

Dual role of low-weight Pauli propagation: A flawed simulator but a powerful initializer for variational quantum algorithms

Zong-Liang Li^{1,2} and Shi-Xin Zhang^{1,*}

¹*Institute of Physics, Chinese Academy of Sciences & Beijing National Laboratory for Condensed Matter Physics, Beijing 100190, China*

²*School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*

 (Received 18 August 2025; revised 4 January 2026; accepted 18 February 2026; published 9 March 2026)

Variational quantum algorithms are often hindered by rugged optimization landscapes. In this work, we investigate the low-weight Pauli propagation (LWPP) algorithm and find that it serves as an unreliable energy estimator for variational circuits. However, we reveal a counterintuitive insight: The Pauli-weight truncation acts as a spectral filter, effectively smoothing out high-frequency local minima while preserving the global basin of attraction in the landscape. We identify this mechanism as landscape alignment, where the approximate landscape becomes a superior navigator compared to the rugged exact landscape. Benchmarks across diverse spin models and molecular systems demonstrate that LWPP-initialized optimization yields order-of-magnitude improvements in accuracy, often finding solutions inaccessible to direct exact optimization. This work reframes LWPP from a flawed simulator into a vital preoptimizer that serves not only as a cheap classical substitute but also as an essential tool for addressing quantum optimization challenges.

DOI: [10.1103/4kyq-n8jb](https://doi.org/10.1103/4kyq-n8jb)

I. INTRODUCTION

Variational quantum algorithms (VQAs) represent a leading strategy for harnessing near-term quantum devices [1–27]. As hybrid quantum-classical methods, VQAs employ a parametrized quantum circuit to prepare a trial state and a classical optimizer to iteratively tune the circuit parameters, seeking to minimize a given cost function. The VQA optimization landscapes are notoriously difficult, plagued by barren plateaus [28–46] and extensive local minima [47,48]. This places immense importance on the choice of initial parameters, as a poor starting point can derail the entire optimization process.

Recent theoretical breakthroughs in the low-weight Pauli propagation (LWPP) algorithm [49,50] have shown that for typical random quantum circuits, expectation values can be estimated efficiently. These results, often based on average-case analysis over an ensemble of circuits [49–61], suggest the possibility of replacing quantum hardware with classical simulators for VQA evaluation. However, there is a fundamental difference between these theoretical frameworks and the reality of a VQA optimization. A VQA trajectory does not explore a random ensemble of circuits. Instead, its parameters evolve via directed updates, and average-case guarantees of classical simulability do not automatically apply to such a specific, nonrandom instance. This raises

a critical and practical question: What is the true utility of classical algorithms like LWPP in the context of VQA workflows?

In this work, we first demonstrate that LWPP is a poor estimator for the energy evaluation of VQAs, highlighting the limitations of applying average-case theories to single instances. However, from this failure, we uncover a counterintuitive inversion of this limitation: The numerical inaccuracy of LWPP is the very source of its optimization power. We demonstrate that the Pauli-weight truncation inherent to LWPP acts as a spectral filter, effectively smoothing out the high-frequency local minima that plague the exact quantum landscape. As a result, the approximate LWPP landscape is not merely a computationally cheaper surrogate but a topologically superior navigator. It robustly guides parameters into the global basin of attraction, avoiding the traps that frequently derail direct optimization on the exact, rugged landscape.

Leveraging this insight, we shift our perspective from using LWPP as a simulator to employing it as a classical preoptimizer. By running an initial, purely classical optimization with the LWPP cost function, we generate superior starting parameters for the main VQA optimization loop. Benchmarking this strategy on various quantum many-body models, including various lattice topologies, hierarchical ansätze [37,50,62–65], molecular ground states, and noisy environments, our results show an order-of-magnitude improvement in both final accuracy and convergence speed over standard heuristics [30], and this strategy often uncovers high-quality solutions inaccessible to direct quantum optimization. Our work thus reframes the role of LWPP: it is a flawed simulator but a powerful initializer and superior navigator. This practical insight offers a path to significantly accelerate VQA performance by offloading the exploratory optimization phase to classical

*Contact author: shixinzhang@iphy.ac.cn

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

computers, thereby reducing requirements on quantum computational resources.

II. THE LOW-WEIGHT PAULI PROPAGATION ALGORITHM

The LWPP algorithm operates in the Heisenberg picture by tracking the backward evolution of an observable O under the circuit's unitary transformation U . The simulation begins by decomposing O into a weighted sum of Pauli strings, $O = \sum_i c_i P_i$, where each P_i is a tensor product of n single-qubit Pauli operators $\{I, X, Y, Z\}$. It then propagates this sum gate by gate to compute the expectation value with respect to the initial state, $\text{Tr}[\rho_{\text{initial}}(U^\dagger O U)]$, where U represents the unitary evolution of the quantum circuit composed of a sequence of gates.

The complexity arises from non-Clifford rotation gates $R_G(\theta) = e^{-i\theta G/2}$, which cause a Pauli string P to branch if it does not commute with the generator G :

$$R_G(\theta)[P] = \cos(\theta)P + \sin(\theta)P', \quad \text{for } [P, G] \neq 0 \quad (1)$$

where $P' = i[G, P]/2$. This branching leads to an exponential growth in the number of terms.

To render the simulation tractable, LWPP applies a weight-truncation approximation. The weight of a Pauli string, $P = P_1 \otimes P_2 \otimes \dots \otimes P_n$, is defined as the number of its single-qubit Pauli operators P_i that are not the identity operator (I). At each step, any string whose weight exceeds a predefined cutoff k is discarded. This approximation reduces the time complexity to polynomial scaling $\mathcal{O}(dn^k)$ [49], where n is the number of qubits and d is the circuit depth. While this introduces approximation errors, particularly when branching coefficients ($\sin \theta$) are large, it provides a tunable trade-off between accuracy and computational cost.

III. A FLAWED ESTIMATOR BUT A USEFUL NAVIGATOR

Recent theoretical work has established that for typical quantum circuits, particularly those exhibiting locally scrambling properties, expectation values can be efficiently estimated classically via LWPP [49]. The accuracy is defined on average over an ensemble of circuits with randomly drawn parameters. However, a crucial question for practical applications remains: Do these average-case guarantees translate to the optimization of a single, specific VQA instance? To directly address this gap, we evaluate LWPP not on a random circuit ensemble, but along an explicit VQA optimization trajectory.

Here, we study the variational quantum eigensolver (VQE) applied to the two-dimensional Heisenberg XYZ model on a square lattice as an example,

$$H = \sum_{\langle i,j \rangle} (J_x X_i X_j + J_y Y_i Y_j + J_z Z_i Z_j), \quad (2)$$

where $J_x = 1.0$, $J_y = 0.8$, and $J_z = 0.5$. All simulations are implemented in TensorCircuit-NG [66], with circuit ansatz details provided in the Supplemental Material [67].

We consider two complementary scenarios: (1) Exact optimization + LWPP evaluation, where parameters are updated via the exact energy gradient while the LWPP estimate is

passively recorded; and (2) LWPP optimization + Exact evaluation, where the optimizer is guided solely by the LWPP approximation while the true energy is tracked for validation. Both random ($\theta \in [-\pi, \pi]$) and near-identity ($\theta \in 0.01 \times [-\pi, \pi]$) initializations are examined.

The results shown in Fig. 1 reveal a sharp distinction between average-case promises and single-instance performance. For random initialization, LWPP fails dramatically as a static estimator. As shown in Fig. 1(a), the LWPP-evaluated energy on the true optimization path diverges from the exact value and collapses to a nonphysical value near zero. This collapse occurs because the repeated truncation of terms branching from $\sin(\theta)$ [as shown in Eq. (1)] causes the estimate to decay exponentially with circuit depth. This confirms that LWPP cannot serve as a reliable single-instance cost estimator for VQE.

It is worth noting that this numerical failure does not preclude optimization utility. As shown in Figs. 1(b) and 1(d), minimizing the LWPP cost function consistently guides parameters toward regions of lower true energy. This suggests that despite its numerical inaccuracy, the LWPP landscape may capture sufficient structural features to act as an effective navigator toward high-quality basins of attraction. This insight reframes the role of LWPP from an unreliable simulator into a potential classical preoptimizer. Rather than seeking precise energy estimation, a task where average-case guarantees often fail for specific VQA instances, we propose leveraging the LWPP landscape to identify promising initial parameters. In the following, we benchmark this strategy on Heisenberg XYZ models to evaluate whether such preoptimization effectively improves the optimization accuracy and efficiency.

IV. LWPP-INITIALIZED OPTIMIZATION FROM RANDOM PARAMETERS

To test the navigation efficacy of LWPP, we compare (1) Direct random optimization, which performs 1500 steps of optimization using the exact cost function, against (2) a two-stage LWPP-initialized optimization protocol, which consists of 1500 classical preoptimization steps on the LWPP landscape followed by 1500 steps of exact VQA optimization. In both cases, the optimization is initiated from the same set of random parameters $\theta \in [-\pi, \pi]$.

We benchmark a 3×4 lattice (circuit depth $d = 6$) in Fig. 2. While direct optimization consistently traps parameters in suboptimal local minima [Fig. 2(a)], LWPP-initialized optimization robustly navigates to the global basin [Fig. 2(b)]. However, the internal dynamics in Figs. 2(c) and 2(d) reveal an underlying challenge: The LWPP energy signal is initially suppressed to magnitudes as low as 10^{-11} due to heavy truncation, creating a barren plateau-like landscape. Although the optimizer manages to escape at this intermediate depth, statistical benchmarks for $d > 8$ show that this signal loss eventually becomes insurmountable as circuit depth increases (see the Supplemental Material [67]). This limitation suggests that a near-identity initialization could be employed, which mitigates the gradient vanishing by keeping rotation angles small. We thus evaluate whether LWPP initialization is a powerful navigator when combined with this strong heuristic.

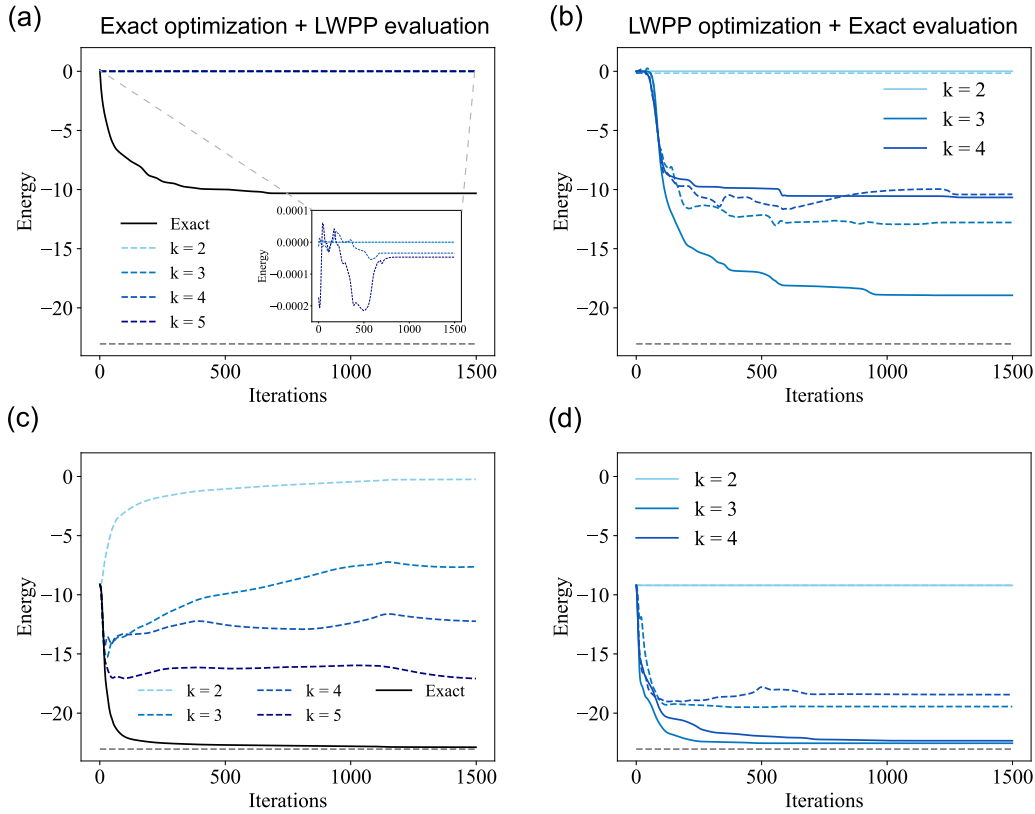


FIG. 1. Parameter-dependent accuracy of the LWPP estimation. Performance analysis on a 3×4 antiferromagnetic Heisenberg XYZ model with circuit depth $d = 4$ and varying truncation cutoffs k . The top row [panels (a) and (b)] starts from random initialization, while the bottom row [panels (c) and (d)] starts from near-identity initialization. (a), (c) Exact optimization + LWPP evaluation: The optimizer is driven by the exact energy gradient $\nabla_{\theta} E$ (solid black line), while the LWPP energy estimate E_{LWPP} (blue dashed lines) is passively recorded at each iteration to benchmark its accuracy. The inset in panel (a) provides a magnified view of the LWPP evaluations, which collapse to near-zero values. (b), (d) LWPP optimization + exact evaluation: The optimizer is driven by the approximate LWPP gradient $\nabla_{\theta} E_{\text{LWPP}}$ (solid colored lines) as the cost function. The true exact energy E (dashed colored lines) is passively recorded to verify the quality of the state. Despite the numerical inaccuracy, minimizing the LWPP cost robustly guides the true energy down, suggesting its potential utility as a navigator.

V. LWPP-INITIALIZED OPTIMIZATION FROM NEAR-IDENTITY PARAMETERS

We now benchmark two strategies, both starting from the same near-identity parameter set ($\theta \in 0.01 \times [-\pi, \pi]$). We also compare the standard (1) Direct near-identity optimization, which utilizes the exact cost function immediately, against our proposed (2) LWPP-initialized optimization. The latter employs a two-stage protocol where near-identity parameters are first preoptimized via the LWPP landscape to seed the subsequent exact VQA loop.

We first examine a representative example on a 3×4 antiferromagnetic Heisenberg XYZ model as shown in Fig. 3. As expected, the direct near-identity optimization performs well, achieving a final relative error of approximately 10^{-3} after 1500 iterations. However, the LWPP-initialized strategy proves to be even more effective. The initial LWPP preoptimization (with $k = 3$) quickly finds a parameter set with a relative error of around 10%. Starting from this improved point, the subsequent exact optimization converges to a final relative error of nearly 10^{-4} . This order-of-magnitude enhancement demonstrates that even when starting from a good heuristic, the LWPP preoptimization phase is still highly superior.

We then performed a statistical analysis across various lattice sizes (from 3×3 to 3×6) and circuit depths (from $d = 2$ to $d = 6$), reporting the decile-based accuracy (top 10% of runs). As shown in Figs. 4(a)–4(d), LWPP-initialized optimization consistently improves the final optimization accuracy across almost all tested configurations. Notably, the margin of improvement often grows with increasing circuit depth, a regime with the emergence of barren plateaus where finding high-fidelity solutions is more challenging.

Beyond final accuracy, LWPP-initialized optimization also offers a dramatic improvement in computational speed. To quantify this, we set a target accuracy defined by the final relative error achieved after 1500 iterations of the direct near-identity optimization. We then measure the relative steps required for the LWPP-initialized runs to reach this target, where the relative steps are calculated as the number of required iterations normalized by 1500. The results, shown in Figs. 4(e)–4(h), demonstrate that LWPP-initialized optimizations consistently reach the target accuracy in a small fraction of the iterations required by the direct method—often requiring only around 10% of the steps. This represents nearly an order-of-magnitude reduction in the number of costly queries to the quantum processor. In summary, our findings consistently demonstrate that LWPP initialization provides

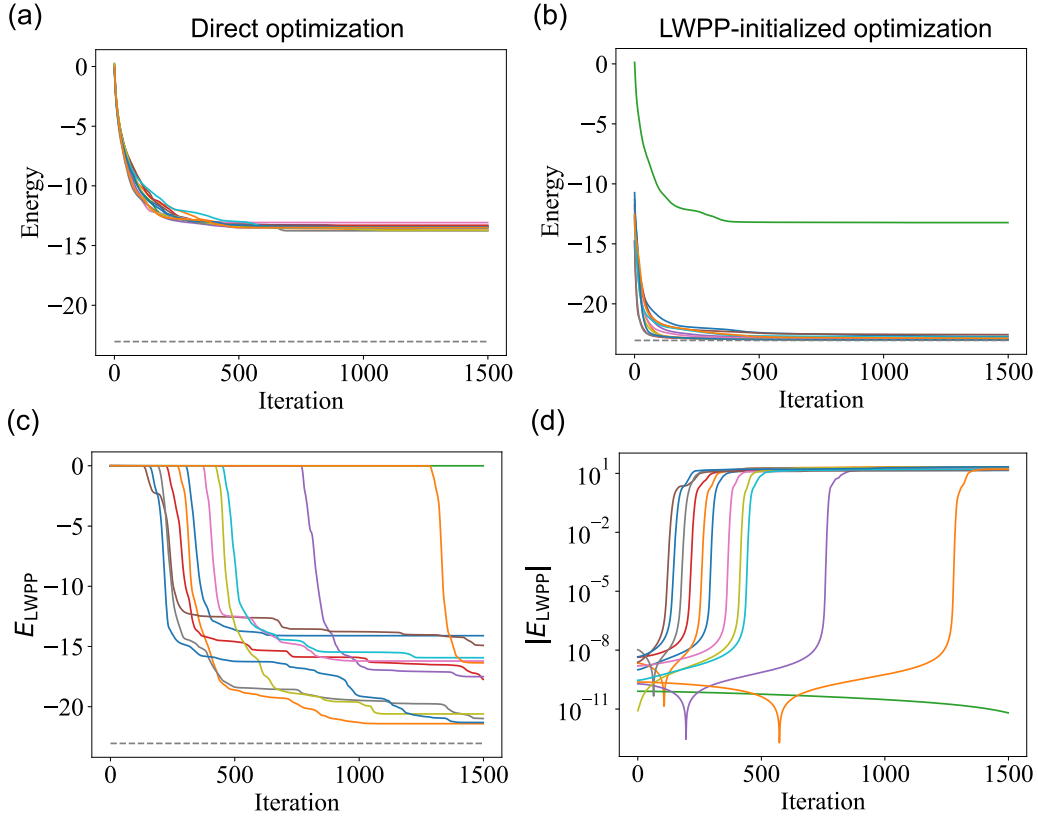


FIG. 2. An illustrative example of LWPP-initialized optimization from a random start. VQA optimization of a $d = 6$ VQA circuit on a 3×4 lattice, starting from the same random parameters. The figure contains 12 independent optimization runs, each initiated from a randomly sampled parameter set and represented by a different color. (a) Direct optimization consistently fails, converging to a high-energy local minimum. (b) LWPP-initialized optimization succeeds in finding a low-energy state. (c), (d) The corresponding LWPP preoptimization dynamics, with panel (d) showing the energy deviation from zero on a log scale.

a substantial advantage even when benchmarked against the strong near-identity initialization heuristic.

VI. LANDSCAPE ALIGNMENT

Given the remarkable and consistent improvements observed across various scenarios, a fundamental question arises regarding the mechanism behind this success. Here, we propose a physical picture to explain the empirical success as

landscape alignment. Standard VQA optimization landscapes are characterized by severe ruggedness. As illustrated in Fig. 5(a), the exact energy landscape is densely populated with high-energy local minima. This ruggedness fundamentally arises from the scrambling nature of deep quantum circuits, where the Heisenberg-evolved observable expands into a sum of exponentially many Pauli strings. High-weight Pauli terms, resulting from extensive operator scrambling, carry high-frequency parameter dependencies that generate narrow

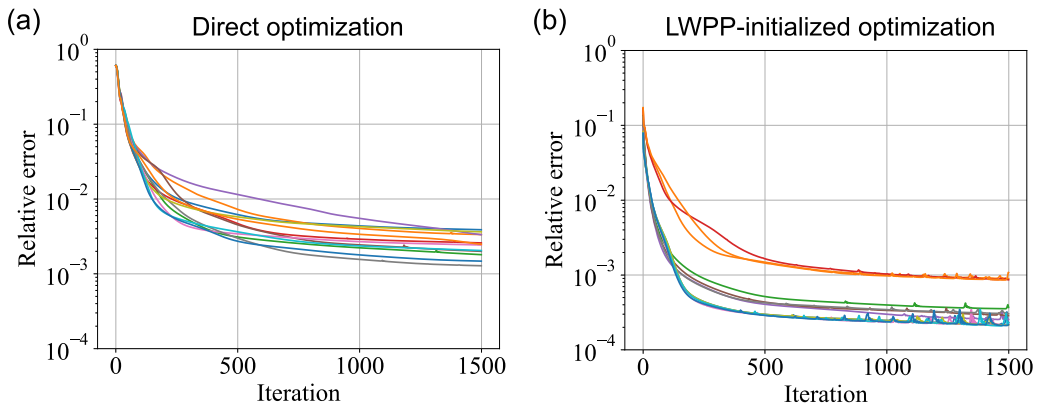


FIG. 3. A comparative example for LWPP from near-identity parameters. Comparison of optimization trajectories of a $d = 6$ circuit on a 3×4 lattice, both starting from the same near-identity parameter set. Direct optimization performs well [panel (a)], but the LWPP-initialized run [panel (b)], which includes a preoptimization phase, achieves a significantly better final accuracy.

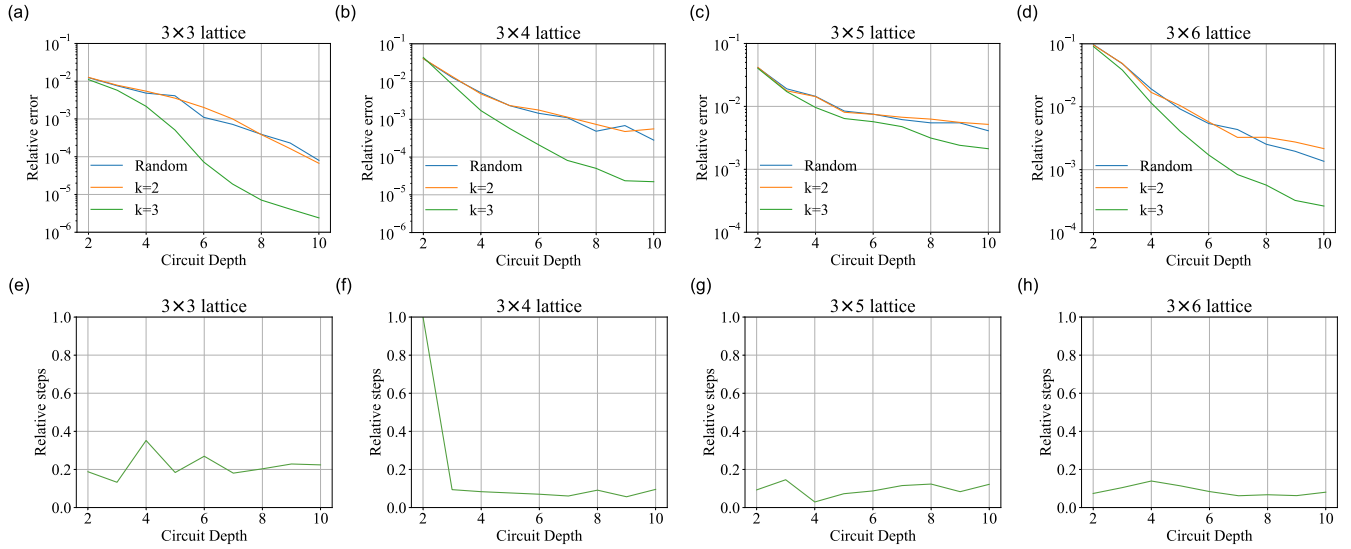


FIG. 4. Statistical comparison of initialization strategies with near-identity heuristic. Results for antiferromagnetic models on lattices ranging from 3×3 to 3×6 . (a)–(d) Final optimization accuracy after 1500 optimization steps: The final accuracy achieved with LWPP initialization with $k = 3$ (green line) is consistently superior to direct near-identity optimization, with the advantage often increasing with circuit depth. (e)–(h) Optimization speedup: The relative steps required for LWPP-initialized runs ($k = 3$) to reach the target accuracy, where the target is defined as the final accuracy reached by 1500 steps of direct near-identity optimization. The relative steps are calculated as the number of required iterations normalized by 1500. Lower values indicate faster convergence; LWPP-initialized optimization consistently achieves the target accuracy in a small fraction of the iterations required by the direct method, often providing a tenfold speedup.

local traps. Consequently, direct optimization ($\nabla_{\theta} E$) typically leaves parameters trapped in these suboptimal valleys.

LWPP instead acts as a spectral filter by truncating terms with higher Pauli weight $k > k_{\text{cutoff}}$. LWPP implicitly discards the fine-grained ruggedness, yielding a smoothed surrogate (E_{LWPP}) of the true energy landscape. As illustrated in Fig. 5, this filtering mechanism fundamentally alters the optimization dynamics. By removing the shallow local minima, the smoothed landscape E_{LWPP} robustly guides the parameters into the broad basin of attraction of the global minimum.

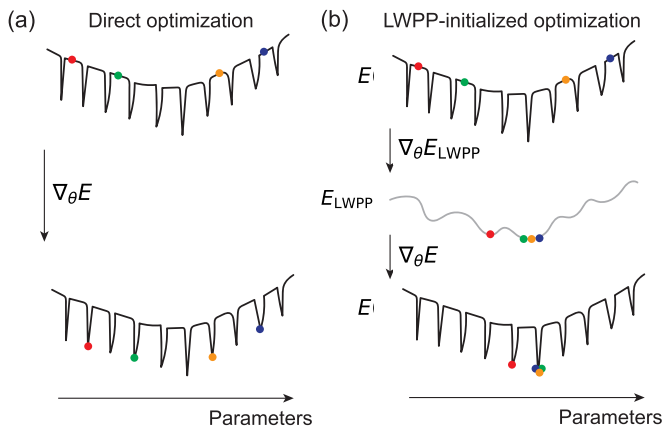


FIG. 5. Schematic of landscape alignment. (a) Direct optimization: This optimization typically traps parameters (colored dots) in high-energy local minima. (b) LWPP-initialized optimization: The exact landscape (black) is rugged due to high-frequency operator spreading. The LWPP approximation (gray) acts as a smoothed surrogate by filtering out high-weight terms. Optimization on this coarse-grained landscape ($\nabla_{\theta} E_{\text{LWPP}}$) avoids narrow traps and robustly guides parameters into a promising basin of attraction.

To substantiate this conceptual framework, we performed additional numerical investigations as detailed in the Supplemental Material [67]. Through a basin probing experiment, we observed that the high-quality optima of the exact landscape are indeed located deep within the basin of the LWPP landscape. Furthermore, we demonstrated that this navigation capability is unique to the LWPP approximation. A comparison with other classical approximations, such as matrix product states (MPS), shows that classical tractability alone does not guarantee a favorable optimization landscape: unlike LWPP, MPS preoptimization fails to guide parameters toward better basins in our benchmarks (see the Supplemental Material [67]).

VII. GENERALITY AND ROBUSTNESS

To establish the universality of our strategy, we performed extensive validations across diverse problem domains (see the Supplemental Material for details [67]). Beyond standard square lattices, we extended our benchmarks to the honeycomb lattice and other spin models as well as to molecular systems, confirming that the method's efficacy is independent of specific quantum many-body models. Notably, for hierarchical barren-plateau-free ansätze like the multiscale entanglement renormalization ansatz [65], LWPP-initialized optimization still achieved a fivefold reduction in optimization steps. Furthermore, under realistic noise where direct gradients are often obscured, the coarse-grained LWPP landscape preserved the navigation signal, maintaining a significant advantage over direct optimization. The consistent superior results observed across such a broad spectrum of scenarios strongly suggest that these successes are universally rooted in the mechanism of landscape alignment, where the smoothed

landscape consistently aligns with the true global basin regardless of model details.

Finally, we emphasize the nontriviality of our strategy with two key findings. First, the preoptimization captures essential high-dimensional parameter correlations, rather than merely identifying a favorable independent probability distribution. As rigorously verified in the Supplemental Material [67], re-initializing the optimization with parameters randomly resampled from the preoptimized distribution leads to a substantial degradation in final accuracy. Second, and most critically, we reveal that the approximate LWPP landscape is topologically superior to the exact quantum landscape for navigation. Rather than merely serving as a computationally cheaper classical proxy, the smoothed landscape actively filters out high-frequency local minima that entrap direct quantum optimization. This is evidenced by the fact that a hybrid protocol (e.g., 500 LWPP classical steps followed by 500 quantum steps) consistently yields lower energy solutions than a direct optimization of 1000 exact quantum steps. These results corroborate the physical picture of landscape alignment: The inaccurate classical surrogate effectively smooths the rugged optimization path, guiding the system into a global basin of attraction that is often inaccessible via the exact quantum landscape alone.

VIII. CONCLUSION AND DISCUSSION

In this work, we have established a paradoxical yet powerful role for the LWPP algorithm in VQA scenarios. We

demonstrated that despite its unreliability for directly simulating the VQA cost function, LWPP serves as an effective tool for preoptimizing VQA parameters, leading to significant improvements in both final optimization accuracy and convergence speed.

The success of this strategy can be attributed to the LWPP optimization landscape. The landscape is not numerically exact, but effectively guides parameters toward favorable basins of attraction, providing a high-quality starting point. We anticipate that this initialization method can become a standard and powerful tool in practical VQA workflows. By offloading the difficult exploratory phase of optimization to an efficient classical algorithm, LWPP-initialized optimization significantly reduces quantum-processor calls while yielding superior optimization accuracy, making the path to practical quantum advantage more attainable.

ACKNOWLEDGMENTS

We acknowledge helpful discussion with Shuo Liu and Hao-Kai Zhang. S.-X.Z. acknowledges the support from Quantum Science and Technology-National Science and Technology Major Project (No. 2024ZD0301700) and the National Natural Science Foundation of China (No. 12574546).

DATA AVAILABILITY

The data that support the findings of this article are openly available [68].

-
- [1] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
 - [2] 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).
 - [3] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, *arXiv:1411.4028*.
 - [4] 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).
 - [5] J. Li, X. Yang, X. Peng, and C.-P. Sun, Hybrid quantum-classical approach to quantum optimal control, *Phys. Rev. Lett.* **118**, 150503 (2017).
 - [6] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, *Phys. Rev. X* **7**, 021050 (2017).
 - [7] J.-G. Liu and L. Wang, Differentiable learning of quantum circuit born machines, *Phys. Rev. A* **98**, 062324 (2018).
 - [8] X. Yuan, S. Endo, Q. Zhao, Y. Li, and S. C. Benjamin, Theory of variational quantum simulation, *Quantum* **3**, 191 (2019).
 - [9] S. Khatri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles, Quantum-assisted quantum compiling, *Quantum* **3**, 140 (2019).
 - [10] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, *Nature (London)* **567**, 209 (2019).
 - [11] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, Training deep quantum neural networks, *Nat. Commun.* **11**, 808 (2020).
 - [12] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, *et al.*, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).
 - [13] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, *Rev. Mod. Phys.* **94**, 015004 (2022).
 - [14] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, *et al.*, The variational quantum eigensolver: A review of methods and best practices, *Phys. Rep.* **986**, 1 (2022).
 - [15] 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).
 - [16] J. Stokes, J. Isaac, N. Killoran, and G. Carleo, Quantum natural gradient, *Quantum* **4**, 269 (2020).
 - [17] Y.-Q. Chen, S.-X. Zhang, C.-Y. Hsieh, and S. Zhang, Non-Hermitian ground-state-searching algorithm enhanced by a variational toolbox, *Phys. Rev. A* **107**, 042418 (2023).
 - [18] 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).
 - [19] J. Miao, C.-Y. Hsieh, and S.-X. Zhang, Neural-network-encoded variational quantum algorithms, *Phys. Rev. Appl.* **21**, 014053 (2024).

- [20] S.-X. Zhang, J. Miao, and C.-Y. Hsieh, Variational post-selection for ground states and thermal states simulation, *Quantum Sci. Technol.* **10**, 015028 (2025).
- [21] H.-L. Huang, X.-Y. Xu, C. Guo, G. Tian, S.-J. Wei, X. Sun, W.-S. Bao, and G.-L. Long, Near-term quantum computing techniques: Variational quantum algorithms, error mitigation, circuit compilation, benchmarking and classical simulation, *Sci. China: Phys. Mech. Astron.* **66**, 250302 (2023).
- [22] L. Cheng, Y.-Q. Chen, S.-X. Zhang, and S. Zhang, Quantum approximate optimization via learning-based adaptive optimization, *Commun. Phys.* **7**, 83 (2024).
- [23] Z.-X. Shang, M.-C. Chen, X. Yuan, C.-Y. Lu, and J.-W. Pan, Schrödinger-Heisenberg variational quantum algorithms, *Phys. Rev. Lett.* **131**, 060406 (2023).
- [24] S. Zhang, Z. Wan, C. Hsieh, H. Yao, and S. Zhang, Variational quantum-neural hybrid error mitigation, *Adv. Quantum Technol.* **6**, 2300147 (2023).
- [25] Y.-Q. Chen and S.-X. Zhang, Superior resilience to poisoning and amenability to unlearning in quantum machine learning, [arXiv:2508.02422](https://arxiv.org/abs/2508.02422).
- [26] S.-X. Zhang and L. Wang, A unified variational framework for quantum excited states, [arXiv:2504.21459](https://arxiv.org/abs/2504.21459).
- [27] P. Zhang, Correcting a noisy quantum computer using a quantum computer, [arXiv:2506.08331](https://arxiv.org/abs/2506.08331).
- [28] 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).
- [29] M. Ostaszewski, E. Grant, and M. Benedetti, Structure optimization for parameterized quantum circuits, *Quantum* **5**, 391 (2021).
- [30] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, An initialization strategy for addressing barren plateaus in parametrized quantum circuits, *Quantum* **3**, 214 (2019).
- [31] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, and P. J. Coles, Effect of barren plateaus on gradient-free optimization, *Quantum* **5**, 558 (2021).
- [32] 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).
- [33] M. Cerezo and P. J. Coles, Higher order derivatives of quantum neural networks with barren plateaus, *Quantum Sci. Technol.* **6**, 035006 (2021).
- [34] C. Ortiz Marrero, M. Kieferová, and N. Wiebe, Entanglement-induced barren plateaus, *PRX Quantum* **2**, 040316 (2021).
- [35] 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).
- [36] 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).
- [37] 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).
- [38] X. Liu, G. Liu, J. Huang, H.-K. Zhang, and X. Wang, Mitigating barren plateaus of variational quantum eigensolvers, *IEEE Trans. Quantum Eng.* **5**, 3103719 (2024).
- [39] 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).
- [40] A. Arrasmith, Z. Holmes, M. Cerezo, and P. J. Coles, Equivalence of quantum barren plateaus to cost concentration and narrow gorges, *Quantum Sci. Technol.* **7**, 045015 (2022).
- [41] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, Connecting ansatz expressibility to gradient magnitudes and barren plateaus, *PRX Quantum* **3**, 010313 (2022).
- [42] Z. Holmes, A. Arrasmith, B. Yan, P. J. Coles, A. Albrecht, and A. T. Sornborger, Barren plateaus preclude learning scramblers, *Phys. Rev. Lett.* **126**, 190501 (2021).
- [43] J.-E. Kim, J. Kim, and D. Rosa, Universal effectiveness of high-depth circuits in variational eigenproblems, *Phys. Rev. Res.* **3**, 023203 (2021).
- [44] S. Liu, S.-X. Zhang, S.-K. Jian, and H. Yao, Training variational quantum algorithms with random gate activation, *Phys. Rev. Res.* **5**, L032040 (2023).
- [45] H.-K. Zhang, S. Liu, and S.-X. Zhang, Absence of barren plateaus in finite local-depth circuits with long-range entanglement, *Phys. Rev. Lett.* **132**, 150603 (2024).
- [46] M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes, and M. Cerezo, Barren plateaus in variational quantum computing, *Nat. Rev. Phys.* **7**, 174 (2025).
- [47] L. Bittel and M. Kliesch, Training variational quantum algorithms is NP-hard, *Phys. Rev. Lett.* **127**, 120502 (2021).
- [48] E. R. Anschuetz and B. T. Kiani, Quantum variational algorithms are swamped with traps, *Nat. Commun.* **13**, 7760 (2022).
- [49] A. Angrisani, A. Schmidhuber, M. S. Rudolph, M. Cerezo, Z. Holmes, and H.-Y. Huang, Classically estimating observables of noiseless quantum circuits, *Phys. Rev. Lett.* **135**, 170602 (2025).
- [50] P. Bermejo, P. Braccia, M. S. Rudolph, and M. Cerezo, Quantum convolutional neural networks are (effectively) classically simulable, *PRX Quantum* (2026), doi:10.1103/8qt9-72ts.
- [51] N. A. Nemkov, E. O. Kiktenko, and A. K. Fedorov, Fourier expansion in variational quantum algorithms, *Phys. Rev. A* **108**, 032406 (2023).
- [52] D. Aharonov, X. Gao, Z. Landau, Y. Liu, and U. Vazirani, A polynomial-time classical algorithm for noisy random circuit sampling, in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, edited by B. Saha and R. A. Servedio (Association for Computing Machinery (ACM), New York, 2023), pp. 945–957.
- [53] E. Fontana, M. S. Rudolph, R. Duncan, I. Rungger, and C. Cîrstoiu, Classical simulations of noisy variational quantum circuits, *npj Quantum Inf.* **11**, 84 (2025).
- [54] M. S. Rudolph, E. Fontana, Z. Holmes, and L. Cincio, Classical surrogate simulation of quantum systems with LOWESA, [arXiv:2308.09109](https://arxiv.org/abs/2308.09109).
- [55] G. González-García, J. I. Cirac, and R. Trivedi, Pauli path simulations of noisy quantum circuits beyond average case, *Quantum* **9**, 1730 (2025).
- [56] Y. Shao, F. Wei, S. Cheng, and Z. Liu, Simulating noisy variational quantum algorithms: A polynomial approach, *Phys. Rev. Lett.* **133**, 120603 (2024).
- [57] S. Lerch, R. Puig, M. S. Rudolph, A. Angrisani, T. Jones, M. Cerezo, S. Thanasilp, and Z. Holmes, Efficient quantum-enhanced classical simulation for patches of quantum landscapes, [arXiv:2411.19896](https://arxiv.org/abs/2411.19896).

- [58] A. Angrisani, A. A. Mele, M. S. Rudolph, M. Cerezo, and Z. Holmes, Simulating quantum circuits with arbitrary local noise using Pauli propagation, [arXiv:2501.13101](#).
- [59] T. Begušić, K. Hejazi, and G. K. Chan, Simulating quantum circuit expectation values by Clifford perturbation theory, *J. Chem. Phys.* **162**, 154110 (2025).
- [60] V. Martinez, A. Angrisani, E. Pankovets, O. Fawzi, and D. S. França, Efficient simulation of parametrized quantum circuits under nonunitary noise through Pauli backpropagation, *Phys. Rev. Lett.* **134**, 250602 (2025).
- [61] P. Rall, D. Liang, J. Cook, and W. Kretschmer, Simulation of qubit quantum circuits via Pauli propagation, *Phys. Rev. A* **99**, 062337 (2019).
- [62] I. Cong, S. Choi, and M. D. Lukin, Quantum convolutional neural networks, *Nat. Phys.* **15**, 1273 (2019).
- [63] Z. Liu, P.-X. Shen, W. Li, L.-M. Duan, and D.-L. Deng, Quantum capsule networks, *Quantum Sci. Technol.* **8**, 015016 (2023).
- [64] R. A. Bravo, K. Najafi, X. Gao, and S. F. Yelin, Quantum reservoir computing using arrays of Rydberg atoms, *PRX Quantum* **3**, 030325 (2022).
- [65] L. Cincio, J. Dziarmaga, and M. M. Rams, Multiscale entanglement renormalization ansatz in two dimensions: Quantum Ising model, *Phys. Rev. Lett.* **100**, 240603 (2008).
- [66] S.-X. Zhang, J. Allcock, Z.-Q. Wan, S. Liu, J. Sun, H. Yu, X.-H. Yang, J. Qiu, Z. Ye, Y.-Q. Chen, *et al.*, Tensorcircuit: A quantum software framework for the NISQ era, *Quantum* **7**, 912 (2023).
- [67] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/4kyq-n8jb> for numerical validation of the landscape alignment mechanism, gradient scaling analysis, extended benchmarks across diverse models and architectures, robustness analysis under noise, and details on circuit implementations and optimization workflows.
- [68] Z.-L. Li and S.-X. Zhang, GitHub (2025), https://github.com/ZongliangLi/lwpp_init.