

Challenges in Barren Plateau Mitigation with Dynamic Parameterized Quantum Circuits

Sumeet Shirgure Efehan Kökcü Siyuan Niu

Department of Electrical and Computer Engineering, University of Central Florida
{sumeetparameshwar.shirgure, efekan.kokcu, siyuan.niu}@ucf.edu

Abstract

Variational quantum algorithms (VQAs) are a promising paradigm for quantum advantage, yet their trainability is severely hampered by *barren plateaus* (BPs). Several works have proposed using *dynamic parameterized quantum circuits* (DPQCs), which intersperse unitary layers with parameterized CPTP maps (e.g. engineered dissipation, feedforward gadgets, or periodic resets), as a potential route around BPs. We unite this class of circuits into a formalization for DPQCs. We identify constraints on the nature and the structure of DPQCs if they are to prevent a significant number of parameters from becoming untrainable. We further show via purification and Pauli path analysis, a mechanism with which cost function anti-concentrates in DPQCs while still suffering from untrainability of a significant number of parameters. Our analysis reveals ways to design DPQCs that do not have an exponentially concentrated cost function, and our results suggest that BP mitigation via DPQCs is at least as hard as designing BP-free unitaries.

1 Motivation

A VQA [1] encodes a computational problem as an optimization problem of a cost function $C(\boldsymbol{\theta}) = \text{Tr}[HU(\boldsymbol{\theta})\rho_0U(\boldsymbol{\theta})^\dagger]$ which is the expectation value of an observable H in the state that a parameterized circuit $U(\boldsymbol{\theta})$ prepares from a fixed input ρ_0 . The central obstacle to scaling VQAs is the *barren plateau* (BP) phenomenon [4, 5]: a parameter is *untrainable* when its cost gradient varies exponentially little in the qubit number n , so resolving it from shot noise takes exponentially many circuit evaluations. A closely related fact is that when all gradients are exponentially suppressed the cost itself concentrates exponentially about its mean [8]. The converse is weaker: lower-bounding the cost variance as done in [9, 10] shows only that the cost is *anti-concentrated*, not that all parameters are trainable.

Many strategies are known for mitigating the effects of BPs, shallow circuits, small Dynamical Lie Algebras, local cost functions and non-trivial initialization strategies. Beyond these, a separate line of work uses explicitly non-unitary operations: engineered dissipation [11, 12], feedforward gadgets [9], and periodic resets [13]. These have been studied in a disparate way and are not as well understood as their unitary counterparts [5, 6]. We close this gap with a unifying perspective in our paper [25]. We work with error-free, precisely engineered non-unitary channels and set aside the effects of noise, since noisy VQAs are known to be classically simulable in many settings [15, 19, 20].

2 The DPQC formalism

We model the basic building blocks as *parameterized dynamic gadgets*: local CPTP maps $\mathcal{E}(\sigma)$ acting on $\mathcal{O}(1)$ qubits, realized by entangling a few ($\mathcal{O}(1)$) ancillas with the system qubits and discarding them [21, 22]. Fig. 1 shows an example of such a gadget and its purification. A *dynamic parameterized quantum circuit* (DPQC) $\mathcal{U}(\boldsymbol{\theta}, \boldsymbol{\sigma})$ is obtained from an ordinary circuit $U(\boldsymbol{\theta})$ by inserting layers of such gadgets $\mathcal{E}_j(\boldsymbol{\sigma})$ between its unitary blocks; the cost $C(\boldsymbol{\theta}, \boldsymbol{\sigma})$ is

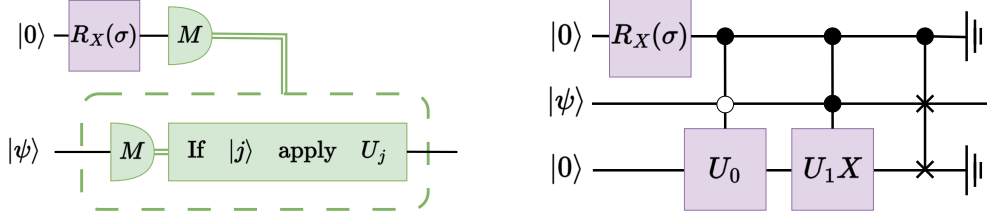


Figure 1: (Left) σ -parameterized “feedforward gadget” from [9]. If the ancilla above is measured to be 1, the circuit in the dotted green box is applied. Which in turn measures the data qubit and applies U_j if outcome is $|j\rangle$. (Right) Purification of this gadget using an extra ancilla below.

defined with respect to the full channel, with θ the original unitary parameters and σ the gadget parameters.

We impose two natural conditions. **Expressivity preservation:** each gadget reduces to the identity when its parameter is zero, so σ acts as a knob interpolating between the unitary circuit ($\sigma = 0$) and the fully dynamic one. **Faithfulness:** a DPQC is *faithful* when turning the knob slightly changes the cost by at most an exponentially small amount, $|C(\theta, \sigma) - C(\theta, 0)| \in \mathcal{O}(e^{-n})$ for all θ and small σ . This is stringent but natural: the original cost is already exponentially concentrated, so any larger deformation risks destroying the encoding of the original problem.

3 When do DPQCs leave parameters untrainable?

Suppose the underlying $U(\theta)$ already exhibits a barren plateau. Two results constrain the nature and the placement of the gadgets.

Lemma 1 (Untrainability of faithful DPQCs). *For a faithful DPQC built from $U(\theta)$, if U has no trainable parameters with respect to (H, ρ_0) , then no θ parameter is trainable in the DPQC.*

Faithfulness and trainability cannot be reconciled for free: a faithful DPQC inherits the untrainability of the unitary it was built from. Escaping it means giving up faithfulness, so that the augmented cost genuinely encodes the solution along the new σ directions. A practical consequence is that effective DPQCs must be *tailored* to the problem; a single one-size-fits-all gadget is unlikely to mitigate BPs across arbitrary ansatzes and Hamiltonians.

Lemma 2 (Untrainability of sparse DPQCs). *If any sufficiently long sub-array of $U(\theta)$ forms a 2-design, then for a DPQC whose consecutive gadget layers are separated by such a sub-array, no θ parameter inside that intervening stretch is trainable.*

Lemma 2 rules out a popular design choice: inserting only a constant number of gadget layers near the end of the circuit [9, 14]. The 2-design assumption is mild: most parameterized families are translation invariant [23, 24, 2], and deep random circuits and the hardware-efficient ansatz form 2-designs once deep enough [7], so every sufficiently long window has the same properties.

4 Anti-concentration without trainability: Pauli-path analysis

We focus on a single gadget layer inserted in the unitary PQC, and analyze such a DPQC using the Pauli-path framework [15, 16, 17, 18]. Any unitary acting after the layer can be “absorbed” into the CPTP map, so without loss of generality we write the DPQC as $\mathcal{E}(\sigma) \circ U(\theta)$. Purifying $\mathcal{E}(\sigma)$ into a unitary $V(\sigma)$ on an enlarged space $\mathcal{H}_s \otimes \mathcal{H}_a$ and discarding the ancillas reproduces the channel [21] (Figure 2 (left)); by construction $V(\sigma)$ carries all parameters of the circuit from the gadget layer onward, while $U(\theta)$ collects those before it. The key observation is that although H is initially supported on the system qubits, its Heisenberg evolution through $V(\sigma)$ generates Pauli terms living entirely on the ancillas, whose coefficients depend only on σ and are

not scrambled by $U(\theta)$. All existing methods of achieving cost anti-concentration [9, 13, 14] rely on assumptions that $V(\sigma)$ has $O(\log(n))$ depth and has local gates, however this is precisely the regime in which we can analyze V classically.

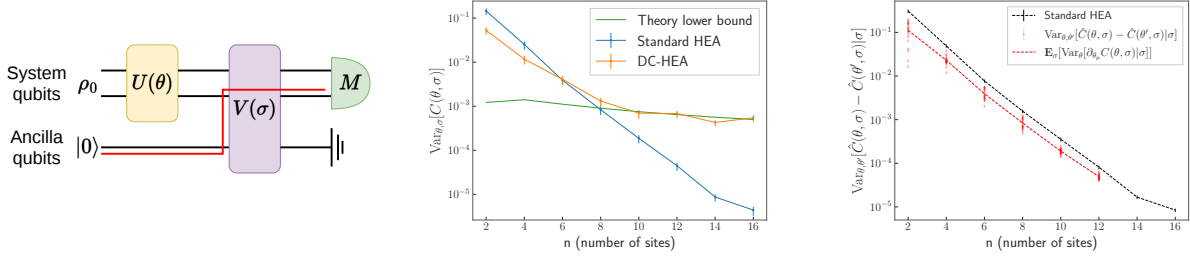


Figure 2: (Left) Purification of a DPQC with one non-unitary layer. The red line denotes Pauli strings in the linear sum representing the observable H , evolving into strings supported solely on ancillas. (Middle) 1D Ising model VQE cost anti-concentration $\text{Var}_{\sigma, \theta}[C] > \Omega(1/n)$ for a DPQC constructed from a unitary hardware efficient ansatz (HEA) and feedforward gadgets from [9] as shown in Figure 1. (Right) A statistic that shows the exponential concentration of partial gradients $\partial_{\theta_\mu} C$ in the same setting. This shows that cost function anti-concentration can co-exist with untrainability of parameters.

Lemma 3 (Cost anti-concentration from ancilla supported Pauli terms). *Let $U(\theta)$ be a PQC that forms a 1-design over θ , and let $\mathcal{U}(\theta, \sigma)$ be a purified DPQC as above. Suppose $V(\sigma)$ is such that the Heisenberg evolution of H through V can be computed classically as a sum of polynomially many terms*

$$V(\sigma)^\dagger H V(\sigma) = \sum_{P=I_s \otimes P_a} f_P(\sigma) I_s \otimes P_a + \sum_{Q=Q_s \otimes Q_a, Q_s \neq I_s} f_Q(\sigma) Q_s \otimes Q_a.$$

Define the classically computable function $F(\sigma) = \sum_P f_P(\sigma) \text{Tr}[P_a |0\rangle\langle 0|_a]$. If we can design $V(\sigma)$ such that $\text{Var}_\sigma[F(\sigma)] \geq g(n)$ for some threshold $g(n)$, then $\text{Var}_{\theta, \sigma}[C(\theta, \sigma)] \geq \text{Var}_\sigma[F(\sigma)] \geq g(n)$.

Lemma 3 gives a concrete certificate: once the circuit preceding the gadget behaves like a 1-design, one only needs to lower-bound the variance of the classically analyzable $F(\sigma)$, for which we provide a symbolic Pauli-Heisenberg evolution engine, `sympauli`. It also explains an asymmetry seen in earlier work [9, 13, 14]: gadgets restore trainability for the parameters *after* them (σ) but not the upstream θ parameters, since the injected variance is controlled by σ alone. This is the *illusion* of a trainable landscape: the cost varies appreciably in the σ directions while the θ gradients stay exponentially suppressed. Pushing the gadget layer too far from the end risks the subarray coming after that layer forming a 2-design and becoming untrainable as well.

5 Discussion

Taken together, DPQCs exhibit a tension between expressivity and trainability that gadgets do not seem able to resolve: faithful DPQCs inherit the plateaus they are built from (Lemma 1), sparse ones strand a significant fraction of the parameters on a BP (Lemma 2), and the anti-concentration they provide comes from ancilla-supported variance that may potentially never reach the upstream θ parameters (Lemma 3). In the purified (unitary) picture, designing a BP-free DPQC reduces to designing a BP-free unitary circuit [21] which is at least as hard as the original problem. Numerical experiments on VQE [3] and Max-Cut QAOA [2] in the companion paper [25] confirm this: cost anti-concentration coexists with untrainable θ parameters, and joint optimization yields no improvement over standard unitary ansatzes. The most promising direction, in our view, is the design of DPQCs whose structure is adapted to the target Hamiltonian rather than appended generically. Code and data are at <https://github.com/sumeetshirgure/dynbp>, and the `sympauli` engine at <https://pypi.org/project/sympauli/>.

References

- [1] M. Cerezo *et al.*, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).
- [2] E. Farhi, J. Goldstone, S. Gutmann, A quantum approximate optimization algorithm, arXiv:1411.4028 (2014).
- [3] A. Peruzzo *et al.*, A variational eigenvalue solver on a photonic quantum processor, *Nat. Commun.* **5**, 4213 (2014).
- [4] J. R. McClean *et al.*, Barren plateaus in quantum neural network training landscapes, *Nat. Commun.* **9**, 4812 (2018).
- [5] M. Larocca *et al.*, Barren plateaus in variational quantum computing, *Nat. Rev. Phys.* **7**, 174 (2025).
- [6] M. Ragone *et al.*, A Lie-algebraic theory of barren plateaus for deep parameterized quantum circuits, *Nat. Commun.* **15**, 7172 (2024).
- [7] M. Cerezo *et al.*, Cost-function-dependent barren plateaus in shallow parametrized quantum circuits, *Nat. Commun.* **12**, 1791 (2021).
- [8] A. Arrasmith, Z. Holmes, M. Cerezo, P. J. Coles, Equivalence of quantum barren plateaus to cost concentration and narrow gorges, *Quantum Sci. Technol.* **7**, 045015 (2022).
- [9] A. Deshpande *et al.*, Dynamic parameterized quantum circuits: expressive and barren-plateau free, arXiv:2411.05760 (2024).
- [10] G. Crognaletti, M. Grossi, A. Bassi, Estimates of loss function concentration in noisy parametrized quantum circuits, *PRX Quantum* **7**, 020336 (2026).
- [11] A. Sannia *et al.*, Engineered dissipation to mitigate barren plateaus, *npj Quantum Inf.* **10**, 81 (2024).
- [12] S. Cichy, P. K. Faehrmann, S. Khatri, J. Eisert, Perturbative gadgets for gate-based quantum computing: nonrecursive constructions without subspace restrictions, *Phys. Rev. A* **109**, 052624 (2024).
- [13] E. Zapusek, I. Rojkov, F. Reiter, Scaling quantum algorithms via dissipation: avoiding barren plateaus, arXiv:2507.02043 (2025).
- [14] Z. Chen, Y. Shao, Z. Liu, Z. Wei, Taming barren plateaus in arbitrary parameterized quantum circuits without sacrificing expressibility, arXiv:2511.13408 (2025).
- [15] G. González-García, J. I. Cirac, R. Trivedi, Pauli path simulations of noisy quantum circuits beyond average case, *Quantum* **9**, 1730 (2025).
- [16] S. Bravyi, D. Gosset, Improved classical simulation of quantum circuits dominated by Clifford gates, *Phys. Rev. Lett.* **116**, 250501 (2016).
- [17] D. Aharonov, X. Gao, Z. Landau, Y. Liu, U. Vazirani, A polynomial-time classical algorithm for noisy random circuit sampling, in *Proc. 55th Annual ACM Symposium on Theory of Computing* (2023), pp. 945–957.
- [18] A. Angrisani, A. Schmidhuber, M. S. Rudolph, M. Cerezo, Z. Holmes, H.-Y. Huang, Classically estimating observables of noiseless quantum circuits, *Phys. Rev. Lett.* **135**, 170602 (2025).
- [19] A. A. Mele *et al.*, Noise-induced shallow circuits and the absence of barren plateaus, *Nat. Phys.* (2026).
- [20] M. Cerezo *et al.*, Does provable absence of barren plateaus imply classical simulability?, *Nat. Commun.* **16**, 7907 (2025).
- [21] W. F. Stinespring, Positive functions on C*-algebras, *Proc. Amer. Math. Soc.* **6**, 211 (1955).
- [22] M. A. Nielsen, I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).

- [23] A. W. Harrow, R. A. Low, Random quantum circuits are approximate 2-designs, *Commun. Math. Phys.* **291**, 257 (2009).
- [24] A. Kandala *et al.*, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, *Nature* **549**, 242 (2017).
- [25] Shigure, S., Kökcü, E. and Niu, S., 2026. Challenges in Barren Plateau Mitigation with Dynamic Parameterized Quantum Circuits. arXiv preprint arXiv:2606.23751.