Letter

# Training variational quantum algorithms with random gate activation

Shuo Liu [1], Shi-Xin Zhang,[2,*] Shao-Kai Jian,[3,†] and Hong Yao[1,‡]
[1]Institute for Advanced Study, Tsinghua University, Beijing 100084, China
[2]Tencent Quantum Laboratory, Tencent, Shenzhen, Guangdong 518057, China
[3]Department of Physics & Engineering Physics, Tulane University, New Orleans, Louisiana 70118, USA

(Received 29 March 2023; revised 30 May 2023; accepted 30 August 2023; published 18 September 2023)

Variational quantum algorithms (VQAs) hold great potential for near-term applications and are promising to achieve quantum advantage in practical tasks. However, VQAs suffer from severe barren plateau problems and have a significant probability of being trapped in local minima. In this Research Letter, we propose a training algorithm with random quantum gate activation for VQAs to efficiently address these two issues. This algorithm processes effectively many fewer training parameters than the conventional plain optimization strategy, which efficiently mitigates barren plateaus with the same expressive capability. Additionally, by randomly adding two-qubit gates to the circuit ansatz, the optimization trajectories can escape from local minima and reach the global minimum more frequently due to more sources of randomness. In real quantum experiments, the training algorithm can also reduce the quantum computational resources required and be more quantum noise resilient. We apply our training algorithm to solve variational quantum simulation problems for ground states and present convincing results that showcase the advantages of our strategy, where better performance is achieved by the combination of mitigating barren plateaus, escaping from local minima, and reducing the effect of quantum noise.

DOI: 10.1103/PhysRevResearch.5.L032040

*Introduction.* Recently, various quantum-classical hybrid variational algorithms, such as the variational quantum eigensolver (VQE) [1,2], the quantum approximate optimization algorithm (QAOA) [3], and quantum neural networks [4], have been proposed with the vision of establishing valuable quantum killer apps in the noisy intermediate-scale quantum (NISQ) era [5]. In the standard variational quantum algorithm (VQA) setting, we minimize the expectation value of the objective function $O$ designed for the target problem with respect to a variational state (ansatz) $|\psi(\vec{\theta})\rangle$ prepared by a parametrized quantum circuit (PQC). This hybrid scheme is accomplished with a feedback loop between classical computers and quantum devices: The quantum devices repeatedly prepare the variational state $|\psi(\vec{\theta})\rangle$ for estimation of the expectation value of the objective function $\langle\psi(\vec{\theta})|O|\psi(\vec{\theta})\rangle$, while the classical computers are utilized to optimize the parameters $\vec{\theta}$ based on classical optimization strategies such as gradient descent.

To guarantee that the VQA solution is close enough to the exact solution for the target problem, high expressibility of the quantum ansatz $U$ is required. In principle, we can increase the depth of the PQC with more quantum gates and training parameters or apply neural network postprocessing modules [6,7] to achieve a higher expressibility. For VQAs on real quantum hardware, trainability is also an important factor. Unfortunately, there is a trade-off between expressibility and trainability [8], and the performance of VQAs is severely limited by optimization issues such as barren plateaus [9], which can be induced by the entanglement in the quantum circuit [10–12] and the noise on NISQ devices [13]. The gradients vanish exponentially with the depth of the PQC. Therefore exponential computational resources are required to accurately estimate the value of the gradients, and a large number of iterations is required for the VQA solution to converge. Even the gradient-free optimization approaches are also suppressed by the barren plateaus [14]. To mitigate the barren plateaus and achieve better performance from the VQA, a series of strategies have been proposed, including parameter initialization methods [15–23], a local objective function [24–26], and a special quantum circuit ansatz [27–34].

Apart from the barren plateaus, the nonconvexity of energy landscapes and the existence of a large number of local minima also strongly limit the trainability of VQAs [35] and render training VQAs unscalable. The VQA solutions provided by gradient descent can be easily trapped in local minima which are correlated with the initialization position and far away from the global minimum [36]. To escape from the local minima, exponential trials with random initialization of parameters, i.e., exponential optimization trajectories, are required in the general case [37]. Although some strategies have been investigated [38–42], more effective and universal strategies are still required to better improve the trainability of VQAs.

In this Research Letter, we propose a training algorithm for VQAs with trainable gates activated randomly and

*shixinzhang@tencent.com
†sjian@tulane.edu
‡yaohong@tsinghua.edu.cn

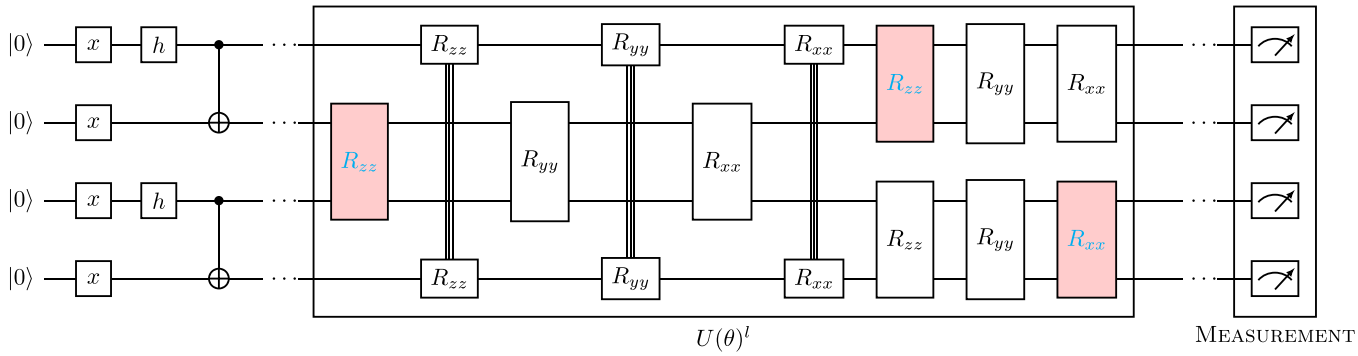


FIG. 1. The HVA PQC for an XXZ model with four qubits: Red rectangles represent the activated two-qubit gates in which the parameters are updated in VQE iterative optimization; white rectangles represent the unactivated two-qubit gates, which can be regarded as identity and do not need to be implemented in real experiments. For the training procedure, firstly, we randomly activate 10% two-qubit gates with random initial parameters $\theta_i \in [0, 2\pi]$, while other unactivated two-qubit gates' parameters remain as 0. During training, more and more two-qubit gates are randomly activated, and their parameters will unfreeze and update. $l$ is the PQC depth, i.e., how many times the Hamiltonian block repeats.

progressively to overcome the above challenges. We use random activation (RA) to represent this strategy. In our approach, the number of trainable quantum gates is effectively much lower, and thus the barren plateaus can be efficiently mitigated with the expressibility unchanged. Additionally, the randomness of activating quantum gates increases the probability of escaping from the local minima. Moreover, the quantum computational resources required and the negative effect of quantum noise can be both greatly reduced when experimentally realizing the VQA on real quantum hardware.

Without loss of generality, we focus on the VQE task in this Research Letter, while our approach is applicable to other types of VQAs. The VQE has been exploited in a variety of contexts from quantum chemistry [43–49] and many-body physics [2,50–56] to lattice gauge theories [1,57] and is specified by a triplet $(O, |\psi_i\rangle, U(\vec{\theta}))$ including an objective function $O$, an initial quantum state $|\psi_i\rangle$, and a PQC ansatz $U(\vec{\theta})$. The objective function is usually chosen as the expectation or the variance [58–60] of the system Hamiltonian operator $H$, whose solution gives the ground state or the excited state, respectively. There are many choices for PQC structures such as the hardware-efficient ansatz (HEA) [44], the Hamiltonian variational ansatz (HVA) [53,61–63], and the unitary coupled cluster (UCC) ansatz [64–71].

*Hamiltonian and circuit ansatz.* In this Research Letter, we utilize the proposed training algorithm to solve the ground state energy problem of the one-dimensional (1D) antiferromagnetic XXZ model with periodic boundary conditions, a representative lattice spin model in quantum many-body physics. With convincing numerical results and detailed ablation studies, we demonstrate the improved optimization results using our approach and pin down the contributing factors as the mitigation of the barren plateaus and a better chance to reach the global minimum.

The Hamiltonian reads

$$H = \sum_i X_i X_{i+1} + Y_i Y_{i+1} + J_z Z_i Z_{i+1}, \tag{1}$$

where $X, Y, Z$ are Pauli matrices and $J_z$ is the $zz$ interaction strength. We choose the HVA, which is rather effective for several quantum many-body models [53,61,62], as the PQC ansatz in this Research Letter. Specifically, a depth-$l$ HVA for the XXZ model is shown in Fig. 1, and the corresponding circuit unitary is

$$U(\vec{\theta}) = \prod_{i=1}^{l} R_{zz}^{i,\mathrm{odd}} R_{yy}^{i,\mathrm{odd}} R_{xx}^{i,\mathrm{odd}} R_{zz}^{i,\mathrm{even}} R_{yy}^{i,\mathrm{even}} R_{xx}^{i,\mathrm{even}}, \tag{2}$$

where $R_{\sigma\sigma}^{i,\mathrm{odd}} = \prod_j \exp(i\theta_{i,2j+1,\sigma}\sigma_{2j+1}\sigma_{2j+2})$ and $R_{\sigma\sigma}^{i,\mathrm{even}} = \prod_j \exp(i\theta_{i,2j,\sigma}\sigma_{2j}\sigma_{2j+1})$ $(\sigma = x, y, z)$. The parameters $\theta$ of quantum gates are independent tuning parameters.

Applying the PQC $U(\vec{\theta})$ to the initial prepared state $|\psi_i\rangle$ [$|\Psi^-\rangle$ Bell state, i.e., $\bigotimes_{i=0}^{\frac{N}{2}-1} \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)_{2i,2i+1}$], we obtain the output state $|\psi(\vec{\theta})\rangle = U(\vec{\theta})|\psi_i\rangle$ and the VQE energy $E(\vec{\theta}) = \langle\psi(\vec{\theta})|H|\psi(\vec{\theta})\rangle$ as the objective function. For plain training, all two-qubit gates are activated from the beginning of optimization with random initial parameters $\vec{\theta}$ sampled uniformly from $[0, 2\pi]$. Then the parameters $\vec{\theta}$ are optimized by gradient descent to reach the lowest energy $\langle\psi(\vec{\theta}^*)|H|\psi(\vec{\theta}^*)\rangle$, where $\vec{\theta}^*$ are the optimal parameters. In principle, a deeper PQC has higher expressibility and a lower converged VQE energy. However, barren plateau problems become more severe with large PQC depth, posing a fundamental challenge to identifying the optimal parameters $\vec{\theta}^*$ as discussed above.

*Training with random gate activation.* To overcome the challenges presented in the VQA training procedure, we introduce the training algorithm with incrementally random gate activation. Different from the plain training method, only a small fraction of the two-qubit gates, for example, 10%, are activated for the initial optimization iterations (i.e., the two-qubit gates with the red color shown in Fig. 1). Meanwhile, other two-qubit gates stay as identity gates, i.e., $\theta = 0$ for these unactivated gates. During the optimization, the parameters $\vec{\theta}_{10\%}$ of the initial 10% activated gates will be updated to approach the optimal parameters $\vec{\theta}^*_{10\%}$ corresponding to the lowest VQE energy estimation $\langle\psi(\vec{\theta}^*_{10\%})|H|\psi(\vec{\theta}^*_{10\%})\rangle \leqslant \langle\psi(\vec{\theta}_{10\%})|H|\psi(\vec{\theta}_{10\%})\rangle$. Then other random 10% unactivated two-qubit gates are activated. Since the parameters for these newly activated gates are set to 0 previously, the VQE energy has no sudden change for the incremental activation

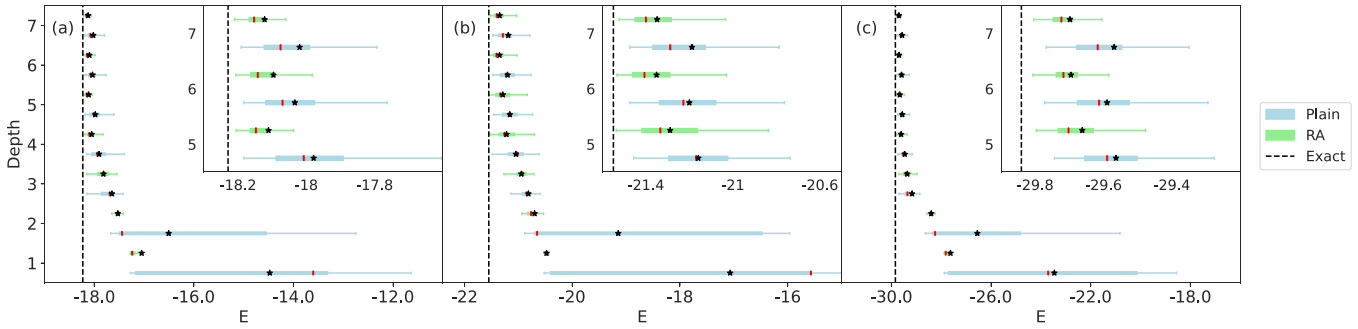


FIG. 2. The converged VQE energy in an $N = 12$ system for the RA training (green) and plain strategy (blue) with different PQC depths $l$: (a) $J_z = 0.5$, (b) $J_z = 1.0$, and (c) $J_z = 2.0$. The black stars represent the averaged VQE energy, and the red lines represent the median VQE energy across 500 independent optimization trials. The insets are zooms of VQE results with deeper PQC. The outliers beyond the caps, which are much larger than the ground truth, are not shown (see the SM for more details about the standard box plots [73]). The performance of RA training is much better: The mean and median of VQE energy from RA is substantially lower than that obtained from plain training, as more trials are trapped in the local minima for plain training.

procedure. Now the parameters of the 20% activated gates are optimized together, and this procedure repeats until all parametrized gates are activated and the optimal parameters $\vec{\theta}^*_{100\%}$ are obtained. This approach has effectively fewer parametrized gates and can hopefully mitigate the barren plateaus with better VQE performance.

*Vqe results comparison.* We provide numerical analyses of the performance of different training strategies, with numerical implementation based on the TENSORCIRCUIT package [72]. The results of the plain training and the RA training are shown in Fig. 2 with varying PQC depths and Hamiltonian parameters (the training dynamics can be found in the Supplemental Material (SM) [73]). The results consist of 500 independent optimization trials on a 12-qubit system. We use the ADAM [74] optimizer with hyperparameters learning_rate $= 0.01$, decay_rate $= 0.9$, and decay_steps $=$ 100, and the percentage of two-qubit gates activated each time is 10% (results are similar for other reasonable hyperparameter choices, and details can be found in the SM [73]. We keep the maximal number of iterations maxiter $= 5000$, which is large enough to guarantee that the two different training approaches achieve their best performance. This is different from the usual approach, where the converge speed instead of the final converged position is the focus [28,31,32]. We believe our comparison is more suitable to exploit the potential quantum advantage of VQAs.

For the RA training, there are many fewer trials trapped in the bad local minima, and the averaged VQE energy is much lower than that obtained from the plain training as shown in Fig. 2. There are two sources for the advantageous performance for the RA training approach, and we will investigate the two factors in detail below: escaping from the local minima more frequently and mitigating the barren plateaus.

The improvement of averaged VQE energy exists for different PQC depths and is particularly significant for shallower quantum circuits with $l = 1, 2$. Note that the barren plateaus are not severe in such a shallow setup. Considering an optimization trajectory of the plain training with random initial parameters $\vec{\theta}$, the converged local minimum of the trajectory is solely determined by the initialization [36] for plain training. As long as the trajectory is trapped in a local minimum, the VQE performance can be rather bad, provided that there are local minima far away from the exact ground state. Instead, if we randomly activate the gate during optimization, for the fixed initial parameters, there are multiple different random optimization trajectories caused by the randomness of how to activate two-qubit gates. The extra randomness sources greatly enlarge the parameter space that can be explored for the optimization trajectories and efficiently help the optimization trajectories escape from local minima more frequently.

Since the effective circuit depth (the ratio of the number of activated quantum gates to the number of qubits) of our algorithm with only a fraction of the number of gates is much lower than that of a plain training algorithm, our algorithm also helps to mitigate barren plateaus. To observe this quantitatively, the averaged variance of the VQE energy gradients with respect to the parameters of the activated gates is shown in Fig. 3. With the increase in the number of activated gates and the system size, the averaged variance decreases exponentially. The case of 100% density corresponds to the plain VQE training where all gates are activated. Such a plain training strategy suffers from the most severe barren plateau problems. As shown in Fig. 3, the gradient variance can be efficiently increased during the RA training procedure. The mitigation

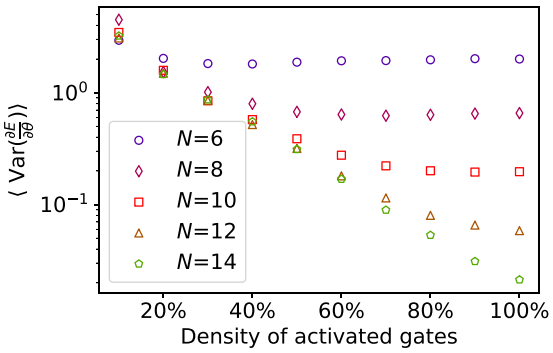


FIG. 3. The averaged variance of energy gradients with respect to the parameters of the activated two-qubit gates. The depth of the PQC is $l = 7$, and the Hamiltonian parameter $J_z = 1.0$. The barren plateau problems become worse with more activated gates.

of barren plateaus can also be confirmed via the better median VQE energy of the training. The median energy is not affected by the outlier local minima trials, and the improvement on the median is more likely a result from mitigating barren plateaus instead of escaping from a local minimum.

In sum, the contribution from escaping the local minimum can be found via the mean of the VQE energy where the distribution of converged energy matters, and the contribution from mitigating barren plateaus can be identified via the median of the VQE energy. Additionally, we have confirmed that the improvements are statistically significant via nonparametric tests (since the converged-value data do not meet the normal distribution).

*Layerwise activation.* We carry out further ablation studies to confirm our analysis of the advantages of RA training. A straightforward question is what will happen if there is no randomness in the gate activation procedure, e.g., activating the two-qubit gates in a layerwise fashion. The barren plateaus can also be mitigated due to the effective lower depth for the layerwise activation idea. However, the extra randomness source for escaping from local minima now disappears. We consider two classes of layerwise optimization: One is the training with layerwise appending activation (LAA), namely, we append the new activated identity layer to the end of the activated layers; the other is the training with layerwise prepending activation (LPA), namely, we prepend the new identity layer to the activated ones.

The distribution of the converged VQE results for different training strategies (plain, RA, LAA, and LPA) is shown in Fig. 4. When the quantum circuit is shallow, almost all the trajectories of RA training successfully escape from the local minima, while many trials of the other three training strategies are trapped in local minima and lead to a larger fluctuation in the results due to the absence of the randomness in the activation. Therefore the randomness of activating two-qubit gates is vital and relevant for better escaping from local minima. When the quantum circuit is deep, the VQE performance of RA training is still much better than that of plain training as discussed above. We have also observed that the VQE performances of RA and LPA are very close. This result indicates that the improvement in VQE performance with deep circuits is mainly due to the mitigation of barren plateaus, which RA and LPA are both capable of.

*Discussion.* We have demonstrated that the RA training can efficiently mitigate the barren plateau problem and help the optimization trajectories escape from local minima based on convincing numerical results. As discussed in the SM [73], we have also offered a potential explanation for the effectiveness of RA training based on the perspective of the entanglement phase transition. This algorithm can be naturally applied to other practical tasks and other flavors of VQAs due to its universal form. Additionally, it can be integrated with other strategies [15–25,27–31,38,39,75,76] to further improve performance and the solution quality of VQAs.

From a practical point of view, the training algorithm presented here brings more benefits when considering implementation on real quantum hardware in the NISQ era. In experiments, the circuit gradients are evaluated via the parameter shift rule [77–80]. For a PQC with $p$ training parameters, we have to evaluate the observable expectation on $2p$ sets of

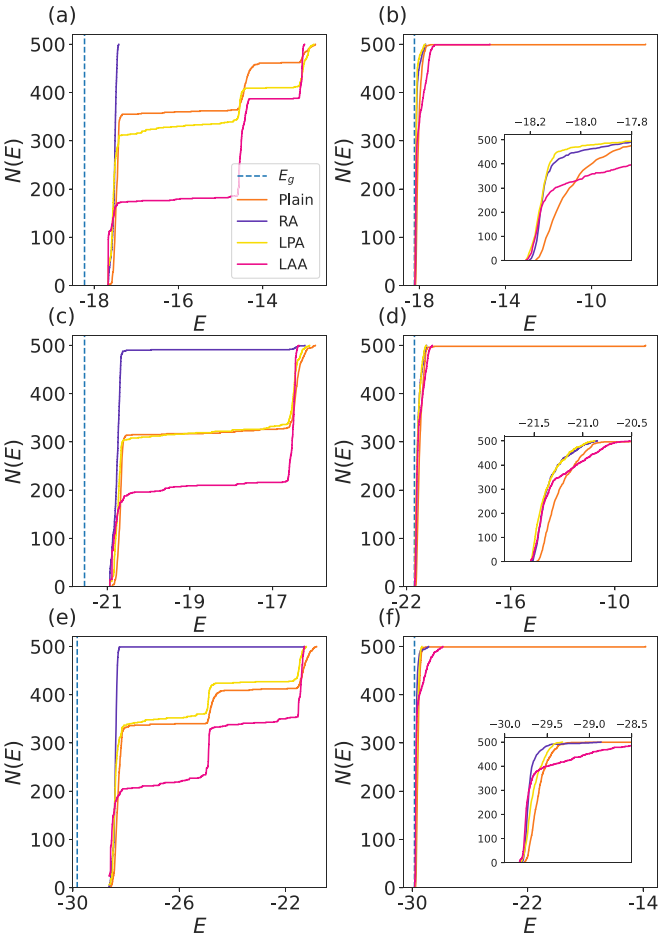


FIG. 4. The accumulated distribution of 500 converged energies from independent optimization trials using four training strategies based on the ADAM optimizer with decay_steps = 100 in an $N = 12$ system: (a) $J_z = 0.5$, $l = 2$; (b) $J_z = 0.5$, $l = 7$; (c) $J_z = 1.0$, $l = 2$; (d) $J_z = 1.0$, $l = 7$; (e) $J_z = 2.0$, $l = 2$; and (f) $J_z = 2.0$, $l = 7$. $N(E)$ is the number of trials whose converged energies are less than $E$. When the quantum circuit is shallow, the VQE performance of RA training is the best because this strategy can escape from local minima more easily. When the quantum circuit is deep, the VQE performances of RA and LPA are very close for $J_z = 0.5$ and $J_z = 1.0$, and RA has the best VQE performance of the four different training strategies for $J_z = 2.0$.

parameters. In the RA training scheme, the effective number of training parameters during the optimization is much smaller than the plain training, which greatly reduces the required number of measurement shots. Moreover, the required number of measurement shots for each observable is also lowered due to the mitigated barren plateaus. Meanwhile, the number of two-qubit gates that one is required to implement on the hardware in RA training is also lower than in the plain training case. Therefore the efficiency in terms of the practical quantum computational resources required is a big advantage of the RA training approach. For the experiments we conducted in this Research Letter, roughly 500 times fewer quantum gates were used for RA training (see the SM for more details about the quantum resource efficiency analysis [73]).

In terms of hardware implementation, RA training also has better noise resilience. On the one hand, we can stop the activation iteration earlier to achieve a better trade-off between the circuit expressive power and the accumulated quantum noise. On the other hand, the effectively shallower circuits also suppress barren plateaus induced by the quantum noise [13]. We show experimental results for noisy VQE settings, and the obtained energies are substantially improved when using RA training (the numerical results and data can be found in the SM [73]). Therefore RA training is a practical and stand-alone training approach that we recommend trying on noisy quantum devices.

*Acknowledgments.* We thank Zhou-Quan Wan for helpful discussions. This work is supported in part by the NSFC under Grant No. 11825404 (S.-X.Z., S.L., and H.Y.), by the MOSTC under MOSTC Grants No. 2021YFA1400100 (H.Y.), and by the NSFC under Grant No. NSFC12042505.

[1] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[2] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New J. Phys. **18**, 023023 (2016).

[3] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv:1411.4028.

[4] E. Farhi and H. Neven, Classification with quantum neural networks on near term processors, arXiv:1802.06002.

[5] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[6] S.-X. Zhang, Z.-Q. Wan, C.-K. Lee, C.-Y. Hsieh, S. Zhang, and H. Yao, Variational Quantum-Neural Hybrid Eigensolver, Phys. Rev. Lett. **128**, 120502 (2022).

[7] S.-X. Zhang, Z.-Q. Wan, C.-Y. Hsieh, H. Yao, and S. Zhang, Variational quantum-neural hybrid error mitigation, Adv. Quantum Technol. 2300147 (2023).

[8] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, Connecting Ansatz Expressibility to Gradient Magnitudes and Barren Plateaus, PRX Quantum **3**, 010313 (2022).

[9] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 4812 (2018).

[10] T. L. Patti, K. Najafi, X. Gao, and S. F. Yelin, Entanglement devised barren plateau mitigation, Phys. Rev. Res. **3**, 033090 (2021).

[11] K. Sharma, M. Cerezo, L. Cincio, and P. J. Coles, Trainability of Dissipative Perceptron-Based Quantum Neural Networks, Phys. Rev. Lett. **128**, 180505 (2022).

[12] C. Ortiz Marrero, M. Kieferová, and N. Wiebe, Entanglement-induced barren plateaus, PRX Quantum **2**, 040316 (2021).

[13] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, Nat. Commun. **12**, 6961 (2021).

[14] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, and P. J. Coles, Effect of barren plateaus on gradient-free optimization, Quantum **5**, 558 (2021).

[15] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, An initialization strategy for addressing barren plateaus in parametrized quantum circuits, Quantum **3**, 214 (2019).

[16] G. Verdon, M. Broughton, J. R. McClean, K. J. Sung, R. Babbush, Z. Jiang, H. Neven, and M. Mohseni, Learning to learn with quantum neural networks via classical neural networks, arXiv:1907.05415.

[17] H.-Y. Liu, T.-P. Sun, Y.-C. Wu, Y.-J. Han, and G.-P. Guo, Mitigating barren plateaus with transfer-learning-inspired parameter initializations, New J. Phys. **25**, 013039 (2023).

[18] A. Rad, A. Seif, and N. M. Linke, Surviving the barren plateau in variational quantum circuits with Bayesian learning initialization, arXiv:2203.02464.

[19] A. Kulshrestha and I. Safro, BEINIT: Avoiding barren plateaus in variational quantum algorithms, arXiv:2204.13751.

[20] F. Sauvage, S. Sim, A. A. Kunitsa, W. A. Simon, M. Mauri, and A. Perdomo-Ortiz, FLIP: A flexible initializer for arbitrarily-sized parametrized quantum circuits, arXiv:2103.08572.

[21] S. H. Sack, R. A. Medina, A. A. Michailidis, R. Kueng, and M. Serbyn, Avoiding Barren Plateaus Using Classical Shadows, PRX Quantum **3**, 020365 (2022).

[22] H. R. Grimsley, G. S. Barron, E. Barnes, S. E. Economou, and N. J. Mayhall, Adaptive, problem-tailored variational quantum eigensolver mitigates rough parameter landscapes and barren plateaus, npj Quantum Inf. **9**, 19 (2023); **9**, 24(E) (2023).

[23] K. Zhang, M.-H. Hsieh, L. Liu, and D. Tao, Gaussian initializations help deep variational quantum circuits escape from the barren plateau, arXiv:2203.09376.

[24] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nat. Commun. **12**, 1791 (2021).

[25] A. V. Uvarov and J. D. Biamonte, On barren plateaus and cost function locality in variational quantum algorithms, J. Phys. A: Math. Theor. **54**, 245301 (2021).

[26] Z. Liu, L.-W. Yu, L.-M. Duan, and D.-L. Deng, Presence and Absence of Barren Plateaus in Tensor-Network Based Machine Learning, Phys. Rev. Lett. **129**, 270501 (2022).

[27] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, Absence of Barren Plateaus in Quantum Convolutional Neural Networks, Phys. Rev. X **11**, 041011 (2021).

[28] X. Liu, G. Liu, J. Huang, H.-K. Zhang, and X. Wang, Mitigating barren plateaus of variational quantum eigensolvers, arXiv:2205.13539.

[29] M. Bilkis, M. Cerezo, G. Verdon, P. J. Coles, and L. Cincio, A semi-agnostic ansatz with variable structure for quantum machine learning, arXiv:2103.06712.

[30] Y. Du, T. Huang, S. You, M.-H. Hsieh, and D. Tao, Quantum circuit architecture search for variational quantum algorithms, npj Quantum Inf. **8**, 62 (2022).

[31] K. Zhang, M.-H. Hsieh, L. Liu, and D. Tao, Toward trainability of deep quantum neural networks, arXiv:2112.15002.

[32] A. Anand, L. B. Kristensen, F. Frohnert, S. Sim, and A. Aspuru-Guzik, Information flow in parameterized quantum circuits, arXiv:2207.05149.

[33] J. Kim, J. Kim, and D. Rosa, Universal effectiveness of high-depth circuits in variational eigenproblems, Phys. Rev. Res. **3**, 023203 (2021).

[34] E. C. Martín, K. Plekhanov, and M. Lubasch, Barren plateaus in quantum tensor network optimization, Quantum **7**, 974 (2023).

[35] L. Bittel and M. Kliesch, Training Variational Quantum Algorithms Is NP-Hard, Phys. Rev. Lett. **127**, 120502 (2021).

[36] E. R. Anschuetz, Critical points in quantum generative models, arXiv:2109.06957.

[37] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, Quantum Approximate Optimization Algorithm: Performance, Mechanism, and Implementation on Near-Term Devices, Phys. Rev. X **10**, 021067 (2020).

[38] J. Rivera-Dean, P. Huembeli, A. Acín, and J. Bowles, Avoiding local minima in variational quantum algorithms with neural networks, arXiv:2104.02955.

[39] L. Slattery, B. Villalonga, and B. K. Clark, Unitary block optimization for variational quantum algorithms, Phys. Rev. Res. **4**, 023072 (2022).

[40] M. Larocca, N. Ju, D. García-Martín, P. J. Coles, and M. Cerezo, Theory of overparametrization in quantum neural networks, Nat. Comput. Sci. **3**, 542 (2023).

[41] D. Faílde, J. D. Viqueira, M. M. Juane, and A. Gómez, Using differential evolution to avoid local minima in variational quantum algorithms, arXiv:2303.12186.

[42] Z. Wang, P.-L. Zheng, B. Wu, and Y. Zhang, Quantum dropout: On and over the quantum approximate optimization algorithm, Phys. Rev. Res. **5**, 023171 (2023).

[43] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus *et al.*, Scalable Quantum Simulation of Molecular Energies, Phys. Rev. X **6**, 031007 (2016).

[44] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature (London) **549**, 242 (2017).

[45] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi, Computation of Molecular Spectra on a Quantum Processor with an Error-Resilient Algorithm, Phys. Rev. X **8**, 011021 (2018).

[46] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik, Quantum chemistry in the age of quantum computing, Chem. Rev. **119**, 10856 (2019).

[47] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, An adaptive variational algorithm for exact molecular simulations on a quantum computer, Nat. Commun. **10**, 3007 (2019).

[48] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, Rev. Mod. Phys. **92**, 015003 (2020).

[49] J. Tilly, G. Jones, H. Chen, L. Wossnig, and E. Grant, Computation of molecular excited states on IBM quantum computers using a discriminative variational quantum eigensolver, Phys. Rev. A **102**, 062425 (2020).

[50] J.-G. Liu, Y.-H. Zhang, Y. Wan, and L. Wang, Variational quantum eigensolver with fewer qubits, Phys. Rev. Res. **1**, 023025 (2019).

[51] Z. Cai, Resource Estimation for Quantum Variational Simulations of the Hubbard Model, Phys. Rev. Appl. **14**, 014059 (2020).

[52] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, Phys. Rev. A **92**, 042303 (2015).

[53] W. W. Ho and T. H. Hsieh, Efficient variational simulation of non-trivial quantum states, SciPost Phys. **6**, 029 (2019).

[54] P.-L. Dallaire-Demers, J. Romero, L. Veis, S. Sim, and A. Aspuru-Guzik, Low-depth circuit ansatz for preparing correlated fermionic states on a quantum computer, Quantum Sci. Technol. **4**, 045005 (2019).

[55] S. Tamiya, S. Koh, and Y. O. Nakagawa, Calculating nonadiabatic couplings and Berry's phase by variational quantum eigensolvers, Phys. Rev. Res. **3**, 023244 (2021).

[56] A. Uvarov, J. Biamonte, and D. Yudin, Variational quantum eigensolver for frustrated quantum systems, Phys. Rev. B **102**, 075104 (2020).

[57] N. Klco, E. F. Dumitrescu, A. J. McCaskey, T. D. Morris, R. C. Pooser, M. Sanz, E. Solano, P. Lougovski, and M. J. Savage, Quantum-classical computation of Schwinger model dynamics using quantum computers, Phys. Rev. A **98**, 032331 (2018).

[58] D.-B. Zhang, B.-L. Chen, Z.-H. Yuan, and T. Yin, Variational quantum eigensolvers by variance minimization, Chin. Phys. B **31**, 120301 (2022).

[59] F. Zhang, N. Gomes, Y. Yao, P. P. Orth, and T. Iadecola, Adaptive variational quantum eigensolvers for highly excited states, Phys. Rev. B **104**, 075159 (2021).

[60] S. Liu, S.-X. Zhang, C.-Y. Hsieh, S. Zhang, and H. Yao, Probing many-body localization by excited-state variational quantum eigensolver, Phys. Rev. B **107**, 024204 (2023).

[61] C. Cade, L. Mineh, A. Montanaro, and S. Stanisic, Strategies for solving the Fermi-Hubbard model on near-term quantum computers, Phys. Rev. B **102**, 235122 (2020).

[62] R. Wiersema, C. Zhou, Y. de Sereville, J. F. Carrasquilla, Y. B. Kim, and H. Yuen, Exploring entanglement and optimization within the hamiltonian variational ansatz, PRX Quantum **1**, 020319 (2020).

[63] A. C. Y. Li, M. S. Alam, T. Iadecola, A. Jahin, J. Job, D. M. Kurkcuoglu, R. Li, P. P. Orth, A. B. Özgüler, G. N. Perdue, and N. M. Tubman, Benchmarking variational quantum eigensolvers for the square-octagon-lattice Kitaev model, Phys. Rev. Res. **5**, 033071 (2023).

[64] W. Kutzelnigg, Quantum chemistry in Fock space. I. The universal wave and energy operators, J. Chem. Phys. **77**, 3081 (1982).

[65] W. Kutzelnigg and S. Koch, Quantum chemistry in fock space. II. effective hamiltonians in fock space, J. Chem. Phys. **79**, 4315 (1983).

[66] W. Kutzelnigg, Quantum chemistry in Fock space. IV. The treatment of permutational symmetry. Spin-free diagrams with symmetrized vertices, J. Chem. Phys. **82**, 4166 (1985).

[67] R. J. Bartlett, S. A. Kucharski, and J. Noga, Alternative coupled-cluster ansätze II. The unitary coupled-cluster method, Chem. Phys. Lett. **155**, 133 (1989).

[68] A. G. Taube and R. J. Bartlett, New perspectives on unitary coupled-cluster theory, Int. J. Quantum Chem. **106**, 3393 (2006).

[69] T. Yanai, Y. Kurashige, D. Ghosh, and G. K.-L. Chan, Accelerating convergence in iterative solution for large-scale complete active space self-consistent-field calculations, Int. J. Quantum Chem. **109**, 2178 (2009).

[70] T. Yanai and G. K.-L. Chan, Canonical transformation theory for multireference problems, J. Chem. Phys. **124**, 194106 (2006).

[71] G. Harsha, T. Shiozaki, and G. E. Scuseria, On the difference between variational and unitary coupled cluster theories, J. Chem. Phys. **148**, 044107 (2018).

[72] S.-X. Zhang, J. Allcock, Z.-Q. Wan, S. Liu, J. Sun, H. Yu, X.-H. Yang, J. Qiu, Z. Ye, Y.-Q. Chen, C.-K. Lee, Y.-C. Zheng, S.-K. Jian, H. Yao, C.-Y. Hsieh, and S. Zhang, TensorCircuit: a quantum software framework for the NISQ era, Quantum **7**, 912 (2023).

[73] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.5.L032040 for details, including the following: (1) box plots, (2) analysis of the barren plateaus, (3) best VQE energy analysis, (4) numerical results with the ADAM optimizer and different decay steps, (5) numerical results with the stochastic gradient descent (SGD) optimizer, (6) numerical results with different rates of random gate activation, (7) training dynamics, (8) comparison with the initialization strategy, (9) robustness against quantum noise, (10) efficiency comparison for the plain and RA strategies, and (11) relation to the entanglement phase transition. References [81–88] are cited therein.

[74] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.

[75] S.-X. Zhang, C.-Y. Hsieh, S. Zhang, and H. Yao, Neural predictor based quantum architecture search, Mach. Learn.: Sci. Technol. **2**, 045027 (2021).

[76] S.-X. Zhang, C.-Y. Hsieh, S. Zhang, and H. Yao, Differentiable quantum architecture search, Quantum Sci. Technol. **7**, 045023 (2022).

[77] J. Li, X. Yang, X. Peng, and C.-P. Sun, Hybrid Quantum-Classical Approach to Quantum Optimal Control, Phys. Rev. Lett. **118**, 150503 (2017).

[78] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, Phys. Rev. A **98**, 032309 (2018).

[79] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, Phys. Rev. A **99**, 032331 (2019).

[80] L. Banchi and G. E. Crooks, Measuring analytic gradients of general quantum evolution with the stochastic parameter shift rule, Quantum **5**, 386 (2021).

[81] Y. Li, X. Chen, and M. P. A. Fisher, Quantum Zeno effect and the many-body entanglement transition, Phys. Rev. B **98**, 205136 (2018).

[82] B. Skinner, J. Ruhman, and A. Nahum, Measurement-Induced Phase Transitions in the Dynamics of Entanglement, Phys. Rev. X **9**, 031009 (2019).

[83] Y. Li, X. Chen, and M. P. A. Fisher, Measurement-driven entanglement transition in hybrid quantum circuits, Phys. Rev. B **100**, 134306 (2019).

[84] S. Choi, Y. Bao, X.-L. Qi, and E. Altman, Quantum Error Correction in Scrambling Dynamics and Measurement-Induced Phase Transition, Phys. Rev. Lett. **125**, 030505 (2020).

[85] M. J. Gullans and D. A. Huse, Dynamical Purification Phase Transition Induced by Quantum Measurements, Phys. Rev. X **10**, 041020 (2020).

[86] S.-K. Jian, C. Liu, X. Chen, B. Swingle, and P. Zhang, Measurement-Induced Phase Transition in the Monitored Sachdev-Ye-Kitaev Model, Phys. Rev. Lett. **127**, 140601 (2021).

[87] S. Liu, M.-R. Li, S.-X. Zhang, S.-K. Jian, and H. Yao, Universal Kardar-Parisi-Zhang scaling in noisy hybrid quantum circuits, Phys. Rev. B **107**, L201113 (2023).

[88] L. Bottou, *Stochastic gradient descent tricks, in Neural Networks: Tricks of the Trade, Lecture Notes in Computer Science* Vol. 7700 (Springer, Berlin, 2012), pp. 421–436.