\theta}_{\mathrm{best}}) H_{XY}]$, and $\rho(\boldsymbol{\theta}_{\mathrm{best}})$ is the density matrix generated by the noisy circuit using the same optimized parameters $\boldsymbol{\theta}_{\mathrm{best}}$. The results are shown in Fig. A6, confirming the linear dependence of the energy density on the amount of noise present in the circuits, as [cf. Eq. (A26), restated here]:

$$\nu_{\mathrm{noisy}} - \nu_{\mathrm{XY}} \approx c_E (p_1 T + 2 p_2 M). \tag{A26}$$

where c_E is a model-dependent constant ($c_E \approx 0.7$ for the XY model).

C.2 Details of Gradiant variance calculation

To study the scaling of gradient variance for noisy PS circuits, we parameterize each two-qubit gate using CNOT gates and single-qubit rotations [62], as shown in Fig. A7. An arbitrary single-qubit unitary A is parametrized as

$$A(\boldsymbol{\theta}) = e^{i\alpha} R_z(\beta)\, R_y(\gamma)\, R_z(\delta) \quad \text{with} \quad R_\alpha(\theta) = \exp\!\left(-i\sigma_\alpha \theta/2\right) \tag{A27}$$

where $\{\sigma_\alpha\}_{\alpha=x,y,z}$ are Pauli matrices. Thus, each two-qubit gate contains 15 parameters, and all parameters of the circuit are collectively denoted as $\boldsymbol{\theta}$. We use automatic differentiation implemented in PastaQ.jl [54] to evaluate the gradient variance of the cost function $E_{\mathrm{noisy}}(\boldsymbol{\theta}) = \mathrm{Tr}[\rho(\boldsymbol{\theta}) H_{XY}]$ for individual random realizations of the circuit. The gradient variance is then averaged over all realizations and parameters $\boldsymbol{\theta}$ to compute

$$V_E^{\mathrm{noisy}}(M, T, p_1, p_2) \equiv \langle (\partial_{\theta_j} E_{\mathrm{noisy}})^2 \rangle_{\boldsymbol{\theta}} - \langle \partial_{\theta_j} E_{\mathrm{noisy}} \rangle_{\boldsymbol{\theta}}^2, \tag{A28}$$

as numerically illustrated in Fig. 4(a).

In Fig. A8(a), we verify the scaling of V_E^{ideal} for noiseless PS circuits with $q = 1$ [cf. Fig. A9(d)] [3], and Fig. A8(b, c) confirms its scaling with the amount of error density $p_1 T$ and $2 p_2 M$. The prefactors for the XY model are $\alpha \approx 0.6$ and $c_V \approx 5.5$. Therefore, Eq. (11) provides a good qualitative description of the gradient variance in PS circuits and suggests that PS circuits with constant M layers exhibit better trainability than sequential circuits and brickwall circuits of the same depth, as discussed in the main text [cf. Fig. 4(a)].

C.3 Details of error propagation calculation

C.3.1 Review on the error propagation setup in Ref. [44]

As mentioned in the main text, to study the effect of error propagation, we consider the setup proposed in Ref. [44], which we briefly review here, and refer the reader to Ref. [44] for more details.

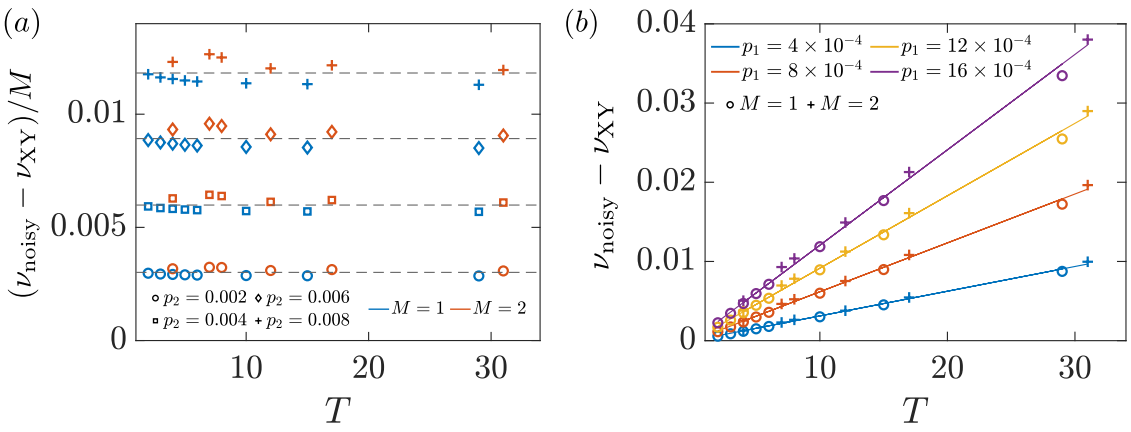


Figure A6: Numerical verification of the scaling of the energy density difference between noisy and noiseless PS circuits, $\nu_{\text{noisy}} - \nu_{\text{XY}}$ [Eq. (A26)], for the cases of (a) various circuit layouts with two-qubit gate errors ($p_1 = 0$) and (b) various circuit layouts with idling errors ($p_2 = 0$).

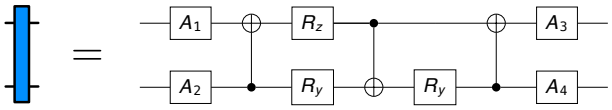


Figure A7: Circuit decomposition of an arbitrary two-qubit gate. Here $A_1, \ldots, A_4$ are arbitrary single-qubit rotations, and R_z, R_y are single-qubit rotations along the z and y directions [cf. Eq. (A27)].

We consider a noisy random circuit evolution that begins from an initial product state $|0\rangle^{\otimes N}$, followed by an entangling evolution U and the corresponding disentangling evolution $U^\dagger$, while experiencing single-qubit depolarizing noise with a rate of p_1. Here, U represents a random circuit of depth T with a given layout; the circuit depth of $U^\dagger U$ is consequentially $2T$.

In the absence of errors, the final state is again $|0\rangle^{\otimes N}$. If a single-qubit depolarizing error occurs somewhere in the circuit, it may end up affecting more than a single qubit, because the random gates can propagate errors. In particular, if we consider the average channel obtained by acting with a two-qubit gate $\mathcal{U}$ some channel $\mathcal{M}$ and the inverse gate $\mathcal{U}^\dagger$

$$\mathcal{E}_d(\rho) = \int_{\mathcal{U}} d\mathcal{U} \left[\mathcal{U}^\dagger \mathcal{M} \mathcal{U}\right](\rho), \tag{A29}$$

we find that depending on whether $\mathcal{M}$ depolarizes zero, one, or two qubits, there are the three possible outcomes [44]:

1. If no error occurs, i.e. $\mathcal{M}(\rho) = \rho$, then $\mathcal{E}_d(\rho) = \rho$.
2. If both qubits are depolarized, i.e., $\mathcal{M}(\rho) = \mathrm{Tr}(\rho)\mathbb{1}^{\otimes 2}/4$, then $\mathcal{E}_d(\rho) = \mathrm{Tr}(\rho)\mathbb{1}^{\otimes 2}/4$.
3. If an error occurs in only one of the qubits, i.e., $\mathcal{M}(\rho) = \mathbb{1}/2 \otimes \mathrm{Tr}_1(\rho)$ or $\mathcal{M}(\rho) = \mathrm{Tr}_2(\rho) \otimes \mathbb{1}/2$, then
$$\mathcal{E}_d(\rho) = \frac{1}{5}\rho + \frac{4}{5}\mathrm{Tr}(\rho)\frac{\mathbb{1}^{\otimes 2}}{4}.$$

Equation A30 describes the error propagation, where there is a probability of 4/5 that a single error evolves into two errors after being processed by the random unitary and a 1/5 chance that it disappears. This enables modeling the error propagation process as a Markov chain, where the system is represented as a string of zeros and ones: zeros denote noiseless qubits, and ones denote depolarized qubits. Using Markov-chain Monte Carlo, the average number of depolarized qubits, $\langle\eta\rangle$, can be computed at the end of the circuit evolution [44], as shown in Fig. 4(b).

In the following, we show that the ratio of depolarized qubits $\langle\eta\rangle/N$ in multi-layer sequential circuits is dominated by a term proportional to $p_1 MT$ for $M > 1$, which notably coincides with the behavior of brickwall circuits, where $M = T/2$ and $\langle\eta\rangle/N \propto p_1 T^2$. We then confirm numerically that PS circuit exhibit the same behavior. Thus for $p_1 MT \ll 1$, we find that $\langle\eta\rangle$ interpolates between the brickwall and sequential limits and find that it is well approximated as

$$\langle\eta\rangle/N \approx p_1 T(c_{\eta,1} + c_{\eta,2} M), \tag{A30}$$

with prefactors $c_{\eta,1}, c_{\eta,2}$ of order 1 and mildly dependent on T, M, but very weak dependence on system size.

C.3.2 Verification of the error propagation scaling [Eq. (12)]

Derivation and evidence for the sequential circuits.— We first consider a single depolarizing error that occurs in the single-layer sequential circuit at site i in bulk, as illustrated in Fig. A9(a). Following the Markov chain evolution described in Eq. (A29), each two-qubit gate receiving one depolarized qubit as input propagates the error with a probability of 4/5 (successful), while the error disappears with a probability of 1/5 (unsuccessful). From the layout of the circuit [cf. Fig. A9(a)], it is evident that error propagation leads to a string of errors whose length depends

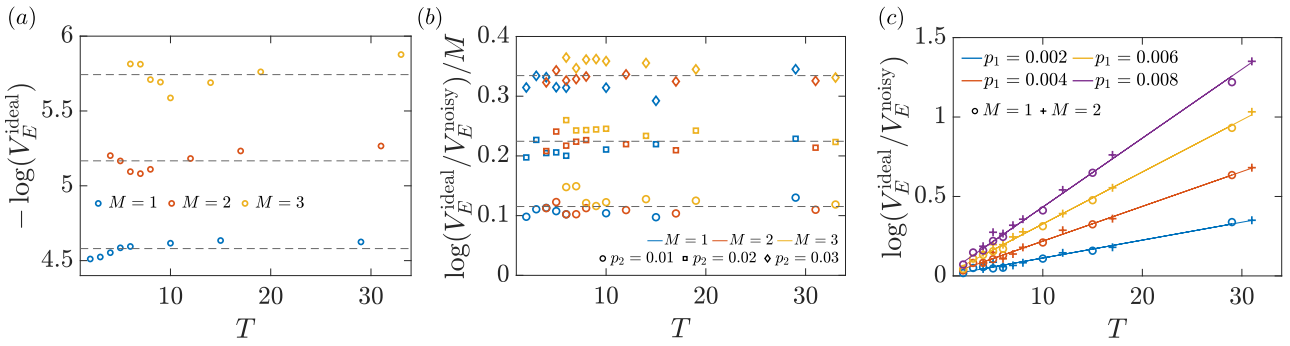


Figure A8: Numerical verification of the scaling of the gradient variance V_E [cf. Eq. (11)]. (a) The scaling of the gradient variance in noiseless circuits (denoted as V_E^{ideal}) for various circuit depths T and numbers of layers M. The dashed lines represent the mean value of the data for the same M. (b) The scaling of the ratio between noisy V_E and noiseless V_E^{ideal} for various T and M, with different two-qubit gate noise levels p_2, and fixed $p_1 = 0$. We also perform a data collapse by dividing the exponent by M. The dashed lines represent the mean values of the data for the same p_2. (c) Same as (b), but for various p_1 and fixed $p_2 = 0$. The lines represent linear fits to the data for the same p_1.

on how often the error was propagated successfully.

Given a error string of arbitrary length k, applying a sequential circuit extends this string by an amount $r \in \{-1, 0, 1, 2, \dots\}$ with probability $(4/5)^{r+1}/5$. On average, the length of the string thus increases by

$$\frac{1}{5}\sum_{k=0}^{\infty}\left(\frac{4}{5}\right)^{k}\cdot(k-1) = 3 \tag{A31}$$

for each application of a sequential circuit. However, if the string length hits 0, the dynamics gets stuck there, since subsequent sequential circuits do nothing to the string length. This dynamics is similar to a biased random walk with absorbing boundary condition at 0. In our case, the probability of being absorbed at 0 approaches $1/4$ as M approaches infinity. After the M^{th} sequential circuit, we thus have a probability distribution of error string lengths l and that yields some expected error string length $L(M)$. The first few values of $L(M)$, are $L(0) = 1$, $L(1) = 4$, $L(2) = 6.4$, $L(3) = 8.704$, $L(4) = 10.97728$.

Asymptotically, we could estimate that the asymptotic error length is the probability of there still being an error $(3/4)$ times the average length if there were no absorbing boundary: $L(M) \to (3/4)(1+3M)$. This result is almost correct, but it neglects the fact that the surviving trajectories grow faster than 3 at early times, because of the postselection on surviving. The true result to 10 digits of numerical precision in simulations for $M = 50$ and $N = 5000$ is $L(M) \to 2 + (9/4)M$.

In the setup described in the main text, the system experiences idle errors at a rate of p_1 [3]. For simplicity, we neglect the interaction between error propagation at different sites, which is valid in the regime of sparse errors $p_1 TM \ll 1$. We also take $M \ll T$, which allows us to neglect errors produced a qubit between applying the first and last gate. The qubit at location $j \in [2, N]$ accumulates a depolarizing probability $p_d(j) = 1 - (1-p_1)^{2(j-1)} \approx 2p_1(j-1)$ before being acted upon by the gates in the sequential circuit [cf. Fig. A9(a)] ($p_d(1) = p_d(2)$, because the qubits are acted on simultaneously). Thus, the average probability that a qubit has an error when it is hit by the first gate is $\approx p_1 T$. Together with our result in the previous paragraph, the number of errors after M layers of the sequential circuit will be $\approx p_1 TL(M)$. After the last gate of the M layers was applied, the qubits idle for at time $N - j$, and during this time they may again depolarize, which happens with an average probability $p_1 T$.

Putting everything together, we find that $\langle \eta \rangle / N \approx p_1 T(L(M) + 1)$ for multi-layer sequential circuits.

Extending to PS circuits.—In Fig. 4(b), we studied the error propagation effect in PS circuits with a constant number of layers M, where the depth T is controlled by the chunk length l, with a fixed $q = 1$ (no extended gates) [4]. This setup corresponds to an interpolation between brickwall and sequential circuits, as illustrated in Fig. A9(d). Notably, the PS circuit can be viewed as multiple parallelized sections of the sequential circuit covering the entire chain, where each section has a size of l. If l is large, we expect the behavior to follow the one of sequential circuits, and when l is small, it approaches the brickwall limit. The dominant contribution to either limit is proportional to $p_1 MT$, with slightly different prefactors. Ultimately, the behavior in Eq. (A30) can serve as a good guide for general PS circuits, which we numerically verify in Fig. A9(b, c).

[3] Note that in this error propagation analysis, we set the two-qubit gate error $p_2 = 0$ to ensure a fair comparison between different circuit layouts, as we focus solely on the effects of error propagation here.

[4] Since M-layer PS circuits with a finite $1 \le q \le l$ always contain fewer gates than the corresponding $M+1$-layer PS circuits with $q = 1$, the results for $q = 1$ are sufficient to represent the random circuit properties for other q.

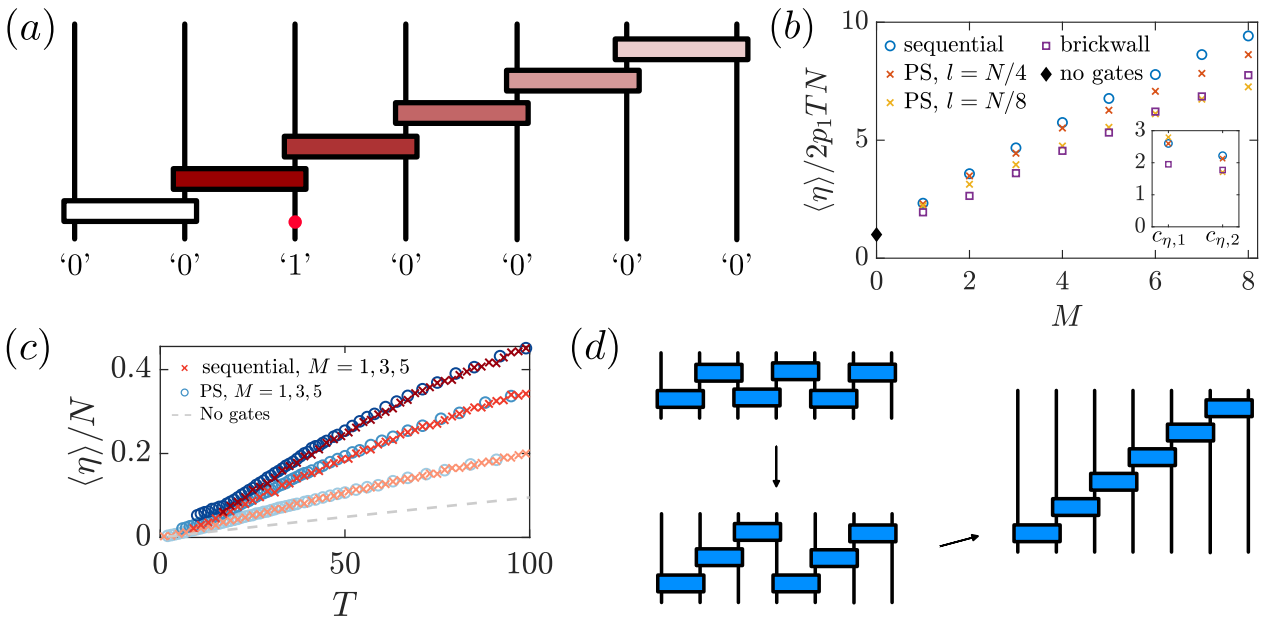


Figure A9: (a) Illustration of the propagation of a single error (denoted as '1' and the red dot) in the single-layer sequential circuit. The brown color represents the gates that propagate errors following the Markov-chain evolution Eq. (A29), with gradually lighter shades indicating that the probability of error propagation decreases with distance. (b) Scaling of the ratio of depolarized qubits, $\langle\eta\rangle/N$ (normalized by $2p_1T$), with the number of layers M for PS circuits with different layouts, including the sequential and brickwall circuits. The inset show the extracted coefficient $c_{\eta,1}$ and $c_{\eta,2}$ [cf. Eq. (A30)]. (c) Scaling of $\langle\eta\rangle/N$ with circuit depth T for PS circuits [same data as in Fig. 4(b)] and sequential circuits, where T is scaled by varying the system size N, with $T = N + 2M - 3$. The noise rate $p_1 = 5 \times 10^{-4}$. (d) Illustration of PS circuit layouts ($q = 1$) interpolating between brickwall and sequential circuits by tuning the chunk length l.

References

[1] M A Nielsen and I L Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2000.

[2] Guifré Vidal. Efficient simulation of one-dimensional quantum many-body systems. In: *Physical Review Letters* 93.4 (2004), pp. 1–4.

[3] C. Schön et al. Sequential generation of entangled multiqubit states. In: *Physical Review Letters* 95.11 (2005), pp. 1–4.

[4] M. C. Bañuls et al. Sequentially generated states for the study of two-dimensional systems. In: *Physical Review A* 77.5 (2008), pp. 1–9.

[5] Michael P. Zaletel and Frank Pollmann. Isometric Tensor Network States in Two Dimensions. In: *Physical Review Letters* 124.3 (2020), p. 37201.

[6] Zhi Yuan Wei, Daniel Malz, and J Ignacio Cirac. Sequential Generation of Projected Entangled-Pair States. In: *Physical Review Letters* 128.1 (2022), pp. 1–14.

[7] Yu-Jie Liu, Kirill Shtengel, Adam Smith, and Frank Pollmann. Methods for Simulating String-Net States and Anyons on a Digital Quantum Computer. In: *PRX Quantum* 3 (4 2022), p. 040315.

[8] Xie Chen et al. Sequential quantum circuits as maps between gapped phases. In: *Physical Review B* 109.7 (2024), p. 075116.

[9] Y. Y. Shi, L. M. Duan, and G Vidal. Classical simulation of quantum many-body systems with a tree tensor network. In: *Physical Review A* 74.2 (2006), pp. 1–4.

[10] Adam Smith, Bernhard Jobst, Andrew G Green, and Frank Pollmann. Crossing a topological phase transition with a quantum computer. In: *Physical Review Research* 4.2 (2022), p. L022020.

[11] Shi Ju Ran. Encoding of matrix product states into quantum circuits of one- A nd two-qubit gates. In: *Physical Review A* 101.3 (2020), pp. 1–7.

[12] Sheng-Hsuan Lin et al. Real-and imaginary-time evolution with compressed quantum circuits. In: *PRX Quantum* 2.1 (2021), p. 10342.

[13] Reza Haghshenas, Johnnie Gray, Andrew C Potter, and Garnet Kin-Lic Chan. Variational power of quantum circuit tensor networks. In: *Physical Review X* 12.1 (2022), p. 11047.

[14] Manuel S Rudolph et al. Decomposition of matrix product states into shallow quantum circuits. In: *Quantum Science and Technology* 9.1 (2023), p. 015012.

[15] Marco Cerezo et al. Variational quantum algorithms. In: *Nature Reviews Physics* 3.9 (2021), pp. 625–644.

[16] Kishor Bharti et al. Noisy intermediate-scale quantum algorithms. In: *Reviews of Modern Physics* 94.1 (2022), p. 15004.

[17] Jules Tilly et al. The variational quantum eigensolver: a review of methods and best practices. In: *Physics Reports* 986 (2022), pp. 1–128.

[18] Matthew B Hastings. An area law for one-dimensional quantum systems. In: *Journal of Statistical Mechanics: Theory and Experiment* 2007.08 (2007), P08024.

[19] Jens Eisert, Marcus Cramer, and Martin B Plenio. Colloquium: Area laws for the entanglement entropy. In: *Reviews of modern physics* 82.1 (2010), p. 277.

[20] M. B. Hastings. Locality in Quantum and Markov Dynamics on Lattices and Networks. In: *Phys. Rev. Lett.* 93 (14 2004), p. 140402.

[21] Daniel Malz, Georgios Styliaris, Zhi-Yuan Wei, and J Ignacio Cirac. Preparation of matrix product states with log-depth quantum circuits. In: *Physical Review Letters* 132.4 (2024), p. 040404.

[22] Silvano Garnerone, Thiago R de Oliveira, Stephan Haas, and Paolo Zanardi. Statistical properties of random matrix product states. In: *Physical Review A* 82.5 (2010), p. 052312.

[23] Jonas Haferkamp, Christian Bertoni, Ingo Roth, and Jens Eisert. Emergent statistical mechanics from properties of disordered random matrix product states. In: *PRX Quantum* 2.4 (2021), p. 40308.

[24] Cécilia Lancien and David Pérez-Garcia. Correlation length in random MPS and PEPS. In: 23.1 (2022), pp. 141–222.

[25] D Perez-Garcia, F Verstraete, M M Wolf, and J I Cirac. Matrix Product State Representations. In: *Quantum Info. Comput.* 7.5 (2007), pp. 401–430.

[26] Michael M. Wolf, Gerardo Ortiz, Frank Verstraete, and J. Ignacio Cirac. Quantum phase transitions in matrix product systems. In: *Physical Review Letters* 97.11 (2006), pp. 1–4.

[27] Carlos Bravo-Prieto, Josep Lumbreras-Zarapico, Luca Tagliacozzo, and José I. Latorre. Scaling of variational quantum circuit depth for condensed matter systems. In: *Quantum* 4 (2020).

[28] Mohsin Iqbal, David Muñoz Ramo, and Henrik Dreyer. Preentangling Quantum Algorithms–the Density Matrix Renormalization Group-assisted Quantum Canonical Transformation. In: *arXiv:2209.07106* (2022).

[29] Prithvi Gundlapalli and Junyi Lee. Deterministic and entanglement-efficient preparation of amplitude-encoded quantum registers. In: *Physical Review Applied* 18.2 (2022), p. 024013.

[30] Ar A Melnikov et al. Quantum state preparation using tensor networks. In: *Quantum Science and Technology* 8.3 (2023), p. 035027.

[31] Jason Iaconis, Sonika Johri, and Elton Yechao Zhu. Quantum state preparation of normal distributions using matrix product states. In: *npj Quantum Information* 10.1 (2024), p. 15.

[32] Vladyslav Bohun et al. Scalable and shallow quantum circuits encoding probability distributions informed by asymptotic entanglement analysis. In: *arXiv:2412.05202* (2024).

[33] Javier Gonzalez-Conde, Thomas W Watts, Pablo Rodriguez-Grasa, and Mikel Sanz. Efficient quantum amplitude encoding of polynomial functions. In: *Quantum* 8 (2024), p. 1297.

[34] Yuichi Sano and Ikko Hamamura. Quantum State Preparation for Probability Distributions with Mirror Symmetry Using Matrix Product States. In: *arXiv:2403.16729* (2024).

[35] Peng-Fei Zhou, Rui Hong, and Shi-Ju Ran. Automatically differentiable quantum circuit for many-qubit state preparation. In: *Phys. Rev. A* 104 (4 2021), p. 042601.

[36] Glen Evenbly and Guifré Vidal. Algorithms for entanglement renormalization. In: *Physical Review B* 79.14 (2009), p. 144108.

[37] Akel Hashim et al. Randomized Compiling for Scalable Quantum Computing on a Noisy Superconducting Quantum Processor. In: *Phys. Rev. X* 11 (4 2021), p. 041039.

[38] Simon J Evered et al. High-fidelity parallel entangling gates on a neutral-atom quantum computer. In: *Nature* 622.7982 (2023), pp. 268–272.

[39] Steven A Moses et al. A race-track trapped-ion quantum processor. In: *Physical Review X* 13.4 (2023), p. 041052.

[40] Rajeev Acharya et al. Quantum error correction below the surface code threshold. In: *arXiv:2408.13687* (2024).

[41] Marco Cerezo et al. Cost function dependent barren plateaus in shallow parametrized quantum circuits. In: *Nature communications* 12.1 (2021), pp. 1–12.

[42] Hao-Kai Zhang, Shuo Liu, and Shi-Xin Zhang. Absence of barren plateaus in finite local-depth circuits with long-range entanglement. In: *Physical Review Letters* 132.15 (2024), p. 150603.

[43] Samson Wang et al. Noise-induced barren plateaus in variational quantum algorithms. In: *Nature communications* 12.1 (2021), pp. 1–11.

[44] Guillermo González-García, Rahul Trivedi, and J. Ignacio Cirac. Error Propagation in NISQ Devices for Solving Classical Optimization Problems. In: *PRX Quantum* 3 (4 2022), p. 040326.

[45] Sergio Boixo et al. Characterizing quantum supremacy in near-term devices. In: *Nature Physics* 14.6 (2018), pp. 595–600.

[46] Frank Arute et al. Quantum supremacy using a programmable superconducting processor. In: *nature* 574.7779 (2019), pp. 505–510.

[47] Fernando G S L Brandão et al. Models of quantum complexity growth. In: *PRX Quantum* 2.3 (2021), p. 30316.

[48] Dominik Hangleiter and Jens Eisert. Computational advantage of quantum random sampling. In: *Rev. Mod. Phys.* 95.3 (2023), p. 035001.

[49] Hans J Briegel et al. Measurement-based quantum computation. In: *Nature Physics* 5.1 (2009), pp. 19–26.

[50] Matthew B Hastings and Tohru Koma. Spectral gap and exponential decay of correlations. In: *Communications in mathematical physics* 265 (2006), pp. 781–804.

[51] Jaš Bensa and Marko Žnidarič. Fastest local entanglement scrambler, multistage thermalization, and a non-Hermitian phantom. In: *Physical Review X* 11.3 (2021), p. 031019.

[52] Harper R. Grimsley, Sophia E. Economou, Edwin Barnes, and Nicholas J. Mayhall. An adaptive variational algorithm for exact molecular simulations on a quantum computer. In: *Nature Communications* 10.1 (2019).

[53] Matthew Fishman, Steven White, and Edwin Stoudenmire. The ITensor software library for tensor network calculations. In: *SciPost Physics Codebases* (2022), p. 004.

[54] Giacomo Torlai and Matthew Fishman. *PastaQ: A Package for Simulation, Tomography and Analysis of Quantum Computers*. 2020.

[55] Ulrich Schollwöck. The density-matrix renormalization group in the age of matrix product states. In: *Annals of Physics* 326.1 (2011), pp. 96–192.

[56] C. Schön et al. Sequential generation of matrix-product states in cavity QED. In: *Physical Review A* 75.3 (2007), pp. 1–10.

[57] Zhi-Yuan Wei, Daniel Malz, and J Ignacio Cirac. Efficient adiabatic preparation of tensor network states. In: *Physical Review Research* 5.2 (2023), p. L022037.

[58] Raban Iten et al. Quantum circuits for isometries. In: *Phys. Rev. A* 93.3 (2016), p. 32318.

[59] Péter Rakyta and Zoltán Zimborás. Approaching the theoretical limit in quantum gate decomposition. In: *Quantum* 6 (2022), p. 710.

[60] Lennart Bittel and Martin Kliesch. Training variational quantum algorithms is np-hard. In: *Physical Review Letters* 127.12 (2021), p. 120502.

[61] Lorenzo Piroli, Georgios Styliaris, and J. Ignacio Cirac. Quantum Circuits assisted by LOCC: Transformations and Phases of Matter. In: *Physical Review Letters* 127.22 (2021), p. 220503.

[62] A. Barenco et al. Elementary gates for quantum computation. In: *Phys. Rev. A* 52.5 (1995), pp. 3457–3467.