One Number Predicts Barren Plateaus: The Lindblad Spectral Gap as Trainability Bound

Don't run the circuit. Compute | |. Know the answer.

Tam\{}\{}'as Nagy, Ph.D.

tnagyphd@gmail.com

Draft

Abstract

We show that the trainability of variational quantum circuits is determined by a single number: the spectral gap $|\lambda_1|$ of the Lindblad noise generator. The gradient variance of any parametrized circuit with n qubits at depth d satisfies $\mathrm{Var}[\partial_\theta C] \sim 2^{-n} \cdot \exp(-2|\lambda_1| \cdot d \cdot t_{\mathrm{eff}})$, where t_{eff} is the effective gate time per layer. This gives an explicit critical depth:

$$d^* = \frac{\log(10^4 \cdot 2^n)}{2\,|\lambda_1| \cdot t_{\mathrm{eff}} \cdot n}$$

Beyond d^*: the gradient vanishes below any useful threshold. The circuit is untrainable. Below d^*: gradients are usable and optimization can proceed.

For IBM Eagle (2024, $T_1 = 300\,\mu$s): $d^*(4 \text{ qubits}) = 2{,}643$, $d^*(50 \text{ qubits}) = 774$. The noise-induced barren plateau is NOT the dominant effect for shallow circuits on current hardware — the expressibility plateau (from Haar-random initialization) dominates at depth \$ \$50. But for 50+ qubit systems at depth \$>\$500 — the regime of practical quantum advantage — the noise-induced plateau becomes the binding constraint, and $|\lambda_1|$ is the number that predicts it.

The spectral gap $|\lambda_1|$ is computable from a 64×64 cluster Lindblad matrix in 0.001 seconds, without running any quantum circuit. This enables: (1) BEFORE optimization: predict whether the circuit is trainable; (2) DURING design: choose the maximum depth for a given hardware; (3) ACROSS hardware: rank quantum computers by their trainability depth d^*.

---

1. Introduction

1.1 The Barren Plateau Problem

Variational quantum algorithms (VQE, QAOA, quantum machine learning) optimize parametrized circuits by gradient descent. The **barren plateau** phenomenon (McClean et al., 2018): for sufficiently deep or wide circuits, the gradient variance $\mathrm{Var}[\partial_\theta C]$ vanishes exponentially in n, making optimization impossible.

Two sources of barren plateaus are known:

1. **Expressibility plateau** (McClean et al., 2018; Cerezo et al., 2021): deep random circuits sample uniformly from the Haar measure. The cost function gradient concentrates at zero: $\mathrm{Var}[\partial_\theta C] \sim 2^{-n}$.
2. **Noise-induced plateau** (Wang et al., 2021): noise during circuit execution further suppresses gradients: $\mathrm{Var}[\partial_\theta C] \sim 2^{-n} \cdot e^{-\alpha \cdot n \cdot d}$ for some noise rate α.

The constant α is typically estimated from heuristic noise models (depolarizing, amplitude damping). We show it has an exact spectral characterization.

1.2 Our Result

$$\alpha = 2\,|\lambda_1| \cdot t_{\text{eff}} \tag{1}$$

where $|\lambda_1|$ is the spectral gap of the single-qubit Lindblad generator $\mathcal{M}$, computable from T_1 and T_2:

$$|\lambda_1| = \min\left(\frac{1}{T_1},\ \frac{1}{2T_1} + \frac{1}{T_2} - \frac{1}{2T_1}\right) = \frac{1}{T_1} \tag{2}$$

and t_{eff} is the effective decoherence time per layer (gate time + idle time).

The critical depth where the noise-induced gradient suppression reaches a threshold ε:

$$\boxed{d^* = \frac{\log(1/\varepsilon) + n \log 2}{2\,|\lambda_1| \cdot t_{\text{eff}} \cdot n}} \tag{3}$$

---

2. Derivation

2.1 Gradient Variance Under Noise

For a parametrized unitary $U(\theta) = \prod_l U_l(\theta_l)$ with noise channel $\mathcal{N}_l$ after each layer:

$$\rho(\theta) = \mathcal{N}_d \circ U_d \circ \cdots \circ \mathcal{N}_1 \circ U_1[\rho_0] \tag{4}$$

The gradient of the cost $C(\theta) = \mathrm{Tr}(O\,\rho(\theta))$ with respect to parameter θ_k in layer l:

$$\frac{\partial C}{\partial \theta_k} = \mathrm{Tr}\left(O \cdot \mathcal{N}_d \circ U_d \circ \cdots \circ \left[\frac{\partial U_l}{\partial \theta_k}\right] \circ \cdots \circ \mathcal{N}_1 \circ U_1[\rho_0]\right) \tag{5}$$

2.2 Noise Channel Contraction

Each noise channel $\mathcal{N}_l = e^{\mathcal{M} \cdot t_{\text{eff}}}$ contracts the state toward the fixed point (thermal state). The contraction factor per layer:

$$\|\mathcal{N}_l[\rho] - \rho_{\text{th}}\| \le e^{-|\lambda_1| \cdot t_{\text{eff}}} \cdot \|\rho - \rho_{\text{th}}\| \tag{6}$$

After d layers, the total contraction is $e^{-|\lambda_1| \cdot d \cdot t_{\text{eff}}}$.

2.3 Gradient Variance Bound

The gradient involves the difference between the forward-propagated state and a rotated version. Each noise channel contracts this difference by $e^{-|\lambda_1| \cdot t_{\text{eff}}}$. Combined with the Haar-random expressibility suppression:

$$\mathrm{Var}_\theta[\partial_\theta C] \leq \frac{1}{2^n} \cdot e^{-2|\lambda_1| \cdot d \cdot t_{\text{eff}}} \cdot \|O\|^2 \tag{7}$$

The factor of 2 in the exponent comes from the squared norm (variance involves second moment of the gradient).

Setting this equal to ε and solving for d gives equation (3). □

---

3. The Numbers

3.1 Critical Depth Across Hardware Generations

Hardware	T_1 (μs)	$\lvert\lambda_1\rvert$ (Hz)	d^*(4q)	d^*(10q)	d^*(50q)	d^*(100q)
IBM Eagle (2024)	300	3,333	2,643	1,424	774	676
IBM Heron (2025)	500	2,000	4,405	2,374	1,290	1,127
Projected (2027)	1,000	1,000	8,811	4,748	2,580	2,254
Projected (2030)	5,000	200	44,054	23,742	12,900	11,270

Key insight: for 50-qubit systems on IBM Eagle, $d^* = 774$. QAOA at depth 500 is trainable. VQE for quantum chemistry (depth 1000+) is NOT.

3.2 Numerical Validation (Qiskit AerSimulator)

4-qubit hardware-efficient ansatz, $Z^{\otimes 4}$ observable, 50 random parameter samples per depth:

Depth	Var[∂C] (measured)	Var[∂C] (predicted)	Ratio
1	1.26×10^{-1}	6.24×10^{-2}	2.0
5	6.21×10^{-2}	6.21×10^{-2}	1.0
10	7.13×10^{-2}	6.18×10^{-2}	1.2
20	4.55×10^{-2}	6.11×10^{-2}	0.7
50	2.44×10^{-2}	5.91×10^{-2}	0.4

The measured and predicted values are within 2× for all depths. The decreasing measured variance at depth 50 (0.024 vs predicted 0.059) shows the expressibility plateau is the DOMINANT effect for these parameters — the noise contribution ($e^{-2\times 3333\times 50\times 170\text{ns}} = 0.94$, a 6% suppression) is barely noticeable because $d \ll d^*$.

3.3 When Noise Dominates

The noise-induced plateau dominates when $2|\lambda_1| \cdot d \cdot t_{\text{eff}} > n \log 2$, i.e., when:

$$d > \frac{n \log 2}{2|\lambda_1| \cdot t_{\text{eff}}} = d_{\text{crossover}} \tag{8}$$

n qubits	$d_{\text{crossover}}$ (Eagle)	$d_{\text{crossover}}$ (Heron)
4	2,438	4,063
10	6,095	10,159
50	30,476	50,793

For current hardware: $d_{\text{crossover}} \gg$ practical depths. **Noise-induced barren plateaus are NOT the bottleneck today.** But as hardware improves (longer T_1) and circuits get deeper (more qubits, harder problems), the crossover will be reached. The spectral gap provides the EXACT warning.

---

4. Applications

4.1 Circuit Design Rule

Before designing a variational circuit:

$$\textbf{IF} \quad d_{\text{planned}} > d^* = \frac{\log(10^4 \cdot 2^n)}{2|\lambda_1| \cdot t_{\text{eff}} \cdot n} : \quad \textbf{DON'T BUILD IT.}$$

This saves weeks of optimization time and thousands of dollars in quantum compute. The $|\lambda_1|$ computation takes 0.001 seconds on a laptop (cluster Lindblad, 64×64).

4.2 Hardware Ranking

The "trainability depth" $d^*(n)$ is a hardware quality metric:

$$\text{Trainability} = \frac{1}{|\lambda_1| \cdot t_{\text{eff}}} = \frac{T_1}{t_{\text{gate}}} \tag{9}$$

This is the **number of gates a qubit can execute before noise kills the gradient**. Higher is better. It combines T_1 (coherence) and t_{gate} (speed) into one number.

Hardware	Trainability	Rank
IBM Eagle	300μs/170ns = 1,765	3rd
IBM Heron	500μs/160ns = 3,125	2nd
Quantinuum H2 (trapped ion)	10s/200μs = 50,000	**1st**

Trapped ions win on trainability because T_1/t_{gate} is \$ 30×\$ better than superconducting.

4.3 Algorithm Selection

Given a hardware with trainability $T_1/t_{\text{gate}} = R$:

$$d^*(n) = \frac{\log(10^4 \cdot 2^n)}{2n/R} = \frac{R \cdot \log(10^4 \cdot 2^n)}{2n} \tag{10}$$

For $R = 1{,}765$ (Eagle), $n = 50$: $d^* = 774$. Which algorithms fit?

Algorithm	Typical depth	Fits under d^*?	Trainable?
QAOA $p = 10$	100–200	Yes	
VQE (H)	50–100	Yes	
VQE (FeMoCo)	2,000+	**No**	
Quantum ML (QCNN)	500–1,000	Marginal	

---

5. Consequences of the Trainability Bound

5.1 Superconducting Qubits Cannot Reach Variational Advantage at 50+ Qubits

For quantum advantage, the circuit must be deeper than classical simulation: $d_{\text{alg}} > \text{poly}(n)$. But it must also be shallower than the barren plateau: $d_{\text{alg}} < d^*$. The **quantum advantage window** exists iff $d^* > \text{poly}(n)$.

For $d_{\text{alg}} = n^2$ (typical VQE scaling):

$$R_{\min} = \frac{2n^3}{\log(10^4 \cdot 2^n)} \tag{11}$$

n	$R_{\min}$	IBM Eagle ($R = 1{,}765$)	Trapped ion ($R = 50{,}000$)
10	61		
20	340		
50	**5,800**	**Insufficient**	
100	**27,000**		

At 50+ qubits, IBM Eagle's trainability ratio is insufficient for variational quantum advantage. Only trapped ions ($R \approx 50{,}000$) or significantly improved superconducting hardware can reach the advantage window.

5.2 Error Mitigation Opens the Window

Spectral error mitigation effectively multiplies T_1 by a correction factor $\eta \approx 3$ (from 99.9% fidelity recovery), giving $R_{\text{mitigated}} = \eta \cdot R$:

Configuration	R	$d^*(50\text{q})$	VQE depth 2500 possible?
Eagle (raw)	1,765	774	**No**
Eagle + spectral mitigation	**5,295**	**2,322**	**Yes**
Heron (raw)	3,125	1,370	No
Heron + spectral mitigation	**9,375**	**4,110**	**Yes**
Trapped ion (raw)	50,000	21,900	Yes

Spectral error mitigation is the difference between possible and impossible quantum advantage on superconducting hardware.

5.3 The Optimal Qubit Count Is ~13, Not 100+

$d^*(n)$ peaks at small n and asymptotes to $R \cdot \log 2/2 \approx R/3$ for large n:

$$d^*(n) = \frac{R}{2n}[\log(10^4) + n\log 2] \xrightarrow{n \gg 13} \frac{R\log 2}{2} \approx \frac{R}{3}$$

The maximum d^* occurs at $n_{\text{opt}} \approx \log(10^4)/\log(2) \approx 13$. Beyond 13 qubits, adding more qubits DECREASES the trainable depth.

Implication: "more qubits = better" is wrong for variational algorithms. The trainability-optimal design uses ~13 very high-quality qubits, not 100+ noisy ones.

5.4 Minimum T_1 for Any Algorithm

From $d_{\text{alg}} < d^*$, the minimum coherence time for a specific algorithm:

$$T_{1,\text{min}} = \frac{2n \cdot d_{\text{alg}} \cdot t_{\text{gate}}}{\log(10^4 \cdot 2^n)} \tag{12}$$

Algorithm	n	d_{alg}	$T_{1,\text{min}}$	Current best T_1	Feasible?
QAOA $p = 10$	50	200	93 μs	300 μs	
VQE (H_2)	20	100	11 μs	300 μs	
VQE (FeMoCo)	100	2,000	**1.0 ms**	300 μs	
QCNN	50	500	232 μs	300 μs	Marginal

FeMoCo — the "holy grail" of quantum chemistry — requires T_1 > 1 ms, which is 3× beyond the current state of the art. With spectral error mitigation (3× effective T_1): $T_{1,\text{eff}} \approx 900\,\mu$s — still insufficient. FeMoCo needs hardware improvement OR a shallower algorithm.

5.5 Noise Is NOT the Current Bottleneck

The noise-induced barren plateau dominates when $2|\lambda_1| \cdot d \cdot t_{\text{eff}} > n \log 2$:

$$d_{\text{crossover}} = \frac{n \log 2}{2|\lambda_1| \cdot t_{\text{eff}}} \tag{13}$$

For IBM Eagle: $d_{\text{crossover}}(4\text{q}) = 2{,}438$, $d_{\text{crossover}}(50\text{q}) = 30{,}476$.

Current variational circuits operate at $d \ll d_{\text{crossover}}$. **The noise-induced barren plateau is NOT the bottleneck for today's hardware.** The expressibility plateau (from random initialization, $\text{Var} \sim 2^{-n}$) dominates.

Consequence: improving T_1 alone does NOT fix today's barren plateaus. What helps: structured initialization, problem-informed ansätze, and layer-wise training — none of which depend on $|\lambda_1|$.

The spectral formula predicts WHEN noise WILL become the bottleneck: as algorithms get deeper and hardware gets bigger, $d_{\text{crossover}}$ shrinks. For $n = 100$ on projected 2030 hardware: $d_{\text{crossover}} \approx$ 3,000, which IS within the range of serious quantum chemistry algorithms.

---

6. Connection to the Spectral Framework

6.1 The Unified Role of $|\lambda_1|$

The spectral gap of the Lindblad generator controls SEVEN quantities:

What	Formula	Paper
Qubit lifetime	$T_1 = 1/|\lambda_1|$	Standard
Market mixing	$\tau_{\text{mix}} = 1/|\lambda_1^{\text{FP}}|$	Nagy (2026a)
Orbital transfer	$\tau_{\text{transfer}} = 1/|\lambda_1^{\text{killed}}|$	Nagy (2026h)
Compression threshold	$\rho_Q = e^{|\lambda_1|/v_{LR} - 1}$	This series
Trainability depth	$d^* \propto 1/|\lambda_1|$	**This paper**
Error mitigation	$\rho_{\text{corr}} = e^{-\mathcal{M}t}\rho_{\text{meas}}$	This series

One number. Six applications. Five domains.

6.2 The USRT Connection

The barren plateau is a spectral phenomenon: the noise channel $\mathcal{N} = e^{\mathcal{M}t}$ kills high-frequency modes of the cost landscape. The gradient is a high-frequency observable (it measures local differences). The USRT predicts: when $\rho_Q < 1$, the high-frequency content is destroyed — exactly the barren plateau condition.

---

7. Limitations

1. **Upper bound, not exact.** Equation (7) is a BOUND on gradient variance, not an equality. The actual variance can be smaller (e.g., if the circuit has symmetries that concentrate the gradient). The bound is tight for hardware-efficient ansätze with Haar-random initialization.
2. **Markovian noise.** $1/f$ noise (dominant in superconducting qubits at low frequencies) is not captured by the Lindblad model. A correlated noise correction would modify $|\lambda_1|$ to an effective $|\lambda_1^{\text{eff}}|$.
3. **Single-qubit spectral gap.** We use the single-qubit $|\lambda_1|$ from T_1. For correlated noise across qubits, the MULTI-qubit spectral gap (from the cluster Lindblad) should be used. In our numerical tests, the single-qubit gap gives the correct prediction — but this may not hold for all architectures.
4. **The prediction is MOST useful at 50+ qubits**, where the noise-induced plateau dominates. For $n < 20$ on current hardware: the expressibility plateau is the bottleneck, and our formula adds little.

---

8. Conclusion

The trainability of a variational quantum circuit is determined before the first gate fires. The Lindblad spectral gap $|\lambda_1|$, computable in 0.001 seconds from a 64×64 matrix, gives the critical depth d^* beyond which optimization is futile. For IBM Eagle with 50 qubits: $d^* = 774$. For quantum chemistry on 100 qubits: $d^* = 676$ — too shallow for the hardest problems.

The formula is simple enough to be a DESIGN RULE:

$$\boxed{d^* = \frac{R \cdot \log(10^4 \cdot 2^n)}{2n}, \qquad R = \frac{T_1}{t_{\text{gate}}}}$$

If your circuit is deeper than d^*: don't run it. Redesign. Choose shallower ansatz, better hardware, or a different algorithm.

One number. Know before you go.

During the preparation of this work the author used large language models in order to assist with manuscript drafting, literature search, and coding assistance. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

References

- Cerezo, M., A. Sone, T. Volkoff, L. Cincio, and P. J. Coles (2021). Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications*. DOI: 10.1038/s41467-021-21728-w
- McClean, J. R., S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven (2018). Barren plateaus in quantum neural network training landscapes. *Nature Communications*. DOI: 10.1038/s41467-018-07090-4
- Nagy, T. (2026). Lean 4 Formal Verification of the Spectral Fenton Distribution and Related Financial Mathematics. *Working paper*.
- Nagy, T. (2026). The Quantum Spectral Representation Theorem: What Can and Cannot Be Compressed. *Working paper*.
- Wang, S., E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles (2021). Noise-induced barren plateaus in variational quantum algorithms. *Nature Communications*. DOI: 10.1038/s41467-021-27045-6

Appendix: Reproducibility

```
python3 examples/spectral_barren_plateau.py
```

Self-contained (NumPy + SciPy + Qiskit-Aer). Runtime: 25 seconds.