POLYNOMIALLY TRAINABLE VARIATIONAL QUANTUM CIRCUITS WITH BQP-HARD COSTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Understanding when parameterized quantum circuits can be both trainable and computationally advantageous is an important challenge in assessing the potential of variational quantum algorithms. Avoiding exponentially vanishing gradients is essential for scalable optimization, but a trainable landscape is of limited computational interest if its cost function can also be efficiently evaluated classically. Of particular interest is the intermediate regime beyond the immediate neighborhood of *Clifford* circuits, where inverse-polynomial gradients and the breakdown of known efficient classical approximations had previously been observed numerically, but their coexistence had not been proved. In this work, we establish this phenomenon rigorously under the standard assumption that $\mathsf{BQP} \not\subseteq \mathsf{BPP}$. Our results provide a theoretical basis for seeking computationally useful trainable quantum models.

1 INTRODUCTION

Parameterized quantum circuits provide a flexible framework for hybrid quantum–classical computation (Kandala et al., 2017; Havlíček et al., 2019; Sim et al., 2019; Schuld et al., 2020; Endo et al., 2021; Callison & Chancellor, 2022). A quantum processor evaluates a cost function determined by a circuit with tunable parameters, while a classical optimization routine updates those parameters using information about the cost landscape (McClean et al., 2016; Sweke et al., 2020). This paradigm underlies a broad class of variational quantum algorithms and quantum learning models, including early proposals for quantum circuit learning (Farhi et al., 2014; Peruzzo et al., 2014; Mitarai et al., 2018; Benedetti et al., 2019). For many commonly used parameterized gates, derivatives of expectation values can be evaluated directly on quantum hardware through parameter-shift identities (Mitarai et al., 2018; Schuld et al., 2019). Consequently, the usefulness of such methods is tied not only to what states or functions a circuit can represent, but also to whether its cost landscape contains gradients that can be resolved with feasible resources (Harrow & Napp, 2021).

A perennial obstruction is the barren-plateau phenomenon identified by McClean et al. (McClean et al., 2018). In a barren plateau, relevant gradients become exponentially suppressed with increasing problem size, so that estimating a useful optimization direction can require exponentially increasing precision or sampling effort (Wang et al., 2021). Subsequent work has shown that the appearance of barren plateaus depends strongly on structural properties of the circuit and objective. In particular, Cerezo et al. (Cerezo et al., 2021) demonstrated that the choice between global and local cost functions can qualitatively change gradient scaling in shallow parameterized circuits. The mechanisms, manifestations, and proposed approaches to barren plateaus have since developed into a substantial literature (Sack et al., 2022; Larocca et al., 2025). Avoiding exponentially small gradients is therefore a basic scalability requirement for variational quantum computation (Arrasmith et al., 2021).

Trainability, however, is only one side of the computational question. A parameterized quantum model whose relevant costs can be reproduced efficiently by a classical computer does not provide a computational separation, even if its optimization landscape is well behaved. This issue is especially visible near Clifford circuits. Stabilizer and Clifford circuits admit efficient classical simulation by the Gottesman–Knill theorem (Gottesman, 1998; Aaronson & Gottesman, 2004), while circuits containing non-Clifford resources can remain tractable when those resources are sufficiently restricted, as illustrated by classical simulation algorithms for Clifford-dominated circuits (Pashayan et al., 2015; Bravyi & Gosset, 2016; Bravyi et al., 2019). Perturbative constructions around Clif-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

ford circuits have likewise been used for efficient benchmarking and initialization of broad classes of variational circuits (Mitarai et al., 2022). More generally, the existence of classical surrogate models can constrain the sense in which a parameterized quantum model may offer a computational advantage (Rudolph et al., 2023; Schreiber et al., 2023).

These observations motivate a fundamental tension. Structural restrictions that make a parameterized circuit trainable may simultaneously make its behavior accessible to classical computation (Basheer et al., 2025). Cerezo et al. (Cerezo et al., 2025) recently examined this question directly, collecting evidence that several known mechanisms used to establish the absence of barren plateaus also confine the relevant computation to structures that admit efficient classical or quantum-assisted classical simulation. Their analysis makes clear that absence of barren plateaus alone is not a sufficient criterion for computational advantage. The important question is instead whether a regime can support gradients that remain large enough to be useful while retaining cost functions that are intrinsically difficult to evaluate classically.

Meyer et al. (Meyer et al., 2026) investigated this trainability–simulability tension in a near-Clifford setting. They consider parameterized circuits built from Clifford gates interleaved with one-qubit Pauli rotations, with costs given by observable expectation values. Around the Clifford point, this model connects naturally to the efficient simulation of stabilizer circuits (Gottesman, 1998; Aaronson & Gottesman, 2004) and to methods exploiting restricted departures from Clifford structure (Bravyi & Gosset, 2016; Mitarai et al., 2022; Begušić et al., 2025). Meyer et al. introduced a linear Clifford encoder that provides constant-scaling gradients in a classically accessible region near the Clifford point. Their numerical results then indicated a possible intermediate transition regime: as the initialization region is enlarged beyond the characteristic $D^{-1/2}$ scale, gradients may cease to remain constant yet decay only polynomially, while the corresponding landscape moves outside the efficiently accessible regime established by their analysis. Importantly, Meyer et al. presented this coexistence as preliminary empirical evidence and identified a formal proof as an open direction.

Our contributions. This work establishes such a coexistence result for an explicit family within the same model. For every fixed $0 < r < 1/2$, we consider a parameter region of width D^{-r}, which is asymptotically larger than the $D^{-1/2}$ near-Clifford scale. We construct parameterized costs for which a designated gradient component has an inverse-polynomial second moment over product-uniform initialization on this entire region, and for which the corresponding gradient variance is also inverse polynomial. Thus the relevant slope remains polynomially resolvable rather than becoming exponentially suppressed in the sense associated with barren plateaus (McClean et al., 2018; Cerezo et al., 2021; Larocca et al., 2025). At the same time, the same parameter region contains explicitly constructible points at which efficient classical evaluation of the cost for the whole family would imply $\mathsf{BQP} \subseteq \mathsf{BPP}$. Under the standard assumption $\mathsf{BQP} \not\subseteq \mathsf{BPP}$, the family therefore exhibits polynomially visible gradients and classically hard cost evaluation within one and the same near-Clifford patch.

Disclaimer. The goal of the present work is to prove the theoretical existence of a region where trainability and classical hardness coexist. Our construction does not provide a practically relevant variational quantum model.

2 NOTATIONS

We write $R_V(\theta) := \exp(-\mathrm{i}\theta V/2)$ for a one-qubit Pauli rotation $V \in \{X, Y, Z\}$, and call a circuit of fixed Clifford gates interleaved with independently parameterized Pauli rotations a Pauli-rotated Clifford (PRC) circuit, which takes the form $U = RW \cdots RW$ (see Appendix C.1 for a formal definition). For an N-qubit PRC circuit $U(\boldsymbol{\theta})$ and Hermitian observable H, its cost is $C(\boldsymbol{\theta}) := \langle 0|U(\boldsymbol{\theta})^\dagger H U(\boldsymbol{\theta})|0\rangle$, where $|0\rangle$ denotes the all-zero state. ∇C, $\|\cdot\|$, and $\|\cdot\|_1$ denote the parameter gradient, operator norm (Euclidean norm for vectors when subscripted by 2), and ℓ_1 norm, respectively. Initialization $\boldsymbol{\theta} \sim \mathrm{Unif}([-\sigma, \sigma]^D)$ means that the D coordinates are independent and uniform on $[-\sigma, \sigma]$. We use $\mathrm{dist}(x, S) := \inf_{s\in S} |x - s|$, with $(\pi/2)\mathbb{Z}$ the set of Clifford rotation angles. A Clifford+T circuit Q is a circuit over Clifford gates and $T, T^\dagger$, and $p(Q)$ denotes its probability of outputting 1 on the designated output qubit o, with $\Pi_o := |1\rangle\langle 1|_o$ denoting the corresponding output-1 projector. We write QCA (quantum circuit acceptance) for the promise problem of deciding whether $p(Q) \geq 2/3$ or $p(Q) \leq 1/3$. BPP and BQP denote standard bounded-

error polynomial-time complexity classes for classical and quantum computation, respectively, and BQP-hardness is under deterministic polynomial-time many-one reductions. Detailed definitions, preliminaries, and analytic tools are provided in Appendices A through E.

3 CONSTRUCTION OF THE TRANSITION-PATCH FAMILY

Fix once and for all

$$r \in \mathbb{Q}, \qquad 0 < r < \frac{1}{2}, \qquad \mu \in \mathbb{N}. \tag{1}$$

The integer μ tunes the polynomial gradient exponent. Define $\lambda_\mu := \mu/(3\ln 2)$. By increasing μ, one can obtain arbitrarily large positive inverse-polynomial exponents.

3.1 PADDING THE SOURCE QUANTUM CIRCUIT

Let Q be an instance of QCA, as in Definition D.4, and let s be the binary description length of Q. Let t_0 be the number of T or $T^\dagger$ gates appearing in Q. Since each gate requires at least one symbol in the description, $t_0 \le s$.

Append identity pairs $TT^\dagger$ on any existing qubit. If

$$q = \left\lceil \frac{s - t_0}{2} \right\rceil, \tag{2}$$

then the padded circuit has

$$t = t_0 + 2q \quad \text{with} \quad s \le t < s + 2. \tag{3}$$

The unitary and acceptance probability are unchanged because each appended pair is the identity. We may append finitely many additional pairs so that t exceeds any fixed constant needed for later asymptotic inequalities. The padded description still has length $O(s)$, and $t = \Theta(s)$.

3.2 SIZE PARAMETERS

Set

$$D = \left\lceil t^{1/(1-r)} \right\rceil, \qquad \sigma_D = D^{-r}. \tag{4}$$

Let

$$c_\star = \frac{7}{8}, \qquad K = \lceil c_\star D^r \rceil, \qquad \delta_D = \frac{\pi}{4K}. \tag{5}$$

The number of spectator qubits is

$$M = \mu \left\lceil D^{2r} \right\rceil \lceil \log_2 D \rceil. \tag{6}$$

Equation (6) is computable by elementary integer arithmetic once the rational exponent r is fixed.

Observation 3.1 (Asymptotic parameter counting)**.** *The quantities in Eqs.* (4) *to* (6) *satisfy*

$$D^{1-r} = t(1 + o(1)), \tag{7}$$

$$\frac{K}{D^r} \longrightarrow c_\star, \qquad \frac{tK}{D} \longrightarrow c_\star, \tag{8}$$

$$\frac{M}{D} \longrightarrow 0, \tag{9}$$

$$M\sigma_D^2 = \mu \log_2 D + O(1) + o(1), \tag{10}$$

$$\frac{\delta_D}{\sigma_D} \longrightarrow \frac{\pi}{4c_\star} = \frac{2\pi}{7} < 1. \tag{11}$$

Consequently, if

$$q_D := tK + K + M, \qquad R_D := D - q_D, \tag{12}$$

then

$$\frac{q_D}{D} \longrightarrow \frac{7}{8}, \qquad \frac{R_D}{D} \longrightarrow \frac{1}{8}. \tag{13}$$

In particular, $R_D \ge 1$ for all sufficiently large D, and $q_D = \Theta(D)$.

Because $K \ge c_\star D^r$, we also have the non-asymptotic inequality

$$0 < \delta_D \le \frac{\pi}{4c_\star} D^{-r} = \frac{2\pi}{7}\sigma_D < \sigma_D. \tag{14}$$

3.3 THE HARD COMPUTATION BLOCK

Replace every T gate of the padded circuit Q by a block of K independently parameterized rotations

$$R_Z(\beta_{j,K}) \cdots R_Z(\beta_{j,1}), \tag{15}$$

and every $T^\dagger$ gate by another such block. All original Clifford gates remain fixed. Denote the resulting parameterized circuit by $V_Q(\boldsymbol{\beta})$, where $\boldsymbol{\beta} \in \mathbb{R}^{tK}$.

Define the distinguished assignment $\boldsymbol{\beta}^\star$ by

$$\beta^\star_{j,\ell} = \begin{cases} +\delta_D, & \text{if block } j \text{ replaces a } T \text{ gate,} \\ -\delta_D, & \text{if block } j \text{ replaces a } T^\dagger \text{ gate.} \end{cases} \tag{16}$$

Observation 3.2. *At $\boldsymbol{\beta} = \boldsymbol{\beta}^\star$, every block in Eq.* (15) *equals the original T or $T^\dagger$ gate up to a global phase. Hence*

$$p_Q(\boldsymbol{\beta}^\star) := \langle 0|V_Q(\boldsymbol{\beta}^\star)^\dagger \Pi_o V_Q(\boldsymbol{\beta}^\star)|0\rangle = p(Q). \tag{17}$$

At $\boldsymbol{\beta} = 0$, the circuit $V_Q(0)$ is Clifford.

By Eq. (14), every coordinate of $\boldsymbol{\beta}^\star$ lies in $[-\sigma_D, \sigma_D]$.

3.4 THE TRAINABILITY HANDLE

Introduce one readout qubit a_0 and M spectator qubits $a_1, \ldots, a_M$. Let

$$U_A(\boldsymbol{\phi}, \boldsymbol{\vartheta}) = R_Y^{(a_0)}(\phi_K) \cdots R_Y^{(a_0)}(\phi_1) \bigotimes_{j=1}^{M} R_Y^{(a_j)}(\vartheta_j), \tag{18}$$

where $\boldsymbol{\phi} \in \mathbb{R}^K$ and $\boldsymbol{\vartheta} \in \mathbb{R}^M$. Define the Pauli string

$$P_A = X_{a_0} \otimes Z_{a_1} \otimes \cdots \otimes Z_{a_M}. \tag{19}$$

The R_D remaining parameters from Eq. (12) are denoted by $\boldsymbol{\gamma}$. Each is assigned to a Z rotation on its own dummy qubit. The observable will act as the identity on all dummy qubits, so these rotations do not affect the cost. We order one dummy rotation as the final circuit gate, ensuring the hypothesis of Observation C.2.

The total parameter vector is

$$\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\vartheta}, \boldsymbol{\beta}, \boldsymbol{\gamma}) \in \mathbb{R}^D. \tag{20}$$

The full circuit is the tensor-product computation

$$U_Q(\boldsymbol{\theta}) = U_A(\boldsymbol{\phi}, \boldsymbol{\vartheta}) \otimes V_Q(\boldsymbol{\beta}) \otimes U_{\mathrm{dum}}(\boldsymbol{\gamma}), \tag{21}$$

serialized into a single gate sequence. Gates on disjoint registers commute, so serialization does not change the tensor-product unitary. By Observation C.2, this is a PRC circuit. At $\boldsymbol{\theta} = 0$, it is Clifford.

Lemma 3.3 (Qubit count)**.** *The number N_D of qubits in the construction satisfies*

$$N_D = \Theta(D). \tag{22}$$

Proof. See Appendix F.1. □

3.5 A NORMALIZED THREE-PAULI OBSERVABLE

On the handle register define

$$A_A = \frac{I + P_A}{2}. \tag{23}$$

On the output qubit define

$$B_o = \frac{I + \Pi_o}{2} = \frac{3I - Z_o}{4}, \tag{24}$$

where Eq. (76) was used. Set

$$H_Q = \frac{8}{5}\left(A_A \otimes B_o - \frac{3}{8} I\right) = \frac{1}{5}\left(-Z_o + 3P_A - P_A Z_o\right). \tag{25}$$

Identity operators on registers not displayed are implicit.

Proposition 3.4. *The observable H_Q in Eq.* (25) *is Hermitian, contains exactly three nonzero Pauli terms, and satisfies*

$$\|H_Q\| = 1. \tag{26}$$

Proof. See Appendix F.2. □

3.6 Cost and gradient formulas

Define

$$S(\boldsymbol{\phi}) = \sum_{\ell=1}^{K} \phi_\ell, \qquad p_Q(\boldsymbol{\beta}) = \langle 0|V_Q(\boldsymbol{\beta})^\dagger \Pi_o V_Q(\boldsymbol{\beta})|0\rangle. \tag{27}$$

Also define

$$g(\boldsymbol{\phi}, \boldsymbol{\vartheta}) = \sin S(\boldsymbol{\phi}) \prod_{j=1}^{M} \cos \vartheta_j. \tag{28}$$

Our cost function and its gradient have the form

$$C_Q(\boldsymbol{\theta}) = \frac{2}{5}\left(1 + g(\boldsymbol{\phi}, \boldsymbol{\vartheta})\right)\left(1 + p_Q(\boldsymbol{\beta})\right) - \frac{3}{5}.$$

and

$$\partial_{\phi_\ell} C_Q(\boldsymbol{\theta}) = \frac{2}{5} \cos S(\boldsymbol{\phi}) \left(\prod_{j=1}^{M} \cos \vartheta_j\right)\left(1 + p_Q(\boldsymbol{\beta})\right),$$

respectively, which implies $2/5 \le \partial_{\phi_\ell} C_Q(0) \le 4/5$ (see Appendix F.3 for details).

4 Gradient scaling on and beyond the local Clifford scale

All statements in this section hold uniformly over the source circuit Q. The hard block enters only through the bounded factor

$$1 \le 1 + p_Q(\boldsymbol{\beta}) \le 2. \tag{29}$$

This is the reason the gradient analysis does not require any statistical assumption about the embedded quantum computation.

4.1 Inverse-polynomial second moment

Theorem 4.1 (Gradient second moment on the D^{-r} patch)**.** *Let*

$$\boldsymbol{\theta} \sim \mathrm{Unif}([-\sigma_D, \sigma_D]^D), \qquad \sigma_D = D^{-r}, \tag{30}$$

with independent coordinates. For every readout coordinate ϕ_ℓ,

$$\mathbb{E}\Big[\left(\partial_{\phi_\ell} C_Q(\boldsymbol{\theta})\right)^2\Big] = \Theta(D^{-\lambda_\mu}), \qquad \lambda_\mu = \frac{\mu}{3\ln 2}. \tag{31}$$

Consequently,

$$\mathbb{E}\Big[\|\nabla C_Q(\boldsymbol{\theta})\|_2^2\Big] = \Omega(D^{-\lambda_\mu}) = \Omega(N_D^{-\lambda_\mu}). \tag{32}$$

Proof. See Appendix F.4. □

4.2 Inverse-polynomial variance

The second moment in Theorem 4.1 already matches the diagnostic emphasized by (Meyer et al., 2026). The next result shows that the variance of the gradient is also inverse polynomial.

Proposition 4.2 (Gradient variance lower bound)**.** *Under the initialization in Eq.* (30)*, every readout component satisfies*

$$\mathrm{Var}(\partial_{\phi_\ell} C_Q(\boldsymbol{\theta})) = \Omega\Big(D^{-(\lambda_\mu + 2r)} \log D\Big). \tag{33}$$

Proof. See Appendix F.5. □

4.3 TRANSITION PROFILE

The same circuit can be initialized with a width τ_D different from the design width σ_D. This reveals a transition from constant to polynomial to superpolynomial gradient suppression.

Theorem 4.3. *Let $\tau_D > 0$ satisfy $\tau_D \to 0$, and initialize every parameter independently and uniformly on $[-\tau_D, \tau_D]$. Then, uniformly over Q,*

$$\mathbb{E}\Big[(\partial_{\phi_\ell} C_Q)^2\Big] = \Theta\left(\exp\left[-\frac{M\tau_D^2}{3} + O(M\tau_D^4)\right]\right). \tag{34}$$

In particular:

$$\tau_D = D^{-1/2} \quad \Longrightarrow \quad \mathbb{E}[(\partial_{\phi_\ell} C_Q)^2] = \Theta(1), \tag{35}$$

$$\tau_D = D^{-r} \quad \Longrightarrow \quad \mathbb{E}[(\partial_{\phi_\ell} C_Q)^2] = \Theta(D^{-\lambda_\mu}), \tag{36}$$

$$\tau_D = D^{-r'}, \quad 0 < r' < r \quad \Longrightarrow \quad \mathbb{E}[(\partial_{\phi_\ell} C_Q)^2] = D^{-\lambda_\mu D^{2(r-r')}(1+o(1))}. \tag{37}$$

The last quantity is smaller than D^{-a} for every fixed $a > 0$ when D is sufficiently large.

Proof. See Appendix F.6. □

Remark. The $D^{-1/2}$ scale has constant gradients in this construction. The larger D^{-r} scale, with $0 < r < 1/2$, has inverse-polynomial gradients. Widening still further to $D^{-r'}$ with $r' < r$ produces superpolynomial suppression. The transition follows from the formula in Eq. (34).

5 A CLASSICALLY HARD POINT WITH AN INVERSE-POLYNOMIAL GRADIENT

5.1 THE DISTINGUISHED TRANSITION POINT

Define

$$J_D = \lceil 2D^r \rceil, \qquad \eta_D = 2\arctan\left(\frac{1}{J_D}\right). \tag{38}$$

The angle η_D is chosen so that it is asymptotic to the patch width while its cosine has a simple rational expression.

Lemma 5.1 (Properties of the spectator angle)**.** *For all sufficiently large D,*

$$0 < \eta_D \le \sigma_D, \tag{39}$$

$$\frac{\eta_D}{\sigma_D} \longrightarrow 1, \tag{40}$$

$$\cos\eta_D = \frac{J_D^2 - 1}{J_D^2 + 1}. \tag{41}$$

If

$$a_D = (\cos\eta_D)^M, \tag{42}$$

then

$$a_D = \Theta(D^{-3\lambda_\mu/2}) \quad \text{and} \quad a_D^2 = \Theta(D^{-3\lambda_\mu}). \tag{43}$$

Proof. See Appendix F.7. □

Define the distinguished point

$$\boldsymbol{\theta}_Q^\star = (\boldsymbol{\phi}^\star, \boldsymbol{\vartheta}^\star, \boldsymbol{\beta}^\star, \boldsymbol{\gamma}^\star) \tag{44}$$

by

$$\phi_\ell^\star = \delta_D \quad (1 \le \ell \le K), \qquad \vartheta_j^\star = \eta_D \quad (1 \le j \le M), \qquad \boldsymbol{\gamma}^\star = 0, \tag{45}$$

and by the hard-block assignment $\boldsymbol{\beta}^\star$ from Eq. (16). At this point, the cost and designated gradient are

$$C_Q(\boldsymbol{\theta}_Q^\star) = \frac{2}{5}\left(1 + \frac{a_D}{\sqrt{2}}\right)(1 + p(Q)) - \frac{3}{5},$$

$$\partial_{\phi_1} C_Q(\boldsymbol{\theta}_Q^\star) = \frac{2a_D}{5\sqrt{2}}(1 + p(Q))$$

(see Appendix F.8 for details).

5.2 HARDNESS OF COST ESTIMATION

Before emitting the final reduction instance, perform the following deterministic padding loop. Compute the rational number a_D from Eqs. (41) and (42). If

$$a_D^2 > \frac{1}{512}, \tag{46}$$

append one identity pair $TT^\dagger$ to the source circuit and recompute t, D, K, M, J_D, and a_D. Repeat until Eq. (46) fails. Exact comparison is possible by integer exponentiation because $a_D = ((J_D^2 - 1)/(J_D^2 + 1))^M$ is rational. The loop terminates because Eq. (43) implies $a_D \to 0$ as $t \to \infty$. Moreover, for fixed r and μ there is a constant threshold beyond which no iteration is needed. Hence the number of added gates is bounded by a constant on the finitely many smaller input sizes and the entire padding procedure is polynomial time. We may therefore assume, without changing the accepted language or any asymptotic estimate, that

$$\frac{a_D}{\sqrt{2}} \le \frac{1}{32}. \tag{47}$$

Definition 5.2 (Transition-patch cost-sign problem)**.** Fix r and μ. An instance is the finite description of the circuit U_Q, the observable H_Q, and the point $\boldsymbol{\theta}_Q^\star$ produced from a Clifford+T circuit Q satisfying the promise in Eq. (89). The two nonrational angle magnitudes are represented by the finite expression strings $\pi/(4K)$ and $2\arctan(1/J_D)$, together with their integer arguments. The task is to decide whether

$$C_Q(\boldsymbol{\theta}_Q^\star) \ge \frac{1}{15} \quad \text{or} \quad C_Q(\boldsymbol{\theta}_Q^\star) \le -\frac{1}{20}. \tag{48}$$

Theorem 5.3 (Worst-case BQP-hardness inside the polynomial-gradient patch)**.** *For every fixed $r \in \mathbb{Q} \cap (0, 1/2)$ and $\mu \in \mathbb{N}$, the problem in Definition 5.2 is* BQP*-hard under deterministic polynomial-time many-one reductions. Consequently, if a randomized classical polynomial-time algorithm estimated every value $C_Q(\boldsymbol{\theta}_Q^\star)$ to additive error at most $1/100$ with success probability at least $2/3$, then*

$$\mathsf{BQP} \subseteq \mathsf{BPP}. \tag{49}$$

Under Eq. (90)*, no such classical estimator exists.*

Proof. See Appendix F.9. □

5.3 HARDNESS OF HIGH-PRECISION GRADIENT EVALUATION

The cost-hardness theorem is sufficient for the claimed coexistence. The same construction also makes a particular gradient component hard to evaluate at its natural inverse-polynomial scale.

Define

$$G_D = \frac{2a_D}{5\sqrt{2}}. \tag{50}$$

By Eq. (41), $a_D = ((J_D^2-1)/(J_D^2+1))^M$ is rational and can be computed by integer exponentiation. G_D can therefore be approximated to any polynomial number of bits in deterministic polynomial time.

Corollary 5.4 (Gradient-evaluation hardness)**.** *Suppose an algorithm, on every constructed instance, estimates $\partial_{\phi_1} C_Q(\boldsymbol{\theta}_Q^\star)$ to additive error at most $G_D/12$ with success probability at least $2/3$. Then the algorithm decides* QCA *with the same success probability. In particular, such gradient estimation is* BQP*-hard, and the required precision is inverse polynomial in D.*

Proof. See Appendix F.10. □

5.4 FINITE RATIONAL ANGLE DESCRIPTIONS

The ideal point was written using π and $\arctan$. Such expressions have finite symbolic descriptions, but a conventional input to a Turing machine is most cleanly formulated with rational numbers. We therefore make the conversion explicit and retain both the hardness gap and the inverse-polynomial gradient.

Lemma 5.5. *For every sufficiently large D, the reduction can compute a rational vector $\widehat{\boldsymbol{\theta}}_Q^\star \in \mathbb{Q}^D$ with $O(\log D)$ fractional bits per coordinate such that*

$$\widehat{\boldsymbol{\theta}}_Q^\star \in [-\sigma_D, \sigma_D]^D, \qquad \Delta_D := \left\| \widehat{\boldsymbol{\theta}}_Q^\star - \boldsymbol{\theta}_Q^\star \right\|_1 \le D^{-(\mu+3)}. \tag{51}$$

At this rational point,

$$\left| C_Q(\widehat{\boldsymbol{\theta}}_Q^\star) - C_Q(\boldsymbol{\theta}_Q^\star) \right| \le \Delta_D, \tag{52}$$

$$\left| \partial_{\phi_1} C_Q(\widehat{\boldsymbol{\theta}}_Q^\star) - \partial_{\phi_1} C_Q(\boldsymbol{\theta}_Q^\star) \right| \le \Delta_D. \tag{53}$$

Consequently,

$$\left(\partial_{\phi_1} C_Q(\widehat{\boldsymbol{\theta}}_Q^\star) \right)^2 = \Theta(D^{-3\lambda_\mu}), \tag{54}$$

$$\left\| \widehat{\boldsymbol{\theta}}_Q^\star \right\|_1 = \Theta(D^{1-r}), \tag{55}$$

$$\# \left\{ j : \frac{\pi}{14} D^{-r} \le \operatorname{dist}\left(\widehat{\theta}_{Q,j}^\star, \frac{\pi}{2}\mathbb{Z} \right) \le \frac{2\pi}{7} D^{-r} \right\} \ge \frac{D}{2} \quad \text{for all sufficiently large } D. \tag{56}$$

Furthermore, the rational-point promise problem

$$\text{yes: } C_Q(\widehat{\boldsymbol{\theta}}_Q^\star) \ge \frac{3}{50}, \qquad \text{no: } C_Q(\widehat{\boldsymbol{\theta}}_Q^\star) \le -\frac{1}{25} \tag{57}$$

is BQP*-hard. Estimating its cost to additive error $1/100$ is therefore* BQP*-hard. Estimating the designated gradient at the rational point to additive error $G_D/24$ is also* BQP*-hard.*

Proof. See Appendix F.11. □

6 Main result

Theorem 6.1. *Fix $r \in \mathbb{Q} \cap (0, 1/2)$ and $\mu \in \mathbb{N}$, and let $\lambda_\mu = \mu/(3 \ln 2)$. There is a deterministic polynomial-time construction that maps every Clifford+T circuit Q with description length s to a PRC parameterized cost C_Q with the following properties.*

1. *The circuit has $D = \mathrm{poly}(s)$ parameters and $N_D = \Theta(D)$ qubits, satisfies that $U_Q(0)$ is Clifford, and uses an observable H_Q containing exactly three nonzero Pauli terms with $\|H_Q\| = 1$.*

2. *For $\boldsymbol{\theta} \sim \mathrm{Unif}([-D^{-r}, D^{-r}]^D)$,*

$$\mathbb{E}\left[(\partial_{\phi_1} C_Q(\boldsymbol{\theta}))^2 \right] = \Theta(D^{-\lambda_\mu}),$$
$$\mathrm{Var}(\partial_{\phi_1} C_Q(\boldsymbol{\theta})) = \Omega(D^{-(\lambda_\mu + 2r)} \log D),$$
$$\mathbb{E}\left[\|\nabla C_Q(\boldsymbol{\theta})\|_2^2 \right] = \Omega(D^{-\lambda_\mu}).$$

   Thus the family is free of an exponential barren plateau according to both Definitions C.4 and C.5.

3. *The same family has the transition profile in Theorem 4.3: the designated second moment is constant at width $D^{-1/2}$, inverse polynomial at width D^{-r}, and superpolynomially small at every width $D^{-r'}$ with $0 < r' < r$.*

4. *The cube $[-D^{-r}, D^{-r}]^D$ contains an explicitly computable rational point $\widehat{\boldsymbol{\theta}}_Q^\star$ with $O(\log D)$ bits after the binary point and with*

$$\left\| \widehat{\boldsymbol{\theta}}_Q^\star \right\|_1 = \Theta(D^{1-r}),$$
$$(\partial_{\phi_1} C_Q(\widehat{\boldsymbol{\theta}}_Q^\star))^2 = \Theta(D^{-3\lambda_\mu}),$$

   and, for all sufficiently large D, with at least $D/2$ coordinates whose distance from the nearest Clifford angle lies between $(\pi/14)D^{-r}$ and $(2\pi/7)D^{-r}$.

5. *Estimating $C_Q(\widehat{\boldsymbol{\theta}}^\star_Q)$ to additive error $1/100$ for every source circuit Q is* BQP*-hard. Consequently, under* BQP $\not\subseteq$ BPP*, no bounded-error randomized classical polynomial-time algorithm evaluates every such cost on the patch. Estimating the designated gradient to its natural inverse-polynomial precision is also* BQP*-hard.*

Proof. See Appendix F.12. □

7 LIMITATIONS

The results of this work are existential. They establish that polynomially visible gradients and classically hard cost evaluation are mathematically compatible within the same near-Clifford parameter region, but they do not imply that this coexistence is generic, experimentally robust, or sufficient for successful variational optimization.

The principal complexity-theoretic limitation is that the hardness result is worst-case within the parameter patch. Consequently, the present analysis does not show that a randomly initialized parameter vector yields a classically hard instance.

Our circuit family is designed to separate the roles of trainability and computational hardness. This modular structure is useful for proving the compatibility theorem, but it is not intended as a model of a typical hardware-efficient variational architecture. Demonstrating analogous coexistence results for other natural ansatz families remains an open problem.

These limitations identify several natural directions in which the present existence result could be strengthened. Particularly important targets are average-case hardness over a physically motivated initialization distribution and constructions with local low-weight observables and geometrically local circuits. Establishing these properties would move the result from an existential compatibility theorem toward a stronger characterization of the trainability–simulability transition in practically motivated variational quantum models.

8 CONCLUSION

We establish that trainability and classical intractability can coexist within the same near-Clifford region of a parameterized quantum cost landscape. Motivated by the transition-zone picture, we constructed a uniform family of static parameterized quantum circuits for which the gradient behavior can be characterized across initialization scales. The resulting family exhibits a progression from constant gradients at the near-Clifford scale, to inverse-polynomial gradients on a parametrically larger patch, and ultimately to superpolynomial suppression. In particular, the intermediate regime provides a rigorous realization of a landscape region in which gradients remain polynomially resolvable rather than becoming barren. This establishes the transition-zone phenomenon as a mathematically consistent possibility.

An important direction is to determine how broadly the same coexistence phenomenon persists under more natural architectural constraints and in representative regions of practically motivated circuit families. Establishing such results would clarify which structural properties, rather than trainability alone, ultimately govern the boundary between classically accessible and genuinely quantum parameterized landscapes.

AI USE STATEMENT

In this work, we used generative AI tools for grammar checking, correctness checking, drafting sections, and identifying relevant prior theoretical work. We have not used generative AI tools for the main proof mechanism, and data collection, experimental design, and figure generation are not applicable to this work. We have reviewed all AI-assisted work. For example, we verified the correctness and applicability of some previously established lemmas identified by AI as potentially useful for our proof. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

Reproducibility statement

This work is theoretical and does not rely on numerical experiments or external datasets. All assumptions, circuit models, parameters, and optimization objectives used in our results are explicitly defined in the text. Complete proofs of the results are provided in the appendix.

References

Scott Aaronson and Daniel Gottesman. Improved simulation of stabilizer circuits. *Physical Review A*, 70(5):052328, 2004.

Dorit Aharonov, Michael Ben-Or, Elad Eban, and Urmila Mahadev. Interactive proofs for quantum computations. *arXiv preprint arXiv:1704.04487*, 2017.

Andrew Arrasmith, Marco Cerezo, Piotr Czarnik, Lukasz Cincio, and Patrick J Coles. Effect of barren plateaus on gradient-free optimization. *Quantum*, 5:558, 2021.

Leonardo Banchi and Gavin E Crooks. Measuring analytic gradients of general quantum evolution with the stochastic parameter shift rule. *Quantum*, 5:386, 2021.

Afrad Basheer, Yuan Feng, Christopher Ferrie, Sanjiang Li, and Hakop Pashayan. On the trainability and classical simulability of learning matrix product states variationally. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 15498–15506, 2025.

Tomislav Begušić, Kasra Hejazi, and Garnet Kin Chan. Simulating quantum circuit expectation values by Clifford perturbation theory. *The Journal of Chemical Physics*, 162(15), 2025.

Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4):043001, 2019.

Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing*, pp. 11–20, 1993.

P Oscar Boykin, Tal Mor, Matthew Pulver, Vwani Roychowdhury, and Farrokh Vatan. A new universal and fault-tolerant quantum basis. *Information Processing Letters*, 75(3):101–107, 2000.

Sergey Bravyi and David Gosset. Improved classical simulation of quantum circuits dominated by Clifford gates. *Physical Review Letters*, 116(25):250501, 2016.

Sergey Bravyi, Dan Browne, Padraic Calpin, Earl Campbell, David Gosset, and Mark Howard. Simulation of quantum circuits by low-rank stabilizer decompositions. *Quantum*, 3:181, 2019.

Adam Callison and Nicholas Chancellor. Hybrid quantum-classical algorithms in the noisy intermediate-scale quantum era and beyond. *Physical Review A*, 106(1):010101, 2022.

Marco Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications*, 12(1): 1791, 2021.

Marco Cerezo, Martin Larocca, Diego García-Martín, Nelson L Diaz, Paolo Braccia, Enrico Fontana, Manuel S Rudolph, Pablo Bermejo, Aroosa Ijaz, Supanut Thanasilp, et al. Does provable absence of barren plateaus imply classical simulability? *Nature Communications*, 16(1): 7907, 2025.

Gavin E Crooks. Gradients of parameterized quantum gates using the parameter-shift rule and gate decomposition. *arXiv preprint arXiv:1905.13311*, 2019.

Christopher M Dawson and Michael A Nielsen. The solovay-kitaev algorithm. *arXiv preprint quant-ph/0505030*, 2005.

Jeroen Dehaene and Bart De Moor. Clifford group, stabilizer states, and linear and quadratic operations over GF(2). *Physical Review A*, 68(4):042318, 2003.

Suguru Endo, Zhenyu Cai, Simon C Benjamin, and Xiao Yuan. Hybrid quantum-classical algorithms and quantum error mitigation. *Journal of the Physical Society of Japan*, 90(3):032001, 2021.

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

Daniel Gottesman. The Heisenberg representation of quantum computers. *arXiv preprint quant-ph/9807006*, 1998.

Edward Grant, Leonard Wossnig, Mateusz Ostaszewski, and Marcello Benedetti. An initialization strategy for addressing barren plateaus in parametrized quantum circuits. *Quantum*, 3:214, 2019.

Aram W Harrow and John C Napp. Low-depth gradient measurements can improve convergence in variational hybrid quantum-classical algorithms. *Physical Review Letters*, 126(14):140502, 2021.

Vojtěch Havlíček, Antonio D Córcoles, Kristan Temme, Aram W Harrow, Abhinav Kandala, Jerry M Chow, and Jay M Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, 2019.

Zoë Holmes, Kunal Sharma, Marco Cerezo, and Patrick J Coles. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. *PRX Quantum*, 3(1):010313, 2022.

Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M Chow, and Jay M Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, 2017.

Martin Larocca, Piotr Czarnik, Kunal Sharma, Gopikrishnan Muraleedharan, Patrick J Coles, and Marco Cerezo. Diagnosing barren plateaus with tools from quantum optimal control. *Quantum*, 6:824, 2022.

Martin Larocca, Supanut Thanasilp, Samson Wang, Kunal Sharma, Jacob Biamonte, Patrick J Coles, Lukasz Cincio, Jarrod R McClean, Zoë Holmes, and Marco Cerezo. Barren plateaus in variational quantum computing. *Nature Reviews Physics*, 7(4):174–189, 2025.

Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18(2):023023, 2016.

Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1):4812, 2018.

Sabri Meyer, Francesco Scala, Francesco Tacchino, and Aurelien Lucchi. Gradient scalability and Taylor surrogation of quantum cost landscapes. *Physical Review Research*, 8(2):023325, 2026.

Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. Quantum circuit learning. *Physical Review A*, 98(3):032309, 2018.

Kosuke Mitarai, Yasunari Suzuki, Wataru Mizukami, Yuya O Nakagawa, and Keisuke Fujii. Quadratic Clifford expansion for efficient benchmarking and initialization of variational quantum algorithms. *Physical Review Research*, 4(3):033012, 2022.

M Nest. Classical simulation of quantum computation, the Gottesman-Knill theorem, and slightly beyond. *arXiv preprint arXiv:0811.0898*, 2008.

Michael A Nielsen, Isaac L Chuang, et al. *Quantum computation and quantum information*, volume 1. Cambridge University Press, 2000.

Hakop Pashayan, Joel J Wallman, and Stephen D Bartlett. Estimating outcome probabilities of quantum circuits using quasiprobabilities. *Physical Review Letters*, 115(7):070501, 2015.

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, 2014.

Arthur Pesah, Marco Cerezo, Samson Wang, Tyler Volkoff, Andrew T Sornborger, and Patrick J Coles. Absence of barren plateaus in quantum convolutional neural networks. *Physical Review X*, 11(4):041011, 2021.

Manuel S Rudolph, Enrico Fontana, Zoë Holmes, and Lukasz Cincio. Classical surrogate simulation of quantum systems with LOWESA. *arXiv preprint arXiv:2308.09109*, 2023.

Stefan H Sack, Raimel A Medina, Alexios A Michailidis, Richard Kueng, and Maksym Serbyn. Avoiding barren plateaus using classical shadows. *PRX Quantum*, 3(2):020365, 2022.

Franz J Schreiber, Jens Eisert, and Johannes Jakob Meyer. Classical surrogates for quantum learning models. *Physical Review Letters*, 131(10):100803, 2023.

Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Physical Review A*, 99(3):032331, 2019.

Maria Schuld, Alex Bocharov, Krysta M Svore, and Nathan Wiebe. Circuit-centric quantum classifiers. *Physical Review A*, 101(3):032308, 2020.

Sukin Sim, Peter D Johnson, and Alán Aspuru-Guzik. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. *Advanced Quantum Technologies*, 2(12):1900070, 2019.

Andrea Skolik, Jarrod R McClean, Masoud Mohseni, Patrick Van Der Smagt, and Martin Leib. Layerwise learning for quantum neural networks. *Quantum Machine Intelligence*, 3(1):5, 2021.

Ryan Sweke, Frederik Wilde, Johannes Meyer, Maria Schuld, Paul K Fährmann, Barthélémy Meynard-Piganeau, and Jens Eisert. Stochastic gradient descent for hybrid quantum-classical optimization. *Quantum*, 4:314, 2020.

AV Uvarov and Jacob D Biamonte. On barren plateaus and cost function locality in variational quantum algorithms. *Journal of Physics A: Mathematical and Theoretical*, 54(24):245301, 2021.

Tyler Volkoff and Patrick J Coles. Large gradients via correlation in random parameterized quantum circuits. *Quantum Science & Technology*, 6(2):025008, 2021.

Samson Wang, Enrico Fontana, Marco Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J Coles. Noise-induced barren plateaus in variational quantum algorithms. *Nature Communications*, 12(1):6961, 2021.

David Wierichs, Josh Izaac, Cody Wang, and Cedric Yen-Yu Lin. General parameter-shift rules for quantum gradients. *Quantum*, 6:677, 2022.

A SYMBOLS

Symbol	Meaning
N_D	Number of qubits in the constructed circuit.
D	Total number of independent real circuit parameters.
r	Patch exponent, fixed in $(0, 1/2)$. The patch width is D^{-r}.
μ	Positive integer controlling the number of spectator qubits.
λ_μ	Polynomial gradient exponent $\mu/(3\ln 2)$.
σ_D	Initialization width D^{-r}.
Q	Source Clifford+T quantum circuit.
$p(Q)$	Output-one acceptance probability of Q.
t	Number of T or $T^\dagger$ gates after identity padding.
K	Number of small Z rotations replacing each T or $T^\dagger$ gate.
δ_D	Split-gate angle $\pi/(4K)$.
M	Number of spectator qubits and spectator parameters.
R_D	Number of dummy parameters and dummy qubits.
$\boldsymbol{\phi}$	Readout-angle block of length K.

Symbol	Meaning
$\boldsymbol{\vartheta}$	Spectator-angle block of length M.
$\boldsymbol{\beta}$	Hard-computation angle block of length tK.
$\boldsymbol{\gamma}$	Dummy-angle block of length R_D.
P_A	Handle Pauli string $X_{a_0} Z_{a_1} \cdots Z_{a_M}$.
H_Q	Three-Pauli normalized observable in Eq. (25).
g	Solvable handle expectation in Eq. (28).
C_Q	Full parameterized quantum cost.
$\boldsymbol{\theta}_Q^\star$	Symbolic hard point with inverse-polynomial gradient.
$\widehat{\boldsymbol{\theta}}_Q^\star$	Dyadic rational hard point.
BPP	Bounded-error probabilistic polynomial time.
BQP	Bounded-error quantum polynomial time.

B PRELIMINARIES

If $v, w \in \mathbb{C}^d$, their inner product is $\langle v, w\rangle = v^\dagger w = \sum_{j=1}^d \overline{v_j} w_j$, where the bar denotes complex conjugation and $\dagger$ denotes conjugate transpose. The Euclidean norm is $\|v\|_2 = \sqrt{\langle v, v\rangle}$. A square matrix $U \in \mathbb{C}^{d\times d}$ is *unitary* if $U^\dagger U = UU^\dagger = I_d$. A matrix H is *Hermitian* if $H^\dagger = H$.

The operator norm of a matrix A is

$$\|A\| = \sup_{v\neq 0} \frac{\|Av\|_2}{\|v\|_2}. \tag{58}$$

For Hermitian H, the operator norm equals the largest absolute value of an eigenvalue. We will repeatedly use

$$|\langle v, Hv\rangle| \le \|H\|\, \|v\|_2^2 . \tag{59}$$

B.1 TENSOR PRODUCTS

If $v \in \mathbb{C}^d$ and $w \in \mathbb{C}^k$, their tensor product $v \otimes w \in \mathbb{C}^{dk}$ is the vector whose entries are all products $v_i w_j$ in the standard lexicographic order. For matrices, the tensor product is defined blockwise and satisfies

$$(A \otimes B)(v \otimes w) = (Av) \otimes (Bw), \qquad (A \otimes B)^\dagger = A^\dagger \otimes B^\dagger. \tag{60}$$

Fact B.1. *Let $v_j \in \mathbb{C}^{d_j}$ and $A_j \in \mathbb{C}^{d_j \times d_j}$ for $j = 1, \ldots, m$. Then*

$$\left\langle \bigotimes_{j=1}^m v_j, \left(\bigotimes_{j=1}^m A_j\right) \bigotimes_{j=1}^m v_j \right\rangle = \prod_{j=1}^m \langle v_j, A_j v_j\rangle. \tag{61}$$

B.2 RANDOM VARIABLES

The notation $X \sim \mathrm{Unif}([-a, a])$ means that X is uniformly distributed on the interval $[-a, a]$, with density $1/(2a)$. If $X_1, \ldots, X_D$ are independent copies, then $(X_1, \ldots, X_D) \sim \mathrm{Unif}([-a, a]^D)$ denotes the product-uniform distribution on the cube.

Lemma B.2 (Chebyshev's inequality)**.** *Let X be a real random variable with finite variance. For every $t > 0$,*

$$\mathbb{P}\big(|X - \mathbb{E}X| \ge t\big) \le \frac{\mathrm{Var}(X)}{t^2}. \tag{62}$$

B.3 QUBITS AND COMPUTATIONAL BASIS STATES

A single qubit is a unit vector in $\mathbb{C}^2$. The standard basis vectors are

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{63}$$

An N-qubit pure state is a unit vector in $(\mathbb{C}^2)^{\otimes N} \cong \mathbb{C}^{2^N}$. For a bit string $z = z_1 \cdots z_N \in \{0,1\}^N$, write

$$|z\rangle = |z_1\rangle \otimes \cdots \otimes |z_N\rangle. \tag{64}$$

The all-zero state is abbreviated as $|0^N\rangle = |0\rangle^{\otimes N}$, or simply $|0\rangle$ when the number of qubits is clear.

A quantum gate for a closed system is represented by a unitary matrix. Applying a circuit U to $|0\rangle$ prepares the state $U|0\rangle$. Multiplying a state or a whole circuit by a scalar $\mathrm{e}^{\mathrm{i}\varphi}$ of unit modulus changes only its *global phase*. This phase has no effect on expectation values.

B.4 Pauli matrices and Pauli strings

The one-qubit Pauli matrices are

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{65}$$

They are Hermitian and unitary, and satisfy

$$X^2 = Y^2 = Z^2 = I. \tag{66}$$

An N-qubit *Pauli string* is a tensor product

$$P = P_1 \otimes \cdots \otimes P_N, \qquad P_j \in \{I, X, Y, Z\}. \tag{67}$$

Every Pauli string is Hermitian, has eigenvalues in $\{+1, -1\}$, and has operator norm one.

B.5 Pauli rotations

For $V \in \{X, Y, Z\}$ and $\theta \in \mathbb{R}$, define

$$R_V(\theta) = \exp\left(-\frac{\mathrm{i}\theta}{2} V\right). \tag{68}$$

Fact B.3 (Nielsen et al. (2000))**.** *For every $V \in \{X, Y, Z\}$ and $\theta, \varphi \in \mathbb{R}$,*

$$R_V(\theta) = \cos(\theta/2) I - \mathrm{i} \sin(\theta/2) V, \tag{69}$$

$$R_V(\theta) R_V(\varphi) = R_V(\theta + \varphi), \tag{70}$$

$$R_V(0) = I. \tag{71}$$

Moreover,

$$R_Z(\pi/4) = \mathrm{e}^{-\mathrm{i}\pi/8} T, \qquad T := \mathrm{diag}(1, \mathrm{e}^{\mathrm{i}\pi/4}). \tag{72}$$

Direct multiplication using Eq. (69) gives

$$\langle 0 | R_Y(\theta)^\dagger X R_Y(\theta) | 0 \rangle = \sin\theta, \tag{73}$$

$$\langle 0 | R_Y(\theta)^\dagger Z R_Y(\theta) | 0 \rangle = \cos\theta. \tag{74}$$

B.6 Measurements and expectation values

A Hermitian matrix H represents an observable. In a normalized state $|\psi\rangle$, its expectation value is

$$\langle H \rangle_\psi = \langle \psi | H | \psi \rangle. \tag{75}$$

If $H = P$ is a Pauli string, a projective measurement returns $+1$ or -1, and the mean outcome is Eq. (75). The projector onto $|1\rangle$ on a designated output qubit o is

$$\Pi_o = |1\rangle\langle 1|_o = \frac{I - Z_o}{2}. \tag{76}$$

Thus, for a circuit Q, the probability of observing output bit one is

$$p(Q) = \langle 0 | Q^\dagger \Pi_o Q | 0 \rangle. \tag{77}$$

By construction, $0 \le p(Q) \le 1$.

B.7 CLIFFORD GATES

The one-qubit Hadamard and phase gates are

$$H_{\mathrm{Had}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad S = \mathrm{diag}(1, \mathrm{i}). \tag{78}$$

Together with the two-qubit controlled-NOT (CNOT) gate, they generate the Clifford group. Equivalently, a unitary W is Clifford when conjugation maps every Pauli string to a Pauli string, up to an overall phase (Dehaene & De Moor, 2003). Stabilizer circuits, consisting only of Clifford operations and computational-basis preparation and measurement, admit efficient classical simulation by the Gottesman–Knill theorem (Gottesman, 1998; Aaronson & Gottesman, 2004; Nest, 2008). This motivates the Clifford-centered surrogate of Meyer et al. (Meyer et al., 2026). Our hardness proof does not assume that a particular stabilizer-based simulator is optimal.

The gate set consisting of Clifford gates together with the non-Clifford T gate is universal for quantum computation (Boykin et al., 2000). We use this standard finite-gate-set model when defining BQP.

The special angles used to measure departure from a Clifford point are indeed Clifford angles. By Eq. (72), $R_Z(\pi/2)$ equals the Clifford gate S up to global phase. Moreover,

$$R_X(\pi/2) = H_{\mathrm{Had}} R_Z(\pi/2) H_{\mathrm{Had}} \quad \text{up to global phase}, \qquad R_Y(\pi/2) = H_{\mathrm{Had}} Z,$$

where $Z = S^2$. Since Clifford operators form a group, every $R_V(k\pi/2)$ with $k \in \mathbb{Z}$ and $V \in \{X, Y, Z\}$ is Clifford up to global phase.

C PARAMETERIZED CIRCUITS AND TRAINABILITY

C.1 THE PAULI-ROTATED CLIFFORD CIRCUIT CLASS

Definition C.1. A Pauli-rotated Clifford (PRC) circuit on N qubits with D real parameters is a unitary family

$$U(\boldsymbol{\theta}) = R_{V_D}(\theta_D) W_D \cdots R_{V_1}(\theta_1) W_1, \qquad \boldsymbol{\theta} = (\theta_1, \ldots, \theta_D) \in \mathbb{R}^D, \tag{79}$$

where every V_j is a one-qubit Pauli operator and every W_j is an N-qubit Clifford unitary. A PRC cost is

$$C(\boldsymbol{\theta}) = \langle 0 | U(\boldsymbol{\theta})^\dagger H U(\boldsymbol{\theta}) | 0 \rangle, \tag{80}$$

where H is Hermitian.

The ordering in Eq. (79) is only a bookkeeping convention. Any circuit containing fixed Clifford gates and D one-parameter Pauli rotations can be placed in this form provided its final gate is a parameterized rotation. In our construction, an otherwise irrelevant dummy rotation will be placed last to ensure this condition.

Observation C.2. *Suppose a circuit is built from fixed Clifford gates and D one-parameter Pauli rotations, and suppose its final gate in time order is one of the parameterized rotations. Then its unitary can be written in the form Eq.* (79) *after multiplying consecutive fixed Clifford gates together.*

At $\boldsymbol{\theta} = 0$, every rotation is the identity by Eq. (71). Hence $U(0)$ is Clifford.

C.2 INITIALIZATION PATCHES

Definition C.3 (Near-Clifford cube)**.** For $D \in \mathbb{N}$ and $\sigma > 0$, define

$$\mathcal{P}_D(\sigma) = [-\sigma, \sigma]^D. \tag{81}$$

The product-uniform initialization on this patch is

$$\boldsymbol{\theta} \sim \mathrm{Unif}(\mathcal{P}_D(\sigma)), \tag{82}$$

meaning that the coordinates are independent and each is uniform on $[-\sigma, \sigma]$.

For a real number x and a nonempty set $S \subseteq \mathbb{R}$, we write

$$\operatorname{dist}(x, S) = \inf_{s \in S} |x - s| . \tag{83}$$

The word "near-Clifford" refers to the fact that $U(0)$ is Clifford and each coordinate is small. A rotation angle in $(\pi/2)\mathbb{Z}$ is itself Clifford up to global phase, so distance from the set $(\pi/2)\mathbb{Z}$ also measures departure from a Clifford angle.

C.3 Gradients and barren-plateau criteria

The gradient of a differentiable cost is

$$\nabla C(\boldsymbol{\theta}) = (\partial_{\theta_1} C(\boldsymbol{\theta}), \ldots, \partial_{\theta_D} C(\boldsymbol{\theta})) . \tag{84}$$

The squared Euclidean norm is

$$\|\nabla C(\boldsymbol{\theta})\|_2^2 = \sum_{j=1}^{D} \left(\partial_{\theta_j} C(\boldsymbol{\theta})\right)^2 . \tag{85}$$

There are several related definitions of a barren plateau in the literature. The original random-circuit formulation emphasizes exponentially small gradient variance (McClean et al., 2018). Later work studies dependence on circuit structure, observables, noise, and initialization (Grant et al., 2019; Cerezo et al., 2021; Pesah et al., 2021; Skolik et al., 2021; Uvarov & Biamonte, 2021; Volkoff & Coles, 2021; Holmes et al., 2022; Larocca et al., 2022; 2025). Meyer et al. use the expected squared gradient norm as their scaling diagnostic (Meyer et al., 2026). To avoid ambiguity, we state both notions.

Definition C.4 (Second-moment trainability)**.** A sequence of cost landscapes C_D with initialization distributions $\mathcal{D}_D$ has an inverse-polynomial gradient second moment if there exists a constant $a > 0$ such that

$$\mathbb{E}_{\boldsymbol{\theta} \sim \mathcal{D}_D} \left[\|\nabla C_D(\boldsymbol{\theta})\|_2^2 \right] = \Omega(D^{-a}). \tag{86}$$

Definition C.5 (Variance trainability)**.** A designated component $\partial_{\theta_k} C_D$ has inverse-polynomial variance if there exists a constant $b > 0$ such that

$$\operatorname{Var}_{\boldsymbol{\theta} \sim \mathcal{D}_D} (\partial_{\theta_k} C_D(\boldsymbol{\theta})) = \Omega(D^{-b}). \tag{87}$$

D Complexity theory

D.1 Decision and promise problems

A decision problem asks for a yes-or-no answer as a function of a finite binary input. A *promise problem* specifies two disjoint sets of valid inputs, called yes-instances and no-instances. Inputs outside both sets need not be handled correctly. Promise problems are natural for quantum acceptance probabilities because one usually promises a gap, such as probability at least $2/3$ versus at most $1/3$.

An algorithm runs in polynomial time when the number of elementary computational steps is bounded by a polynomial in the input length.

D.2 The classes BPP and BQP

Definition D.1 (Bounded-error probabilistic polynomial time)**.** A promise problem is in BPP if a randomized classical algorithm runs in polynomial time and outputs the correct answer with probability at least $2/3$ on every promised input.

The constant $2/3$ is conventional. Independent repetition followed by majority vote can amplify any fixed success probability strictly larger than $1/2$.

Definition D.2 (Bounded-error quantum polynomial time)**.** A promise problem is in BQP if there is a polynomial-time uniform family of polynomial-size quantum circuits over a fixed universal gate set, with one designated output qubit, such that measuring that qubit gives the correct answer with probability at least $2/3$ on every promised input.

We use the fixed gate alphabet

$$\mathcal{G} = \{H_{\mathrm{Had}}, S, S^\dagger, \mathrm{CNOT}, T, T^\dagger\}. \tag{88}$$

In Definition D.2, the fixed gate set is taken to be $\mathcal{G}$. The inverse symbols are redundant (for example, $S^\dagger = S^3$ and $T^\dagger = T^7$), but including them makes the gate list easier to read. Throughout this work, a "Clifford+T circuit" means a finite gate list over this fixed alphabet. The equivalence of standard universal finite gate sets is a foundational result of quantum complexity theory (Bernstein & Vazirani, 1993; Boykin et al., 2000; Dawson & Nielsen, 2005). "Uniform" means that an ordinary deterministic polynomial-time program can generate the gate list from the input.

Definition D.3. A promise problem A reduces to a promise problem B if a deterministic polynomial-time classical map sends every promised instance x of A to a promised instance $f(x)$ of B, preserving yes and no answers. A problem is BQP-hard if every problem in BQP reduces to it.

D.3 Quantum circuit acceptance

Definition D.4 (Quantum circuit acceptance, QCA)**.** An instance is a polynomial-size circuit Q over the alphabet $\mathcal{G}$ in Eq. (88), acting on the all-zero input and having a designated output qubit. The input encoding lists the gates and their operand labels. We canonically relabel the finitely many mentioned qubits by consecutive integers. Consequently, the number of qubits is at most a constant multiple of the binary description length. Let $p(Q)$ be the probability of measuring one on the designated qubit. The promise is

$$\text{yes: } p(Q) \geq \frac{2}{3}, \qquad \text{no: } p(Q) \leq \frac{1}{3}. \tag{89}$$

The task is to decide whether the instance is yes or no.

Proposition D.5 (Aharonov et al. (2017), QCA is BQP-complete)**.** *The promise problem in Definition D.4 belongs to* BQP *and is* BQP*-hard under the reductions of Definition D.3.*

To conclude that no bounded-error classical polynomial-time algorithm evaluates every hard cost, we assume

$$\boxed{\mathsf{BQP} \not\subseteq \mathsf{BPP}.} \tag{90}$$

This statement is widely believed but unproved. It says that at least one efficiently solvable quantum decision problem is not efficiently solvable by a randomized classical computer.

Our reduction proves that a classical evaluator for our costs would imply the opposite containment $\mathsf{BQP} \subseteq \mathsf{BPP}$.

E Analytic tools for Pauli-rotation costs

E.1 Parameter-shift identities and derivative bounds

The parameter-shift rule is standard in variational quantum algorithms (Mitarai et al., 2018; Crooks, 2019; Schuld et al., 2019; Banchi & Crooks, 2021; Wierichs et al., 2022).

For $k \in \{1, \ldots, D\}$, let e_k denote the k-th standard basis vector of $\mathbb{R}^D$.

Lemma E.1 (One-parameter sinusoidal form)**.** *Fix all parameters of a PRC cost except θ_k. Then there exist real constants a, b, c, depending on the fixed parameters but not on θ_k, such that*

$$C(\boldsymbol{\theta}) = a + b\cos\theta_k + c\sin\theta_k. \tag{91}$$

Proof. Write the circuit as $U(\boldsymbol{\theta}) = AR_V(\theta_k)B$, where A and B contain all gates after and before the k-th parameterized rotation, respectively, and V is the corresponding Pauli operator. Set $|\phi\rangle = B|0\rangle$ and $G = A^\dagger H A$. By Eq. (69), with $u = \theta_k/2$,

$$R_V(\theta_k) = \cos u\, I - \mathrm{i}\sin u\, V, \qquad R_V(\theta_k)^\dagger = \cos u\, I + \mathrm{i}\sin u\, V.$$

Expanding $\langle\phi|R_V(\theta_k)^\dagger G R_V(\theta_k)|\phi\rangle$ gives a linear combination of $\cos^2 u$, $\sin^2 u$, and $\sin u\cos u$, with real coefficients because the total expression is real. The identities

$$\cos^2 u = \frac{1+\cos(2u)}{2}, \qquad \sin^2 u = \frac{1-\cos(2u)}{2}, \qquad 2\sin u\cos u = \sin(2u)$$

convert this expression into Eq. (91), since $2u = \theta_k$. □

Lemma E.2 (Multivariate parameter-shift rule)**.** *Let C be a PRC cost. For every multi-index $\alpha = (\alpha_1, \ldots, \alpha_D) \in \mathbb{N}_0^D$,*

$$D^\alpha C(\boldsymbol{\theta}) = \frac{1}{2^{|\alpha|}} \sum_{0 \le j \le \alpha} (-1)^{|j|} \binom{\alpha}{j} C\Big(\boldsymbol{\theta} + \frac{\pi}{2}(\alpha - 2j)\Big), \tag{92}$$

where

$$|\alpha| = \sum_{k=1}^{D} \alpha_k, \quad \binom{\alpha}{j} = \prod_{k=1}^{D} \binom{\alpha_k}{j_k}, \quad 0 \le j \le \alpha \text{ means } 0 \le j_k \le \alpha_k \text{ for all } k.$$

Consequently,

$$\sup_{\boldsymbol{\theta} \in \mathbb{R}^D} |D^\alpha C(\boldsymbol{\theta})| \le \|H\|. \tag{93}$$

Proof. By Lemma E.1, differentiation with respect to θ_k gives

$$\partial_{\theta_k} C(\boldsymbol{\theta}) = -b \sin\theta_k + c \cos\theta_k.$$

On the other hand,

$$\frac{1}{2}\left[C\Big(\boldsymbol{\theta} + \frac{\pi}{2}e_k\Big) - C\Big(\boldsymbol{\theta} - \frac{\pi}{2}e_k\Big)\right] = -b \sin\theta_k + c \cos\theta_k.$$

Thus

$$\partial_{\theta_k} C(\boldsymbol{\theta}) = \frac{1}{2}\left[C\Big(\boldsymbol{\theta} + \frac{\pi}{2}e_k\Big) - C\Big(\boldsymbol{\theta} - \frac{\pi}{2}e_k\Big)\right]. \tag{94}$$

Applying Eq. (94) repeatedly α_k times in coordinate k yields

$$\partial_{\theta_k}^{\alpha_k} C(\boldsymbol{\theta}) = \frac{1}{2^{\alpha_k}} \sum_{j_k=0}^{\alpha_k} (-1)^{j_k} \binom{\alpha_k}{j_k} C\Big(\boldsymbol{\theta} + \frac{\pi}{2}(\alpha_k - 2j_k)e_k\Big).$$

Partial derivatives commute because C is analytic, so applying this identity in all coordinates and multiplying the coefficients gives Eq. (92).

For every parameter point $\boldsymbol{\eta}$, Eq. (59) and unitarity imply

$$|C(\boldsymbol{\eta})| = \left|\langle 0|U(\boldsymbol{\eta})^\dagger H U(\boldsymbol{\eta})|0\rangle\right| \le \|H\|.$$

Taking absolute values in Eq. (92) therefore gives

$$|D^\alpha C(\boldsymbol{\theta})| \le \frac{\|H\|}{2^{|\alpha|}} \sum_{0 \le j \le \alpha} \binom{\alpha}{j}.$$

The product form of the binomial theorem gives

$$\sum_{0 \le j \le \alpha} \binom{\alpha}{j} = \prod_{k=1}^{D} \sum_{j_k=0}^{\alpha_k} \binom{\alpha_k}{j_k} = \prod_{k=1}^{D} 2^{\alpha_k} = 2^{|\alpha|}.$$

Substitution proves Eq. (93). □

Corollary E.3 (Lipschitz bounds)**.** *For every two parameter vectors $\boldsymbol{\theta}, \boldsymbol{\varphi} \in \mathbb{R}^D$ and every coordinate k,*

$$|C(\boldsymbol{\theta}) - C(\boldsymbol{\varphi})| \le \|H\| \, \|\boldsymbol{\theta} - \boldsymbol{\varphi}\|_1, \tag{95}$$

$$|\partial_{\theta_k} C(\boldsymbol{\theta}) - \partial_{\theta_k} C(\boldsymbol{\varphi})| \le \|H\| \, \|\boldsymbol{\theta} - \boldsymbol{\varphi}\|_1. \tag{96}$$

Proof. Define the line segment $\gamma(t) = \boldsymbol{\varphi} + t(\boldsymbol{\theta} - \boldsymbol{\varphi})$ for $t \in [0, 1]$. The chain rule gives

$$C(\boldsymbol{\theta}) - C(\boldsymbol{\varphi}) = \int_0^1 \sum_{j=1}^{D} (\theta_j - \varphi_j) \partial_{\theta_j} C(\gamma(t))\, dt.$$

By Eq. (93) with $\alpha = e_j$, every first derivative has absolute value at most $\|H\|$. Taking absolute values proves Eq. (95). Applying the same argument to the function $\partial_{\theta_k} C$ and using Eq. (93) for the second derivatives $\partial_{\theta_j}\partial_{\theta_k} C$ proves Eq. (96). □

E.2 Uniform trigonometric moments

Lemma E.4. *Let $X \sim \mathrm{Unif}([-\tau, \tau])$ with $\tau > 0$. Define*

$$c(\tau) = \frac{\sin \tau}{\tau}, \qquad b(\tau) = \frac{1}{2} + \frac{\sin(2\tau)}{4\tau}. \tag{97}$$

Then

$$\mathbb{E}[\cos X] = c(\tau), \tag{98}$$

$$\mathbb{E}[\cos^2 X] = b(\tau). \tag{99}$$

If $X_1, \ldots, X_K$ are independent copies and $S_K = \sum_{\ell=1}^{K} X_\ell$, then

$$\mathbb{E}[\cos S_K] = c(\tau)^K, \tag{100}$$

$$\mathbb{E}[\cos^2 S_K] = \frac{1}{2}\left[1 + \left(\frac{\sin(2\tau)}{2\tau}\right)^K\right]. \tag{101}$$

For $0 < \tau < \pi/2$, the quantity in Eq. (101) lies between $1/2$ and 1.

Proof. Direct integration gives

$$\mathbb{E}[\cos X] = \frac{1}{2\tau}\int_{-\tau}^{\tau} \cos x \, dx = \frac{\sin \tau}{\tau} = c(\tau).$$

Using $\cos^2 x = (1 + \cos 2x)/2$ gives

$$\mathbb{E}[\cos^2 X] = \frac{1}{2\tau}\int_{-\tau}^{\tau} \frac{1 + \cos(2x)}{2}\, dx = \frac{1}{2} + \frac{1}{4\tau}\int_{-\tau}^{\tau} \cos(2x)\, dx = \frac{1}{2} + \frac{\sin(2\tau)}{4\tau} = b(\tau).$$

By symmetry, $\mathbb{E}[\sin X] = 0$, so

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}X}] = \mathbb{E}[\cos X] + \mathrm{i}\mathbb{E}[\sin X] = c(\tau).$$

Independence implies

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}S_K}] = \prod_{\ell=1}^{K} \mathbb{E}[\mathrm{e}^{\mathrm{i}X_\ell}] = c(\tau)^K.$$

Taking real parts proves Eq. (100). Replacing X by $2X$ and using $\cos^2 S_K = (1 + \cos 2S_K)/2$ proves Eq. (101). If $0 < \tau < \pi/2$, then $0 < \sin(2\tau)/(2\tau) \le 1$, establishing the final bounds. □

E.3 Small-angle asymptotics

Lemma E.5. *As $u \to 0$,*

$$\log b(u) = -\frac{u^2}{3} + \frac{u^4}{90} + O(u^6), \tag{102}$$

$$2\log c(u) = -\frac{u^2}{3} - \frac{u^4}{90} + O(u^6), \tag{103}$$

$$\log b(u) - 2\log c(u) = \frac{u^4}{45} + O(u^6), \tag{104}$$

$$\log(\cos u) = -\frac{u^2}{2} - \frac{u^4}{12} + O(u^6). \tag{105}$$

Proof. Taylor's theorem at zero gives

$$\sin u = u - \frac{u^3}{6} + \frac{u^5}{120} + O(u^7),$$

$$\sin(2u) = 2u - \frac{(2u)^3}{6} + \frac{(2u)^5}{120} + O(u^7),$$

$$\cos u = 1 - \frac{u^2}{2} + \frac{u^4}{24} + O(u^6).$$

Substitution into Eq. (97) yields

$$c(u) = 1 - \frac{u^2}{6} + \frac{u^4}{120} + O(u^6),$$
$$b(u) = 1 - \frac{u^2}{3} + \frac{u^4}{15} + O(u^6).$$

For $z \to 0$, $\log(1+z) = z - z^2/2 + O(z^3)$. Applying this formula to the expansions above gives

$$\log b(u) = \left(-\frac{u^2}{3} + \frac{u^4}{15}\right) - \frac{1}{2}\left(\frac{u^4}{9}\right) + O(u^6) = -\frac{u^2}{3} + \frac{u^4}{90} + O(u^6),$$
$$\log c(u) = \left(-\frac{u^2}{6} + \frac{u^4}{120}\right) - \frac{1}{2}\left(\frac{u^4}{36}\right) + O(u^6) = -\frac{u^2}{6} - \frac{u^4}{180} + O(u^6),$$
$$\log(\cos u) = \left(-\frac{u^2}{2} + \frac{u^4}{24}\right) - \frac{1}{2}\left(\frac{u^4}{4}\right) + O(u^6) = -\frac{u^2}{2} - \frac{u^4}{12} + O(u^6).$$

Doubling the second line and subtracting it from the first proves Eqs. (103) and (104). □

Fact E.6. *If $x_D > 0$ and $x_D \to 0$, then*

$$1 - \mathrm{e}^{-x_D} = x_D(1 + o(1)). \tag{106}$$

F Additional statements and proofs

F.1 Proof of Lemma 3.3

Proof. There is one dummy qubit for each of the R_D dummy parameters, so $N_D \geq R_D$. By Eq. (13), $R_D/D \to 1/8$, giving $N_D = \Omega(D)$.

For the upper bound, the hard circuit uses at most s qubits, hence $O(t) = O(D^{1-r})$ qubits by Eqs. (3) and (7). The handle uses $M + 1 = o(D)$ qubits by Eq. (9). The dummy register uses $R_D \leq D$ qubits. Therefore $N_D = O(D)$. □

F.2 Proof of Proposition 3.4

Proof. The operators Z_o, P_A, and $P_A Z_o$ are three distinct Hermitian Pauli strings, so Eq. (25) is a Hermitian three-term Pauli decomposition with coefficients $-1/5$, $3/5$, and $-1/5$.

Because $P_A^2 = I$, the operator $A_A = (I + P_A)/2$ is the orthogonal projector onto the $+1$ eigenspace of P_A, and its eigenvalues are 0 and 1. The projector Π_o has eigenvalues 0 and 1, so $B_o = (I + \Pi_o)/2$ has eigenvalues $1/2$ and 1. The two operators act on disjoint registers. Consequently, the eigenvalues of $A_A \otimes B_o$ belong to

$$\left\{0, \frac{1}{2}, 1\right\},$$

and all three values occur when the registers are nontrivial. The eigenvalues of $A_A \otimes B_o - 3I/8$ therefore belong to

$$\left\{-\frac{3}{8}, \frac{1}{8}, \frac{5}{8}\right\}.$$

Multiplication by $8/5$ gives eigenvalues among $\{-3/5, 1/5, 1\}$, including 1. Since H_Q is Hermitian, its operator norm is the largest absolute eigenvalue, namely one. □

F.3 Cost and gradient formulas

Theorem F.1. *For the circuit in Eq. (21) and observable in Eq. (25),*

$$C_Q(\boldsymbol{\theta}) = \frac{2}{5}\left(1 + g(\boldsymbol{\phi}, \boldsymbol{\vartheta})\right)\left(1 + p_Q(\boldsymbol{\beta})\right) - \frac{3}{5}. \tag{107}$$

For every readout coordinate ϕ_ℓ,

$$\partial_{\phi_\ell} C_Q(\boldsymbol{\theta}) = \frac{2}{5} \cos S(\boldsymbol{\phi}) \left(\prod_{j=1}^{M} \cos \vartheta_j \right) (1 + p_Q(\boldsymbol{\beta})) \,. \tag{108}$$

In particular,

$$\frac{2}{5} \le \partial_{\phi_\ell} C_Q(0) \le \frac{4}{5}. \tag{109}$$

Proof. By Eq. (70), the K rotations on qubit a_0 combine to $R_Y(S(\boldsymbol{\phi}))$. Applying Eqs. (73) and (74), followed by the product factorization in Fact B.1, gives

$$\langle P_A \rangle = \sin S(\boldsymbol{\phi}) \prod_{j=1}^{M} \cos \vartheta_j = g(\boldsymbol{\phi}, \boldsymbol{\vartheta}).$$

On the hard output qubit, Eq. (76) implies

$$\langle Z_o \rangle = 1 - 2p_Q(\boldsymbol{\beta}).$$

The handle state and hard-block state are tensor factors, so Fact B.1 also gives

$$\langle P_A Z_o \rangle = g(\boldsymbol{\phi}, \boldsymbol{\vartheta}) \, (1 - 2p_Q(\boldsymbol{\beta})) \,.$$

Substituting these three expectations into Eq. (25),

$$\begin{aligned} C_Q(\boldsymbol{\theta}) &= \frac{1}{5} \left[-(1 - 2p_Q) + 3g - g(1 - 2p_Q) \right] \\ &= \frac{1}{5} \left[-1 + 2p_Q + 2g + 2gp_Q \right] \\ &= \frac{2}{5}(1 + g)(1 + p_Q) - \frac{3}{5}, \end{aligned}$$

which proves Eq. (107). Differentiating Eq. (28) with respect to any ϕ_ℓ gives

$$\partial_{\phi_\ell} g = \cos S(\boldsymbol{\phi}) \prod_{j=1}^{M} \cos \vartheta_j.$$

Differentiating Eq. (107) proves Eq. (108). At the origin, all cosine factors equal one and $0 \le p_Q(0) \le 1$, giving Eq. (109). □

F.4 Proof of Theorem 4.1

Proof. From Eq. (108), write

$$\partial_{\phi_\ell} C_Q = \frac{2}{5} ABY, \tag{110}$$

where

$$A = \cos S(\boldsymbol{\phi}), \qquad B = 1 + p_Q(\boldsymbol{\beta}), \qquad Y = \prod_{j=1}^{M} \cos \vartheta_j. \tag{111}$$

The three random variables depend on disjoint coordinate blocks and are therefore independent. By Lemma E.4,

$$\mathbb{E}[A^2] = \frac{1}{2} \left[1 + \left(\frac{\sin(2\sigma_D)}{2\sigma_D} \right)^K \right], \qquad \mathbb{E}[Y^2] = b(\sigma_D)^M. \tag{112}$$

For sufficiently large D, $0 < \sigma_D < \pi/2$, so

$$\frac{1}{2} \le \mathbb{E}[A^2] \le 1. \tag{113}$$

By Eq. (29),

$$1 \le \mathbb{E}[B^2] \le 4. \tag{114}$$

Independence, Eq. (110), and Eqs. (112) to (114) imply the two-sided bound

$$\frac{2}{25}b(\sigma_D)^M \le \mathbb{E}\Big[(\partial_{\phi_\ell} C_Q)^2\Big] \le \frac{16}{25}b(\sigma_D)^M. \tag{115}$$

It remains to evaluate $b(\sigma_D)^M$. By Eq. (102),

$$M \log b(\sigma_D) = -\frac{1}{3}M\sigma_D^2 + O(M\sigma_D^4).$$

By Eq. (10),

$$-\frac{1}{3}M\sigma_D^2 = -\frac{\mu}{3}\log_2 D + O(1) + o(1) = -\lambda_\mu \ln D + O(1) + o(1). \tag{116}$$

Furthermore,

$$M\sigma_D^4 = (M\sigma_D^2)\sigma_D^2 = O(D^{-2r}\log D) = o(1). \tag{117}$$

Exponentiating Eq. (116) and using Eq. (117) gives

$$b(\sigma_D)^M = \Theta(D^{-\lambda_\mu}). \tag{118}$$

Substitution into Eq. (115) proves Eq. (31).

Finally, Eq. (85) implies

$$\|\nabla C_Q\|_2^2 \ge (\partial_{\phi_\ell} C_Q)^2.$$

Taking expectations proves the first lower bound in Eq. (32). The conversion to qubit count follows from $N_D = \Theta(D)$ in Lemma 3.3. □

F.5 Proof of Proposition 4.2

Proof. Continue using A, B, Y from Eq. (111). Independence gives

$$\begin{aligned}\mathrm{Var}(ABY) &= \mathbb{E}[A^2]\mathbb{E}[B^2]\mathbb{E}[Y^2] - (\mathbb{E}A)^2(\mathbb{E}B)^2(\mathbb{E}Y)^2 \\ &= \left(\mathbb{E}[A^2]\mathbb{E}[B^2] - (\mathbb{E}A)^2(\mathbb{E}B)^2\right)\mathbb{E}[Y^2] + (\mathbb{E}A)^2(\mathbb{E}B)^2\mathrm{Var}(Y).\end{aligned}$$

The first term is nonnegative because $\mathbb{E}[A^2] \ge (\mathbb{E}A)^2$ and $\mathbb{E}[B^2] \ge (\mathbb{E}B)^2$. Therefore

$$\mathrm{Var}(ABY) \ge (\mathbb{E}A)^2(\mathbb{E}B)^2\mathrm{Var}(Y). \tag{119}$$

By Eq. (100), $\mathbb{E}A = c(\sigma_D)^K$. Using $\log c(u) = -u^2/6 + O(u^4)$ from the proof of Lemma E.5,

$$K \log c(\sigma_D) = O(K\sigma_D^2) = O(D^{-r}) \longrightarrow 0,$$

because $K = \Theta(D^r)$. Hence $\mathbb{E}A \to 1$, so $\mathbb{E}A \ge 1/2$ for all sufficiently large D. Also, $B \ge 1$ pointwise, so $\mathbb{E}B \ge 1$.

It remains to estimate $\mathrm{Var}(Y)$. By independence and Lemma E.4,

$$\mathrm{Var}(Y) = b(\sigma_D)^M - c(\sigma_D)^{2M}. \tag{120}$$

Define

$$x_D := M\left(\log b(\sigma_D) - 2\log c(\sigma_D)\right). \tag{121}$$

By Eq. (104),

$$x_D = \frac{M\sigma_D^4}{45}\left(1 + O(\sigma_D^2)\right) \tag{122}$$

$$= \Theta(D^{-2r}\log D), \tag{123}$$

and in particular $x_D > 0$ for large D and $x_D \to 0$. Factoring Eq. (120),

$$\mathrm{Var}(Y) = b(\sigma_D)^M\left(1 - \mathrm{e}^{-x_D}\right). \tag{124}$$

By Fact E.6, $1 - \mathrm{e}^{-x_D} = x_D(1 + o(1))$. Combining this with Eqs. (118) and (123) gives

$$\mathrm{Var}(Y) = \Theta\Big(D^{-(\lambda_\mu + 2r)}\log D\Big). \tag{125}$$

Finally, Eqs. (110) and (119) and the bounds $\mathbb{E}A \ge 1/2$, $\mathbb{E}B \ge 1$ yield

$$\mathrm{Var}(\partial_{\phi_\ell} C_Q) = \frac{4}{25}\mathrm{Var}(ABY) \ge \frac{1}{25}\mathrm{Var}(Y).$$

Substituting Eq. (125) proves Eq. (33). □

F.6 Proof of Theorem 4.3

Proof. The proof of Theorem 4.1 did not use the special choice $\tau_D = \sigma_D$ until the final asymptotic step. Replacing σ_D by τ_D in Eq. (115) gives

$$\frac{2}{25} b(\tau_D)^M \le \mathbb{E}[(\partial_{\phi_\ell} C_Q)^2] \le \frac{16}{25} b(\tau_D)^M \tag{126}$$

for all sufficiently large D, since then $\tau_D < \pi/2$. By Eq. (102),

$$M \log b(\tau_D) = -\frac{M\tau_D^2}{3} + O(M\tau_D^4).$$

Exponentiating and applying Eq. (126) proves Eq. (34).

If $\tau_D = D^{-1/2}$, then

$$M\tau_D^2 = O(D^{2r-1} \log D) = o(1)$$

and $M\tau_D^4 = o(1)$, so Eq. (34) is $\Theta(1)$. This proves Eq. (35).

The case $\tau_D = D^{-r}$ is Theorem 4.1, proving Eq. (36).

Finally, let $\tau_D = D^{-r'}$ with $0 < r' < r$. From Eq. (6),

$$\frac{M\tau_D^2}{3} = \frac{\mu}{3} D^{2(r-r')} \log_2 D \,(1 + o(1)) = \lambda_\mu D^{2(r-r')} \ln D \,(1 + o(1)). \tag{127}$$

Moreover,

$$\frac{M\tau_D^4}{M\tau_D^2} = \tau_D^2 \to 0,$$

so the fourth-order term is lower order. Substituting Eq. (127) into Eq. (34) gives Eq. (37). Since $D^{2(r-r')} \to \infty$, its exponent eventually exceeds every fixed $a > 0$. □

F.7 Proof of Lemma 5.1

Proof. For $x > 0$, $\arctan x < x$. Since $J_D \ge 2D^r$,

$$0 < \eta_D < \frac{2}{J_D} \le D^{-r} = \sigma_D,$$

proving Eq. (39). Furthermore,

$$\frac{\eta_D}{\sigma_D} = 2D^r \arctan(1/J_D) = \frac{2D^r}{J_D} \left[J_D \arctan(1/J_D) \right].$$

The first factor tends to one because $J_D/(2D^r) \to 1$, and the second tends to one because $\arctan x / x \to 1$ as $x \to 0$. This proves Eq. (40).

Let $u = \arctan(1/J_D)$, so $\tan u = 1/J_D$ and $\eta_D = 2u$. The double-angle identity $\cos(2u) = (1 - \tan^2 u)/(1 + \tan^2 u)$ gives

$$\cos \eta_D = \frac{1 - J_D^{-2}}{1 + J_D^{-2}} = \frac{J_D^2 - 1}{J_D^2 + 1},$$

proving Eq. (41).

By Eq. (105),

$$\log a_D = M \log(\cos \eta_D) = -\frac{1}{2} M\eta_D^2 + O(M\eta_D^4). \tag{128}$$

From Eq. (40), $\eta_D^2 = \sigma_D^2(1 + o(1))$. The convergence rate is sufficient after multiplication by $\log D$. Indeed, $J_D = 2D^r + O(1)$ and $\arctan x = x + O(x^3)$ imply $\eta_D/\sigma_D = 1 + O(D^{-r})$, so $M\eta_D^2 = M\sigma_D^2 + o(1)$ because $D^{-r} \log D \to 0$. Using Eq. (10),

$$M\eta_D^2 = \mu \log_2 D + O(1) + o(1) = 3\lambda_\mu \ln D + O(1) + o(1). \tag{129}$$

Also, $M\eta_D^4 = O(D^{-2r} \log D) = o(1)$. Substitution into Eq. (128) yields

$$\log a_D = -\frac{3\lambda_\mu}{2} \ln D + O(1) + o(1).$$

Exponentiating proves both statements in Eq. (43). □

F.8 PROPERTIES OF THE DISTINGUISHED HARD POINT

Theorem F.2. *The point $\boldsymbol{\theta}_Q^\star$ belongs to $\mathcal{P}_D(\sigma_D)$ and satisfies*

$$\left\|\boldsymbol{\theta}_Q^\star\right\|_1 = \Theta(D^{1-r}), \tag{130}$$

$$\#\left\{j : \frac{\pi}{7}D^{-r} \le \operatorname{dist}\Big(\theta_{Q,j}^\star, \frac{\pi}{2}\mathbb{Z}\Big) \le \frac{2\pi}{7}D^{-r}\right\} \ge \frac{D}{2} \quad \text{for all sufficiently large } D. \tag{131}$$

Moreover, with $p(Q)$ from Eq. (77)*,*

$$C_Q(\boldsymbol{\theta}_Q^\star) = \frac{2}{5}\left(1 + \frac{a_D}{\sqrt{2}}\right)(1 + p(Q)) - \frac{3}{5}, \tag{132}$$

$$\partial_{\phi_1} C_Q(\boldsymbol{\theta}_Q^\star) = \frac{2a_D}{5\sqrt{2}}(1 + p(Q)), \tag{133}$$

and hence

$$\left(\partial_{\phi_1} C_Q(\boldsymbol{\theta}_Q^\star)\right)^2 = \Theta(D^{-3\lambda_\mu}). \tag{134}$$

Proof. The readout and hard-block angles have magnitude $\delta_D < \sigma_D$ by Eq. (14). The spectator angles satisfy $\eta_D \le \sigma_D$ by Eq. (39), and the dummy angles are zero. Thus $\boldsymbol{\theta}_Q^\star \in \mathcal{P}_D(\sigma_D)$.

The ℓ_1 contribution from the tK hard-block parameters is

$$tK\delta_D = tK\frac{\pi}{4K} = \frac{\pi}{4}t.$$

The readout contribution is $K\delta_D = \pi/4$. The spectator contribution is

$$M\eta_D = O(D^{2r}\log D)O(D^{-r}) = O(D^r \log D).$$

Since $r < 1/2$, one has $D^r \log D = o(D^{1-r})$. By Eq. (7), $t = \Theta(D^{1-r})$. Therefore the hard-block contribution is $\Theta(D^{1-r})$ and dominates the other contributions, proving Eq. (130).

There are tK hard-block coordinates, each equal to $\pm\delta_D$. By Eq. (8), $tK/D \to 7/8$, so $tK \ge D/2$ for all sufficiently large D. Also,

$$K \le \frac{7}{8}D^r + 1 \le \frac{7}{4}D^r$$

for all sufficiently large D, while $K \ge (7/8)D^r$. Hence

$$\frac{\pi}{7}D^{-r} \le \delta_D = \frac{\pi}{4K} \le \frac{2\pi}{7}D^{-r}.$$

For large D, $0 < \delta_D < \pi/4$, so the nearest element of $(\pi/2)\mathbb{Z}$ is zero. The $tK \ge D/2$ hard-block coordinates therefore satisfy the two-sided bounds in Eq. (131).

At the point in Eq. (45),

$$S(\boldsymbol{\phi}^\star) = K\delta_D = \frac{\pi}{4}, \qquad \prod_{j=1}^{M} \cos\vartheta_j^\star = (\cos\eta_D)^M = a_D.$$

Thus Eq. (28) gives

$$g(\boldsymbol{\phi}^\star, \boldsymbol{\vartheta}^\star) = \frac{a_D}{\sqrt{2}}. \tag{135}$$

By Observation 3.2, $p_Q(\boldsymbol{\beta}^\star) = p(Q)$. Substituting these identities into Eq. (107) proves Eq. (132). Substitution into Eq. (108), using $\cos(\pi/4) = 1/\sqrt{2}$, proves Eq. (133). Finally, $1 \le 1 + p(Q) \le 2$ and Eq. (43) imply Eq. (134). □

F.9 Proof of Theorem 5.3

Proof. We reduce from QCA, which is BQP-complete by Proposition D.5. The reduction maps Q to the finite description of the circuit U_Q, the three-term observable H_Q, and the point $\boldsymbol{\theta}^\star_Q$.

First we verify the promise. Let $g_D = a_D/\sqrt{2}$, so $0 \le g_D \le 1/32$ by Eq. (47). If Q is a yes-instance, then $p(Q) \ge 2/3$, and Eq. (132) gives

$$C_Q(\boldsymbol{\theta}^\star_Q) \ge \frac{2}{5} \cdot 1 \cdot \frac{5}{3} - \frac{3}{5} = \frac{1}{15}.$$

If Q is a no-instance, then $p(Q) \le 1/3$, and

$$\begin{aligned} C_Q(\boldsymbol{\theta}^\star_Q) &\le \frac{2}{5}(1+g_D)\frac{4}{3} - \frac{3}{5} \\ &= -\frac{1}{15} + \frac{8}{15} g_D \\ &\le -\frac{1}{15} + \frac{1}{60} = -\frac{1}{20}. \end{aligned}$$

Thus the map preserves yes and no instances.

The reduction runs in polynomial time. Indeed, $D = t^{1/(1-r)} + O(1)$ is a fixed power of $t = \Theta(s)$; $K = O(D^r)$; $M = O(D^{2r} \log D)$; and the constructed circuit contains $O(D)$ parameterized gates by Observation 3.1. Since r is a fixed rational, integer ceilings of powers such as D^r can be computed by binary search using fixed-degree integer comparisons. The angles $\pi/(4K)$ and $2\arctan(1/J_D)$ have the finite symbolic encodings specified in Definition 5.2, of length $O(\log D)$. Thus the output description has polynomial length. A stronger conventional formulation using only dyadic rational angles is proved in Lemma 5.5.

Therefore QCA reduces to the transition-patch cost-sign problem, proving BQP-hardness.

Now suppose a randomized classical polynomial-time estimator returns $\widetilde{C}$ with $\left|\widetilde{C} - C_Q(\boldsymbol{\theta}^\star_Q)\right| \le 1/100$ with probability at least $2/3$. On that event, a yes-instance satisfies

$$\widetilde{C} \ge \frac{1}{15} - \frac{1}{100} > 0,$$

whereas a no-instance satisfies

$$\widetilde{C} \le -\frac{1}{20} + \frac{1}{100} < 0.$$

The sign of $\widetilde{C}$ therefore decides QCA with probability at least $2/3$ in classical polynomial time. By BQP-completeness of QCA, this implies Eq. (49). Under Eq. (90), the estimator cannot exist. □

F.10 Proof of Corollary 5.4

Proof. By Eq. (133), the gradient equals

$$G_D(1 + p(Q)).$$

If Q is a yes-instance, this is at least $5G_D/3$. If Q is a no-instance, it is at most $4G_D/3$. Let $\widetilde{g}$ be an estimate with error at most $G_D/12$. Then, on the event of a successful estimation,

$$Q \text{ yes} \quad \Longrightarrow \quad \widetilde{g} \ge \left(\frac{5}{3} - \frac{1}{12}\right) G_D = \frac{19}{12} G_D > \frac{3}{2} G_D,$$

while

$$Q \text{ no} \quad \Longrightarrow \quad \widetilde{g} \le \left(\frac{4}{3} + \frac{1}{12}\right) G_D = \frac{17}{12} G_D < \frac{3}{2} G_D.$$

Comparison with the efficiently computable threshold $3G_D/2$ decides QCA. By Eq. (43), $G_D = \Theta(D^{-3\lambda_\mu/2})$, so $G_D/12$ is inverse polynomial. □

F.11 Proof of Lemma 5.5

Observation F.3 (Polynomial-time dyadic approximation of the required angles)**.** *Given integers $n \geq 2$ and $B \geq 1$, a deterministic classical algorithm can produce rational intervals of width less than 2^{-B} containing $\arctan(1/n)$ and π, respectively. The endpoints are dyadic rationals with $O(B)$ fractional bits, and the running time is polynomial in $B + \log n$.*

Proof of Lemma 5.5. Let

$$B_D = (\mu + 4)\lceil \log_2 D \rceil . \tag{136}$$

Invoke Observation F.3 with precision $B_D + 3 + \lceil \log_2(4K) \rceil$ for π, and divide the resulting interval by $4K$. Invoke it with precision $B_D + 4$ for $\arctan(1/J_D)$, and multiply that interval by two. In both cases the resulting interval around the positive ideal magnitude $\delta_D = \pi/(4K)$ or $\eta_D = 2\arctan(1/J_D)$ has width less than $2^{-(B_D+2)}$. These extra precision requests are still $O(\log D)$. For sufficiently large D the lower endpoints are positive.

From each interval, choose the largest nonnegative multiple of $2^{-(B_D+1)}$ that is no greater than the lower endpoint. Restore the sign of each hard-block coordinate, and leave every zero coordinate equal to zero. The chosen magnitude is no greater than the exact magnitude, and its error is less than $2^{-(B_D+2)} + 2^{-(B_D+1)} < 2^{-B_D}$. Hence the rounded vector stays inside the original cube.

Since $2^{-\lceil \log_2 D \rceil} \leq D^{-1}$,

$$\Delta_D < D\, 2^{-B_D} \leq D\, D^{-(\mu+4)} = D^{-(\mu+3)},$$

which proves Eq. (51). The observable has norm one by Proposition 3.4. The two Lipschitz bounds in Corollary E.3 therefore prove Eqs. (52) and (53).

By Eqs. (43) and (50),

$$G_D = \Theta(D^{-3\lambda_\mu/2}) = \Theta(D^{-\mu/(2\ln 2)}).$$

The integral bound $\ln 2 = \int_1^2 x^{-1} dx > 1/2$ implies $\mu/(2\ln 2) < \mu < \mu + 3$. Thus

$$\Delta_D = o(G_D). \tag{137}$$

For all sufficiently large D, we may in particular assume $\Delta_D \leq \min\{1/2000, G_D/100\}$. The ideal gradient lies in $[G_D, 2G_D]$ by Eq. (133). Combining this fact with Eq. (53) proves Eq. (54).

The perturbation $\Delta_D = o(D^{1-r})$ and the reverse triangle inequality leave the ℓ_1 scaling in Eq. (130) unchanged, proving Eq. (55). Each coordinate changes by less than $D^{-(\mu+4)} = o(D^{-r})$. For all sufficiently large D this is at most $(\pi/14)D^{-r}$. Combining this fact with Eq. (131) shows that the same $D/2$ split-T coordinates have rounded distance at least $(\pi/14)D^{-r}$ and at most $(2\pi/7)D^{-r}$ from the nearest Clifford angle. This proves Eq. (56).

For a yes-instance, the ideal-point promise proved in Theorem 5.3, together with Eq. (52), gives

$$C_Q(\widehat{\boldsymbol{\theta}}^\star_Q) \geq \frac{1}{15} - \frac{1}{2000} > \frac{3}{50}.$$

For a no-instance,

$$C_Q(\widehat{\boldsymbol{\theta}}^\star_Q) \leq -\frac{1}{20} + \frac{1}{2000} < -\frac{1}{25}.$$

Thus the same polynomial-time map from QCA establishes the rational promise in Eq. (57). A cost estimate with error at most $1/100$ has positive sign in the yes case and negative sign in the no case, so cost estimation is BQP-hard.

Finally, in the yes case the ideal gradient is at least $5G_D/3$, and in the no case it is at most $4G_D/3$. Rounding changes it by at most $G_D/100$. If a further estimator has error at most $G_D/24$, then its output is at least

$$\left(\frac{5}{3} - \frac{1}{100} - \frac{1}{24}\right) G_D > \frac{3}{2} G_D$$

in the yes case, and at most

$$\left(\frac{4}{3} + \frac{1}{100} + \frac{1}{24}\right) G_D < \frac{3}{2} G_D$$

in the no case. Comparison with $3G_D/2$ decides QCA, proving the gradient-hardness statement. For a finite computation, a_D is rational by Eq. (41), and $1/\sqrt{2}$ can be approximated to any requested number of bits by bisection on $[0, 1]$ using rational square comparisons. Thus the reduction can compute a rational approximation to the threshold with error, for example, at most $G_D/1000$ in polynomial time. This additional error is far smaller than the separation above. □

F.12 Proof of Theorem 6.1

Proof. The circuit and observable are constructed in Section 3. Polynomial size follows from Observation 3.1. The PRC normal form follows from Observation C.2. $U_Q(0)$ is Clifford by Observation 3.2 and the construction of the handle and dummy registers. The qubit count follows from Lemma 3.3, and the observable properties follow from Proposition 3.4. This proves item (i).

Item (ii) is Theorem 4.1 and Proposition 4.2, together with $N_D = \Theta(D)$. Item (iii) is Theorem 4.3. Item (iv) follows from Theorem F.2 and Lemma 5.5. Item (v) follows from Theorem 5.3, Corollary 5.4, and Lemma 5.5. □