

Scalable quantum dynamics compilation via quantum machine learning

Yuxuan Zhang^{1,2,*}, Roeland Wiersema^{2,3,*}, Juan Carrasquilla^{4,†}, Lukasz Cincio^{5,‡} and Yong Baek Kim^{1,§}

¹*Department of Physics and Centre for Quantum Information and Quantum Control, University of Toronto, 60 Saint George Street, Toronto, Ontario M5S 1A7, Canada*

²*Vector Institute, W1140-108 College Street, Schwartz Reisman Innovation Campus, Toronto, Ontario M5G 0C6, Canada*

³*Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*

⁴*Institute for Theoretical Physics, ETH Zürich, 8093 Zurich, Switzerland*

⁵*Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*



(Received 19 November 2024; revised 5 June 2025; accepted 4 February 2026; published 7 May 2026)

Quantum dynamics compilation is an important task for improving quantum simulation efficiency: It aims to synthesize multiqubit target dynamics into a circuit consisting of as few elementary gates as possible. Compared to deterministic methods such as Trotterization, variational quantum compilation (VQC) methods employ variational optimization to reduce gate costs while maintaining high accuracy. In this work, we make a paradigm shift in VQC's framework design. First, we use a cost function from recently established out-of-distribution generalization results in quantum machine learning: By learning the action of a given many-body dynamics on a small dataset of product states, we can obtain a unitary circuit that generalizes to highly entangled states such as the Haar random states. Moreover, a key observation is that the low complexity in the data states allows us to use tensor-network methods to compress such time-evolved product states by exploiting their low entanglement features. Utilizing warm start methods, our approach exceeds state-of-the-art compilation results in both system size and accuracy in one dimension. For the first time, we extend VQC to systems to two-dimensional systems, demonstrating a significant resource advantage over standard Trotterization methods. The flexibility of our framework highlights the method's promise for advancing quantum simulation tasks on near-term quantum processors.

DOI: [10.1103/psqv-nq6d](https://doi.org/10.1103/psqv-nq6d)

I. INTRODUCTION

The simulation of quantum dynamics was among the initial applications of quantum computers envisioned by Richard Feynman over four decades ago [1]. Today, it remains one of quantum computers' most promising applications, as the exponential size of Hilbert space restricts classical computers. Apart from a pure demonstration of quantum advantage [2–11], the importance of quantum dynamics simulation is twofold: On the one hand, quantum dynamical processes are of broad interest for physicists in studying problems like the scattering in quantum field theory [12,13], molecular reactions in quantum chemistry [14,15], and thermalization and dynamical phase transition in many-body physics [16–18]; on the other hand, the faithful implementation of dynamics is at the core of executing quantum algorithms like quantum approximate optimization algorithm [19–23], quantum

phase estimation [24], and may even provide a quantum-enhancement to Markov chain Monte Carlo [25].

The conventional approach to simulating quantum dynamics, known as Suzuki-Trotterization [26,27], involves decomposing the unitary evolution generated by a many-body Hamiltonian into small time steps and approximating each step with products of local unitary gates. To achieve a high accuracy in the simulation, it is important that each slice is as accurate as possible. While Trotterization is widely used, it is well known that this type of slicing is far from optimal [28–36]: For example, randomly reordering Trotter terms can already lead to lower errors [37]. Such inefficiencies raise an important question: Can we develop more effective techniques that minimize circuit depth, a critical factor for near-term quantum simulations [38,39]?

Although there are analytical fast-forwarding methods for integrable models that can compress long-time simulations into circuits of constant depth [40,41], it has been proven that these methods do not apply to most generic Hamiltonians [42]. Given this challenge, variational quantum compilation methods present a promising alternative. By leveraging variational optimization, these methods can potentially reduce the gate count and circuit depth while maintaining the accuracy of the quantum simulation, which has been demonstrated by various hardware implementations [43].

The primary challenge in quantum circuit compilation is accurately evaluating the Hilbert-Schmidt product between the target unitary operation U and a parametrized quantum

*These authors contributed equally to this work.

†Contact author: jcarrasquill@ethz.ch

‡Contact author: lcincio@lanl.gov

§Contact author: ybkim@physics.utoronto.ca

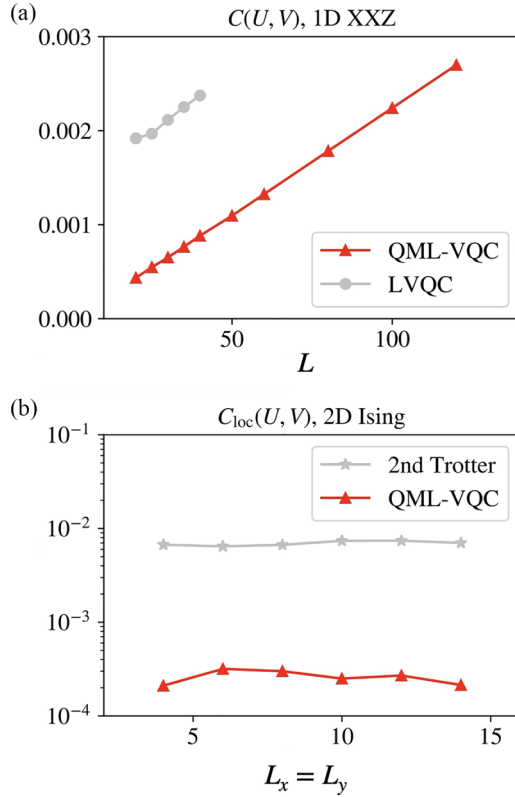


FIG. 1. Comparison with previous VQC and Trotterization results. (a) We compare the generalization risk of our QML-inspired VQC to a local-cost-function-based VQC result reported in Ref. [46]. The model is a Heisenberg chain ($h = 0, t = 0.5$) without the disorder term [as defined in Eq. (11)], and the circuit ansatz is a depth-10 brickwall circuit with charge-conserving gates. The LVQC data are extracted from Fig. 6 in Ref. [46] (up to $L = 40$), and the QML-VQC data are trained on five random product states. The linear growth of the HST cost function with L suggests that the dynamics are being learned in the thermodynamical limit and the finite-size effect is vanishing. (b) We compare the fidelity of 2D transversal field Ising model on a torus (no extra longitudinal field: $g = 0, t = 0.3$) with a widely used second-order Trotterization at $\tau = 2$. We use the Trotterization circuit's architecture and finetune gate parameters variationally. The state is represented with a projected pair entangled state and the contraction is conducted with by making use of belief propagation algorithm.

circuit (PQC) $V(\theta)$, which is crucial for ensuring that the compiled circuit closely approximates the desired quantum dynamics. First, computing the overlap with conventional methods, such as the Hilbert-Schmidt test (HST), is both classically and quantumly costly to compute (see Fig. 2 for a diagrammatic representation). In fact, for large system sizes, even explicitly writing U as a matrix becomes challenging. Next, variational algorithms that evaluate the full unitary cost face another challenge known as the barren plateau (BP) problem, since U and a randomly initialized V are likely to have exponentially vanishing overlaps. As such, early compiling papers [28–32,44,45] have either been limited by the system size studied, the total time of the dynamics, or focused on the time evolution of a few specific initial states. Additionally, to the best of our knowledge, all previous works

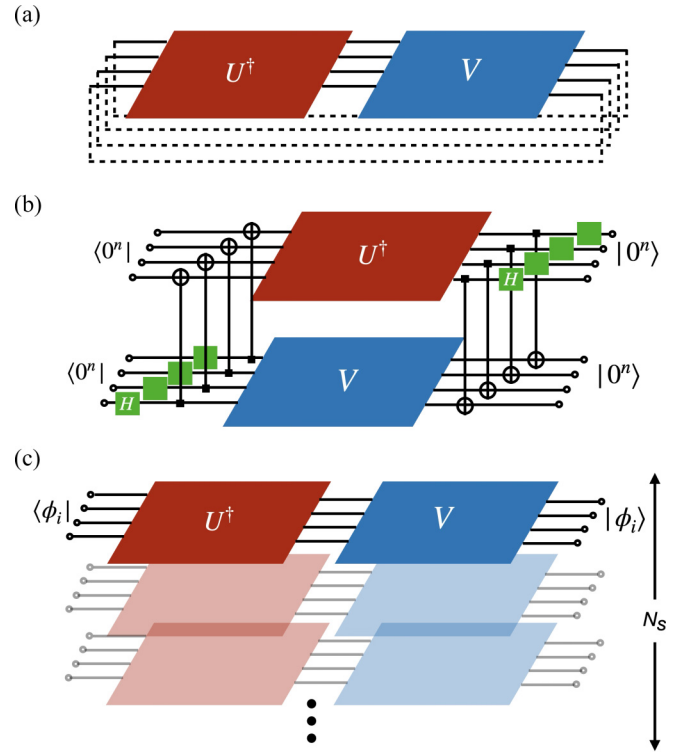


FIG. 2. Tensor-network diagrams for different VQC cost functions. (a) The most naive cost function for computing two unitaries involves computing $\text{Tr}[U^\dagger V]$. For classical simulations, $U^\dagger$ can be costly to write down as anMPO. (b) The HST circuit is introduced to compute $\text{Tr}[U^\dagger V]$ on a quantum computer and is sometimes considered as a classical cost function [46]. Green squares represent Hadamard gates. This circuit, however, requires doubling the number of qubits of the systems and introduces highly nonlocal gates. (c) In this work, we consider a QML-generated cost function [Eq. (5)]. The cost function avoids the nonlocality introduced by HST and has only a linear dependence on the number of samples N_s , which makes it suitable for both noisy intermediate-scale quantum (NISQ) implementations and numerical simulations. In this work, we focus on classical simulation as a primary tool for developing and testing our methods. However, it would be of future interest to extend these techniques to perform compilation on a quantum computer.

consider one-dimensional (1D) Hamiltonians with no mention of two-dimensional (2D) topologies. As shown in Fig. 1(b), our data-driven method allows us to compile circuits on ~ 120 qubits in 1D, and up to 14×14 in two dimensions, marking a significant step forward in what is possible with variational quantum compilation (VQC).

Some recent works address the challenge by leveraging translation invariance (TI) and locality of quantum systems [46–48]. However, the application of local HST-based cost functions can come with limitations: On the one hand, it relies on an explicit light-cone structure of the physical model and could fail for certain long-range interaction Hamiltonians even when their entanglement growth is slow [49]; on the other hand, not using a global cost function could result in a biased and suboptimal training result, as Fig. 1(a) demonstrates. Others proposed alternative cost functions such make use of the conservation of physical quantities [50], yet these

works fail to provide any guarantee on the true unitary fidelity, as it is easier to score high on some correlation function benchmark than simulating the dynamics with high fidelity. In this work, we overcome these limitations by leveraging recent advances in quantum machine learning [52–56] and making a paradigm shift by combining a QML cost function with powerful numerical simulation tools. The key insight is to rephrase the quantum compilation problem as a supervised machine learning task: Suppose, with a PQC, V , we want to learn some unitary dynamics U , which is provided by a black box that outputs the time-evolved state for any input. The question then arises: What set of input states should we use, and how much data are required to accurately learn U ? It turns out that the PQC only requires very few input states randomly drawn from the set of all random product states. Interestingly, if the trained PQC can learn the dynamics with small errors, then it would reproduce the dynamics accurately on samples drawn not only from the *same* distribution of random product states but also for *other* distributions like the Haar random states: In the machine learning literature, these generalization errors are referred to as the “in-distribution” and “out-of-distribution” generalization risk [57,58], concepts we adopt in this context. Most importantly, the out-of-distribution generalization risk effectively measures the Hilbert-Schmidt product between U and V , which is the ultimate target of VQC.

Previous works [52–55,59] have laid the groundwork for employing QML in VQC. However, their numerical simulations were limited to relatively small system sizes. We resolve this issue by employing tensor-network techniques: Focusing on short-time quantum dynamics for locally interacting Hamiltonians, we observe that the training process involves computing overlaps between low entanglement states only, which can be computed efficiently with tensor network techniques. Further, inspired by concepts in machine learning [60], we devised initialization strategies that effectively mitigate the barren plateau problem, which can be otherwise severe even at constant circuit depth for the cost function we study [61]. The success of our warm start methods aligns with recent analytical results [62,63] (see Appendix D 1 for more details). The combination of tensor-network techniques with a QML-based approach allows us to explore a wide range of system sizes with enhanced performance. Benchmarking with a 1D Heisenberg model, for the case where $n = 40$ and $t = 0.5$, at the same variational circuit depth ($\tau = 10$), our method decreases unitary infidelity from previous local-cost-function-based VQC (LVQC) results, which is confirmed with a tensor-network implementation of the HST algorithm.

For large systems, calculating the infidelity directly between U and V can be computationally hard, especially in higher dimensions. To verify the out-of-distribution generalization of the variationally trained PQCs, we repeatedly applied the circuits to fixed initial states to perform long dynamical simulations. Comparing the time-evolution result with dense Trotterization circuits, we find the VQC-generated circuit accurately captures physical correlations. Finally, a resource comparison with optimized Trotterization showed that VQC circuits could be orders of magnitude more accurate (see Table I), reinforcing the effectiveness of VQC in the dynamical simulation of large, complex quantum systems across one and higher dimensions.

TABLE I. For the nonintegrable Ising model [Eq. (12) at $g = -1$ and $\kappa = 0$] on a 2D rectangular strip and a 1D chain, we compare the generalization risk on random product input states between quantum circuits derived from VQC and those from optimized Trotterization, using the same CNOT gate budget. See Fig. 7 for a detailed comparison.

n	t	CNOTS	VQC (this work)	Trotter [51]
4×10	0.5	550	5.7×10^{-4}	4.2×10^{-3}
80	0.5	950	1.0×10^{-5}	1.8×10^{-2}

II. RESULTS

We consider an n -qubit system on a Hilbert space $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$, which has dimension $N = 2^n$. Quantum computers are expected to excel the most in simulating dynamical problems, where one is interested in simulating the time evolution $U(t) = \exp -iHt$ of some state $|\psi\rangle$ generated by some Hamiltonian H . Throughout the work, we focus on 1D nearest-neighbor and next-nearest-neighbor Hamiltonians.

In all hardware, quantum operations are executed through a set of minimum operations called the universal gate set [64], usually consisting of single- and two-qubit operations only. Thus, it is a fundamental problem to find a scalable and efficient way of compiling U into products of elementary gates. Conventionally, this could be done deterministically, as canonically exemplified by the Trotterization method [26]. However, analytical and numerical evidence has shown that such deterministic methods are far from optimal. In fact, previous results [28,29,42,65] have established evidence that the time evolution of various types of Hamiltonians can be compressible.

A. The QML cost function for learning quantum dynamics

For brevity, we now state the main ingredients for our optimization algorithm but will defer the mathematical details to Appendix A. Given two unitary operators U and V , the cost function

$$C(U, V) = 1 - \mathcal{F}(U, V) = 1 - \frac{1}{N^2} |\text{Tr}[U^\dagger V]|^2 \quad (1)$$

is a measure of distance between U and V where we ignore a global phase. Note that $0 \leq C(U, V) \leq 1$ for all U, V .

Consider a parametrized unitary circuit $V(\theta)$ with variational parameters $\theta = \{\theta_1, \dots, \theta_d\} \in \mathbb{R}^d$. We want to approximate U with $V(\theta)$, for the case where $U = \exp -iHT$ and H is some local Hamiltonian. We can achieve this by solving the optimization problem

$$\min_{\theta} C(U, V(\theta)). \quad (2)$$

Note that the cost function $C(U, V(\theta))$ is in many cases prohibitively expensive to evaluate, since for a large number of qubits, the $2^n \times 2^n$ unitary matrices quickly become too large to store in memory. However, recent work in QML theory suggests that one can drastically simplify this optimization procedure [53,55].

First, we can show that the optimization problem of Eq. (2) is equivalent to minimizing the *expected risk*

$$\min_{\theta} R_{\mathcal{P}_{\text{Haar}}}(\theta) = \mathbb{E}[1 - |\langle \psi | U^\dagger V(\theta) | \psi \rangle|^2]_{\psi \sim \mathcal{P}_{\text{Haar}}}, \quad (3)$$

if the test states are drawn from the n -qubit Haar random distribution $\mathcal{P}_{\text{Haar}} = \mathcal{S}_{\text{Haar}_n}$ [66]. For larger systems, however, the Haar random states require exponential bond dimensions to write down classically, so we instead simplify the cost function further with three observations:

(1) Quantum dynamics at a short time has low circuit complexity.

(2) Circuits of low complexity can be learned sample-efficiently with a high generalization accuracy on product states (in-distribution generalization), which are classically easy to compute.

(3) Most importantly, circuits learned with product states will also generalize to Haar random states (out-of-distribution generalization).

The main learning result that establishes the second and third points comes from Refs. [53,55], which we restate here:

Proposition 1. (*Equivalence of in- and out-of-distribution risks*). If a PQC is trained on samples drawn from the random product state, $\mathcal{Q} = \mathcal{S}_{\text{Haar}_1^{\otimes n}}$, then

$$\frac{1}{2} R_{\mathcal{P}_{\text{Haar}}}(\theta) \leq \frac{N}{N+1} R_{\mathcal{Q}}(\theta) \leq R_{\mathcal{P}_{\text{Haar}}}(\theta). \quad (4)$$

We call $R_{\mathcal{Q}}$ the *in-distribution risk* and $R_{\mathcal{P}_{\text{Haar}}}$ the *out-of-distribution risk*. Proposition 1 is a powerful result, which states that any PQC trained on the random product states, $\mathcal{Q} = \mathcal{S}_{\text{Haar}_1^{\otimes n}}$ (we use this convention throughout the manuscript), will have an in-distribution risk that is within a factor 1/2 of the out-of-distribution risk over the Haar measure. Of course, in a real experiment, one could only obtain a finite number of samples to estimate the expected risk. This empirical estimation is denoted as the *empirical risk* [55]:

$$C_{\mathcal{D}_{\mathcal{Q}}}(\theta) = 1 - \frac{1}{N_s} \sum_{i=1}^{N_s} |\langle \psi_i | U^\dagger V(\theta) | \psi_i \rangle|^2, \quad (5)$$

for a dataset $\mathcal{D}_{\mathcal{Q}} = \{|\psi_i\rangle\}$ of N_s states $|\psi_i\rangle \sim \mathcal{Q}$. As we increase $N_s \rightarrow \infty$, the empirical risk converges to the expected risk, and for finite N_s , the empirical risk still yields an upper bound, as we detail in Appendix A. For a fixed PQC, the tightness of this bound scales as $O(N_s^{-1/2})$, although in practice we find that a small number of states ($N_s < 50$) suffices. We have reduced the optimization of Eq. (1) to a QML-based cost function

$$\min_{\theta} C_{\mathcal{D}_{\mathcal{Q}}}(\theta). \quad (6)$$

A complementary, locality-aware objective we also use is

$$C_{\text{loc}, \mathcal{D}_{\mathcal{Q}}}(\theta) := 1 - \frac{1}{N_s n} \sum_{i=1}^n \sum_{|\psi\rangle \in \mathcal{D}_{\mathcal{Q}}} F(\rho_i^U(\psi), \rho_i^{V(\theta)}(\psi)), \quad (7)$$

where $F(\cdot, \cdot)$ is the Uhlmann fidelity, and

$$\rho_i^U(\psi) = \text{Tr}_{\bar{i}}[U|\psi\rangle\langle\psi|U^\dagger], \quad (8)$$

$$\rho_i^{V(\theta)}(\psi) = \text{Tr}_{\bar{i}}[V(\theta)|\psi\rangle\langle\psi|V(\theta)^\dagger] \quad (9)$$

corresponds to the local density matrix on qubit i that we obtain by tracing out all $\bar{i} \neq i$ qubits. In practice, $C_{\text{loc}, \mathcal{D}_{\mathcal{Q}}}$ has lower variance and cheaper estimation than the global cost (6). The two cost functions are also related by the following relationship [29,55]:

$$\frac{1}{2} R_{\text{loc}}(\theta) \leq C_{\text{glob}}(\theta) \leq 2n R_{\text{loc}}(\theta). \quad (10)$$

B. Numerical methods considered in this work

1. TEBD on MPS for (quasi-)1D training

Compared with Eq. (1), the cost function above, as depicted in Fig. 2(c), is significantly more efficient to evaluate on both classical and quantum computers. To see this, first observe that the fidelity function in Eq. (1) can be computed by either performing a direct contraction or casting the fidelity as an equivalent HST circuit, as illustrated by Figs. 2(a) and 2(b), respectively. For locally interacting Hamiltonians in one dimension and for narrow 2D strips (also denoted as quasi-1D), matrix product methods offer a natural route to scalable simulation. One may either compress the time-evolution operator as a matrix product operator (MPO) [67–72] [as in Fig. 2(a)] or evolve and compress the state itself as a matrix product state (MPS) [Figs. 2(b) and 2(c)].

In practice, two considerations favor the data-driven, state-only cost function in our setting: (1) MPO bond dimensions typically grow faster than those required to represent $U|\psi\rangle$ for product inputs at the same accuracy and (2) Hilbert-Schmidt tests of overlaps entail n ancillas together with long-range Bell-pair entanglers, which are costly on local hardware. Leveraging the QML cost in Eq. (5), we therefore evaluate $U|\psi_i\rangle$ directly and compress the result as a high-precision MPS using time-evolving block decimation (TEBD) [67,68]. We employ a Trotterized evolution with nearest-neighbor two-site gates and singular value decomposition (SVD)-based truncation after each gate. Throughout, we fix the discarded-weight tolerance to 10^{-11} and adapt the bond dimension χ accordingly. For many 1D and quasi-1D quenches, the entanglement entropy grows linearly in time, implying $\chi \sim \exp(O(t))$ [73]; consequently, time-evolved states at $t = O(\log n)$ remain asymptotically efficient to compute. This regime is precisely where the combination of TEBD with our learning objective provides an accurate and scalable surrogate for exact contraction in the contractions of Figs. 2(b) and 2(c).

2. Belief propagation on PEPS for 2D training

To target unitaries on general graphs beyond 1D chains, we represent the random product training states and their evolved images with a projected entangled-pair state (PEPS). Exact PEPS contractions are #P-hard in general [74,75], so we resort to approximate, message-passing-based contraction via quantum belief propagation (BP) [76–78]. In brief, BP introduces local “cavity” channels called “messages” on edges and iteratively enforces consistency of single-site reduced states with their neighborhoods, corresponding to stationary conditions of a Bethe free energy on the interaction graph. BP can be

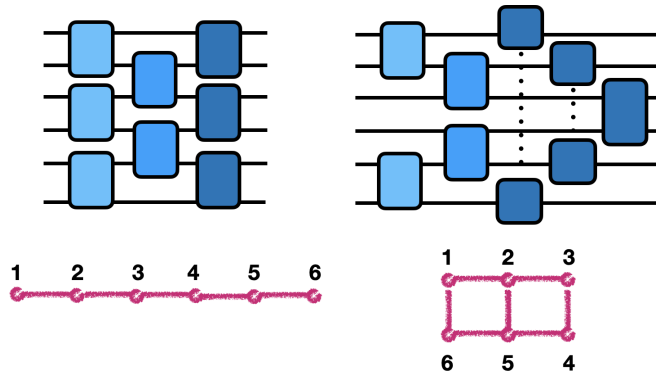


FIG. 3. Geometry of the PQC. 1D (left) and 2D (right) ansatz used for the PQCs are shown here, both with depth $\tau = 3$. For quasi-1D treatment of 2D, we use a “snake” indexing. Each block represents an $SU(4)$ unitary and unitaries with the same color are considered within the same circuit layer. For the TI circuit ansätze, we set the parameters in each layer equal to each other.

done accurately for states with treelike correlations [79]; on loopy lattices, it is a controlled heuristic that is accurate when correlations are locally treelike or when loop corrections are weak [80].

To compute expectation values of the form $\langle \psi | U^\dagger O U | \psi \rangle$, we follow the *mixed picture* of Ref. [81]: factor the desired time-evolution circuit as $U = U_H U_S$, evolve the state forward in the Schrödinger picture $|\psi_S\rangle := U_S |\psi\rangle$, and push the remaining unitary onto the observable in the Heisenberg picture $O_H := U_H^\dagger O U_H$, so that the target contraction reads $\langle \psi_S | O_H | \psi_S \rangle$. Both the PEPS and PEP operator (PEPO) generated by these partial evolutions are then compressed using 2-norm belief propagation [82–84]. The final PEPS-PEPO-PEPS contraction is then evaluated by the so-called “lazy 1-norm BP,” which approximates the overlap via the Bethe free-entropy built from the converged messages [85]. Choosing the split to balance entanglement growth on state and operator sides enables stable estimation at effective bond dimensions far beyond exact contraction [81], up to $\chi = 64$, which is far beyond what 1D tensor-network (TN) could represent in reality. By making use of the local cost function of Eq. (7), we successfully improve the accuracy for the dynamics of the 2D transverse field Ising model for systems up to 14×14 [see Fig. 1(b)] when compared to standard Trotterization.

C. Training and generalization

Throughout our work, we parametrize the ansatz circuit $V(\theta)$ with a brickwall structure of general nearest-neighbor $SU(4)$ gates, as shown in Fig. 3. Since the gradient of the cost $C_{\mathcal{D}_Q}(\theta)$ will be exponentially vanishing in the system size n [61,86,87], it is crucial that we give the optimization a “warm start” by initializing the variational gates in a regime of the parameter space where the gradient is barren-plateau-free [62,63,88]. Utilizing the structure of quantum dynamics, we outline and compare three different warm start methods to achieve this in Appendix D 1.

To assess the in-distribution generalization properties of the learned circuit $V(\theta)$, we perform training via gradient

ALGORITHM 1. Scalable variational compilation.

```

Input:  $N_s, H, t, V(\theta)$ 
for  $i \in [1, N_s]$  do
     $|\phi_i\rangle \sim \mathcal{S}_{\text{Haar}_1^{\otimes n}}$ 
     $|\psi_i\rangle \leftarrow \text{TEBD}(H, t, |\phi_i\rangle)$ 
end
step  $\leftarrow 0$ 
while step  $< 1000$  do
     $C_{\mathcal{D}_{\text{Train}}}(\theta) \leftarrow 1 - \frac{1}{N_s} \sum_{i=1}^{N_s} |\langle \psi_i | V(\theta) | \phi_i \rangle|^2$ 
     $\delta\theta \leftarrow \nabla C_{\mathcal{D}_Q}(\theta)$ 
     $\theta \leftarrow \theta - \delta\theta$ 
    step  $\leftarrow$  step + 1
end
for  $i \in [1, 100]$  do
     $|\phi_i\rangle \sim \mathcal{S}_{\text{Haar}_1^{\otimes n}}$ 
     $|\psi_i\rangle \leftarrow \text{TEBD}(H, t, |\phi_i\rangle)$ 
     $C_{\mathcal{D}_{\text{Test}}}(\theta) \leftarrow 1 - \frac{1}{N_s} \sum_{i=1}^{N_s} |\langle \psi_i | V(\theta) | \phi_i \rangle|^2$ 
end

```

descent on a dataset $\mathcal{D}_{\text{Train}}$ with N_s states $|\phi_i\rangle$ drawn from $\mathcal{S}_{\text{Haar}_1^{\otimes n}}$. We apply the TEBD algorithm to obtain the numerically exact dynamics $e^{iH} |\phi_i\rangle$. We use a second-order Trotterization scheme with time slices $\delta t = t/20$ for all t . We repeat this procedure to generate test dataset $\mathcal{D}_{\text{Test}}$ of 100 unseen states (also drawn from $\mathcal{S}_{\text{Haar}_1^{\otimes n}}$). This validation set will help us track how well the circuit is performing on in-distribution data. In Algorithm 1, we show the pseudocode for our optimization procedure. In practice, we use an early stopping criterion to terminate the optimization if the cost has converged, and we use adaptive moment estimation (ADAM) [89] for the gradient descent update step. We provide code written in QUIMB [90] and PyTorch [91] to reproduce our results at Ref. [92].

We examine the accuracy of training for two widely studied physical models: A Heisenberg model with on-site disorder [93–95]

$$H_{\text{Heisenberg}} = \sum_{\langle i,j \rangle} (\sigma_i^x \sigma_j^x + \sigma_i^y \sigma_j^y + \sigma_i^z \sigma_j^z) + \sum_{i=1}^n h_i \sigma_i^z, \quad (11)$$

where $h_i \sim \text{i.i.d.}[-h, h]$, and the transversal field Ising model with uniform on-site and next-nearest neighbor perturbations:

$$H_{\text{Ising}} = - \left(\sum_{\langle i,j \rangle} \sigma_i^x \sigma_j^x + \sum_{i=1}^n \sigma_i^z \right) + g \sum_{i=1}^n \sigma_i^x + \kappa \left(\sum_{\langle i,j \rangle} \sigma_i^z \sigma_j^z + \sum_{\langle\langle i,j \rangle\rangle} \sigma_i^x \sigma_j^x \right). \quad (12)$$

We start by examining the in-distribution generalization for 1D systems with open boundary conditions. Three trends emerge when examining the three panels shown in Fig. 4: First, fixing the ansatz to be nontranslation invariant, the difference between the objective function and the in-distribution risk decreases as one increases the number of samples. Moreover, perhaps surprisingly, panel (b) shows that, despite the open boundary condition of the 1D chains, the translation invariant PQC’s generalization outperforms that of the

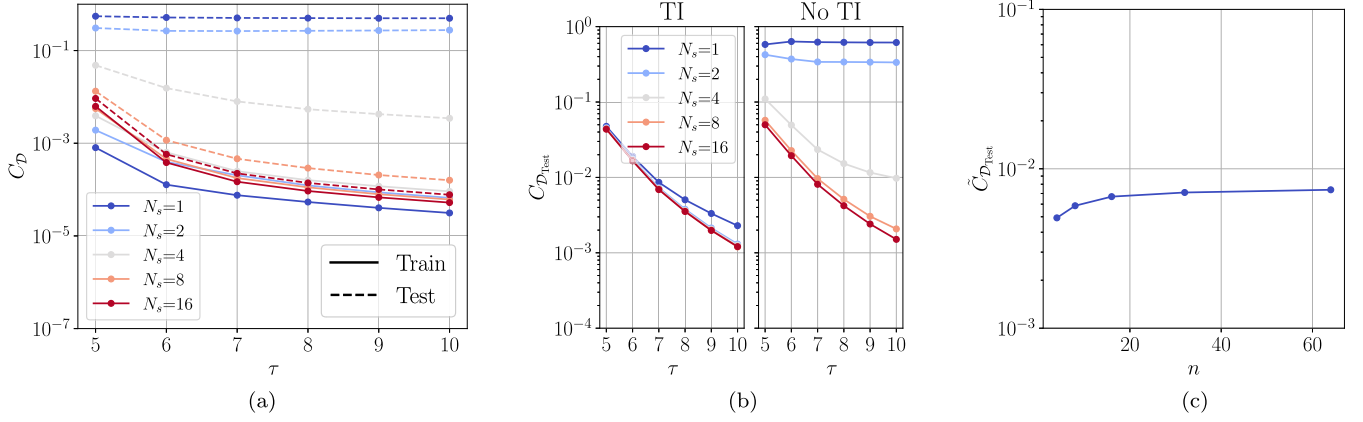


FIG. 4. In-distribution generalization. (a) Ising, $g = -1$, $\kappa = 0$, $n = 80$, $t = 1$. We first train a PQC with no translation invariance on N_s samples drawn i.i.d. from the random product state ensemble and test it on 100 states drawn i.i.d. from the same distribution. The training and testing loss results are shown on the solid and dashed lines, respectively. As we increase the circuit depth τ , the training and testing loss both get lower. As we increase N_s , we see that the training loss curves converge, and the testing loss curve approaches the training one. This indicates that an increased number of data samples improves the in-distribution generalization accuracy. (b) Heisenberg, $h = 0$, $n = 60$, $t = 0.5$. We compare the testing loss results between PQCs TI (left) and ones without (right) (as defined in Fig. 3). The faster convergence of the training loss curves of the translation invariant PQCs can be reasoned from its smaller number of variational parameters. Additionally, despite the Hamiltonian's open boundary condition, both PQCs converge to similar test losses. (c) Ising, $g = 0$, $\kappa = 0.2$, $t = 1$. Fixing the PQC depth at $\tau = 4$, we now train our PQC on various system sizes. We observe that, as we increase the system size, the “per-site” cost, $\tilde{C}_{D_Q}$, as defined in Eq. (13) remains relatively unchanged for $n > 20$. This shows our VQC results are scalable.

nontranslation invariant one due to its fewer parameters. It appears that the extra parameters of the non-TI PQC did not increase its accuracy. Finally, fixing the total evolution time and PQC depth while varying the system size, we plot a “per-site” risk:

$$\tilde{C}_{D_Q} = 1 - (1 - C_{D_Q})^{1/n}, \quad (13)$$

against n to gauge the fidelity loss by adding more lattice sites (otherwise, the fidelity will decrease exponentially even if one trivially combines quantum states together). We find that the per-site risk function flattens already at $n > 20$, showing that the QML method can be extended to larger systems even in the presence of next-neighbor interaction.

Proposition 1 gives a theoretical lower bound for out-of-distribution generalization: For any Haar random ensemble, the risk of generalization [which is equivalent to the unitary cost $C(U, V)$] is at most twice the in-distribution risk. One could verify this by numerically computing $C(U, V)$ for smaller systems, but this quickly becomes infeasible as explained in Sec. II A. As we show in Appendix B, one could accurately simulate long-time dynamics for $T \gg t$ by repeatedly applying the compiled circuit, which is similar to how Trotterization is executed in practice [96,97]. To demonstrate that our algorithm already yields useful results for long-time quantum simulation, we test out-of-distribution generalization by using our compiled unitary to simulate the long-time quench physics of the 1D and 2D systems (see also, e.g., Ref. [98] for an experimental demonstration in long-time, small systems). We also assess the quality of the final compilation using an HST between the model $V(\theta)$ and a high-accuracy Trotterization of the target U .

In Fig. 5, we compare the dynamics generated by the Heisenberg model with and without disorder, starting with a

product state $|0 \dots 10 \dots 10 \dots \rangle$, i.e., a state analogous to two localized particles (the two spins up) in a vacuum. Comparing the magnetization plots between a TEBD simulation and the VQC circuit, we find that the VQC result accurately captures the localization of the particles (spins up) induced by disorder. In panel (c), we also illustrate the bound of Proposition 1 and show that the true infidelity $C(U, V)$ is close to the test risk C_{D_Q} in practice, making it a suitable cost function for VQC.

Similarly, we compile the Heisenberg model on a cylinder where the initial state is a strip of spins down surrounded by spins up that is evolved by time evolution under the Heisenberg Hamiltonian (see Fig. 6). This simulates a system of strongly interacting bosons expanding on an optical lattice, similar to the theoretical and experimental studies in Refs. [99–103]. We demonstrate the quench dynamics of such a system in Fig. 6.

D. Resource comparison with Trotterization

One immediate application for VQC is to reduce the number of elementary gates for dynamical simulations compared to Trotterization. Fixing the interaction graph, a variety of methods exist to Trotterize a Hamiltonian [104,105]. Nevertheless, for Hamiltonians with a particular graph structure and a small number of noncommuting terms, finding the optimal Trotterization at a certain p is possible. For the 1D nearest-neighbor geometry we consider in this paper, we used the optimized parameters in Ref. [51] for two and three noncommuting Hamiltonian terms, which we use for a fair comparison to our variational compilation scheme. We compare the estimated gate cost between VQC and Trotterization circuits by implementing Ising-Hamiltonian dynamics with various parameters and geometries. We report the gate cost in the number of nearest-neighbor CNOT gates in Table II.

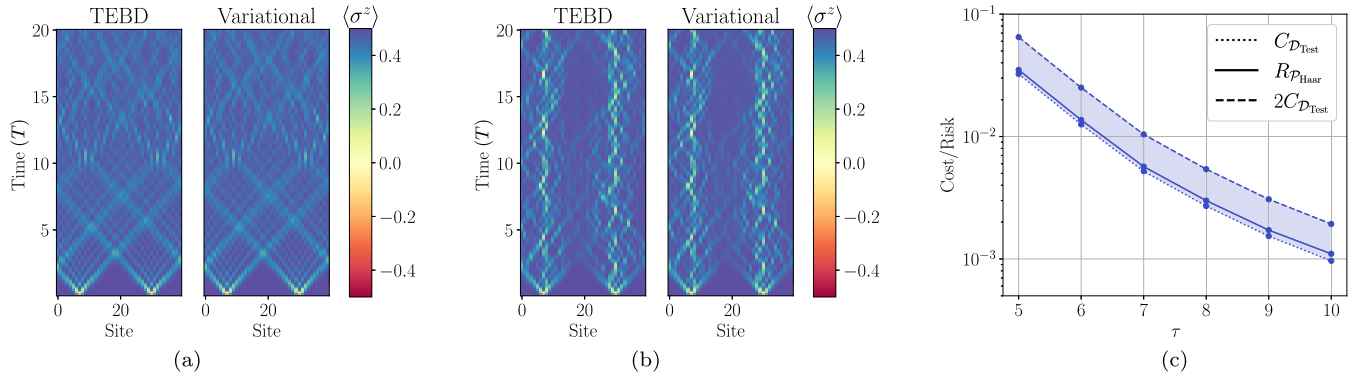


FIG. 5. Out-of-distribution generalization. To test out out-of-distribution generalization, we repeatedly apply a VQC trained at time $t = 0.1$ to perform a long-time evolution of a computational basis state for total evolution time $T = 20$. We see that we can obtain accurate dynamics at long times for an $n = 40$ (a) Heisenberg, $h = 0$. Heisenberg chain and (b) Heisenberg, $h = 1$. Heisenberg chain with random fields on each spin. The fidelity of the variational simulation drops from about ≈ 0.99999 to ≈ 0.99 between $T = 0.1$ and $T = 20$, respectively. (c) Heisenberg, $h = 0$. Using the HST, we compare the actual unitary infidelity to the upper and lower bounds of Proposition 1 for the $n = 40$ Heisenberg model with $t = 0.5$. We see that the test cost is close to the true Haar random risk in practice.

Figure 7 shows that for the same test risk, the QML-compiled circuit requires significantly fewer CNOT gates as compared to Trotterization. In the 1D case with nearest-neighbor interactions only, the PQC's generalization risk already drops below 10^{-4} with merely 500 cnots, whereas the $p = 6$ Trotterization reaches similar risk with a gate count of over 3000. Turning on next-nearest-neighbor interactions and setting $t = 1$, the variational circuit matches the $p = 6$ Trotterization result with 1/5th of the gate cost. In 2D, the resource cost difference between VQC and Trotterization shrinks, which could be the result of the faster scrambling in 2D geometries or due to the finite-size effect of our system. Nevertheless, compared with $p = 4$ Trotterization, the PQC-learned circuit still reduces the generalization risk by a factor of 7 while using fewer CNOT gates.

III. DISCUSSION

In this work, we develop a scalable quantum dynamics compilation algorithm for both 1D and quasi-1D systems by reformulating the compilation task as a machine learning problem, employing tensor-network techniques, and introducing initialization strategies. Extensive simulations demonstrate that the learned unitary circuits can accurately reproduce long-time dynamics and significantly reduce computational costs compared to highly optimized Trotterization methods. Our integration of quantum compilation with simulation techniques holds great potential for near-term quantum

TABLE II. A list of the (nearest-neighbor) CNOT costs of gates considered in this work. For the Ising Hamiltonian, the XX gate costs only 2 cnots instead of 3. The nonlocal gate requires us to use a swap, which triples the CNOT count.

Gate	No. of cnots
Swap	3
Nearest-neighbor XX gate	2
Next nearest-neighbor XX gate	6
SU(4)	3

computing applications. It is worth noting that the QML cost function is an improvement over the HST-based cost function itself, so every VQC technique developed for HST-based cost functions almost automatically applies to our method, except

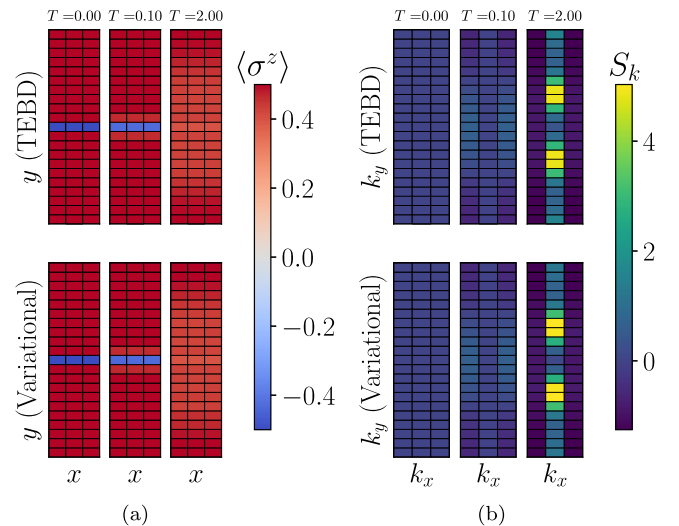


FIG. 6. Dynamics of 2D Heisenberg model on a cylinder with a variationally compiled circuit. Here, we target the simulation of the expansion dynamics of ultracold hardcore bosons on a quasi-2D optical lattice after suddenly turning off a strong confining potential [100,102]. We use the mapping between the hard core boson and spin operators, $S^+ = b^\dagger$, $S^- = b$, $S^z = b^\dagger b - 1/2$. We compile the time evolution of a $t = 0.1$ Heisenberg model on a 3×21 lattice with periodic boundaries in the x direction. Then, we quench the state $|0 \dots 111 \dots 0\rangle$, where the $|1\rangle$ states are exactly in the middle of the lattice corresponding to the confined bosonic cloud. (Left) Magnetization; In real space, we see that the hardcore bosons rapidly expand away from the center of the lattice in both directions. (Right) Structure factor; We calculate the structure factor $S_k = \sum_r e^{ir \cdot k} \langle S_j^+ S_j^- \rangle$ for all time steps and show that the dynamics of the momentum distribution exhibits two peaks corresponding to the expanding wings moving up and down in the lattice at roughly constant velocity.

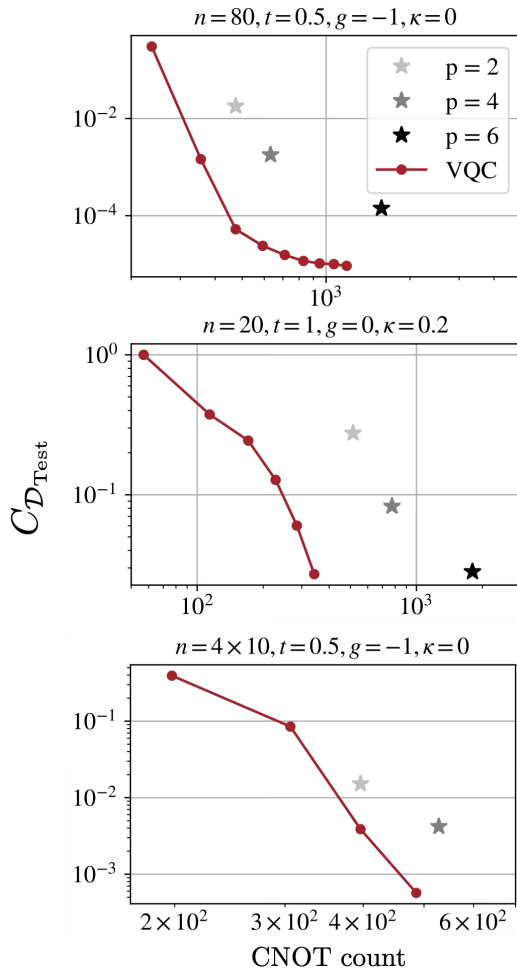


FIG. 7. A resource comparison. Here, we show a resource and accuracy comparison between implementing quantum dynamics with VQC and optimized Trotterization [51] of various orders p [106], tested in 1D (top and middle) and 2D (bottom) Ising models. Compared to the Trotterization circuits at the same nearest-neighbor CNOT cost, the VQC circuit would generate circuit with significantly lower generalization risk on random product states, and thus have lower unitary infidelity.

that the cost function can be evaluated much more efficiently. This shows the versatility of our method.

Our work also opens up several promising directions for future research. The first natural question is how to extend our method to larger system sizes, particularly in 2D and above. One possibility is to adopt a local update scheme, where each layer of the quantum circuit is optimized sequentially, similar to the approach used in density matrix renormalization group algorithms [107] for solving ground states. However, as we demonstrate in Appendix D 2, local updates can sometimes struggle to find the global minimum of the objective function. Nevertheless, the evaluation of the cost function could be improved by optimizing the contraction order of the tensor network. On the other hand, instead of the quasi-1D treatment, one may consider representing the time-evolved state with methods that are more natural to higher physical dimensions such as the 2D isometric tensor networks [108]. In Ref. [109], we show that the QML-VQC framework can also be used in the

Heisenberg representation by making use of Pauli Propagation techniques [2]. Finally, we plan a more in-depth study of the belief propagation method for VQC in a future work. Our algorithm can directly benefit from the incorporation of these tensor network techniques.

Next, symmetries naturally emerge in physical systems, as illustrated by the U(1)-conserving Heisenberg model considered in this work. An important question is whether these symmetries can be exploited to further enhance compilation quality. Recent advancements in variational quantum computing have explored incorporating symmetries when constructing circuit ansätze [110–112], but can such symmetries be leveraged in the context of compilation? Intuitively, learning the dynamics within a specific charge sector seems more straightforward than learning the entire unitary, as less information is to be captured. This raises the next important question: What is the appropriate ensemble to use for training? To our initial surprise, as proven in Appendix C, we show that training samples for Heisenberg dynamics cannot be generated by any U(1)-conserving random circuit with depth less than n , as such circuits fail to form a local scrambling ensemble, even within the desired charge sector. This contrasts sharply with the non-charge-conserving case, where a depth-1 random circuit suffices. This observation leads to an open problem: Are there symmetric Hamiltonians whose training samples could be generated by a shallow random circuit? If so, can they be trained on fewer samples?

Finally, since quantum computation can be implemented on various platforms with different gate sets and connectivities, experimentalists must tailor their circuits to specific hardware constraints and capabilities. For a device with nearest-neighbor coupling, dynamical simulation of long-range interactions using Trotterization often requires additional compilation steps, leading to significant overhead due to the extensive use of swap gates. In contrast, PQCs offer greater flexibility by allowing the circuit to be trained with a desired architecture and gate set. In this work, we only considered SU(4) gates with 1D and 2D brickwall connectivities, but it is of great interest to explore PQCs designed specifically for various platforms. For example, could certain quasilocal quantum dynamics benefit significantly from all-to-all connectivity, as in trapped ion quantum systems [113]? Can variational algorithms be leveraged to optimize the depth of T gates in a Clifford + T circuit [114]? Optimizing PQCs for such tailored tasks could further enhance performance on specific hardware architectures.

ACKNOWLEDGMENTS

We would like to thank Dima Abanin, Matteo D’Anna, Hong-Ye Hu, Tim Hsieh, Roger Luo, Andrew Potter, and Yifan Zhang for their insightful discussions. R.W. would like to thank Max Hunter Gordon for his help in 2022 during the Quantum Computing Summer School in Los Alamos. J.C. acknowledges support from the Shared Hierarchical Academic Research Computing Network (SHARCNET), Compute Canada, and the Canadian Institute for Advanced Research (CIFAR) AI chair program. Y.Z., J.C., and Y.B.K. were supported by the Natural Science and Engineering Research Council (NSERC) of Canada. Y.Z. and Y.B.K.

acknowledge support from the Center for Quantum Materials at the University of Toronto. Y.Z. was further supported by a CQIQC fellowship at the University of Toronto, and in part by grant NSF PHY-2309135 to the Kavli Institute for Theoretical Physics (KITP). L.C. was supported by Laboratory Directed Research and Development program of Los Alamos National Laboratory under Project No. 20230049DR. Resources used in preparing this research were provided, in part, by the Province of Ontario, the Government of Canada through CIFAR, and companies sponsoring the Vector Institute [115]. The simulation code in this work can be found at Ref. [92].

DATA AVAILABILITY

The data that support the findings of this article are openly available [92].

APPENDIX A: LEARNING DYNAMICS VIA OUT-OF-DISTRIBUTION GENERALIZATION

In this Appendix, we review the key quantum machine learning theorems used in the main text. We start with the Hilbert-Schmidt norm,

$$\langle A, B \rangle_{\text{HS}} := \frac{1}{N} \text{Tr}[A^\dagger B], \quad (\text{A1})$$

for $A, B \in \mathbb{C}^{N \times N}$. We see that the norm decomposes as follows:

$$\begin{aligned} \|A - B\|_{\text{HS}}^2 &= \|A\|_{\text{HS}}^2 + \|B\|_{\text{HS}}^2 - \langle A, B \rangle_{\text{HS}} - \langle B, A \rangle_{\text{HS}} \\ &= \|A\|_{\text{HS}}^2 + \|B\|_{\text{HS}}^2 - 2\Re[\langle A, B \rangle_{\text{HS}}]. \end{aligned} \quad (\text{A2})$$

Note that for a unitary U , we have $\langle U, U \rangle_{\text{HS}} = \|U\|_{\text{HS}}^2 = N$. Hence, for two unitary matrices U and V , we find

$$\frac{1}{2N} \|U - V\|_{\text{HS}}^2 = 1 - \frac{1}{N} \Re[\langle U, V \rangle_{\text{HS}}]. \quad (\text{A3})$$

Note that if we have $V = e^{i\phi}U$, we get

$$\frac{1}{2N} \|U - V\|_{\text{HS}}^2 = 1 - \Re[e^{i\phi}]. \quad (\text{A4})$$

If we choose the to ignore the global phase, a more natural cost function would be

$$C(U, V) = 1 - \frac{1}{N^2} |\text{Tr}[U^\dagger V]|^2. \quad (\text{A5})$$

We now take V to be a parametrized unitary circuit $V(\theta)$ with parameters $\theta = \{\theta_1, \dots, \theta_d\} \in \mathbb{R}^d$.

Note that that the cost in Eq. (1) can be related to the average Haar fidelity [66, 116]

$$C(U, V(\theta)) = \frac{d+1}{d} (1 - \bar{F}(U, V(\theta))), \quad (\text{A6})$$

$$\begin{aligned} \bar{F}(U, V(\theta)) &= \mathbb{E}_{\psi \sim \mathcal{P}_{\text{Haar}}} [|\langle \psi | U^\dagger V(\theta) | \psi \rangle|^2] \\ &= \int_{\psi} d\psi |\langle \psi | U^\dagger V(\theta) | \psi \rangle|^2, \end{aligned} \quad (\text{A7})$$

where $\mathcal{P}_{\text{Haar}}(\mathcal{H})$ is the uniform probability distribution of states in $\mathcal{H}$ and $d\psi$ is the Haar measure on $\mathcal{H}$. Instead of using the Haar measure, we can write down a cost called the

expected risk,

$$R_{\mathcal{Q}}(\theta) = \mathbb{E}_{\psi \sim \mathcal{Q}} [1 - |\langle \psi | U^\dagger V(\theta) | \psi \rangle|^2], \quad (\text{A8})$$

for some distribution of quantum states $\mathcal{Q}$. Note that $R_{\mathcal{Q}}$ has values $0 \leq R_{\mathcal{Q}} \leq 1$ and that for $\mathcal{Q} = \mathcal{P}_{\text{Haar}}$, we have $R_{\mathcal{Q}}(\theta) = 1 - \bar{F}(U, V(\theta))$. Given a dataset $\mathcal{D}_{\mathcal{Q}} = \{|\psi_k\rangle\}$ of K states $|\psi_k\rangle \sim \mathcal{Q}$, we can write down the empirical risk function

$$C_{\mathcal{D}_{\mathcal{Q}}}(\theta) = 1 - \frac{1}{K} \sum_{k=1}^K |\langle \psi_k | U^\dagger V(\theta) | \psi_k \rangle|^2, \quad (\text{A9})$$

which is an unbiased estimator for the true expected risk $R_{\mathcal{Q}}(\theta)$. Note that this cost function is much easier to handle than Eq. (1) since we now have to measure the overlap between states sampled from $\mathcal{Q}$. In an ideal world, we would be able to sample from the Haar measure on $\mathcal{H}$, so that we get the equality $R_{\mathcal{Q}}(\theta) = 1 - \bar{F}(U, V(\theta))$, however, this is not possible in practice. The question is then, is there a “simple” distribution $\mathcal{Q}$, s.t. $R_{\mathcal{Q}}(\theta)$ approximates $R_{\mathcal{P}_{\text{Haar}}}(\theta)$? As shown in Ref. [55], the answer is that such distributions do indeed exist. First, we define the notion of a *locally scrambling ensemble*.

Definition A1 (Locally scrambling unitary ensemble). A unitary ensemble $\mathcal{U}_{\text{LS}}$ is locally scrambled iff for $U \sim \mathcal{U}_{\text{LS}}$ and for any fixed $U_1, \dots, U_n \in \mathcal{U}(\mathbb{C}^2)$ also $U \cdot (\bigotimes_{i=1}^n U_i) \sim \mathcal{U}_{\text{LS}}$. Here, $\mathcal{U}(\mathbb{C}^2)$ is the unitary Haar measure on a single qubit.

In other words, acting with a random local unitary should to change the properties of the ensemble. We can use the above definition to define a locally scrambled state.

Definition A2 (Locally scrambling ensemble of states).

$$\mathcal{S}_{\text{LS}} = \mathcal{U}_{\text{LS}} |0\rangle^{\otimes n}. \quad (\text{A10})$$

Given such a locally scrambling ensemble, the authors of Ref. [55] prove the following theorem:

Theorem A1. (Equivalence of locally scrambled risks). If $\mathcal{Q}$ is a locally scrambling ensemble, then

$$\frac{1}{2} R_{\mathcal{P}_{\text{Haar}}}(\theta) \leq \frac{N}{N+1} R_{\mathcal{Q}}(\theta) \leq R_{\mathcal{P}_{\text{Haar}}}(\theta). \quad (\text{A11})$$

What this means is that any ensemble $\mathcal{Q}$ that is locally scrambling will give a risk that is within a factor 1/2 of the empirical risk over the full Haar measure.

The question we are then left with is: If we achieve a low empirical cost $C_{\mathcal{D}_{\mathcal{Q}}}$ on the data, how close to the actual expected risk $R_{\mathcal{P}}$ can we get? We have the answer with the following corollary from Ref. [55].

Corollary A1 (Locally scrambled out-of-distribution generalization from in-distribution generalization). Let $\mathcal{P} \in \mathcal{S}_{\text{LS}}$ and $\mathcal{Q} \in \mathcal{S}_{\text{LS}}$. Let U be an unknown n -qubit unitary. Let $V(\theta)$ be an n -qubit quantum circuit with N_p parametrized local gates. When trained using training data $\mathcal{D}_{\mathcal{Q}}$ containing K input-output pairs, with inputs drawn from the ensemble $\mathcal{Q}$. Then, the out-of-distribution risk w.r.t. $\mathcal{P}$ with optimal parameters θ^* after training satisfies

$$R_{\mathcal{P}}(\theta^*) \leq 2C_{\mathcal{D}_{\mathcal{Q}}}(\theta^*) + O\left(\sqrt{\frac{N_p \log N_p}{K}}\right). \quad (\text{A12})$$

Hence, the true expected risk is of a locally scrambled ensemble can be upper bounded by the empirical cost and a

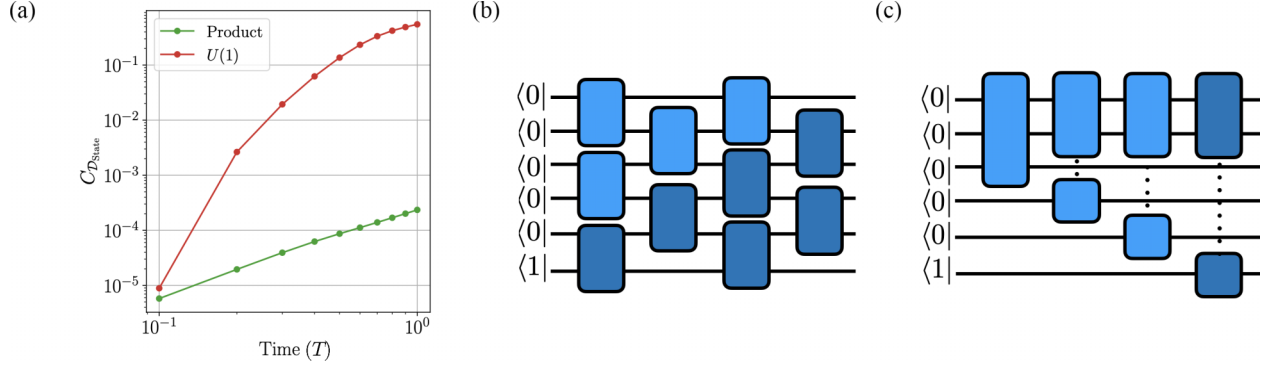


FIG. 8. Difficulty of generalization within $U(1)$ -preserving ensembles. (a) We compile a unitary of time $t = 0.1$ and perform a long time evolution up to $T = 1$ for the $n = 20$ Heisenberg model. Note that this Hamiltonian has a $U(1)$ symmetry to preserve the particle sector of a state under unitary evolution. As training data we use product states from $S_{\text{Haar}^{\otimes n}}$ (green line) and $S_{U(1)}$ within the half-filling charge sector (red line). We perform the evolution of a random $U(1)$ state with half filling and calculate the overlap $C_{\mathcal{D}_{\text{State}}}$ with TEBD simulation. We see that the generalization fails if we use training data from $S_{U(1)}$. Only when the unitary is compiled with random product states as data do, we observe the correct long-time dynamics. (b) We consider a 1D brickwall-shaped RQC where each gate can be a $U(1)$ charge-conserving gate. In a shallow circuit, due to the light-cone structure, the charge only remains inside the navy blue region, and thus the output state ensemble cannot be a locally scrambled one. (c) An MPS can be written as a sequential circuit [117–123], and a similar light-cone argument can be made.

factor that scales with the number of gates in the circuit and the number of states in the data.

APPENDIX B: GENERALIZATION TO LONG-TIME DYNAMICS

Theorem B1. Let $U, V \in \text{SU}(N)$. Then, for any input quantum state ρ , we have

$$\|U^m \rho U^{m\dagger} - V^m \rho V^{m\dagger}\|_1 \leq m \|U \rho U^\dagger - V \rho V^\dagger\|_1.$$

Proof. For any two unitary channels $S_U(\rho) = U \rho U^\dagger$ and $S_V(\rho) = V \rho V^\dagger$, one has the operator algebra fact

$$S_U^m - S_V^m = \sum_{k=0}^{m-1} S_V^k (S_U - S_V) S_U^{m-1-k}.$$

Applied to ρ , this reads

$$\begin{aligned} U^m \rho U^{m\dagger} - V^m \rho V^{m\dagger} &= \sum_{k=0}^{m-1} V^k (U \rho U^\dagger - V \rho V^\dagger) (U^{m-1-k})^\dagger. \end{aligned}$$

Taking the trace norm and using

$$\begin{aligned} \left\| \sum_k X_k \right\|_1 &\leq \sum_k \|X_k\|_1, \\ \|V^k (U \rho U^\dagger - V \rho V^\dagger) (U^{m-1-k})^\dagger\|_1 &= \|U \rho U^\dagger - V \rho V^\dagger\|_1 \quad (\text{unitary invariance}), \end{aligned}$$

we get

$$\begin{aligned} \|U^m \rho U^{m\dagger} - V^m \rho V^{m\dagger}\|_1 &\leq \sum_{k=0}^{m-1} \|U \rho U^\dagger - V \rho V^\dagger\|_1 \\ &= m \|U \rho U^\dagger - V \rho V^\dagger\|_1, \end{aligned}$$

which means that the infidelity of a pure state should grow, in the worst case, quadratically with $m = T/t$. ■

APPENDIX C: TRAINING FOR SYSTEMS WITH $U(1)$ SYMMETRIES

A natural extension of the training protocol is considering a system with specific symmetries and restricting the training protocol to a specific symmetry sector. At first glance, training should be easier because the symmetry sector contains less information than the whole unitary. The next natural question is then: What training dataset should we use?

In this Appendix, we consider the case of a system with a global $U(1)$ symmetry. This symmetry can be implemented with random quantum circuit (RQC) or MPS. For computational efficiency, we require the dataset being generated from an RQC with a fixed depth $\mathcal{D}_{\mathcal{Q}} \sim \mathcal{S}_{U(1), \text{RQC}}$ or from a random MPS with a fixed bond dimension χ [124] $\mathcal{D}_{\mathcal{Q}} \sim \mathcal{S}_{U(1), \text{MPS}}$. Here, $\mathcal{S}_{U(1)} = \mathcal{U}_{U(1)}|s\rangle$, where $|s\rangle$ is a computational basis whose charge (Hamming weight) we want to preserve. In Appendix C, we report the numerical results tested with the Heisenberg model. To our surprise, the generalization is actually *worse* than the PQC trained on product state, even when tested on trial states within the same charge sector of the training data (see Fig. 8).

We can understand this result analytically. In fact, the failure of generalization is due to neither of these unitary ensembles being locally scrambling. A key requirement for the out-of-distribution generalization to work is that an ensemble shares the same expectation with the Haar random unitary ensemble when acted for all density matrices ρ , up to the second moment.

$$\begin{aligned} \mathbb{E}_{U \sim \mathcal{U}_{\text{RQC}}} [U \rho U^\dagger] &= \mathbb{E}_{U \sim \mathcal{U}_{\text{Haar}_n}} [U \rho U^\dagger], \\ \mathbb{E}_{U \sim \mathcal{U}_{\text{RQC}}} [U^{\otimes 2} \rho (U^\dagger)^{\otimes 2}] &= \mathbb{E}_{U \sim \mathcal{U}_{\text{Haar}_n}} [U^{\otimes 2} \rho (U^\dagger)^{\otimes 2}]. \end{aligned} \quad (\text{C1})$$

This however cannot be true in a shallow $U(1)$ charge-conserving circuit. We sketch the proof for the RQC case but a similar argument can be made to the fixed- χ MPS case.

Theorem C1 (Shallow U(1) RQCs are not locally scrambling ensemble). An RQC cannot form a locally scrambling ensemble with depth $\tau < n$.

Here, we prove by contradiction:

Proof. Consider ρ is a product state $|1000 \dots\rangle\langle 1000 \dots|$. The Haar random circuit would scramble the charge and thus the charge should spread evenly in the system; however, for the RQC, the charge spreading is limited by the circuit depth and the light cone of the $|1\rangle$ site. In order for the light cone to cover the whole system, the circuit depth has to be at least n .

Therefore, there exists a ρ s.t. $\mathbb{E}_{U \sim \mathcal{U}_{\text{RQC}}}[U \rho U^\dagger] \neq \mathbb{E}_{U \sim \mathcal{U}_{\text{Haar}}}[U \rho U^\dagger]$ for any RQC depth $\tau < n$. ■

The training set generated by a shallow circuit therefore cannot generalize well, matching what we observed numerically. Hence to achieve generalization within a U(1) charge sector, the training set has to be generated by a random circuit with depth at least $\tau = \Omega(n)$. Therefore, in all Heisenberg model training results, we did not use a U(1)-conserving state ensemble.

APPENDIX D: OPTIMIZATION METHODS FOR PQC

1. Mitigatable barren plateaus in sample efficient learning

Despite the success of variational compilation for many problems, the optimization can suffer from barren plateau issues. Roughly speaking, BPs occur when the landscape of the cost function has an average gradient variance that is exponentially small in the system size [86,88]. For local cost functions such as a single Pauli observable, this is known to happen in a $\text{poly}(n)$ circuit depth; whereas for global cost functions even $O(\log(n))$ -depth circuits can be problematic to optimize [61], which could prevent the successful optimization of fidelity functions.

One approach to mitigate BPs, similar previous works such as [29], [125], is to consider an alternative local cost function:

$$C_{\mathcal{D}}^{\text{local}}(\theta) = 1 - \frac{1}{K} \sum_{k=1}^K \frac{1}{n} \sum_i^n \text{Tr}[(|\psi_{k,i}\rangle\langle\psi_{k,i}| \otimes \mathbb{I}) \times V^\dagger(\theta) U |\psi_k\rangle\langle\psi_k| U^\dagger V(\theta)], \quad (\text{D1})$$

where $|\psi_{k,i}\rangle\langle\psi_{k,i}|$ is the reduced density matrix of $|\psi_k\rangle\langle\psi_k|$ on site i . When Eq. (D1) equals 0, Eq. (3) is guaranteed to be 0. On paper, it seems like Eq. (D1) reduces the global cost function to a local cost function where BPs are less severe. However, in practice, this introduces an extra summation over each site i in the cost function, which adds extra computational costs. Some work reduces the contraction cost by restricting the cost function to the lightcone of the targeted evolution [46]. However, we found that this cost is still too prohibitive for our numerical studies. Furthermore, it is known that the learning is less efficient if it is not “end-to-end” [126], meaning that when the target cost function is not directly aligned with the desired outcome, the learning process becomes less effective.

Nevertheless, the time evolution we are trying to learn here has a lot of structure that can be exploited. A series of “warm start” strategies have been proposed and proven successful in synthesizing quantum dynamics. For example, one could truncate the time evolution into slices. Starting from

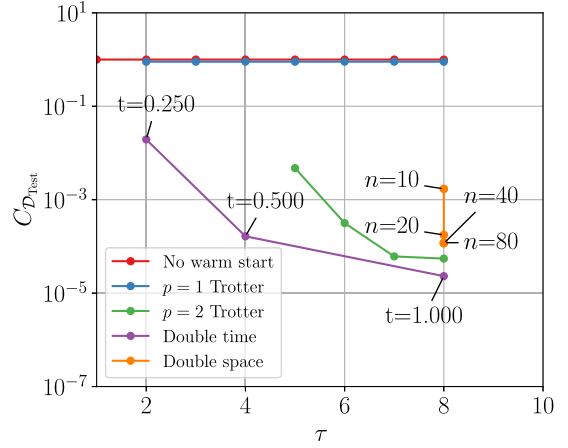


FIG. 9. The importance of warm starts. We show the results of a variational compilation procedure for the one-dimensional Ising model at $g = 1$ for $n = 80$ and $t = 1.0$. All circuits are translation invariant (same parameters for all gates per layer) and trained with $N_s = 16$.

a single time slice and iteratively “grow” the dynamics, using the parameters from a previous iteration as the initialization [31]. In a recent work [63], the authors analytically study this iterative initialization strategy and prove such a variational algorithm has at least an inverse polynomial gradient in n at each time step. By establishing convexity guarantees for these regions, their work suggests polynomial-size time steps could be trainable. While the initialization strategy considered in Ref. [63] is successful, it requires iterations linear in the total evolution time and can be costly to implement numerically. In this work, we use three different initialization strategies to overcome BPs, inspired by ideas in transfer learning [60]:

(1) *Trotterization.* We start the optimization from an order p Trotterized time evolution. This has the advantage that it is completely deterministic and gives us control over the starting fidelity of the optimization. This has the benefit that it is completely deterministic. However, sometimes Trotterization does not give a circuit that matches the desired PQC architecture so other initialization methods are considered in our work.

(2) *Double time.* Here, we start with a very short time evolution t_{init} with a circuit of depth τ_{init} . For short times, even for large systems, this should result in a finite overlap between the training states $U|\psi_i\rangle$ and targets $V_{\text{init}}(\theta)|\psi_i\rangle$. After optimization, we obtain the parameters θ^* . Then, we double the time $t_1 = 2t_{\text{init}}$, and initialize the circuit with two times the previous solution: $V_1(\theta) = V_{\text{init}}(\theta^*)V_{\text{init}}(\theta^*)$. This initialization will again have a nonzero fidelity, hence, we can reoptimize θ . We repeat this process until we reach some target time $t_{\text{final}} = t_{\text{init}} \times 2^i$.

(3) *Double space.* Similar to the “Double time” strategy, we start with a system size n_{init} for a fixed time t and fixed depth τ . We solve the optimization problem and find a solution $V_{\text{init}}(\theta^*)$. Then, we double the system size, $n_1 = 2n_{\text{init}}$ and initialize the circuit with the tensor product of the previous solution: $V_1(\theta) = V_{\text{init}}(\theta^*) \otimes V_{\text{init}}(\theta^*)$. We reoptimize the parameters θ for this system size and keep repeating this procedure until we reach a target system size $n_{\text{final}} = n_{\text{init}} \times 2^i$.

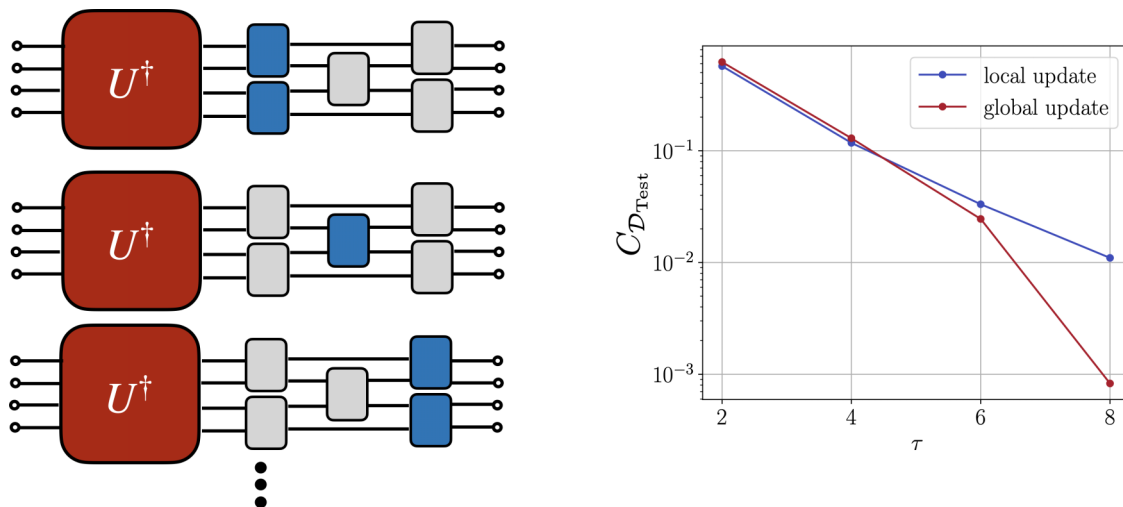


FIG. 10. Comparing local and global update schemes. Left: A demonstration of the local sweeping scheme. In this method, the PQC is translation invariant, and only one layer of the circuit is updated at a time. Right: Performance comparison of the two update schemes on a 4×4 Ising model with parameters $t = 0.5$, $g = -1$, and $\kappa = 0$.

In Fig. 9, we show how these three different initialization strategies behave in a typical setting. First, we see that without any warm start, the optimization will get stuck at $C_{D_{\text{Test}}} = 1$. Similarly, a $p = 1$ Trotterization is not enough to get the optimization going. Only when we initialize a $\tau = 5$ circuit with a $p = 2$ Trotterization, we do get a large enough signal to start the optimization. We see that both the “double time” and “double space” approaches result in large enough signals to start the optimization. Interestingly, we see that different warm start approaches lead to different solutions that can vary almost an order of magnitude in infidelity.

2. Global update versus local update

In our study, we implemented a global parameter update scheme where all variational parameters in the PQCs were updated simultaneously. Here, we compare this approach to a local update scheme, which is akin to the density matrix

renormalization group method [127], and has been considered in recent works such as those by Lin *et al.* and Drudis *et al.* [31,63]. In the local update method, the PQC is sequentially optimized by adjusting one layer of tensors at a time, sweeping from left to right and then back again, and this process is repeated until convergence is achieved.

In Fig. 10, our results indicate that for shallow circuits (depth $\tau \leq 4$), the outcomes of the local update scheme are comparable to those of the global update scheme. However, as the circuit depth increases, the local update method struggles to locate the global minimum, particularly at a depth of $\tau = 8$, where the global update outperforms the local update by more than an order of magnitude. This highlights a key difference between ground state finding problems and our scenario, suggesting that local updates may not always be effective in achieving the global optimum in more complex optimization landscapes.

-
- [1] R. P. Feynman, Simulating physics with computers, *Int. J. Theor. Phys.* **21**, 467 (1982).
 - [2] S. Aaronson and A. Arkhipov, The computational complexity of linear optics, in *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing*, edited by L. Fortnow and S. P. Vadhan (ACM, New York, NY, 2011), pp. 333–342.
 - [3] J. Preskill, Quantum computing and the entanglement frontier, [arXiv:1203.5813](#).
 - [4] S. Aaronson and A. Arkhipov, Bosonsampling is far from uniform, [arXiv:1309.7460](#).
 - [5] A. M. Childs, D. Gosset, and Z. Webb, Universal computation by multiparticle quantum walk, *Science* **339**, 791 (2013).
 - [6] S. Aaronson and L. Chen, Complexity-theoretic foundations of quantum supremacy experiments, [arXiv:1612.05903](#).
 - [7] A. Bouland, B. Fefferman, C. Nirkhe, and U. Vazirani, On the complexity and verification of quantum random circuit sampling, *Nat. Phys.* **15**, 159 (2019).
 - [8] R. Movassagh, Efficient unitary paths and quantum computational supremacy: A proof of average-case hardness of random circuit sampling, [arXiv:1810.04681](#).
 - [9] S. Aaronson and Y. Zhang, On verifiable quantum advantage with peaked circuit sampling, [arXiv:2404.14493](#).
 - [10] H. Gharibyan, M. Z. Mullath, N. E. Sherman, V. P. Su, H. Tepanyan, and Y. Zhang, Heuristic quantum advantage with peaked circuits, [arXiv:2510.25838](#).
 - [11] Y. Zhang, Complexity and hardness of random peaked circuits, [arXiv:2510.00132](#).
 - [12] S. P. Jordan, H. Krovi, K. S. Lee, and J. Preskill, BQP-completeness of scattering in scalar quantum field theory, *Quantum* **2**, 44 (2018).
 - [13] M. Kreshchuk, W. M. Kirby, G. Goldstein, H. Beauchemin, and P. J. Love, Quantum simulation of quantum field theory in the light-front formulation, *Phys. Rev. A* **105**, 032418 (2022).
 - [14] I. Kassal, S. P. Jordan, P. J. Love, M. Mohseni, and A. Aspuru-Guzik, Polynomial-time quantum algorithm for the simulation

- of chemical dynamics, *Proc. Natl. Acad. Sci. USA* **105**, 18681 (2008).
- [15] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik, Quantum chemistry in the age of quantum computing, *Chem. Rev.* **119**, 10856 (2019).
- [16] R. Nandkishore and D. A. Huse, Many-body localization and thermalization in quantum statistical mechanics, *Annu. Rev. Condens. Matter Phys.* **6**, 15 (2015).
- [17] I.-D. Potirniche, A. C. Potter, M. Schleier-Smith, A. Vishwanath, and N. Y. Yao, Floquet symmetry-protected topological phases in cold-atom systems, *Phys. Rev. Lett.* **119**, 123601 (2017).
- [18] N. Y. Yao, A. C. Potter, I.-D. Potirniche, and A. Vishwanath, Discrete time crystals: Rigidity, criticality, and realizations, *Phys. Rev. Lett.* **118**, 030401 (2017).
- [19] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, [arXiv:1411.4028](https://arxiv.org/abs/1411.4028).
- [20] E. Farhi and A. W. Harrow, Quantum supremacy through the quantum approximate optimization algorithm, [arXiv:1602.07674](https://arxiv.org/abs/1602.07674).
- [21] G. G. Guerreschi and A. Y. Matsuura, QAOA for Max-Cut requires hundreds of qubits for quantum speed-up, *Sci. Rep.* **9**, 6903 (2019).
- [22] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices, *Phys. Rev. X* **10**, 021067 (2020).
- [23] Y. Zhang, R. Zhang, and A. C. Potter, QED driven QAOA for network-flow optimization, *Quantum* **5**, 510 (2021).
- [24] S. Kanno, K. Sugisaki, H. Nakamura, H. Yamauchi, R. Sakuma, T. Kobayashi, Q. Gao, and N. Yamamoto, Tensor-based quantum phase difference estimation for large-scale demonstration, *Proc. Natl. Acad. Sci. USA* **122**, e2425026122 (2025).
- [25] D. Layden, G. Mazzola, R. V. Mishmash, M. Motta, P. Wocjan, J.-S. Kim, and S. Sheldon, Quantum-enhanced Markov chain Monte Carlo, *Nature (London)* **619**, 282 (2023).
- [26] H. F. Trotter, On the product of semi-groups of operators, *Proc. Am. Math. Soc.* **10**, 545 (1959).
- [27] M. Suzuki, Generalized Trotter's formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems, *Commun. Math. Phys.* **51**, 183 (1976).
- [28] L. Cincio, Y. Subaşı, A. T. Sornborger, and P. J. Coles, Learning the quantum algorithm for state overlap, *New J. Phys.* **20**, 113022 (2018).
- [29] S. Khatri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles, Quantum-assisted quantum compiling, *Quantum* **3**, 140 (2019).
- [30] C. Cîrstoiu, Z. Holmes, J. Iosue, L. Cincio, P. J. Coles, and A. Sornborger, Variational fast forwarding for quantum simulation beyond the coherence time, *npj Quantum Inf.* **6**, 82 (2020).
- [31] S.-H. Lin, R. Dilip, A. G. Green, A. Smith, and F. Pollmann, Real- and imaginary-time evolution with compressed quantum circuits, *PRX Quantum* **2**, 010342 (2021).
- [32] Y.-X. Yao, N. Gomes, F. Zhang, C.-Z. Wang, K.-M. Ho, T. Iadecola, and P. P. Orth, Adaptive variational quantum dynamics simulations, *PRX Quantum* **2**, 030307 (2021).
- [33] E. Kökçü, D. Camps, L. Bassman Oftelie, J. K. Freericks, W. A. de Jong, R. Van Beeumen, and A. F. Kemper, Algebraic compression of quantum circuits for Hamiltonian evolution, *Phys. Rev. A* **105**, 032420 (2022).
- [34] E. Kökçü, T. Steckmann, Y. Wang, J. K. Freericks, E. F. Dumitrescu, and A. F. Kemper, Fixed depth Hamiltonian simulation via Cartan decomposition, *Phys. Rev. Lett.* **129**, 070501 (2022).
- [35] C. Mc Keever and M. Lubasch, Classically optimized Hamiltonian simulation, *Phys. Rev. Res.* **5**, 023146 (2023).
- [36] C. Mc Keever and M. Lubasch, Towards adiabatic quantum computing using compressed quantum circuits, *PRX Quantum* **5**, 020362 (2024).
- [37] A. M. Childs, A. Ostrander, and Y. Su, Faster quantum simulation by randomization, *Quantum* **3**, 182 (2019).
- [38] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
- [39] Y. Zhang, D. Niu, A. Shabani, and H. Shapourian, Quantum volume for photonic quantum processors, *Phys. Rev. Lett.* **130**, 110602 (2023).
- [40] A. Cervera-Lierta, Exact Ising model simulation on a quantum computer, *Quantum* **2**, 114 (2018).
- [41] B. Peng, S. Gulania, Y. Alexeev, and N. Govind, Quantum time dynamics employing the Yang-Baxter equation for circuit compression, *Phys. Rev. A* **106**, 012412 (2022).
- [42] Y. Atia and D. Aharonov, Fast-forwarding of Hamiltonians and exponentially precise measurements, *Nat. Commun.* **8**, 1572 (2017).
- [43] Y. Zhang, J. Carrasquilla, and Y. B. Kim, Observation of a non-Hermitian supersonic mode on a trapped-ion quantum computer, *Nat. Commun.* **16**, 3286 (2025).
- [44] J. P. Palao and R. Kosloff, Forcing a unitary transformation by an external field: Comparing two approaches based on optimal control theory, [arXiv:quant-ph/0208147](https://arxiv.org/abs/quant-ph/0208147).
- [45] Y. Zhang, S. Jahanbani, A. Riswadkar, S. Shankar, and A. C. Potter, Sequential quantum simulation of spin chains with a single circuit QED device, *Phys. Rev. A* **109**, 022606 (2024).
- [46] K. Mizuta, Y. O. Nakagawa, K. Mitarai, and K. Fujii, Local variational quantum compilation of large-scale Hamiltonian dynamics, *PRX Quantum* **3**, 040302 (2022).
- [47] S. Kanasugi, S. Tsutsui, Y. O. Nakagawa, K. Maruyama, H. Oshima, and S. Sato, Computation of Green's function by local variational quantum compilation, *Phys. Rev. Res.* **5**, 033070 (2023).
- [48] S. Kanasugi, Y. Hidaka, Y. O. Nakagawa, S. Tsutsui, N. Matsumoto, K. Maruyama, H. Oshima, and S. Sato, Subspace-based local compilation of variational quantum circuits for large-scale quantum many-body simulation, *Phys. Rev. Res.* **7**, 023298 (2025).
- [49] J. Schachenmayer, B. P. Lanyon, C. F. Roos, and A. J. Daley, Entanglement growth in quench dynamics with variable range interactions, *Phys. Rev. X* **3**, 031015 (2013).
- [50] H. Zhao, M. Bukov, M. Heyl, and R. Moessner, Adaptive trotterization for time-dependent Hamiltonian quantum dynamics using piecewise conservation laws, *Phys. Rev. Lett.* **133**, 010603 (2024).

- [51] T. Barthel and Y. Zhang, Optimized Lie–Trotter–Suzuki decompositions for two and three non-commuting terms, *Ann. Phys. (NY)* **418**, 168165 (2020).
- [52] L. Banchi, J. Pereira, and S. Pirandola, Generalization in quantum machine learning: A quantum information standpoint, *PRX Quantum* **2**, 040321 (2021).
- [53] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, Power of data in quantum machine learning, *Nat. Commun.* **12**, 2631 (2021).
- [54] M. C. Caro, H.-Y. Huang, M. Cerezo, K. Sharma, A. Sornborger, L. Cincio, and P. J. Coles, Generalization in quantum machine learning from few training data, *Nat. Commun.* **13**, 4919 (2022).
- [55] M. C. Caro, H.-Y. Huang, N. Ezzell, J. Gibbs, A. T. Sornborger, L. Cincio, P. J. Coles, and Z. Holmes, Out-of-distribution generalization for learning quantum dynamics, *Nat. Commun.* **14**, 3751 (2023).
- [56] S. Jerbi, J. Gibbs, M. S. Rudolph, M. C. Caro, P. J. Coles, H.-Y. Huang, and Z. Holmes, The power and limitations of learning quantum dynamics incoherently, [arXiv:2303.12834](https://arxiv.org/abs/2303.12834).
- [57] L. G. Valiant, A theory of the learnable, in *Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing—STOC’84*, edited by R. A. DeMillo (ACM Press, New York, NY, 1984), pp. 436–445.
- [58] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, A theory of learning from different domains, *Mach. Learn.* **79**, 151 (2010).
- [59] T. Haug and M. S. Kim, Generalization of quantum machine learning models using quantum Fisher information metric, *Phys. Rev. Lett.* **133**, 050603 (2024).
- [60] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, A comprehensive survey on transfer learning, *Proc. IEEE* **109**, 43 (2021).
- [61] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, *Nat. Commun.* **12**, 1791 (2021).
- [62] A. A. Mele, G. B. Mbeng, G. E. Santoro, M. Collura, and P. Torta, Avoiding barren plateaus via transferability of smooth solutions in a Hamiltonian variational ansatz, *Phys. Rev. A* **106**, L060401 (2022).
- [63] R. Puig, M. Drudis, S. Thanasilp, and Z. Holmes, Variational quantum simulation: A case study for understanding warm starts, *PRX Quantum* **6**, 010317 (2025).
- [64] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2001), Vol. 2.
- [65] N. F. Berthussen, T. V. Trevisan, T. Iadecola, and P. P. Orth, Quantum dynamics simulations beyond the coherence time on noisy intermediate-scale quantum hardware by variational Trotter compression, *Phys. Rev. Res.* **4**, 023097 (2022).
- [66] M. A. Nielsen, A simple formula for the average gate fidelity of a quantum dynamical operation, *Phys. Lett. A* **303**, 249 (2002).
- [67] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, Matrix product density operators: Simulation of finite-temperature and dissipative systems, *Phys. Rev. Lett.* **93**, 207204 (2004).
- [68] G. Vidal, Efficient simulation of one-dimensional quantum many-body systems, *Phys. Rev. Lett.* **93**, 040502 (2004).
- [69] I. P. McCulloch, From density-matrix renormalization group to matrix product states, *J. Stat. Mech.* (2007) P10014.
- [70] B. Pirvu, V. Murg, J. I. Cirac, and F. Verstraete, Matrix product operator representations, *New J. Phys.* **12**, 025012 (2010).
- [71] J. Gibbs and L. Cincio, Deep circuit compression for quantum dynamics via tensor networks, *Quantum* **9**, 1789 (2025).
- [72] I. N. M. Le, S. Sun, and C. B. Mendl, Riemannian quantum circuit optimization based on matrix product operators, *Quantum* **9**, 1833 (2025).
- [73] M. B. Hastings, An area law for one-dimensional quantum systems, *J. Stat. Mech.* (2007) P08024.
- [74] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, Computational complexity of projected entangled pair states, *Phys. Rev. Lett.* **98**, 140506 (2007).
- [75] J. Haferkamp, D. Hangleiter, J. Eisert, and M. Gluza, Contracting projected entangled pair states is average-case hard, *Phys. Rev. Res.* **2**, 013010 (2020).
- [76] M. S. Leifer and D. Poulin, Quantum graphical models and belief propagation, *Ann. Phys. (NY)* **323**, 1899 (2008).
- [77] C. Guo, D. Poletti, and I. Arad, Block belief propagation algorithm for two-dimensional tensor networks, *Phys. Rev. B* **108**, 125111 (2023).
- [78] G. Park, J. Gray, and G. K. Chan, Simulating quantum dynamics in two-dimensional lattices with tensor network influence functional belief propagation, *Phys. Rev. B* **112**, 174310 (2025).
- [79] J. Tindall and M. Fishman, Gauging tensor networks with belief propagation, *SciPost Phys.* **15**, 222 (2023).
- [80] S. Midha and Y. F. Zhang, Beyond belief propagation: Cluster-corrected tensor network contraction with exponential convergence, [arXiv:2510.02290](https://arxiv.org/abs/2510.02290).
- [81] T. Begušić, J. Gray, and G. K.-L. Chan, Fast and converged classical simulations of evidence for the utility of quantum computing before fault tolerance, *Sci. Adv.* **10**, eadk4321 (2024).
- [82] H.-C. Jiang, Z.-Y. Weng, and T. Xiang, Accurate determination of tensor network state of quantum lattice models in two dimensions, *Phys. Rev. Lett.* **101**, 090603 (2008).
- [83] M. Lubasch, J. I. Cirac, and M.-C. Banuls, Unifying projected entangled pair state contractions, *New J. Phys.* **16**, 033014 (2014).
- [84] M. Lubasch, J. I. Cirac, and M.-C. Banuls, Algorithms for finite projected entangled pair states, *Phys. Rev. B* **90**, 064425 (2014).
- [85] N. Pancotti and J. Gray, One-step replica symmetry breaking in the language of tensor networks, [arXiv:2306.15004](https://arxiv.org/abs/2306.15004).
- [86] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, *Nat. Commun.* **9**, 4812 (2018).
- [87] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, and P. J. Coles, Effect of barren plateaus on gradient-free optimization, *Quantum* **5**, 558 (2021).
- [88] M. Ragone, B. N. Bakalov, F. Sauvage, A. F. Kemper, C. Ortiz Marrero, M. Larocca, and M. Cerezo, A Lie algebraic theory of barren plateaus for deep parameterized quantum circuits, *Nat. Commun.* **15**, 7172 (2024).
- [89] D. P. Kingma and J. Ba, ADAM: A method for stochastic optimization, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 79, 2015, Conference Track Proceedings*, edited by Y. Bengio and Y. LeCun (International Conference on Learning Representations (ICLR), San Diego, CA, 2015).

- [90] J. Gray, quimb: A Python library for quantum information and many-body calculations, *J. Open Source Software* **3**, 819 (2018).
- [91] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, PyTorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems*, Vol. 32 (Curran Associates, Inc., Red Hook, NY, 2019), pp. 8024–8035.
- [92] R. Wiersema and Y. Zhang, Scalable quantum dynamics compilation via quantum machine learning (2024), https://github.com/therooler/variational_compilation.
- [93] M. Žnidarič, T. Prosen, and P. Prelovšek, Many-body localization in the Heisenberg XXZ magnet in a random field, *Phys. Rev. B* **77**, 064426 (2008).
- [94] J. H. Bardarson, F. Pollmann, and J. E. Moore, Unbounded growth of entanglement in models of many-body localization, *Phys. Rev. Lett.* **109**, 017202 (2012).
- [95] M. Serbyn, Z. Papić, and D. A. Abanin, Universal slow growth of entanglement in interacting strongly disordered systems, *Phys. Rev. Lett.* **110**, 260601 (2013).
- [96] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, Evidence for the utility of quantum computing before fault tolerance, *Nature (London)* **618**, 500 (2023).
- [97] R. Haghshenas, E. Chertkov, M. Mills, W. Kadow, S.-H. Lin, Y.-H. Chen, C. Cade, I. Niesen, T. Begušić, M. S. Rudolph, *et al.*, Digital quantum magnetism at the frontier of classical simulations, [arXiv:2503.20870](https://arxiv.org/abs/2503.20870).
- [98] J. Gibbs, Z. Holmes, M. C. Caro, N. Ezzell, H.-Y. Huang, L. Cincio, A. T. Sornborger, and P. J. Coles, Dynamical simulation via quantum machine learning with provable generalization, *Phys. Rev. Res.* **6**, 013241 (2024).
- [99] M. Rigol and A. Muramatsu, Fermionization in an expanding 1D gas of hard-core bosons, *Phys. Rev. Lett.* **94**, 240403 (2005).
- [100] I. Hen and M. Rigol, Strongly interacting atom lasers in three-dimensional optical lattices, *Phys. Rev. Lett.* **105**, 180401 (2010).
- [101] M. Jreissaty, J. Carrasquilla, F. A. Wolf, and M. Rigol, Expansion of Bose-Hubbard Mott insulators in optical lattices, *Phys. Rev. A* **84**, 043610 (2011).
- [102] A. Jreissaty, J. Carrasquilla, and M. Rigol, Self-trapping in the two-dimensional Bose-Hubbard model, *Phys. Rev. A* **88**, 031606(R) (2013).
- [103] J. M. Wilson, N. Malvania, Y. Le, Y. Zhang, M. Rigol, and D. S. Weiss, Observation of dynamical fermionization, *Science* **367**, 1461 (2020).
- [104] A. M. Childs, Y. Su, M. C. Tran, N. Wiebe, and S. Zhu, Theory of Trotter error with commutator scaling, *Phys. Rev. X* **11**, 011020 (2021).
- [105] S. Paeckel, T. Köhler, A. Swoboda, S. R. Manmana, U. Schollwöck, and C. Hubig, Time-evolution methods for matrix-product states, *Ann. Phys. (NY)* **411**, 167998 (2019).
- [106] A Trotterization of order p is such that the leading term of the error grows with $\sim t^{p+1}$.
- [107] F. Verstraete and J. I. Cirac, Renormalization algorithms for quantum-many body systems in two and higher dimensions, [arXiv:cond-mat/0407066](https://arxiv.org/abs/cond-mat/0407066).
- [108] M. P. Zaletel and F. Pollmann, Isometric tensor network states in two dimensions, *Phys. Rev. Lett.* **124**, 037201 (2020).
- [109] M. D’Anna, Y. Zhang, R. Wiersema, M. S. Rudolph, and J. Carrasquilla, Circuit compression for 2D quantum dynamics, [arXiv:2507.01883](https://arxiv.org/abs/2507.01883).
- [110] J. J. Meyer, M. Mularski, E. Gil-Fuster, A. A. Mele, F. Arzani, A. Wilms, and J. Eisert, Exploiting symmetry in variational quantum machine learning, *PRX Quantum* **4**, 010328 (2023).
- [111] Q. T. Nguyen, L. Schatzki, P. Braccia, M. Ragone, P. J. Coles, F. Sauvage, M. Larocca, and M. Cerezo, Theory for equivariant quantum neural networks, *PRX Quantum* **5**, 020328 (2024).
- [112] R. Wiersema, A. F. Kemper, B. N. Bakalov, and N. Killoran, Geometric quantum machine learning with horizontal quantum gates, *Phys. Rev. Res.* **7**, 013148 (2025).
- [113] M. Malinowski, D. T. C. Allcock, and C. J. Ballance, How to wire a 1000-qubit trapped-ion quantum computer, *PRX Quantum* **4**, 040313 (2023).
- [114] Y. Zhang and Y. Zhang, Classical simulability of quantum circuits with shallow magic depth, *PRX Quantum* **6**, 010337 (2025).
- [115] <https://vectorinstitute.ai/#partners>.
- [116] M. Horodecki, P. Horodecki, and R. Horodecki, General teleportation channel, singlet fraction, and quasidistillation, *Phys. Rev. A* **60**, 1888 (1999).
- [117] C. Schön, E. Solano, F. Verstraete, J. I. Cirac, and M. M. Wolf, Sequential generation of entangled multiqubit states, *Phys. Rev. Lett.* **95**, 110503 (2005).
- [118] I. H. Kim, Holographic quantum simulation, [arXiv:1702.02093](https://arxiv.org/abs/1702.02093).
- [119] J.-G. Liu, Y.-H. Zhang, Y. Wan, and L. Wang, Variational quantum eigensolver with fewer qubits, *Phys. Rev. Res.* **1**, 023025 (2019).
- [120] T. Soejima, K. Siva, N. Bultinck, S. Chatterjee, F. Pollmann, and M. P. Zaletel, Isometric tensor network representation of string-net liquids, *Phys. Rev. B* **101**, 085117 (2020).
- [121] M. Foss-Feig, D. Hayes, J. M. Dreiling, C. Figgatt, J. P. Gaebler, S. A. Moses, J. M. Pino, and A. C. Potter, Holographic quantum algorithms for simulating correlated spin systems, *Phys. Rev. Res.* **3**, 033002 (2021).
- [122] S. Anand, J. Hauschild, Y. Zhang, A. C. Potter, and M. P. Zaletel, Holographic quantum simulation of entanglement renormalization circuits, *PRX Quantum* **4**, 030334 (2023).
- [123] Y. Zhang, S. Jahanbani, D. Niu, R. Haghshenas, and A. C. Potter, Qubit-efficient simulation of thermal states with quantum tensor networks, *Phys. Rev. B* **106**, 165126 (2022).
- [124] S. Singh, R. N. C. Pfeifer, and G. Vidal, Tensor network states and algorithms in the presence of a global U(1) symmetry, *Phys. Rev. B* **83**, 115125 (2011).
- [125] T. Jones and S. C. Benjamin, Robust quantum compilation and circuit optimisation via energy minimisation, *Quantum* **6**, 628 (2022).
- [126] X.-Z. Luo, D. Luo, and R. G. Melko, Operator learning renormalization group, [arXiv:2403.03199](https://arxiv.org/abs/2403.03199).
- [127] S. R. White, Density matrix formulation for quantum renormalization groups, *Phys. Rev. Lett.* **69**, 2863 (1992).