

Variational Power of Quantum Circuit Tensor Networks

Reza Haghshenas^{1,*}, Johnnie Gray^{1,†}, Andrew C. Potter,^{2,3} and Garnet Kin-Lic Chan^{1,‡}

¹*Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA*

²*Department of Physics, University of Texas at Austin, Austin, Texas 78712, USA*

³*Department of Physics and Astronomy, and Quantum Matter Institute, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z1*



(Received 21 July 2021; revised 10 December 2021; accepted 13 January 2022; published 11 March 2022)

We characterize the variational power of quantum circuit tensor networks in the representation of physical many-body ground states. Such tensor networks are formed by replacing the dense block unitaries and isometries in standard tensor networks by local quantum circuits. We explore both quantum circuit matrix product states and the quantum circuit multiscale entanglement renormalization *Ansatz*, and introduce an adaptive method to optimize the resulting circuits to high fidelity with more than 10^4 parameters. We benchmark their expressiveness against standard tensor networks, as well as other common circuit architectures, for the 1D and 2D Heisenberg and 1D Fermi-Hubbard models. We find quantum circuit tensor networks to be substantially more expressive than other quantum circuits for these problems, and that they can even be more compact than standard tensor networks. Extrapolating to circuit depths which can no longer be emulated classically, this suggests a region of advantage in quantum expressiveness in the representation of physical ground states.

DOI: 10.1103/PhysRevX.12.011047

Subject Areas: Computational Physics
Condensed Matter Physics
Quantum Physics

I. INTRODUCTION

Advances in digital quantum computing have led to renewed interest in quantum circuit (QC) representations of many-body states. For this purpose, it is crucial to understand the representational power and trainability of different circuit architectures. Out of the many possible architectures, one promising choice is circuits derived from tensor network states used in classical simulations of quantum states with limited entanglement. Because of the close connections between tensor networks and quantum circuits, and the significant numerical experience using them in classical simulations, they provide a natural setting to define the boundary between classical and quantum capabilities for quantum simulation. The simplest measure of the classical complexity of a tensor network is the tensor bond dimension. Consequently, one can construct a tensor network that is hard to simulate classically by devising a

quantum circuit that maps to a tensor network with a very large bond dimension, in a small number of circuit elements. For example, one can construct quantum circuits that generate tensors with bond dimensions exponential in the circuit depth. This is the idea behind “deep” or quantum circuit tensor networks which have been of interest for different applications of quantum devices [1–7]. Further, when combined with midcircuit measurements and qubit reuse, some of these methods allow simulation of large-scale quantum systems with few physical qubits [4,8,9].

However, constructing a class of states that is hard to represent classically is not sufficient to advance the simulation of physical systems. In the context of physical quantum many-body states, we must address additional questions. (i) Are sparsely parametrized quantum “circuit” tensor networks capable of representing physical states more efficiently than the “dense” tensor networks (i.e., where all elements of the tensors can be independently varied) traditionally used in classical simulation? (ii) And for this task, what are the optimal circuit architectures and optimization protocols? The answers have potential implications not only for quantum simulations, but also for classical simulations with tensor networks. For example, an affirmative answer to (i) would suggest that it may sometimes be better to classically simulate with the quantum circuit tensor network directly, rather than via the classical dense tensor network, so long as the circuit depth and

*haqshena@caltech.edu

†jgray@caltech.edu

‡garnetc@caltech.edu

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

tensor values support efficient classical contraction and/or approximation.

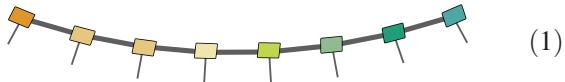
Some work to answer questions (i) and (ii) has already appeared, such as in Refs. [10–13]. In this work, our focus is on establishing the variational power of quantum tensor networks to capture quantum many-body ground states. This is an application where traditional dense tensor networks do well, and is thus in some sense the hardest test for quantum circuit tensor networks to pass. We focus in particular on understanding the resources (e.g., number of variational parameters) required, optimization strategies, and influence of circuit architecture on the results. Because of the large number of numerical experiments required, we will mainly focus on one-dimensional quantum many-body states, although we present suggestive findings on two-dimensional problems also. As we shall demonstrate, with careful optimization strategies, quantum circuit tensor networks are very expressive, and in some cases even more expressive than classical dense tensor networks. This suggests a regime where a quantum advantage in the sense of *expressiveness* may be observed in physical ground-state simulations.

The structure of the paper is as follows. We first introduce quantum circuit tensor networks, review the mapping between two common tensor networks, the matrix product state and the multiscale entanglement renormalization *Ansatz* (MERA) [14], to block unitary circuits [4,7], and introduce different architectures for the local unitary representations. We also provide some intuition regarding the expressiveness of different structures of quantum circuit *Ansätze*. We then examine optimization strategies for such circuits. We next carry out a detailed comparison between the quantum circuit tensor networks, classical dense tensor networks, and reference circuit classes studied in the literature, to evaluate their variational efficiency and power, for both energies and correlation functions. We finish with a discussion of our findings in the context of computational quantum advantage.

II. TENSOR NETWORKS AND QUANTUM CIRCUITS

A. Canonical form of the matrix product state

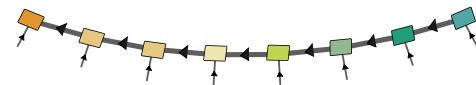
A matrix product state (MPS) [15–17] is a tensor network consisting of a tensor for each site, connected by bonds in a one-dimensional geometry. It is represented diagrammatically by



where each square tensor denotes a $D \times D \times d$ -dimensional tensor. The open bonds index the d -dimensional physical Hilbert space of the site. (For example, if $d = 2$, we can identify each site with a qubit.)

The connected “virtual” bonds of dimension D then control the number of parameters and thus the variational power of the MPS (or more physically, the maximum bipartite entanglement at each bipartition in the network).

The individual tensors in a MPS are not in unique correspondence with a given quantum state due to gauge degrees of freedom: the state is invariant under insertion of a gauge matrix and its inverse G, G^{-1} between two tensors (along a virtual bond). To fix the gauge degrees of freedom, a MPS can be recast into a canonical form [18]. In canonical form, the tensors satisfy additional isometric or normalization constraints, but for a (normalized) MPS, this imposes no loss of representational power. A simple algorithm to obtain the canonical form is to perform a sequence of QR (LQ) decompositions of the tensors; doing this from left to right (right to left) brings an arbitrary MPS into left (right) canonical form [19]. For example, the right canonical form of the MPS in Eq. (1) can be represented by the following diagram:



where every square tensor with arrows is an isometric tensor. The isometric condition is satisfied for contractions performed on the incoming bonds = , where the left-hand side of the equality is the identity matrix. The isometric constraint fixes the gauge freedom up to permutations of bond indices. In a similar way, the right canonical form is defined by tensors which satisfy an isometric condition = , while a “mixed canonical form” [central to the density matrix renormalization group (DMRG) [20]] is obtained by combining left and right canonical forms around a given central site, with the central tensor satisfying the condition = 1. When the tensors in a canonical MPS satisfy isometric conditions, the MPS is an example of an “isometric” tensor network [21,22]. The number of independent variational parameters in a canonical MPS of length L , with all elements real, scales asymptotically as $\sim L \times D(3D - 1)/2$. We refer to these standard MPSs as “dense” MPS (DMPS), since the tensors assume their most flexible parametrization, in contrast to the “sparse” parametrization by a quantum circuit used later. Because of the close relationship between the standard MPS formulation and the DMRG, we will sometimes use the term DMRG.

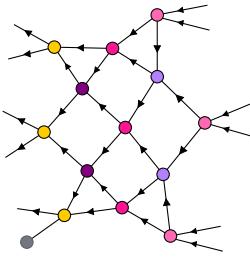
B. Quantum circuit MPS

Given the canonical form of the MPS, the mapping to a block unitary quantum circuit follows a simple procedure [4,7,23]. This is because isometric tensors can be viewed as arising from the application of a unitary tensor to a partial set of inputs. Concretely, the steps are as follows. (i) The

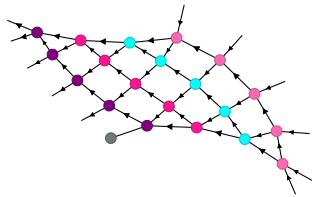
virtual bonds of dimension D (thick lines) are viewed as q qubit bonds (aggregate dimension $D = 2^q$). Graphically, this is the relation $\text{---} \equiv \text{---} \rightarrow$. (Note that, following standard MPS conventions, the arrows on these diagrams indicate right canonical form, and are opposite to the execution-time direction in the associated quantum circuit). (ii) The isometric tensors are viewed as columns of a block unitary matrix with one fixed output qubit (denoted by a black dot, here assuming $d = 2$), i.e., $\text{---} \equiv \text{---} \rightarrow$. This mapping generates the MPS via a block unitary circuit, where each block unitary is a matrix of dimension $dD \times dD$.

In the above mapping, the variational space of normalized states generated by the block unitary circuit and the standard dense MPS are exactly the same. However, we can imagine replacing the block unitary by a local circuit of two-qubit gates of given depth. One can then create a block unitary that acts on a set of q qubits, using as few as $O(q)$ two-qubit gates. This allows us to generate an entangled state that lives in the variational space of a $D = 2^q$ DMPS, with as few as $O(q)$ variational parameters per site. We refer to a MPS where the block unitary is encoded as a local circuit as a “quantum circuit MPS” (QMPS).

There is much freedom to choose the structure of the local circuit. Here, we explore several local circuit structures. (i) A brick-wall circuit, denoted graphically as



with densely connected nearest-neighbor two-body unitary gates (circle tensors $\text{---} \otimes \text{---}$). Throughout, colors visually distinguish different circuit layers, but each gate of a given color implements a distinct gate. For the brick-wall circuit, we refer to a layer of even gates and a layer of odd gates as two layers, thus the above circuit has a circuit depth of $\tau = 6$. In a brick-wall circuit, correlations spread slowly with increasing τ , as the effective correlation length ξ grows linearly with circuit depth $\xi \sim \tau$. (ii) A ladder circuit, for example,



where we label the circuit above as depth $\tau = 4$. Correlations propagate more efficiently in this structure:

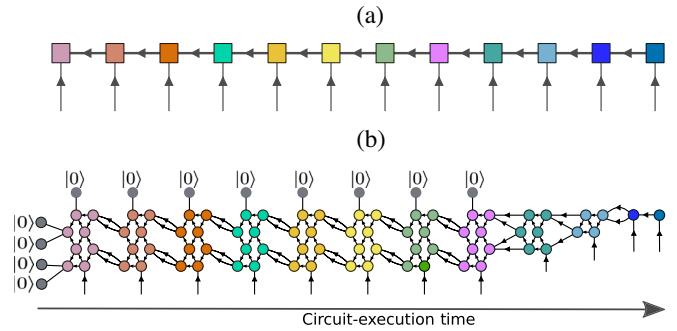
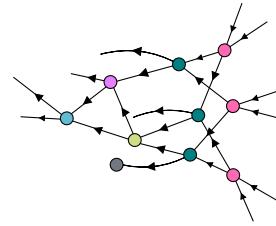


FIG. 1. Schematic representation of a matrix product state and its quantum circuit in (a) right canonical form for a system with $L = 12$ (or equivalently 12 qubits). The square tensors represent isometric tensors. (b) A quantum circuit MPS with four bond qubits ($q = 4$) and a local brick-wall circuit structure with depth $\tau = 4$. The circle tensors are two-qubit unitary gates, while the black tensors denote register qubits initialized in the $|0\rangle$ state. Note that the arrows associated with the MPS and quantum circuit MPS follow the tensor network convention rather than the circuit convention; i.e., they are in the *opposite* direction of circuit-execution time.

with $\tau = 1$, the first and last qubits are already correlated. (iii) A MERA structure,



which represents a binary MERA [24] with depth $\tau = 5$. Properties of MERA circuits in general are discussed more in the section below. Note that here, however, the MERA structure is only being used for the local circuit (i.e., a MERA quantum circuit, rather than a quantum circuit MERA) and globally, the *Ansatz* is still a QMPS. An example of the final circuit structure of the QMPS using a local brick-wall circuit is shown in Fig. 1. Corresponding figures for QMPS with local ladder and MERA circuits are shown in Fig. 12.

Overall, the variational power of the QMPS *Ansatz* is determined by three factors: (i) number of qubits q on which each block unitary circuit acts (placing the QMPS in the variational space of a dense MPS with $D = 2^q$), (ii) the number of gates in the local circuit, (iii) the internal structure of the local circuit. Note that the number of gates in each layer differs between the local circuit structures; thus τ should not be directly compared between the different structures. Instead, the number of gates (or equivalently, variational parameters) asymptotically behaves as $\sim \frac{1}{2}\tau L(q+1)$, $\sim \tau L(q+1)$, and $\sim 2\tau L(q+1)$ for the brick-wall, ladder, and MERA structures,

respectively. In the numerical simulations, we refer to these three kinds of QMPS circuits as QMPS-B, QMPS-L, and QMPS-M, respectively.

C. Quantum circuit MERA

Another commonly used tensor network is the MERA. This is a tensor network state where the tensors are arranged to introduce bipartite entanglement at multiple scales. In this *Ansatz*, isometric tensors perform coarse graining while unitary disentanglers are applied to remove short-range entanglement at the different length scales. We show a binary form of MERA in Fig. 2(a) with unitary disentanglers and isometric coarse-graining tensors  distinguished by different colors in each layer. We refer to the standard form of MERA where all tensors are assumed dense as dense MERA (DMERA).

Since MERAs are isometric tensor networks by construction, like in the MPS canonical form, a quantum circuit MERA (i.e., a circuit whose global structure is derived from the MERA), i.e., QMERA, can be straightforwardly obtained by decomposing both the block isometric and unitary tensors into local circuits with a finite depth τ and given internal structure as already discussed. A graphical illustration of a QMERA is shown in Fig. 2(b). An important difference between a QMERA and a QMPS is the presence of a structured set of long-range unitary gates. In 1D, this allows the QMERA to capture critical entanglement beyond the area law with only a polynomial number of gates [25].

Similarly to the QMPS, the variational power of the QMERA *Ansatz* is determined by three factors: i.e., q , the number of gates, and the internal structure of local circuits. In this work, we only consider QMERA with local brick-wall circuits as depicted in Fig. 2(b) [26], which we refer to as QMERA-B in the numerical studies.

D. Global quantum circuit *Ansatz*

To place the performance of the quantum circuit tensor networks in context, we also consider global brick-wall and ladder circuit *Ansätze*, as depicted in Fig. 3. These are referred to as QC-B and QC-L in the numerical studies below.

E. Properties of different quantum circuit *Ansatz*

All the above *Ansätze* are universal approximators in the sense that with sufficient numbers of parameters (for example, controlled by q and τ in the quantum circuit tensor networks) they can represent any state. Certain types of ground state might be more compactly represented by one *Ansatz* than another, but it is difficult to make general statements without numerical studies, as performed below. However, here we briefly provide some intuition into the relationship between circuit structure and expressivity, and the connection between the different *Ansätze*.

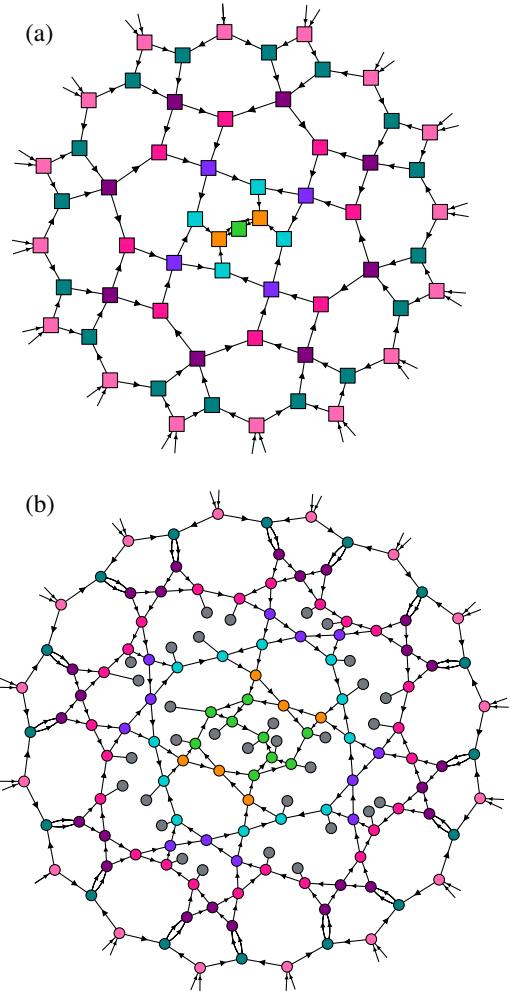


FIG. 2. (a) A binary MERA wave function for $L = 32$ qubits and (b) its quantum circuit, i.e., QMERA, where the local quantum circuit is chosen to have a brick-wall structure with $q = 2$ and $\tau = 2$. The different layers of the MERA are shown in different colors.

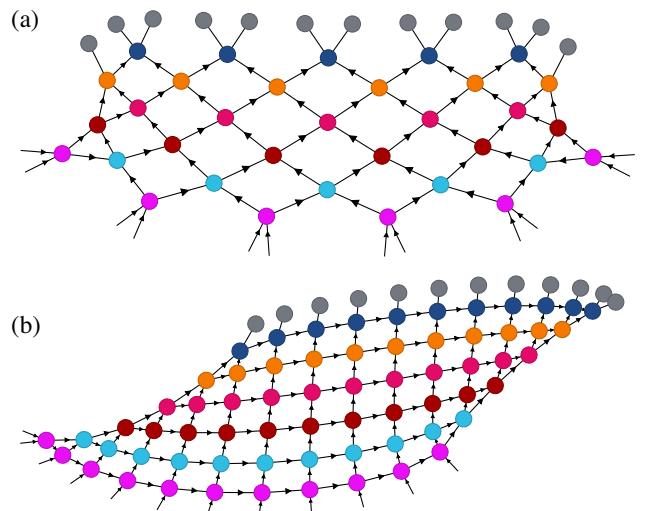


FIG. 3. Global quantum circuits with (a) brick-wall and (b) ladder structures. Execution time flows from top to bottom.

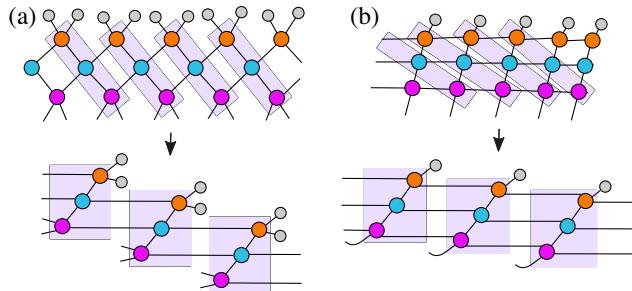


FIG. 4. Mapping from global quantum circuit *Ansatz* to QMPS for the (a) brick-wall and (b) ladder structures. Grouping the gates in the global *Ansatz* (top) defines block unitaries in the QMPS (bottom). Execution time flows from top to bottom.

We start with the relationship between the global quantum circuits and quantum MPS. Both the global brick-wall and ladder circuits can be directly transcribed into QMPS by grouping gates into block unitaries (Fig. 4). QC-B and QC-L circuits with τ layers map to QMPS with $q = \tau - 1$ and $q = \tau$ block unitaries, respectively. Each block unitary has the sparsest possible entangled parametrization with τ gates arranged in a ladder structure, with the ladder ascending in the opposite direction to the ascending direction of the QMPS block unitaries. In fact, the only difference between the global brick-wall and global ladder circuits when viewed from their corresponding QMPS is that the QC-B block unitaries overlap on only $q - 1$ qubits rather than the usual q qubits. This choice of nonmaximal overlap is what gives rise to the specific brick-wall light cone, where correlations cannot spread as quickly as in a general QMPS or a global ladder circuit.

Mapping QC-B and QC-L to QMPS reveals that the circuit structures prioritize reaching block unitaries of large size q (large MPS bond dimension $D = 2^q$) at the expense of accurately representing each unitary, as each block unitary is only minimally connected. If we assume, as seems likely, that the part of the dense MPS variational space of bond dimension $D = 2^q$ required to represent a large variety of quantum ground states of physical interest is not fully captured by these minimal local circuits, then QC-B and QC-L do not efficiently cover the variational space. One can see the influence of the block unitary circuit depth most dramatically in the expectation values of operators acting on the leftmost site in the QMPS diagram in Fig. 4(a). Because of the circuit ordering, such expectation values depend *only* on the parametrization of the first block unitary. In the QC-B and QC-L *Ansätze* one improves the leftmost local expectation values by increasing the sizes (q) of the block unitaries, simultaneously increasing the block unitary circuit depth and global circuit depth.

In the more general form of the QMPS, τ and q can be independently varied. Thus, QMPS is a superset of QC-B or QC-L and is more expressive, although the balance between τ and q will be problem specific. The question is

then whether the local circuit depth can be significantly reduced from that required to exactly parametrize a block unitary over q qubits, which is exponential in q . There is room for optimism, as there are other powerful variational states which map to sparse parametrizations of dense MPS. For example, correlator product states [27–30], entangled plaquette states [31,32], and neural network quantum states [33] can be viewed as variational states parametrized by nonunitary gates. These can map to dense MPS with large D (for example, capturing volume law scaling of entanglement with a polynomial number of variational parameters). However, numerical studies have shown that the number of variational parameters required in these *Ansätze* for physical ground state problems can be fewer than in a dense MPS [27].

Many of the above points also apply to the QMERA, in particular, the potential for sparse circuit representation of the unitaries and isometries arising in the dense MERA. In addition, the special geometric structure of the MERA means that in 1D it spreads correlations while capturing logarithmic corrections to the entanglement law. This is an important formal distinction from the other quantum circuit structures considered here, although its importance for capturing the energies and correlation functions of finite systems in numerical studies must be established empirically.

III. NUMERICAL OPTIMIZATION STUDIES

A. Algorithms

The first question to answer in a numerical assessment of the variational power of an *Ansatz* is how to optimize it. In this section, we investigate how to optimize the quantum circuit tensor network and global circuit *Ansatz* considered in this work, by optimizing the parameters of the two-body unitary gates \otimes . We assume each two-body unitary is a general $SO(4)$ unitary (i.e., real-valued unitary) with a six-dimensional variational space [34]. We use two optimization algorithms: (i) a local “DMRG-like” optimization, where we sweep through the unitary gates \otimes , optimizing them one at a time while holding the others fixed, and (ii) a global gradient-based optimization, where all variational parameters are updated at the same time. In the local optimization scheme, a linearization of the problem, similar to the one used in MERA optimizations [24], is used to find locally optimal gates. In the global gradient-based scheme, the global gradient [i.e., the first-order derivative with respect to all variational parameters $\partial(\dots, \otimes, \otimes, \dots)$] is analytically calculated by automatic differentiation as supported in quimb [35]. The energy is computed via exact contraction [36] of either a matrix product operator representation of the Hamiltonian (MPS) or a sum of local terms (MERA). The unitary constraints are enforced by differentiating through a function that maps the gate parameters to a

unitary matrix. The cost-function minimization is then performed using either the conjugate gradient (CG) or limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS-B) algorithms [37,38]. The algorithms are stopped once the relative change in energy is less than 10^{-8} . The above choice of gradient descent algorithms can be modified if the gradient is susceptible to sampling noise (see Sec. V for a theoretical discussion of sampling overhead); for example, stochastic gradient methods or gradient with momentum methods may be preferred. However, in our limited studies, the different choices of gradient descent methods reach the same minimum, and thus do not impact our studies of expressivity. For some examples of quantum circuit tensor network optimization with sampling noise, see Ref. [39].

B. Model Hamiltonians

We choose to study the 1D Heisenberg and Fermi-Hubbard models:

$$\begin{aligned}\mathcal{H}_H &= J \sum_i \vec{S}_i \cdot \vec{S}_{i+1}, \\ \mathcal{H}_{FH} &= -t \sum_{i,\sigma} (c_{i\sigma}^\dagger c_{i+1\sigma} + \text{H.c.}) + U \sum_i c_{i\uparrow}^\dagger c_{i\uparrow} c_{i\downarrow}^\dagger c_{i\downarrow} \\ &\quad - \mu \sum_{i\sigma} c_{i\sigma}^\dagger c_{i\sigma},\end{aligned}$$

where \vec{S} are spin-1/2 operators and $c_\sigma^\dagger, c_\sigma$ are spin-1/2 fermionic creation and annihilation operators, respectively. For the Heisenberg model we use $J = 1$, and for the Hubbard model we use $t = 1$, $U = 3$, and $\mu = U/10$. In both cases, the ground states are gapless in the thermodynamic limit with algebraically decaying correlation functions, although in practice we will simulate finite chains with open boundary conditions.

C. Local optimization versus global optimization

We first compare local DMRG style optimization versus global gradient-based optimization for the infidelity cost function $\bar{\mathcal{F}} = 1 - |\langle \Psi | \psi \rangle|$, where $|\Psi\rangle$ is the ground-state wave function of the model and $|\psi\rangle$ is the *Ansatz* state. $|\Psi\rangle$ is obtained by the standard DMRG algorithm using a dense MPS of sufficiently large bond dimension ($D \sim 400$) so that any error in $|\Psi\rangle$ is negligible. When $\bar{\mathcal{F}} = 0$, then the circuit *Ansatz* is identical to the ground-state wave function.

The result of minimizing $\bar{\mathcal{F}}$ for the different *Ansatz* and optimization methods is shown in Fig. 5. We show data from the quantum circuit MPS with a brick-wall local *Ansatz* (QMPS-B) ($L = 24$, $q = 4$, $\tau = 6$) and global brick-wall quantum circuit (QC-B) ($L = 16$, $\tau = 6$) as representative examples. We find that in all cases, the local DMRG style optimization converges to the local minimum faster than the global gradient-based optimization using either the CG or L-BFGS-B algorithms. In addition, we find that in all cases, the L-BFGS-B algorithm

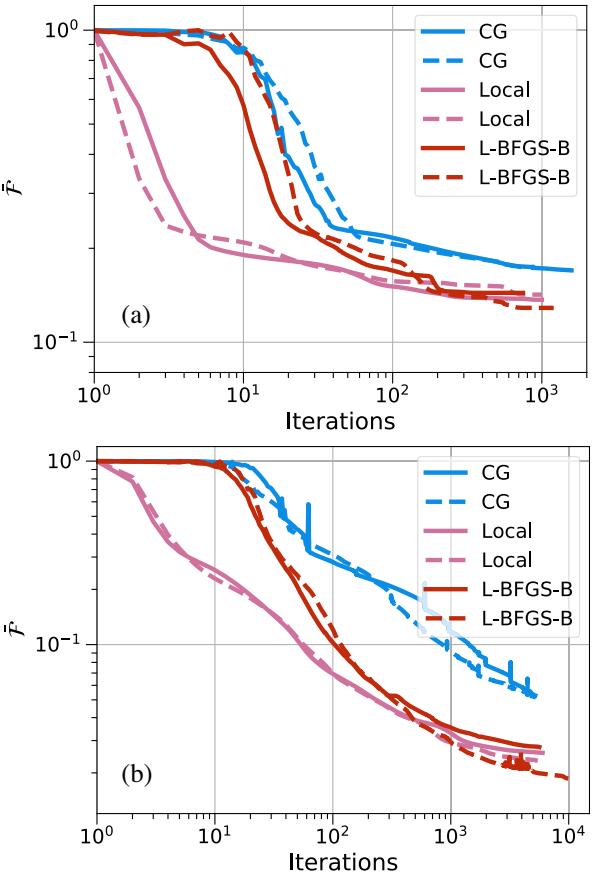


FIG. 5. The influence of the choice of optimization method. (a) The infidelity $\bar{\mathcal{F}}$ versus iteration number for the global brick-wall *Ansatz* QC-B with $L = 16$, $\tau = 6$, optimized by CG, L-BFGS-B, and local methods. The solid and dashed lines represent two different initial starting states, chosen from a uniform random distribution for the tensors. (b) The same plot for a QMPS-B with $L = 24$, $q = 4$, $\tau = 4$. The targeted wave function is the ground state of Heisenberg model \mathcal{H}_H .

converges more quickly than the CG algorithm. However, we also see that both the speed of convergence as well as the final converged result have some dependence on the initial guess. As observed in Fig. 5, given a suitable initial guess, the global gradient-based optimization eventually converges to a slightly lower minimum than that found by the local DMRG optimization.

D. Initial guess

The dependence of the optimization on the initial guess is well known in quantum circuit optimization, where poor initial guesses can sometimes give rise to exponentially small gradients (the barren plateau problem [40–44]). We can see a related problem in our circuits. To illustrate this, we show results from optimizing the energy cost function $E = \min \langle \psi | \mathcal{H} | \psi \rangle, |\psi\rangle \in \text{qTN}$ (quantum tensor network) and we report the relative energy error $\delta E = E/E_{\text{exact}} - 1$ versus iteration number in Fig. 7 for several example

circuits. We see in the top panel [Fig. 7(a)] that when starting from a random initial guess for the global QC-B *Ansatz*, we converge to the same relative error for two different circuit depths $\tau = 8$ and $\tau = 10$, despite the significantly larger number of variational parameters for $\tau = 10$ versus $\tau = 8$.

To improve the initialization of larger circuits, we use optimized gates obtained from a shallow circuit to initialize gates at larger circuit depth. The heuristic for this adaptive initialization method is summarized as follows: (i) optimal gates are obtained from a random initial guess for a low depth circuit τ' , (ii) the initial guess for the *Ansatz* with depth $\tau > \tau'$ is given by the optimized gates (from previous calculations) for $\tau \geq \tau - \tau'$ and the identity operator for $\tau < \tau - \tau'$, respectively (measuring depth from the register qubits) (see Fig. 6 for an explicit example), (iii) small random perturbations to all gates in the *Ansatz* are applied to avoid getting stuck in a local minimum, (iv) for a circuit *Ansatz* with larger depth, we repeat steps (ii) and (iii). Empirically, it is found that a gradual increase of the circuit depth by 2–6 layers works well, i.e., $\tau - \tau' = 2$ –6. The identity perturbation strength is also chosen to be of the same order as the local unitary gradient norm.

In Fig. 7(a), we show that optimizing the global QC-B *Ansatz* with $\tau = 10$ starting from optimal gates from $\tau = 8$

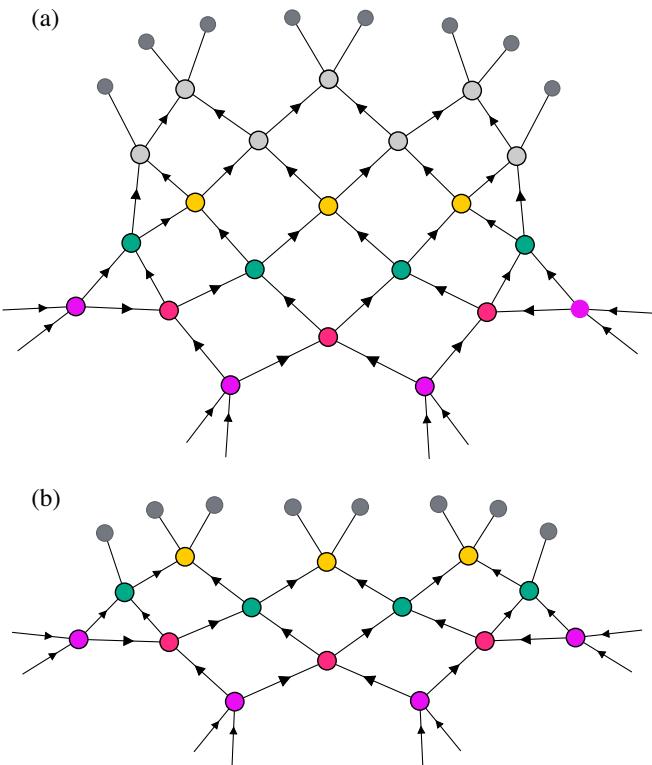


FIG. 6. Adaptive initialization of a QC-B circuit with (a) $\tau = 6$ from an optimized smaller circuit (b) with $\tau = 4$. The gray gates denote identity operators with some small random perturbations. Note that execution time flows from top to bottom (opposite to the arrow direction).

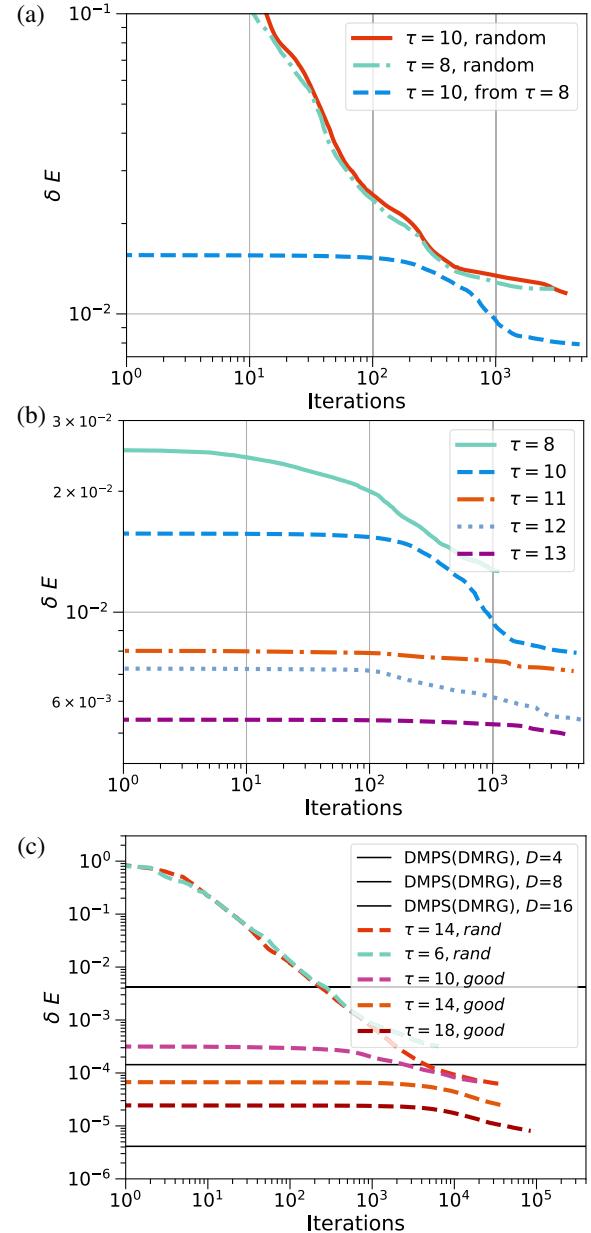


FIG. 7. The role of the initial guess in the optimization. (a) The relative error of the ground-state energy δE (of Heisenberg model \mathcal{H}_H) as a function of iteration number for a global brick-wall quantum circuit (QC-B) with depth τ . When the $\tau = 10$ *Ansatz* is initialized from the optimized $\tau = 8$ parameters, we obtain a better minimum than from a random initial guess. Notice that when using a random initial guess the *Ansatz* optimization can get stuck in a poor minimum, as seen by the $\tau = 8$ (random) and $\tau = 10$ (random) results, which obtain the same minimum. (b) A better initialization procedure using optimized circuit parameters from smaller depths guarantees that the relative error decreases monotonically when increasing circuit depth τ . (c) Similar plot for a QMPS *Ansatz* with $q = 4$. We similarly find the relative error drops monotonically when increasing τ using initial guesses from a smaller τ *Ansatz* (good), while initializing from random guesses (rand) results in optimizations which terminate at poor minima. Reference data from dense MPS (DMRG) also shown.

indeed results in a lower relative error. In Fig. 7(b), we plot the results of optimizing the QC-B *Ansatz* with different τ [each initialized in the manner above, using lower depth circuits as shown in Fig. 6(c)], which shows that we can now achieve a systematic decrease in the relative error δE as a function of increasing τ . Notice that in all cases, it is necessary to first perform many iterations to bring the *Ansatz* out of the local minimum generated by the smaller τ guess, before one observes a significant drop in the relative error. Similar results are shown for the QMPS-B *Ansatz* with fixed $q = 4$ and increasing τ for the local circuit (brick-wall *Ansatz*) in Fig. 7(c); we similarly see that we can achieve a systematic decrease in the relative error when increasing τ . Indeed, as we increase τ in the QMPS-B *Ansatz*, we obtain results that approach the dense MPS (DMRG) result with bond dimension $D = 2^q = 16$. As this is the lower bound for the variational energy of any QMPS with $q = 4$, our optimization heuristic using adaptive initialization thus fully realizes the variational power of the quantum circuit tensor network.

IV. VARIATIONAL POWER OF QUANTUM CIRCUIT TENSOR NETWORKS, DENSE TENSOR NETWORKS, AND GLOBAL QUANTUM CIRCUITS

A. Energies

Using the above optimization strategies, we can now systematically characterize the variational power of the different *Ansätze* discussed in this work for ground-state representation. As the measure of expressiveness, we use the relative energy error δE as a function of the number of variational parameters. (Related measures have been recently used to compare different variational wave functions [45].) The various *Ansätze* are optimized using the global gradient scheme with the L-BFGS-B algorithm, while the dense MPS results are obtained by the DMRG algorithm. The largest parametrized circuit *Ansätze* correspond to the following: (i) for QMPS-B and QMERA-B, the number of bond qubits and local circuit depths are $q = 8$, $\tau = 32$ and $q = 4$, $\tau = 8$, respectively, and (ii) for the global circuit ansatz QC-B and QC-L, the largest circuit depths are $\tau = 14$ and $\tau = 9$, respectively. Despite the large number of circuit parameters, we find that the optimization heuristics work to high accuracy, if sufficient iterations are used. For example, the smallest relative energy errors we find using the QMPS *Ansatz* are 10^{-6} and 10^{-4} for the Heisenberg and Fermi-Hubbard models with $L = 32$, respectively, using $\sim 10^{6-7}$ iterations.

We benchmark the performance of the QMPS, QMERA, and global QC *Ansatz* versus the dense MPS (DMRG) and dense MERA for the 1D Heisenberg and Fermi-Hubbard models with $L = 32$ (2D results are discussed in a later section). The key findings are as follows. (i) Comparing dense MPS with QMPS, we find that for an equivalent

number of variational parameters, QMPS achieves lower energies than the dense MPS [see Figs. 8(a) and 8(d)] in both the Heisenberg and Fermi-Hubbard models. (ii) Similarly, comparing dense MERA with QMERA, we find that for an equivalent number of variational parameters, QMERA achieves lower energies than dense MERA. Taken together with the previous statement, this implies that the appropriate quantum circuit tensor networks are more compact and expressive than their traditional dense counterparts for these problems. In particular, the worst case possibility, that one requires an exponential number of gates to accurately parametrize the local block unitary, does not apply to these physical ground states. (iii) Such expressiveness is not shared by the global brick-wall and ladder circuits [Figs. 8(b) and 8(e)] which are consistently less expressive than the QMPS. This is consistent with the theoretical analysis in Sec. II E, which identifies QC-B and QC-L as constrained versions of the QMPS with minimal parametrizations of the block unitaries. (iv) QMPS is somewhat more expressive than QMERA [Figs. 8(c) and 8(f)]. This suggests that the formal ability to capture logarithmic corrections (which exist in the thermodynamic limit of the 1D Heisenberg model) is unimportant either for the energy or at the system size considered.

Empirically, we can also summarize the data by fitting the relative error to the inverse polynomial $\delta E(n) \sim an^{-b}$, where n is the number of variational parameters. As shown in Fig. 8 in the log-log plot, this form fits all the *Ansätze* reasonably well, with some small systematic deviations, for example, in the case of DMRG at larger n . A linear fit to the log-log data yields an estimate of a and b , as shown in Table I, where b gives the asymptotic scaling for large n . These results further support the findings above: in the large n limit, the QMPS *Ansatz* is the most expressive *Ansatz* and outperforms the dense MPS, while QMERA outperforms the dense MERA, with both also outperforming the global brick-wall and ladder *Ansätze*. We see that the brick-wall local circuit structure for QMPS yields a better overall performance than local ladder or MERA structures (see also the Appendix). The QMERA *Ansatz* performs similarly to the dense MPS algorithm in both models with $b_{\text{QMERA}} \approx b_{\text{DMPS}}$. The ratio $b_{\text{QMPS}}/b_{\text{QMERA}}$ and $b_{\text{QMPS}}/b_{\text{DMPS}}$ is ~ 1.3 in both models. Earlier studies of scale-invariant MERA and infinite MPS have found a similar ratio, $b_{\text{MPS}}/b_{\text{MERA}} \sim 1.2$ [46]. Finally, depending on the model, either QC-B or QC-L performs better: $b_{\text{QC-B}} < b_{\text{QC-L}}$ ($b_{\text{QC-B}} > b_{\text{QC-L}}$) for the Heisenberg (Fermi-Hubbard) models, respectively.

B. Correlation functions

We next study how faithfully the different *Ansätze* capture correlation functions of the Heisenberg and Fermi-Hubbard models. We use the spin-spin correlation

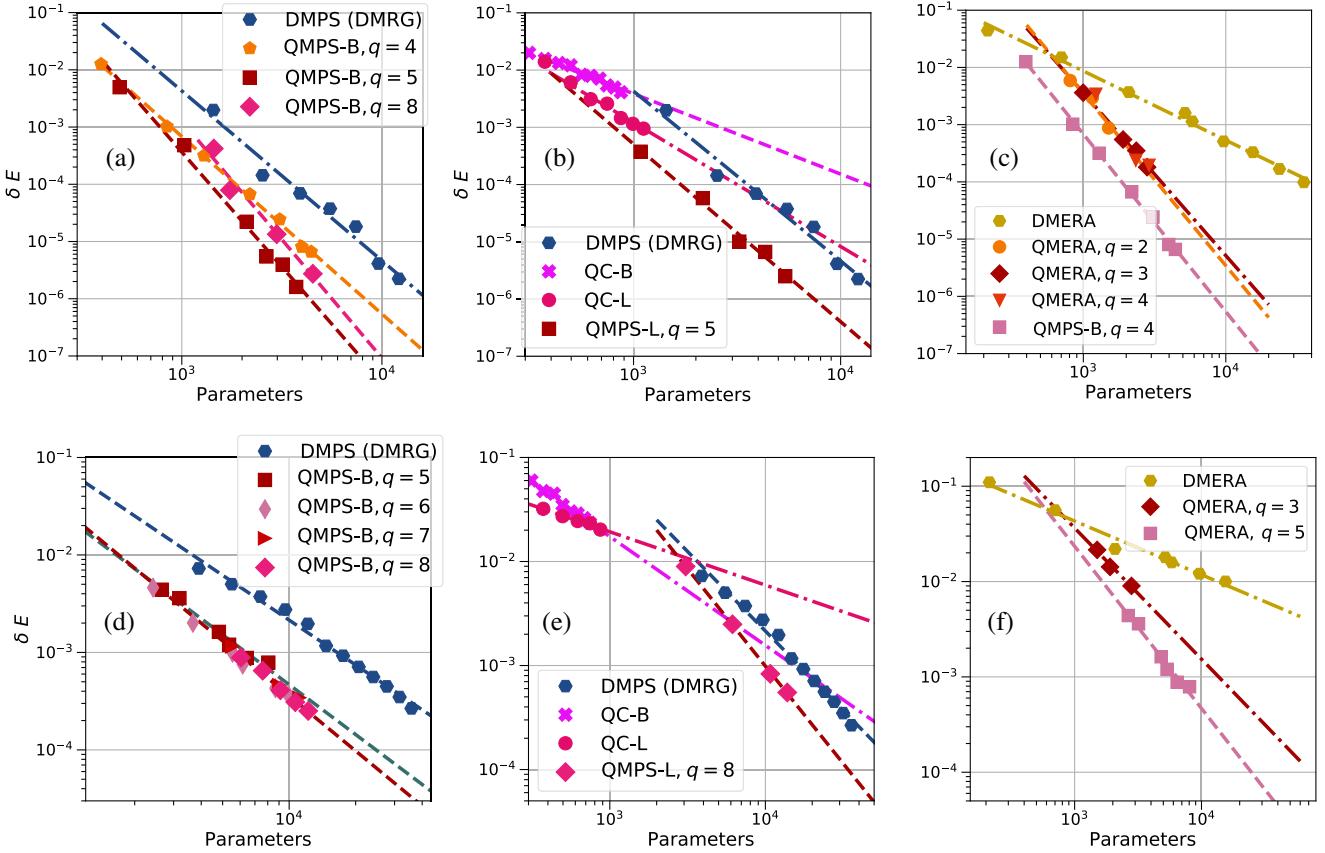


FIG. 8. Comparing the variational power of quantum circuit tensor networks (QMPS, QMERA), dense tensor networks (DMPS and DMERA), and global quantum circuits (QC). We show the relative energy error δE versus the number of variational parameters in the *Ansatz* for the Heisenberg (a)–(c) and Fermi-Hubbard (d)–(f) models with $L = 32$. Indices B and L stand for brick-wall and ladder. (a), (b) Comparison between QMPS with local brick-wall and ladder circuits with DMRG and QC with brick-wall and ladder structures. The largest circuit depth used for QC-B (L) and QMPS-B (L) with $q = 5$ is $\tau = 14$ (9). (c) The performance of QMERA-B versus QMPS-B and DMERA. The largest parameter sets used for the QMERA-B *Ansatz* correspond to $q = 4$, $\tau = 8$. (d)–(f) The same comparison for the Fermi-Hubbard model. The largest circuit depths for QMPS-B with $q = 8$ and QMERA-B with $q = 3$ are $\tau = 32$ and $\tau = 12$, respectively.

function, defined as $\langle \vec{S}_0 \cdot \vec{S}_r \rangle - \langle \vec{S}_0^2 \rangle$ (Heisenberg and Fermi-Hubbard) as a representative example.

In Fig. 9, we show errors in the correlation function $\delta C(r)$ (relative to numerically exact data) for the various *Ansätze*. In Fig. 9(a) (Heisenberg model), we see that

TABLE I. Scaling coefficients (a, b) in the form $\delta E \sim an^{-b}$ for the various *Ansätze* in the Heisenberg and Fermi-Hubbard models. The asymptotic behavior of the relative error δE at large n is controlled by b .

<i>Ansatz</i>	Heisenberg, (a, b)	Hubbard, (a, b)
QMPS-B	(20,4.0)	(9,1.9)
QMPS-L	(14,3.1)	(10,1.9)
QC-B	(4,1.4)	(4.4,1.0)
QC-L	(8,2.2)	(0.4,0.5)
QMERA-B	(15,3.1)	(6.0,1.4)
DMPS (DMRG)	(15,2.9)	(8.0,1.5)
DMERA	(3.5,1.2)	(0.8,0.6)

QMPS-B with $q = 4$, $\tau = 16$ produces the same algebraically decaying correlation function as a dense MPS with $D = 16$ over the full distance range of $r < 22$, despite having fewer variational parameters. In Fig. 9(b) (Hubbard model), we similarly find that QMPS-B with $q = 5$, $\tau = 30$ produces a similar quality correlation function to the dense MPS with $D = 32$, but again with fewer variational parameters. These results are consistent with the greater expressiveness of the quantum circuit tensor network relative to its dense counterpart.

In both models, we see that increasing either τ or q leads to an improvement of the correlation function. However, neither is a dominant factor for convergence. For example, in the Heisenberg model, we find that using a QMPS-B with $q = 5$, $\tau = 14$ provides a lower relative error compared to $q = 4$, $\tau = 18$, despite having fewer variational parameters, but a QMPS-B with $q = 8$, $\tau = 4$ performs similarly to $q = 4$, $\tau = 5$, despite having a large number of variational parameters. Thus a balanced choice of q , τ is needed to obtain the best result.

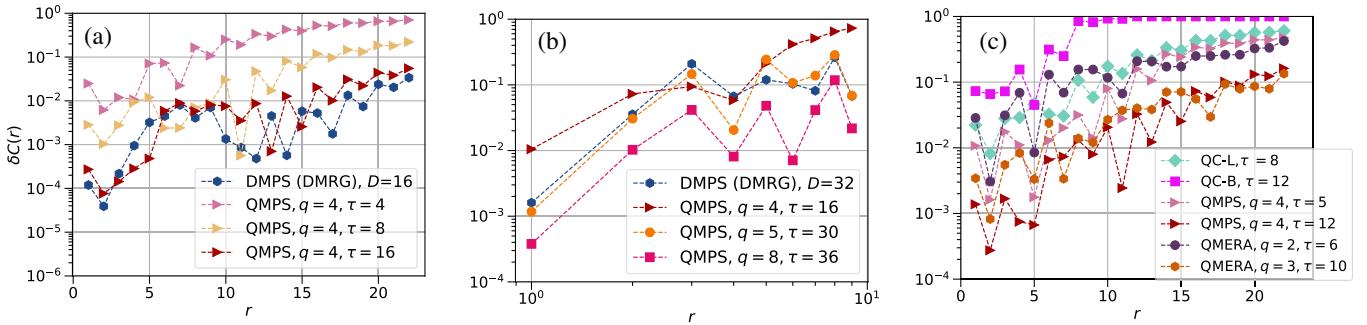


FIG. 9. Relative error in the spin-spin correlation function as a function of distance r . Panels (a) and (c) are for the Heisenberg model, while panel (b) is for the Fermi-Hubbard model (both $L = 32$). Both QMPS and QMERA use a local brick-wall circuit (QMPS-B, QMERA-B). In (a), the QMPS with $q = 4, \tau = 16$ has fewer parameters than the DMPS with $D = 16$, while in (b), the QMPS with $q = 5, \tau = 30$ has fewer parameters than the DMPS with $D = 32$. In (c), the QC-L *Ansatz* has a comparable numbers of parameters to the QMERA-B and QMPS-B with $q = 2, \tau = 6$ and $q = 4, \tau = 5$, respectively. QMPS-B ($q = 4, \tau = 12$) and QMERA-B ($q = 3, \tau = 10$) have comparable numbers of variational parameters. QC-B correlation functions are generally worse than those from the above *Ansatz*.

In Fig. 9(c), we show the QMERA-B, QC-L, and QC-B correlation functions alongside the QMPS-B correlation functions. One expects that QMPS-B will accurately reproduce short-range correlations (up to the MPS correlation length) while QMERA should perform better at long distances. Quantitatively, we find that QMPS-B ($q = 4, \tau = 12$ and $\delta E = 6 \times 10^{-5}$) indeed provides a lower relative error at short distances, while QMERA-B (with $q = 3, \tau = 10$ and $\delta E = 3 \times 10^{-4}$) with a similar number

of variational parameters is more accurate at longer distances ($r > 15$). In addition, QMPS with $q = 4, \tau = 5$ outperforms QC-L with a similar number of parameters, especially at short distances. At long range, QMERA-B with $q = 2, \tau = 6$ is clearly better than QC-L, while at short distances it is comparable. Overall, QMPS and QMERA thus appear to provide more faithful representations of the correlation functions than the general quantum circuit *Ansatz*, again reinforcing the need to balance the spreading of entanglement and the accuracy of the local block unitary representation. In addition, the improved entanglement spreading structure of the QMERA is detected in the correlation functions, even though it is not represented in the energy metric of the previous section.

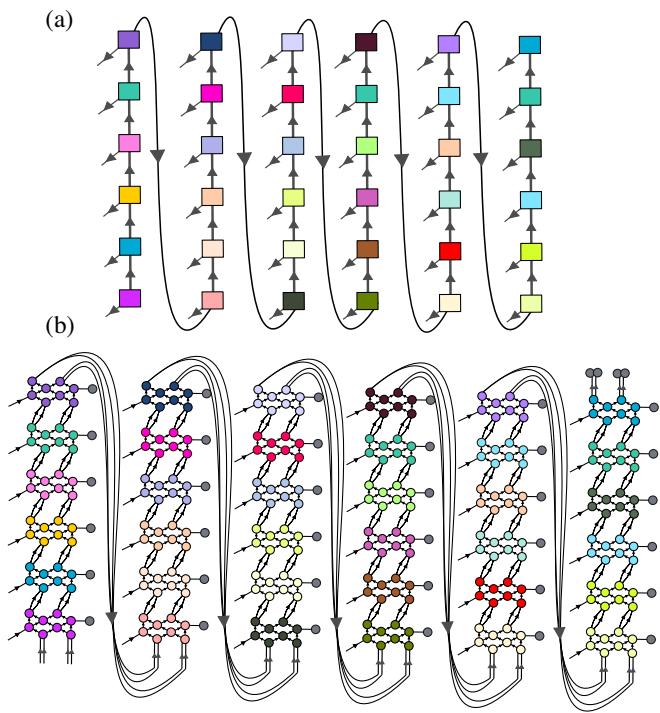


FIG. 10. Schematic representation of (a) a snake DMPS and (b) its quantum circuit (QMPS) used in the two-dimensional simulation of Heisenberg model with $L = 6 \times 6$ sites. The QMPS-B shown has four bond qubits $q = 4$ with depth $\tau = 4$.

C. Two-dimensional systems

We have also carried out a limited set of studies on a two-dimensional Heisenberg model using the QMPS *Ansatz* made in a traditional snake through the two-dimensional lattice as depicted in Fig. 10. In two-dimensional systems, it

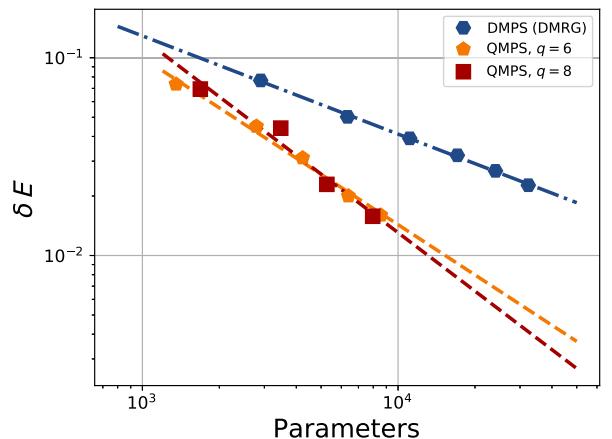


FIG. 11. Comparison between QMPS and DMPS for the two-dimensional Heisenberg model on a 6×6 lattice.

TABLE II. Scaling coefficients (a, b) for QMPS and DMPS in the two-dimensional Heisenberg model on a 6×6 lattice.

Ansatz	2D Heisenberg, (a, b)
QMPS-B	(4.7, 1.1)
DMPS (DMRG)	(1.4, 0.48)

is well known that a traditional dense MPS (arranged as a snake) requires an exponentially large bond dimension in the system width to represent area law ground states. Since states with exponentially large bond dimension can be reached by QMPS with a polynomial circuit depth, one might expect a larger expressivity difference between QMPS and DMPS in 2D. The relative energy error δE is plotted as a function of the number of parameters in Fig. 11, and the corresponding fit of the relative error $\delta E(n) \sim an^{-b}$ is shown in Table II. From both of these, we indeed find that the expressivity gap between QMPS and DMPS is greater in 2D, with $b_{\text{QMPS}}/b_{\text{DMPS}} = 1.4$ in 1D, but $b_{\text{QMPS}}/b_{\text{DMPS}} = 2.3$ in 2D. Of course, in two and more dimensions, the possible choices of circuit architecture are richer, with many types of isometric tensor networks to explore beyond the QMPS [21,22,47,48].

V. DISCUSSION OF QUANTUM ADVANTAGE

Our numerical results clearly show that for ground states of some commonly considered physical models, there is a quantum advantage in the expressivity of quantum circuit tensor networks versus the traditionally employed classical dense tensor networks. The difference in scaling of the achieved accuracy as a function of the number of parameters suggests that the advantage in expressiveness will persist into a regime where the circuits can no longer be contracted efficiently classically. Assuming standard tensor networks are the best classical representation for these problems, this means that in the high accuracy regime, quantum circuit tensor networks may have the (theoretical) potential to achieve quantum advantage also in terms of computational cost.

To explicitly translate the advantage in representation to one of computation, we must consider the cost to compute with the quantum circuit representation (on a quantum device) versus the dense representation on a classical device. We consider the case of MPS as an example. The cost to contract a classical dense MPS is $O(D^3)$, or $O(n^{3/2})$, where n here denotes the number of local parameters in the block unitary. (We ignore scaling with L here and below.) For QMPS, the cost to run the circuit to prepare the state (assuming gates are executed sequentially) is $O(n)$. We then imagine computing the energy by sampling terms in the Hamiltonian; for a relative precision δE , we require $O(1/(\delta E)^2)$ measurements per term. Combining these factors together with the empirical scaling of δE with n , one finds that the cost T to compute the

energy to an accuracy of δE is $T \sim \delta E^{0.52}$ (classical DMPS) and $T \sim \delta E^{2.25}$ (QMPS-B) in the 1D Heisenberg model, and $T \sim \delta E^{3.1}$ (classical DMPS) and $T \sim \delta E^{2.9}$ (QMPS-B) in the 2D Heisenberg model. A similar analysis for MERA finds in the 1D Hubbard model $T \sim \delta E^{3.75}$ (classical DMERA) and $T \sim \delta E^{2.7}$ (QMERA). These small polynomial advantages (where they appear) are perhaps reflective of the challenges of variational quantum algorithms, and whether they are realizable, or persist with improved classical techniques remains to be seen. However, it should be noted that the asymptotic inefficiency of the quantum algorithm stems from the steep cost associated with sampling expectation values. Techniques that trade coherence for reduced sampling, for example, reducing the number of repetitions to as few as $O(\log(1/(\delta E)))$, with a measurement circuit depth proportional to $O(1/(\delta E))$, therefore greatly affect this analysis of computational advantage [49].

Of course, to seize this potential advantage, one would also need to optimize circuit parameters in this classically intractable regime. In the case of gradient optimization, classical algorithms obtain the gradient at the same cost as the energy through backpropagation, but for quantum algorithms using finite differences (for example, using the parameter shift rule) [50,51], the energy evaluation must be repeated $O(n)$ times. In this case, the above polynomial advantages will disappear unless coherent expectation value techniques are used, which may be further combined with coherent techniques for gradient evaluation [52]. Also, we have assumed that the number of optimization iterations needed to find the ground state scales with δE in a comparable way in the quantum and classical computations. While the optimization heuristics discussed in this work successfully find accurate ground states with a tractable number of optimization iterations, it remains to be seen whether this scaling persists in very large circuits.

VI. CONCLUSIONS

In this work, we studied the variational power of quantum circuit tensor networks, and in particular, quantum circuit matrix product states and the quantum circuit multiscale entanglement renormalization *Ansatz*, for representing the ground states of quantum many-particle problems. As we argued, this is a problem where standard tensor networks excel, and is thus a high bar for quantum circuit tensor networks to meet. We found that quantum circuit tensor networks outperform other common global quantum circuit *Ansätze* in variational power, requiring far fewer parametrized gates for a given accuracy. In fact, they appear to be asymptotically even more expressive than the standard tensor networks, in terms of the number of parameters to converge to a comparable accuracy in the variational energy and correlation functions. Our initial

results in 2D suggest that this expressive advantage increases in higher dimensions.

Although all simulations here were carried out classically, the difference in expressiveness of the classical and quantum circuit tensor networks raises the possibility of polynomial quantum advantage in the computation of certain problems. The practical realization of such advantage critically depends on both the performance of optimization heuristics (such as the one proposed here) as well as the cost of estimating expectation values on quantum devices. However, the variational power of the quantum circuit tensor networks provides new motivation to improve the optimization strategies for this class of circuits. It also provides impetus to study related types of *Ansätze* in the context of classical simulations, where they may provide the chance to improve on long-standing tensor network paradigms.

The numerical codes were implemented in the quimb library [35], which is freely available [53,54].

ACKNOWLEDGMENTS

Primary funding for this work (R. H., G. K.-L. C.) was provided by the U.S. Department of Energy, Office of Science, via DE-SC0019374. G. K.-L. C. acknowledges additional support from the Simons Foundation via the Many-Electron Collaboration and via the Simons Investigator program. A. C. P. was supported by the U.S. NSF Convergence Accelerator Track C Grant No. 2040549. Support for J. G. and the development of the quimb library was provided by a gift from Amazon Web Services, Inc.

APPENDIX: QMPS WITH MERA LOCAL CIRCUITS

We benchmark the accuracy of the QMPS-M *Ansatz* by studying the relative error δE for the Heisenberg model \mathcal{H}_H . In general, a systematic study of the *Ansatz* is difficult, as there are many controlling parameters: q, q_m, τ where q_m is the number of bond qubits for the local MERA; and in addition, there is also the choice of the internal structure of the local MERA. To simplify things, we fix the number of bond qubits to $q = 8$ and use a brick-wall structure for the MERA with $q_m = 3$. Empirically, we find that it is difficult to obtain converged results for QMPS-M as it easily gets stuck in local minima. Thus, for the reported data, two different initial states were chosen, one random and one obtained from a smaller optimized circuit. In Fig. 13(a), we compare QMPS-M with QMPS-B for different τ . We find that the QMPS-B with $\tau = 4$ easily outperforms QMPS-M with $\tau = 4$ (which has a larger number of variational parameters). We see that increasing τ in QMPS-M from 4 to 10 only slightly improves the relative error, i.e., from

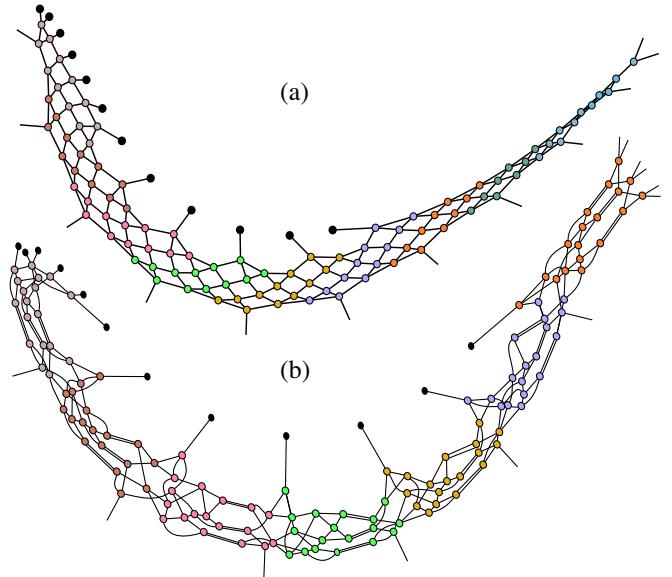


FIG. 12. Schematic representation of a QMPS circuit with (a) ladder and (b) MERA internal local circuits, with local depth $\tau = 2, 4$ and bond qubits $q = 5, 6$, respectively.

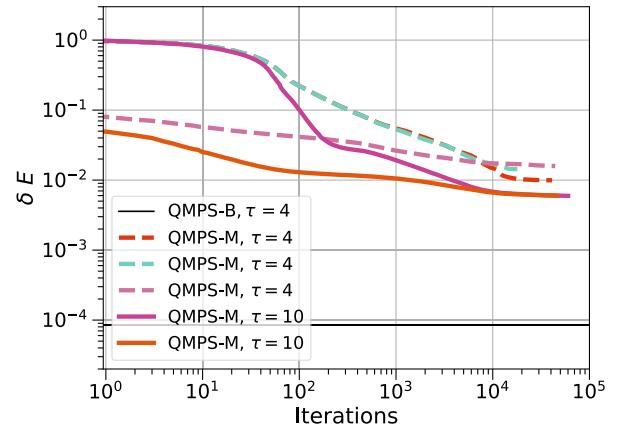


FIG. 13. Comparison between QMPS with MERA and brick-wall local circuits. The solid and dashed lines represent two different initial starting states.

9×10^{-3} to 6×10^{-3} . Overall, the complexity of this circuit structure may require additional improvements in optimization strategy in order to realize its variational power.

- [1] I. H. Kim and B. Swingle, *Robust Entanglement Renormalization on a Noisy Quantum Computer*, arXiv:1711.07500.
- [2] I. Cong, S. Choi, and M. D. Lukin, *Quantum Convolutional Neural Networks*, Nat. Phys. **15**, 1273 (2019).
- [3] W. Huggins, P. Patil, B. Mitchell, K. B. Whaley, and E. M. Stoudenmire, *Towards Quantum Machine Learning with Tensor Networks*, Quantum Sci. Technol. **4**, 024001 (2019).

[4] 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. Research* **3**, 033002 (2021).

[5] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, *Quantum Algorithms for Quantum Chemistry and Quantum Materials Science*, *Chem. Rev.* **120**, 12685 (2020).

[6] M. Lubasch, J. Joo, P. Moinier, M. Kiffner, and D. Jaksch, *Variational Quantum Algorithms for Nonlinear Problems*, *Phys. Rev. A* **101**, 010301(R) (2020).

[7] F. Barratt, J. Dborin, M. Bal, V. Stojovic, F. Pollmann, and A. G. Green, *Parallel Quantum Simulation of Large Systems on Small NISQ Computers*, *npj Quantum Inf.* **7**, 79 (2021).

[8] M. Foss-Feig, S. Ragole, A. Potter, J. Dreiling, C. Figgatt, J. Gaebler, A. Hall, S. Moses, J. Pino, B. Spaun *et al.*, *Entanglement from Tensor Networks on a Trapped-Ion QCCD Quantum Computer*, [arXiv:2104.11235](https://arxiv.org/abs/2104.11235).

[9] E. Chertkov, J. Bohnet, D. Francois, J. Gaebler, D. Gresh, A. Hankin, K. Lee, R. Tobey, D. Hayes, B. Neyenhuis *et al.*, *Holographic Dynamics Simulations with a Trapped Ion Quantum Computer*, [arXiv:2105.09324](https://arxiv.org/abs/2105.09324).

[10] A. Uvarov, J. D. Biamonte, and D. Yudin, *Variational Quantum Eigensolver for Frustrated Quantum Systems*, *Phys. Rev. B* **102**, 075104 (2020).

[11] 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).

[12] L. Slattery, B. Villalonga, and B. K. Clark, *Unitary Block Optimization for Variational Quantum Algorithms*, [arXiv:2102.08403](https://arxiv.org/abs/2102.08403).

[13] R. Haghshenas, *Optimization Schemes for Unitary Tensor-Network Circuit*, *Phys. Rev. Research* **3**, 023148 (2021).

[14] G. Vidal, *Class of Quantum Many-Body States That Can be Efficiently Simulated*, *Phys. Rev. Lett.* **101**, 110501 (2008).

[15] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, *Rigorous Results on Valence-Bond Ground States in Antiferromagnets*, *Phys. Rev. Lett.* **59**, 799 (1987).

[16] F. Verstraete, M. M. Wolf, D. Perez-Garcia, and J. I. Cirac, *Criticality, the Area Law, and the Computational Power of Projected Entangled Pair States*, *Phys. Rev. Lett.* **96**, 220601 (2006).

[17] F. Verstraete, V. Murg, and J. I. Cirac, *Matrix Product States, Projected Entangled Pair States, and Variational Renormalization Group Methods for Quantum Spin Systems*, *Adv. Phys.* **57**, 143 (2008).

[18] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, *Matrix Product State Representations*, *Quantum Inf. Comput.* **7**, 401 (2007).

[19] G. Vidal, *Efficient Classical Simulation of Slightly Entangled Quantum Computations*, *Phys. Rev. Lett.* **91**, 147902 (2003).

[20] S. R. White, *Density Matrix Formulation for Quantum Renormalization Groups*, *Phys. Rev. Lett.* **69**, 2863 (1992).

[21] R. Haghshenas, M. J. O'Rourke, and G. K.-L. Chan, *Conversion of Projected Entangled Pair States into a Canonical Form*, *Phys. Rev. B* **100**, 054404 (2019).

[22] M. P. Zaletel and F. Pollmann, *Isometric Tensor Network States in Two Dimensions*, *Phys. Rev. Lett.* **124**, 037201 (2020).

[23] C. Schön, K. Hammerer, M. M. Wolf, J. I. Cirac, and E. Solano, *Sequential Generation of Matrix-Product States in Cavity QED*, *Phys. Rev. A* **75**, 032311 (2007).

[24] G. Evenbly and G. Vidal, *Algorithms for Entanglement Renormalization*, *Phys. Rev. B* **79**, 144108 (2009).

[25] G. Evenbly and G. Vidal, *Tensor Network States and Geometry*, *J. Stat. Phys.* **145**, 891 (2011).

[26] Empirically, we find that QMERA with local brick-wall and ladder circuits variationally behave very similarly, at least within the range of accessible qubits q and circuit depth τ presented in this work.

[27] H. J. Changlani, J. M. Kinder, C. J. Umrigar, and G. K.-L. Chan, *Approximating Strongly Correlated Wave Functions with Correlator Product States*, *Phys. Rev. B* **80**, 245116 (2009).

[28] E. Neuscamman, H. Changlani, J. Kinder, and G. K.-L. Chan, *Nonstochastic Algorithms for Jastrow-Slater and Correlator Product State Wave Functions*, *Phys. Rev. B* **84**, 205132 (2011).

[29] S. R. Clark, *Unifying Neural-Network Quantum States and Correlator Product States via Tensor Networks*, *J. Phys. A* **51**, 135301 (2018).

[30] S. Al-Assam, S. R. Clark, C. J. Foot, and D. Jaksch, *Capturing Long Range Correlations in Two-Dimensional Quantum Lattice Systems Using Correlator Product States*, *Phys. Rev. B* **84**, 205108 (2011).

[31] F. Mezzacapo, N. Schuch, M. Boninsegni, and J. I. Cirac, *Ground-State Properties of Quantum Many-Body Systems: Entangled-Plaquette States and Variational Monte Carlo*, *New J. Phys.* **11**, 083026 (2009).

[32] J. Thibaut, T. Roscilde, and F. Mezzacapo, *Long-Range Entangled-Plaquette States for Critical and Frustrated Quantum Systems on a Lattice*, *Phys. Rev. B* **100**, 155148 (2019).

[33] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, *Neural-Network Quantum States, String-Bond States, and Chiral Topological States*, *Phys. Rev. X* **8**, 011006 (2018).

[34] A further reduction in the number of parameters is possible by eliminating one parameter along the connected bonds, absorbing a rotational gate into adjacent gates, and also 6 parameters for two gates acting on top of each other. In our reported data, we have considered these reductions, which mainly change the scaling prefactor a , not the important exponent b .

[35] J. Gray, *quimb: A PYTHON Library for Quantum Information and Many-Body Calculations*, *J. Open Source Software* **3**, 819 (2018).

[36] J. Gray and S. Kourtis, *Hyper-Optimized Tensor Network Contraction*, *Quantum* **5**, 410 (2021).

[37] J. Nocedal, *Updating Quasi-Newton Matrices with Limited Storage*, *Math. Comput.* **35**, 773 (1980).

[38] J. Nocedal and S. Wright, *Numerical Optimization* (Springer-Verlag, New York, 2006).

[39] J.-G. Liu, Y.-H. Zhang, Y. Wan, and L. Wang, *Variational Quantum Eigensolver with Fewer Qubits*, *Phys. Rev. Research* **1**, 023025 (2019).

[40] 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).

[41] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, *An Initialization Strategy for Addressing Barren Plateaus in Parametrized Quantum Circuits*, *Quantum* **3**, 214 (2019).

[42] T. Volkoff and P. J. Coles, *Large Gradients via Correlation in Random Parameterized Quantum Circuits*, *Quantum Sci. Technol.* **6**, 025008 (2021).

[43] 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).

[44] C. Zhao and X.-S. Gao, *Analyzing the Barren Plateau Phenomenon in Training Quantum Neural Networks with the ZX-Calculus*, *Quantum* **5**, 466 (2021).

[45] N. H. Stair and F. A. Evangelista, *Exploring Hilbert Space on a Budget: Novel Benchmark Set and Performance Metric for Testing Electronic Structure Methods in the Regime of Strong Correlation*, *J. Chem. Phys.* **153**, 104108 (2020).

[46] A. Avella and F. Mancini, *Strongly Correlated Systems: Numerical Methods* (Springer-Verlag, Berlin, 2013), Vol. 124.

[47] L. Slattery and B. K. Clark, *Quantum Circuits for Two-Dimensional Isometric Tensor Networks*, [arXiv:2108.02792](https://arxiv.org/abs/2108.02792).

[48] I. MacCormack, A. Galda, and A. L. Lyon, *Simulating Large PEPs Tensor Networks on Small Quantum Devices*, [arXiv:2110.00507](https://arxiv.org/abs/2110.00507).

[49] D. Wang, O. Higgott, and S. Brierley, *Accelerated Variational Quantum Eigensolver*, *Phys. Rev. Lett.* **122**, 140504 (2019).

[50] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, *Quantum Circuit Learning*, *Phys. Rev. A* **98**, 032309 (2018).

[51] L. Banchi and G. E. Crooks, *Measuring Analytic Gradients of General Quantum Evolution with the Stochastic Parameter Shift Rule*, *Quantum* **5**, 386 (2021).

[52] A. Gilyén, S. Arunachalam, and N. Wiebe, *Optimizing Quantum Optimization Algorithms via Faster Quantum Gradient Computation*, in *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms* (SIAM, San Diego, CA, 2019), pp. 1425–1444.

[53] <https://github.com/jcmgray/quimb>.

[54] <https://github.com/rezah/qMPS>.