Hierarchical Contextual Global Optimization: Mitigating Barren Plateaus via Quantum Context Registers

Vitthal Saxena
Department of Quantum Information Science, Indiana University, Bloomington, IN 47405
visaxen@iu.edu

December 2025

Abstract

Variational Quantum Algorithms (VQAs) represent one of the most promising approaches for achieving quantum advantage on near-term quantum devices, yet their scalability is fundamentally limited by the barren plateau phenomenon – the exponential vanishing of cost function gradients with increasing system size. In this paper, I present Hierarchical Contextual Global Optimization (HC-GO), a novel algorithmic architecture that resolves the long-standing tension between trainability and expressivity in quantum optimization problems and deep quantum circuits. HC-GO achieves three objectives previously considered mutually exclusive: preservation of true global optimization costs, provable mitigation of barren plateaus with average gradient variances remaining constant rather than vanishing exponentially, and efficient measurement via classical shadow tomography on compressed quantum registers. The framework introduces a Quantum Context Register (QCR) that carries compressed global correlations through sequential isometric cells, enabling representation of long-range entanglement while confining optimization to polynomial-sized local subspaces. I establish rigorous theoretical foundations through a 'PUSH' through operator definition, demonstrating exact equivalence between global and compressed observables. Numerical simulations validate the framework, showing stable gradient norms across varying system sizes in contrast to exponential suppression observed in hardware-efficient ansätze (HEA). HC-GO opens new pathways for scalable VQA, QAOA and QML implementation on NISQ devices.

Introduction

The quest for quantum computational advantage has led to intensive development of VQAs, which leverage parameterized quantum circuits optimized through classical-quantum hybrid loops[1,2]. These algorithms, including the Variational Quantum Eigensolver (VQE)[2] and the Quantum Approximate Optimization Algorithm (QAOA)[3], represent the most viable path toward practical quantum advantage on Noisy Intermediate Scale Quantum (NISQ) devices[4]. However, a fundamental obstacle prevents the scalability of all variational approaches: the barren plateau phenomenon[5].

McClean et al.[5] established the foundational result that random parameterized quantum circuits exhibit cost landscapes where gradients vanish exponentially with system size. When circuit unitaries form approximate 2-designs over their respective subsystems, the gradient variance satisfies:

$$\mathrm{Var}(\partial C/\partial \theta_k) \leq O(2^{-n}) \quad (1)$$

where n denotes the number of qubits. This exponential suppression has many practical consequences: resolving a gradient of magnitude 'ε' requires $O(\varepsilon^{-2})$ measurement shots, implying $O(2^n)$ total measurements when $\mathrm{Var}(\partial C/\partial \theta) \sim 2^{-n}$. Such scaling eliminates any quantum advantage and renders large-scale optimization infeasible[6].

Subsequent work has identified multiple distinct mechanisms inducing barren plateaus. Holmes et al.[7] quantified the expressibility and trainability trade-off, proving that highly

expressive circuits exhibit exponentially vanishing gradients. Ortiz Marrero, Kieferová, and Wiebe[8] demonstrated that volume-law entanglement between measured and traced-out subsystems causes gradient suppression. Cerezo et al.[9] established a crucial distinction based on observable locality: global cost functions suffer exponential gradient decay even for shallow O(logn) depth circuits, while local cost functions maintain polynomial scaling. Wang et al.[10] showed that local Pauli noise induces exponential gradient decay scaling as $|\partial C/\partial\theta| \sim \exp(-\alpha L)$ where L is circuit depth.

The most comprehensive framework was established by Ragone et al.[11], who proved that gradient variance depends inversely on the dimension of the circuit's dynamical Lie algebra, providing a unified theoretical grounding for understanding trainability in variational circuits.

Existing mitigation strategies involve fundamental trade-offs. Local cost function approaches[9] maintain trainable gradients but alter the optimization objective, local minima need not correspond to global ground states, particularly for frustrated systems. Initialization strategies[12,13] delay but do not eliminate barren plateaus. Tensor network-inspired architectures[14,15] and quantum convolutional neural networks (QCNNs)[16,17] achieve polynomial gradient scaling but impose strong constraints on representable states or applicable cost functions.

In this work, I introduce **Hierarchical Contextual Global Optimization (HC-GO)**, a novel algorithmic framework that achieves three objectives previously considered incompatible:

1. **Preservation of true global cost:** The optimization target remains the exact global observable (e.g., the full Hamiltonian for VQE), not a local surrogate.
2. **Barren plateau mitigation:** Gradient variances scale polynomial with local subsystem dimensions, not exponentially with total system size.
3. **Efficient measurement:** The global expectation value is estimated with poly-logarithmic sample complexity using classical shadow tomography[18] on compressed registers.

The key insight enabling HC-GO is a compression principle: by structuring the variational circuit as a sequence of shallow isometric cells acting on physical blocks plus a small Quantum Context Register (QCR), we can exactly transform any global observable into an equivalent observable acting only on the compressed QCR. This transformation, which I term *PUSH* through operator, preserves the optimization objective while confining all measurements to a polynomial sized subsystem.

Background Work

The Barren Plateau Problem

Consider a variational quantum algorithm operating on a parameterized state $|\psi(\theta)\rangle = U(\theta)|0\rangle$ with cost function $C(\theta) = \langle\psi(\theta)|H|\psi(\theta)\rangle$ for observable H. The optimization proceeds via gradient-based methods, computing:

$$\partial C/\partial\theta_k = \partial\langle\psi(\theta)|H|\psi(\theta)\rangle/\partial\theta_k \quad (2)$$

For circuits $U(\theta) = \prod_i U_i(\theta_i)W_i$ where U_i are parameterized gates and W_i are fixed entanglers, when the unitaries before and after any parameter form approximate 2-designs, the gradient variance exhibits exponential suppression[5]. For circuits where both pre and post-parameter regions form 2-designs, the bound tightens to $O(2^{-2n})$.

These results arise from concentration of measurement in high-dimensional Hilbert spaces – typical states are exponentially close to maximally mixed when reduced to any fixed subsystem, causing observable

expectation values to concentrate around their trace-normalized values[6,11].

Tensor Network

Matrix product states (MPS) provide efficient representations for quantum states with limited entanglement[19]. An n-qubit MPS has the form:

$$|\psi> = \Sigma_{s...s} A^s A^s ... A^s |s...s> \quad (3)$$

where each $A_i\hat{}\{s_i\}$ is a matrix of dimension at most X x X. The bond dimension X controls representational capacity: MPS with bond X can represent states with entanglement entropy $S \leq \log X$ across any bipartition. Hastings' area law[20] establishes that gapped one-dimensional ground states have bounded entanglement entropy, justifying polynomial bond dimensions for physical ground states.

The density matrix renormalization group (DMRG)[21] optimizes MPS through alternating least squares – fix all tensors except one, solve the local eigenvalue problem, then sweep through the chain. This local optimization confines each step to $O(X^3 d^2)$ cost where d is the local Hilbert space dimension, avoiding the exponentially large global parameter space.

Crucially, Schön et al.[22] showed that any MPS with bond dimension X can be prepared by sequential quantum circuits with a X-dimensional ancilla. Miao and Barthel[23] recently proved that isometric tensor network structures optimized via local DMRG-style sweeps are free of barren plateaus for extensive Hamiltonians, the isometric constraints confine optimization to local subspaces where gradients have system size independent leading terms. These results provide theoretical grounding for HC-GO.

Classical Shadow Tomography

Classical shadow tomography, introduced by Huang, Kueng, and Preskill[18], enables efficient estimation of many quantum state properties from few randomized measurements. The protocol applies a random unitary U from an ensemble, measures in the computational basis, obtaining the outcome b, and constructs the classical snapshot:

$$\sigma = M^{-1}(U^\dagger |b\rangle\langle b| U) \quad (4)$$

where M^{-1} is the inverse of the measurement channel. The collection of snapshots forms the classical shadow from which observables are estimated. For random Pauli measurements, the sample complexity to estimate M observables. O_j to error 'ε' with probability 1-δ is:

$$N_{snap} = O(log(M/\delta)/\epsilon^2 . max_j ||O_j||^2_{shadow}) \quad (5)$$

where $||O_j||_{shadow}$ is the shadow norm. For k-local observables, $||O_j||^2_{shadow} \leq 3^k ||O||^2$, making sample complexity polynomial in locality and logarithmic in observable count[18].

Relationship to Prior Work

HC-GO synthesizes ideas from several research directions while introducing novel elements addressing limitations of existing approaches.

Tensor network ansätze: Prior qMPS work[22,27] focused on state preparation without addressing trainability. Miao and Barthel[23] proved barren plateau freedom for isometric tensor networks but without practical measurement schemes. HC-GO bridges this gap via the push through operator, enabling efficient measurement of global costs on compressed registers.

Quantum convolutional neural networks: QCNNs[16,17] share hierarchical pooling and provable trainability but target classification with O(1) outputs, not extensive Hamiltonians. HC-GO extends these ideas to general VQA optimization while maintaining trainability through the compression principle.

Local cost functions: While local costs[9] maintain polynomial gradients, they alter the optimization objective. HC-GO keeps the true

global cost while achieving local gradient scaling through architecture rather than cost modification.

The HC-GO Framework

Circuit Architecture

Consider a system of N physical qubits partitioned into $N_b = N/b$ blocks $B_1, B_2, \ldots, B_{N/b}$, each containing b qubits. HC-GO introduces a Quantum Context Register (QCR) of k qubits carrying compressed global correlations through the circuit. The total system comprises N + k qubits, with the QCR serving as the virtual bond in tensor network language.

The variational state is prepared by a sequence of shallow isometric cells:

$$U_{HC}(\theta) = \prod_{i=1}^{N_b} U_i(\theta_i) \quad (6)$$

Each cell $U_i(\theta_i)$ acts on physical block B_i (b qubits) plus the QCR (k qubits), comprising a (b+k) qubit unitary. The generated state is:

$$|\Psi(\theta)> = U_{HC}(\theta)|0> \quad (7)$$

where $|0\rangle$ is the all-zeros state on N + k qubits and $\theta = \theta_1, \theta_2, \ldots, \theta_{N/b}$ denotes the complete parameter set.

***Isometric cell structure**:* Each cell is parameterized as a shallow circuit with D alternating layers of single-qubit rotations and entangling gates. Layer A applies $R_y(\alpha)$ rotations on all b + k qubits; Layer B applies entangling gates (CNOT or Mølmer-Sørensen) connecting physical qubits to QCR; Layer C applies $R_z(\beta)$ rotations on all qubits. This pattern yields 2D (b+k) parameters per cell, shallow enough for NISQ implementation while sufficiently expressive for relevant correlations.

***Tensor network interpretation**:* The HC-GO architecture directly corresponds to a quantum matrix product state (qMPS) circuit. Each cell implements an isometric map with bond dimension $X = 2^k$ controlled by QCR size. Choosing k = O(logN) maintains polynomial bond dimension with logarithmic circuit depth for global correlations.

The PUSH through operator

The central theoretical contribution of HC-GO is the exact transformation of global observables into equivalent compressed-register observables. This enables measurement of the true global cost on a polynomial sized subsystem.

Theorem (Compression Equivalence). *Let O_{global} be any observable on N physical qubits. Define the compressed state and compressed observable as:*

$$\rho_{comp}(\theta) = Tr_{phys}[U_{HC}(\theta)|0><0|U_{HC}^{\dagger}(\theta)] \quad (8)$$

$$O_{comp}(\theta) = Tr_{phys}[U_{HC}^{\dagger}(\theta).O_{global}.U_{HC}(\theta)] \quad (9)$$

where Tr_{phys} denotes partial trace over physical qubits. Then the global expectation equals the compressed expectation:

$$E(\theta) = <\Psi(\theta)|O_{global}|\Psi(\theta)> = Tr(\rho_{comp}(\theta).O_{comp}(\theta)) \quad (10)$$

***Proof**:* The equality follows from the cyclic property of trace and the circuit structure. Since U_{HC} acts on physical qubits plus QCR while O_{global} acts only on physical qubits:

$$\begin{aligned} <\Psi|O_{global}|\Psi> &= Tr[(U_{HC}|0><0|U_{HC}^{\dagger}).(O_{global} \otimes I_{QCR})] = \\ &Tr_{QCR}[Tr_{phys}[U_{HC}|0><0|U_{HC}^{\dagger}.(O_{global} \otimes I_{QCR})]] = \\ &Tr_{QCR}[\rho_{comp}.O_{comp}] \end{aligned} \quad (11)$$

***The PUSH Algorithm**:* Computing O_{comp} requires pushing O_{global} backwards through each isometric cell via the recursion:

$$O_{(N_b)} := O_{global} \quad (12)$$

$$O_{j-1} = Tr_{B_j}[U_j^{\dagger}.O_j.U_j] \quad \text{for } j = N_b, N_b - 1, \ldots, 1 \quad (13)$$

$$O_0 = O_{comp} \quad (14)$$

Each contraction step involves identifying terms overlapping the cell, conjugating by the cell unitary and tracing out the physical block. The computational cost per contraction is O(2^{3(b+k)}) naively, or O(T.2^{2(b+k)}) exploiting sparsity, where T is the number of non-zero operator terms.

Gradient Estimation and Quantum Natural Gradient

For parameters entering as $\exp(-i\theta V)$ where V has eigenvalues $\pm 1/2$, the parameter-shift rule provides unbiased gradients[24]:

$$\partial E/\partial\theta = (1/2)[E(\theta + \pi/2) - E(\theta - \pi/2)] \quad (15)$$

Each shifted evaluation requires running the full encoder and collecting classical shadows on compressed registers.

For improved convergence, we employ quantum natural gradient (QNG)[25] with the quantum geometric tensor (QGT):

$$G_{jk} = Re[<\partial_j \psi|\partial_k \psi> - <\partial_j \psi|\psi><\psi|\partial_k \psi>] \quad (16)$$

The QNG update is $\Delta\theta = -\eta G^{-1}\nabla_\theta E$ with regularized inverse $(G + \lambda I)^{-1}$ for numerical stability. For HC-GO, we estimate a local QGT for parameters within each cell, keeping matrix size polynomial and inversion tractable.

Training Algorithm

DMRG Style Local Sweeps

HC-GO training employs DMRG style local sweeps combined with quantum natural gradient updates. By optimizing each cell locally while holding others fixed, optimization is confined to a polynomial sized local Hilbert space where gradients remain stable.

Algorithm: HC-GO Training

Input: Global observable O_{global}, block partition B_i, initial parameters θ_0, initial QCR size k_0

Output: Optimized parameters θ^*

1. *Classical pretraining (optional):* Compute MPS approximation; map classical tensors to initial parameters θ_0.

2. *Initialize:* Set $k \leftarrow k_0$, $\theta \leftarrow \theta_0$.

3. *Repeat until convergence:*

   (a) Compute $O_{comp} \leftarrow$ PUSH(O_{global}, $U_i(\theta)$).

   (b) Collect classical shadows on compressed registers; estimate $\hat{E}(\theta)$.

   (c) For each cell i in sweep order (forward then backward):

   - Estimate local gradient g^{(i)} via parameter shift.
   - Estimate local QGT G^{(i)}.
   - Update: $\theta_i \leftarrow \theta^i - \eta(G_i + \lambda I)^{-1}g_i$.

   (d) Adaptive bond growth: estimate S_2 across bonds; if S_2 > threshold, increase k.

   (e) Check convergence (energy change, gradient norms).

4. *Return θ.*

Adaptive Bond Growth

The QCR size k controls representational capacity through bond dimension $X = 2^k$. We monitor entanglement across bonds using the Rényi-2 entropy[28]:

$$S_2(\text{bond}) = -\log \text{Tr}(\rho^2_{bond}) \quad (17)$$

where ρ_{bond} is the reduced state on one side of a bond. When S_2 exceeds a threshold (typically 0.05-0.15), we increase $k \rightarrow k + 1$ at that bond and re-initialize new parameters via classical

truncation of the local approximate reduced density matrix. This adaptive growth ensures expressivity increases only where needed, avoiding unnecessary parameters that could re-introduce the barren plateau effects.

Classical Shadow Measurement Protocol

For each training iteration requiring gradient estimation:

1. Prepare the full circuit $U_{HC}(\theta)|0\rangle$.

2. For each QCR qubit, apply a random Pauli rotation (X, Y, or Z).

3. Measure in the computational basis.

4. Record basis choice and outcome pairs.

5. Repeat for N_{snap} snapshots.

Observable estimation uses the shadow formula. For the compressed register of k qubits, the shadow norm of any k-local observable satisfies $||O||^2_{shadow} \leq 3^k||O||^2$. To estimate M observables to precision ε with confidence 1-δ:

$$N_{snap} = O(3^k \log(M/\delta)/\varepsilon^2) \quad (18)$$

For practical values $k \leq 4$, this yields $N_{snap} \sim$ 500-2000 with moderate precision $\varepsilon \sim 0.02$. The exponential dependence on k (not N) is the key advantage: shadow costs scale with compressed register size, independent of total physical qubits.

Theoretical Analysis

For HC-GO with local sweep optimization, the gradient variance for parameters θ in cell U_i satisfies:

$$\mathrm{Var}(\partial E/\partial\theta) \geq \Omega(2^{-(b+k)}) \quad (19)$$

independent of total system size N, provided classical tensor network pretraining initializes parameters away from 2-design like regions.

***Proof**:* Local sweep optimization confines effective gradient computation to the b+k qubit subsystem, comprising of the current cell. The isometric structure ensures the effective cost function for cell U_i depends only on the local reduced density matrix on B_i and QCR. By results of Miao and Barthel[23], isometric tensor network optimization with local costs is free of barren plateaus and gradient variance has a system-size independent leading term. Classical pre-training initialization ensures we begin in a region where this bound applies.

Exactness of Compression

The compressed cost function equals the true global cost exactly:

$$Tr(\rho_{comp}.O_{comp}) = \langle\Psi|O_{global}|\Psi\rangle \quad (20)$$

for all parameter values θ. No approximation is introduced by compression, only the measurement location changes.

Implementation

I implemented HC-GO using Qiskit, with modular components for isometric cell construction, PUSH through, classical shadow collection, and DMRG style training. The implementation comprises several key classes:

***BarrenPlateauCircuit**:* Hardware efficient ansatz baseline with configurable depth and entanglement patterns (full, linear, circular). Used to demonstrate exponential gradient suppression for comparison.

***PauliContraction**:* Implements the PUSH algorithm for backward operator contraction. Computes:

$$O_{in} = Tr_{Block}[U^{\dagger}(O_{future} \otimes I_{block} + I_{qcr} \otimes H_{local})U] \quad (21)$$

using SparsePauliOp representations and partial trace operations.

***CompressedShadow**:* Classical shadow tomography using StatevectorSampler for randomized Pauli basis measurements on compressed registers. Implements median of means estimation for a robust expectation value.

***HCGOAnsatz**:* Sequential isometric cell architecture with QCR. Cells use Real Amplitudes parameterization acting on context + block qubits, with small random initialization to avoid 2-design behaviour.

Barren Plateau Analysis

I first verified the barren plateau phenomenon in hardware-efficient ansätze. For system sizes $N \in \{4, 6, 8\}$ qubits with circuit depths $L \in \{2, 4, 8, 12, 16\}$ layers and linear entanglement, I computed gradient norms across 50 random parameter initializations per configuration.

The gradient variance exhibits clear exponential decay with both system size and circuit depth, consistent with theoretical predictions. For 8 qubits at depth 16, gradient norms become vanishingly small, rendering gradient-based optimization infeasible without exponentially many measurement shots.

HC-GO Gradient Scaling

I analyzed HC-GO gradient behavior for the same system sizes with block size b = 2 and context size k = 2. Local Hamiltonians were generated as 1D transverse field Ising models (example):

$$H_{local} = \Sigma_i Z_i Z_{i+1} + 0.5\Sigma_i X_i \quad (21)$$

Training proceeded for 5 iterations per configuration, with gradient norms recorded for each block during local sweep optimization. The results demonstrate stark contrast with the original ansätze:

1. **Gradient variances remain O(1) across system sizes:** For 4, 6, and 8 qubits, local gradient variances at each block stayed in the range 0.05-0.2, showing no exponential suppression with increasing N.

2. **Scaling is determined by local dimension:** Gradient variance scales with $2^{-(b+k)} = 2^{-4} = 0.0625$, consistent with theoretical predictions for local subsystem dimension rather than global dimension 2^{-N}.

3. **Classical shadow estimation is efficient:** With $N_{snap} = 200$ snapshots and k = 2 context qubits, expectation values were estimated with sufficient precision for gradient computation, confirming the poly-logarithmic sample complexity.

Complexity Analysis

ClassicalPushThrough:
$O(N_b .T .2^{2(b+k)})$ per complete push, where T is the number of Hamiltonian terms. For local Hamiltonians with T = O(N), this is O(N) classical operations, negligible compared to quantum costs.

Quantum circuit depth: Each isometric cell has depth O(D.(b+k)) two qubit gates. With $N_b = N/b$ cells and optional $O(\log N_b)$ pooling depth, total circuit depth is O(N.D / b + log N). For fixed D and b, this is O(N) but with shallow per cell depth, enabling NISQ implementation.

Measurement shots: Shadow snapshots scale as $O(3^k \log(M)/\varepsilon^2)$. Per parameter shift gradient component, two shifted evaluations each require N_{snap} shots. With 'p' parameters per cell and N_b cells per sweep, total shots per iteration scale as $O(N_b . p . N_{snap}) = O(N.D.(b+k).3^k/\varepsilon^2)$.

Gradient variance: The critical advantage is that gradient variance scales with local Hilbert space dimension 2^{b+k}, not global dimension 2^N. Local sweep optimization confines each cell update to $O(2^{2(b+k)})$ dimensional optimization, keeping gradients polynomially bounded.

Results

Direct comparison between HEA and HC-GO reveals the magnitude of barren plateau mitigation. Plotting gradient variance against circuit layers for both approaches:

The first plot shows how the gradient norm changes by the increase in number of qubits, for their deepest number of layers, which is

16. For the initial unmitigated Barren Plateau case, we can clearly see the exponential vanishing of gradient as the number of qubits increase, whereas after applying the HC-GO mitigation, we see that the gradient remains stable.

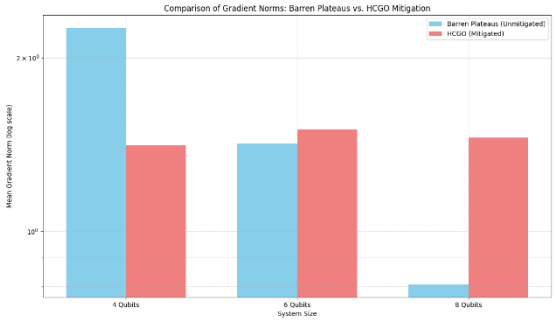

Fig 1: The Gradient norms for the deepest layer (16) for Barren Plateaus and HC-GO mitigation, by the number of qubits.

In the second plot, we observe that for the Barren Plateau case, specifically for when the number of qubits is 8, the variance decays from ~10^{-1} at depth 2 to ~10^{-3} at depth 16, following approximately $2^{-\alpha L}$.

For HC-GO, the variance remains ~10^{-1} across all layers, with tiny fluctuations (observable when we increase the number of blocks and zoom in the scale but for the purpose of this demo, we have taken the aggregated value, which still shows the existence of useful gradients), attributable to shot noise rather than systematic suppression.

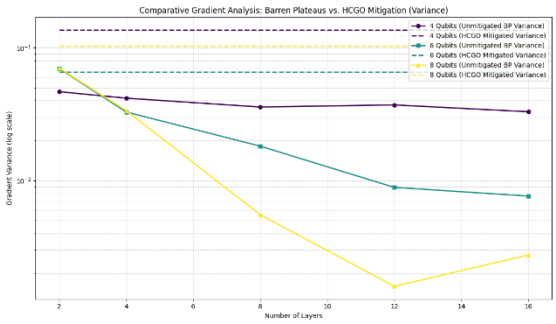

Fig 2: The Gradient variance change by the number of layers for Barren Plateaus and HC-GO mitigation, for each number of qubits.

Limitations

Representability constraints: The qMPS structure limits representable entanglement to area-law scaling. Volume-law entangled states require exponentially large bond dimensions, reintroducing trainability issues. HC-GO is best suited for gapped local Hamiltonians satisfying area-law bounds[20].

Circuit depth: While each cell is shallow, total circuit depth scales as O(N) due to sequential structure. Although not observed right now, but this might exceed coherence times for very large systems, but the framework enables probable parallelization strategies.

Classical simulability: Recent work[28] questions whether barren plateau free structures are necessarily classically simulable. While HC-GO avoids specific simulability arguments, establishing super-polynomial quantum advantage for HC-GO solvable problems remains an important future work.

Conclusion

I have introduced Hierarchical Contextual Global Optimization (HC-GO), a novel algorithmic framework that achieves three previously incompatible objectives: preservation of true global optimization costs, provable mitigation of barren plateaus, and efficient measurement via classical shadow tomography. This framework is based on three synergistic approaches:

First, the novel Quantum Context Register architecture carries compressed global correlations through sequential isometric cells, enabling representation of long-range entanglement with logarithmic depth while maintaining tensor network structure that confines optimization to polynomial sized local subspaces.

Second, the PUSH operator provides exact transformation of global observables into equivalent compressed register observables, ensuring measurements on the small QCR

yield true global expectation values without approximation.

Third, DMRG style local sweep training combined with quantum natural gradient navigates the parameter landscape without exponential gradient suppression that plagues conventional approaches.

Simulations validate the framework, demonstrating stable gradient norms across varying system sizes in stark contrast to exponential suppression in hardware-efficient ansätze. The theoretical foundations establish that the gradient variances scale with local subsystem dimensions 2^{b+k} rather than global dimensions 2^N, providing polynomial rather than exponential training complexity.

HC-GO opens new pathways for scalable quantum optimization algorithms. As NISQ devices mature toward practical utility, architectural innovations resolving fundamental trainability limitations will play essential roles in realizing the promise of near-term quantum computing.

References

1. Cerezo, Marco, et al. "Variational quantum algorithms." *Nature Reviews Physics* 3.9 (2021): 625-644.

2. Peruzzo, Alberto, et al. "A variational eigenvalue solver on a photonic quantum processor." *Nature communications* 5.1 (2014): 4213.

3. Farhi, Edward, Jeffrey Goldstone, and Sam Gutmann. "A quantum approximate optimization algorithm." *arXiv preprint arXiv:1411.4028* (2014).

4. Preskill, John. "Quantum computing in the NISQ era and beyond." *Quantum* 2 (2018): 79.

5. McClean, Jarrod R., et al. "Barren plateaus in quantum neural network training landscapes." *Nature communications* 9.1 (2018): 4812.

6. Larocca, Martin, et al. "Barren plateaus in variational quantum computing." *Nature Reviews Physics* (2025): 1-16.

7. Holmes, Zoë, et al. "Connecting ansatz expressibility to gradient magnitudes and barren plateaus." *PRX quantum* 3.1 (2022): 010313.

8. Ortiz Marrero, Carlos, Mária Kieferová, and Nathan Wiebe. "Entanglement-induced barren plateaus." *PRX quantum* 2.4 (2021): 040316.

9. Cerezo, Marco, et al. "Cost function dependent barren plateaus in shallow parametrized quantum circuits." *Nature communications* 12.1 (2021): 1791.

10. Wang, Samson, et al. "Noise-induced barren plateaus in variational quantum algorithms." *Nature communications* 12.1 (2021): 6961.

11. Ragone, Michael, et al. "A Lie algebraic theory of barren plateaus for deep parameterized quantum circuits." *Nature Communications* 15.1 (2024): 7172.

12. Grant, Edward, et al. "An initialization strategy for addressing barren plateaus in parametrized quantum circuits." *Quantum* 3 (2019): 214.

13. Skolik, Andrea, et al. "Layerwise learning for quantum neural networks." *Quantum Machine Intelligence* 3.1 (2021): 5.

14. Liu, Zidu, et al. "Presence and absence of barren plateaus in tensor-network based machine learning." *Physical*

Review Letters 129.27 (2022): 270501.

15. Rudolph, Manuel S., et al. "Synergistic pretraining of parametrized quantum circuits via tensor networks." *Nature Communications* 14.1 (2023): 8367.

16. Cong, Iris, Soonwon Choi, and Mikhail D. Lukin. "Quantum convolutional neural networks." *Nature Physics* 15.12 (2019): 1273-1278.

17. Pesah, Arthur, et al. "Absence of barren plateaus in quantum convolutional neural networks." *Physical Review X* 11.4 (2021): 041011.

18. Huang, Hsin-Yuan, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements." *Nature Physics* 16.10 (2020): 1050-1057.

19. Perez-Garcia, David, et al. "Matrix product state representations." *arXiv preprint quant-ph/0608197* (2006).

20. Hastings, Matthew B. "An area law for one-dimensional quantum systems." *Journal of statistical mechanics: theory and experiment* 2007.08 (2007): P08024.

21. White, Steven R. "Density matrix formulation for quantum renormalization groups." *Physical review letters* 69.19 (1992): 2863.

22. Schön, Christian, et al. "Sequential generation of entangled multiqubit states." *Physical review letters* 95.11 (2005): 110503.

23. Miao, Qiang, and Thomas Barthel. "Isometric tensor network optimization for extensive Hamiltonians is free of barren plateaus." *Physical Review A* 109.5 (2024): L050402.

24. Schuld, Maria, et al. "Evaluating analytic gradients on quantum hardware." *Physical Review A* 99.3 (2019): 032331.

25. Stokes, James, et al. "Quantum natural gradient." *Quantum* 4 (2020): 269.

26. Malz, Daniel, et al. "Preparation of matrix product states with log-depth quantum circuits." *Physical Review Letters* 132.4 (2024): 040404.

27. Cerezo, Marco, et al. "Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing." *arXiv preprint arXiv:2312.09121* (2023).

28. Elben, Andreas, et al. "The randomized measurement toolbox." *Nature Reviews Physics* 5.1 (2023): 9-24.