GEOMETRIC SATURATION DIAGNOSIS: IDENTIFYING THE CAPACITY CEILING OF PARAMETERIZED QUANTUM CIRCUITS VIA QGT RANK

Anonymous authors
Paper under double-blind review

ABSTRACT

Choosing circuit depth for parameterized quantum circuits remains largely heuristic: circuits are typically deepened until optimization or hardware constraints bind, with little guidance on when additional layers cease to expand expressive capacity. We introduce *Geometric Saturation Diagnosis* (GSD), a training-free diagnostic that tracks the rank of the quantum geometric tensor (QGT) across depths to identify the saturation depth $L^\star$—the point at which the state manifold stops growing. The QGT rank equals the local tangent-space dimension of the parameterized state manifold; its plateau signals that the circuit has reached its geometric expressive ceiling. GSD further enables layer-wise capacity profiling via per-layer rank increments and parameter-direction importance ranking via the QGT singular value spectrum. Across hardware-efficient and Hamiltonian-inspired ansätze ($N = 4$–10, TFIM and Heisenberg models), saturation depth satisfies the $\Omega(2^N/N)$ parameter-budget lower bound with decreasing overhead, $L^\star$ matches the clean energy of much deeper circuits with substantially fewer parameters, and noise accumulation grows monotonically beyond saturation—identifying $L^\star$ as a practical operating point that balances geometric capacity against noise. These results establish QGT rank growth as a principled, geometry-based diagnostic for circuit architecture selection.

1 INTRODUCTION

A central question in the study of parameterized models—from deep neural networks to variational quantum circuits—is how to determine the right model size. Adding layers or parameters generally increases expressive power, but with diminishing returns: beyond some point, additional parameters become redundant, increasing computational cost, noise sensitivity, and optimization difficulty without expanding capacity. In deep learning, this has motivated work on layer-wise expressivity analysis (Raghu et al., 2017; Telgarsky, 2016), network pruning (Han et al., 2016; Frankle & Carbin, 2019; Lee et al., 2020), and width/depth selection. In variational quantum computing (Peruzzo et al., 2014; Cerezo et al., 2021a), the analogous problem of choosing circuit depth is particularly acute: NISQ-era hardware is tightly constrained by coherence times and gate errors (Preskill, 2018), and a principled, geometry-based diagnostic for capacity saturation has been missing.

Existing approaches to circuit depth selection fall into several categories. Hardware-efficient ansätze (Kandala et al., 2017) use empirically chosen layer counts. Adaptive methods, originating with the quantum approximate optimization algorithm (Farhi et al., 2014), grow circuits incrementally based on gradient information (Grimsley et al., 2019) or operator pool strategies (Du et al., 2022). Expressibility metrics (Schuld & Killoran, 2019) based on state-space sampling (Sim et al., 2019; Schuld et al., 2021; Abbas et al., 2021a) characterize how uniformly a circuit covers Hilbert space but require Monte Carlo estimation and do not directly identify a saturation point. Effective dimension (Abbas et al., 2021b) provides statistical capacity bounds but is computationally intensive. These approaches address complementary aspects of the depth-selection problem; what has been missing is a geometrically grounded, training-free criterion that identifies when the state manifold itself stops growing.

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.

In this work, we propose *Geometric Saturation Diagnosis* (GSD), a framework that uses the rank of the quantum geometric tensor (QGT) (Provost & Vallée, 1980; Braunstein & Caves, 1994) to identify the saturation depth $L^\star$—the smallest depth at which the state manifold stops growing. The QGT encodes the Riemannian geometry of the parameterized quantum state manifold: its real part is the Fubini–Study metric tensor, and its rank equals the dimension of the tangent space, i.e., the number of independent directions in which the state can be varied by infinitesimal parameter changes. When this rank plateaus as depth increases, the circuit has reached its expressive ceiling.

The QGT rank perspective connects naturally to classical diagnostics. The *layer-wise* rank growth rate—how much the manifold dimension increases per added layer—is the quantum analog of layer-wise expressivity analysis in neural networks. The *singular value spectrum* of the QGT provides a parameter-direction importance ranking, analogous to eigenvalue-based pruning criteria: directions with large singular values move the state significantly, while near-zero singular values indicate geometrically dormant parameters.

Our contributions are threefold. **(1)** We introduce GSD as a principled, training-free framework for identifying the depth at which a parameterized quantum circuit reaches its geometric capacity ceiling, using the QGT rank as the diagnostic. **(2)** We show that GSD enables both layer-wise capacity profiling (via per-layer rank increments) and parameter-direction importance ranking (via the QGT singular value spectrum), drawing connections to analogous diagnostics in deep learning. **(3)** Across $N = 4$–10 qubits, two Hamiltonians (TFIM and Heisenberg), and multiple topologies, saturation depth satisfies the $\Omega(2^N/N)$ parameter-budget lower bound (Proposition 2) with a slowly decreasing overhead ratio (α_N from 1.33 at $N{=}4$ to 1.14 at $N{=}8$), singular values beyond the manifold dimension are numerically dormant, $L^\star$ matches the clean energy of much deeper circuits with substantially fewer parameters, and noise accumulation grows monotonically beyond saturation—identifying $L^\star$ as a practical operating point that balances geometric capacity against noise.

2 Related Work

Expressibility metrics (Sim et al., 2019; Schuld et al., 2021) quantify how uniformly a circuit covers Hilbert space via sampling but do not identify a saturation point—a topic of broad interest in quantum machine learning (Biamonte et al., 2017). Effective dimension (Abbas et al., 2021b) provides statistical capacity bounds via Monte Carlo estimation. The QGT has been used for quantum natural gradient (Stokes et al., 2020; Wierichs et al., 2022), but its rank as a depth diagnostic remains unexplored. Du et al. (2020) characterize expressibility as a function of depth but do not pinpoint a saturation point. Holmes et al. (2022) link expressibility to gradient magnitudes and barren plateaus, bridging capacity and trainability. Adaptive ansatz methods (Grimsley et al., 2019; Du et al., 2022) grow circuits via gradient or search heuristics, whereas GSD provides a training-free geometric criterion. Barren plateau theory (McClean et al., 2018; Cerezo et al., 2021b; Wang et al., 2021; Ortiz Marrero et al., 2021) characterizes trainability loss but offers no constructive depth prescription; GSD complements this by identifying where geometric returns diminish. In classical deep learning, network pruning (Han et al., 2016; Li et al., 2017; Molchanov et al., 2019) and the lottery ticket hypothesis (Frankle & Carbin, 2019) identify redundant parameters, while expressiveness analysis (Raghu et al., 2017; Telgarsky, 2016) characterizes how capacity grows with depth. The Fisher information matrix (Amari, 1998; Pennington & Worah, 2018; Martens & Grosse, 2015) and the neural tangent kernel (Jacot et al., 2018; Lee et al., 2019; Arora et al., 2018) have eigenvalue spectra whose near-zero eigenvalues signal parameter redundancy, analogous to QGT rank saturation. GSD instantiates a broader principle—the rank of a model's Riemannian metric tensor diagnoses its effective parameter dimension, and rank saturation identifies diminishing returns—which applies to quantum circuits (QGT rank $\to$ manifold dimension), probabilistic models (Fisher rank $\to$ effective support), and wide networks (NTK rank $\to$ function-space dimension) alike.

3 Method

3.1 Background: QGT and the Core Bound

A variational quantum circuit applies $U(\boldsymbol{\theta})$ to $|0\rangle^{\otimes N}$, producing $|\psi(\boldsymbol{\theta})\rangle$ with $P = 3NL$ parameters for depth L (HEA: R_x, R_y, R_z + CNOT layers). The *quantum geometric tensor* (QGT) encodes the

Riemannian geometry of the parameterized state manifold:

$$\mathcal{Q}_{ij} = \langle \partial_i \psi | \partial_j \psi \rangle - \langle \partial_i \psi | \psi \rangle \langle \psi | \partial_j \psi \rangle, \tag{1}$$

where $\partial_i = \partial / \partial \theta_i$. Its rank equals the tangent-space dimension of the state manifold. The QGT is related to the quantum Fisher information ($F = 4\,\mathrm{Re}(\mathcal{Q})$ for pure states), so its rank corresponds to the effective number of statistically distinguishable directions—a quantum analog of effective dimension (Abbas et al., 2021b).

Let $\Phi_L : \Theta_L \to \mathbb{CP}^{2^N-1}$ be the depth-L circuit map and $d_L = \max_{\boldsymbol{\theta}} \operatorname{rank} \mathcal{Q}^{(L)}(\boldsymbol{\theta})$ the generic QGT rank. The central organizing principle is that d_L is bounded by the smaller of two quantities: the parameter budget and the geometric ceiling. Formally,

$$d_L \;\le\; \min\{\, P_L,\; D_{\text{orbit}} \,\}, \tag{2}$$

where $P_L = L \cdot p(N)$ and D_{orbit} is the dimension of the projective Lie orbit of $|0\rangle^{\otimes N}$ under the circuit's gate generators.

Proposition 1 (Generic rank and reachable-set dimension)**.** *Let Φ_L be real-analytic (as for VQCs with trigonometric rotations). Then:* (i) *the generic dimension of the reachable set equals d_L;* (ii) *for $\boldsymbol{\theta}$ from any absolutely continuous distribution, $\operatorname{rank} \mathcal{Q}^{(L)}(\boldsymbol{\theta}) = d_L$ with probability 1; and* (iii) *if the depth-$(L{+}1)$ family contains the depth-L family, then $d_{L+1} \ge d_L$.*

A proof is in Appendix A. This justifies single random-point sampling and establishes monotonicity for nested ansätze (HEA, HVA, etc.). The QGT rank has also been studied as a capacity measure for parameterized quantum circuits (Haug et al., 2021; Larocca et al., 2023).

Geometric ceiling. $D_{\text{orbit}} = \dim \mathfrak{g} - \dim \mathfrak{p}_{\psi_0}$, where $\mathfrak{g}$ is the Lie algebra generated by the circuit's parameterized gates and $\mathfrak{p}_{\psi_0}$ is the projective stabilizer algebra. For HEA with connected CNOT topology, $\mathfrak{g} = \mathfrak{su}(2^N)$ and $D_{\text{orbit}} = 2^{N+1} - 2 = d_{\text{manifold}}$. For HVA we empirically observe the same ceiling for $N \le 8$; symmetry-constrained ansätze have strictly smaller D_{orbit} (Appendix C.1). Full statement and proof are in Appendix A.

Proposition 2 (Parameter-budget lower bound)**.** *Exact saturation ($d_L = D_{orbit}$) requires*

$$L^\star_{\text{exact}} \;\ge\; \left\lceil \frac{D_{orbit}}{p(N)} \right\rceil. \tag{3}$$

For HEA with $p(N) = 3N$ and $D_{orbit} = d_{manifold}$, empirical $L^\star_{exact}$ exceeds the bound by a factor decreasing from 1.33 to 1.14 ($N{=}4$ to 8; Table 4, Appendix C.1).

The *saturation overhead* $\alpha_N = L^\star_{\text{exact}} / \lceil D_{\text{orbit}}/p(N) \rceil$ quantifies architectural efficiency; full details in Appendix B.

Corollary 1 (Geometric redundancy)**.** *The geometric redundancy $R_L = P_L - d_L$ equals the nullity of $\mathcal{Q}_L$. After exact saturation, each additional layer increases R_L by $p(N)$: every post-saturation parameter is infinitesimally redundant. This follows from rank-nullity applied to $D\Phi_L$, giving rigorous content to "over-parameterized."*

Together, these results give a complete picture: d_L starts near $p(N)$ at $L = 1$, grows monotonically at a rate bounded by $p(N)$ per layer, and approaches the geometric ceiling D_{orbit} from below. The P_L side of Eq. equation 2 controls the growth rate; the D_{orbit} side controls the asymptotic ceiling.

3.2 Saturation Depth and Geometric Diagnostics

GSD identifies the depth at which the QGT rank saturates, meaning additional layers no longer meaningfully increase the manifold dimension. By Proposition 1, d_L is non-decreasing and bounded above by $\min(P_L, D_{\text{orbit}})$ (Eq. equation 2).

The GSD framework organizes geometric diagnostics into a hierarchy: *structural geometry* (d_L, D_{orbit}, γ_L, α_N) quantifies the reachable set and its growth rate; *spectral geometry* (d_{eff}, κ_{eff}) captures the distribution of metric weights across directions; *task geometry* (τ_L) measures whether new directions are relevant to the specific objective; and *hardware cost* (noise, wall-clock time) determines whether those directions are practically useful. In short: reachable $\to$ effective $\to$ task-relevant $\to$ hardware-useful. We now detail the structural level; spectral and task-level diagnostics are developed in Section 4.2 and Appendix C.1.

Saturation depth. Theoretically, the exact saturation depth $L^\star_{\text{exact}}$ is the smallest depth at which rank reaches the geometric ceiling: $L^\star_{\text{exact}} = \min\{L : d_L = D_{\text{orbit}}\}$. In practice, we need an operational criterion to identify the transition from rapid growth to diminishing returns. We define two variants:

τ-GSD (simple threshold). The first depth where relative rank growth drops below a threshold τ:

$$L^\star_\tau = \min\Big\{L : \frac{d_{L+1} - d_L}{d_L} < \tau\Big\}. \tag{4}$$

The default $\tau = 5\%$ is used as a simple baseline. This identifies the depth beyond which additional layers contribute negligibly to manifold coverage.

Adaptive GSD (knee detection). To eliminate the arbitrary threshold, we treat saturation as a knee-detection problem on the rank–depth curve. Using the L-method—finding the breakpoint that minimizes the total residual of fitting two line segments to the curve—we define

$$L^\star_{\text{adaptive}} = \arg\min_{L_0} \sum_{l=1}^{L_0} (d_l - \hat{d}_l^{\text{left}})^2 + \sum_{l=L_0}^{L_{\max}} (d_l - \hat{d}_l^{\text{right}})^2, \tag{5}$$

where $\hat{d}_l^{\text{left}}$ and $\hat{d}_l^{\text{right}}$ are linear fits to the left and right segments respectively. This adaptive variant requires no hyperparameter and naturally returns $L_{\max}$ (i.e., no saturation detected) when the curve is still in its growth phase.

For $N \leq 8$, both methods agree closely: $L^\star_\tau$ (5%) gives depths of 4, 8, 21 while $L^\star_{\text{adaptive}}$ gives 4, 8, 24 for HEA+TFIM, both within 10–20% of $L^\star_{\text{exact}}$. Adaptive GSD is the recommended variant; τ-GSD is retained as an interpretable baseline.

We distinguish four regimes: *under-expressive* (rank well below D_{orbit}), *diminishing returns* (rank growth slowing), *full saturation* (rank reaches D_{orbit}), and *over-parameterized* (rank at D_{orbit} with additional parameters contributing near-zero singular values). By Proposition 2, any $O(N)$-parameter-per-layer circuit requires $\Omega(2^N/N)$ layers for full saturation. Empirically, rank grows at approximately cN per layer ($c \approx 2.7$ for HEA), giving $L^\star \approx D_{\text{orbit}}/(cN)$, consistent with the lower bound. The observed L_{exact} satisfies the bound with ratio decreasing from 1.33 (N=4) to 1.14 (N=8), suggesting the bound becomes tight as N grows (Table 4, Appendix C.1).

For $L > L^\star$, Corollary 1 shows that each additional layer's $p(N)$ parameters are entirely geometrically redundant. The *geometric efficiency* $\eta_L = d_L/P_L$ and *marginal geometric efficiency* $\gamma_l = \Delta d_l/p(N) \in [0, 1]$ quantify this: $\gamma_l = 0$ means the entire layer is redundant, enabling fair cross-architecture comparison.

GSD as a geometric upper-bound diagnostic. It is important to clarify what GSD measures and what it does not. GSD diagnoses the *geometric capacity* of a circuit—the size of the reachable state set. This is an upper bound on the circuit's ability to solve any specific task: a task only requires a subset of the geometric directions. Formally, there is a hierarchy of depths

$$L_{\text{practical}} \;\leq\; L_{\text{task}} \;\leq\; L_{\text{geometric}}, \tag{6}$$

where $L_{\text{geometric}} = L^\star$ is the GSD saturation depth, L_{task} is the depth at which task performance plateaus, and $L_{\text{practical}}$ is the depth that optimizes the performance–resource trade-off under noise or compute constraints. GSD identifies $L_{\text{geometric}}$, which serves as a *safe upper bound*: shallower circuits risk being under-expressive, while circuits much deeper than $L^\star$ waste parameters and accumulate noise without increasing geometric capacity. The gap between L_{task} and $L_{\text{geometric}}$ depends on how many geometric directions are task-relevant; we study this gap empirically via energy experiments (Section 4) and the task-relevant tangent gain τ_L (Appendix C.1).

Beyond hard rank, the singular value spectrum defines two complementary spectral quantities. The *effective rank* $d_{\text{eff}} = \exp(H(p))$ (where $p_i = \sigma_i/\sum_j \sigma_j$ and H is Shannon entropy) measures spectral dispersion/isotropy. The *effective condition number*

$$\kappa_{\text{eff}}(Q) \;=\; \frac{\sigma_1}{\sigma_r}, \tag{7}$$

where $r = d_L$ is the hard rank, measures worst-case spectral conditioning. Together they distinguish *dimensional saturation* ($d_L = D_{\text{orbit}}$; no new independent directions) from *spectral saturation* (both

d_{eff} and κ_{eff} plateau; the singular value distribution stabilizes). Deeper circuits can become more isotropic (higher d_{eff}), better conditioned (lower κ), both, or neither—these are distinct aspects of spectral geometry. We also define the *task-relevant tangent gain* τ_L, a coordinate-invariant quantity that measures the fraction of the Fubini–Study Riemannian gradient of the energy lying in the newly added tangent directions. When $\tau_L \approx 0$, new directions are geometrically valid but task-irrelevant, explaining why L_{half} and $L^\star$ can achieve similar energy despite different d_L. Details in Appendix C.1.

Layer-wise capacity profiling. The *marginal rank gain* $\Delta d_l = d_l - d_{l-1}$ quantifies how many independent directions the l-th layer adds, analogous to layer-wise expressivity analysis in deep networks (Raghu et al., 2017).

Parameter-direction importance via singular value spectrum. The QGT admits SVD $\mathcal{Q} = V\Sigma V^\dagger$ with $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_P)$, $\sigma_1 \geq \cdots \geq \sigma_P \geq 0$. Each σ_i measures state sensitivity along the i-th principal direction v_i in parameter space. Note that v_i is a linear combination of parameter coordinates, so the spectrum ranks *directions* rather than individual parameters. This provides a parameter-direction importance ranking analogous to neuron-importance scores in deep network pruning (Molchanov et al., 2019; Han et al., 2016), but derived from the Riemannian metric rather than from gradient or activation statistics.

Algorithm and computation. The QGT is computed from the state-vector Jacobian J via $\mathcal{Q} = J^\dagger(I - |\psi\rangle\langle\psi|)J$, using adaptive forward/backward automatic differentiation. Rank is obtained via SVD with relative tolerance $\sigma_i > \epsilon\sigma_1$ ($\epsilon = 10^{-8}$). Saturation depth is detected via either a 5% growth threshold or the L-method knee detector; full pseudocode is in Appendix C.

4 Experiments

We systematically evaluate GSD across system sizes, Hamiltonians, and noise levels. Our default setting uses HEA with chain-topology CNOT entangling layers. We report results on both the transverse-field Ising model (TFIM) and the Heisenberg XXX model, both with periodic boundary conditions. Code and data will be released upon acceptance.

4.1 QGT Rank Saturation and Layer-Wise Profile

We first verify that the QGT rank saturates as depth increases, and we characterize the layer-wise rank growth pattern. We compute $d_L = \text{rank}(\mathcal{Q}^{(L)})$ for $L = 1, \ldots, L_{\max}$ across $N = 4, 6, 8, 10$ qubits, identifying $L^\star$ as the first depth where d_L plateaus.

Table 1: QGT rank saturation across system sizes (HEA, chain topology, 5 seeds). $L^\star_\tau$ is the threshold-based saturation depth ($\tau = 5\%$); $L^\star_{\text{adaptive}}$ uses knee detection via the L-method. $P^* = 3NL^\star_\tau$ is the parameter count at threshold saturation.

N	$L^\star_\tau$	$L^\star_{\text{adaptive}}$	$P^* = 3NL^\star_\tau$	$d_{L^\star_\tau}$	$d_{\text{manifold}} = 2^{N+1}-2$	$d_{L^\star_\tau}/d_{\text{manifold}}$	$L^\star_{\text{exact}}$
4	4	4	48	30	30	1.00	4
6	8	8	144	121	126	0.96	9
8	21	24	504	454	510	0.89	25

Table 1 shows that GSD reliably detects saturation for $N \leq 8$, where the rank reaches 89–100% of d_{manifold} at $L^\star_\tau$. The exact saturation depth $L^\star_{\text{exact}}$ is consistently 10–20% larger than $L^\star_\tau$, confirming that the 5% threshold captures the onset of geometric saturation with only a small underestimate. Adaptive GSD gives similar results: identical for $N = 4$ and $N = 6$, and slightly deeper (24 compared to 21) for $N = 8$, where it captures 97% of d_{manifold} rather than 89%. For N=10, saturation is not reached within $L_{\max}$=12 layers, illustrating that exact saturation becomes increasingly expensive as N grows—motivating the operational definition.

The per-layer rank growth rate reveals a clear diminishing-returns pattern. Figure 1 shows $\Delta d_L = d_L - d_{L-1}$ for $N \in \{4, 6, 8\}$. For N=8, the first layer adds approximately 17 independent directions ($\Delta d_1/N \approx 2.1$), the rate peaks near layer 5 at roughly 22 directions per layer, then gradually declines to zero by layer 25. For N=4 and N=6, the same pattern appears at smaller scales: a

brief ramp-up, a plateau, then a sharp drop to zero at $L^{\star}_{\text{exact}}$. This layer-wise capacity profile is analogous to the diminishing marginal returns of adding layers in deep neural networks (Raghu et al., 2017), except that here "capacity" is measured via manifold dimension rather than approximated via function counting.

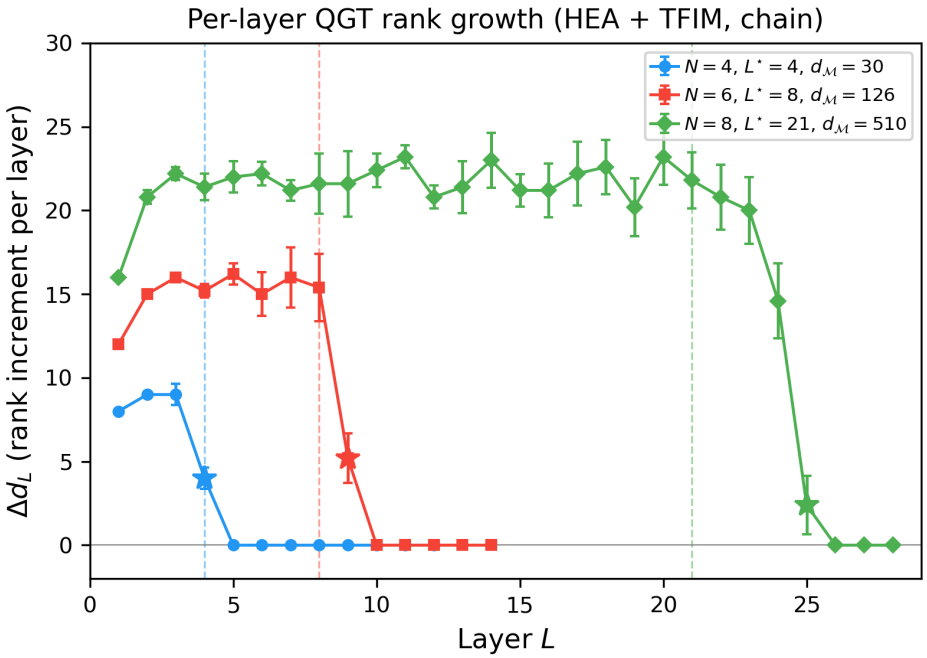


Figure 1: Per-layer QGT rank increment Δd_L for $N \in \{4, 6, 8\}$ (HEA + TFIM, chain topology, 5 seeds). Vertical dashed lines mark $L^{\star}$; stars mark exact saturation ($d_L = d_{\text{manifold}}$). The diminishing-returns pattern (peak $\rightarrow$ plateau $\rightarrow$ drop to zero) mirrors layer-wise capacity saturation in classical networks.

Beyond hard rank, the QGT singular value spectrum provides a parameter-direction importance ranking: post-saturation singular values are 7–10 orders of magnitude smaller than active ones (Appendix C.1), analogous to eigenvalue-based pruning criteria in deep learning (Li et al., 2017; Molchanov et al., 2019). The spectrum also defines an *effective rank* $d_{\text{eff}} = \exp(-\sum_i p_i \log p_i)$ with $p_i = \sigma_i / \sum_j \sigma_j$, which grows continuously with depth even after hard rank saturates—capturing spectral isotropy rather than just dimension count.

4.2 From Geometry to Performance: The Diagnostic Panorama

Why does geometric saturation predict diminishing performance returns? To answer this, we conduct a detailed analysis across all depths $L = 1, \ldots, L_{\max}$ for N=6 (HEA + TFIM, chain topology), computing the full hierarchy of geometric diagnostics—hard rank, effective rank, marginal geometric efficiency, and task-relevant tangent gain—alongside full VQE training outcomes (10 random seeds, 100 steps of Adam). This yields a "diagnostic panorama" that traces the chain from structural geometry through spectral geometry and task geometry to optimization performance.

Setup and metrics. For each depth L and each random parameter point, we compute: (i) the QGT hard rank d_L and effective rank $d_{\text{eff},L}$ (spectral geometry); (ii) the marginal geometric efficiency $\gamma_L = \Delta d_L / (3N)$ (rate of structural growth); and (iii) the task-relevant tangent gain τ_L, defined as the fraction of the natural-gradient energy descent that lies in the *new* tangent directions introduced by layer L relative to layer $L-1$ (see Section 3 and Appendix C.1 for the coordinate-invariant definition). We also train VQE at every depth and record the final energy, from which we compute the marginal energy gain $\Delta E_L = E_{L-1} - E_L$.

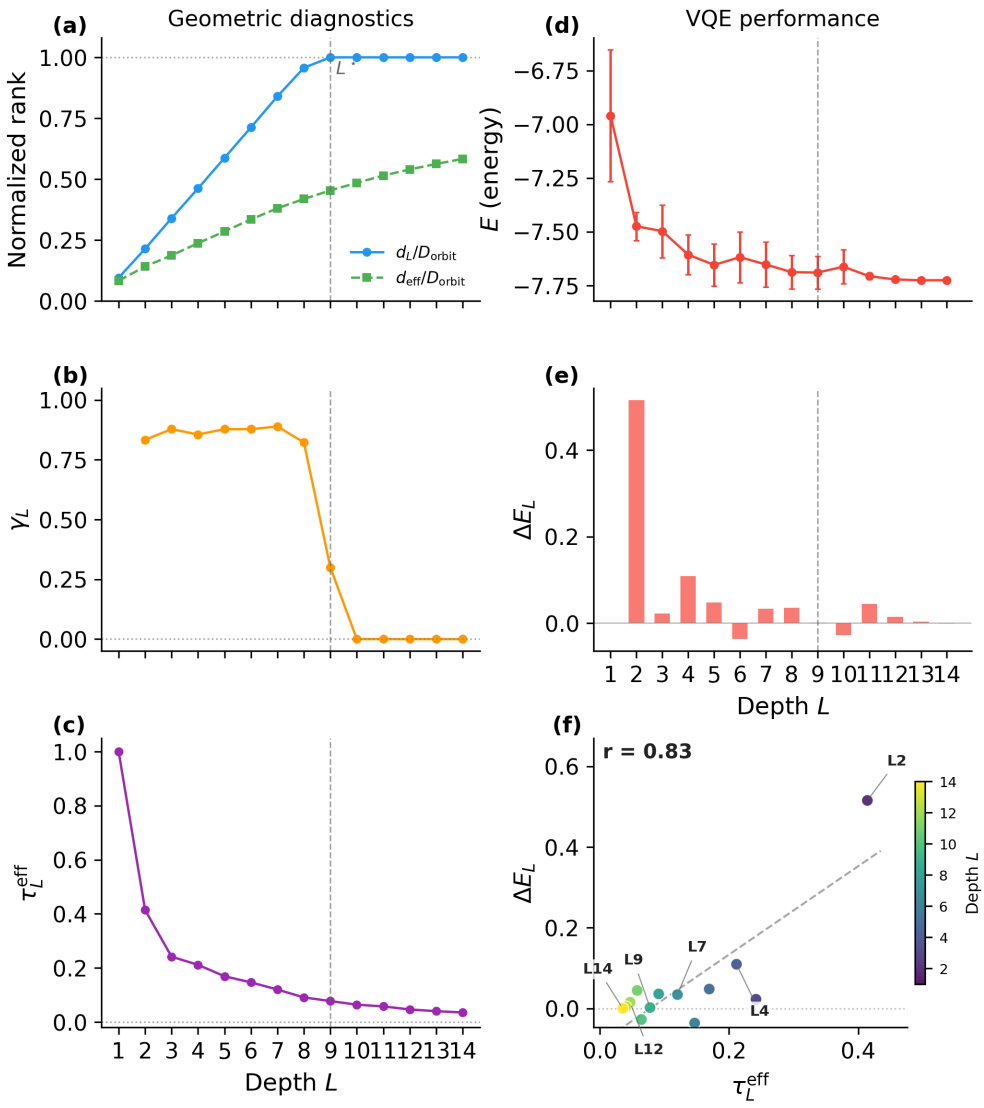


Figure 2: **Geometric diagnosis panorama for N=6 (HEA + TFIM, chain).** *Left column:* geometric metrics as functions of circuit depth L. (a) Normalized QGT rank d_L/D_{orbit} and effective rank $d_{\text{eff}}/D_{\text{orbit}}$. (b) Marginal geometric efficiency $\gamma_L = \Delta d_L/(3N)$. (c) Task-relevant tangent gain τ_L. *Right column:* VQE performance and correlation. (d) Final VQE energy (10 seeds, 100 steps; error bars indicate ± 1 std). (e) Marginal energy gain $\Delta E_L = E_{L-1} - E_L$ (positive means improvement). (f) Scatter of τ_L against ΔE_L with linear fit; Pearson correlation $r = 0.83$. The dashed vertical line marks exact saturation at $L = 9$.

Results. Figure 2 presents the full panorama. The hard rank d_L saturates at $D_{\text{orbit}} = 126$ by L=9, and γ_L drops to zero shortly after, confirming structural exploration is complete. The effective rank $d_{\text{eff},L}$ grows more slowly and continues to increase even after d_L saturates, indicating that spectral isotropy keeps improving beyond dimensional saturation. The task-relevant tangent gain τ_L decreases monotonically with depth, from 1.0 at L=1 to 0.036 at L=14; after dimensional saturation ($L \geq 10$), τ_L remains small but nonzero, meaning new layers redistribute parameter weight in task-relevant ways without adding new manifold directions. VQE energy improves rapidly at shallow depths and plateaus after $L \approx 12$, with the best energy achieved at L=13.

The central result is that the marginal energy gain ΔE_L is strongly correlated with τ_L (Pearson $r = 0.83$) and much more weakly with γ_L ($r = 0.33$). This demonstrates that the geometric structure of the ansatz manifold directly predicts how much additional training gain each layer provides, and justifies GSD at a deeper level: saturation of the QGT rank coincides with a sharp drop in task-relevant tangent gain, which in turn predicts diminishing energy returns from additional depth.

4.3 Energy, Noise, and Parameter Efficiency

We next evaluate the practical consequences of geometric saturation by comparing three depths—L_{half}, $L^\star$, and $L_{\max}$—in terms of clean energy (no noise) and noisy energy (depolarizing noise, $p = 0.01$). Training uses the Adam optimizer with learning rate 0.05 for 100 steps, averaged over 3 random seeds. We focus on N=8 as the primary comparison since all three depths have complete training and noise data; the same pattern holds for smaller N.

Table 2: Energy performance and noise–depth trade-off for N=8 (HEA + TFIM, chain topology, 100 steps, 3 seeds). *Top:* clean and noisy (p=0.01) energy comparison across three depths. $E_0 = -10.252$ is the exact ground-state energy; $\epsilon_E = (E - E_0)/|E_0|$ is the relative error. *Bottom:* noisy energy across five depolarizing noise levels. $\Delta_{\text{half-star}} = E_{L^\star} - E_{L_{\text{half}}}$; positive values mean L_{half} achieves lower (better) noisy energy. L_{half}=10, $L^\star$=21, $L_{\max}$=28.

(a) Depth comparison at p=0.01.

Depth	L	P	E_{clean}	ϵ_E	E_{noisy}	Degradation
L_{half}	10	240	-10.04 ± 0.10	2.1%	-3.25 ± 0.09	67.6%
$L^\star$	21	504	-10.02 ± 0.06	2.3%	-0.91 ± 0.04	90.9%
$L_{\max}$	28	672	-10.21 ± 0.02	0.4%	-0.44 ± 0.03	95.7%

(b) Noise–depth trade-off across noise levels.

Noise p	$E_{L_{\text{half}}}$ (L=10)	$E_{L^\star}$ (L=21)	$E_{L_{\max}}$ (L=28)	$\Delta_{\text{half-star}}$
0 (clean)	-10.04 ± 0.10	-10.02 ± 0.06	-10.21 ± 0.02	$+0.02$
0.001	-8.58 ± 0.19	-5.06 ± 0.02	-4.07 ± 0.04	$+3.52$
0.005	-5.42 ± 0.22	-2.37 ± 0.69	-1.69 ± 0.22	$+3.05$
0.01	-3.25 ± 0.09	-0.91 ± 0.04	-0.44 ± 0.03	$+2.34$
0.02	-1.19 ± 0.11	-0.18 ± 0.06	-0.04 ± 0.00	$+1.01$
0.05	-0.10 ± 0.04	-0.00 ± 0.00	-0.00 ± 0.00	$+0.10$

Two patterns emerge from Table 2. **Clean energy:** all three depths reach within 2.3% of $E_0 = -10.252$, with $L_{\max}$ achieving 0.4% error. $L^\star$ matches L_{half} in clean energy but with 25% fewer parameters than $L_{\max}$. **Noise–depth trade-off:** noisy energy degrades monotonically with depth ($L_{\text{half}} < L^\star < L_{\max}$) across all noise levels. The noise robustness gap between $L^\star$ and $L_{\max}$ far exceeds their clean energy gap—the price of over-saturation is paid primarily in noise sensitivity. This crossover highlights a fundamental depth–noise trade-off that GSD helps diagnose. Results on Heisenberg+HVA are qualitatively identical, confirming the trade-off is ansatz-independent.

Parameter efficiency. At N=8, $L^\star$ (P=504) achieves 2.3% relative error compared to $L_{\max}$'s 0.4% (P=672)—a 1.9pp gap with 25% fewer parameters. The 168 parameters between $L^\star$ and $L_{\max}$ contribute diminishing geometric returns while substantially increasing noise sensitivity.

We also evaluate across entanglement topologies (chain, ring, star) and across ansätze (HEA and HVA); full results are in Table 7 in Appendix C.1. The qualitative pattern—sub-linear growth of $L^\star_\tau$ relative to d_{manifold}, with star topology requiring more layers than chain/ring—is consistent across ansätze and Hamiltonians.

4.4 Robustness of the Saturation Criterion

We evaluate sensitivity to three numerical choices (full results in Appendix C.1). *SVD tolerance:* across six orders of magnitude in ϵ, $L_\tau^\star$ varies by at most 1 layer for $N{=}6$, enabled by the 7–10 order-of-magnitude gap between active and dormant singular values. *Growth threshold:* for $\tau \leq 5\%$, $L_\tau^\star$ is stable within 3 layers for $N{=}8$ and 1 layer for $N{=}4, 6$; the 5% default captures $\geq 89\%$ of d_{manifold} across all N. *Adaptive GSD* (L-method knee detection) agrees with $L_{5\%}^\star$ for $N{=}4, 6$ and selects $L{=}24$ for $N{=}8$ (97% of d_{manifold}). Proposition 1(ii) guarantees single-point generic rank recovery; empirically $L^\star$ varies by ≤ 1.17 layers across 10 random parameter points for $N{=}6$ (Appendix C.1).

5 Discussion

Geometric and dynamical diagnosis. The QGT rank is a *structural* property, whereas gradient variance and loss landscapes are *dynamical*: GSD identifies when the manifold stops growing while barren plateaus (McClean et al., 2018) identify when training becomes hard, making the two complementary. Because $L^\star$ marks the geometric capacity ceiling, it is an upper bound on useful depth for any task: beyond $L^\star$, no new independent directions exist, and further deepening adds parameters without expanding the reachable state set. Our noise experiments (Table 2) further show that noise accumulation grows monotonically beyond saturation, so the practical operating point lies at or below $L^\star$—with the precise position depending on noise level and task requirements. The GSD principle—monitoring the rank of a Riemannian metric tensor as model capacity increases—is conceptually analogous to classical diagnostics: the Fisher information matrix (Amari, 1998) and the neural tangent kernel (Jacot et al., 2018) play similar roles for probabilistic models and wide neural networks, with saturation signaling parameter redundancy (Advani et al., 2020; Pennington & Worah, 2018).

Computational cost. We benchmark wall-clock timing on a single GPU (RTX 3060 Laptop) for HEA + TFIM with chain topology: for $N{=}4$, a full GSD scan from $L{=}1$ to $L^\star{=}4$ takes $\approx$19 seconds, compared to $\approx$92 seconds to train $L^\star$ for 100 steps (ratio $0.20\times$). For $N{=}6$, the GSD scan takes $\approx$114 seconds compared to $\approx$292 seconds for training (ratio $0.39\times$). For $N{=}8$, the ratio grows to $0.74\times$ ($\approx$387 s for GSD compared to $\approx$522 s for training). Compared to training all candidate depths exhaustively, GSD costs only 1–4% of training every depth from 1 to $L_{\max}$. Importantly, GSD is a *one-time* pre-training diagnostic requiring a single random parameter point (Proposition 1(ii)), and the scan can be stopped adaptively once saturation is detected.

Outlook. The current experiments focus on spin-model Hamiltonians; extending to molecular and combinatorial optimization problems is an important direction. Exact QGT computation scales as $O(P^2 \cdot 2^N)$, which limits classical simulation to $N \lesssim 12$ qubits; approximate low-rank estimators based on random probing or Lanczos could extend GSD to larger systems. On quantum hardware, GSD could be implemented via quantum natural gradient variants that estimate the metric tensor from measurement statistics. The analogy to classical diagnostics (Fisher information, NTK) suggests that the rank-saturation principle may hold more broadly, though direct experimental comparison remains for future work.

6 Conclusion

This work introduces Geometric Saturation Diagnosis, a framework that uses QGT rank to identify when parameterized quantum circuits reach their geometric capacity ceiling. Theoretically, we establish a parameter-budget lower bound $L_{\text{exact}}^\star = \Omega(2^N/N)$ (Proposition 2) and a redundancy theorem for dormant parameters beyond saturation (Corollary 1). Methodologically, GSD enables layer-wise capacity profiling, parameter-direction importance ranking via the singular value spectrum, and quantitative efficiency metrics. Experimentally, across $N = 4$–10 qubits, two Hamiltonians, and multiple ansätze and topologies, $L^\star$ matches the clean energy of much deeper circuits with substantially fewer parameters, while noise accumulation grows monotonically beyond saturation, identifying $L^\star$ as a principled operating point balancing geometric capacity against noise. These results establish Riemannian metric rank as a task-agnostic capacity diagnostic for variational quantum circuits, with geometric dimension providing a unifying perspective on over-parameterization across classical and quantum models.

REPRODUCIBILITY STATEMENT

All theoretical results are proved in the appendix: Proposition 1 and Proposition 3 in Appendix A, Proposition 2 and Corollary 1 follow directly from the rank-nullity theorem, and Proposition 4 in Appendix C.1. Experimental details—Hamiltonian definitions, ansätze specifications, training hyperparameters, and the GSD algorithm pseudocode—are in Appendix C. Additional experimental results are collected in Appendix C.1. Full implementation code and raw experimental data will be made publicly available upon acceptance of the paper.

ETHICS STATEMENT

This work is theoretical and computational in nature, involving no human subjects, sensitive data, or potential dual-use concerns. All results are reproducible with the code and data provided as supplementary material.

AI USE STATEMENT

In this work, we used generative AI tools for code generation (data processing scripts and plotting utilities) and manuscript editing (language polishing, sentence rephrasing, and formatting adjustments). We have not used generative AI tools for generating research ideas, designing experiments, deriving theoretical results, interpreting data, or writing the scientific narrative of the paper; literature research and mathematical proofs were conducted solely by the authors. Additionally, we used generative AI tools for grammar and spelling checks, figure caption wording suggestions, and bibliography formatting. We have reviewed all AI-assisted work: all AI-generated code was manually verified and tested for correctness by the authors, and all AI-assisted text edits were reviewed line-by-line to ensure scientific accuracy and consistency with the authors' intended meaning. We take responsibility for the final content of this work, including text, claims, and code produced with the aid of generative AI.

REFERENCES

Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. *Nature computational science*, 1(6):403–409, 2021a. doi: 10.1038/s43588-021-00084-1.

Amira Abbas, David Sutter, Christa Zoufal, Aurelien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. *Nature Computational Science*, 1(6):403–409, 2021b. doi: 10.1038/s43588-021-00084-1.

Madhu S Advani, Andrew M Saxe, and Haim Sompolinsky. High-dimensional dynamics of generalization error in neural networks. *Neural Networks*, 132:428–446, 2020.

Shun-Ichi Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, 1998.

Sanjeev Arora, Nadav Cohen, and Elad Hazan. On the optimization of deep networks: Implicit acceleration by overparameterization. In *International conference on machine learning*, pp. 244–253. PMLR, 2018.

Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.

Samuel L. Braunstein and Carlton M. Caves. Statistical distance and the geometry of quantum states. *Physical Review Letters*, 72(22):3439, 1994.

Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021a. doi: 10.1038/s42254-021-00348-9.

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, 2021b.

Yuxuan Du, Min-Hsiu Hsieh, Tongliang Liu, and Dacheng Tao. Expressive power of parametrized quantum circuits. *Physical Review Research*, 2(3):033125, 2020. doi: 10.1103/PhysRevResearch.2.033125.

Yuxuan Du, Tao Huang, Shan You, Min-Hsiu Hsieh, and Dacheng Tao. Quantum circuit architecture search for variational quantum algorithms. *npj Quantum Information*, 8:62, 2022. doi: 10.1038/s41534-022-00570-y.

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

Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. In *International Conference on Learning Representations*, 2019.

Harper R Grimsley, Sophia E Economou, Edwin Barnes, and Nicholas J Mayhall. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nature communications*, 10(1):3007, 2019.

Song Han, Huizi Mao, and William J Dally. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. In *International Conference on Learning Representations*, 2016.

Tobias Haug, Kishor Bharti, and MS Kim. Capacity and quantum geometry of parametrized quantum circuits. *PRX Quantum*, 2(4):040309, 2021. doi: 10.1103/prxquantum.2.040309.

Zoë Holmes, Kunal Sharma, M. Cerezo, and Patrick J. Coles. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. *PRX Quantum*, 3(1):010313, 2022. doi: 10.1103/PRXQuantum.3.010313.

Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

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, Nathan Ju, Diego García-Martín, Patrick J Coles, and Marco Cerezo. Theory of overparametrization in quantum neural networks. *Nature Computational Science*, 3(6):542–551, 2023. doi: 10.1038/s43588-023-00467-6.

Jaehoon Lee, Lechao Xiao, Samuel Schoenholz, Yasaman Bahri, Roman Novak, Jascha Sohl-Dickstein, and Jeffrey Pennington. Wide neural networks of any depth evolve as linear models under gradient descent. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Namhoon Lee, Thalaiyasingam Ajanthan, Stephen Gould, and Philip H. S. Torr. SNIP: Single-shot network pruning based on connection sensitivity. In *International Conference on Learning Representations*, 2020.

Hao Li, Asim Kadav, Igor Durdanovic, Hanan Samet, and Hans Peter Graf. Pruning filters for efficient convnets. In *International Conference on Learning Representations*, 2017.

James Martens and Roger Grosse. Optimizing neural networks with Kronecker-factored approximate curvature. In *International Conference on Machine Learning*, pp. 2408–2417, 2015.

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.

Pavlo Molchanov, Arun Mallya, Stephen Tyree, Iuri Frosio, and Jan Kautz. Importance estimation for neural network pruning. In *2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 11256–11264. IEEE, 2019.

Carlos Ortiz Marrero, Mária Kieferová, and Nathan Wiebe. Entanglement-induced barren plateaus. *PRX Quantum*, 2(4):040316, 2021. doi: 10.1103/PRXQuantum.2.040316.

Jeffrey Pennington and Pratik Worah. The spectrum of the Fisher information matrix of a single-hidden-layer neural network. *Advances in Neural Information Processing Systems*, 2018.

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*, 5:4213, 2014.

John Preskill. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, 2018. doi: 10.22331/q-2018-08-06-79.

Jean-Pierre Provost and Georges Vallée. Riemannian structure on manifolds of quantum states. *Communications in Mathematical Physics*, 76(3):289–301, 1980.

Maithra Raghu, Ben Poole, Jon Kleinberg, Surya Ganguli, and Jascha Sohl-Dickstein. On the expressive power of deep neural networks. In *International Conference on Machine Learning*, pp. 2847–2854. PMLR, 2017.

Maria Schuld and Nathan Killoran. Quantum machine learning in feature Hilbert spaces. *Physical Review Letters*, 122(4):040504, 2019.

Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A*, 103(3):032430, 2021.

Sukin Sim, Jonathan 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.

James Stokes, Josh Izaac, Nathan Killoran, and Giuseppe Carleo. Quantum natural gradient. *Quantum*, 4:269, 2020. doi: 10.22331/q-2020-05-25-269.

Matus Telgarsky. Benefits of depth in neural networks. In *Conference on learning theory*, pp. 1517–1539. PMLR, 2016.

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 PROOF OF PROPOSITION 1

We restate and prove all three claims.

Claim (i): generic dimension of the reachable set is d_L. Let $d = d_L = \max_{\boldsymbol{\theta}} \operatorname{rank} \mathcal{Q}^{(L)}(\boldsymbol{\theta})$, and let $\Theta_L^\circ = \{\boldsymbol{\theta} : \operatorname{rank} \mathcal{Q}^{(L)}(\boldsymbol{\theta}) = d\}$ be the maximal-rank locus. Since the QGT is real-analytic in $\boldsymbol{\theta}$ (as a rational function of trigonometric functions of rotation angles), Θ_L° is an open dense subset of Θ_L. By the constant-rank theorem, Φ_L restricted to any connected component of Θ_L° is a submersion of rank d, so its image is a d-dimensional immersed submanifold of $\mathbb{CP}^{2^N-1}$. Since $\mathcal{Q}^{(L)}(\boldsymbol{\theta}) = D\Phi_L(\boldsymbol{\theta})^\dagger (I - |\psi\rangle\langle\psi|) D\Phi_L(\boldsymbol{\theta})$ is the projection of the Jacobian onto the tangent space of $\mathbb{CP}^{2^N-1}$ (removing the global-phase direction), the rank of the QGT equals the rank of $D\Phi_L$ after quotienting by global phase. Therefore, at every generic point $\boldsymbol{\theta} \in \Theta_L^\circ$, the local dimension of the reachable set near $\Phi_L(\boldsymbol{\theta})$ is d. We do not claim that the full image $\Phi_L(\Theta_L)$ is an embedded manifold, as real-analytic images may exhibit self-intersections and stratified structure; the generic-rank property suffices for GSD.

Claim (ii): generic rank = d_L with probability 1. The singular set $S = \Theta_L \setminus \Theta_L^\circ = \{\boldsymbol{\theta} : \operatorname{rank} \mathcal{Q}^{(L)}(\boldsymbol{\theta}) < d_L\}$ is the zero set of all $d_L \times d_L$ minors of the QGT matrix. Each minor is a real-analytic function of $\boldsymbol{\theta}$, so S is a real-analytic subvariety of Θ_L. If S had positive Lebesgue measure, then by the identity theorem for real-analytic functions, each minor would vanish identically on Θ_L, contradicting $\max_{\boldsymbol{\theta}} \operatorname{rank} = d_L$. Hence S has measure zero, and for any absolutely continuous distribution on Θ_L, $\operatorname{rank} \mathcal{Q}^{(L)}(\boldsymbol{\theta}) = d_L$ with probability 1.

Claim (iii): monotonicity under nesting. If the depth-$(L{+}1)$ circuit family contains the depth-L family—i.e., for every $\boldsymbol{\theta} \in \Theta_L$ there exists $\boldsymbol{\theta}' \in \Theta_{L+1}$ such that $U_{L+1}(\boldsymbol{\theta}') = U_L(\boldsymbol{\theta})$ (achieved by setting the additional layer parameters to implement the identity, which is possible for ansätze where rotations have zero-angle identity points)—then $\mathcal{M}_L \subseteq \mathcal{M}_{L+1}$. Therefore $\dim \mathcal{M}_{L+1} \geq \dim \mathcal{M}_L$, and by (i), $d_{L+1} \geq d_L$.

Proposition 2 (Parameter-budget lower bound). The QGT at depth L is a $P_L \times P_L$ matrix, so $d_L = \operatorname{rank}(\mathcal{Q}_L) \leq P_L$. With $P_L = L \cdot p(N)$ parameters, exact saturation $d_L = D_{\text{orbit}}$ requires $D_{\text{orbit}} \leq L \cdot p(N)$, hence $L \geq D_{\text{orbit}}/p(N)$.

Proposition 3 (Lie-orbit geometric ceiling)**.** *Let $\mathfrak{g}$ be the Lie algebra generated by the circuit's parameterized generators under commutator closure, let $G = \exp(\mathfrak{g}) \subseteq \mathrm{SU}(2^N)$ be the corresponding connected Lie subgroup, and let $\mathcal{O}_{\psi_0} = \{[U|\psi_0\rangle] : U \in G\} \subseteq \mathbb{CP}^{2^N-1}$ be the* projective orbit *of the initial state $|\psi_0\rangle = |0\rangle^{\otimes N}$, where $[\cdot]$ denotes the equivalence class modulo global phase. Define the* projective stabilizer algebra

$$\mathfrak{p}_{\psi_0} = \big\{X \in \mathfrak{g} : X|\psi_0\rangle = i\phi\,|\psi_0\rangle \text{ for some } \phi \in \mathbb{R}\big\}, \tag{8}$$

which includes both operators that annihilate $|\psi_0\rangle$ and those that act as a global phase on $|\psi_0\rangle$. Then the orbit dimension is

$$D_{orbit} = \dim \mathcal{O}_{\psi_0} = \dim \mathfrak{g} - \dim \mathfrak{p}_{\psi_0}. \tag{9}$$

For any finite depth L, the reachable set satisfies

$$\mathcal{M}_L \subseteq \mathcal{O}_{\psi_0}, \tag{10}$$

and consequently $d_L \leq D_{orbit}$. Equality $d_L = D_{orbit}$ may be approached asymptotically as $L \to \infty$ if the circuit's generators generate $\mathfrak{g}$ densely, so that repeated products of layer unitaries fill the group G.

Proof. The circuit's parameterized unitaries form a connected Lie subgroup $G = \exp(\mathfrak{g})$ of $\mathrm{SU}(2^N)$, where $\mathfrak{g}$ is the Lie algebra generated by the circuit's rotation generators under commutator closure. The reachable state set in projective Hilbert space is the orbit $\mathcal{O}_{\psi_0} = \{[U|\psi_0\rangle] : U \in G\} \subseteq \mathbb{CP}^{2^N-1}$, where $[\cdot]$ denotes the equivalence class modulo global phase. By the orbit-stabilizer theorem for Lie group actions on projective space, $\dim \mathcal{O}_{\psi_0} = \dim G - \dim G_{\psi_0}^{\text{proj}}$, where

$G_{\psi_0}^{\text{proj}} = \{U \in G : U|\psi_0\rangle = e^{i\phi}|\psi_0\rangle \text{ for some } \phi \in \mathbb{R}\}$ is the projective stabilizer (stabilizing the state up to global phase). At the Lie algebra level, the projective stabilizer algebra is

$$\mathfrak{p}_{\psi_0} = \{X \in \mathfrak{g} : X|\psi_0\rangle = i\phi\,|\psi_0\rangle \text{ for some } \phi \in \mathbb{R}\}, \tag{11}$$

which includes both operators that annihilate $|\psi_0\rangle$ and those proportional to the identity on $|\psi_0\rangle$ (i.e., global-phase generators). Then $\dim \mathcal{O}_{\psi_0} = \dim \mathfrak{g} - \dim \mathfrak{p}_{\psi_0}$. Since QGT rank measures the tangent dimension of this orbit at generic parameter points (by Proposition 1(i)), and the orbit dimension is a topological invariant, $D_{\text{orbit}} = \dim \mathcal{O}_{\psi_0} = \dim \mathfrak{g} - \dim \mathfrak{p}_{\psi_0}$. For any finite depth L, the reachable set $\mathcal{M}_L = \Phi_L(\Theta_L)$ is a subset of $\mathcal{O}_{\psi_0}$, so $d_L \leq D_{\text{orbit}}$. As $L \to \infty$, the parameterized family may densely fill G (if the generators generate $\mathfrak{g}$), in which case $d_L \to D_{\text{orbit}}$.

For HEA with a connected CNOT topology (e.g., chain, ring, star), the single-qubit rotation generators $\{X_i, Y_i, Z_i\}_{i=1}^N$ together with CNOT conjugation generate all N-qubit Pauli strings. Concretely, $\text{CNOT}_{ij}\, X_i\, \text{CNOT}_{ij}^\dagger = X_i X_j$, and iterating across a connected graph produces all weight-k Pauli strings for $k = 1, \ldots, N$. Since the $4^N - 1$ traceless Pauli operators form a basis of $\mathfrak{su}(2^N)$, we have $\mathfrak{g} = \mathfrak{su}(2^N)$ and $\dim \mathfrak{g} = 4^N - 1$. The projective stabilizer satisfies $\dim \mathfrak{p}_{\psi_0} = \dim \mathfrak{su}(2^N) - \dim \mathcal{O}_{\psi_0} = (4^N - 1) - (2^{N+1} - 2) = 4^N - 2^{N+1} + 1$, so $D_{\text{orbit}} = 2^{N+1} - 2 = d_{\text{manifold}}$. For HVA, we empirically observe that d_L approaches d_{manifold} for $N \leq 8$, consistent with the conjecture that HVA also generates the full $\mathfrak{su}(2^N)$ on a connected topology; we do not prove this and treat it as an empirical observation.

B Saturation Overhead and Efficiency Hypothesis

Definition 1 (Saturation overhead)**.** *Define the* saturation overhead

$$\alpha_N \;=\; \frac{L^\star_{\text{exact}}}{\lceil D_{orbit}/p(N) \rceil}, \tag{12}$$

which measures how many layers beyond the parameter-counting lower bound are needed to achieve exact geometric saturation. By Proposition 2, $\alpha_N \geq 1$. An architecture with $\alpha_N = 1$ is geometrically optimal*: essentially every newly added parameter contributes an independent tangent direction until the orbit ceiling is reached.*

The observed ratio α_N decreasing from 1.33 (N=4) to 1.14 (N=8) for HEA suggests the following hypothesis, which can be tested across architectures and topologies:

Saturation-efficiency hypothesis. *Sufficiently mixing hardware-efficient ansätze approach the parameter-budget lower bound as N grows ($\alpha_N \to 1$), while symmetry-constrained or poorly connected architectures exhibit a nonvanishing saturation overhead ($\alpha_N > 1$ uniformly in N).*

Under this hypothesis, the ratio α_N provides a principled, architecture-comparable metric for geometric efficiency: lower α_N means each layer contributes more independent directions on the way to the ceiling.

C Experiment Details

GSD algorithm pseudocode. The adaptive GSD procedure is given below. τ-GSD (threshold-based) is a simpler alternative: return the smallest L such that $d_L/d_{L-1} - 1 < \tau$.

Algorithm: Geometric Saturation Diagnosis (Adaptive GSD)

Input: Number of qubits N; entanglement topology $\mathcal{E}$; max depth $L_{\max}$; SVD tolerance ϵ.
Output: Saturation depth $L^\star_{\text{adaptive}}$.

1. **for** $L = 1, 2, \ldots, L_{\max}$ **do**
2. Sample random $\boldsymbol{\theta} \in [0, 2\pi)^{3NL}$ (justified by Proposition 1(ii)).
3. Compute state-vector Jacobian J via automatic differentiation.
4. Compute $\mathcal{Q}^{(L)} = J^\dagger(I - |\psi\rangle\langle\psi|)J$.
5. Compute $d_L = \text{rank}_\epsilon(\mathcal{Q}^{(L)})$ via SVD with relative tolerance $\sigma_i > \epsilon\sigma_1$.
6. **end for**
7. **return** $L^\star_{\text{adaptive}} = \arg\min_{L_0} \left[\sum_{l \le L_0}(d_l - \hat{d}_l^{\text{left}})^2 + \sum_{l > L_0}(d_l - \hat{d}_l^{\text{right}})^2\right]$
(L-method knee detection; τ-GSD with τ=5% is a simpler alternative.)

Hamiltonians. We use two spin-model Hamiltonians on a 1D chain with periodic boundary conditions. The transverse-field Ising model (TFIM) is $H_{\text{TFIM}} = -\sum_i Z_i Z_{i+1} - h\sum_i X_i$ with $h = 1$. The Heisenberg XXX model is $H_{\text{XXX}} = \sum_i (X_i X_{i+1} + Y_i Y_{i+1} + Z_i Z_{i+1})$. Ground-state energies are obtained by exact diagonalization.

Ansatze. The hardware-efficient ansatz (HEA) uses R_x, R_y, R_z per qubit per layer followed by CNOT gates on adjacent pairs (chain topology), giving $P = 3NL$ parameters. The Hamiltonian variational ansatz (HVA) uses three parameterized rotations per qubit per layer with alternating $XX + YY + ZZ$ entangling terms, giving $P = 3NL$ parameters.

Training. All VQE training uses the Adam optimizer with learning rate 0.05 for 100 steps, initialized from random parameters in $[-\pi, \pi]$. Results are averaged over 3 seeds unless otherwise noted. Noisy simulations use depolarizing noise on each qubit after every gate.

C.1 Additional Experimental Results

QGT singular value spectrum. The full QGT singular value spectrum reveals three regimes: under-parameterized (all directions active), geometrically saturated (rank matches D_{orbit}), and over-parameterized (dormant directions with negligible singular values). Figure 3 shows the sorted spectrum for N=6 (HVA + Heisenberg) at three depths: L_{half}=4 (P=72), $L^\star$=8 (P=144), and $L_{\max}$=14 (P=252). At L_{half}, the spectrum spans only $\sim$ 55 non-negligible indices before dropping to machine precision; at $L^\star$, it extends to $\sim$ 126 matching d_{manifold}; at $L_{\max}$, the additional 108 parameters contribute singular values 7–10 orders of magnitude smaller than the active ones. This provides a geometry-based parameter-direction importance ranking, analogous to eigenvalue-based pruning criteria in deep learning but derived from the Riemannian metric.

Robustness of the saturation criterion. Table 3 shows the full robustness data: SVD tolerance sensitivity, growth threshold sweep, and adaptive GSD comparison.

Table 3: Robustness of the saturation criterion (HEA + TFIM). *Left:* $L^\star_\tau$ under four SVD tolerances ϵ for N=6 (3 seeds). *Right:* $L^\star_\tau$ under four growth thresholds τ for N=4, 6, 8.

SVD tolerance (N=6)				Growth threshold τ				
ϵ	$L^\star_\tau$	$d_{L^\star}$	d/d_{manifold}	N	1%	2%	5%	10%
10^{-6}	9	126	100%	4	4	4	4	4
10^{-8}	8	121	96%	6	9	9	8	8
10^{-10}	8	123	98%	8	24	24	21	11
10^{-12}	8	123	98%					

Effective rank: supplementary data. As defined in Section 3, the effective rank $d_{\text{eff}} = \exp(H(p))$ with $p_i = \sigma_i / \sum_j \sigma_j$ provides a continuous capacity measure. For HVA+Heisenberg at N=6, d_{eff} rises from 10 at L=1 to 53 at $L^\star$=8 and 74 at $L_{\max}$=14, while d_{hard} saturates at 126 by L=9. The growing gap between d_{hard} and d_{eff} indicates that late layers add directions with progressively smaller metric weight, confirming the spectral improvement predicted by the two-saturation framework.

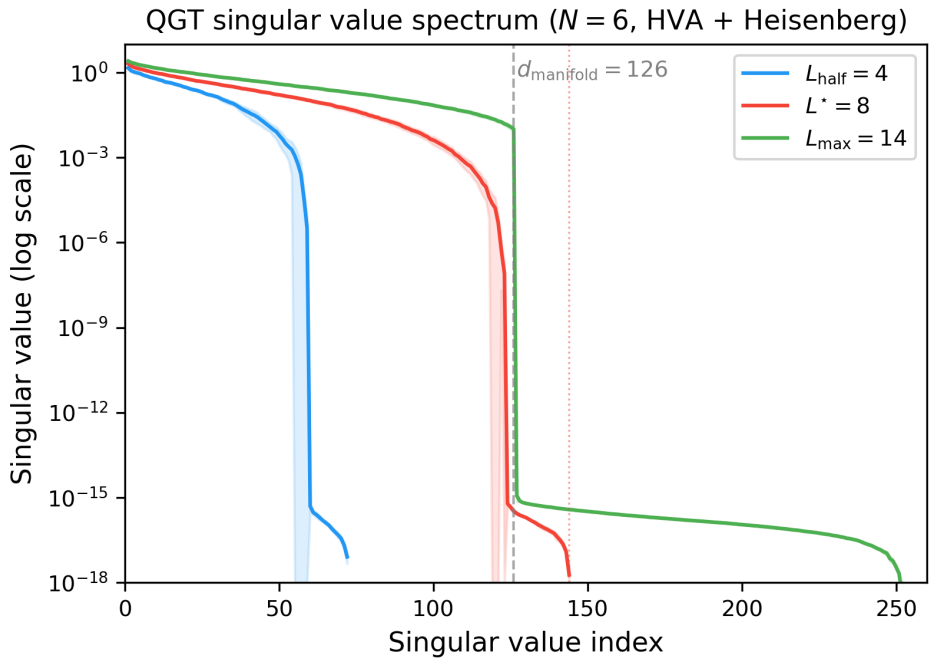


Figure 3: QGT singular value spectrum for N=6 (HVA + Heisenberg) at three depths: L_{half}=4 (P=72), $L^\star$=8 (P=144), and L_{max}=14 (P=252). Singular values are sorted in descending order (5-seed mean with shaded ± 1 std band) and plotted on a log scale. The dashed line marks $d_{\text{manifold}} = 126$, separating active directions (left) from redundant directions with negligible metric weight (right).

Two-saturation framework and task-relevant tangent gain. Hard rank d_L counts directions with nonzero metric weight, but the QGT singular value spectrum reveals that not all directions are equally accessible. Two complementary spectral quantities characterize the distribution:

- *Effective rank*: $d_{\text{eff}}(L) = \exp(H(p))$ where $p_i = \sigma_i / \sum_j \sigma_j$ and H is Shannon entropy. This measures *spectral dispersion* or isotropy—how evenly the metric weight is distributed across directions.
- *Effective condition number*: $\kappa_{\text{eff}}(Q) = \sigma_1/\sigma_r$ where $r = d_L$. This measures *worst-case conditioning*—the ratio between the most and least sensitive active directions.

Both grow and evolve continuously with depth, even after hard rank saturates. This distinguishes two saturation phenomena: *dimensional saturation* (no new independent directions; $d_L = D_{\text{orbit}}$) and *spectral saturation* (both d_{eff} and κ_{eff} plateau; the singular value distribution stabilizes). The former is a structural ceiling; the latter indicates that the manifold's spectral distribution has stabilized, meaning additional layers do not significantly improve the geometry of optimization. Deeper circuits can become more isotropic (higher d_{eff}), better conditioned (lower κ), both, or neither—these are distinct aspects of spectral geometry. For HVA at N=6, dimensional saturation occurs at L=9 ($d_L = 126$) while d_{eff} continues rising from 53 at $L^\star$=8 to 74 at L_{max}=14, indicating persistent spectral improvement even after dimensional saturation.

The *task-relevant tangent gain* is defined geometrically on the projective Hilbert space. Let $T_L = T_{[\psi]}\mathcal{M}_L$ denote the tangent space of the depth-L state manifold at $[\psi]$ (a subspace of the tangent space of $\mathbb{CP}^{2^N-1}$), and define the *newly added tangent subspace*

$$V_L^{\text{new}} = T_L \ominus T_{L-1}, \tag{13}$$

where orthogonality is with respect to the Fubini–Study metric (the restriction of the real part of the QGT to the tangent space). Let $\nabla_{\text{FS}} E$ be the Riemannian gradient of the energy E with respect

to the Fubini–Study metric, i.e., the unique tangent vector such that $dE(v) = \langle \nabla_{\mathrm{FS}} E, v \rangle_{\mathrm{FS}}$ for all tangent vectors v. Then

$$\tau_L \;=\; \frac{\| P_{V_L^{\mathrm{new}}} \nabla_{\mathrm{FS}} E \|_{\mathrm{FS}}^2}{\| \nabla_{\mathrm{FS}} E \|_{\mathrm{FS}}^2}, \tag{14}$$

where $P_{V_L^{\mathrm{new}}}$ denotes orthogonal projection onto V_L^{new} under the Fubini–Study metric, and $\| \cdot \|_{\mathrm{FS}}$ is the Fubini–Study norm. This definition is manifestly coordinate-invariant, consistent with the geometric philosophy of the paper.

In practice, τ_L can be computed in parameter coordinates via the QGT. Let $\nabla_\theta E = \partial E / \partial \boldsymbol{\theta}$ be the parameter-space gradient. The Riemannian gradient pushes forward as $D\Phi_L(\boldsymbol{\theta})\, \mathcal{Q}_L^+ \nabla_\theta E$, and the Fubini–Study norm squared of the gradient is $\nabla_\theta E^\top \mathcal{Q}_L^+ \nabla_\theta E$ (where $\mathcal{Q}_L^+$ is the Moore–Penrose pseudoinverse). The coordinate expression for τ_L is then

$$\tau_L = \frac{\nabla_\theta E^\top \mathcal{Q}_L^{+1/2} P_{\widehat{V}_L^{\mathrm{new}}} \mathcal{Q}_L^{+1/2} \nabla_\theta E}{\nabla_\theta E^\top \mathcal{Q}_L^+ \nabla_\theta E}, \tag{15}$$

where $\widehat{V}_L^{\mathrm{new}}$ is the parameter-space representation of the newly added directions. For practical computation, a simpler approximation projects the QGT-weighted gradient onto the subspace of parameter directions corresponding to the newly added principal components of $\mathcal{Q}_L$ not already present at depth $L-1$.

When $\tau_L \approx 0$, the new geometric directions are nearly orthogonal to the energy gradient in the Fubini–Study metric—the manifold expanded, but not in the task-relevant direction. This explains why L_{half} and $L^\star$ can achieve similar energy despite very different d_L: the directions added between L_{half} and $L^\star$ contribute geometric capacity but progressively less to energy reduction. The τ_L metric bridges the gap between task-agnostic geometric capacity (GSD) and task-specific depth selection, and connects to the classical notion of *effective dimension* in statistical learning (Advani et al., 2020).

The following proposition makes precise the sense in which τ_L controls the optimization value of additional layers:

Proposition 4 (Task-relevant tangent gain and first-order descent)**.** *Let $v \in V_L^{\mathrm{new}}$ be a unit-norm tangent vector in the newly added subspace ($\|v\|_{\mathrm{FS}} = 1$). The maximal first-order decrease in energy achievable by moving along any direction in V_L^{new} with step size η is*

$$\max_{v \in V_L^{\mathrm{new}},\, \|v\|_{\mathrm{FS}}=1} \big[E([\psi]) - E([\psi] + \eta v) \big] \;=\; \eta \, \| P_{V_L^{\mathrm{new}}} \nabla_{\mathrm{FS}} E \|_{\mathrm{FS}} \;+\; O(\eta^2). \tag{16}$$

Consequently, if $\tau_L = 0$, then adding layer L provides no new first-order descent capability *at that state, even if $\Delta d_L > 0$ (i.e., even though the new directions are geometrically independent). In particular,*

$$\Delta d_L > 0 \quad \not\Longrightarrow \quad \textit{task improvement}. \tag{17}$$

Proof. The directional derivative of E along v is $dE(v) = \langle \nabla_{\mathrm{FS}} E, v \rangle_{\mathrm{FS}}$. The first-order Taylor expansion is $E([\psi] + \eta v) = E([\psi]) + \eta \langle \nabla_{\mathrm{FS}} E, v \rangle_{\mathrm{FS}} + O(\eta^2)$. The energy decrease is $E([\psi]) - E([\psi] + \eta v) = -\eta \langle \nabla_{\mathrm{FS}} E, v \rangle_{\mathrm{FS}} + O(\eta^2)$. Maximizing over unit vectors $v \in V_L^{\mathrm{new}}$ gives $-\eta \min_{v \in V_L^{\mathrm{new}}, \|v\|=1} \langle \nabla_{\mathrm{FS}} E, v \rangle_{\mathrm{FS}} = \eta \| P_{V_L^{\mathrm{new}}} \nabla_{\mathrm{FS}} E \|_{\mathrm{FS}}$, where the equality follows from the Cauchy–Schwarz inequality (the minimum inner product is the negative of the norm of the projected gradient). If $\tau_L = 0$, then $\| P_{V_L^{\mathrm{new}}} \nabla_{\mathrm{FS}} E \|_{\mathrm{FS}} = 0$, so the first-order decrease is zero. □

This proposition rigorously separates two notions of "adding capacity": structural capacity ($\Delta d_L > 0$, new geometric directions exist) and task-relevant capacity ($\tau_L > 0$, those directions reduce the objective to first order). The gap between them is precisely why a geometric diagnostic like GSD cannot, by itself, determine the task-optimal depth—it identifies the safe upper bound, while task-specific quantities like τ_L quantify how much of that bound is actually useful.

Parameter-point robustness. Proposition 1(ii) guarantees that a single random parameter draw recovers the generic rank with probability 1. We empirically verify this by computing $\mathrm{rank}(\mathcal{Q})$ at 10 independent random parameter points for $N{=}6$ across all depths $L{=}1, \ldots, 14$. The standard deviation across points is at most 1.17 (at $L{=}8$, where borderline singular values cause occasional ± 1 rank fluctuations) and is exactly 0 for $L \leq 2$ and $L \geq 9$. The saturation depth $L^\star$ is 8 or 9 for all 10 points, confirming that the generic rank assumption holds in practice.

Verification of the parameter-budget lower bound. Table 4 verifies Proposition 2 for both HEA and HVA ($p(N)$=$3N$). The empirical L_{exact} satisfies the bound in all cases, with ratio $L_{\text{exact}}/L_{\text{bound}}$ decreasing from 1.33 to 1.14 as N grows, suggesting the bound becomes tight.

Table 4: Verification of the parameter-budget lower bound (Proposition 2). $L_{\text{bound}} = \lceil d_{\text{manifold}}/(3N) \rceil$. Both HEA and HVA have $p(N) = 3N$ parameters per layer.

N	d_{manifold}	$p(N)$=$3N$	L_{bound}	L_{exact}	Ratio
4	30	12	3	4	1.33
6	126	18	7	9	1.29
8	510	24	22	25	1.14
10	2046	30	69	>12 (unsaturated)	—

Full saturation results. Table 5 reports saturation depths for both ansätze (HEA+TFIM and HVA+Heisenberg) across all tested system sizes. The qualitative pattern—sub-linear growth of L^* relative to d_{manifold}, with N=10 remaining unsaturated at the maximum tested depth—is consistent across both ansätze.

Table 5: Full saturation results for HEA+TFIM and HVA+Heisenberg ansatze across all system sizes. L^* uses the 5% relative growth criterion; L_{exact} is the depth at which $\text{rank}(\mathcal{Q}) = d_{\text{manifold}}$.

		HEA + TFIM			**HVA + Heisenberg**		
N	d_{man}	L^*	d_{L^*}	L_{exact}	L^*	d_{L^*}	L_{exact}
4	30	4	30	4	5	30	4
6	126	8	121	9	9	126	9
8	510	21	454	25	21	450	25
10	2046	12	326	>max	12	325	>max

Topology comparison. Table 6 compares saturation depths across three CNOT topologies: chain, ring, and star. Ring topology saturates slightly faster for small N due to higher per-layer CNOT count, while star topology requires more layers for $N = 6$ because its all-to-all pattern wastes entanglement capacity on non-adjacent pairs that contribute fewer independent directions early on.

Table 6: Saturation depth for three CNOT topologies (chain, ring, star) at $N = 4$ and $N = 6$, HEA ansatz.

	$N = 4$ ($d_{\text{man}} = 30$)			$N = 6$ ($d_{\text{man}} = 126$)		
Topology	L^*	d_{L^*}	CNOTs/layer	L^*	d_{L^*}	CNOTs/layer
chain	4	30	3	8	121	5
ring	3	30	4	8	126	6
star	4	30	3	11	126	5

Cross-architecture summary. Table 7 summarizes saturation depths across all tested ansätze, Hamiltonians, and topologies. The qualitative pattern is consistent: $L^\star_\tau$ grows sub-linearly with d_{manifold}, and star topology requires more layers than chain/ring for N=6.

Full noise robustness results. Table 8 extends the noise analysis to $N = 4$ and $N = 6$ across all four noise levels. The pattern observed for $N = 8$ in the main text—shallower circuits exhibit better noisy energy, with the gap growing with noise strength—holds consistently across all system sizes.

Training convergence curves. Figure 4 compares training convergence for three circuit depths on HVA+Heisenberg $N = 8$: L_{half} (half of L^*), L^* (GSD-identified saturation depth), and $L_{\max}$ (maximum tested depth). While $L_{\max}$ achieves the lowest final energy, L^* reaches within 2.6% of $L_{\max}$'s energy with 29% fewer parameters and substantially faster initial convergence. L_{half} converges quickly but plateaus at a noticeably higher energy, confirming that going below L^* sacrifices meaningful expressive capacity.

Table 7: Saturation depths across ansätze, Hamiltonians, and topologies. All results use τ-GSD with 5% threshold and 5 seeds. $d_{L^\star}/d_{\text{manifold}}$ is the fraction of reachable-set dimension captured at $L^\star_\tau$. For N=8, only chain topology was tested.

			N=4		N=6		N=8
Ansatz	Hamiltonian	Topology	$L^\star_\tau$	d/d_{manifold}	$L^\star_\tau$	d/d_{manifold}	$L^\star_\tau$
HEA	TFIM	chain	4	100%	8	96%	21
HEA	TFIM	ring	3	100%	8	100%	—
HEA	TFIM	star	4	100%	11	100%	—
HVA	Heisenberg	chain	4	100%	8	100%	20

Table 8: Noisy energy (mean $\pm$ std) and noise gap for L^*, L^*_{+2}, and $L_{\max}$ across all noise levels, for $N = 4$ and $N = 6$, HEA+TFIM.

	$N = 4$			$N = 6$		
Noise p	L^* (4)	L^*_{+2} (6)	$L_{\max}$ (12)	L^* (8)	L^*_{+2} (10)	$L_{\max}$ (14)
Noisy final energy						
0.005	-4.59 ± 0.03	-4.22 ± 0.11	-3.54 ± 0.06	-5.64 ± 0.17	-5.35 ± 0.14	-4.12 ± 0.28
0.01	-4.12 ± 0.05	-3.63 ± 0.06	-2.71 ± 0.07	-4.39 ± 0.14	-4.06 ± 0.02	-3.09 ± 0.02
0.02	-3.31 ± 0.09	-2.72 ± 0.01	-1.53 ± 0.09	-2.95 ± 0.18	-2.33 ± 0.10	-1.30 ± 0.06
0.05	-1.84 ± 0.04	-1.20 ± 0.05	-0.37 ± 0.05	-0.95 ± 0.02	-0.51 ± 0.15	-0.24 ± 0.07
Noise gap (clean − noisy)						
0.005	0.62	1.00	1.69	1.98	2.35	3.61
0.01	1.09	1.60	2.51	3.23	3.64	4.63
0.02	1.90	2.50	3.70	4.68	5.37	6.43
0.05	3.37	4.02	4.86	6.68	7.18	7.48

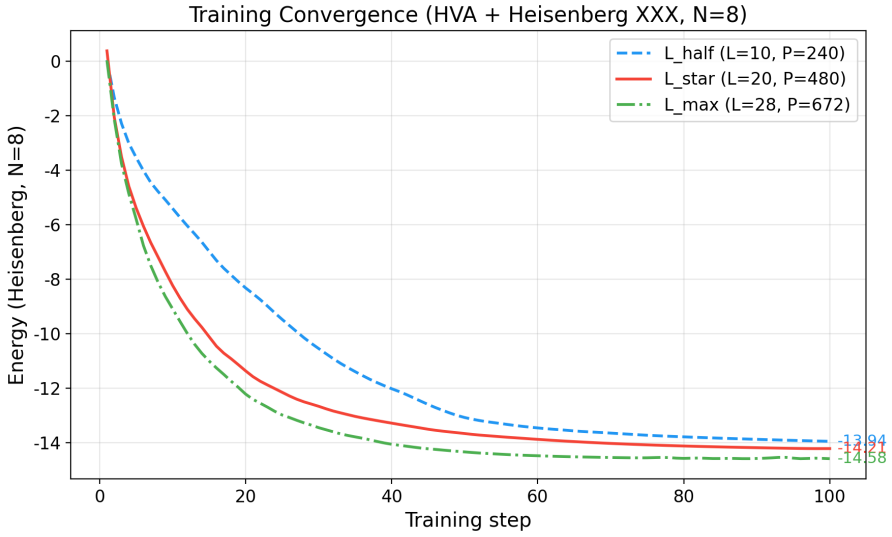


Figure 4: Training convergence (energy as a function of optimization steps) for HVA+Heisenberg $N = 8$ at three depths: $L_{\text{half}} = 10$ (240 params), $L^* = 20$ (480 params), and $L_{\max} = 28$ (672 params). Curves are averaged over 10 random initializations; shaded regions show ± 1 standard deviation.

HVA+Heisenberg noise robustness. Table 9 extends the noise analysis to the HVA+Heisenberg setting for $N = 4$, 6, and 8. The qualitative pattern—shallower circuits are more noise-robust ($L_{\text{half}} > L^* > L_{\max}$ in noisy energy)—matches the HEA+TFIM results in Table 8, confirming that the noise–depth trade-off is ansatz-independent.

Table 9: Noisy energy (mean ± std) for L_{half}, L^*, and $L_{\max}$ across five noise levels, HVA+Heisenberg. Top: $N = 4$ and $N = 6$; bottom: $N = 8$.

	$N = 4$			$N = 6$		
p	L_{half} (2)	L^* (4)	$L_{\max}$ (12)	L_{half} (4)	L^* (8)	$L_{\max}$ (14)
0.001	-6.34 ± 0.00	-7.54 ± 0.01	-6.63 ± 0.01	-10.18 ± 0.11	-9.46 ± 0.08	-8.05 ± 0.02
0.005	-5.80 ± 0.00	-6.02 ± 0.05	-3.55 ± 0.08	-7.94 ± 0.08	-5.77 ± 0.17	-3.88 ± 0.19
0.01	-5.20 ± 0.00	-4.68 ± 0.07	-1.94 ± 0.01	-5.94 ± 0.10	-3.54 ± 0.18	-1.78 ± 0.09
0.02	-4.15 ± 0.05	-2.92 ± 0.09	-0.65 ± 0.06	-3.67 ± 0.05	-1.39 ± 0.16	-0.39 ± 0.10
0.05	-2.20 ± 0.01	-0.84 ± 0.04	-0.05 ± 0.00	-1.03 ± 0.01	-0.14 ± 0.04	-0.01 ± 0.01

	$N = 8$		
p	L_{half} (10)	L^* (20)	$L_{\max}$ (28)
0.001	-11.16 ± 0.23	-7.25 ± 0.05	-5.54 ± 0.07
0.005	-6.29 ± 0.38	-2.32 ± 1.13	-1.29 ± 0.55
0.01	-3.38 ± 0.25	-0.92 ± 0.11	-0.34 ± 0.07
0.02	-1.15 ± 0.07	-0.16 ± 0.04	-0.02 ± 0.00
0.05	-0.09 ± 0.01	-0.00 ± 0.00	0.00 ± 0.00

QGT rank saturation curves. Figure 5 visualizes QGT rank growth as a function of circuit depth for $N = 4$, 6, and 8 (HEA+TFIM, chain topology, 5 seeds). In all three cases, the rank grows rapidly in early layers, approaches d_{manifold} (red dashed line), and then plateaus. The GSD-identified L^* (green dotted line) consistently marks the transition from rapid growth to the plateau region, confirming that the 5% relative growth criterion captures the geometric saturation point.

Geometric capacity and task performance. Figure 6 directly tests whether QGT rank growth predicts task performance: for HVA+Heisenberg at $N = 4$ and $N = 6$, we train VQE at *every* depth $L = 1, \ldots, L_{\max}$ (3 seeds each) and plot relative ground-state error against the fraction of geometric capacity (d_L/d_{manifold}). The results show a clear monotone correspondence: as the QGT rank saturates, the energy error drops by orders of magnitude and then plateaus. For $N = 4$, energy convergence ($< 10^{-4}$) is achieved exactly at $L^* = 4$ where rank reaches d_{manifold}. For $N = 6$, energy approaches the ground state around $L^* = 8$ and shows diminishing returns beyond, with intermediate depths ($L = 9$) occasionally performing worse due to optimization difficulty in the larger parameter space. This validates GSD's central claim: geometric rank saturation is a reliable predictor of when additional circuit depth ceases to improve task performance.

Symmetry-constrained ansatz. Proposition 3 predicts that an ansatz whose generators preserve a symmetry should saturate at a lower orbit dimension $D_{\text{orbit}} < d_{\text{manifold}}$. To test this, we construct a particle-number-conserving ansatz using Ising-type $XX{+}YY$ entangling gates (which preserve excitation number) plus single-qubit R_z rotations, initialized in a state with k excitations. For N qubits and k excitations, the full k-excitation sector has real dimension $2\binom{N}{k} - 2$, providing a theoretical upper bound on D_{orbit} for any ansatz that preserves particle number.

Table 10 reports the results. For $N{=}4$ with $k{=}2$ excitations, the symmetric ansatz saturates at rank 8 (27% of $d_{\text{manifold}}{=}30$), compared to 30 for HEA. For $N{=}6$ with $k{=}3$, the symmetric ansatz saturates at 18 (14% of $d_{\text{manifold}}{=}126$), compared to 126 for HEA. The symmetric ansatz also reaches saturation much faster ($L^\star{=}3$ compared to 4 for $N{=}4$, and 4 compared to 8 for $N{=}6$), as expected from its lower per-layer parameter count and lower ceiling. In both cases the saturated rank lies strictly below the full symmetric-sector dimension ($D_{\text{sym}} = 2\binom{N}{k} - 2$), indicating that the $XX{+}YY$ plus R_z gate set is not universal on the k-excitation sector—it generates only a subalgebra. This demonstrates that GSD can diagnose not only full-manifold saturation but also the

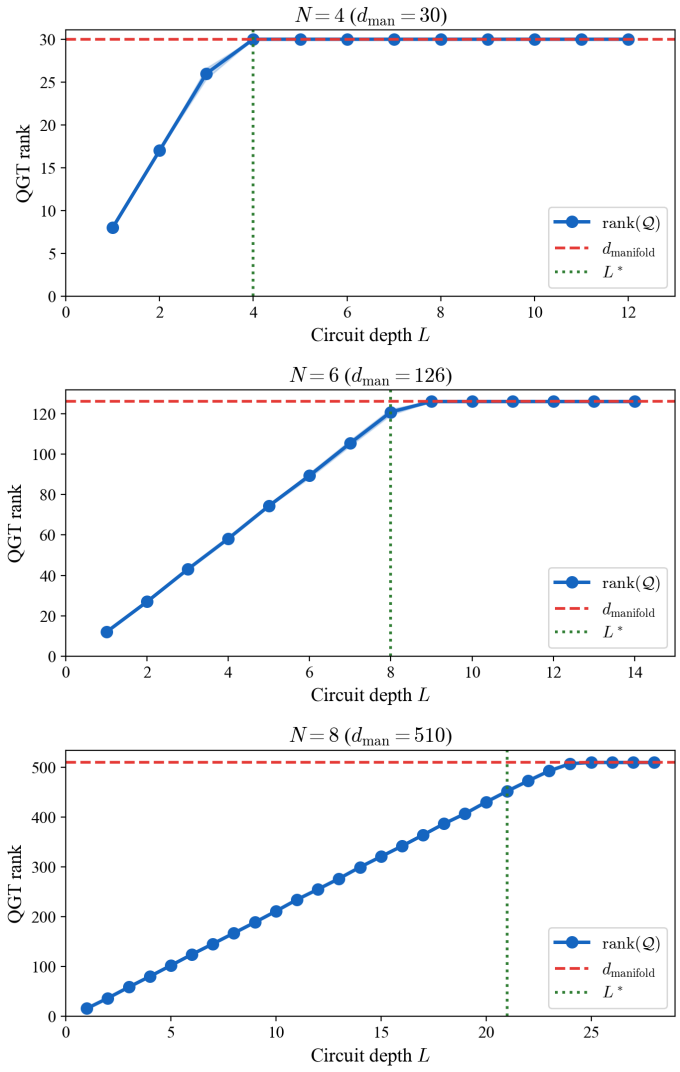


Figure 5: QGT rank saturation curves for HEA+TFIM at $N = 4, 6, 8$ (chain topology, 5 seeds). Blue circles: mean rank with ± 1 std shading; red dashed: d_{manifold}; green dotted: L^*.

Table 10: Saturation of symmetry-constrained (particle-number-conserving) ansatz compared to standard HEA. $D_{\text{sym}} = 2\binom{N}{k} - 2$ is the dimension of the full k-excitation symmetric sector. D_{sat} is the empirically observed saturated QGT rank. The symmetric ansatz saturates at a manifold dimension dramatically lower than d_{manifold} (73–86% reduction), confirming that GSD detects symmetry-imposed orbit ceilings. The fact that $D_{\text{sat}} < D_{\text{sym}}$ indicates that this specific gate set is not universal on the symmetric sector—it generates only a sub-algebra of the full $\mathfrak{su}(\binom{N}{k})$.

N	k	Ansatz	Params/layer	D_{sym}	D_{sat}	$L^\star$	$D_{\text{sat}}/d_{\text{manifold}}$
4	2	Symmetric	7	10	8	3	27%
4	2	HEA	12	30	30	4	100%
6	3	Symmetric	11	38	18	4	14%
6	3	HEA	18	126	126	8	100%

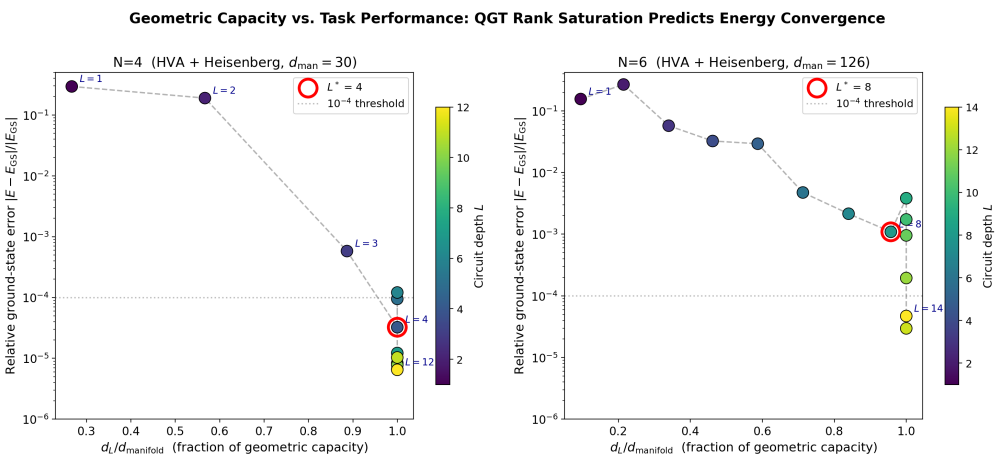


Figure 6: Geometric capacity and task performance for HVA+Heisenberg at $N = 4$ (left) and $N = 6$ (right). Each point is one circuit depth L (color-coded), with x-axis showing QGT rank as fraction of d_{manifold} and y-axis showing relative ground-state error. Red circle: L^* (GSD-identified saturation depth). The monotone relationship supports that geometric rank saturation predicts energy convergence.

actual orbit dimension of any symmetry-constrained or architecture-constrained ansatz, providing an empirical measurement of how much of a symmetric sector a given gate set can actually explore.