PAPER

Fast decay of classification error in variational quantum circuits

Quantum Science and Technology

PAPER

Fast decay of classification error in variational quantum circuits

RECEIVED
13 December 2021

REVISED
9 May 2022

ACCEPTED FOR PUBLICATION
18 May 2022

PUBLISHED
6 June 2022

Bingzhi Zhang[1,2] and Quntao Zhuang[2,3,4,*]

[1] Department of Physics, University of Arizona, Tucson, AZ 85721, United States of America
[2] Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721, United States of America
[3] James C. Wyant College of Optical Sciences, University of Arizona, Tucson, AZ 85721, United States of America
[4] Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: zhuangquntao@gmail.com

Keywords: quantum state discrimination, variational quantum circuit, quantum many-body systems

Abstract

Variational quantum circuits (VQCs) have shown great potential in near-term applications. However, the discriminative power of a VQC, in connection to its circuit architecture and depth, is not understood. To unleash the genuine discriminative power of a VQC, we propose a VQC system with the optimal classical post-processing—maximum-likelihood estimation on measuring all VQC output qubits. Via extensive numerical simulations, we find that the error of VQC quantum data classification typically decays exponentially with the circuit depth, when the VQC architecture is extensive—the number of gates does not shrink with the circuit depth. This fast error suppression ends at the saturation towards the ultimate Helstrom limit of quantum state discrimination. On the other hand, non-extensive VQCs such as quantum convolutional neural networks are sub-optimal and fail to achieve the Helstrom limit, demonstrating a trade-off between ansatz complexity and classification performance in general. To achieve the best performance for a given VQC, the optimal classical post-processing is crucial even for a binary classification problem. To simplify VQCs for near-term implementations, we find that utilizing the symmetry of the input properly can improve the performance, while oversimplification can lead to degradation.

1. Introduction

Quantum computation promises to solve classically intractable problems with a speedup in performance [1]. However, as scalable error-corrected quantum computers are not available, quantum information processing is limited to protocols using noisy intermediate-scale quantum (NISQ) [2] technology. The technological constraints also call for an alternative route towards a quantum advantage. Among the candidates, variational quantum circuits (VQCs) are a class of quantum–classical hybrid systems applicable to various tasks, including optimization [3], state preparation [4, 5], auto-encoding [6, 7], eigen-solvers [8–14], unsampling and state approximation [15, 16], state classification [17–20], state tomography [21], sensor networks [22, 23], solving partial-differential equations [24], quantum simulation [25–27] and machine learning in general [28–36].

Despite various applications, the fundamental understanding of the capability of VQCs in connection to circuit depth and circuit architecture is still missing. Recent progresses unveil the notion of depth efficiency on expressive and discriminative power in VQCs' classical counterpart—neural networks [37–40]; VQCs' discriminative power on quantum data has also attracted much attention recently, showing great potential in the classification of few-qubit states [18] and quantum phases of many-body systems [19, 20], even in the presence of noise [17]. For VQCs, recent works propose to quantify the expressivity via the effective dimension from an information geometry perspective and show some quantum advantage in expressivity of VQCs compared with their classical counterparts [41]. At the same time, the expressivity of VQC ansatzs is also connected to the barren plateau phenomena [42–44], where the variance of gradient decays

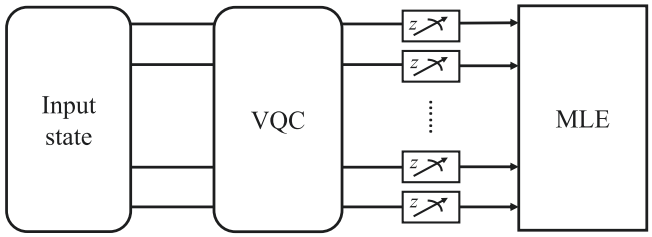

Figure 1. Schematic illustration of the MLE-VQC system. As any single qubit rotation can be combined into the VQC, we can fix each single-qubit measurement to be Pauli Z without loss of generality. An MLE strategy is utilized on the measurement results to make the decision.

exponentially with the number of qubits in the system and therefore highly limits the training performance of large-scale VQCs. However, regardless of these challenges in training, it is still important to understand a VQC's discriminative power in connection to its depth and architecture.

In this paper, we take a further step to unveil how VQCs' discriminative power quantitatively connects to circuit depth and architecture, and compares with the ultimate limit [45]. First, to unleash the genuine discriminative power of VQCs, we go beyond the popular approach—a single-qubit measurement on VQC output [48, 49]—to measure all qubits and then perform maximum-likelihood estimation (MLE) on the measurement results, even for a binary classification problem between two states. Our numerical results show that the MLE-VQC approach offers an order-of-magnitude smaller deviation from the Helstrom limit than the single-qubit approach. While both approaches can face the dilemma of barren plateau that limits the trainability of large circuits [42–44], numerical evaluation shows that our MLE-VQC approach has larger gradients relative to the single-qubit approach. With the full discriminative power of VQCs in hand, we proceed to explore its connection to circuit architecture and depth.

When discriminating between complex quantum states, we find that the discrimination error is exponentially suppressed with the continuous increase of the VQC depth, until a saturation to the minimum given by the Helstrom limit [46, 47]. When such a continuous increase of depth is forbidden by the non-extensive architectures, e.g., tree tensor network (TTN), multi-scale entanglement renormalization ansatz (MERA) [49] and quantum convolutional neural networks (QCNNs) [19], the discriminative error deviates from the Helstrom limit, even for translation-invariant (TI) or less entangled input states. Indeed, these non-extensive architectures enable better trainability [50] at a cost. For extensive architectures that allows the continuous increase of VQC depth, we find that the discriminative power is closely connected to the scrambling power of VQCs.

To reduce the complexity in experimental implementation, we consider simplified VQCs with less parameters or gates. Given the same VQC architecture, for symmetric input states, assuming symmetric VQC gate parameters makes the VQC much easier to train while still competitive in the error probability performance; for real ground states from many-body systems, restricting the VQC to implement a real unitary significantly reduces the number of gates required to achieve the optimal performance. Indeed, simplification of VQCs helps state discrimination only when properly utilizing the symmetry of the input states.

2. Circuit architecture and main results

As shown in figure 1, to perform state discrimination, our MLE-VQC system utilizes a VQC to process the input state, and then performs measurement. Different from existing approaches [17–20], we consider a measurement on all qubits and optimal MLE post-processing to perform the state discrimination task. In this paper, we will consider multiple-qubit input states, while infinite-dimensional quantum states can be potentially considered generalizing the approach in references [22, 23].

A VQC is determined by the type of allowed gates and the circuit architecture. To discriminate between general states, we allow each (two-qubit) gate in the VQC to be universal, composed of single-qubit rotations and CNOT gates. As each gate only acts on at most two qubits, the spread and processing of quantum information is determined by the circuit architecture. We will start with the simple 1D 'brickwall' local circuits (see figure 2(a) or figure 7(a)) with interchanging between gates acting on two set of neighboring pairs. In section 5, we benchmark between different circuit architectures, including extensive ones (brickwall, prism and polygon [52]) and non-extensive ones (QCNN [19, 20], TTN and MERA [49]), as shown in figure 7. We also explore restricted gate sets to simplify VQCs for near-term implementations.

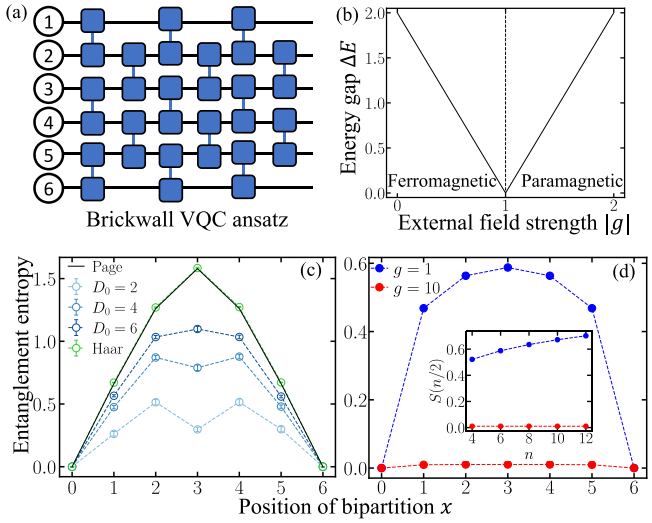

Figure 2. (a) The open boundary local random unitary circuit, 'brickwall'. Every pair of connected boxes represent local two-qubit unitary gate. Here we show an example of a depth $D_0 = 6$ circuit on $n = 6$ qubits input. (b) Phase diagram for TFIM. The black curve represents the analytical expression for energy gap as $\Delta E = 2\left(|g| - 1\right)$, and the dashed line indicates the critical point. (c) Ensemble-averaged bipartite von Neumann entanglement entropy (see equation (4)) of states prepared by circuit in (a). Black curve is the Page curve of Haar random states defined in equation (5). (d) Bipartite entanglement entropy curve for $n = 6$ spins TFIM ground states with $g = 1, 10$. Inset shows the maximal bipartite entanglement entropy with different n.

In order to retrieve the maximal information from the states output by the VQC, we perform simultaneous single-qubit measurements on all n qubits in the system, as shown in figure 1. After the measurement, a decision is made on the input state. The performance of such a VQC state discrimination system is described by the error probability. For example, when discriminating between a pair of equal-prior pure states $\{\psi_0, \psi_1\}$, the error probability refers to the probability of events where the decision $\tilde{H}$ is not equal to the true state label H, namely

$$P_E(U_D; \psi_0, \psi_1) = 1 - \frac{1}{2}\sum_{h=0}^{1} P_{\tilde{H}|H}(h|h), \tag{1}$$

where $P_{\tilde{H}|H}(\tilde{h}|h)$ is the conditional probability of making the decision of $\tilde{H} = \tilde{h}$ (the state is $\psi_{\tilde{h}}$) while the true label is $H = h$ (the state is ψ_h) and the dependence on the VQC unitary U_D is implicit. In this paper, we take the MLE decision strategy, where the decision on the input state is chosen to maximize the posterior probability of the measurement outcome (see appendix A for details). Given the VQC and the final measurements, MLE is known to be the optimal decision strategy that minimizes the error probability.

The minimum 'Helstrom' error probability [46, 47] further optimizes over the measurement bases, leading to the ultimate error probability of state discrimination (see appendix A). When discriminating between a pair of equal-prior pure states $\{\psi_0, \psi_1\}$, the Helstorm limit has a simple closed-form

$$P_H(\psi_0, \psi_1) = \frac{1}{2}\left[1 - \sqrt{1 - |\langle\psi_0|\psi_1\rangle|^2}\right]. \tag{2}$$

We train the VQC unitary U_D to achieve the lowest error probability $P_E(U_D; \psi_0, \psi_1)$ in state discrimination ('dis') between states $\{\psi_0, \psi_1\}$. The corresponding cost function for training is chosen to be

$$\mathcal{C}_{\text{dis}}(U_D; \psi_0, \psi_1) \equiv P_E(U_D; \psi_0, \psi_1) - P_H(\psi_0, \psi_1). \tag{3}$$

Details for the VQC training process can be found in appendix F. Below we summarize the organization and main results in this paper, which are obtained on VQCs after a sufficiently long period of training.

In section 3, we introduce the states being considered for the discrimination task. To make our results on discrimination representative, we consider states generated from random local circuits and ground states of many-body systems.

In section 4.1, we show that our MLE-VQC strategy exponentially suppresses the error with the growth of the circuit depth until the saturation to the Helstrom limit, when discriminating between complex quantum states. The VQC circuit does so by engineering the output state to be highly entangled such that local measurements can realize complex positive-valued operator measure (POVM) elements. In section 4.2, we further demonstrate the importance of the MLE decision strategy. Given the same circuit

architecture and gate set, the performance achievable by a single qubit measurement deviates from the Helstrom limit by an order of magnitude more, compared to the deviation of the MLE case. In addition, the MLE strategy also makes the training easier by increasing the gradients.

In section 4.3, we explore the error decay for states generated by random circuits with different depths. When the complexity of input states—quantified by the preparation circuit depth D_0—increases, the depth of the VQC circuit required to achieve close to the Helstrom limit increases linearly with D_0.

Compared with other tasks such as state generation, we find that the VQC state discrimination task is easier, both in terms of the error and trainability, as detailed in section 4.4. In addition, symmetry in the inputs makes the Helstrom limit larger, but otherwise preserves all other characters in a VQC state discrimination task.

Section 5.1 addresses the benchmark between VQC architectures. Regardless of whether the random input states are symmetric or not, the extensive architectures with a constant number of gates per depth (brickwall, prism, polygon) work much better than those with a limited depth (QCNN, TTN and MERA). This shows a limitation of those over-simplified ansatzs, despite their advantage in trainability [50]. Due to nonlocal gates, prism and polygon have slight advantages in error probability over the brickwall architecture, consistent with their scrambling powers [52–54, 56, 57], as verified by operator size calculations in section 5.2.

Section 5.3 addresses the simplification of VQCs for near-term implementation. When the input is symmetric, given the same circuit architecture, the symmetric VQC ansatz works almost as good as the general VQC ansatz, and is much easier to train due to larger gradients. For ground states of a time-reversal symmetric Hamiltonian that have real wave functions, assuming a real matrix representation of the VQC circuit offers similar performance advantages. However, simplifications not based on symmetry and structure of the input can harm the performance.

Finally, we conclude with some additional discussions in section 6.

3. Ensemble of states under discrimination

In this section, we explain the ensemble of states being considered to benchmark the performance of our MLE-VQC systems in quantum state discrimination. In order to make the benchmark representative, we consider a wide range of applications for quantum state discrimination, in quantum communication, quantum sensing and many-body physics, which involves different types of states. In quantum communication, the decoding of classical information can be considered as a quantum state discrimination task, and the direct coding part of capacity theorems are often obtained via a hypothesis testing approach [58–62]. There, the state involved can be simple, for example coherent states in optical communication, and can also be entangled across a large number of inputs in more advanced encoding. In quantum sensing, distributed sensing [63, 64] and other applications [65–68] involve entangled state in a complex form. In many-body physics, people are interested in detecting complex quantum phases of matter, which involves ground states that can be highly entangled [19, 20].

3.1. States generated from local quantum circuits

To represent different classes of states involved, we consider quantum states generated by inputting a trivial product state $|\mathbf{0}\rangle = |0\rangle^{\otimes n}$ to local quantum circuits, composed of general two-qubit gates acting on neighboring qubits (see figure 2(a)). As we choose the gates randomly, the ensemble is characterized by the preparation circuit depth D_0, and therefore denoted as $\mathbb{H}(D_0)$. We utilize entanglement entropy as a measure on the complexity of the states generated [69, 70]. For a quantum system in state ρ_{AB} divided into subsystem A and B, the von-Neumann entanglement entropy between A and B is

$$S(A,B) = -\operatorname{Tr}\left(\rho_A \log \rho_A\right), \tag{4}$$

where $\rho_A = \operatorname{Tr}_A(\rho_{AB})$ is the reduced density matrix of subsystem A. Typically, the entanglement of states in $\mathbb{H}(D_0)$ and thus circuit complexity grows linearly with depth D_0 [69, 70] before saturation. As shown in figure 2(c), when $D_0 < n$ is a fixed constant, the states generated have area-law entanglement where the bipartite entanglement $S(A,B)$ in a system only depends on the boundary, a constant in 1D system; while when $D_0 \propto n$ is large, the states are typically highly entangled under a volume-law where entanglement $S(A,B)$ is characterized by the size (number of qubits) of subsystem. The growth of entanglement is also analytically characterized by reference [53]. Indeed, $\mathbb{H}(D_0)$ well captures the different problems of interest. In quantum communication, states in $\mathbb{H}(D_0)$ are used as the random encoding [7] to achieve capacity; in many-body physics, the depth D_0 will control the bond dimension of the matrix product representation of

states. Moreover, $\mathbb{H}(D_0)$ are also studied in quantum information scrambling [53, 54] and t-design complexity [71–74].

For $D_0 \to \infty$, the ensemble $\mathbb{H}(\infty)$ approaches Haar random, where the average entanglement between subsystem A and B, known as Page curve, is [55]

$$S_{\mathbb{H}(\infty)}(A,B) \simeq \log d_A - \frac{d_A}{2d_B} \tag{5}$$

where d_A, d_B are the dimension of subsystem A and B with the assumption that $d_A \leqslant d_B$. The typical Helstrom limit between states $\psi_0, \psi_1 \in \mathbb{H}(\infty)$ can be evaluated

$$\langle P_{\mathrm{H}}(\psi_0,\psi_1)\rangle_{\mathbb{H}(\infty)} = \frac{1}{2\left(2^{n+1}-1\right)} \sim \frac{1}{2^{n+2}}, \tag{6}$$

For a finite D_0, when $2^{n+1} \gg 1$, equation (6) still holds to the leading order (see appendix B for details).

As many-body systems often have a translational symmetry, we therefore also consider the subset of TI states $\mathbb{S}(D_0)$, which is prepared by the periodic boundary TI local random unitary circuit with depth D_0. The typical Helstrom limit for $\mathbb{S}(D_0)$ is larger than equation (6) for $\mathbb{H}(D_0)$, however still independent of D_0 to the leading order (see figure A1). The increase of typical Helstrom limit for ensembles with symmetry can be understood as follows. The Helstrom limit for two pure states is directly related to the overlap between them; when there is symmetry or other constraints on the random states, the states come from a smaller Hilbert space and therefore have a larger typical overlap. This larger overlap then leads to larger Helstrom limit from equation (2).

3.2. Ground states of many-body systems

We also consider ground states of many body systems. We focus on the well-known toy model, transverse-field Ising model (TFIM), whose Hamiltonian is

$$H_{\mathrm{TFIM}} = -\sum_i Z_i Z_{i+1} + g\sum_i X_i, \tag{7}$$

where Z_i, X_i are Pauli matrices at site i and g is the strength of external field relative to the coupling strength. To reduce the finite-size effects, we consider a periodic boundary condition. As depicted in figure 2(b), when $|g| < 1$, the system stays in an ordered ferromagnetic phase; as $|g|$ increases, it transits to disordered paramagnetic phase. In both phases, the ground states of the system show area-law entanglement (figure 2(d)). At the critical point, $|g| = 1$, the system undergoes a quantum phase transition. The entanglement entropy shows a logarithmic scaling behavior, which can also be described by conformal field theory with central charge $c = 1/2$ [75–77].

4. Performance of the brickwall VQC in state discrimination

With the VQC circuit depth increasing, the performance of the MLE-VQC system will eventually approach the ultimate Helstrom limit for pure state discrimination. This is because for an ensemble of pure states being considered in this paper, the optimal POVM elements are also rank-one projectors [78, 79]; therefore additional ancilla is not necessary in the measurement. At the same time, however, the training of the circuit becomes harder as the number of parameters increases and the gradient decays [42, 43, 81–85].

In this section, we explore the error probability performance and trainability of the MLE-VQC system with the open-boundary brickwall VQC ansatz. The periodic-boundary VQC shows similar results, as we shown in appendix D. To understand the performance with a finite depth circuit, in section 4.1 we evaluate the decay of error probability towards the Helstrom limit for different ensemble of input states. Then we compare the performance and trainability between the MLE approach and single-qubit measurement approach in section 4.2. Afterwards, we explore the connection between the performance to the input state ensemble complexity in section 4.3. We close the section by a comparison between generative tasks and discriminative tasks for VQCs in section 4.4.

4.1. Fast error decay

To begin with, we consider the average error probability for complex states discriminated by VQCs with different depth D. In figure 3(a), we consider Haar random states (blue dots) and find a fast error suppression—the error probability decays exponentially with D before saturation to the Helstrom limit (blue dashed horizontal line). To represent the symmetric case, we further consider the set of states $\mathbb{S}(2^n)$ prepared by TI local quantum circuits with a large enough depth $2^n \gg 1$. While symmetry increases the

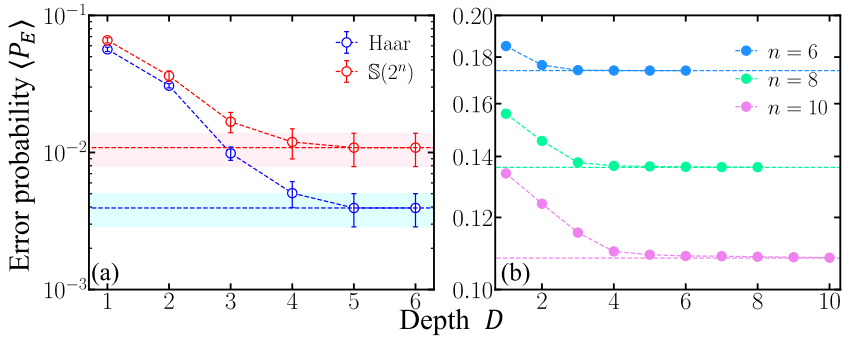

Figure 3. Error probability of binary state discrimination (between two quantum states) with trained MLE-VQC systems. (a) State discrimination between Haar random states or complex TI states $\mathbb{S}(2^n)$ in a system of $n = 6$ qubits. Horizontal dashed lines represent the ensemble-averaged Helstrom limit $\langle P_H \rangle$ and light color area represents the amount of ensemble fluctuation. (b) State discrimination between ground states of TFIM with $g = 1$ and $g = 10$ in a system of $n = 6, 8, 10$ qubits, where the y-axis is in a logarithmic scale. Horizontal dashed lines show the Helstrom limit in each case. We use the open-boundary brickwall VQC ansatz in all cases.

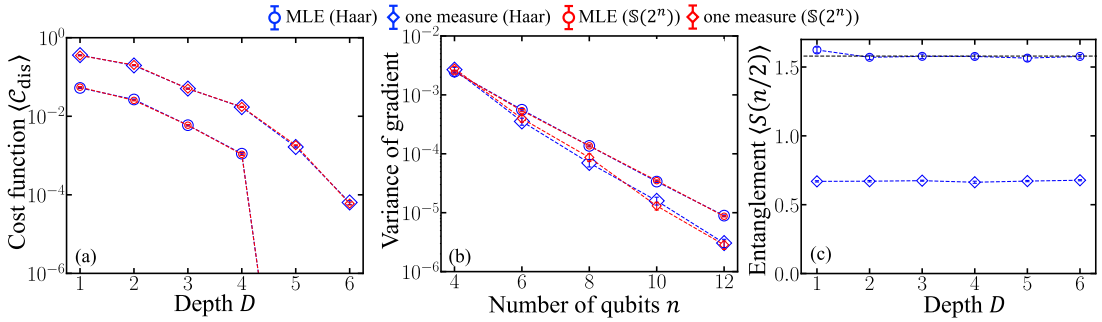

Figure 4. MLE versus single-qubit measurement approach in a brickwall ansatz VQC for discriminating between Haar random states and TI states sampled from $\mathbb{S}(2^n)$. We consider a system of $n = 6$ qubits. We show the ensemble-averaged cost function $\langle \mathcal{C}_{\rm dis} \rangle$ in (a) and the parameter-averaged variance of gradient $\langle \mathrm{Var}\,(g_i) \rangle_i$ for ansatz with depth $D = 2$ in (b). We show the half bipartite von Neumann entanglement entropy $\langle S(n/2) \rangle$ of VQC output states with MLE (circles) and single-qubit measurement (diamonds) in (c). The input states are Haar random states. Black dashed line indicates the Page entanglement (see equation (5)). In (a), the line going below the plot region decreases to zero at the machine precision; we choose the plot range to make the trends clear.

Helstrom limit (red dashed horizontal line), it does not change the exponential suppression of error probability with depth D (red dots). To extend the results beyond random states, we consider the discrimination between two ground states of TFIM with different parameters g in figure 3(b). We see as the number of qubits n increases, the Helstrom limit decreases and the error probability shows an exponentially suppressed trend with VQC depth D. Although the number of qubits is limited due to the increasing level of difficulty in the training, we see the depth required to saturate the Helstrom limit scales linearly with the system size.

It is worthy to point out that the amount of entanglement in states before the final measurement is high in the MLE-VQC approach. For the Haar random input states, the optimal VQC of different depth D preserves the bipartite entanglement entropy $\langle S(n/2) \rangle$ at the Page curve value, as shown by the purple line in figure 5(b). Note that for less entangled inputs $\mathbb{H}(D_0)$ prepared by random local circuits of depth D_0, the VQC circuit increases the level of entanglement $\langle S(n/2) \rangle$ at the output side before the final measurements (green, red, blue for $D_0 = 2, 4, 6$). From this, we see that the VQC is essentially sorting and increasing entanglement between the qubits to enable the best performance on the final separable measurement. In contrast, for the single-qubit measurement approach, as shown in figure 4(c), the entanglement entropy is much lower. This is because when the final measurement is only performed on a single qubit, the circuit tries to concentrate all valuable information onto a single-qubit. Therefore, there is no incentive for the VQC to entanglement the final qubits, rather it tries to disentangle that qubit being measured from the rest of the system.

4.2. MLE's superiority over single measurement schemes

The simultaneous single-qubit measurements on all qubits and optimal MLE decision rule are crucial for our MLE-VQC approach to unleash the full power of the brickwall ansatz. To demonstrate such, we

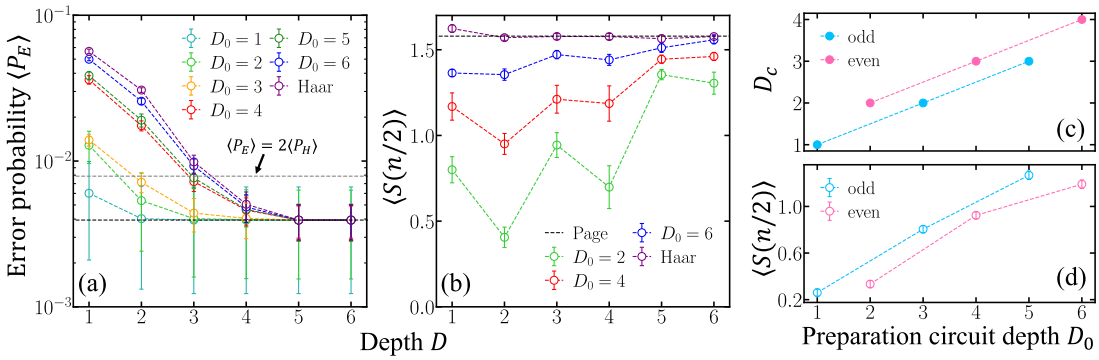

Figure 5. (a) Discrimination error probability using the brickwall ansatz between a pair of random states sampled from $\mathbb{H}(D_0)$ for $n = 6$ qubits. Black and grey dashed horizontal lines show the Helstrom limit $\langle P_H \rangle$ and a critical value $\langle P_E \rangle = 2\langle P_H \rangle$. Note the relative large error bar in the error probability in states $\mathbb{H}(D_0)$ generated by a shallow depth VQC is due to the lack of self-averaging. (b) Average maximal bipartite entanglement entropy $\langle S(n/2) \rangle$ of VQC output states before measurements for $\mathbb{H}(D_0)$ discrimination. (c) The critical depth D_c to achieve $\langle P_E \rangle = 2\langle P_H \rangle$ versus input complexity D_0. (d) Maximal bipartite entanglement entropy $\langle S(n/2) \rangle$ of the input states to be distinguished. In (c) and (d), we plot odd and even D_0 separately.

benchmark the MLE-VQC approach against the VQC with only a single-qubit Z-measurement at the center. To show their difference, we still focus on two ensembles, Haar random states and complex TI states $\mathbb{S}(2^n)$. As shown in figure 4(a), the residual error utilizing MLE-VQC approach is around an order of magnitude smaller than that of the single-qubit approach at a given D, which shows the power of MLE to gather the full information from all qubits. Moreover, the MLE case shows a sharp drop in the error at a large enough VQC depth D, while the single-qubit case has a consistent decay. As for trainability, although the variance of gradient for both approaches decreases exponentially with the number of qubits n, the MLE approach typically has a larger gradient and therefore is easier to train (see figure 4(b)). Note that the local two-qubit gates in VQC ansatzs we study are universal, and thus the depth of VQCs falls in the range of barren plateau for both single-qubit approach and MLE strategy, whose trainability are much different in shallow VQCs according to their corresponding local and global cost function [43].

4.3. Linear growth of complexity

Here we explore how the complexity of the input state ensemble affects the error probability. As explained in section 3.1, we can tune the complexity of output states $\mathbb{H}(D_0)$ produced in a depth-D_0 local random circuit by controlling the depth D_0; therefore, we study the discrimination between states sampled from $\mathbb{H}(D_0)$.

For states sampled from $\mathbb{H}(D_0)$, the fast suppression of error probability still holds, as shown in figure 5(a). With increasing input states complexity, the discrimination task becomes harder, leading to an increasing error probability for a fixed VQC depth D. Similar to results in figure 5(a), as for binary discrimination between TI states sampled from $\mathbb{S}(D_0)$, the universal exponential suppression still holds with a slightly increased Helstrom limit, as shown in figure D1.

As the saturation towards the Helstrom limit has a long tail, we consider the number of layers D_c required to achieve an error probability $P_E = 2P_H$. In figure 5(c), we see a linear growth of D_c with D_0, as expected. This can also be explained by a sub-optimal strategy mimicking the Kennedy receiver [86], which implements the POVM element $\Pi_0 = |\psi_0\rangle$ with a depth $D \sim D_0$ VQC to achieve the error probability $|\langle\psi_0|\psi_1\rangle|^2/2 \simeq 2P_H$ when $P_H \ll 1$. One can also understand the increase from the increase of entanglement in the input state, which also shows a linear trend, as depicted in figure 5(d). Our error probability results are obtained from a finite system of six qubits, however, extending to larger systems to further consolidate the conclusion is challenging, due to the exponential decay of gradient shown in section 4.4.

4.4. Comparison between state generation and discrimination: performance and trainability

To understand the level of difficulty of state discrimination ('dis'), we benchmark with the most relevant task of state generation ('gen') [15, 16]. In an n-qubit state generation task, the VQC performs a unitary U_D on a trivial product state $|\mathbf{0}\rangle = |0\rangle^{\otimes n}$ to approximate a target state $|\psi\rangle$. Similar to the discrimination case in equation (3), we utilize the following cost function

$$\mathcal{C}_{\text{gen}}(U_D; \psi) \equiv 1 - |\langle\psi|U_D|\mathbf{0}\rangle|^2, \tag{8}$$

as a function of the VQC unitary U_D; details for the numerical optimization process can be found in appendix F.

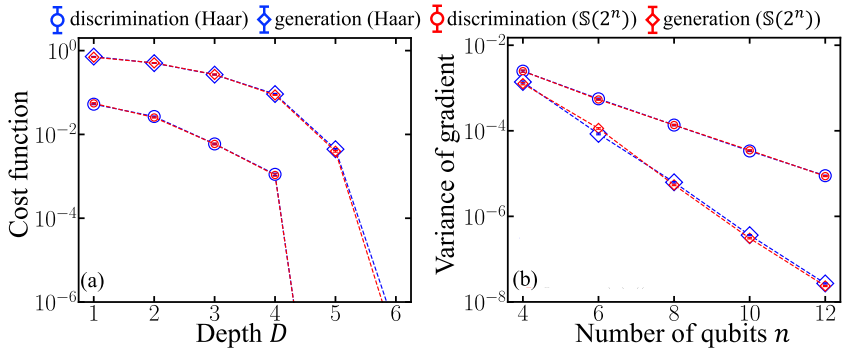

Figure 6. Performance and gradient of the open-boundary brickwall VQC. We consider two different state ensemble, Haar random states and TI states $\mathbb{S}(2^n)$. (a) Ensemble-averaged cost functions for discrimination and generation. Blue and red curves show ensemble of Haar random states and TI states separately in a system of $n = 6$ qubits. (b) Parameter-averaged variance of gradient $\langle \mathrm{Var}\left(g_i\right)\rangle_i$ in the $D = 2$ ansatz. In (a), the lines going below the plot region decrease to zero at the machine precision; we choose the plot range to make the trends clear.

From figure 6(a), we can identify a sharp transition in the cost function $\mathcal{C}$ for both discrimination and generation, where $\mathcal{C}$ is exponentially suppressed before reaching an extremely small value. Although we can find that $\mathcal{C}_{\mathrm{dis}}$ is about an order of magnitude smaller than $\mathcal{C}_{\mathrm{gen}}$, due to the different cost function definition, the gap does not necessarily indicate one task is harder than the other. The same conclusion is also confirmed for random states sampled from $\mathbb{H}(D_0)$ and $\mathbb{S}(D_0)$ with a finite D_0 (see figure B3). Although we have focused on the brickwall ansatz in figure 6(a) in this section, the same gap between discrimination and generation exists in other architectures as we will discuss in section 5.

To explore the trainability of VQCs, we evaluate the gradient of the cost functions to the parameters. For both the generation and discrimination tasks, the cost function gradient g_i with respect to parameter θ_i can be obtained numerically from a central finite-difference (see appendix F). Note that as the gradient can be positive or negative, we evaluate the variance $\mathrm{Var}(g_i)$ among different positions to get a sense of the magnitude of gradients, similar to reference [42]. Moreover, we take an average over the different gradient directions to obtain the average variance of gradient $\langle \mathrm{Var}(g_i)\rangle_i$. In figure 6(b), the parameter-averaged variance of gradient decays exponentially with the number of qubits n, predicted by the well-known barren plateau phenomena [42]. The barren plateau, combined with the high average entanglement from t-design, makes training specifically hard in large systems for the t-design VQCs we study here; to mitigate the difficulty, we discuss the simplification of VQC ansatzs and utilization of symmetry in section 5.3.

5. Performance benchmarks of different ansatz

In section 4, we employ the brickwall VQC ansatz, where each two-qubit gate is applied to pairs of nearest neighbors in one dimension. As we mentioned in section 2, various other architectures have been proposed for different tasks. Since we find that VQCs relying on increasing the amount of entanglement in the quantum states to approach the Helstrom limit, we expect that architectures with non-local gates might improve the discriminative power of VQCs. In section 5.1, we offer a benchmark between different architectures to confirm this. Then we provide insights into the different performance by evaluating the scrambling power of the VQCs in section 5.2.

Despite the different architectures, in the NISQ era, VQC implementations are limited in the circuit depth and number of gates, due to the accumulation of device imperfections. Therefore, we further explore simplification of the gate sets in section 5.3. In particular, we find that symmetry in the input states allows VQCs to be simplified.

5.1. Comparison between different architectures

In this section, we benchmark the various VQC architectures (see figure 7) for both state discrimination and generation tasks. An architecture determines the layout of the quantum gates and therefore constrains the information flow in the VQC. In section 4, we have focused on the local brickwall ansatz. To extend the interactions beyond nearest-neighbor, prism and polygon [52] architectures generalize the line geometry to different shapes. These three architectures are extensive—they have the number of gates per layer roughly unchanged as the depth of the circuit increases. Other popular choices of architectures have a fixed depth and therefore are not extensive, including QCNN [19, 20] and tensor network architectures (TTN and MERA) [49].

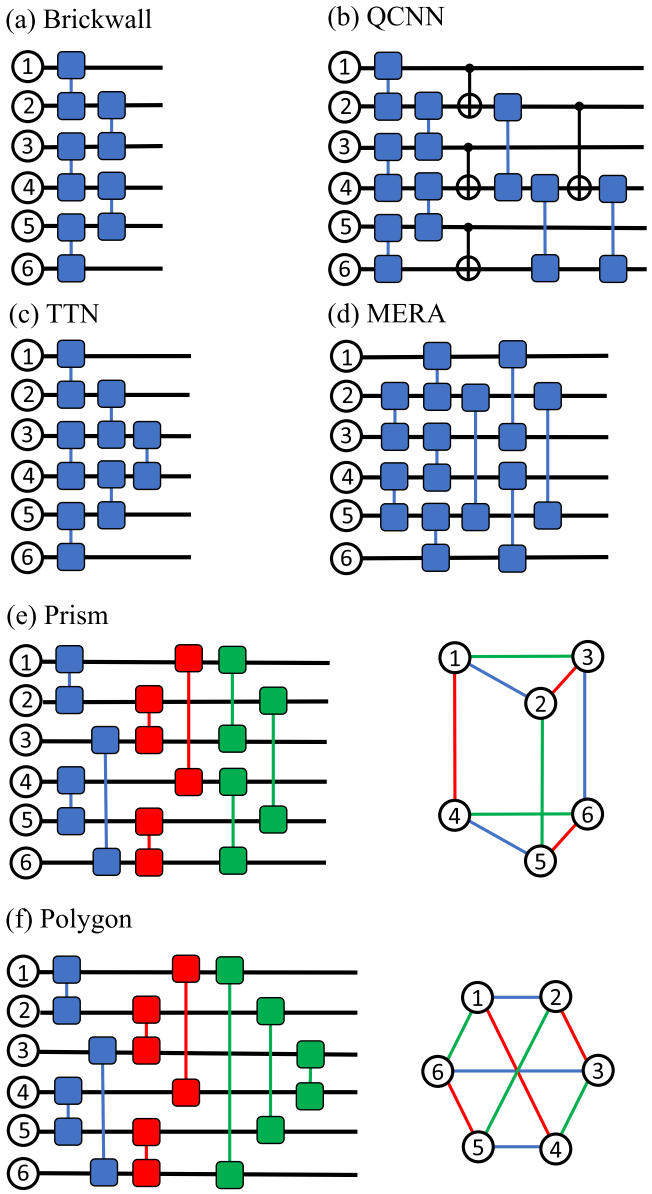

Figure 7. Topological architectures of VQCs. Two connected boxes represents a universal two-qubit gate. For (a) brickwall, (e) prism and (f) polygon ansatzs, we show the case with depth $D = 2, 3, 3$ separately.

In figure 8(a), we begin the benchmark with the error probability performance in discriminating Haar random state pairs. We see that the extensive architectures (brickwall, prism and polygon) provide a better performance over the non-extensive architectures (QCNN, TTN and MERA) at the same depth D. In particular, the extensive ones saturate the Helstrom limit at $D = 5$ exactly, while the non-extensive ones are far from the Helstrom limit even at a larger depth. Among the three extensive architectures, we find prism and polygon to be slightly better than the brickwall architecture at a finite depth D, due to non-local gates more efficiently processing the global quantum information. The same conclusions also generalize to the complex TI inputs $\mathbb{S}(2^n)$, as shown in figure 8(c). In some sense, it is expected that QCNN, TTN and MERA do not work well, as they are developed for problems with specific structure and symmetry.

We also consider the task of state generation, and find similar conclusions to hold—the relative ordering of the error is identical to that in state discrimination, as shown in figures 8(b) and (d) for the Haar ensemble and TI ensemble. This shows a consistent ordering of quantum information processing power among the architectures, which also agrees with the quantum information scrambling capabilities explored in section 5.2. Comparing between the state generation task in figures 8(b) and (d) and the state discrimination task in figures 8(a) and (c), we also extend the previous conclusion in section 4.4 that state discrimination is easier than state generation to VQCs with different architectures.

Comparing figure 7 bottom panels (c) and (d) with the top panels (a) and (b), we can see that for a general VQC, translational symmetry in the input merely increases the Helstrom limit, while at any depth

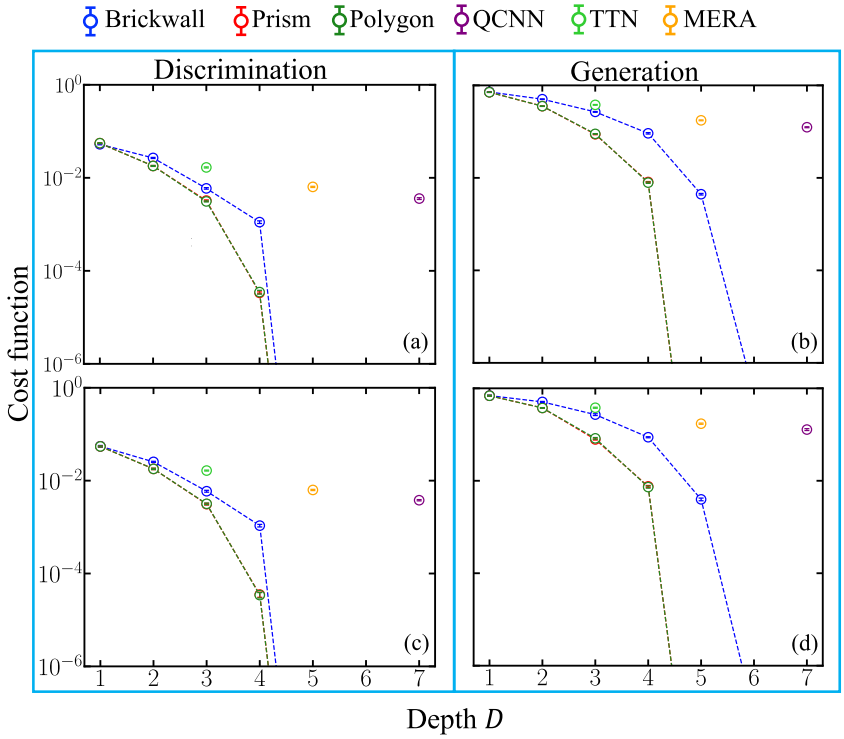

Figure 8. Cost functions for the discrimination (left) and generation (right) among Haar random states (top) or complex TI states $\mathbb{S}(2^n)$ (bottom) in a system of $n = 6$ qubits. We compare performance with different VQC architectures. For comparison, the Helstrom limit $\langle P_H \rangle \sim 10^{-2.4}$ for (a) and $\langle P_H \rangle \sim 10^{-2}$ for (c), as shown in figure 3. The relative differences in the cost functions between top and bottom panels are below 3%. The lines going below the plot region decreases to zero at the machine precision; we choose the plot range to keep the trends clear. Note that each of the non-extensive architectures (TTN, MERA, QCNN) have a single fixed depth and therefore is represented by a single dot.

the deviations to the Helstrom limit (the cost function) are almost identical with and without input symmetry. Although here symmetry does not make much difference, it will allow simplifications of the VQC, as we will explain in section 5.3.

Finally, we want to emphasize that even for less complex states $\mathbb{H}(D_0)$ and $\mathbb{S}(D_0)$ prepared by low-depth circuits with and without symmetry, the non-extensive architectures (QCNN, TTN and MERA) are still worse than the extensive architectures at the same depth, as shown in figures B3 and B4 in appendix E. In this sense, these non-extensive architectures trade-off performance for the smaller number of quantum gates and parameters. To obtain the optimal performance, extensive architectures are preferable.

5.2. Discriminative power versus scrambling power

As the circuit depth increases, a VQC generates more entangled outputs in order to achieve a better error probability in state discrimination (see figure 5(b)). The expressibility and entangling power of VQCs are also studied in terms of the connection pattern of two-qubit gates in a relatively small system from a general perspective of view [51]. Because entanglement growth is also an important indicator of quantum information scrambling in the circuit, we evaluate the scrambling power of the VQCs utilized in section 5, in comparison to their performances.

Similar to reference [52], we choose the operator size [56] as the metric to evaluate the scrambling power of VQCs. To define operator size, we consider a Pauli-Z operator $M_0 = I_1 \otimes \cdots Z_{n/2} \otimes \cdots \otimes I_n$ initially located at the center, where I_k is an identity operator acting on the kth qubit. Under the VQC represented by unitary U_D, the operator evolves to $M_D = U_D^\dagger M_0 U_D$. In general, M_D can be expanded in the Pauli bases, i.e., $M_D = \sum_{\mathcal{S}} \alpha_{\mathcal{S}} \mathcal{S}$, where $\mathcal{S} = \otimes_{k=1}^n \sigma_k$ is a Pauli string with σ_k being one of the four Pauli operator I_k, X_k, Y_k, Z_k at kth qubit. The size of the evolved operator

$$\text{Size}\,(M_D) = \sum_{\mathcal{S}} |\alpha_{\mathcal{S}}|^2 L\,(\mathcal{S})\,, \tag{9}$$

where $L\,(\mathcal{S})$ is the number of non-identity elements in the Pauli string $\mathcal{S}$. The operator size starts from the minimum value of unity when the operator M_D is a single-qubit local operator, and saturates to the maximum value of $3n/4$ when it is uniformly distributed in the space spanned by Pauli strings. We

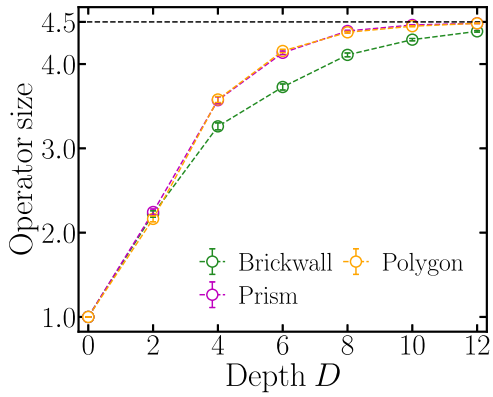

Figure 9. Ensemble-averaged operator size growth with the VQC depth for Z initially localized at the $n/2$th qubit in a $n = 6$ system.

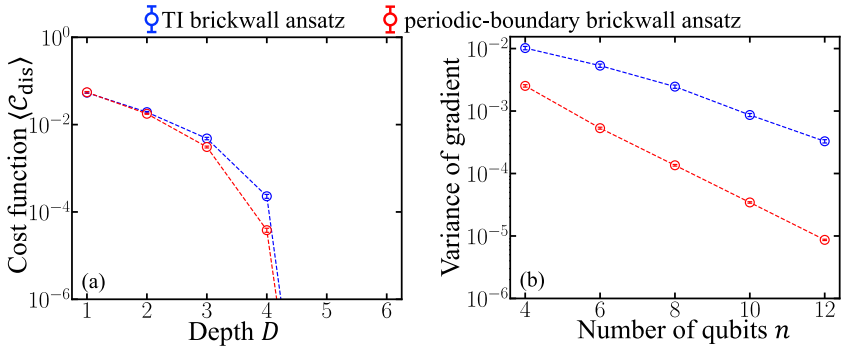

Figure 10. Cost function $\langle \mathcal{C}_{\rm dis} \rangle$ (a) and average variance of gradient $\langle \mathrm{Var}\left(g_i\right) \rangle_i$ (b) of different brickwall ansatz in discriminating between random states sampled from $\mathbb{S}(2^n)$ with number of qubits $n = 6$. We take TI and periodic-boundary brickwall ansatz. In (b) all ansatzs are set to be $D = 2$. In (a), the lines going below the plot region decrease to zero at the machine precision; we choose the plot range to make the trends clear.

numerically study the ensemble-averaged operator size with different VQC architectures in figure 9, when each two-qubit gate is randomly chosen. Comparing the operator size growth with the performances in state discrimination and generation of figure 8, we find the same ordering for all VQC architectures. This consistency confirms the connection between the scrambling power and the discrimination power of VQCs.

5.3. Performance of simplified gate sets

In the NISQ era [2], quantum circuit implementations are limited due to device imperfections. In particular, the imperfections accumulate with the increase of the number of quantum gates and the depth of the circuit. In all the VQC architectures being explored in section 5.1, the realization of a universal two-qubit gate in fact requires three CNOT gates and additional qubit rotations (see appendix C), creating extra burdens in the VQC implementation. Another major constraint comes from the vanishing gradient due to barren plateau [43] that prevents the efficient training of VQCs, which limits the scale of the implementations. In this section, we consider different ways to simplify the quantum gates in the brickwall VQC and probe the induced change of the performance and trainability in state discrimination.

As we are often considering state discrimination between TI states, a natural attempt to simplify the VQC is to enforce TI on each layer of the VQC, including the gate parameters and a periodic boundary. As the TI symmetry reduces the number of parameters, we expect TI VQCs to be more trainable, which is confirmed by the gradient evaluations in figure 10(b): the TI VQC typically shows a much larger gradient. Although assuming symmetry might lose some performance, however, as show in figure 10(a), we find that when the input is symmetric, the TI brickwall ansatz (blue) provides almost identical performance to the periodic-boundary brickwall ansatz without the symmetry constraint (red). This shows that enforcing the VQC to have the same symmetry as the input simplifies the training of the VQC, while not losing much performance. Similar benefits from simplification are being explored in other tasks [80].

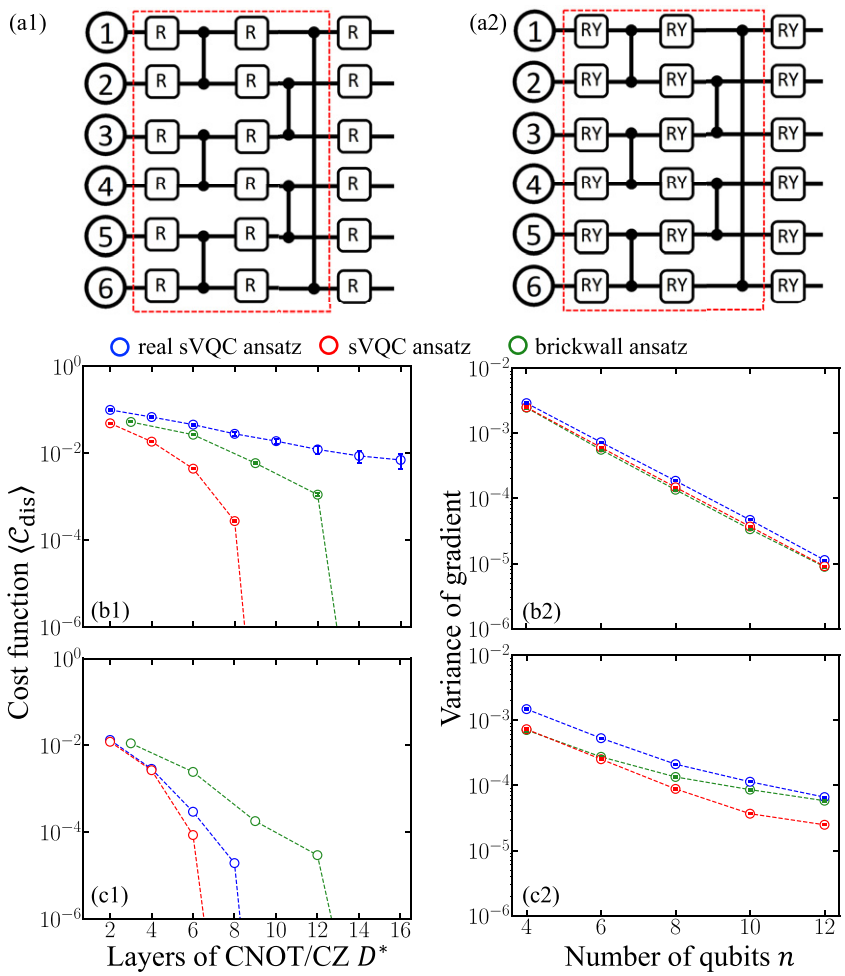

Figure 11. Architectures and performance for NISQ VQCs. (a) Layout of an $n = 6$-qubit sVQC ansatz and real sVQC ansatz, plotted in (a1) and (a2) separately. The sVQC ansatz consists of CZ gates and generic single qubit rotations, and the real sVQC ansatz consists of CZ gates and RY rotations. The circuits surrounded by the red dashed box represent $D^\star = 2$ layers and at the end of circuit each qubit is applied with a rotation. (b1) and (b2) Cost function $\langle \mathcal{C}_{\rm dis} \rangle$ and average variance of gradient $\langle \mathrm{Var}\,(g_i) \rangle_i$ for discriminating between Haar random states. (c1) and (c2) Cost function and average variance of gradient for discriminating between TFIM ground states with $g = 1, 10$. We benchmark in a system of $n = 6$ qubits. In (b2) and (c2) we take $D^\star = 6$ for all ansatz. Note for the brickwall ansatz, it corresponds to $D = 2$. In (b1) and (c1), the lines going below the plot region decrease to zero at the machine precision; we choose the plot range to keep the trends clear.

Next, we consider simplifying the set of gates in VQCs. We replace each universal two-qubit gate in figure 7(a) (which requires three CNOT gates) with a single CZ gate[4], and insert general single-qubit rotations in between each layer of CNOT gates, leading to the simple VQC (sVQC) of figure 11(a1) similar to the designs in references [9, 33]. In an sVQC, each single-qubit rotation $R(\theta_1, \theta_2, \theta_3) \equiv \mathrm{e}^{-\mathrm{i}\theta_1 Z/2}\, \mathrm{e}^{-\mathrm{i}\theta_2 Y/2}\, \mathrm{e}^{-\mathrm{i}\theta_3 Z/2}$ is characterized by three angles θ_1, θ_2 and θ_3, where Z, Y are Pauli matrices. We can further reduce the complexity of each single-qubit gate by constraining its matrix representation to be real (via fixing $\theta_1 = \theta_3 = 0$), leading to the real sVQC architecture which only implements real unitary U_D in figure 11(a2). Real sVQCs are widely utilized in eigen-solvers [13, 14], as the ground state of a time-reversal-symmetric Hamiltonian can be taken as real. The overall number of parameters (which equals the number of single qubit gates) and the number of CNOT/CZ gates are listed in table 1 for comparison: both sVQC and real sVQC reduce the number of CNOT/CZ gates by a factor of three; while sVQC reduces the number of parameters roughly by half, the real sVQC reduces the number of parameters roughly by a factor of seven.

To ensure a fair comparison of error probability performance and trainability, instead of the depth D of VQCs in terms of universal two-qubit gates that we consider in most of the paper, we count the number of layers of CNOT/CZ gates $D^\star$ in the final physical implementations in the original (brickwall) ansatz, sVQC ansatz and real sVQC ansatz. In figure 11, we find that sVQC outperforms the original ansatz consistently at the same $D^\star$, while the trainability barely changes. This is due to a certain level of redundancy in requiring

[4] An extra CZ gate is inserted every two layers to form a periodic boundary condition.

Table 1. The number of parameters and CNOT/CZ gates of brickwall ansatz, sVQC and real sVQC ansatz to the leading order $O(nD)$ with depth D in a system of n qubits.

Number	Brickwall	sVQC	Real sVQC
Parameters	$\sim \frac{15}{2}nD$	$3n(D+1)$	$n(D+1)$
CNOT/CZ gates	$\sim \frac{3}{2}nD$	$\sim \frac{1}{2}nD$	$\sim \frac{1}{2}nD$

each two-qubit gate to be universal in the original ansatz. The real sVQC further restricts the unitary implemented by the VQC to be real without losing much performance compared to sVQC in the discrimination between real ground states of TFIM, as shown in figure 11(c1); indeed, the trainability of the real sVQC improves due to the further simplification, as indicated by the larger gradients shown in figure 11(c2). While for Haar random states, due to the lack of symmetry of the input, such a restriction to real unitaries harms the performance while not improving the trainability, as shown in figures 11(b1) and (b2).

6. Discussions and conclusions

To conclude, we propose an MLE-VQC scheme for quantum data classification, which shows an enormous error probability advantage over the single-qubit measurement approach. As the depth of the VQC increases in an extensive way, the error probability of VQCs decreases exponentially towards the Helstrom limit. Despite being popular choices, non-extensive VQCs such as QCNN and MERA are sub-optimal in their error probability performances. The proposed MLE-VQC scheme can be implemented on near-term quantum devices, and has the potential to be used in various applications. It is an important future direction to explore the MLE-VQC's performance in different applications, as in each application the symmetry and structure of the problem vary and may allow additional simplifications.

Finally, we discuss about some important messages of our work. Adopting VQCs to the symmetry of the problem will maintain the discriminative power for the particular problem while simplifying the VQC implementations. For example, when the input is TI, constraining the VQC to be TI not only reduces the number of parameters and improves the trainability, but also preserves the performance; similar advantages apply to assuming real VQCs for real ground states of TFIM. However, oversimplification can be problematic. As we have seen in figure 10, constraining the unitary to be real will significantly harm the performance when discriminating between general complex quantum states. As in the typical case, powerful VQCs are hard to train due to small gradients [43], one needs to utilize the structure and symmetry of the task to simplify the circuit and enable efficient training. Our results indicate that such a simplification needs to be tailored with caution.

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Superconducting Quantum Materials and Systems Center (SQMS) under the Contract No. DE-AC02-07CH11359. QZ acknowledges support from the Defense Advanced Research Projects Agency (DARPA) under Young Faculty Award (YFA) Grant No. N660012014029 and National Science Foundation (NSF) CAREER Award CCF-2142882.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Appendix A. Preliminary of state discrimination

A.1. General Helstrom limit

The minimum 'Helstrom' error probability [46, 47] for the discrimination between m states $\{\rho_i\}_{i=0}^{m-1}$ with prior probability $\{p_i\}_{i=0}^{m-1}$

$$P_H\left(\{\rho_i, p_i\}\right) = 1 - \max_{\sum_i \Pi_i = I} \sum_i p_i \,\mathrm{Tr}\left(\rho_i \Pi_i\right), \tag{A.1}$$

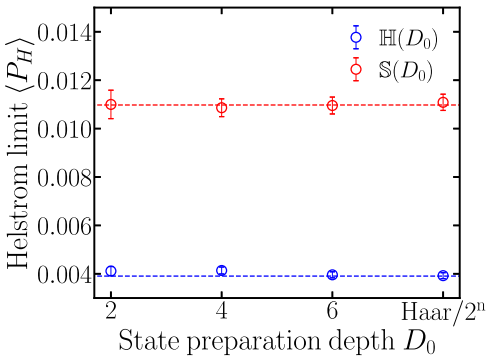

Figure A1. Ensemble-averaged Helstrom limit $\langle P_H \rangle$ for $n = 6$ qubits states sampled from $\mathbb{H}(D_0)$ and $\mathbb{S}(D_0)$. The blue dashed line represents $\langle P_H \rangle \sim 1/2^{n+2}$ and the red dashed line is the average over red dots.

where the POVM element Π_i corresponds to the hypothesis that the state is ρ_i. For the binary pure-state case, equation (A.1) can be reduced to equation (2).

A.2. MLE decision for general state discrimination

We use the notation $P_{A|B}(a|b)$ for the probability of event $A = a$ to happen conditioned on event $B = b$ already happened. For example, in the main text $P_{\tilde{H}|H}(\tilde{h}|h)$ in equation (1) denotes the event of decision $\tilde{H} = h$ conditioned on the true label $H = h$.

Consider the general state discrimination described above, when the VQC implements the unitary U_D on input ρ_i, the measurement result M is equal to j with the probability

$$P_{M|H}(j|i) = \mathrm{Tr}\left(|j\rangle\langle j|U_D\rho_i U_D^\dagger\right), \tag{A.2}$$

where $\{j\}_{j=0}^{2^n-1}$ forms the set of all possible measurement results. Conditioned on the measurement result j, MLE strategy minimizes the average error probability by making the decision on the state $\rho_{\tilde{i}}$ via

$$\tilde{i}(j) = \mathrm{argmax}_i\, p_i P_{M|H}(j|i), \tag{A.3}$$

leading to the minimum error probability for a fixed measurement choice as

$$P_E\left(\{\rho_i, p_i\}\right) = 1 - \sum_{i=0}^{m-1} p_i P_{\tilde{H}|H}(i|i). \tag{A.4}$$

Here the conditional correct probability $P_{\tilde{H}|H}(i|i) = \sum_{\tilde{i}(j)=i} P_{M|H}(j|i)$.

In this paper, we focus on the binary pure state case with equal prior and the MLE error probability is reduced to

$$P_E = \frac{1}{2}\left\{1 - \sum_{j:P_{M|H}(j|0)\geqslant P_{M|H}(j|1)} \left[P_{\tilde{H}|H}(j|0) - P_{\tilde{H}|H}(j|1)\right]\right\}. \tag{A.5}$$

Appendix B. Evaluation of $\langle P_H(\psi_0, \psi_1)\rangle_{\mathbb{H}(D_0)}$

B.1. $n \gg 1$ limit of finite D_0

When $n \gg 1$, we expect the typical state overlap $|\langle\psi_0|\psi_1\rangle|^2$ to be small [73], therefore

$$P_H(\psi_0, \psi_1) = \frac{1}{2}\left[1 - \sqrt{1 - |\langle\psi_0|\psi_1\rangle|^2}\right] \sim \frac{1}{4}|\langle\psi_0|\psi_1\rangle|^2. \tag{B.1}$$

Below we show that regardless of D_0, within the state ensemble $\mathbb{H}(D_0)$, the typical Helstrom limit between states is $\langle P_{\mathrm{H}}(\psi_0, \psi_1)\rangle_{\mathbb{H}(D_0)} \sim 1/2^{n+2}$, which simply follows from the typical overlap $\langle|\langle\psi_0|\psi_1\rangle|^2\rangle_{\mathbb{H}(D_0)} = 1/2^n$. We also show it numerically in figure A1.

Consider an n-qubit system, suppose one considers an initial product state $|0\rangle^{\otimes n}$, for two local quantum circuits implementing unitaries U_0, U_1, the overlap is $|\langle\psi_0|\psi_1\rangle|^2 = \langle\mathbf{0}|U|\mathbf{0}\rangle\langle\mathbf{0}|U|\mathbf{0}\rangle$ where $U = U_0^\dagger U_1$. As U and $U^\dagger$ each appears only once, taking the ensemble average over a one-design [71–73] suffices to produce the Haar value. In our case, we consider random local quantum circuits for U_0 and U_1, with each two-qubit

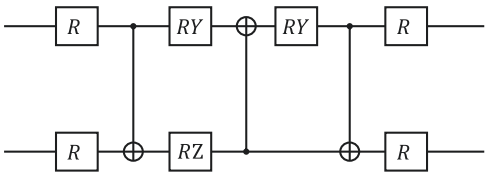

Figure B1. A decomposition of a general two-qubit gate. The R gate represents an arbitrary single qubit rotation with three independent parameters as $R(\theta_1, \theta_2, \theta_3) = RZ(\theta_1)RY(\theta_2)RZ(\theta_3)$.

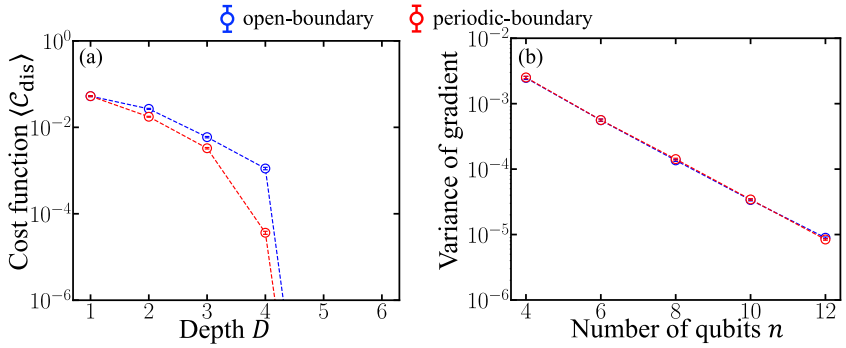

Figure B2. (a) Cost function $\langle C_{\rm dis}\rangle$ and (b) average variance of gradient $\langle \mathrm{Var}\left(g_i\right)\rangle_i$ for discriminating between $n = 6$ qubits Haar random states using open-boundary and periodic-boundary brickwall ansatzs. In (b) we evaluate both ansatzs at the depth of $D = 2$. In (a), the lines going below the plot region decrease to zero at the machine precision; we choose the plot range to make the trends clear.

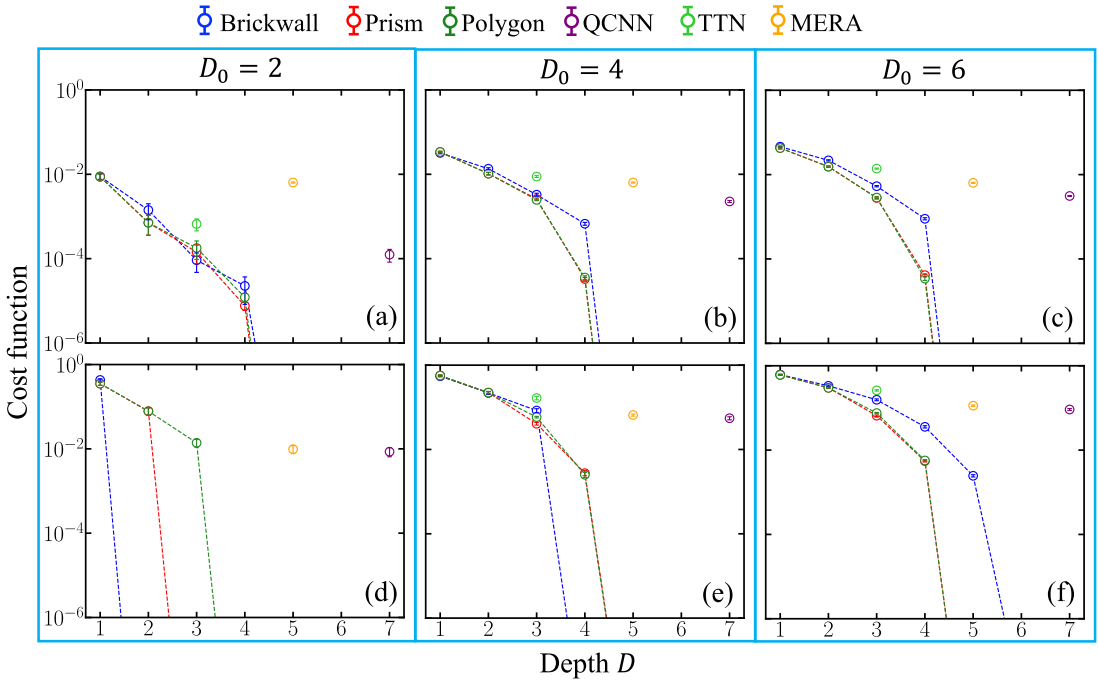

Figure B3. Cost functions of different VQC architectures in the discrimination (top) and generation (bottom) of random states sampled from $\mathbb{H}(D_0)$ with $D_0 = 2, 4, 6$ in a system of $n = 6$ qubits. The lines going below the plot region decrease to zero at the machine precision; we choose the plot range to make the trends clear. Note that each of the non-extensive architectures (TTN, MERA, QCNN) have a single fixed depth and therefore is represented by a single dot. The same applies to figure B4.

unitary Haar random. Regardless of the number of layers of the circuit, the ensemble $\{U_0^\dagger U_1\}$ forms a one-design, therefore

$$\int_{\rm Haar} \mathrm{d}U \langle \boldsymbol{0}|U^\dagger|\boldsymbol{0}\rangle\langle\boldsymbol{0}|U|\boldsymbol{0}\rangle = \frac{1}{2^n}, \tag{B.2}$$

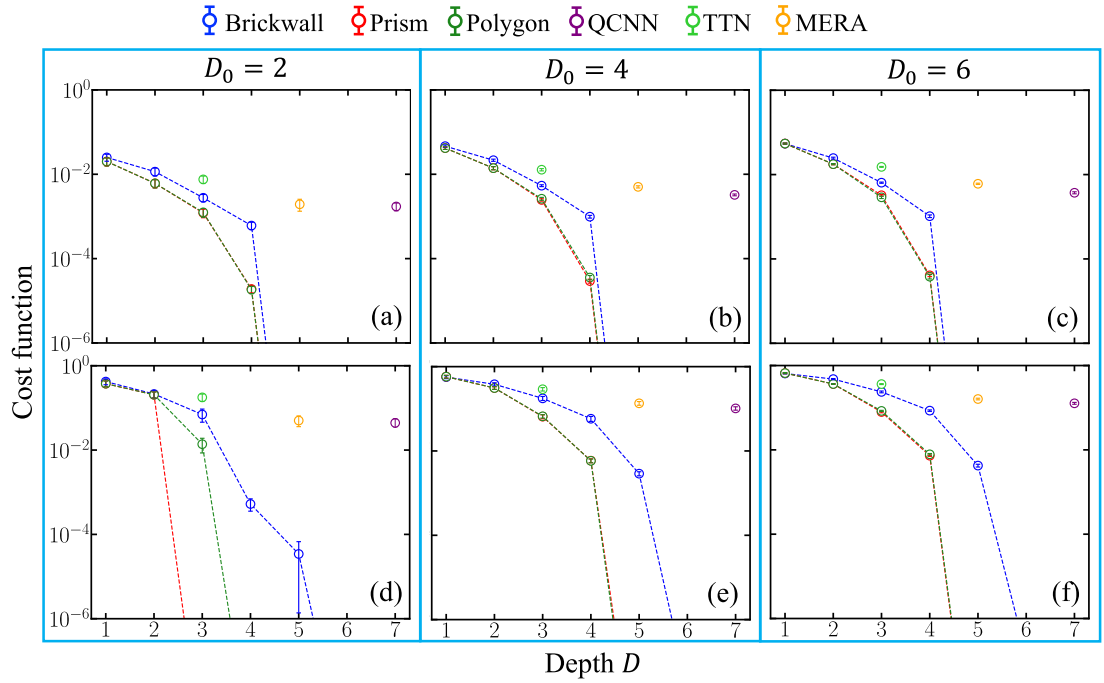

Figure B4. Cost functions of different VQC architectures in discrimination (top) and generation (bottom) of TI random states sampled from $\mathbb{S}(D_0)$ with $D_0 = 2, 4, 6$ in a system of $n = 6$ qubits.

where we utilized the Haar average identity for two elements of a $U(d)$ matrix,

$$\langle |\langle\psi_0|\psi_1\rangle|^2\rangle_{\mathbb{H}(D_0)} = \int_{\text{Haar}} \mathrm{d}U\, U_{\alpha a} U^*_{\beta b} = \frac{1}{d}\delta_{\alpha\beta}\delta_{ab}. \tag{B.3}$$

where d is its dimension.

B.2. $D_0 \gg 1$ limit

The exact $\langle P_{\mathrm{H}}(\psi_0, \psi_1)\rangle_{\mathbb{H}(D_0)}$ is not easy to compute for a finite D_0. Here we consider the case of $D_0 \gg 1$, where the ensemble $\mathbb{H}(D_0)$ is simply Haar random. In this case, the distribution of the overlap $x = |\langle\psi_0|\psi_1\rangle|^2$ can be analytically obtained [87].

Consider $d = 2^n$ complex numbers $\{\alpha_i\}_{i=1}^d$ as the complex amplitudes of the states ψ_1 with normalization condition $\sum_{i=1}^d |\alpha_i|^2 = 1$, it indicates that the amplitudes $\{\alpha_i\}_{i=1}^d$ forms 2D-sphere. The other state ψ_0 we can choose it to be $|\psi_0\rangle = |\mathbf{0}\rangle$ with the freedom in choosing the Haar unitary. The probability distribution for $x = |\langle\psi_0|\psi_1\rangle|^2 = |\langle\mathbf{0}|\psi_1\rangle|^2 = |\alpha_1|^2 = \gamma$ is

$$\begin{aligned} P(x=\gamma) &= \frac{\int_{-\infty}^{\infty}\prod_i \mathrm{d}^2\alpha_i\, \delta\left(\gamma - |\alpha_1|^2\right)\delta\left(1-\sum_i|\alpha_i|^2\right)}{\int_{-\infty}^{\infty}\prod_i \mathrm{d}^2\alpha_i\, \delta\left(1-\sum_i|a_i|^2\right)} = \frac{\int_{-\infty}^{\infty}\prod_{i>1} \mathrm{d}^2\alpha_i\, S_1\left(\sqrt{\gamma}\right)\delta\left(1-\gamma-\sum_{i>1}|\alpha_i|^2\right)}{2\sqrt{\gamma}S_{2d-1}(1)/2} \\ &= \frac{S_1\left(\sqrt{\gamma}\right)S_{2d-3}\left(\sqrt{1-\gamma}\right)}{2\sqrt{\gamma}\sqrt{1-\gamma}S_{2d-1}(1)} = (d-1)(1-\gamma)^{d-2}. \end{aligned} \tag{B.4}$$

Here we utilize the N-dimensional sphere surface area formula

$$S_N(R) \equiv \int_{-\infty}^{\infty}\prod_{l=1}^{N+1}\mathrm{d}x_l\,\delta\left(R-\sqrt{\sum_{l=1}^{N+1}x_l^2}\right) = 2R\int_{-\infty}^{\infty}\prod_{l=1}^{N+1}\mathrm{d}x_l\,\delta\left(R^2-\sum_{l=1}^{N+1}x_l^2\right) = \frac{2\pi^{\frac{N+1}{2}}}{\Gamma\left(\frac{N+1}{2}\right)}R^N, \tag{B.5}$$

where we use the expansion $2a\delta(x^2-a^2) = \delta(x-a)+\delta(x+a)$. One can also check that the probability $\int_0^1 \mathrm{d}\gamma\, P(x=\gamma) = 1$ is normalized. From the probability distribution, we have the Haar average Helstrom limit as

$$\langle P_H\rangle_{\text{Haar}} = \int_0^1 \mathrm{d}\gamma\,\frac{1}{2}\left(1-\sqrt{1-\gamma}\right)p(x=\gamma) = \frac{d-1}{2\left(d^2-3d+1\right)} = \frac{1}{2\left(2^{n+1}-1\right)}. \tag{B.6}$$

When $n \gg 1$, the Haar average Helstrom limit $\langle P_H\rangle_{\text{Haar}} \sim 1/2^{n+2}$.

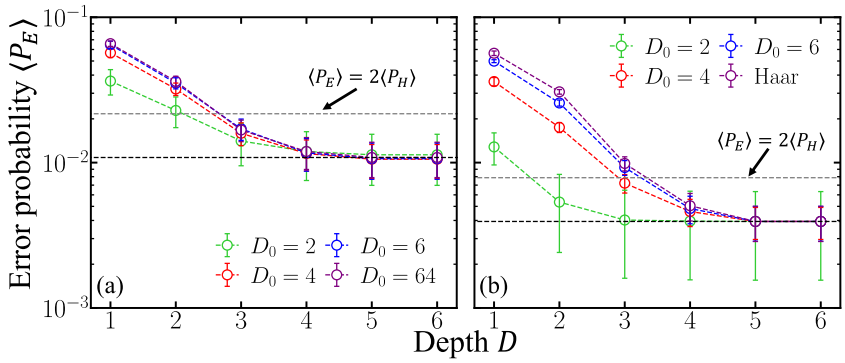

Figure D1. Error probability for discriminating between random states from $\mathbb{S}(D_0)$ (left) and $\mathbb{H}(D_0)$(right) using brickwall ansatz in an $n = 6$ system. Black and grey dashed lines in (a) and (b) show ensemble-averaged Helstrom limit $\langle P_H \rangle$ and $2\langle P_H \rangle$.

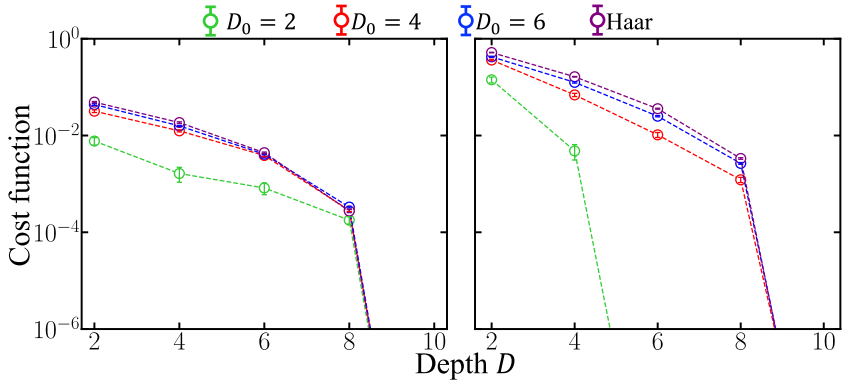

Figure D2. Cost functions of the sVQC ansatz in discrimination $\langle \mathcal{C}_{\text{dis}} \rangle$ (a) and generating $\langle \mathcal{C}_{\text{gen}} \rangle$ (b) of random states from $\mathbb{H}(D_0)$ with $D_0 = 2, 4, 6$ and Haar random states in a system of $n = 6$ qubits. In this case, the depth $D = D^*$ equals the layers of CNOT/CZ. The lines going below the plot region decrease to zero at the machine precision; we choose the plot range to make the trends clear.

Appendix C. Local random gates construction

As shown in figure 7, various architectures of VQC are constructed from local two-qubit gates. In general, a two-qubit gate, in the form of 4×4 unitary, includes 15 independent parameters (up to a global phase). Such a gate can be decomposed into single qubit rotations and up to three CNOT gate [88], as we show in figure B1.

Appendix D. Boundary condition for VQCs

In section 4, we assume the brickwall VQC ansatz to be open-boundary, which is a local quantum circuit. However, one can also assume the brickwall ansatz to have a periodic-boundary. In figure B2, we compare the performance and gradient of open-boundary and periodic-boundary brickwall ansatzs in discriminating between Haar random states. Just as we expect, the periodic-boundary brickwall ansatz performs slightly better than the open-boundary one in the low depth region, due to the extra gates on the first and last qubit that help process global quantum information, while the gradient is nearly the same which leads to the same difficulty of training. Therefore, without specific clarification we utilize open-boundary brickwall ansatz in the main text to restrict the locality of VQCs.

Appendix E. More results of VQCs performance

The VQC architectures utilized in section 5 are applied to the discrimination and generation of random states sampled from $\mathbb{H}(D_0)$ and $\mathbb{S}(D_0)$ with $D_0 \leqslant n$. With D_0 increases, the random states become more complex and entangled, as shown in figure 2(b). Similar to the Haar case, we can also confirm the advantage for extensive architectures towards non-extensive ones in small D_0 cases from figure B3. From

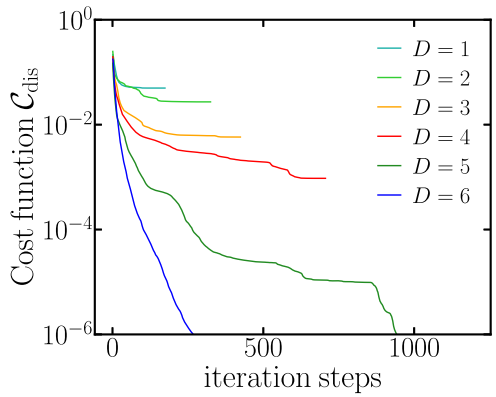

Figure F1. The optimization history of cost function $\mathcal{C}_{\rm dis}$ in discriminating between a pair of Haar random states. We apply the brickwall ansatz with different depth D in a system of $n = 6$ qubits.

figures B3(a)–(c), as states become more complex, for a fixed D the residual error also increases with extensive architectures, brickwall, prism and polygon. Compared to generation tasks shown in figures B3(d)–(f), discrimination still shows better performance within the depth range that $\mathcal{C}_{\rm gen}(U_D, |\psi\rangle)$ is non-zero.

Unlike the Haar case shown in figure 8, we can find that as D_0 decreases, the advantage of prism and polygon ansatzs over brickwall ansatz is reduced. Note that the information of input states is locally stored as D_0 is small, and thus the non-local gates in those architectures (which help to processes global information) could not perform as efficiently as in the large D_0 cases. The above discussions can also be extended to the performance of VQCs in TI states from $\mathbb{S}(D_0)$, as we show in figure B4.

In figure 6(a), we see that the discrimination residual error for random states from Haar ensemble and $\mathbb{S}(2^n)$ roughly agrees when D is low. We plot the discrimination error probability directly in figure D1. Compared to the typical random states, the Helstrom limit of TI random states is larger and independent of D_0, as we already confirmed in figure A1. Except for that, for all D_0, the error probability for TI states is still exponential suppressed with the depth D.

Finally, we provide the performance of the sVQC anstaz, in figure D2. Due to the limited number of CNOT/CZ gates and parameters, the depth required to approach the ultimate limit increases for the sVQC ansatz.

Appendix F. Computation details

To implement different VQC ansatzs, we use Qulacs [89], a high performance VQC simulator for Python. We utilize the BFGS algorithm [90], a gradient based quasi-Newton method for numerical optimization on parameters of VQCs. Compared to regular gradient descent training methods like stochastic gradient descent (SGD) and Adam used in classical neural networks, the VQC can be optimized in less iteration steps using the BFGS algorithm. We choose the best one among the 40 optimization results from random initializations of parameters. An example of optimization history is shown in figure F1.

We evaluate the gradient of cost functions with respect to each parameter using the central finite difference method. Another formalism, parameter-shift rule, is also proposed to provide an analytical expression for the gradient [91]. For a given observable O, the gradient of the mean of the observable is $g^i_{\rm PSR} \equiv \partial_i \langle O\rangle = \frac{1}{2}\left(\langle O\rangle^+ - \langle O\rangle^-\right)$, where $\langle O\rangle^\pm = \langle O\rangle_{\theta_i \pm \frac{\pi}{2}}$. Note that with the MLE strategy, the classical post-processing makes the observable dependent on the choice of each parameter separately and implicitly, for convenience and a fair comparison in all cases, we have adopted a finite-difference numerical approach in calculations of gradient. The step size Δs is also a free parameter in the definition of finite difference, and here we apply $\Delta s = 10^{-6}$.

ORCID iDs

Bingzhi Zhang https://orcid.org/0000-0002-5427-8185
Quntao Zhuang https://orcid.org/0000-0002-9554-3846

References

[1] Shor P W 1997 Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer *SIAM J. Comput.* **26** 1484
[2] Preskill J 2018 Quantum computing in the NISQ era and beyond *Quantum* **2** 79
[3] Farhi E, Goldstone J and Gutmann S 2014 A quantum approximate optimization algorithm (arXiv:1411.40282)
[4] Wecker D, Hastings M B and Troyer M 2015 Progress towards practical quantum variational algorithms *Phys. Rev.* A **92** 042303
[5] Chen M-C *et al* 2020 Demonstration of adiabatic variational quantum computing with a superconducting quantum coprocessor *Phys. Rev. Lett.* **125** 180501
[6] Romero J, Olson J P and Aspuru-Guzik A 2017 Quantum autoencoders for efficient compression of quantum data *Quantum Sci. Technol.* **2** 045001
[7] Gullans M J, Krastanov K, Huse D A, Jiang L and Flammia S T 2020 Quantum coding with low-depth random circuits (arXiv:2010.09775)
[8] 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
[9] 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
[10] McClean J R, Romero J, Babbush R and Aspuru-Guzik A 2016 The theory of variational hybrid quantum–classical algorithms *New J. Phys.* **18** 023023
[11] O'Malley P J J *et al* 2016 Scalable quantum simulation of molecular energies *Phys. Rev.* X **6** 031007
[12] Colless J I, Ramasesh V V, Dahlen D, Blok M S, Kimchi-Schwartz M E, McClean J R, Carter J, de Jong W A and Siddiqi I 2018 Computation of molecular spectra on a quantum processor with an error-resilient algorithm *Phys. Rev.* X **8** 011021
[13] Bravo-Prieto C, Lumbreras-Zarapico J, Tagliacozzo L and Latorre J I 2020 Scaling of variational quantum circuit depth for condensed matter systems *Quantum* **4** 272
[14] 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
[15] Carolan J *et al* 2020 Variational quantum unsampling on a quantum photonic processor *Nat. Phys.* **16** 322
[16] Benedetti M, Grant E, Wossnig L and Severini S 2019 Adversarial quantum circuit learning for pure state approximation *New J. Phys.* **21** 043023
[17] Patterson A, Chen H, Wossnig L, Severini S, Browne D and Rungger I 2021 Quantum state discrimination using noisy quantum neural networks *Phys. Rev. Res.* **3** 031063
[18] Chen H, Wossnig L, Severini S, Neven H and Mohseni M 2021 Universal discriminative quantum neural networks *Quantum Mach. Intell.* **3** 1
[19] Cong I, Choi S and Lukin M D 2019 Quantum convolutional neural networks *Nat. Phys.* **15** 1273
[20] MacCormack I, Delaney C, Galda A, Aggarwal N and Narang P 2020 Branching quantum convolutional neural networks (arXiv:2012.14439)
[21] Liu Y *et al* 2020 Variational quantum circuits for quantum state tomography *Phys. Rev.* A **101** 052316
[22] Zhuang Q and Zhang Z 2019 Physical-layer supervised learning assisted by an entangled sensor network *Phys. Rev.* X **9** 041023
[23] Xia Y, Li W, Zhuang Q and Zhang Z 2021 Quantum-enhanced data classification with a variational entangled sensor network *Phys. Rev.* X **11** 021047
[24] Lubasch M, Joo J, Moinier P, Kiffner M and Jaksch D 2020 Variational quantum algorithms for nonlinear problems *Phys. Rev.* A **101** 010301
[25] Li Y and Benjamin S C 2017 Efficient variational quantum simulator incorporating active error minimization *Phys. Rev.* X **7** 021050
[26] Dumitrescu E F, McCaskey A J, Hagen G, Jansen G R, Morris T D, Papenbrock T, Pooser R C, Dean D J and Lougovski P 2018 Cloud quantum computing of an atomic nucleus *Phys. Rev. Lett.* **120** 210501
[27] McArdle S, Jones T, Endo S, Li Y, Benjamin S C and Yuan X 2019 Variational ansatz-based quantum simulation of imaginary time evolution *npj Quantum Inf.* **5** 75
[28] Schuld M, Sinayskiy I and Petruccione F 2015 An introduction to quantum machine learning *Contemp. Phys.* **56** 172
[29] Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N and Lloyd S 2017 Quantum machine learning *Nature* **549** 195
[30] Dunjko V and Briegel H J 2018 Machine learning & artificial intelligence in the quantum domain: a review of recent progress *Rep. Prog. Phys.* **81** 074001
[31] Rebentrost P, Bromley T R, Weedbrook C and Lloyd S 2018 Quantum Hopfield neural network *Phys. Rev.* A **98** 042308
[32] Killoran N, Bromley T R, Arrazola J M, Schuld M, Quesada N and Lloyd S 2019 Continuous-variable quantum neural networks *Phys. Rev. Res.* **1** 033063
[33] Havlíček V, Córcoles A D, Temme K, Harrow A W, Kandala A, Chow J M and Gambetta J M 2019 Supervised learning with quantum-enhanced feature spaces *Nature* **567** 209
[34] Schuld M and Killoran N 2019 Quantum machine learning in feature Hilbert spaces *Phys. Rev. Lett.* **122** 040504
[35] Du Y, Hsieh M-H, Liu T and Tao D 2020 Expressive power of parametrized quantum circuits *Phys. Rev. Res.* **2** 033125
[36] Yang C, Garner A, Liu F, Tischler N, Thompson J, Yung M-H, Gu M and Dahlsten O 2021 Provable superior accuracy in machine learned quantum models (arXiv:2105.14434)
[37] Eldan R and Shamir O 2016 The power of depth for feedforward neural networks *Conf. Learning Theory (PMLR)* pp 907–40
[38] Telgarsky M 2016 Benefits of depth in neural networks *Conf. Learning Theory (PMLR)* pp 1517–39
[39] Rolnick D and Tegmark M 2017 The power of deeper networks for expressing natural functions (arXiv:1705.05502)
[40] Lu Z, Pu H, Wang F, Hu Z and Wang L 2017 The expressive power of neural networks: a view from the width (arXiv:1709.02540)
[41] Abbas A, Sutter D, Zoufal C, Lucchi A, Figalli A and Woerner S 2021 The power of quantum neural networks *Nat. Comput. Sci.* **1** 403–9
[42] 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
[43] Cerezo M, Sone A, Volkoff T, Cincio L and Coles P J 2021 Cost function dependent barren plateaus in shallow parametrized quantum circuits *Nat. Commun.* **12** 1791

[44] Holmes Z, Sharma K, Cerezo M and Coles P J 2022 Connecting ansatz expressibility to gradient magnitudes and barren plateaus *PRX Quantum* **3** 010313
[45] Helstrom C 1976 *Quantum Detection and Estimation Theory* (New York: Academic)
[46] Helstrom C W 1967 Minimum mean-squared error of estimates in quantum statistics *Phys. Lett.* A **25** 101
[47] Helstrom C 1976 *Quantum Detection and Estimation Theory* (*Mathematics in Science and Engineering: A Series of Monographs and Textbooks*) (New York: Academic)
[48] Schuld M, Bocharov A, Svore K M and Wiebe N 2020 Circuit-centric quantum classifiers *Phys. Rev.* A **101** 032308
[49] Grant E, Benedetti M, Cao S, Hallam A, Lockhart J, Stojevic V, Green A G and Severini S 2018 Hierarchical quantum classifiers *npj Quantum Inf.* **4** 65
[50] Pesah A, Cerozo M, Wang S, Volkoff T, Sornborger A T and Coles P J 2021 Absence of barren plateaus in quantum convolutional neural networks *Phys. Rev.* X **11** 041011
[51] Sam 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
[52] Wu Y, Zhang P and Zhai H 2021 Scrambling ability of quantum neural network architectures *Phys. Rev. Res.* **3** L032057
[53] Nahum A, Ruhman J, Vijay S and Haah J 2017 Quantum entanglement growth under random unitary dynamics *Phys. Rev.* X **7** 031016
[54] Nahum A, Vijay S and Haah J 2018 Operator spreading in random unitary circuits *Phys. Rev.* X **8** 021014
[55] Page D N 1993 Average entropy of a subsystem *Phys. Rev. Lett.* **71** 1291
[56] Zhuang Q, Schuster T, Yoshida B and Yao N Y 2019 Scrambling and complexity in phase space *Phys. Rev.* A **99** 062334
[57] Zhang B and Zhuang Q 2021 Entanglement formation in continuous-variable random quantum networks *npj Quantum Inf.* **7** 33
[58] Datta N, Mosonyi M, Hsieh M-H and Brandao F G S L 2013 A smooth entropy approach to quantum hypothesis testing and the classical capacity of quantum channels *IEEE Trans. Inf. Theory* **59** 8014
[59] Wang L and Renner R 2012 One-shot classical–quantum capacity and hypothesis testing *Phys. Rev. Lett.* **108** 200501
[60] Anshu A, Devabathini V K and Jain R 2017 Quantum communication using coherent rejection sampling *Phys. Rev. Lett.* **119** 120506
[61] Anshu A, Jain R and Warsi N A 2018 Building blocks for communication over noisy quantum networks *IEEE Trans. Inf. Theory* **65** 1287
[62] Hayashi M and Nagaoka H 2003 General formulas for capacity of classical–quantum channels *IEEE Trans. Inf. Theory* **49** 1753
[63] Zhuang Q, Zhang Z and Shapiro J H 2018 Distributed quantum sensing using continuous-variable multipartite entanglement *Phys. Rev.* A **97** 032329
[64] Ge W, Jacobs K, Eldredge Z, Gorshkov A V and Foss-Feig M 2018 Distributed quantum metrology with linear networks and separable input *Phys. Rev. Lett.* **121** 043604
[65] Tan S-H, Erkmen B I, Giovannetti V, Guha S, Lloyd S, Maccone L, Pirandola S and Shapiro J H 2008 Quantum illumination with Gaussian states *Phys. Rev. Lett.* **101** 253601
[66] Zhuang Q 2021 Quantum ranging with Gaussian entanglement *Phys. Rev. Lett.* **126** 240501
[67] Pirandola S 2011 Quantum reading of a classical digital memory *Phys. Rev. Lett.* **106** 090504
[68] Shi H, Zhang Z, Pirandola S and Zhuang Q 2020 Entanglement-assisted absorption spectroscopy *Phys. Rev. Lett.* **125** 180502
[69] Eisert J 2021 Entangling power and quantum circuit complexity (arXiv:2104.03332)
[70] Haferkamp J, Faist P, Kothakonda N B T, Eisert J and Halpern N Y 2022 Linear growth of quantum circuit complexity *Nat. Phys.* **18** 528–32
[71] Gross D, Audenaert K and Eisert J 2007 Evenly distributed unitaries: on the structure of unitary designs *J. Math. Phys.* **48** 052104
[72] Ambainis A and Emerson J 2007 Quantum t-designs: t-wise independence in the quantum world *22nd Annual IEEE Conf. Computational Complexity (CCC'07)* pp 129–40
[73] Roberts D A and Yoshida B 2017 Chaos and complexity by design *J. High Energy Phys.* JHEP04(2017)121
[74] Brandão F G S L, Harrow A W and Horodecki M 2016 Local random quantum circuits are approximate polynomial-designs *Commun. Math. Phys.* **346** 397
[75] Holzhey C, Larsen F and Wilczek F 1994 Geometric and renormalized entropy in conformal field theory *Nucl. Phys.* B **424** 443
[76] Latorre J I, Rico E and Vidal G 2003 Ground state entanglement in quantum spin chains (arXiv:quant-ph/0304098)
[77] Calabrese P and Cardy J 2004 Entanglement entropy and quantum field theory *J. Stat. Mech.* P06002
[78] Kennedy R S 1973 On the optimum receiver for the m-ary linearly independent pure state problem *Q. Prog. Rep.* **110** 142
[79] Eldar Y C, Megretski A and Verghese G C 2003 Designing optimal quantum detectors via semidefinite programming *IEEE Trans. Inf. Theory* **49** 1007
[80] Volkoff T and Coles P J 2021 Large gradients via correlation in random parameterized quantum circuits (arXiv:2005.12200)
[81] 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)
[82] Arrasmith A, Cerezo M, Czarnik P, Cincio L and Coles P J 2020 Effect of barren plateaus on gradient-free optimization (arXiv:2011.12245)
[83] Cerezo M and Coles P J 2021 Higher order derivatives of quantum neural networks with barren plateaus *Quantum Sci. Technol.* **6** 035006
[84] Arrasmith A, Holmes Z, Cerezo M and Coles P J 2021 Equivalence of quantum barren plateaus to cost concentration and narrow gorges (arXiv:2104.05868)
[85] Holmes Z, Arrasmith A, Yan B, Coles P J, Albrecht A and Sornborger A T 2021 Barren plateaus preclude learning scramblers *Phys. Rev. Lett.* **126** 190501
[86] Kennedy R S 1973 A near-optimum receiver for the binary coherent state quantum channel *Q. Prog. Rep.* **108** 219
[87] Zhuang Q and Wu B 2013 Equilibration of quantum chaotic systems *Phys. Rev.* E **88** 062147
[88] Vatan F and Williams C 2004 Optimal quantum circuits for general two-qubit gates *Phys. Rev.* A **69** 032315
[89] Suzuki Y *et al* 2020 Qulacs: a fast and versatile quantum circuit simulator for research purpose (arXiv:2011.13524)
[90] Virtanen P *et al* 2020 SciPy 1.0: fundamental algorithms for scientific computing in python *Nat. Methods* **17** 261
[91] Mitarai K, Negoro M, Kitagawa M and Fujii K 2018 Quantum circuit learning *Phys. Rev.* A **98** 032309