

<https://doi.org/10.1038/s42005-025-02261-4>

# Variational optimization for quantum problems using deep generative networks

Check for updates

Lingxia Zhang <sup>1,2,7</sup>, Xiaodie Lin <sup>3,4,7</sup>, Peidong Wang<sup>1,2</sup>, Kaiyan Yang<sup>1,2</sup>, Xiao Zeng<sup>1,2</sup>, Zhaohui Wei <sup>5,6</sup> & Zizhu Wang <sup>1,2</sup>

Optimization drives advances in quantum science and machine learning, yet most generative models aim to mimic data rather than to discover optimal answers to challenging problems. Here we present a variational generative optimization network that learns to map simple random inputs into high quality solutions across a variety of quantum tasks. We demonstrate that the network rapidly identifies entangled states exhibiting an optimal advantage in entanglement detection when allowing classical communication, attains the ground state energy of an eighteen spin model without encountering the barren plateau phenomenon that hampers standard hybrid algorithms, and—after a single training run—outputs multiple orthogonal ground states of degenerate quantum models. Because the method is model agnostic, parallelizable and runs on current classical hardware, it can accelerate future variational optimization problems in quantum information, quantum computing and beyond.

Mathematical optimization is ubiquitous in modern science and technology. Spanning diverse fields like economics, chemistry, physics, and various engineering areas, its applications abound<sup>1</sup>. In quantum information theory, many problems relating to the approximation and characterization of quantum correlations can be formulated as convex optimization problems<sup>2–4</sup>, which is a particular kind of mathematical optimization with provable global optimality guarantees. For quantum problems where convexity is hard to come by or the global optimality of the solution is a secondary consideration when compared to the efficiency of the algorithm, variational optimization provides a rich toolbox. When solutions are expected to be quantum, hybrid quantum-classical variational algorithms are popular choices. In these algorithms, variational optimization is carried out on the classical parameters, while quantum gates and measurements are implemented in the corresponding quantum circuit<sup>5–10</sup>.

Optimization is also at the core of every machine learning algorithm<sup>11</sup>. Recently, machine learning algorithms have opened a new way to address scientific problems spanning a broad spectrum, accelerating the integration of AI into the scientific discovery process<sup>12,13</sup>. In mathematics, they help humans discover new results<sup>14</sup> and develop faster solutions to problems<sup>15</sup>. In biology, they help with drug developments<sup>16</sup>. Particularly, generative models have seen explosive growth in the form of large language models<sup>17,18</sup>, which are transforming the way humans interact with machines. Applying these models to science has enabled new solutions to mathematical problems to be

discovered<sup>19</sup>. Meanwhile, generative models have also been widely applied to quantum physics. For example, many-body quantum models can be efficiently solved by restricted Boltzmann machines<sup>20,21</sup>, lattice gauge theories can be simulated using normalizing flows<sup>22,23</sup>, quantum states can be more efficiently represented by variational autoencoders (VAEs)<sup>24–27</sup>, and quantum circuits with desired properties can be generated by the generative pre-trained transformer<sup>28</sup>.

However, despite these encouraging advances, current applications of generative models to quantum problems usually focus on learning certain features from training data sets, and then generating new data with similar features. In the scenario where a classical (i.e., not quantum) generative model is used to solve a quantum problem, the training data may be quantum states or complex correlation information contained therein, and a neural network is expected to generate new quantum states or information resembling the training set.

In order to extend the possibility of applying generative models to quantum problems beyond this scenario, inspired by the classical VAE, we propose a method called the variational generative optimization network (VGON), whose output does not just resemble the input, but can be (nearly) optimal solutions to general variational optimization problems. VGON contains a pair of deep feed-forward neural networks connected by a stochastic latent layer, and a problem-specific objective function. The intrinsic randomness in the model can be leveraged both in

<sup>1</sup>Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, China. <sup>2</sup>Key Laboratory of Quantum Physics and Photonic Quantum Information, Ministry of Education, University of Electronic Science and Technology of China, Chengdu, China. <sup>3</sup>Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China. <sup>4</sup>College of Computer and Data Science, Fuzhou University, Fuzhou, China. <sup>5</sup>Yau Mathematical Sciences Center, Tsinghua University, Beijing, China. <sup>6</sup>Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Beijing, China. <sup>7</sup>These authors contributed equally: Lingxia Zhang, Xiaodie Lin. e-mail: [weizhaohui@mail.tsinghua.edu.cn](mailto:weizhaohui@mail.tsinghua.edu.cn); [zizhu@uestc.edu.cn](mailto:zizhu@uestc.edu.cn)

its training and testing stages. During the training stage, we have not encountered any issues with the optimization getting trapped in local minima. We believe this can be partially explained by having random inputs, which effectively gives the optimization multiple starting points, and the architecture of our model, especially the existence of the latent layer, which regularizes the input and leads to good trainability. In the testing stage, the randomness allows VGON to produce multiple optimal solutions to the objective functions simultaneously, even after only a single stage of training.

We apply VGON to a variety of quantum problems to showcase its potential. We first demonstrate that it outperforms stochastic gradient descent (SGD) by avoiding entrapment in local optima in variational optimization problems of modest size, while also converging orders of magnitude faster. For larger problems with tens of thousands of parameters, we show that VGON can substantially alleviate the problem of barren plateaux in parameterized quantum circuits. Since generative models allow multiple optimal solutions to be found and generated simultaneously, a capability that deterministic algorithms lack, we use VGON to explore the ground state space of two quantum many-body models known to be degenerate. We show that VGON can successfully identify the dimensionality of the ground state space and generate a variety of orthogonal or linearly independent ground states spanning the entire space.

## Results

### The VGON model

The architecture of VGON, shown in Fig. 1, consists of two deep feed-forward neural networks, the encoder  $E_\omega$  and the decoder  $D_\phi$  are connected via a latent layer  $\mathcal{Z}$  containing a normal distribution  $\mathcal{N}(\mu(z), \sigma^2(z))$ , where the mean  $\mu$  and the standard deviation  $\sigma$  are provided by  $E_\omega$ . During the training stage, input data  $x_0$  is sampled from a distribution  $P(x_0)$ , which in all our tests is the uniform distribution over the parameter space. It is then mapped to the latent distribution  $\mathcal{N}(\mu(z), \sigma^2(z))$  by the encoder network  $E_\omega$ . Next the decoder network  $D_\phi$  maps data  $z$  sampled from the latent distribution  $\mathcal{N}(\mu(z), \sigma^2(z))$  to a distribution minimizing the objective function  $h(x)$ . This minimization is achieved by iteratively updating the parameters  $\omega$  and  $\phi$  in  $E_\omega$  and  $D_\phi$ , respectively. Due to the existence of a stochastic latent layer, the gradients cannot be propagated backwards in the network. We solve this issue by using the reparameterization trick<sup>24</sup>.

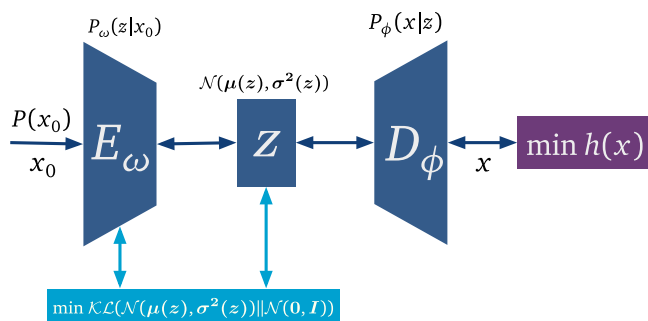
The key difference between VGON and VAE lies in the objective function (also called the loss function in the machine learning literature): instead of asking the output data distribution to approximate the input distribution by maximizing a given similarity measure, VGON simply

requires the output data to minimize any objective function that is appropriate for the target problem. In addition, the Kullback-Leibler (KL) divergence between the latent distribution  $\mathcal{N}(\mu(z), \sigma^2(z))$  and the standard normal distribution  $\mathcal{N}(\mathbf{0}, \mathbf{I})$  is also minimized during training, as part of the objective function.

After the objective function converges to within a given tolerance, the training stage is complete. To utilize the trained model, the encoder network is disabled, and data sampled from a standard normal distribution  $\mathcal{N}(\mathbf{0}, \mathbf{I})$  are fed into the decoder network. Depending on the characteristics of the objective function, the corresponding output distribution can be tightly centered around one or multiple optimum values. Moreover, it is worth mentioning that requiring the latent layer to follow a normal distribution not only facilitates efficient optimization of the objective function but also simplifies the sampling process, since the KL divergence between two normal distributions can be analytically evaluated and sampling from a normal distribution is computationally efficient.

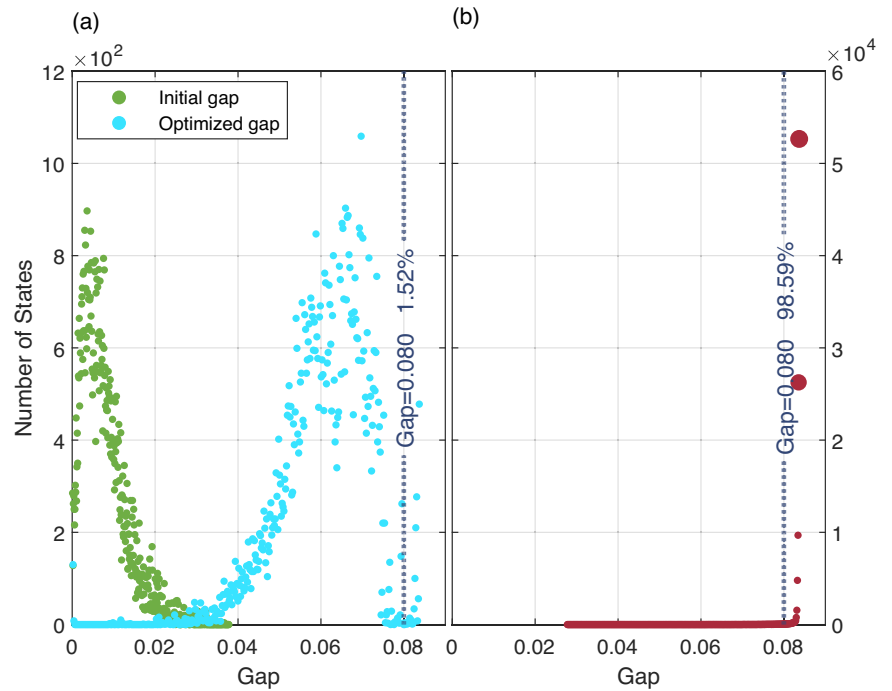
The goal of VGON can be seen as finding a way which maps a simple distribution defined over the latent space, i.e., the Gaussian, to a complicated one, i.e., a distribution whose samples can minimize the objective function with high probability<sup>29</sup>. This shares the spirit of transport theory, where given two probability measures  $\mu$  and  $\nu$  on spaces  $X$  and  $Y$ , we call a map  $T: X \rightarrow Y$  a transport map if  $T_*(\mu) = \nu$ , where  $T_*(\mu)$  is the pushforward of  $\mu$  by  $T$ , representing the process of transferring (or “pushing forward”) the measure  $\mu$  from  $X$  to  $Y$  via the measurable function  $T$ . In an optimal transport problem<sup>30</sup>, one is interested in finding a map  $T$  that minimizes the transport cost  $\int c(x, T_*(x))d\mu(x)$ , subject to the constraint that the pushforward measure satisfies  $T_*(\mu) = \nu$ <sup>31</sup>. In addition to optimal transport problems, there are problems that do not have an explicitly defined target distribution, where the task is evaluating the loss function directly on the generated samples. In these problems, an optimal  $T$  can either be found analytically, such as in inverse transform sampling, where both distributions are one-dimensional<sup>32</sup>, or  $T$  can be learned/optimized from a parameterized  $T_\theta$  on training data, such as generative model like normalizing flows<sup>22,33</sup>, where both distributions are high-dimensional but  $T$  is invertible and constructed using neural networks. In VGON, the latent space usually has a much lower dimension than the output layer, making  $T$  surjective, which means every point in the target space can be reached by applying the decoder to some latent input<sup>34</sup>. This surjectivity relaxes the requirement for invertibility and enables VGON to easily cover the complex target distributions. In our experience, the best optimization results come from training the encoder-latent space-decoder triple as a whole, even though it is possible to achieve good results without including the encoder in the training process.

To show that VGON can work well, we first use it to solve a variational optimization problem with a known unique optimal solution: finding the minimum ground state energy density among a class of quantum many-body models that matches the lower bound certified by an SDP relaxation<sup>35</sup>. More specifically, we consider a class of infinite 1D translation-invariant (TI) models with fixed couplings<sup>36</sup>, and the optimization variables are the local observables. The ground state energy density of this class of models has a lower bound certified by a variant of the NPA hierarchy<sup>36</sup>. However, there is no guarantee that any infinite TI quantum many-body Hamiltonian, if couplings are fixed but the local observables can be arbitrary, can achieve this bound. Meanwhile, by optimizing 3-dimensional local observables with SGD and computing the ground state with uniform matrix product state algorithms, a Hamiltonian whose ground state energy density matches the above lower bound to seven significant digits has been found. We replace SGD with VGON to conduct the same optimization, and find that the converged model can (almost) deterministically generate Hamiltonians whose ground state energy density matches the NPA lower bound to eight significant digits, reaching the precision limit of commercial SDP solvers (see Supplementary Note 1 for more technical details). Below we apply VGON to several more complicated problems.



**Fig. 1 | The framework of variational generative optimization network.** The network is composed of an encoder network  $E_\omega$ , a latent space  $\mathcal{Z}$ , and a decoder network  $D_\phi$ . Training data  $x_0$  sampled from  $P(x_0)$  is first mapped into a latent distribution  $\mathcal{N}(\mu(z), \sigma^2(z))$  by  $E_\omega(x_0)$ . Then a latent variable  $z$  sampled from  $\mathcal{N}(\mu(z), \sigma^2(z))$  is transformed to the output  $x$  by  $D_\phi(z)$ . The parameters  $\phi$  and  $\omega$  are updated iteratively to minimize the objective function  $h(x)$ , together with the Kullback-Leibler (KL) divergence between the latent distribution  $\mathcal{N}(\mu(z), \sigma^2(z))$  and the standard normal distribution  $\mathcal{N}(\mathbf{0}, \mathbf{I})$ .

**Fig. 2 | Comparisons of stochastic gradient descent (SGD) and variational generative optimization network (VGON) in generating states with large gaps. a** Most of the 79,663 random initial states for SGD exhibit small gaps around 0.0036, while after optimization 1.52% of states have gaps larger than 0.08, which is indicated by the dashed line. The largest gap is 0.083722. **b** Over 98.59% of the 100,000 states generated by a trained VGON model have gaps larger than 0.08, which is presented by the dashed line. In particular, over 50% of these states are tightly centered around 0.0837.



### Finding the optimal state for entanglement detection

Entanglement detection plays a central role in quantum tasks such as secure communication and distributed computing, where entanglement serves as a fundamental resource. Suppose two players, Alice and Bob, receive a bipartite quantum state  $\rho$  from a source, then they want to determine whether  $\rho$  is entangled, with the smallest statistical error.

They can either perform the experiment independently in their respective laboratories and subsequently communicate the outcomes from Alice to Bob, or choose to forgo communication entirely. In the first scenario they are implementing a local operations and one-way classical communication (1-LOCC) protocol while in the second scenario they are implementing a local operations (LO) one. Implementing the 1-LOCC protocol experimentally requires fast real-time switching of Bob’s measurement settings and a quantum memory to store Bob’s half of the state while Alice performs her measurement and communicates the result. Do these extra experimental complexities yield tangible advantages such as reduced statistical error probabilities? In fact, it has been shown that for some simple states, such an advantage does exist, but it is too small to be useful<sup>37</sup>. In fact, the advantage is highly dependent on the choice of target states and is hard to estimate analytically. The goal of this task is to identify high-dimensional quantum states for which this advantage, defined as the gap between the minimum statistical error probabilities in LO and 1-LOCC protocols, is as large as possible. This gap quantifies the practical advantage of allowing one-way communication between the parties in entanglement detection.

Specifically, given a quantum state  $\rho$ , the advantage is defined to be the gap between the minimum probabilities  $p_2$  of committing false-negative errors (a.k.a. type-II errors, defined as a source distributes an entangled state, but Alice and Bob conclude the state they received is separable) when Alice and Bob employ LO and 1-LOCC protocols, and the two protocols have the same probability of making false-positive errors, i.e., type-I error, denoted  $p_1$  and defined as the scenario in which they conclude that they have received an entangled state, while the source actually distributes a separable one. The desired quantity can be computed by solving two SDPs that share the same objective but differ in constraint structure. Specifically, the following SDP is solved twice, once under LO protocol  $P \in \mathcal{P}^{LO}$  and once under 1-LOCC protocol  $P \in \mathcal{P}^{1-LOCC}$ . The final result is obtained by subtracting the

respective optimal values of  $p_2$ <sup>38</sup>:

$$\begin{aligned} \min_P \quad & p_2 \\ \text{s.t.} \quad & \text{tr}(M_N(P)\rho) = p_2, \\ & p_1 \mathbb{I} - M_Y(P) \in \mathcal{S}^*, \\ & P \in \mathcal{P}^{(LO,1-LOCC)}. \end{aligned} \tag{1}$$

Here  $\mathcal{S}^*$  denotes the dual of the separable set  $\mathcal{S}$ . The positive operator-valued measure (POVM) operators  $M_{Y(N)}(P)$  can be constructed as  $M_{Y(N)}(P) = \sum_{x,y,a,b} P(x,y,Y(N)|a,b) A_x^a \otimes B_y^b$ , where  $\{A_x^a\}$  ( $\{B_y^b\}$ ) are predetermined measurements performed by Alice (Bob) with  $x$  ( $y$ ) being measurement labels and  $a$  ( $b$ ) being outcomes, and  $P$  is the shorthand for the distribution  $P(x,y,Y(N)|a,b)$ , which specifies the detection strategy by assigning probabilities to particular combinations of settings, outcomes, and decisions. Here  $Y$  and  $N$  denote the decisions corresponding to the presence or absence of entanglement, respectively.

For a random quantum state  $\rho$ , it turns out that the gap calculated above is usually very small, as shown by the green dots in Fig. 2a. In order to observe the gap under noisy experimental conditions, we focus on a linear optical setup that generates bipartite qutrit states, which also allows us to parameterize the state space. We first employ SGD to maximize the gap by starting from random pure bipartite qutrit states. The results are shown in Fig. 2a. The SGD algorithm gets trapped easily in local maxima and needs to compute the gradient by solving dozens of SDPs. Optimizing the gap for 79,663 random pure states is computationally very costly (see Supplementary Note 2.1 for more details). Most of these states exhibit gaps around 0.0036 before optimization, while the largest gap afterwards is 0.083722.

The results of using VGON to maximize the gap are depicted in Fig. 2b and summarized in Table 1. Based on 3000 sets of initial parameters produced by uniform sampling, the model converges after less than 2 h of training. After that, we use it to generate 10,000 output states. We find that over 98% of them manifest gaps over 0.08, while over 50% of them have gaps larger than 0.0835. A similar performance has been observed even when choosing the initial quantum states from a variational submanifold of the space of all mixed states, where out of 10,000 states generated by the converged VGON model, 83 have gaps larger than 0.0837, with an average purity of 0.99999. For comparison, we also apply seven other global

optimization algorithms and a Multilayer Perceptron (MLP)—a neural network with multiple fully connected layers—to this task, and find that VGON consistently outperforms all the baseline methods (see Table 2 and Supplementary Note 2 for more details). Importantly, using VGON to solve an experiment-relevant instance of this problem allows us to experimentally demonstrate the advantage of 1-LOCC protocols over LO ones in detecting entanglement, where we can observe an error probability that is impossible to achieve with local operations alone<sup>38</sup>.

### Alleviating the effect of barren plateaux in variational quantum algorithms

On problems with a moderate size of optimization variables, VGON has shown its ability to quickly converge to the (nearly) optimal output distribution and generate high quality solutions with high probability. In near-term hybrid quantum-classical algorithms such as the variational quantum eigensolver (VQE)<sup>3</sup>, however, the number of classical parameters can quickly reach thousands or tens of thousands. The performance of such a hybrid algorithm can be hard to predict. On the classical part, the problems of vanishing gradients and having multiple minima are often present<sup>39–43</sup>. On the quantum part, the choice of ansätze greatly affects the expressivity of the quantum circuit, making the certification of global optimality difficult<sup>44–48</sup>.

For example, in a typical VQE algorithm, a parameterized variational circuit  $U(\theta)$  is used to approximately generate the ground state of a target

Hamiltonian  $H$ . The circuit structure usually loosely follows the target Hamiltonian and is often called an ansatz. Then by setting the energy of the output state  $|\psi(\theta)\rangle = U(\theta)|00 \cdots 0\rangle$  with respect to  $H$ , i.e.,  $\langle\psi(\theta)|H|\psi(\theta)\rangle$ , as the objective function, the algorithm iteratively updates the parameters in the quantum circuit by applying gradient-based methods on a classical computer. When the algorithm converges, the output quantum state will likely be very close to the ground state of  $H$ .

However, when the size of quantum systems increases, gradients vanish exponentially. This is primarily because the random initializations of parameterized unitaries conform to the statistics of a unitary 2-design<sup>41,49</sup>, making the working of gradient-based optimization difficult. To overcome the BP problem, several strategies have been proposed and investigated, with the small-angle initialization (VQE-SA) method being identified as an effective technique<sup>50–52</sup>. It initializes parameters  $\theta$  to be close to the zero vector, which differs from the statistics of the parameters from a 2-design and thus may alleviate the BP problem.

The advantage of VGON over VQE-SA in alleviating BPs can be seen when we use them both, with the same parameterized quantum circuit, to compute the ground state energy of the Heisenberg XXZ model. Its Hamiltonian with periodic boundary conditions is given by:

$$H_{\text{XXZ}} = - \sum_{i=1}^N \left( \sigma_x^i \sigma_x^{i+1} + \sigma_y^i \sigma_y^{i+1} - \sigma_z^i \sigma_z^{i+1} \right),$$

where  $\sigma_{x,y,z}^i$  denote the Pauli operators at site  $i$ . The ansatz for the parameterized quantum circuit is inspired by the matrix product state encoding<sup>53</sup>. It consists of sequential blocks of nearest-neighbor unitary gates, each of which is made of 11 layers of single qubit rotations and CNOT gates (see Supplementary Note 3.2 for more details).

By choosing  $N = 18$ , the circuit contains 816 blocks and 12,240 variational parameters. The average ground state energy, computed using exact diagonalization (ED), is  $-1.7828$ . We use both VQE-SA and VGON to compute the same quantity, with each method repeated 10 times. The results are shown in Table 3 and Fig. 3, where the mean values and the 95% confidence intervals of these methods are visualized. The dark-blue and the green lines represent the average energy for VQE-SA and VGON, whose mean values at the last iteration are  $-1.7613$  and  $-1.7802$ , respectively. Furthermore, to compare the performance of the two methods in a more fine-grained manner, we also calculate the fidelity between the states produced by the quantum circuit and the exact ground state. As the purple line depicted, VGON can achieve a 99% fidelity by around 880 iterations, while the VQE-SA method can only achieve 78.25% fidelity within the same number of iterations. We would like to remark that since the quantum computational resource consumed in VGON is similar to that in VQE-SA and classical computational resource is relatively cheap, the wall clock time cost per iteration for VGON can be comparable to that of VQE-SA. Moreover, the batch training of VGON can lead to a more stable convergence. For practical on-hardware quantum optimization, batch evaluation also enhances noise robustness by averaging over multiple circuit executions, thereby mitigating stochastic fluctuations, suppressing outliers, and accelerating convergence. For another comparison, VQE with uniformly random initial parameters can barely provide meaningful results due to the presence of barren plateaux, where the mean value of the average energy is  $-0.1367$  after 1000 iterations across ten repetitions, as illustrated by the light-blue line in Fig. 3.

**Table 1 | Performance comparison between variational generative optimization network (VGON) and stochastic gradient descent (SGD)**

Method	Percentage of states with gap		
	$\geq 0.08$	$\geq 0.08355$	$0.0837 \pm 5 \times 10^{-5}$
SGD with pure states	1.52%	0.57%	0.43%
VGON with pure states	98.59%	52.75%	9.38%
VGON with mixed states	99.32%	57.36%	22.19%

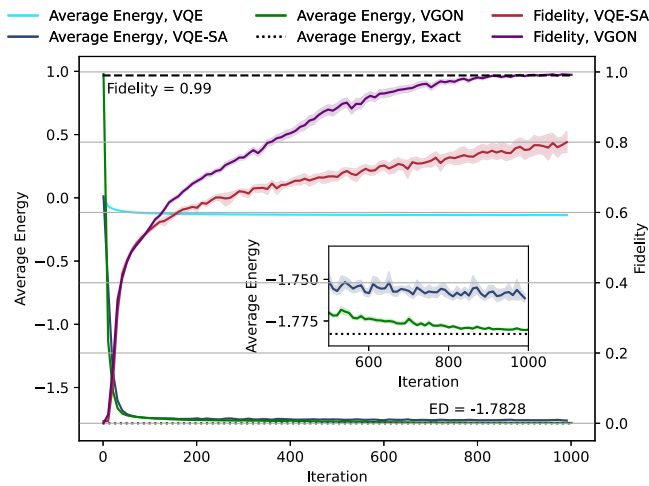
The comparison is carried out on finding states which maximize the advantage of one-way local operations and classical communication (1-LOCC) protocols over local operations (LO) ones.

**Table 2 | Comparisons of variational generative optimization network (VGON) with seven global optimization algorithms and multilayer perceptron (MLP)**

Algorithm	Maximal optimized gap	Run time
GlobalSearch	1.2e−07	24 h
Multistart	0.0726	24 h
Genetic Algorithm	0.0779	24 h
Particle Swarm Optimization	0.0348	24 h
Simulated Annealing	0.0067	24 h
Pattern Search	0.0717	24 h
Surrogate optimization	0.0412	24 h
MLP	0.0384(mean)	3000 iterations
VGON	0.0837	3000 iterations

**Table 3 | Comparison between variational quantum eigensolver (VQE), VQE with small-angle initialization (VQE-SA), and variational generative optimization network (VGON)**

–	Optimal			Mean		
	VQE	VQE-SA	VGON	VQE	VQE-SA	VGON
Average energy	−0.1374	−1.7684	−1.7821	−0.01367	−1.7613	−1.7802
Fidelity	–	90.00%	99.82%	–	80.01%	99.17%



**Fig. 3 | Mean values and 95% confidence intervals of the energy densities and the fidelity to the exact ground state at different iterations.** The light-blue line shows the average energy at different iterations for variational quantum eigensolver (VQE). The dark-blue (red) and the green (purple) lines represent the average energies (fidelity between the produced state and the exact ground state) at different iterations for the variational quantum eigensolver with small-angle initialization (VQE-SA) method and variational generative optimization network (VGON), respectively. The exact average ground state energy is depicted by the black dots. The inset zooms in on the convergence behavior of the average energies for VGON and VQE-SA, showcasing the faster convergence of VGON. Each method is repeated ten times to calculate the mean values and 95% confidence intervals.

### Identifying degenerate ground state space of quantum models

Deterministic gradient-based optimization methods are predisposed to follow a single path, therefore hampering their ability to efficiently detect multiple optima. A unique advantage of generative models is the ability to produce diverse samples of objects, all of which may minimize the objective function. In optimization, this leads to the possibility of finding multiple optimal solutions with a single stage of training. We now show that when appropriately trained, VGON exemplifies such an effective capability for generating multiple (nearly) optimal solutions simultaneously. This capability can be largely ascribed to its integration of randomness and the adoption of batch training. The former facilitates broader exploration within the variational manifold, while the latter, which involves processing subsets of data samples concurrently, supports the collective identification of multiple optimal solutions.

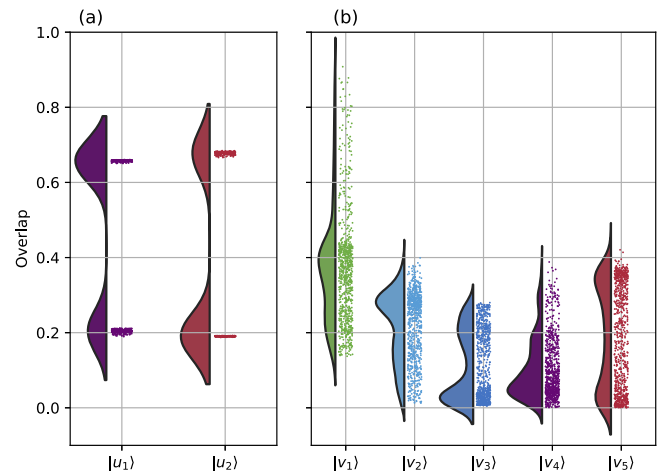
A natural multi-optima problem in quantum many-body physics is the exploration of degenerate ground spaces of quantum many-body Hamiltonians. We apply VGON to two Hamiltonians with known degenerate ground states: the Majumdar-Ghosh (MG)<sup>54,55</sup> model in Eq. (2), and a Heisenberg-like model in Eq. (3) coming from one of the contextuality witnesses presented in ref. 37:

$$H_{MG} = \sum_{i=1}^N \sigma^i \cdot \sigma^{i+1} + \sigma^{i+1} \cdot \sigma^{i+2} + \sigma^i \cdot \sigma^{i+2}, \quad (2)$$

$$H_{232} = \sum_{i=1}^N (2\sigma_x^i \sigma_x^{i+1} + \sigma_x^i \sigma_y^{i+1} - \sigma_y^i \sigma_x^{i+1}), \quad (3)$$

where  $\sigma^i = (\sigma_x^i, \sigma_y^i, \sigma_z^i)$  are Pauli operators at site  $i$ . We take  $N = 10$  for  $H_{MG}$  and  $N = 11$  for  $H_{232}$ , making their ground state spaces five- and two-fold degenerate, respectively. An orthonormal basis for their respective degenerate ground state spaces is computed by the ED method, which outputs five vectors  $|v_1\rangle \dots |v_5\rangle$  spanning the ground state space of  $H_{MG}$  and two vectors  $|u_1\rangle$  and  $|u_2\rangle$  spanning that of  $H_{232}$ .

The overall objective of this problem is similar to the previous one: finding the ground states of  $H_{MG}$  and  $H_{232}$  with variational quantum



**Fig. 4 | The overlap between 1000 states generated by the trained variational generative optimization network and the orthonormal bases of the ground space.** The corresponding orthonormal bases of the ground space are computed by exact diagonalization, with notations  $|u_1\rangle$  and  $|u_2\rangle$  for  $H_{232}$  shown in (a) and  $|v_1\rangle, |v_2\rangle, \dots, |v_5\rangle$  for  $H_{MG}$  shown in (b). The shaded curves show the population densities of the generated states having different overlaps with one of the basis states.

circuits. We maintain the same circuit layout as in the previous problem, and use 36 and 60 blocks of unitary gates for each Hamiltonian respectively. Profiting from the use of mini-batches to estimate gradients, a common technique in training neural networks, VGON can effectively evaluate many different circuits simultaneously. Meanwhile, to enhance intra-batch diversity, a penalty term consisting of the mean cosine similarity among all pairs of sets of circuit parameters in the same batch is added to the objective function. This penalty term, together with the mean energy of the states in the batch, ensures a balance between maintaining the diversity of the generated outputs and minimizing the energy. Further details can be found in Supplementary Note 3.1.

As a result, unlike VQE-based algorithms aiming to generate multiple energy eigenstates, the objective function of VGON is model-agnostic. In other words, with no prior knowledge of the degeneracy of the ground space, VGON is capable of generating orthogonal or linearly independent states in it. In comparison, to achieve a diversity of outputs with VQE-based algorithms<sup>56,57</sup>, it is essential to provide diverse inputs for the VQE model. However, attaining this diversity can result in barren plateaux within the optimization landscape. Though employing VQE-SA may address this problem, it could significantly diminish the diversity of inputs, as it tends to constrain inputs to values near zero.

We generate 1000 output states for each Hamiltonian using a VGON model trained with the above objective function. We find that the vast majority of these states have energy low enough to be treated as ground states. Figure 4 shows the overlap between the generated states and the basis of their ground state space. In Fig. 4a, the generated states for  $H_{232}$  fall into two orthogonal classes, which form an orthonormal basis of the ground state space. For  $H_{MG}$ , Fig. 4b shows that most of them are linearly independent and span the same space as  $|v_1\rangle \dots |v_5\rangle$ .

### Conclusions

We propose a general approach called variational generative optimization network, or VGON for short, for tackling variational optimization challenges in a variety of quantum problems. This approach combines deep generative models in classical machine learning with sampling procedures and a problem-specific objective function, exhibiting excellent convergence efficiency and solution quality in quantum optimization problems of various sizes. Particularly, it may alleviate the barren plateau problem, a pervasive issue in variational quantum algorithms, and surpasses the performance of the VQE-SA method, an approach designed specifically to

avoid barren plateaux. Additionally, the capability of VGON to identify degenerate ground states of quantum many-body models underscores its efficacy in addressing problems with multiple optima. Beyond the quantum world, generative models are emerging as powerful tools in the field of optimization problems. For instance, diffusion models are now being utilized for combinatorial optimizations<sup>58</sup>. Due to the flexible designs, we also envisage VGON and such algorithms to complement each other in addressing a broader spectrum of optimization challenges.

## Methods

### Details on VGON

A VGON model contains two neural networks, i.e., the encoder  $E_\omega : \mathcal{X} \rightarrow \mathcal{Z}$  and the decoder  $D_\phi : \mathcal{Z} \rightarrow \mathcal{X}$  connected by a latent space  $\mathcal{Z}$ , and they are parameterized by  $\omega$ ,  $\phi$ , respectively. These parameters are iteratively updated to produce a solution distribution  $P_{\phi,\omega}(\mathbf{x})$  such that the expectation of an objective function  $h(\mathbf{x})$  is optimized:

$$\begin{aligned} & \mathbb{E}_{\mathbf{x} \sim P_{\phi,\omega}(\mathbf{x})}[h(\mathbf{x})] \\ &= \int \int h(D_\phi(\mathbf{z}))P_\omega(\mathbf{z}|\mathbf{x}_0)P(\mathbf{x}_0)d\mathbf{x}_0d\mathbf{z} \\ &= \int h(D_\phi(\mathbf{z}))P_\omega(\mathbf{z})d\mathbf{z} \\ &= \mathbb{E}_{\mathbf{z} \sim P_\omega(\mathbf{z})}[h(D_\phi(\mathbf{z}))]. \end{aligned}$$

More specifically, the input data  $\mathbf{x}_0$  is sampled from a given distribution  $P(\mathbf{x}_0)$ , which is then mapped to the latent distribution  $P_\omega(\mathbf{z})$  by the encoding estimator  $E_\omega$ , i.e.,  $P_\omega(\mathbf{z}) = \int P_\omega(\mathbf{z}|\mathbf{x}_0)P(\mathbf{x}_0)d\mathbf{x}_0$ . Next the decoding estimator  $D_\phi(\mathbf{z})$  further maps the latent distribution  $P_\omega(\mathbf{z})$  to a distribution  $P_{\phi,\omega}(\mathbf{x})$  of the target data  $\mathbf{x}$ , which is right the input of the objective function  $h(\mathbf{x})$ .

Additionally, for the convenience of training, we constraint the distribution of the latent space to be a normal distribution  $\mathcal{N}(\boldsymbol{\mu}(\mathbf{z}), \boldsymbol{\sigma}^2(\mathbf{z}))$ , and try to minimize the distance between it and the standard normal distribution  $\mathcal{N}(0, I)$ , measured by the KL divergence. Specifically, the cost function for VGON is formulated as:

$$C(\phi, \omega) = \mathbb{E}_{\mathbf{x} \sim P_{\phi,\omega}(\mathbf{x})}[h(\mathbf{x})] + \beta \cdot \mathcal{D}[\mathcal{N}(\boldsymbol{\mu}(\mathbf{z}), \boldsymbol{\sigma}^2(\mathbf{z}))||\mathcal{N}(0, I)], \quad (4)$$

where the hyperparameter  $\beta$  represents the trade-off between the expectation of the objective function and the above KL divergence.

In our implementations of VGON, all the training procedures are conducted based on the PyTorch framework<sup>59</sup>. To address different tasks, diverse objective functions  $h(\mathbf{x})$  are employed, and each requires a specific interfacing with PyTorch. The configurations of these VGON models will be detailed in the subsequent sections, which can provide us a comprehensive understanding of how VGON is tailored for different optimization challenges.

### Details on finding the optimal states in entanglement detection

To find the quantum state that can exhibit the largest advantage of 1-LOCC protocols over LO protocols in entanglement detection, the problem can be formulated as maximizing the difference between the solutions to the two SDPs introduced in Eq. (1), with the following objective function and parameter space:

- Objective function:  $h(\rho', p_1) = p_2^{LO*}(\rho', p_1) - p_2^{1-LOCC*}(\rho', p_1)$
- Parameter space:  $\{e_1 \in \mathbb{R}, \rho' \in \{\rho_{exp}\} \text{ or } \{\rho\}\}$

Here,  $p_1$  is parameterized as  $p_1 = (\tanh(e_1) + 1)/2$ , and  $\{\rho_{exp}\}$  and  $\{\rho\}$  represent the set of states for the pure case and mixed case, respectively.

Efficient variational optimization for these SDPs and their integration into the PyTorch framework for machine learning requires the use of CVXPY and cvxpylayers<sup>60-62</sup>. The first translates a convex optimization problem into a form that solvers can understand, while the latter allows automatic differentiation of convex optimization problems by computing their gradients and backpropagate them through the neural network.

As we mentioned in the main text, the state space we consider follows closely the linear optical setup which generates arbitrary bipartite qutrit states  $\rho_{exp}$ . Photons generated by the laser source are expressed as  $\sum_i c_i |i\rangle$ , where  $c_i$  are complex numbers satisfying  $\sum_i |c_i|^2 = 1$ . Afterwards, these photons go through spontaneous parametric down-conversion (SDPC), which converts their state to  $|\psi\rangle = \sum_i c_i |ii\rangle$ . In the case of qutrits, the state can be parameterized as<sup>63,64</sup>:

$$|\psi\rangle = \sin \frac{\theta}{2} \cos \frac{\phi}{4} e^{im} |00\rangle + \sin \frac{\theta}{2} \sin \frac{\phi}{4} e^{in} |11\rangle + \cos \frac{\theta}{2} |22\rangle,$$

where  $\phi, m, n \in [0, 2\pi)$ , and  $\theta \in [0, \pi]$  are variational parameters. Subsequently, two local unitaries denoted by  $U_A, U_B$  can be applied on the two subsystems, resulting in the quantum state:

$$\rho_{exp} = (U_A \otimes U_B)|\psi\rangle\langle\psi|(U_A^\dagger \otimes U_B^\dagger).$$

Here,  $U_A (U_B)$  can be parameterized by a set of  $3^2 = 9$  linearly-independent skew-Hermitian matrices  $\{T_j\}$ <sup>65</sup>, i.e.,

$$U_A (U_B) = \exp\left(\sum_{j=1}^9 \lambda_j T_j\right),$$

where  $\lambda_j$ 's are nine real numbers, denoted as  $\lambda_A (\lambda_B)$ . Therefore, the parameterized space for pure states  $\rho_{exp}$  is represented as:

$$\{m, n, \phi, \theta \in \mathbb{R}, \lambda_A, \lambda_B \in \mathbb{R}^9\}.$$

On the other hand, mixed qutrit states  $\rho \in \mathcal{H}^3 \otimes \mathcal{H}^3$  can be parameterized by:

$$\rho = U \Sigma U^\dagger,$$

where  $\Sigma$  is a  $9 \times 9$  diagonal matrix whose diagonal entries are nonnegative and sum to 1, and  $U$  is a unitary matrix that can be parameterized by a set of  $9^2 - 1$  generalized Gell-Mann matrices  $\{T_j\}$ <sup>66</sup>, i.e.,

$$U = \exp\left(i \sum_{j=1}^{9^2-1} \lambda_j T_j\right),$$

where  $\lambda_j$ 's are  $9^2 - 1$  real numbers, denoted as  $\lambda \in \mathbb{R}^{9^2-1}$ . Furthermore, the normalized diagonal matrix  $\Sigma$ , denoted as  $\text{diag}(\sigma_1^2, \dots, \sigma_9^2)$ , can be obtained by ensuring that the Euclidean norm of the vector  $\sigma$  is equal to 1, i.e.,  $|\sigma|_2 = 1$ , where  $\sigma = (\sigma_1, \dots, \sigma_9)$ . Consequently, the parameterized space for mixed states  $\rho$  case is written as:

$$\{\lambda \in \mathbb{R}^{9^2-1}, \sigma \in \mathbb{R}^9 : |\sigma|_2 = 1\}.$$

### Details on alleviating barren plateaus in variational quantum algorithms

A typical VQE algorithm can approximate the ground state of a given Hamiltonian  $H$  using a variational wave function generated by a parameterized quantum circuit (PQC)  $U(\boldsymbol{\theta})$ , represented as  $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|00 \dots 0\rangle$ . This sets up the minimization problem:

- Objective function:  $h(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|H|\psi(\boldsymbol{\theta})\rangle$
- Parameter space:  $\{\boldsymbol{\theta} \in \mathbb{R}^M\}$

The dimension of the parameter space  $M$  is determined by the structure of the PQC.

The simulation of PQCs and the computation of energy are implemented by PennyLane<sup>67</sup>, a software library for quantum machine learning. Its support of the CUDA-based CuQuantum SDK from NVIDIA enables

VGON to handle over 10,000 variational parameters on a consumer grade graphics card. PennyLane also provides seamless integration with PyTorch and its machine learning toolkit. Efficient GPU-accelerated simulation of PQCs is achieved by using the adjoint differentiation method<sup>68</sup> to compute the gradients, after which the parameters are updated by the Adam optimizer.

One of the key differences between VQE, VQE-SA and VGON is the initialization of parameters. For the VQE and VQE-SA, the initial parameters  $\theta$  are uniformly sampled from the range  $[0, 2\pi)$  and  $[0, 0.01)$ , respectively. In VGON, on the other hand, the decoder initialized using PyTorch's default settings generates the circuit's initial parameters  $\theta$ . For more details on these methodologies and the comparisons between their performance, please refer to "Alleviating the effect of barren plateaux in variational quantum algorithms" in "Results" and Supplementary Note 2.

### Details on identifying degenerate ground state space of quantum models

To identify the degenerate ground space of a Hamiltonian  $H$  with VGON, the objective function needs two pivotal components to steer the optimized quantum state  $|\psi(\theta)\rangle$  towards diverse ground states. The first component utilizes a PQC  $U(\theta)$  to generate the state  $|\psi(\theta)\rangle = U(\theta)|00 \cdots 0\rangle$ , targeting the ground space. The second component integrates a cosine similarity measure into the optimization objective, aiming to enhance the diversity among the generated quantum states.

Specifically, for a batch of  $S_b$  states  $\{|\psi(\theta_i)\rangle\}$ , the mean energy is calculated by:

$$\bar{E}(\Theta) = \frac{1}{S_b} \sum_{i=1}^{S_b} \langle \psi(\theta_i) | H | \psi(\theta_i) \rangle,$$

where  $\Theta = (\theta_1, \theta_2, \dots, \theta_{S_b})$ . In addition, a penalty term for the objective function based on the cosine similarity is defined as:

$$\bar{S}_{C_{S_b}^2}(\Theta) = \frac{1}{|C_{S_b}^2|} \sum_{(i,j) \in C_{S_b}^2} \frac{\theta_i \cdot \theta_j}{\|\theta_i\| \|\theta_j\|},$$

where  $C_{S_b}^2$  represents the set of all 2-combinations pairs derived from the elements in  $\{1, 2, \dots, S_b\}$ , and  $\|\cdot\|$  denotes the Euclidean norm. Eventually, the optimization objective is set as minimizing the linear combination of  $\bar{E}(\Theta)$  and  $\bar{S}_{C_{S_b}^2}(\Theta)$  according to a trade-off coefficient  $\gamma$ , i.e.,

- Objective function:  $h(\Theta) = \bar{E}(\Theta) + \gamma \cdot \bar{S}_{C_{S_b}^2}(\Theta)$
- Parameter space:  $\{\Theta \in \mathbb{R}^{S_b M}\}$

We estimate the quality of the generated state by computing the overlap between them and a basis of the ground space. Such computations can be resource-intensive, and therefore we only demonstrate the performance of VGON for 10-qubit systems.

### Data availability

Data sets generated during the current study are available from the corresponding authors on reasonable request.

### Code availability

The complete code of this study is openly accessible via the GitHub repository <https://github.com/zhangjianzhang/VGON>.

Received: 10 February 2025; Accepted: 4 August 2025;

Published online: 15 August 2025

### References

1. Kochenderfer, M. J. & Wheeler, T. A. *Algorithms for Optimization*. (The MIT Press, 2019).

2. Navascués, M., Pironio, S. & Acín, A. Bounding the set of quantum correlations. *Phys. Rev. Lett.* **98**, 010401 (2007).
3. Tavakoli, A., Pozas-Kerstjens, A., Brown, P. & Araújo, M. Semidefinite programming relaxations for quantum correlations. *Rev. Mod. Phys.* **96**, 045006 (2024).
4. Doherty, A. C., Parrilo, P. A. & Spedalieri, F. M. Distinguishing separable and entangled states. *Phys. Rev. Lett.* **88**, 187904 (2002).
5. Peruzzo, A. et al. A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **5**, 4213 (2014).
6. McArdle, S., Endo, S., Aspuru-Guzik, A., Benjamin, S. C. & Yuan, X. Quantum computational chemistry. *Rev. Mod. Phys.* **92**, 015003 (2020).
7. Farhi, E., Goldstone, J. & Gutmann, S. A quantum approximate optimization algorithm. Preprint at *arXiv* <https://doi.org/10.48550/arXiv.1411.4028> (2014).
8. Schuld, M., Bocharov, A., Svore, K. M. & Wiebe, N. Circuit-centric quantum classifiers. *Phys. Rev. A* **101**, 032308 (2020).
9. Romero, J., Olson, J. P. & Aspuru-Guzik, A. Quantum autoencoders for efficient compression of quantum data. *Quantum Sci. Technol.* **2**, 045001 (2017).
10. Havlíček, V. et al. Supervised learning with quantum-enhanced feature spaces. *Nature* **567**, 209–212 (2019).
11. Murphy, K. P. *Probabilistic Machine Learning: An introduction* (MIT Press, 2022).
12. Krenn, M. et al. On scientific understanding with artificial intelligence. *Nat. Rev. Phys.* **4**, 761–769 (2022).
13. Wang, H. et al. Scientific discovery in the age of artificial intelligence. *Nature* **620**, 47–60 (2023).
14. Davies, A. et al. Advancing mathematics by guiding human intuition with AI. *Nature* **600**, 70–74 (2021).
15. Fawzi, A. et al. Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature* **610**, 47–53 (2022).
16. Jiménez-Luna, J., Grisoni, F. & Schneider, G. Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* **2**, 573–584 (2020).
17. Radford, A., Narasimhan, K., Salimans, T. & Sutskever, I. Improving language understanding by generative pre-training. OpenAI Preprint (2017).
18. Vaswani, A. et al. Attention is all you need. In *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems* (eds von Luxburg, U., Guyon, I., Bengio, S., Wallach, H. & Fergus, R.) (Curran Associates Inc., 2017).
19. Romera-Paredes, B. et al. Mathematical discoveries from program search with large language models. *Nature* **625**, 468–475 (2024).
20. Carleo, G. & Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **355**, 602–606 (2017).
21. Melko, R. G., Carleo, G., Carrasquilla, J. & Cirac, J. I. Restricted Boltzmann machines in quantum physics. *Nat. Phys.* **15**, 887–892 (2019).
22. Li, S.-H. & Wang, L. Neural network renormalization group. *Phys. Rev. Lett.* **121**, 260601 (2018).
23. Stornati, P. *Variational Quantum Simulations of Lattice Gauge Theories*. Ph.D. thesis, Humboldt-Universität zu Berlin (2022).
24. Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. *ICLR'14: 2nd International Conference on Learning Representations* (2014).
25. Rocchetto, A., Grant, E., Strelchuk, S., Carleo, G. & Severini, S. Learning hard quantum distributions with variational autoencoders. *npj Quantum Inf.* **4**, 28 (2018).
26. Luchnikov, I. A., Ryzhov, A., Stas, P.-J., Filippov, S. N. & Ouerdane, H. Variational autoencoder reconstruction of complex many-body physics. *Entropy* **21**, 1091 (2019).

27. Carrasquilla, J., Torlai, G., Melko, R. G. & Aolita, L. Reconstructing quantum states with generative models. *Nat. Mach. Intell.* **1**, 155–161 (2019).
28. Nakaji, K. et al. The generative quantum eigensolver (GQE) and its application for ground state search. <https://arxiv.org/abs/2401.09253> (2024).
29. Doersch, C. Tutorial on variational autoencoders. Preprint at *arXiv* <https://doi.org/10.48550/arXiv.1606.05908> (2021).
30. Villani, C. *Optimal Transport: Old and New No. 338 in Grundlehren der mathematischen Wissenschaften* (Springer, 2009).
31. Peyré, G. et al. Computational optimal transport: with applications to data science. *Found. Trends Mach. Learn.* **11**, 355–607 (2019).
32. Devroye, L. *Non-Uniform Random Variate Generation* (Springer, 1986).
33. Zhang, L., E, W. & Wang, L. Monge-Ampère flow for generative modeling. <https://arxiv.org/abs/1809.10188> (2018).
34. Nielsen, D., Jaini, P., Hoogeboom, E., Winther, O. & Welling, M. SurVAE flows: surjections to bridge the gap between VAEs and flows. *Adv. Neural Inf. Process. Syst.* **33**, 12685–12696 (2020).
35. Mironowicz, P. Semi-definite programming and quantum information. *J. Phys. A Math. Theor.* **57**, 163002 (2024).
36. Yang, K. et al. Contextuality in infinite one-dimensional translation-invariant local Hamiltonians. *npj Quantum Inf.* **8**, 89 (2022).
37. Weilenmann, M., Aguilar, E. A. & Navascués, M. Analysis and optimization of quantum adaptive measurement protocols with the framework of preparation games. *Nat. Commun.* **12**, 4553 (2021).
38. Xing, W.-B. et al. Practical advantage of classical communication in entanglement detection. <https://arxiv.org/abs/2504.09791> (2025).
39. Hanin, B. et al. (eds) Which neural net architectures give rise to exploding and vanishing gradients? In *NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems* (eds Bengio, S. et al.) (Curran Associates, Inc., 2018).
40. Kolen, J. F. & Kremer, S. C. *Gradient Flow in Recurrent Nets: The Difficulty of Learning Long-Term Dependencies*, 237–243 (Wiley-IEEE Press, 2001).
41. McClean, J. R., Boixo, S., Smelyanskiy, V. N., Babbush, R. & Neven, H. Barren plateaus in quantum neural network training landscapes. *Nat. Commun.* **9**, 4812 (2018).
42. Cerezo, M., Sone, A., Volkoff, T., Cincio, L. & Coles, P. J. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nat. Commun.* **12**, 1791 (2021).
43. Ortiz Marrero, C., Kieferová, M. & Wiebe, N. Entanglement-induced barren plateaus. *PRX Quantum* **2**, 040316 (2021).
44. Kim, J., Kim, J. & Rosa, D. Universal effectiveness of high-depth circuits in variational eigenproblems. *Phys. Rev. Res.* **3**, 023203 (2021).
45. Larocca, M., Ju, N., García-Martín, D., Coles, P. J. & Cerezo, M. Theory of overparametrization in quantum neural networks. *Nat. Comput. Sci.* **3**, 542–551 (2023).
46. Romero, J. et al. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Sci. Technol.* **4**, 014008 (2018).
47. Taube, A. G. & Bartlett, R. J. New perspectives on unitary coupled-cluster theory. *Int. J. Quantum Chem.* **106**, 3393–3401 (2006).
48. Cerezo, M. et al. Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing. Preprint at *arXiv* <https://doi.org/10.48550/arXiv.2312.09121> (2024).
49. Harrow, A. W. & Low, R. A. Random quantum circuits are approximate 2-designs. *Commun. Math. Phys.* **291**, 257–302 (2009).
50. Holmes, Z., Sharma, K., Cerezo, M. & Coles, P. J. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. *PRX Quantum* **3**, 010313 (2022).
51. Haug, T., Bharti, K. & Kim, M. Capacity and quantum geometry of parametrized quantum circuits. *PRX Quantum* **2**, 040309 (2021).
52. Sack, S. H., Medina, R. A., Michailidis, A. A., Kueng, R. & Serbyn, M. Avoiding barren plateaus using classical shadows. *PRX Quantum* **3**, 020365 (2022).
53. Ran, S.-J. Encoding of matrix product states into quantum circuits of one- and two-qubit gates. *Phys. Rev. A* **101**, 032310 (2020).
54. Majumdar, C. K. & Ghosh, D. K. On next-nearest-neighbor interaction in linear chain. I. *J. Math. Phys.* **10**, 1388–1398 (1969).
55. Majumdar, C. K. & Ghosh, D. K. On Next-nearest-neighbor interaction in linear chain. II. *J. Math. Phys.* **10**, 1399–1402 (1969).
56. Higgott, O., Wang, D. & Brierley, S. Variational quantum computation of excited states. *Quantum* **3**, 156 (2019).
57. Nakanishi, K. M., Mitarai, K. & Fujii, K. Subspace-search variational quantum eigensolver for excited states. *Phys. Rev. Res.* **1**, 033062 (2019).
58. Sanokowski, S., Hochreiter, S. & Lehner, S. A diffusion model framework for unsupervised neural combinatorial optimization. In *ICML'24: Proceedings of the 41st International Conference on Machine Learning*, PMLR **235**, 43346–43367 (2024).
59. Paszke, A. et al. PyTorch: an imperative style, high-performance deep learning library. In *NIPS'19: Proceedings of the 33rd International Conference on Neural Information Processing Systems* (eds Wallach, H. et al.) (Curran Associates, Inc., 2019).
60. Diamond, S. & Boyd, S. CVXPY: a Python-embedded modeling language for convex optimization. *J. Mach. Learn. Res.* **17**, 1–5 (2016).
61. Agrawal, A., Verschuere, R., Diamond, S. & Boyd, S. A rewriting system for convex optimization problems. *J. Control Decis.* **5**, 42–60 (2018).
62. Agrawal, A. et al. Differentiable convex optimization layers. In *NIPS'19: Proceedings of the 33rd International Conference on Neural Information Processing Systems* (eds Wallach, H. M., Larochelle, H., Beygelzimer, A., d'Alché Buc, F. & Fox, E. B.) (Curran Associates, Inc., 2019).
63. Hu, X.-M. et al. Optimized detection of high-dimensional entanglement. *Phys. Rev. Lett.* **127**, 220501 (2021).
64. Hu, X.-M. et al. Beating the channel capacity limit for superdense coding with entangled ququarts. *Sci. Adv.* **4**, eaat9304 (2018).
65. Hyland, S. & Rätsch, G. Learning unitary operators with help from  $u(n)$ . *Proc. AAAI Conf. Artif. Intell.* **31**, 2050–2058 (2017).
66. Bertlmann, R. A. & Krammer, P. Bloch vectors for qudits. *J. Phys. A Math. Theor.* **41**, 235303 (2008).
67. Bergholm, V. et al. PennyLane: automatic differentiation of hybrid quantum-classical computations. Preprint at *arXiv* <https://doi.org/10.48550/arXiv.1811.04968> (2022).
68. Jones, T. & Gacon, J. Efficient calculation of gradients in classical simulations of variational quantum algorithms. Preprint at *arXiv* <https://doi.org/10.48550/arXiv.2009.02823> (2020).

## Acknowledgements

This work is supported by the Sichuan Science and Technology Program (2024YFHZ0371), the National Natural Science Foundation of China (62250073, 62272259, 62332009) and the National Key R&D Program of China (2021YFE0113100). The authors would like to thank Abolfazl Bayat, Dongling Deng, Chu Guo, Zhengfeng Ji, Damian Markham, Miguel Navascués and Ying Tang for helpful comments.

## Author contributions

Z.Z.W. and Z.H.W. conceived the project. L.Z., X.L., P.W., K.Y., X.Z. and Z.Z.W. performed the numerical simulations. All authors contributed to writing and editing of the manuscript.

## Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42005-025-02261-4>.

**Correspondence** and requests for materials should be addressed to Zhaohui Wei or Zizhu Wang.

**Peer review information** *Communications Physics* thanks Ghazi Mohammad Vakili and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025