

Stochastic Pauli-Path Simulator for Large-Scale Quantum Optimization

Kaining Zhang¹, Xinbiao Wang¹, Kunsheng Li¹, Qixin Zhang¹

Yuxuan Du^{1,2}, Min-Hsiu Hsieh³, Dacheng Tao¹

¹NTU, Singapore ²SPMS, NTU, Singapore ³Hon Hai Research Institute, Taipei

Abstract

Pauli-based simulation offers a promising route to large-scale classical simulation of quantum circuits in structured regimes, but existing truncation-based Pauli simulators are mainly designed for forward expectation estimation. When embedded in gradient-driven quantum optimization, these truncations can produce systematically biased gradients and distorted optimization trajectories. We propose the stochastic Pauli-path simulator (SPPS), a Pauli-propagation framework for large-scale quantum optimization. SPPS samples from the full Pauli-path space, uses importance reweighting to obtain unbiased estimates, and applies path automatic differentiation to estimate all gradient components from sampled paths. We prove that truncation can approximate objective values while retaining non-vanishing gradient error, whereas SPPS gives unbiased gradients with high-probability accuracy and convergence guarantees. Experiments on VQE pre-training up to 100 qubits and QNN pre-training up to 40 qubits show faithful optimization dynamics and improved accuracy-runtime trade-offs.

1 Introduction

Advances in classical simulation are indispensable to the progress of quantum computing [1, 2, 3]. Classical simulators provide practical platforms for designing quantum algorithms before large-scale quantum machines are available, and they impose increasingly stringent standards on claims of quantum advantage. More recently, Pauli-based simulation (PBS) methods have opened a complementary route to simulating quantum dynamics by propagating observables in the Heisenberg picture [4, 5, 6, 7, 8]. By tracking Pauli representations rather than full state vectors, these methods can remain effective in regimes that are difficult for tensor-network simulators, especially when low-magic structure rather than low entanglement is the relevant simplifying feature.

Despite this progress, prior PBS methods remain limited in optimization functionality. Most methods target fixed-circuit forward simulation, where the goal is to estimate one expectation value for one parameter setting. Quantum optimization is more demanding: the simulator must repeatedly estimate objectives and gradients along an adaptive trajectory, and small directional errors can accumulate over many updates. This creates a gap between large-qubit forward simulation and reliable large-qubit training simulation.

Truncation-based PBS methods, also known as Pauli-propagation simulators, are a natural first attempt to close this gap because they build differentiable surrogates by discarding Pauli paths according to coefficient magnitude, Fourier level, or Pauli weight. These criteria can control expectation-value error, but quantum optimization depends on gradients. A path with a small value coefficient may have a large derivative with respect to a circuit parameter, so deterministic truncation can remove terms that determine the descent direction. This limitation affects tasks such as pre-training variational quantum eigensolvers (VQEs), quantum neural networks (QNNs), variational state preparation, and circuit compilation [9, 10, 11, 12, 13, 14].

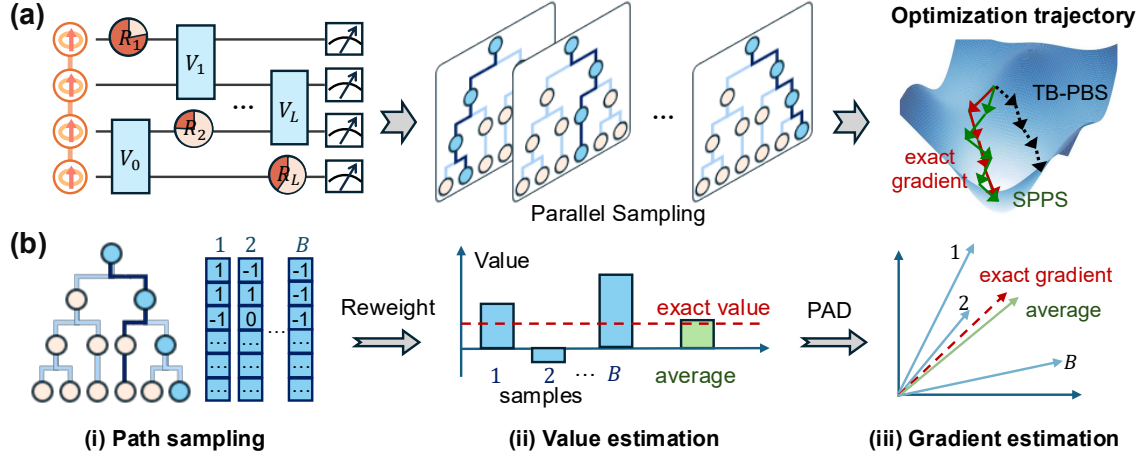


Figure 1: Workflow of SPPS. The simulator samples legal Pauli paths during Heisenberg propagation, corrects sampled contributions by importance reweighting, and applies path automatic differentiation to obtain unbiased stochastic gradients.

The central question is therefore whether Pauli-based simulation can support quantum optimization at scale without inheriting the structural bias of deterministic truncation. We answer positively by developing the stochastic Pauli-path simulator (SPPS). Instead of constructing a truncated subset of paths, SPPS dynamically samples propagation paths from the full path space and corrects every sampled contribution by importance reweighting. A path automatic differentiation (PAD) rule then converts each sampled path into an unbiased gradient contribution. This design changes the failure mode: approximation error is statistical and can be reduced by increasing the sample budget, rather than structural and biased by a fixed truncation rule.

2 Method

For a parameterized circuit $U(\theta)$ and observable O , the expectation value can be written as a Pauli-path expansion $f(\theta) = \text{Tr}[OU(\theta)\rho U(\theta)^\dagger] = \sum_{\omega \in \Omega} h_\omega(\theta)$, where Ω is the legal path set generated by Heisenberg propagation through parameterized Pauli rotations. Each path records the sequence of cosine, sine, and non-branching outcomes encountered while propagating the observable backward through the circuit. The theoretical contribution has two complementary parts: a negative result for deterministic truncation and a positive result for stochastic full-path sampling.

Theorem 1 (Gradient bias of truncation and guarantee of SPPS). *For any $\epsilon \in (0, 1)$, there exist circuits and parameter regions where a truncation-based Pauli simulator satisfies $|\hat{f}(\theta) - f(\theta)| \leq \epsilon$ while $\|\nabla \hat{f}(\theta) - \nabla f(\theta)\|_2 \geq 1 - \epsilon$; consequently, value accuracy alone does not guarantee faithful optimization. In contrast, SPPS samples from the full path space and produces unbiased estimators of both $f(\theta)$ and $\nabla f(\theta)$. With effective branching complexity $\kappa(\theta) = \prod_{j=1}^L (1 + |\sin(2\theta_j)|)$, it estimates the gradient to error ϵ with high probability using $\tilde{O}((1 + 2a)^L \kappa(\theta) L / (a\epsilon^2))$ samples, and the resulting stochastic gradient descent reaches an approximate stationary point under standard smoothness assumptions.*

This theorem separates the two failure modes. Truncation can be accurate for forward values while still deleting derivative-dominant paths, whereas SPPS keeps the full path space in expectation and turns approximation error into controllable sampling error.

Figure 1 summarizes SPPS. The algorithm has three steps. First, it samples legal Pauli paths sequentially without materializing the full propagation tree. At each branching parameterized rotation, the cosine branch is sampled with probability $q_j = (|\cos \theta_j| + a) / (|\cos \theta_j| + |\sin \theta_j| + 2a)$, where $a > 0$ prevents derivative-sensitive branches from receiving vanishing probability. Second, if a path ω is sampled with probability $\text{Pr}(\omega)$, its reweighted contribution $h_\omega(\theta) / \text{Pr}(\omega)$ is an unbiased estimator of $f(\theta)$. This

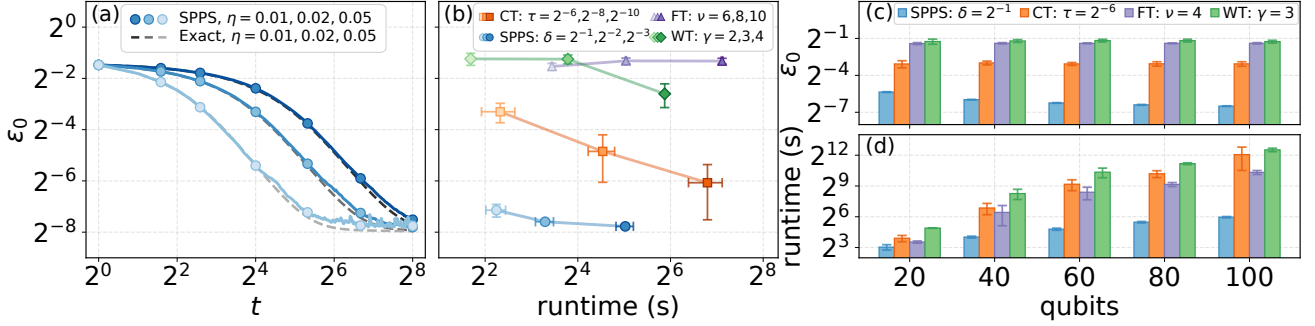


Figure 2: VQE pre-training results. SPPS tracks exact-gradient dynamics on small instances, achieves a better accuracy-runtime trade-off than coefficient, frequency, and Pauli-weight truncation, and scales favorably from 20 to 100 qubits.

importance correction is the key difference from pruning: rare paths are not discarded, but are assigned the correct statistical weight when sampled.

Third, PAD differentiates the trigonometric path coefficient directly. Cosine branches contribute score $-\tan \theta_j$, sine branches contribute score $\cot \theta_j$, and inactive positions contribute zero. Thus a single sampled path produces a vector-valued gradient contribution for all parameters touched by that path, rather than requiring separate shifted simulations for each parameter. The sampled paths are independent and can be evaluated in parallel; for multi-term observables, SPPS applies the estimator term by term and combines the results linearly. When $\kappa(\theta)$ remains moderate along the optimization path, SPPS gives an efficient stochastic simulator with controllable gradient error; when it grows exponentially, the cost reflects the intrinsic hardness of generic quantum simulation rather than a truncation artifact.

3 Experiments

We benchmark SPPS on quantum optimization tasks, focusing on VQE and QNN pre-training. Pre-training uses classical computation to find high-quality initial parameters, reducing the quantum resources needed for subsequent on-device training [15, 10, 16]. The VQE benchmark uses the one-dimensional transverse-field Ising model $H = -J \sum_{i=1}^{n-1} Z_i Z_{i+1} - g \sum_{i=1}^n X_i$ with a five-layer variational ansatz and system sizes up to 100 qubits [17, 18]. The QNN benchmark trains a 40-qubit hardware-efficient circuit on supervised labels generated by a target quantum circuit [19, 13, 14].

On a 15-qubit TFIM instance where exact gradients are available, SPPS closely follows exact gradient descent for learning rates $\eta \in \{0.01, 0.02, 0.05\}$. At $\eta = 0.05$, the normalized energy error decreases from above 2^{-2} to around 2^{-8} within about 64 steps. In contrast, truncation-based baselines often plateau or follow non-monotone trajectories even when less aggressive thresholds are used. Figure 2 reports the resulting accuracy-runtime comparison and scaling behavior.

SPPS consistently improves the trade-off between final error and runtime. On the 15-qubit TFIM with $g = 1.0$, tightening the SPPS adaptive gradient-error threshold reduces final energy error from about 2^{-7} to 2^{-8} while keeping runtime below one minute. The most accurate coefficient-truncation baseline takes roughly 2^7 – 2^8 seconds and reaches only about 2^{-6} error. Across field strengths $g \in \{0.6, 0.8, 1.0, 1.2, 1.4\}$, SPPS gives both lower error and shorter runtime; at $g = 0.6$, for example, it reaches 0.004 ± 0.001 error in 10.0 ± 3.7 seconds, compared with 0.077 ± 0.035 error in 48.4 ± 28.8 seconds for the strongest truncation baseline. Scaling from 20 to 100 qubits, SPPS maintains errors around 2^{-5} – 2^{-7} and becomes more than eight times faster than coefficient truncation at 100 qubits.

The QNN benchmark confirms that the benefit extends to quantum machine learning. SPPS reduces the training mean-squared error below 2^{-7} and reaches a test R^2 around 0.8 on the 40-qubit supervised task. Coefficient and Pauli-weight truncation saturate at higher losses with $R^2 \approx 0.4$, while frequency truncation fails to obtain a positive R^2 . SPPS also finishes in less than 2^9 seconds, compared with runtimes above 2^{11} seconds for the truncation baselines.

References

- [1] Igor L. Markov and Yaoyun Shi. Simulating quantum computation by contracting tensor networks. *SIAM Journal on Computing*, 38(3):963–981, 2008.
- [2] Sergey Bravyi and David Gosset. Improved classical simulation of quantum circuits dominated by clifford gates. *Phys. Rev. Lett.*, 116:250501, Jun 2016.
- [3] Xiaosi Xu, Simon Benjamin, Jianxin Chen, Jinzhao Sun, Xiao Yuan, and Pan Zhang. A herculean task: classical simulation of quantum computers. *Science Bulletin*, 70(23):4104–4112, 2025.
- [4] Xun Gao and Luming Duan. Efficient classical simulation of noisy quantum computation. *arXiv preprint arXiv:1810.03176*, 2018.
- [5] Tomislav Begušić and Garnet Kin-Lic Chan. Real-time operator evolution in two and three dimensions via sparse pauli dynamics. *PRX Quantum*, 6:020302, Apr 2025.
- [6] Yuguo Shao, Fuchuan Wei, Song Cheng, and Zhengwei Liu. Simulating noisy variational quantum algorithms: A polynomial approach. *Phys. Rev. Lett.*, 133:120603, Sep 2024.
- [7] Cheng-Ju Lin, Hrant Gharibyan, and Vincent P. Su. Utility-Scale Quantum State Preparation: Classical Training using Pauli Path Simulation. *Quantum*, 10:2014, March 2026.
- [8] Manuel S Rudolph, Tyson Jones, Yanting Teng, Armando Angrisani, and Zoë Holmes. Pauli propagation: A computational framework for simulating quantum systems. *arXiv preprint arXiv:2505.21606*, 2025.
- [9] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, Sep 2021.
- [10] Daniel J. Egger, Jakub Mareček, and Stefan Woerner. Warm-starting quantum optimization. *Quantum*, 5:479, June 2021.
- [11] Viacheslav V. Kuzmin and Pietro Silvi. Variational quantum state preparation via quantum data buses. *Quantum*, 4:290, July 2020.
- [12] Youle Wang, Guangxi Li, and Xin Wang. Variational quantum gibbs state preparation with a truncated taylor series. *Phys. Rev. Appl.*, 16:054035, Nov 2021.
- [13] Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4):043001, nov 2019.
- [14] Guillaume Verdon, Michael Broughton, Jarrod R McClean, Kevin J Sung, Ryan Babbush, Zhang Jiang, Hartmut Neven, and Masoud Mohseni. Learning to learn with quantum neural networks via classical neural networks. *arXiv preprint arXiv:1907.05415*, 2019.
- [15] Manuel S Rudolph, Jacob Miller, Danial Motlagh, Jing Chen, Atithi Acharya, and Alejandro Perdomo-Ortiz. Synergistic pretraining of parametrized quantum circuits via tensor networks. *Nature Communications*, 14(1):8367, 2023.
- [16] Abid Khan, Bryan K Clark, and Norm M Tubman. Pre-optimizing variational quantum eigensolvers with tensor networks. *arXiv preprint arXiv:2310.12965*, 2023.
- [17] Pierre Pfeuty. The one-dimensional ising model with a transverse field. *Annals of Physics*, 57(1):79–90, 1970.

- [18] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O’Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5(1):4213, Jul 2014.
- [19] Pablo Bermejo, Paolo Braccia, Manuel S. Rudolph, Zoë Holmes, Lukasz Cincio, and M. Cerezo. Quantum convolutional neural networks are effectively classically simulable. *PRX Quantum*, 7:020304, Apr 2026.