OPEN ACCESS

Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

PAPER: Quantum statistical physics, condensed matter, integrable systems

Entanglement diagnostics for efficient VQA optimization

Joonho Kim[1,*] **and Yaron Oz**[1,2]

[1] School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540, United States of America
[2] Raymond and Beverly Sackler School of Physics and Astronomy, Tel-Aviv University, Tel-Aviv 69978, Israel
E-mail: joonhokim@ias.edu

Received 14 December 2021
Accepted for publication 30 May 2022
Published 15 July 2022

Online at stacks.iop.org/JSTAT/2022/073101
https://doi.org/10.1088/1742-5468/ac7791

Abstract. We consider information spreading measures in randomly initialized variational quantum circuits and introduce entanglement diagnostics for efficient variational quantum/classical computations. We establish a robust connection between entanglement measures and optimization accuracy by solving two eigensolver problems for Ising Hamiltonians with nearest-neighbor and long-range spin interactions. As the circuit depth affects the average entanglement of random circuit states, the entanglement diagnostics can identify a high-performing depth range for optimization tasks encoded in local Hamiltonians. We argue, based on an eigensolver problem for the Sachdev–Ye–Kitaev model, that entanglement alone is insufficient as a diagnostic to the approximation of volume-law entangled target states and that a large number of circuit parameters is needed for such an optimization task.

Keywords: entanglement entropies, quantum computation, quantum information

*Author to whom any correspondence should be addressed.

1742-5468/22/073101+30$33.00

Contents

1. Introduction

Noisy intermediate-scale quantum (NISQ) technology is being developed rapidly and poses a great challenge to come up with efficient quantum algorithms [1], which will

J. Stat. Mech. (2022) 073101

operate on the NISQ computers and perform better than classical algorithms. Many real-world use cases are associated with machine learning and optimization, for which variational quantum circuits offer an appropriate framework. See [2] and references therein. The typical optimization tasks can be formulated as a search for the ground state of a Hamiltonian H, which may encode an exact combinatorial problem [3, 4].

The variational quantum algorithms (VQA) consist of two elements [5]. The first part is to construct a parameterized quantum circuit composed of L unitary layers on the product state of n qubits, $|0\rangle^{\otimes n}$. The layer unitaries and quantum gates therein depend on continuous parameters, each initialized with the uniform measure on $[0, 2\pi)$. Denoting all circuit parameters collectively by θ, the variational state is written as

$$|\psi_{\rm c}(\theta)\rangle = U(\theta)|0\rangle^{\otimes n}. \tag{1}$$

The second part of the variational quantum algorithm is to estimate the Hamiltonian expectation value with the variational circuit state, i.e.

$$E(\theta) = \langle\psi_{\rm c}(\theta)|H|\psi_{\rm c}(\theta)\rangle, \tag{2}$$

then minimize it in the nL-dimensional parameter space using the gradient descent method.

Entanglement encodes information in the qubit correlations, which are generated by the successive application of the circuit layers. Given two complementary systems A/B, the Rényi-k entropy of the reduced density matrix

$$\mathcal{R}_A^k = \frac{1}{1-k}\log\,{\rm Tr}\left(\rho_A^k\right) \tag{3}$$

measures their entanglement, so does the von Neumann entropy that corresponds to (3) in the special limit $k \to 1$:

$$S_{EE} = -\,{\rm Tr}\;\rho_A\log\,\rho_A. \tag{4}$$

The reduced density matrix ρ_A is obtained from the full circuit density matrix $\rho_{\rm c}(\theta) = |\psi_{\rm c}(\theta)\rangle\langle\psi_{\rm c}(\theta)|$ by taking a partial trace over the subsystem B.

The performance of the variational quantum algorithm depends largely on whether the quantum circuit can prepare an initial variational state $|\psi_{\rm c}(\theta)\rangle$ that is close to the target ground state $|\psi_{\rm g}\rangle$ of the Hamiltonian. In this paper we argue that the average entanglement entropy (3) or (4) of random circuit states provides a distance measure that can quantify a successful minimization of the energy function. Note that, for their computation, we specifically use the equal partition $n_A = n_B = n/2$ and the binary logarithm.

The evolution of the entanglement entropies and the geometric distance as a function of the circuit depth L is schematically drawn in figure 1. See appendix B2 for a brief introduction to the geometric measure. It is convenient to divide the range of L into three regions A, B, and C.[3] A is where the entanglement entropy continues to grow, while C is where it has saturated to a constant value. As for their scaling behavior in n, the random circuit states in A/C obey the area/volume law scaling of the entanglement entropies,

[3] Such usage of A, B, C should be distinguishable from the other usage of A, B that denotes a subset of n qubits.

J. Stat. Mech. (2022) 073101

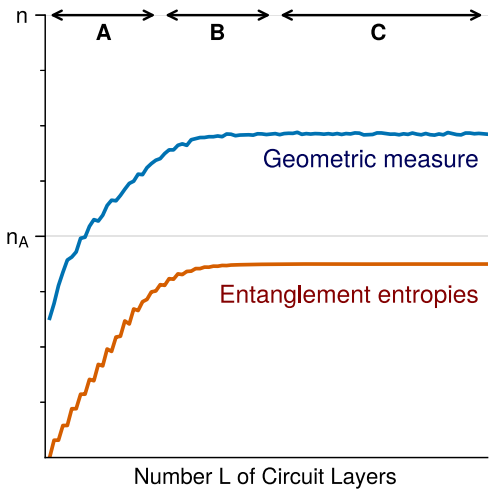


Figure 1. A schematic plot for the growth of mean entanglement entropies and geometric distance of circuit-generated states. We often distinguish the low, intermediate, high-depth circuits in sections 3 and 4 and denote their corresponding depth ranges by A, B, and C, respectively.

respectively. Since the ground states of gapped local Hamiltonians are expected to have an area law entanglement entropy, we expect that an initial variational state $|\psi_c(\theta)\rangle$ in A would lead to efficient VQA optimization in contrast to those circuits in C. We also identify B as a transition region between A and C, where the entanglement entropy has already reached saturation yet the initial random parameter can determine the success/failure of the VQA optimization.

The technical reason why the circuit optimization fails in region C is the vanishing gradient problem. When the circuit distribution is approximately two-design, such that the first and second moments are indistinguishable from those of the Haar distribution, the energy gradient at initial random values cannot deviate from zero, except for an exponentially decaying probability in n [6–9]. It happens for the circuit ensemble in B/C, where the Rényi-2 entropy as a diagnostic of the quantum two-design is closest to $n_A = n/2$ that corresponds to the Haar ensemble.

Until now, we assumed that the entanglement entropy of the target state follows the area law scaling, as in gapped local one-dimensional systems [10]. However, it does not always hold, and the variational circuit in A cannot minimize the circuit energy (2) to the ground level. For the Sachdev–Ye–Kitaev (SYK) model [11–13], whose ground state exhibits a volume law entanglement [14], the optimization does fail no matter to which of $A/B/C$ the variational circuit belongs. Incidentally, we argue that higher-dimensional parameter space can assist the circuit optimization even at high level of circuit state entanglement, so that over-parameterized circuits can offer a high precision approximation of volume law entangled target states including the SYK ground state [15–19].

The rest of this paper is organized as follows: section 2 motivates the entanglement diagnostics as the initialization condition to arrange variational states close to the target. Section 3 studies the average entanglement growth of circuit states as a function of the

J. Stat. Mech. (2022) 073101

circuit depth. Section 4 examines the importance of the entanglement diagnostics in the local gradient search of optimal circuit parameters. Section 5 checks the validity of the entanglement diagnostics by testing them against different circuit architectures and also discuss the impact of shrinking the circuit parameter dimension. The paper concludes with discussion and outlook in section 6. All our simulations are conducted at the level of state vectors, not suffering from the issue of sampling noise or deficiency of quantum gates. Additional details are given in the appendices.

2. Entanglement diagnostics

Using the density matrix of the quantum circuit $\rho_{\mathrm{c}}(\theta)$, the expectation value of the Hamiltonian (2) reads:

$$E(\theta) = \mathrm{Tr}\left(\rho_{\mathrm{c}}(\theta)H\right). \tag{5}$$

Our optimization task is to get as close as possible to the ground state of the Hamiltonian by minimizing (5). It can be achieved by multiple iterations of evaluating the density matrix $\rho_{\mathrm{c}}(\theta)$ and updating the parameters via the gradient descent (26) that will finally stop at $\theta = \theta_{\mathrm{f}}$. We would like to reach the final parameter θ_{f} such that

$$\Delta E \equiv \mathrm{Tr}\left((\rho_{\mathrm{c}}(\theta_{\mathrm{f}}) - \rho_{\mathrm{g}})H\right) \simeq 0 \tag{6}$$

where ρ_{g} is the exact ground state of the Hamiltonian. A simple upper bound of this approximation error ΔE follows from the Cauchy–Schwarz inequality,

$$\mathrm{Tr}\left((\rho_{\mathrm{c}}(\theta_{\mathrm{f}}) - \rho_{\mathrm{g}})H\right) \leqslant \|\rho_{\mathrm{c}}(\theta_{\mathrm{f}}) - \rho_{\mathrm{g}}\|_1 \cdot \|H\|_1, \tag{7}$$

where the trace norm $\|O\|_1$ is the sum of singular values of an operator O, i.e. eigenvalues of $(O^\dagger O)^{1/2}$.

A natural condition for efficient reduction of ΔE is arranging an initial circuit state $\rho_{\mathrm{c}}(\theta_{\mathrm{in}})$ to be in the proximity of the ground state with a small enough trace distance $\|\rho_{\mathrm{c}}(\theta_{\mathrm{in}}) - \rho_{\mathrm{g}}\|_1$. However, we will confront two issues. First, we generally do not know the ground state, thus being unable to estimate the trace distance $\|\rho_{\mathrm{c}}(\theta_{\mathrm{in}}) - \rho_{\mathrm{g}}\|_1$. Second, the trace distance can be considerably large even when the energy difference between quantum states is minuscule. So the above condition is often over-restrictive, discarding most reasonable initial states.

Instead, we want to relax the condition by using the entanglement entropy of an initial circuit state as a distance proxy between $\rho_{\mathrm{c}}(\theta_{\mathrm{in}})$ and ρ_{g}, from which one can expect the success/failure of circuit optimization. It can be motivated as follows: the inequalities on the von Neumann and Rényi-k entropy differences [20–23]:

$$|S_{EE}(\rho_A) - S_{EE}(\sigma_A)| \leqslant n\|\rho_A - \sigma_A\|_1 + (e\ \ln\ 2)^{-1}, \tag{8}$$

$$|\mathcal{R}^k(\rho_A) - \mathcal{R}^k(\sigma_A)| \leqslant k \cdot 2^{n_A(k-1)}\|\rho_A - \sigma_A\|_1, \tag{9}$$

J. Stat. Mech. (2022) 073101

show that, for given two quantum states ρ_A and σ_A, being close in their quantum entropies is necessary for being close in their trace distance $\|\rho_A - \sigma_A\|_1$. Suppose now that ρ_A and σ_A are the density matrices of a subsystem A, which can be combined with a complementary part B to constitute the entire n qubit system, i.e. $\rho_A \equiv \mathrm{Tr}_B(\rho)$ and $\sigma_A \equiv \mathrm{Tr}_B(\sigma)$. Monotonicity of the trace distance under the partial trace,

$$\|\rho_A - \sigma_A\|_1 \leqslant \|\rho - \sigma\|_1, \tag{10}$$

implies in turn that, for the trace distance between two quantum states ρ and σ to be small, the difference in their entanglement entropies should necessarily be small. Hence, the entanglement diagnostics of initial circuit states can be considered as a weaker version of the proximity measure.

We usually cannot estimate the trace distance from the exact ground state ρ_g due to our ignorance of ρ_g. However, we expect that the ground states of gapped local Hamiltonians are far from typical quantum states σ, whose σ_A are approximately maximally mixed [24]. Thus, we require the trace distance between the equiprobable state and the reduced density matrix of the circuit state ρ_A to be large. This requires non-maximal entanglement entropies of the circuit states, i.e. they should not scale with the subsystem size n_A [25]. This is encoded in the following:

Theorem 1. (i) *The trace distance between the reduced density matrix ρ_A and the maximally mixed state $2^{-n_A} I_{n_A}$ satisfies the following inequality:*

$$\mathcal{L}(n_A, \rho_A) \leqslant \frac{1}{2}\left\|\rho_A - \frac{I_A}{2^{n_A}}\right\|_1 \leqslant \mathcal{U}(n_A, \rho_A) \tag{11}$$

with

$$\mathcal{U}(n_A, \rho_A) = \left(\frac{n_A - S_{EE}(\rho_A)}{2}\right)^{1/2} \tag{12}$$

$$\mathcal{L}(n_A, \rho_A) = \frac{1}{2k}\left(2^{(1-k)\mathcal{R}_A^k(\rho_A)} - 2^{(1-k)n_A}\right) \tag{13}$$

where $S_{EE}(\rho_A)$ and $\mathcal{R}_A^q(\rho_A)$ are the von Neumann and Rényi-k entropies of the reduced state ρ_A, respectively.

(ii) In the large size limit $n_A \gg 1$ of the subsystem A, the following lower bound holds asymptotically:

$$1 - \frac{S_{EE}(\rho_A)}{n_A} \lesssim \frac{1}{2}\left\|\rho_A - \frac{I_A}{2^{n_A}}\right\|_1. \tag{14}$$

Proof. (i) We start from the Pinsker's inequality:

$$\frac{1}{2}\|\rho - \sigma\|_1^2 \leqslant S(\rho\|\sigma) \tag{15}$$

J. Stat. Mech. (2022) 073101

on the trace distance between two states ρ and σ and their relative entropy. Plugging $\rho = \rho_A$ and $\sigma = 2^{-n_A} I_A$,

$$S(\rho_A \| \sigma_A) = n_A + \mathrm{Tr}(\rho_A \log \rho_A) = n_A - S_{EE}(\rho_A) \tag{16}$$

such that (15) becomes (12).

The continuity bound of the Tsallis-k entropy implies [21]:

$$|\,\mathrm{Tr}\left(\rho_A^k\right) - \mathrm{Tr}\left(\sigma_A^k\right)\,| \leqslant k\,\|\rho_A - \sigma_A\|_1, \tag{17}$$

which can turn into (13) by inserting $\sigma_A = 2^{-n_A} I_A$ and $\mathrm{Tr}\left(\rho_A^k\right) = 2^{(1-k)\mathcal{R}_A^k(\rho_A)}$.

(ii) Recall the Fannes–Audenaert inequality [22]:

$$|S_{EE}(\rho_A) - S_{EE}(\sigma_A)| \leqslant \frac{1}{2}\|\rho_A - \sigma_A\|_1 \log(2^{n_A} - 1) + H(\frac{1}{2}\|\rho_A - \sigma_A\|_1) \tag{18}$$

where $H(t) \equiv -t \log t - (1-t)\log(1-t)$. Substituting $\sigma_A = 2^{-n_A} I_A$ and taking the large system size limit $n_A \gg 1$, the lhs of (18) becomes $(n_A - S_{EE}(\rho_A))$, which leads to the asymptotic inequality (14). □

We stress that the entanglement diagnostic for circuit states is only a necessary condition to keep the initial and target states close. Remarkably, as we will see in section 4, the gradient-based optimization indeed works efficiently for those variational circuits whose average entanglement entropy scales slower than the volume law. Concerning the circuit depth, this suggests to avoid intermediate-depth and high-depth circuits, respectively corresponding to B/C of figure 1, and favor the circuits with fewer layers that belong to A. We will estimate the critical depth L_{s} that divides A and B/C in the following section 3.

3. Random quantum circuit

In this section, we study the growth of entanglement entropy for the circuit states generated by a random circuit evolution of the initial product state $|0\rangle^{\otimes n}$. Figure 2 is the quantum circuit architecture used in this paper. It defines a $(1+1)$-dimensional discrete quantum system, where the n qubits along the vertical axis represent the space, and the L layers along the horizontal axis span the time. The qubits are arranged identically with period n, i.e. $i \simeq i + n$, imposing a periodic boundary condition along the spatial direction. At each time step, the wavefunction evolves by a chain of the two-qubit unitary gates, acting alternatingly on all neighboring odd–even/even–odd qubit pairs. The two-qubit gate is made of independent Pauli-y rotations acting on single qubits,

$$R(\varphi) = \exp(i\sigma_y\varphi) = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}, \tag{19}$$

followed by the controlled-Z operation

J. Stat. Mech. (2022) 073101

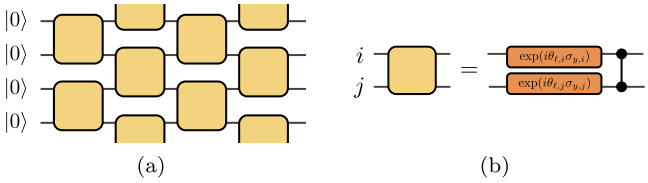


Figure 2. The circuit architecture used in sections 3 and 4. (a) The horizontal axis can be interpreted as the discrete time L. We call the commuting set of simultaneous two-qubit gates as the circuit layer. (b) Each gate consists of the single-qubit Pauli-y rotations (19) followed by the CZ operation (20).

$$CZ = \mathrm{diag}\,(1, 1, 1, -1) \tag{20}$$

that generically creates a pairwise entanglement. We will collectively denote all rotation angles by θ while using $\theta_{l,i}$ to indicate a specific angle that rotates the ith qubit at the lth layer, where $1 \leqslant i \leqslant n$ and $1 \leqslant \ell \leqslant L$. These variables are randomly chosen from $\mathcal{U}(0, 2\pi)$, the uniform distribution between 0 to 2π.

Since this circuit architecture consists of one-qubit rotation gates and adjacent two-qubit entanglers, it belongs to the family of the hardware-efficient ansatzes that can be efficiently implemented on NISQ devices with, e.g. superconducting qubits [26]. In fact, the hardware-efficient ansatz defines a broad category of variational circuits, whose detailed characteristics depend on the choice of specific unitary gates [16, 27, 28]. While we work with a particular form of the hardware-efficient ansatz, the relationship between the entanglement diagnostics and the VQA optimization should hold generically for other Hamiltonian-agnostic architectures. In contrast, the adaptive ansatzes [29, 30] have the advantage of being non-generic with a less entangled quantum state as the initial point of optimization. For these, the entanglement measures would not serve as an indicator of successful VQA optimization.

3.1. Linearity of the initial entanglement growth

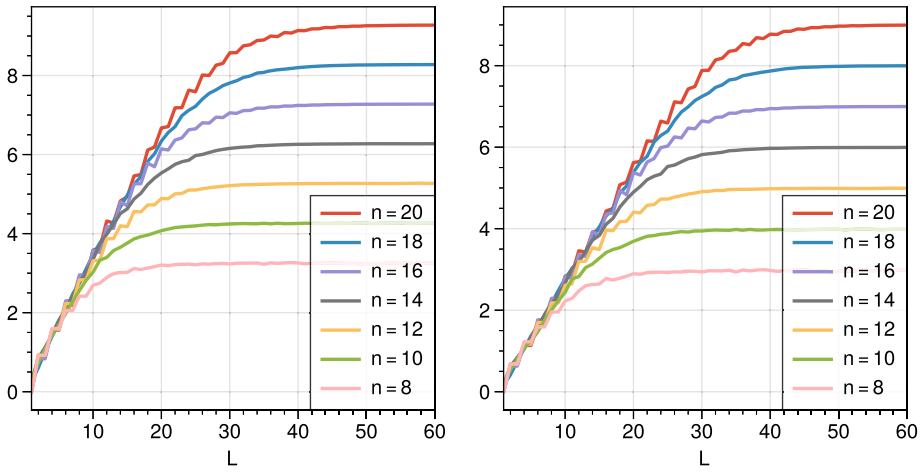
Let us consider the evolution of the n-qubit state under the random circuit unitaries of figure 2, as a function of the number of layers L. We measure the average growth of the bipartite entanglement of random circuit states by decomposing the n qubits into two equal-size subsystems, $n_A = n_B = n/2$, and calculating the sample statistics of various Rényi entropies for different n and L.

Figures 3(a) and (b) show the von Neumann and Rényi-2 entropies averaged over 50 random circuit states with different numbers of qubits n. Figure 4 compares the Rényi entropies of different orders averaged over 50 random circuit states with $n = 12$ and 20 qubits. They all exhibit the linear growth of the entanglement entropies at initial times. The curves then slow down in growth and eventually reach the plateaus [31]. See figures 19–21 in the appendix for the growth curves of several other entanglement quantities with different system sizes n.

The early linear growth of the entanglement entropy,

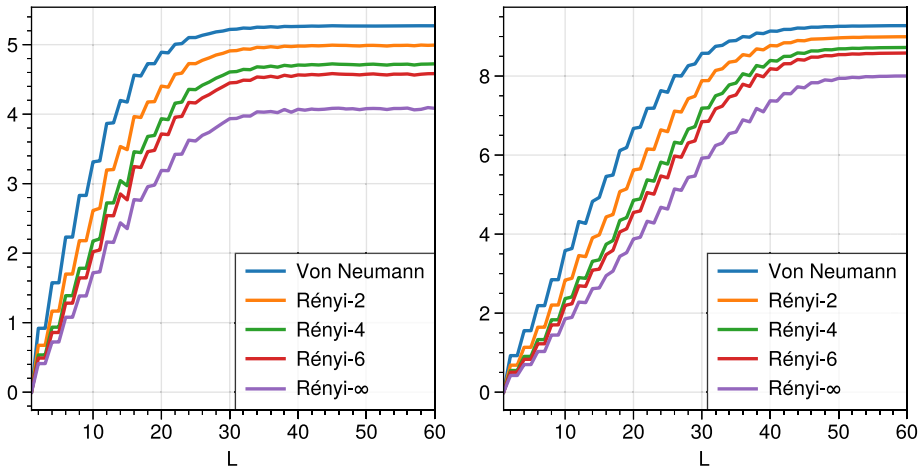
$$\mathcal{R}_A^k(\rho_A) = v_k L, \tag{21}$$

J. Stat. Mech. (2022) 073101



(a) Von Neumann Entropy (b) Rényi-2 Entropy

Figure 3. Von Neumann and Rényi-2 entropies for $8 \leqslant n \leqslant 20$ averaged over 50 independent circuit states, as a function of the circuit depth L.



(a) $n = 12$ (b) $n = 20$

Figure 4. Various Rényi-k entropies for $n = 12$ and $n = 20$ averaged over 50 independent circuit states, as a function of the circuit depth L.

is a characteristic feature of the global quench dynamics [32], which in our case is driven by the successive application of the layer unitaries $U(\theta_L)$ to the n-qubit product state $|0\rangle^{\otimes n}$. The coefficient v_k is known as the entanglement velocity and generally depends on k. We determine v_k by the linear regression of the early-time entropies on the range of $0 \leqslant L \leqslant n/2$. The estimated values of v_k, computed at different n's and k's, are summarized in the third columns of tables 1 and 2. We find that v_k is independent of n except for minor fluctuations, identifying (21) with the area-law entanglement of

J. Stat. Mech. (2022) 073101

Table 1. Phenomenological estimation of the entanglement velocity, saturation value, early-time scale, and late-time scale.

Type	n	v_k	L_l	$r_{n,k}$	L_s	Type	n	v_k	L_l	$r_{n,k}$	L_s
S_{EE}	8	0.3669	7	3.2526	29	$\mathcal{R}_A^2$	8	0.2771	9	2.9722	29
	10	0.3480	6	4.2639	33		10	0.2581	12	3.9821	36
	12	0.3533	13	5.2703	43		12	0.2645	17	4.9896	43
	14	0.3551	9	6.2743	49		14	0.2711	15	5.9944	49
	16	0.3522	17	7.2766	58		16	0.2689	23	6.9973	60
	18	0.3459	12	8.2776	65		18	0.2682	24	7.9986	65
	20	0.3411	23	9.2781	71		20	0.2619	33	8.9993	71

Table 2. Estimation of the entanglement velocity, late-time saturation value, early-time scale, and the late-time scale.

Type	n	v_k	L_l	$r_{n,k}$	L_s	Type	n	v_k	L_l	$r_{n,k}$	L_s
S_max	8	0.8534	6	3.9985	9	S_min	8	0.1724	9	2.2341	23
	10	0.9342	6	4.9983	12		10	0.1611	14	3.1373	33
	12	0.8694	8	5.9974	18		12	0.1651	21	4.0823	39
	14	0.8922	8	6.9965	23		14	0.1739	23	5.0491	46
	16	0.8510	9	7.9956	32		16	0.1715	35	6.0303	53
	18	0.8305	10	8.9947	35		18	0.1726	39	7.0181	60
	20	0.8011	11	9.9931	47		20	0.1686	49	8.0118	66
$\mathcal{R}_A^4$	8	0.2232	9	2.7084	29	$\mathcal{R}_A^6$	8	0.2050	9	2.5798	27
	10	0.2077	12	3.7101	33		10	0.1911	14	3.5708	33
	12	0.2134	19	4.7177	43		12	0.1963	21	4.5764	41
	14	0.2225	19	5.7234	49		14	0.2057	22	5.5823	49
	16	0.2199	29	6.7272	60		16	0.2031	31	6.5866	60
	18	0.2212	32	7.7290	63		18	0.2046	35	7.5887	63
	20	0.2156	41	8.7299	71		20	0.1995	43	8.5899	71

the early-time circuit states. Furthermore, v_k decreases when the order k of the Rényi entropy increases, i.e. $v_{k_1} > v_{k_2}$ for $k_1 < k_2$.

On the other end, at a late time, the Rényi-k entropy saturates to a constant $r_{n,k}$ for any n and k. We compute the saturated value of $\mathcal{R}_A^k$ by averaging it over the time frame $200 \leqslant L \leqslant 250$ and record that in the fifth columns of tables 1 and 2. The resulting constants $r_{n,k}$ manifest the following simple dependency on $n_A = n/2$:

$$\mathcal{R}_A^k(\rho_A) = r_{n,k} \simeq n/2 - c_k = \mathrm{Vol}(A) - c_k \tag{22}$$

implying the volume-law entanglement of the late-time circuit states [33]. We also find that, as the entropy order k increases, the saturated value $r_{n,k}$ declines monotonically, so the shift constant $c_k > 0$ can be only larger.

J. Stat. Mech. (2022) 073101

Combined with the discussion in section 2, the average entanglement curves suggest to refrain from using a variational circuit in the region of the plateau, i.e. $L \geqslant L_{\mathrm{s}}$, in order to prepare an initial circuit state in proximity to the target ground state that follows the area-law entanglement. We now turn to examine the scaling behavior of the early-time and late-time scales, i.e. L_{l} and L_{s}.

3.2. Timescale for the entanglement growth

Let us study the early-time L_{l} and late-time scales L_{s}, respectively, as the depth scales where the linear growth (21) ends and where the saturation (22) begins. We measure L_{l} and L_{s} using the following operational definitions:

$$\begin{aligned} L_{\mathrm{l}} &= \max\left\{L : |\mathcal{R}_A^k(L) - v_k L| \leqslant 2\,\mathrm{RMS}_k(0, n/2)\right\} \\ L_{\mathrm{s}} &= \min\left\{L : |\mathcal{R}_A^k(L) - r_{n,k}| \leqslant 2\,\mathrm{RMS}_k(200, 250)\right\}, \end{aligned} \tag{23}$$

L_{l} is the maximum depth L where the gap $|\mathcal{R}_A^k(L) - v_k L|$ between the Rényi entropy and its linear approximation maintains smaller than two times the RMS deviation,

$$\mathrm{RMS}_k(a, b) = \sqrt{\frac{1}{b-a+1}\sum_{\ell=a}^{b} (\mathcal{R}_A^k(\ell) - v_k \ell)^2}, \tag{24}$$

for $0 \leqslant L \leqslant n/2$. Similarly, L_{s} is the minimum depth L whose difference $|\mathcal{R}_A^k(L) - r_{n,k}|$ between the Rényi entropy and its saturated value remains to be smaller than two times the RMS deviation (24) for $200 \leqslant L \leqslant 250$.

The estimated values of L_{l} and L_{s} for different values of n and k are summarized in the fourth and sixth columns of tables 1 and 2. We make three observations: first, both timescales L_{l} and L_{s} increase as the entropy order k goes higher. Second, the saturation time L_{s} scales linearly in the system size n, i.e. $L_{\mathrm{s}} \sim \mathcal{O}(n)$, because

$$L_{\mathrm{s}} \gtrsim \frac{\mathrm{Vol}(A) - c_k}{v_k \cdot \mathrm{Area}(\partial A)} \sim \frac{n_A}{v_k} \sim \mathcal{O}(n). \tag{25}$$

This is consistent with [33] that a unitary design that maximizes all Rényi entropies can be reached within a linear complexity in the system size n. Third, there exists a transient gap between L_{l} and L_{s}, at least for finite-sized systems, in which the entanglement growth is slower. Details of the entanglement curves in this crossover region are largely model-dependent. See [34] for an example.

4. Circuit optimization

Our focus in this section is on the classical component of the hybrid quantum/classical algorithm. The objective is to find circuit parameters θ^* that closely approximate the

J. Stat. Mech. (2022) 073101

ground state energy, $E(\theta^*) \simeq E_g$, by taking iterative steps proportional to the negative gradient of the energy function (5) at each point θ_τ, i.e.[4]

$$\theta_{\tau+1} = \theta_\tau - \eta \nabla E(\theta_\tau). \tag{26}$$

The learning rate η scales the step size of each update. A too-large η can cause overshooting near the minimum θ^*, while a too-small η can make the optimization trajectory stuck at local minima. We will use $\eta = 0.005$ for most experiments. When the parameters update is small, each step of the gradient descent can reduce the energy by

$$\Delta E(\theta_\tau) \equiv E(\theta_{\tau+1}) - E(\theta_\tau) = -\eta \|\nabla E(\theta_\tau)\|_2^2. \tag{27}$$

Due to the constant decrease of the energy (27), we expect to reach E_g eventually if there are no other obstacles. We will terminate the iteration after updating the circuit parameter $\tau = 10^4$ times in all numerical experiments.

4.1. Results

Let us discuss the eigensolver optimization results that aim to solve the ground state of many-body systems. We specifically consider the 1D transverse-field Ising models with nearest-neighbor and long-range interactions and the SYK models with $n = 12$ qubits. See appendix A for a brief review of their Hamiltonians and ground-state entanglement properties.

4.1.1. The transverse-field Ising models. We search the ground states of interacting 1D spin-chain systems. To break the degeneracy of ground states, we turn on the magnetic coupling g to all the spin variables, choosing it to be $g = 1$ or 2. As we are interested in finding a general correlation between the entanglement diagnostics and the success of optimization, not relying on specific characteristics of Hamiltonians, we study the optimization for the following three Ising models:

(a) The nearest-neighbor spin coupling (A1) with $g = 2$
(b) The nearest-neighbor spin coupling (A1) with $g = 1$
(c) The long-range spin coupling (A3) with $\alpha = g = 1$.

We repeatedly perform the circuit optimization 50 times, to remove fluctuation made by random parameter initialization, and record the circuit outputs in figures 5–7 as a function of the circuit depth L.

Each figure consists of three panels. The left ones represent the energy difference $E(\theta) - E_g$ between the circuit state (5) and the exact ground state (A2). The middle ones show the trace distance between the reduced circuit state $\rho_{c,A}$ and the reduced ground state $\rho_{g,A}$. The right ones display the Rényi-2 entropy of the reduced circuit state. All the orange/blue curves therein represent a corresponding quantity before/after the optimization.

Figure 5 is for the nearest-neighbor Ising model (A1) with $g = 2$. It reveals the relation between the average entanglement entropy of initial circuit states and the success

[4] Estimating the gradient requires the readout of the circuit state ρ_c at shifted gate parameters [35] conducted by repeated measurements of Pauli strings. We will not consider the effect of the readout noise in this paper.

J. Stat. Mech. (2022) 073101

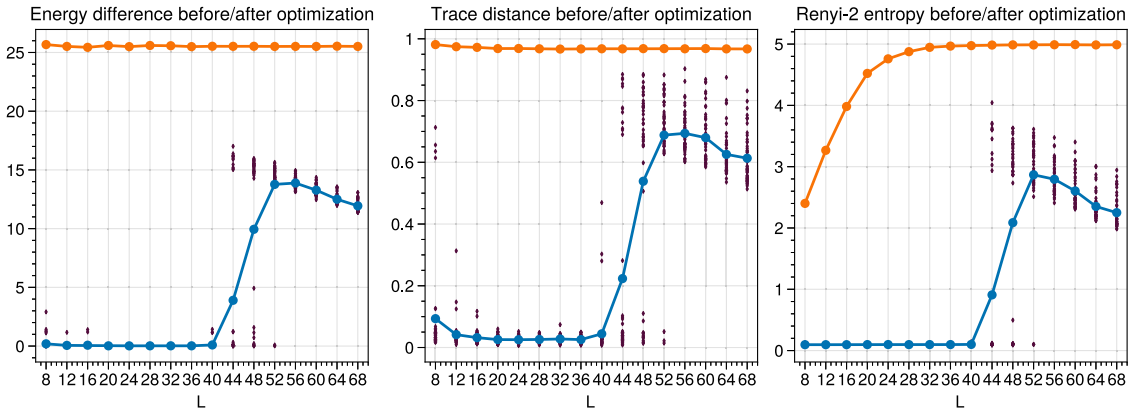


Figure 5. Measurements averaged over 50 independent circuits, before/after the VQA optimization with the nearest-neighbor Ising Hamiltonian (A1) at $n = 12$ and $g = 2$, as a function of the number of circuit layers L. Orange/blue curves denote the sample averages before/after the optimization. Black dots represent individual sample values after the optimization. $\Delta E \sim 4, 8$ corresponds to the relative error rate 7.84%, 15.67% obtained by dividing with the Hamiltonian bandwidth 51.05. The circuit variables are optimized with $\eta = 0.005$ over 10^4 iterations.

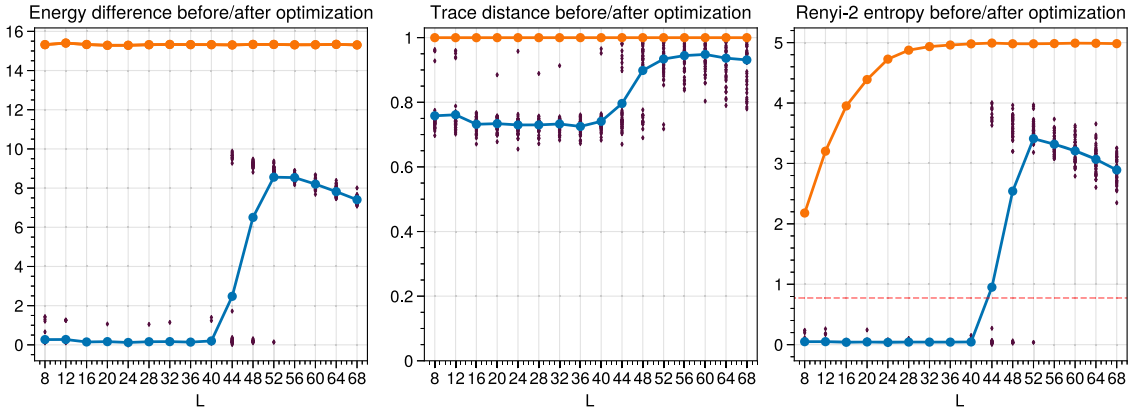


Figure 6. Measurements averaged over 50 independent circuits, before/after the VQA optimization with the nearest-neighbor Ising Hamiltonian (A1) at $n = 12$ and $g = 1$, as a function of the circuit depth L. Orange/blue curves denote the sample averages before/after the optimization. Black dots represent individual sample values after the optimization. The dashed line shows the Rényi-2 entropy of the exact ground state. $\Delta E \sim 4, 8$ corresponds to the relative error rate 13.05%, 26.11% obtained by dividing with the Hamiltonian bandwidth 30.65. The circuit variables are optimized with $\eta = 0.005$ over 10^4 iterations.

of gradient-based optimization: the optimization works well for the circuits with $L \lesssim 36$. However, in the intermediate range of $36 \lesssim L \lesssim 52$, the success rate gradually lowers as the circuit becomes deeper. Beyond that, i.e. $L \gtrsim 52$, it always fails to close the gap between the exact ground state and the circuit state as to their energy and entanglement

J. Stat. Mech. (2022) 073101

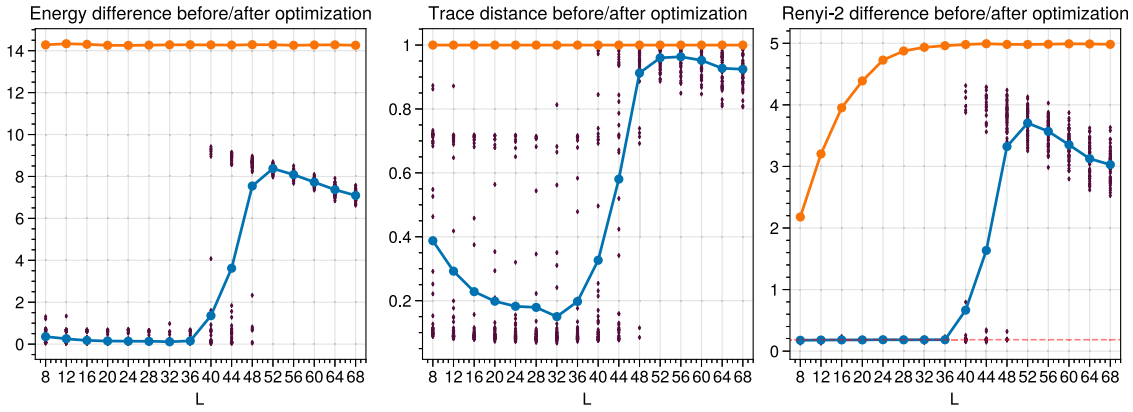


Figure 7. Measurements averaged over 50 independent circuits, before/after the VQA optimization with the non-local Ising Hamiltonian (A3) at $n = 12$ and $\alpha = g = 1$, as a function of the number of the layers L. Orange/blue curves denote the sample averages before/after the optimization. Black dots represent individual sample values after the optimization. The dashed line shows the Rényi-2 entropy of the exact ground state. $\Delta E \sim 4, 8$ corresponds to the relative error rate 9.10%, 18.20% obtained by dividing with the Hamiltonian bandwidth 43.95. The circuit variables are optimized with $\eta = 0.005$ over 10^4 iterations.

entropies. Such relation shows an advantage of using the circuits with $L < L_s$, whose entanglement curve has not reached the plateau. The above pattern also persists in figures 6 and 7, which correspond to the nearest-neighbor and long-range Ising models with $g = 1$.

A notable difference of figure 6 from figures 5 or 7 appears in the trace distance curve, where the optimization fails to narrow the distance even when the circuit energy is close to the exact ground state energy. It is related to the fact that the ground state entanglement entropy in the $g = 1$ nearest-neighbor Ising model is higher than those in other Ising models, as shown in figure 18. When the entanglement entropy of target ground states is higher, the local search of approximating circuit parameters becomes increasingly difficult. Such difficulty leads to deviations between the post-optimization circuit state and the exact ground state, to which the trace distance reflects much more sensitively than the energy and entanglement entropy differences.

4.1.2. The Sachdev–Ye–Kitaev model. We will now discuss the circuit optimization in a situation where the Hamiltonian ground state exhibits a volume law scaling of the entanglement entropy.

The SYK_4 Hamiltonian (A4) defined with an instance of random coupling constants has a ground state that follows the volume law scaling of entanglement [14], as reviewed in appendix A and specifically in figure 18. We optimize the variational circuit to approximate the SYK_4 ground state and summarize the output in figure 8 as a function of the circuit depth L.

Since the approximation target state itself behaves in terms of entanglement like a generic quantum state, the optimization task is now much more challenging. Unlike

J. Stat. Mech. (2022) 073101

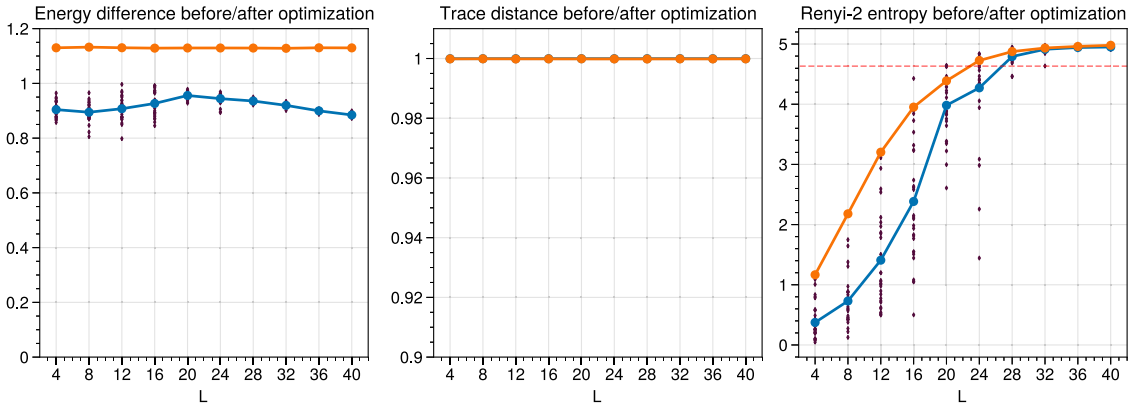


Figure 8. Measurements averaged over 30 independent circuits, before/after the VQA optimization with a particular instance of the SYK_4 Hamiltonian (A4) at $n = 12$, as a function of the circuit depth L. Orange/blue curves denote the sample averages before/after the optimization. Black dots represent individual sample values after the optimization. The dashed line shows the Rényi-2 entropy of the exact ground state. $\Delta E \sim 1$ corresponds to the relative error rate 44.00% obtained by dividing with the Hamiltonian bandwidth 2.27. In the middle panel, the black dots and blue line are entirely overlapped by the orange curve. The circuit variables are optimized with $\eta = 0.005$ over 10^4 iterations.

the optimization towards the Ising ground state, even choosing a less entangled circuit within the range $L \lesssim 36$ does not lead to success. Figure 8 manifests this failure, not only in the trace distance between the circuit and exact ground states but also their differences of energy and entanglement entropy.

4.1.3. Optimization speed. As another indicator of how difficult the circuit optimization is, we draw in figure 9 the evolution of Rényi-2 entanglement entropy $\mathcal{R}_A^2(\rho_{c,A})$ as a function of the number of parameter updates τ. The three curves therein are for the circuits with $L = 12$, 40, 68, which represent the characteristics of low-, intermediate-, high-depth circuits. The entanglement entropy of the ground states is marked by the dashed lines.

Towards the Ising ground states, the gradient descent works efficiently for the $L = 12$ circuit, rapidly reducing the entanglement entropy within 10^3 steps of the update. However, the same gradient descent takes a much longer time for the $L = 40$ circuit and even fails to reach closely the target state for the $L = 68$ circuit. As the average entanglement entropy of initial circuit states increases, the optimization difficulty becomes more evident not only in the trace distance, as in figure 6, but also through the entanglement diagnostics.

Curiously, the optimization search with 12 and 40 layers for the Ising ground state in figure 9 results in a state whose Renyi-2 entropy is lower than the entropy of the exact ground state. This is correlated with the large trace distance in figure 6. It suggests that in this case there is a large set of local minima having various degrees of overlap with the ground state wave function that can approximate well the ground state energy.

The optimization task towards the SYK_4 ground state is inherently more challenging such that all three curves leave a large entanglement gap from the target state.

J. Stat. Mech. (2022) 073101

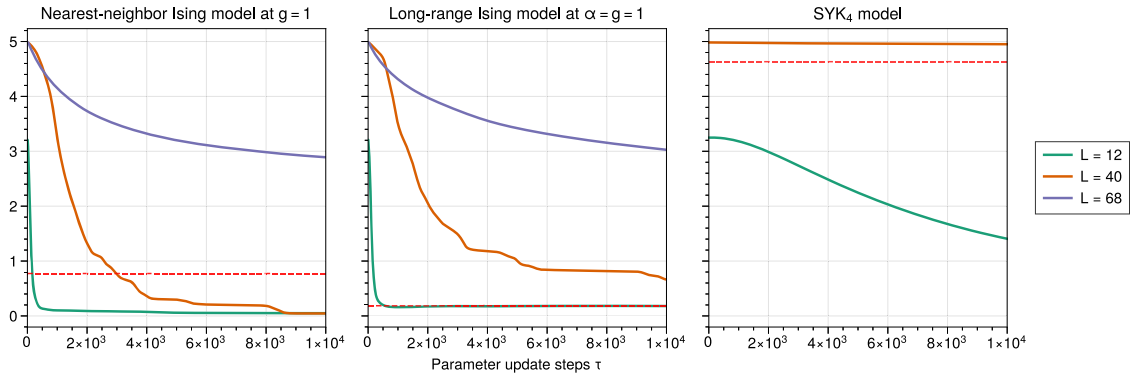


Figure 9. The evolution of the Rényi-2 entropy by the gradient-based VQA optimization with the Ising and SYK_4 Hamiltonians, i.e. (A1), (A3) and (A4), for $n = 12$ qubits. The horizontal axis means the number of parameter updates τ. The VQA optimization uses $\eta = 0.005$ for 10^4 iterations. The dashed line shows the Rényi-2 entropy of the exact ground state for each Hamiltonian.

Interestingly, the gradient descent constantly reduces the entanglement entropy of the $L = 12$ circuit state, enlarging the gap over the optimization steps τ. In general, over-parameterization can assist the circuit optimization that starts from/ends at a highly-entangled typical quantum state [15–19], similarly to how it simplifies the optimization of parametric machine learning models [36, 37]. It was found in [17] that an exponentially high-dimensional parameter space (i.e. $L \gg 1$) allows the high precision VQA approximation of the SYK_4 ground state.

4.2. Entanglement diagnostics and optimization

Our results shown in section 4.1 exemplify the difficulty in finding a successful optimization trajectory that starts from or ends at a typical quantum state that takes up the vast majority of the Hilbert space. This has been best described through the evolution of the entanglement entropies, (3) and (4), over the optimization steps, rather than a more commonly-used sensitive metric such as the trace distance between the circuit and target states.

Suppose we can divide the Hilbert space into two subregions distinguished by their entanglement entropies, say A and B/C, in accordance with figure 1. Generic random states belong to the region B/C whose entanglement entropies are approximately maximal.

For many interesting applications, the target state is a non-generic state that resides in the region A, i.e. following the area-law scaling of the entanglement [4]. Along an optimization path inside the region A, the circuit state entanglement tends to decrease regularly. However, when an initial state $\rho_c(\theta_{in})$ belongs to the region B/C, the local parameter update (26) is unable to cross over to the region A, thus failing to reach the ground state energy. We make these observations from the optimization result in section 4.1.1 that discusses the Ising Hamiltonians.

Even when the target state is maximally entangled and lies in the region B/C, the entanglement entropy of the circuit state $\rho_c(\theta_\tau)$ still tends to decrease on average.

J. Stat. Mech. (2022) 073101

It means that the entanglement gap between the circuit and target state can become larger, if an initial circuit state has a smaller entanglement entropy than the target state, i.e. $\mathcal{R}^k_A(\rho_{c,A}) \lesssim \mathcal{R}^k_A(\rho_{g,A})$. When $\mathcal{R}^k_A(\rho_{c,A}) > \mathcal{R}^k_A(\rho_{g,A})$, the optimization moves towards narrowing the gap, but often failing to match a desired level of the entanglement. These observations are based on the optimization results in section 4.1.2, obtained for the 1D SYK Hamiltonian.

The numerical results suggest that the Hilbert space can be partitioned into multiple layers, distinguished by the supported amount of the bipartite entanglement entropy. It is a very demanding task to move across distant layers via the gradient descent (26), which is doable only for over-parameterized circuits that involve exponentially large parameter space [15–19]. See [18, 19] for how over-parameterization changes the quantum energy landscape.

Note that, while we employ the vanilla gradient descent with specific values of η and τ for the controlled experiments, the effectiveness of the entanglement diagnostics extends generically to other gradient-based VQA optimizations—we have checked that adding the momentum, using the Adam optimizer, or adjusting η does not change the result of whether the ground state is reached or not. It does, however, affects the convergence speed. Even though it is difficult to completely exclude the possibility of finding the ground state with more iterations, regardless of what the entanglement measures suggest, we have seen that the constant scaling of τ is not sufficient to make a change. Another important factor that affects whether the VQA optimization can reach θ^* independently of the entanglement diagnostics is the number of circuit parameters. See section 5.2 and [18] for related studies.

5. Other circuit architectures

We discussed the importance of choosing the circuit to avoid the saturation of its average initial entanglement entropy, for a generic optimization task that finds a target state with the area law entanglement. This section examines if the entanglement diagnostic still serves as an indicator of efficient circuit optimization with different circuit architectures. We also consider the effect of reducing the number of circuit parameters while maintaining a similar degree of entanglement.

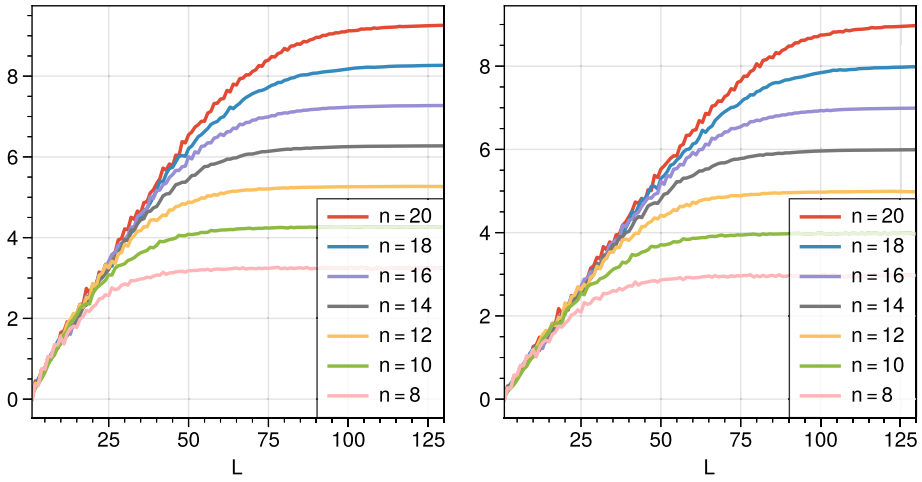
5.1. Random graph architecture

Let us study a simple stochastic variation of the circuit architecture that omits the CZ entangler (20) inside the two-qubit gate of figure 2(b) with a fixed probability $p = 1/2$.

5.1.1. Entanglement growth. Since the average number of the entangler is cut in half, we expect that the entanglement growth rate would be halved. Accordingly, the circuit depth L to reach the saturation of the entanglement entropy would be doubled.

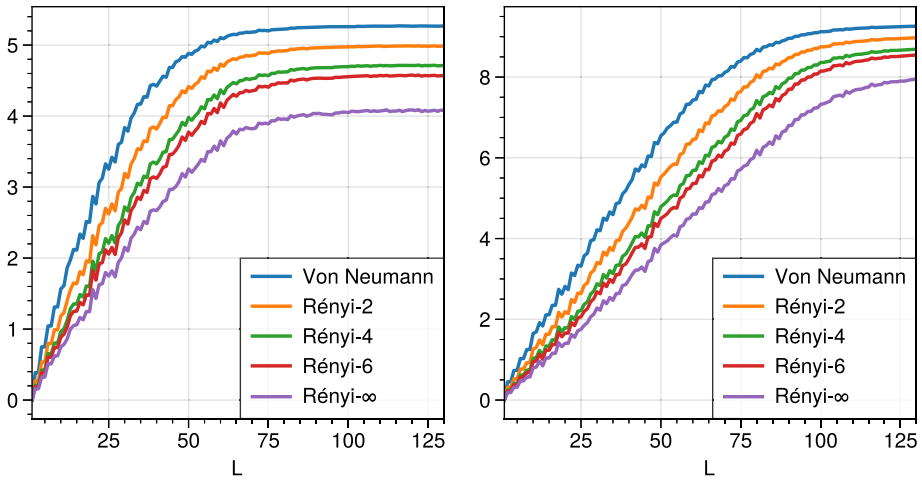
Figures 10 and 11 show the evolution of entanglement diagnostics as a function of the circuit depth L, estimated by the sample averages of 50 random states with $n = 12$ and 20. The overall shape of the curves remains the same, but the growth rate has significantly decreased. Reaching a certain level of the entanglement diagnostics

J. Stat. Mech. (2022) 073101



(a) Von Neumann Entropy (b) Rényi-2 Entropy

Figure 10. Von Neumann and Rényi-2 entropies for $8 \leqslant n \leqslant 20$ averaged over 50 independent $p = 1/2$ circuit states, as a function of the circuit depth L.



(a) $n = 12$ (b) $n = 20$

Figure 11. Various Rényi-k entropies for $n = 12$ and $n = 20$ averaged over 50 independent $p = 1/2$ circuit states, as a function of the circuit depth L.

requires twice the circuit depth compared to the non-stochastic architecture, i.e. $p = 1$, as expected. See figure 21(b) for the curve of the geometric measure at $8 \leqslant n \leqslant 20$ whose growth rate has been halved.

5.1.2. Optimization. Given the optimization task that reaches the nearest-neighbor Ising ground energy (A2) with the background field coupling $g = 1$, the outputs of the $p = 1/2$ stochastic circuit are all collected in figure 12 as a function of L.

J. Stat. Mech. (2022) 073101

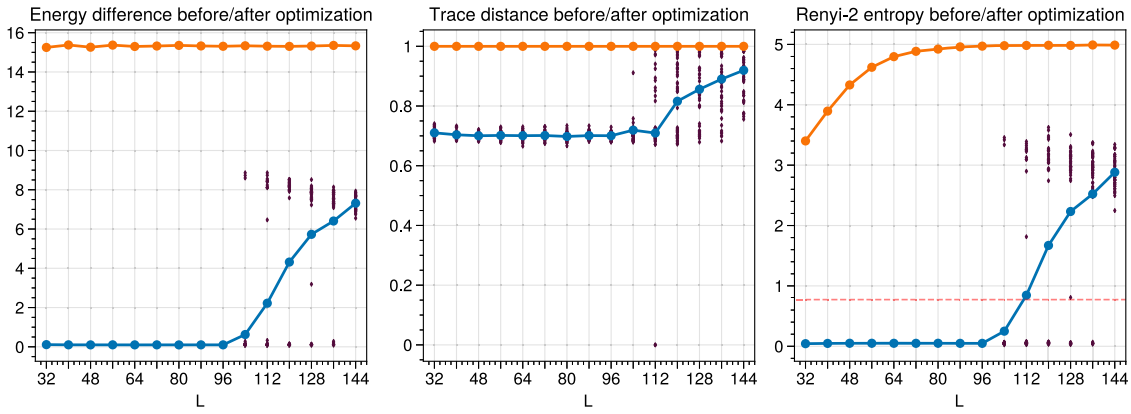


Figure 12. Measurements averaged over 50 independent $p = 1/2$ circuits, before/after the VQA optimization with the nearest-neighbor Ising Hamiltonian (A1) at $n = 12$ and $g = 1$, as a function of the number of circuit layers L. Orange/blue curves denote the sample averages before/after the optimization. Black dots represent individual sample values after the optimization. The dashed line shows the Rényi-2 entropy of the exact ground state.

The depth range of the $p = 1/2$ circuits where the gradient descent remains successful has increased to $L \lesssim 96$. Beyond that, the optimization success rate continues to drop until it reaches 0% at $L \sim 144$ and above. This is consistent with the entanglement growth curves, which maintain the same overall shape as in section 3 but only with a lower growth rate. We remark that the low and intermediate ranges, in which the optimization may succeed with a non-zero probability, has been extended to $L \lesssim 136$, more than mere doubling. It is the impact of the expanded parameter space whose dimension has been doubled, as required for the $p = 1/2$ circuit to hold the same level of entanglement.

Over the entire range of L, unlike the trace distance, the entropy diagnostic holds a robust correlation with the successful minimization of the circuit energy (5), showing its usefulness regardless of circuit-specific details.

5.2. Restricted circuit parametrization

Recall that an additional circuit layer increases both the average entanglement entropy of initial circuit states and the number of classical parameters. To isolate the effect of the classical parameter space, we study the consequence of imposing the following restriction:

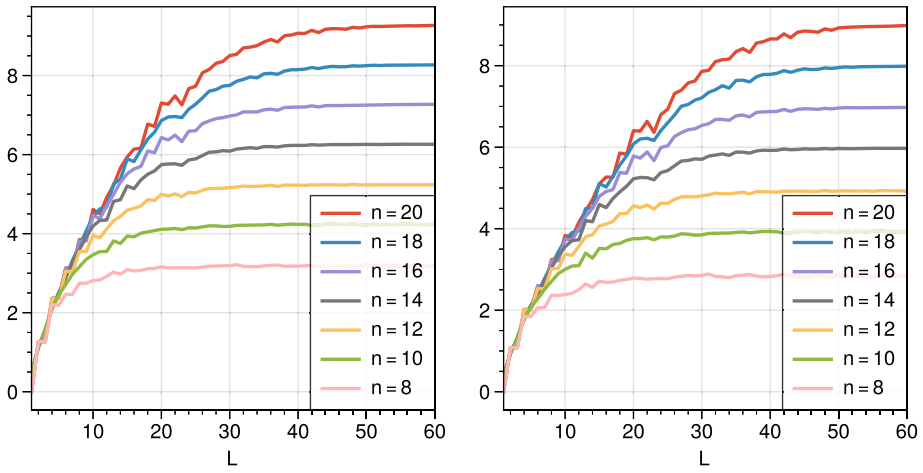
$$\theta_{\ell,1} = \theta_{\ell,2} = \cdots = \theta_{\ell,n} \quad \text{for all } 1 \leqslant \ell \leqslant L, \tag{28}$$

which equates all the parameters in each layer, yet maintains the same growth rate of entanglement diagnostics.

The basic two-qubit gate $\mathcal{O}_{i,j}$ in figure 2(b) reads:

$$\mathcal{O}_{i,j} = CZ_{i,j} \cdot R(\theta_{l,i}) \otimes R(\theta_{l,j}), \tag{29}$$

J. Stat. Mech. (2022) 073101



(a) Von Neumann Entropy (b) Rényi-2 Entropy

Figure 13. Von Neumann and Rényi-2 entropies for $8 \leqslant n \leqslant 20$ averaged over 50 independent restricted circuit states (28), as a function of the circuit depth L.

where $CZ = \mathrm{diag}(1, 1, 1 - 1)$ and $R(\theta)$ is the Pauli rotation (19) around the y-axis. It is curious to note that the constraint (28) is equivalent to imposing $[\mathcal{O}_{i,j}, Q_{i,j}] = 0$ on the Hilbert space of (i, j) qubits, where:

$$Q_{i,j} = \begin{pmatrix} q_1 + q_2 & 0 & 0 & 0 \\ 0 & q_1 & q_2 & 0 \\ 0 & q_2 & q_1 & 0 \\ 0 & 0 & 0 & q_1 + q_2 \end{pmatrix} \tag{30}$$

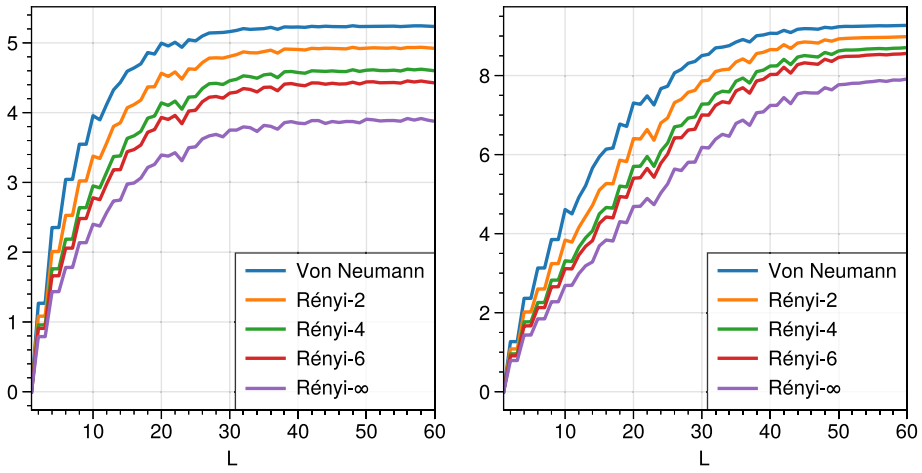
in the computational basis of (i, j) qubits. Still, there is no globally conserved charge written as a tensor product sum of $Q_{i,i+1}$, because it does not generically commute with $\mathcal{O}_{i-1,i}$ and $\mathcal{O}_{i,i+1}$ on the next layer.

5.2.1. Entanglement growth. The entanglement entropies averaged over 50 random circuit states under the parameter space restriction (28) are illustrated in figures 13 and 14 as a function of the number of circuit layers L. Except small extra wiggles, the overall growth shape and speed of the entanglement diagnostics are similar to those of the unconstrained circuit. Such correspondence of the entanglement growth curves renders the restricted circuit an appropriate setup to study separately the effect of the parameter space dimension on the circuit optimization.

Figure 21 shows that the geometric measure discussed in appendix B2 has the same growth pattern as the entanglement entropy curve. However, the evolution curve for the restricted circuit illustrated in figure 21(c) is fluctuating much more in comparison to the unconstrained circuit, while their saturation depth scales remain largely the same.

5.2.2. Optimization. We optimize the restricted circuit to approximate the ground state of the nearest-neighbor Ising model at $g = 1$ using the gradient descent. The results are summarized in figure 15 as a function of L. It is notable that even the circuit

J. Stat. Mech. (2022) 073101



(a) $n = 12$ (b) $n = 20$

Figure 14. Various Rényi-k entropies for $n = 12$ and $n = 20$ averaged over 50 independent restricted circuit states (28), as a function of the circuit depth L.

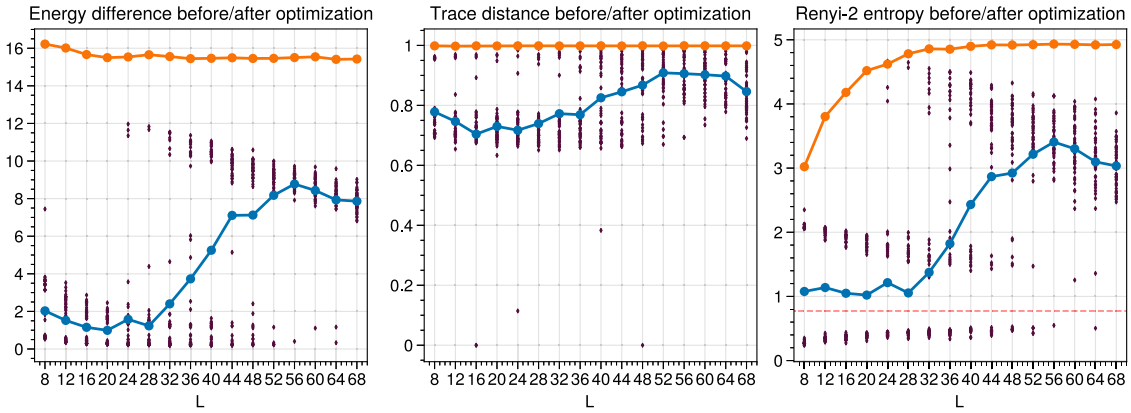


Figure 15. Measurements averaged over 50 independent restricted circuits (28), before/after the VQA optimization with the nearest-neighbor Ising Hamiltonian (A1) at $n = 12$ and $g = 1$, as a function of the circuit depth L. Orange/blue curves denote the sample averages before/after the optimization. Black dots represent individual sample values after the optimization. The dashed line shows the Rényi-2 entropy of the exact ground state.

optimization with $L \lesssim 20$ often stops at $\Delta E \gtrsim 1$, not giving a reliable approximation of the ground energy. Furthermore, starting from $L \gtrsim 24$, an increasingly large proportion of the randomly initialized circuits fails to reach the ground level energy E_g and leave a large gap, i.e. $\Delta E \gtrsim 8$. Such transitional result emerges at a much lower depth than $L = 44$ of the unconstrained circuit.

This trend is not an artifact of the particular choice of optimization hyperparameters, but rather illustrates the importance of having enough parameters in applying the gradient descent to optimization tasks, even for those circuits that remain within a suitable range of the entanglement diagnostics. Interestingly, the upper bound of the transition window between A and C is also expanded for this restricted circuit, such that certain VQA instances at $52 \leqslant L \leqslant 64$ can reach the ground-level energy quite closely. Given that the minimum reachable energy is practically more important than the average energy over different random initializations, it would be worthwhile to explore further the relationship between parameter hyperplanes (e.g. dimensionality, symmetry, and other algebraic properties) and VQA optimization.

6. Discussion and outlook

In this paper we considered the variational circuit model of quantum computation, arguing for the effectiveness of entanglement diagnostics in finding the circuit architecture for efficient parameter optimization that minimizes the Hamiltonian expectation value. Introduced as a distance measure between the circuit and target states, the entanglement diagnostic has shown its usefulness by illustrating that quantum circuit states within a suitable range of entanglement entropies can successfully reach the ground level energy of local Hamiltonians during a finite number of gradient descent steps (specifically, 10^4). It also says that, while entanglement is a valuable non-local resource for quantum computation, circuit states being highly entangled do not necessarily have an advantage but it can be rather the opposite[5].

One way to control the average entanglement entropy of circuit states is to adjust the number of circuit layers. The mean entropy grows linearly with the circuit depth, then gradually slows down, and finally converges to a constant near the theoretical maximum. Denoting by L_{s} the saturation depth beyond which the average entanglement entropy has converged, we divided the depth range into two intervals, $L < L_{\mathrm{s}}$ and $L \geqslant L_{\mathrm{s}}$, and called them respectively A and B/C. A is typically the optimal region that leads to efficient VQA computation, e.g. when we search an area-law entangled target state, while B/C can suffer from the barren plateau problem, one can further differentiate C from B based on whether the optimization success rate has become 0% or not yet.

Although the assumption of an area-law entangled target state covers a large number of interesting VQA applications [4], the ground states of some important Hamiltonians exhibit volume-law entanglement scaling. Moreover, the simulation of excited states [39] and of out of equilibrium dynamical systems [40] require the simulation volume-law entangled states. Matching the entanglement diagnostic alone is not sufficient to approximate such states due to the overwhelming population of highly-entangled quantum states. We need deep variational circuits whose depth L lies in B/C and that are equipped with a large parameter space that can assist high-resolution specification and approximation of the desired target state [17]. Having more circuit parameters can generally help to approximate the ground state better, as exemplified by the decreased

[5] See also the discussion in [38].

J. Stat. Mech. (2022) 073101

accuracy for a reduced number of independent variables[6] in section 5.2, as well as the increased success rate for circuits with extra single-qubit rotation parameters [17, 18].

There are many follow-up directions for further investigation: first, for having additional substantial evidence to the validity of the entanglement diagnostics, it would be crucial to consider 2D gapped local Hamiltonians whose ground states follow the area law entanglement scaling but are difficult to approximate. Second, we would like to explore various circuit architectures, e.g. using other rotation and entangling gates [28], built on different graph structures [42, 43], or conserving diffusive charges [44, 45]. Especially, symmetry-preserving circuits can work efficiently for the VQA optimization if the target state is known to respect the imposed built-in symmetries [46, 47]. Third, the layered circuit defines a discrete dynamical system. We would like to investigate the appearance of quantum chaos in the circuit wavefunction, such as the emergence of random matrix ensemble for the reduced density matrix [48] and the operator spreading [49], relating them to the efficient VQA optimization [50]. Fourth, it is crucial to investigate the implication of entanglement diagnostics with specialized optimizers for quantum circuits, such as quantum natural gradient descent or non-gradient-based methods. Finally, it is important to study different types of noise and analyze how they affect the VQA optimization performance [51, 52].[7]

Acknowledgments

We would like to thank Khen Cohen, Tom Levy, Eun Gook Moon, Muli Safra and Lior Wolf for valuable discussions. We are grateful to Jaedeok Kim and Dario Rosa for the collaboration at the early stage of this project. The work of J K is supported by the NSF Grant PHY-1911298 and the Sivian fund. The work of Y O is supported in part by Israel Science Foundation Center of Excellence, the IBM Einstein Fellowship and John and Maureen Hendricks Charitable Foundation at the Institute for Advanced Study in Princeton. Our Python code for the numerical experiments is written in TensorFlow Quantum [55]. The experimental data are managed by using Weights & Biases [56].

Appendix A. Hamiltonians

Let us consider the following 1D Hamiltonian systems: the nearest-neighbor and long-range Ising models coupled to a transverse magnetic field and the SYK model [11–13]. Here we summarize some of their important characteristics, including the entanglement entropy scaling of their ground states.

[6] It is not a conflict with [41] which reduces parameter space redundancy by identifying the principal directions carefully for maximal expressibility. Our reduction is rather arbitrary and without a guarantee that the circuit with remaining independent parameters can be maximally expressible.

[7] Although a careful analysis of the noise effect deserves a separate comprehensive study, as in [53], one can look into a simplifying case to obtain a rough intuition: when the decoherence can effectively be summarized into applying the following error channel once, $\rho_c \to p\rho_c + (1-p)\mathbf{1}$, the energy function also scales down, $E(\theta) \to pE(\theta)$, due to $\mathrm{Tr}(H) \simeq 0$. It creates an extra numerical difficulty for the parameter optimization, calling for the invention of effective noise mitigation schemes [54].

J. Stat. Mech. (2022) 073101

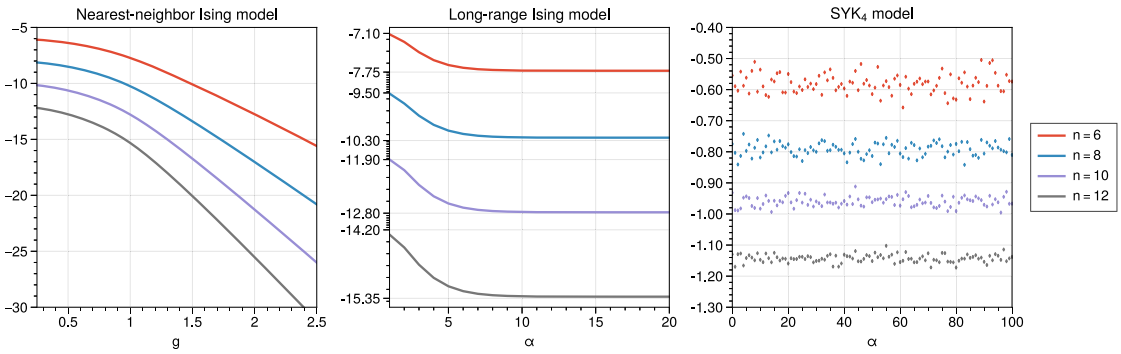


Figure 16. The ground energy of the following Hamiltonian systems of different sizes n: (left) nearest-neighbor Ising Hamiltonian with different g. (Middle) Long-range Ising Hamiltonian at $g = 1$ with different α. (Right) SYK_4 Hamiltonian with 100 different instances of Gaussian random couplings.

A1. Nearest-neighbor Ising model

The 1D Hamiltonian with the nearest-neighbor spin interaction coupled to a transverse magnetic field reads:

$$H = \sum_{i=1}^{n} \sigma_{z,i}\sigma_{z,i+1} + g\sum_{i=1}^{n} \sigma_{x,i}, \tag{A1}$$

where we assume the periodic boundary condition $i \sim i + n$ and the magnetic coupling g being positive. $\sigma_{x/z,i}$ are Pauli-x/z matrices acting on the ith spin, respectively.

One can solve this Hamiltonian exactly, leading to the following ground-level energy [57]:

$$E_{\mathrm{g}} = -\sum_{k=0}^{n-1} \left(1 + g^2 - 2g\,\cos\left(\frac{2\pi k}{n}\right)\right)^{1/2}. \tag{A2}$$

The $g = 0$ ground state is in the anti-ferromagnetic phase where all spin variables are aligned antiparallel to neighboring spins, such that $E_{\mathrm{g}} = -n$. As g grows, the spin-field coupling contributes more significantly to E_{g}. Especially when $g \sim g_{\mathrm{c}}$, the paramagnetic phase transition occurs such that all spins are now aligned in $-x$ direction. As $g \gg g_{\mathrm{c}}$, the ground energy approaches to $E_{\mathrm{g}} \simeq -gn$. See figure 16 for the ground energy curves over the range of the magnetic coupling $0 \leqslant g \leqslant 2.5$ with different n.

We also draw in figure 18 the curves of various Rényi-k entropies of the $n = 12$ Ising ground state for $0 \leqslant g \leqslant 2.5$. Figure 17 contains the scatter plot of energy and Rényi-2 entanglement entropy over all the eigenstates of the $g = 1$ Ising Hamiltonian with $n = 12$ qubits.

J. Stat. Mech. (2022) 073101

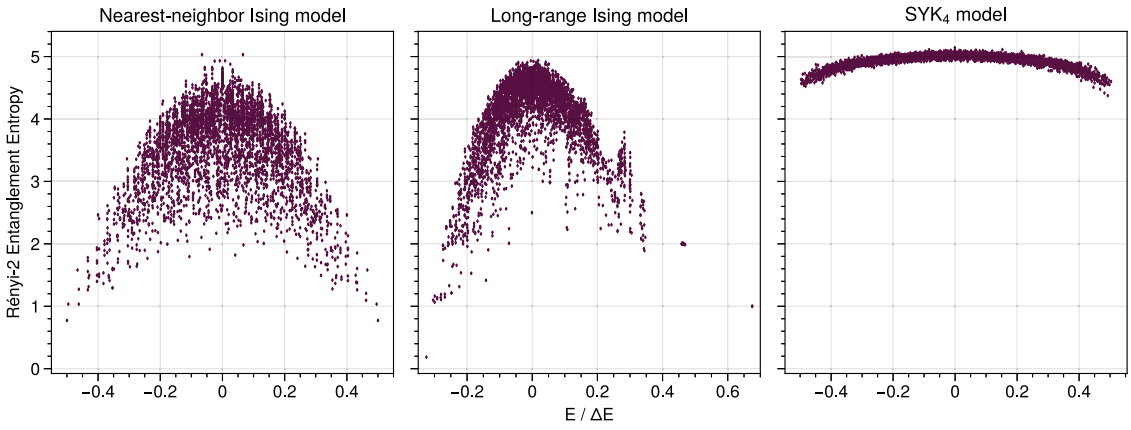


Figure 17. The scatter plot of energy and Rényi-2 entropy over all eigenstates of the following $n = 12$ Hamiltonians: (left) nearest-neighbor Ising Hamiltonian at $g = 1$. (Middle) Long-range Ising Hamiltonian at $g = \alpha = 1$. (Right) SYK_4 Hamiltonian with a particular instance of Majorana fermion couplings sampled from Gaussian distribution.

A2. Long-range Ising model

We add the long-range spin–spin interaction whose couplings decay with the distance. This Hamiltonian reads:

$$H = \sum_{i<j} \frac{1}{d(i,j)^{\alpha}} \sigma_{z,i}\sigma_{z,j} + g\sum_{i=1}^{n} \sigma_{x,i} \tag{A3}$$

where $d(i, j)$ is the shortest distance between the ith and jth spins with the periodic boundary condition $i \sim i + n$. All the non-local interactions vanish in the limit $\alpha \to \infty$, thus the Hamiltonian (A3) reduces to (A1).

As in the nearest-neighbor model, for any $\alpha \geqslant 0$, the long-range Ising model exhibits a transition between the anti-ferromagnetic and paramagnetic phases. Note that the ground state in the anti-ferromagnetic phase can have the entanglement entropy that grows with n, thus violating the area law. Its scaling behavior in n is logarithmic for $\alpha > 1$ and sub-logarithmic for $\alpha < 1$ [58]. Since the matrix product state ansatz can still closely approximate the ground state [58, 59], we expect the mild violation of the area law entanglement would not be a big obstacle of the gradient-based optimization even for large n. Several Rényi-k entropies of the $n = 12$ long-range Ising ground state are shown in figure 18 as a function of $0 \leqslant \alpha \leqslant 20$.

Figure 16 plots the $g = 1$ ground energy as a function of the exponent $0 \leqslant \alpha \leqslant 20$ for different system sizes n. Since the long-range interactions are almost negligible for $\alpha \gtrsim 10$, the curves converge to the energy (A2) of the nearest-neighbor Ising model at $g = 1$. We also draw the scatter plot of energy and Rényi-2 entanglement entropy in

J. Stat. Mech. (2022) 073101

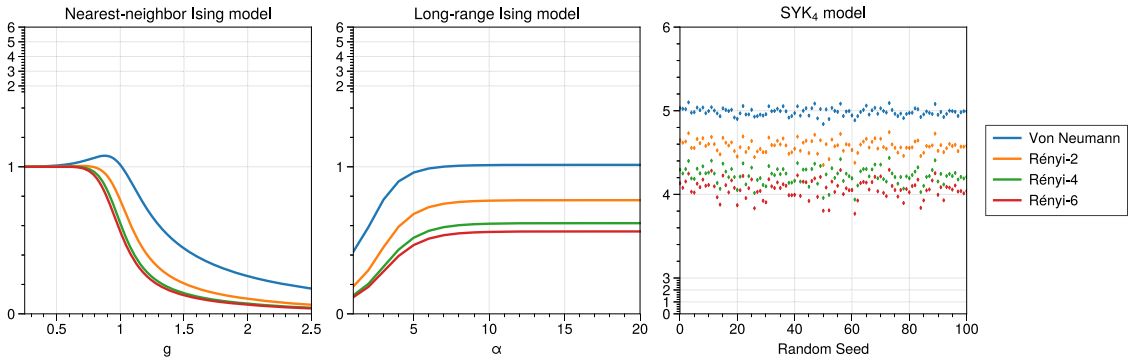


Figure 18. The ground state entanglement entropies of the following Hamiltonian systems of different sizes n: (left) nearest-neighbor Ising Hamiltonian with different g. (Middle) Long-range Ising Hamiltonian at $g = 1$ with different α. (Right) SYK_4 Hamiltonian with 100 different instances of Gaussian random couplings.

figure 17, denoting every eigenstate of the $g = \alpha = 1$ long-range Ising Hamiltonian with $n = 12$ qubits.

A3. The SYK model

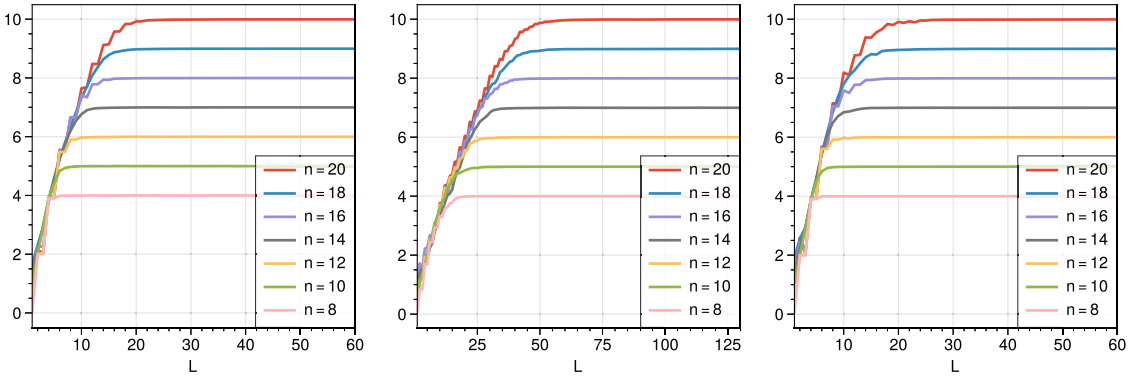
The SYK model [11, 12] consists of random, long-range, all-to-all interactions of n qubits, which correspond to the following random couplings of q Majorana fermions:

$$H = (i)^{q/2} \sum_{1 \leqslant i_1 < \cdots < i_q \leqslant 2n} J_{i_1 \ldots i_q} \gamma_{i_1} \cdots \gamma_{i_q}, \tag{A4}$$

where the Majorana fermions $\{\gamma_i\}_{1 \leqslant i \leqslant 2n}$ satisfy the Clifford algebra $\{\gamma_i, \gamma_j\} = \delta_{ij}$ and can be translated to the spin variables via the Jorgan–Wigner map. The coupling constants $J_{i_1 \ldots i_q}$ are randomly drawn from the Gaussian distribution with mean 0 and variance $(q-1)!/(2n)^{q-1}$. Much attention has been paid to the SYK_q model because it is exactly solvable and exhibits a chaotic dynamics for $q \geqslant 4$ at the same time [12, 13].

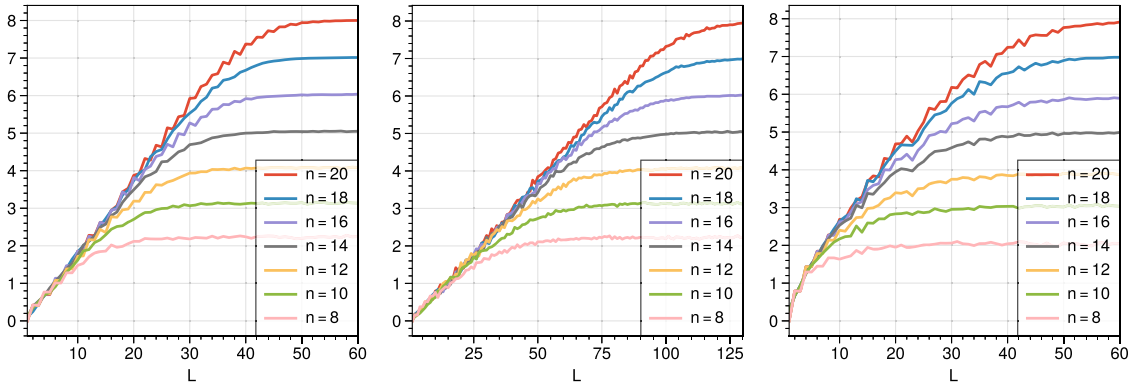
We focus on the SYK_4 model and drop the subscript for brevity. Each random draw of the coupling constants $J_{i_1 \ldots i_4}$ from the Gaussian distribution defines a different instance of the SYK Hamiltonian. The SYK ground energy for 100 individual instances with n qubits, or equivalently, $2n$ Majorana fermions are displayed in figure 16. Similarly, the entanglement entropies of 100 instances of the $n = 12$ SYK ground state are visualized in figure 18, illustrating the SYK ground state is much more highly-entangled than that of the Ising models. More generally, the energy-entropy scatter plot of figure 17 denotes the full spectrum for an instance of the $n = 12$ SYK Hamiltonian, exhibiting the volume-law scaling of the entanglement entropies [14]. Its energy gap is notably smaller than that of the Ising models, thus violating the assumption [10] for the area-law entanglement of the ground state.

J. Stat. Mech. (2022) 073101



(a) The circuit of Figure 1 (b) The $p = 1/2$ circuit (c) The restricted circuit

Figure 19. The max-entropy of the circuit reduced density matrix $\rho_A(\theta)$ as a function of the number of layers L.



(a) The circuit of Figure 1 (b) The $p = 1/2$ circuit (c) The restricted circuit

Figure 20. The min-entropy of the circuit reduced density matrix $\rho_A(\theta)$ as a function of the number of layers L.

Appendix B. Other entanglement measures

B1. Max–min entropies

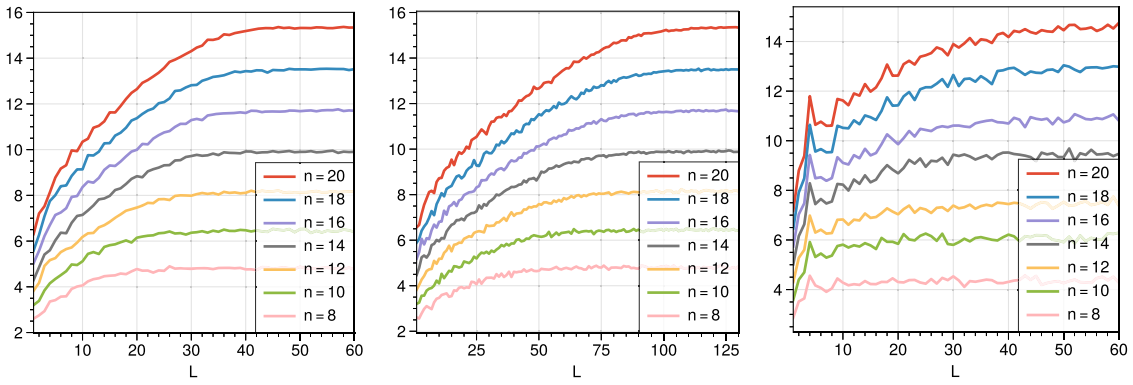
The min-entropy and max-entropy arise as two limiting cases of the Rényi-k entropies, i.e.

$$\begin{aligned} S_{\max}(\rho_A) &= \lim_{k\to 0} \mathcal{R}_A^k(\rho_A) = \log\left(\operatorname{rank}\rho_A\right) \\ S_{\min}(\rho_A) &= \lim_{k\to\infty} \mathcal{R}_A^k(\rho_A) = -\log\left(\lambda_{\max}(\rho_A)\right) \end{aligned} \tag{B1}$$

where $\lambda_{\max}(\rho)$ is the largest eigenvalue of ρ. Given the three architectures in figure 2 and section 5, their average max–min entanglement entropies of 50 random circuit states are given in figures 19 and 20 as a function of L.

The max-entropy shows rapid initial growth caused by the circuit architecture in figure 2(a), which increase the rank of the reduced density matrix ρ_A every second

J. Stat. Mech. (2022) 073101



(a) The circuit of Figure 1 (b) The $p = 1/2$ circuit (c) The restricted circuit

Figure 21. The geometric measure of the circuit reduced density matrix $\rho_A(\theta)$ as a function of the number of layers L.

layer until the rank saturates at the allowed maximum, 2^{n_A}. The non-negativeness and normalization of the reduced density matrix, i.e. $\lambda_i(\rho_A) \geqslant 0$ and $\sum_i \lambda_i(\rho_A) = 1$, then implies the change of the eigenvalue statistics from having one λ being non-zero and having a value of 1 to all λ's being non-zero and having similar values around 2^{-n_A}. Such decrease in the largest eigenvalue of ρ_A is displayed in the min-entropy curve. We note this spectral change of ρ_A can be a contributing factor for the emergence of the random matrix behavior of ρ_A, studied elsewhere [50]. Note that while the min-entropy can provide a diagnostic for effective VQA optimization, this is not the case with the max-entropy.

B2. Geometric measure

Another way to measure the quantum entanglement of circuit states is to study the geometric measure of entanglement, based on the overlap between the circuit state $|\psi(\theta)\rangle$ and its nearest product state [60]. It reads:

$$\mathcal{E}_{\mathrm{g}}(|\psi(\theta)) = -\log \sup_{\alpha \in \mathcal{P}} |\langle \alpha | \psi(\theta) |^2, \tag{B2}$$

where $\mathcal{P}$ is the set of qubit product states. Figure 21 shows that the geometric measure of entanglement grows in the same pattern as of the entanglement entropy curve. However, note that the geometric measure is directly calculated from the full density matrix ρ, while the entanglement entropies are found from the reduced state ρ_A.

References

[1] Preskill J 2018 Quantum computing in the NISQ era and beyond *Quantum* **2** 79
[2] Bharti K *et al* 2022 Noisy intermediate-scale quantum algorithms *Rev. Mod. Phys.* **94** 015004
[3] Farhi E, Goldstone J and Gutmann S 2014 A quantum approximate optimization algorithm (arXiv:1411.4028)
[4] Lucas A 2014 Ising formulations of many np problems *Front. Phys.* **2** 00005
[5] Peruzzo A, McClean J, Shadbolt P, Yung M-H, Zhou X-Q, Love P J, Aspuru-Guzik A and O'Brien J L 2014 A variational eigenvalue solver on a photonic quantum processor *Nat. Commun.* **5** 4213

J. Stat. Mech. (2022) 073101

[6] McClean J R, Boixo S, Smelyanskiy V N, Babbush R and Neven H 2018 Barren plateaus in quantum neural network training landscapes *Nat. Commun.* **9** 4812
[7] Cerezo M, Sone A, Volkoff T, Cincio L and Coles P J 2020 Cost-function-dependent barren plateaus in shallow quantum neural networks (arXiv:2001.00550)
[8] Marrero C O, Kieferová M and Wiebe N 2020 Entanglement induced barren plateaus (arXiv:2010.15968)
[9] Holmes Z, Sharma K, Cerezo M and Coles P J 2021 Connecting ansatz expressibility to gradient magnitudes and barren plateaus (arXiv:2101.02138)
[10] Hastings M B 2007 An area law for one-dimensional quantum systems *J. Stat. Mech.* P08024
[11] Sachdev S and Ye J 1993 Gapless spin-fluid ground state in a random quantum Heisenberg magnet *Phys. Rev. Lett.* **70** 3339–42
[12] Kitaev A 2015 A simple model of quantum holography (http://online.kitp.ucsb.edu/online/entangled15/kitaev/ and http://online.kitp.ucsb.edu/online/entangled15/kitaev2/)
[13] Maldacena J and Stanford D 2016 Remarks on the Sachdev–Ye–Kitaev model *Phys. Rev.* D **94** 106002
[14] Huang Y and Gu Y 2019 Eigenstate entanglement in the Sachdev–Ye–Kitaev model *Phys. Rev.* D **100** 041901
[15] Kiani B T, Lloyd S and Maity R 2020 Learning unitaries by gradient descent (arXiv:2001.11897)
[16] Wiersema R, Zhou C, de Sereville Y, Carrasquilla J F, Kim Y B and Yuen H 2020 Exploring entanglement and optimization within the Hamiltonian variational ansatz *PRX Quantum* **1** 020319
[17] Kim J, Kim J and Rosa D 2020 Universal effectiveness of high-depth circuits in variational eigenproblems (arXiv:2010.00157)
[18] Kim J and Oz Y 2021 Quantum energy landscape and VQA optimization (arXiv:2107.10166)
[19] Larocca M, Ju N, Garc'ia-Mart'in D, Coles P J and Cerezo M 2021 Theory of overparametrization in quantum neural networks (arXiv:2109.11676)
[20] Fannes M 1973 A continuity property of the entropy density for spin lattice systems *Commun. Math. Phys.* **31** 291–4
[21] Raggioa G 1995 Properties of q-entropies *J. Math. Phys.* **36** 4785–91
[22] Audenaert K M R 2007 A sharp continuity estimate for the von Neumann entropy *J. Phys. A: Math. Theor.* **40** 8127–36
[23] Chen Z, Ma Z, Nikoufar I and Fei S-M 2016 Sharp continuity bounds for entropy and conditional entropy *Sci. China: Phys., Mech. Astron.* **60** 020321
[24] Goldstein S, Lebowitz J L, Tumulka R and Zanghì N 2006 Canonical typicality *Phys. Rev. Lett.* **96** 050403
[25] Page D N 1993 Average entropy of a subsystem *Phys. Rev. Lett.* **71** 1291–4
[26] Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow J M and Gambetta J M 2017 Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets *Nature* **549** 242–6
[27] Sim S, Johnson P D and Aspuru-Guzik A 2019 Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms *Adv. Quantum Technol.* **2** 1900070
[28] Haug T, Bharti K and Kim M S 2021 Capacity and quantum geometry of parametrized quantum circuits (arXiv:2102.01659)
[29] Grimsley H R, Economou S E, Barnes E and Mayhall N J 2019 An adaptive variational algorithm for exact molecular simulations on a quantum computer *Nat. Commun.* **10** 3007
[30] Tang H L, Shkolnikov V O, Barron G S, Grimsley H R, Mayhall N J, Barnes E and Economou S E 2019 Qubit-adapt-vqe: an adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor (arXiv:1911.10205)
[31] Geller M R 2018 Sampling and scrambling on a chain of superconducting qubits *Phys. Rev. Appl.* **10** 024052
[32] Calabrese P and Cardy J 2005 Evolution of entanglement entropy in one-dimensional systems *J. Stat. Mech.* P04010
[33] Liu Z-W, Lloyd S, Zhu E and Zhu H 2018 Entanglement, quantum randomness, and complexity beyond scrambling *J. High Energy Phys.* JHEP07(2018)041
[34] Liu H and Suh S J 2014 Entanglement tsunami: universal scaling in holographic thermalization *Phys. Rev. Lett.* **112** 011601
[35] Mitarai K, Negoro M, Kitagawa M and Fujii K 2018 Quantum circuit learning *Phys. Rev.* A **98** 032309
[36] Jacot A, Gabriel F and Hongler C 2018 Neural tangent kernel: convergence and generalization in neural networks (invited paper) *Proc. 53rd Annual ACM SIGACT Symp. on Theory of Computing*
[37] Allen-Zhu Z, Li Y and Song Z 2019 A convergence theory for deep learning via over-parameterization (arXiv:1811.03962)
[38] Gross D, Flammia S T and Eisert J 2009 Most quantum states are too entangled to be useful as computational resources *Phys. Rev. Lett.* **102** 190501
[39] Higgott O, Wang D and Brierley S 2019 Variational quantum computation of excited states *Quantum* **3** 156

J. Stat. Mech. (2022) 073101

[40] Lau J W Z, Lim K H, Bharti K, Kwek L-C and Vinjanampathy S 2022 Convex optimization for non-equilibrium steady states on a hybrid quantum processor (arXiv:2204.03203)
[41] Funcke L, Hartung T, Jansen K, Kühn S and Stornati P 2021 Dimensional expressivity analysis of parametric quantum circuits *Quantum* **5** 422
[42] Harrow A W, Kong L, Liu Z-W, Mehraban S and Shor P W 2019 A separation of out-of-time-ordered correlation and entanglement (arXiv:1906.02219)
[43] Cervera-Lierta A, Latorre J I and Goyeneche D 2019 Quantum circuits for maximally entangled states *Phys. Rev.* A **100** 022342
[44] Rakovszky T, Pollmann F and von Keyserlingk C W 2019 Sub-ballistic growth of Rényi entropies due to diffusion *Phys. Rev. Lett.* **122** 250602
[45] Žnidarič M 2020 Entanglement growth in diffusive systems *Commun. Phys.* **3** 100
[46] Liu J-G, Zhang Y-H, Wan Y and Wang L 2019 Variational quantum eigensolver with fewer qubits *Phys. Rev. Res.* **1** 023025
[47] Gard B T, Zhu L, Barron G S, Mayhall N J, Economou S E and Barnes E 2020 Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm *npj Quantum Inf.* **6** 10
[48] Chen X and Ludwig A W W 2018 Universal spectral correlations in the chaotic wave function and the development of quantum chaos *Phys. Rev.* B **98** 064309
[49] Mezei M and Stanford D 2017 On entanglement spreading in chaotic systems *J. High Energy Phys.* JHEP05(2017)065
[50] Kim J, Oz Y and Rosa D 2022 Quantum chaos and circuit parameter optimization (arXiv:2201.01452)
[51] Barron G S and Wood C J 2020 Measurement error mitigation for variational quantum algorithms (arXiv:2010.08520)
[52] Fontana E, Cerezo M, Arrasmith A, Rungger I and Coles P J 2020 Optimizing parametrized quantum circuits via noise-induced breaking of symmetries (arXiv:2011.08763)
[53] Wang S, Fontana E, Cerezo M, Sharma K, Sone A, Cincio L and Coles P J 2020 Noise-induced barren plateaus in variational quantum algorithms (arXiv:2007.14384)
[54] Gentini L, Cuccoli A, Pirandola S, Verrucchi P and Banchi L 2020 Noise-resilient variational hybrid quantum-classical optimization *Phys. Rev.* A **102** 052414
[55] Broughton M *et al* 2020 TensorFlow quantum: a software framework for quantum machine learning (arXiv:2003.02989)
[56] Biewald L 2020 Experiment tracking with weights and biases (https://wandb.com/)
[57] Fagotti M and Essler F H L 2013 Reduced density matrix after a quantum quench *Phys. Rev.* B **87** 245107
[58] Koffel T, Lewenstein M and Tagliacozzo L 2012 Entanglement entropy for the long-range Ising chain in a transverse field *Phys. Rev. Lett.* **109** 267203
[59] Verstraete F and Cirac J I 2006 Matrix product states represent ground states faithfully *Phys. Rev.* B **73** 094423
[60] Wei T-C and Goldbart P M 2003 Geometric measure of entanglement and applications to bipartite and multipartite quantum states *Phys. Rev.* A **68** 042307

J. Stat. Mech. (2022) 073101