The Expressivity-Trainability Tradeoff

Variational Quantum Algorithms · Module 8 — Expressivity Vs Trainability

Node 8.5

Position in Knowledge Graph

System: Variational Quantum Algorithms **Structural Domain:** Expressivity vs Trainability **Node:** 8.5

This is the **central node** of Module 8 — the one that combines the expressivity machinery (8.1-8.3) and the trainability bounds (8.4) into a single quantitative statement: **a curve relating how much an ansatz can express to how easily it can be optimized**.

The tradeoff is not a heuristic anymore. It is a *theorem*. The Ragone et al. (2024) Lie-algebraic framework provides the cleanest version, and it says (paraphrasing):

> For an ansatz with dynamical Lie algebra dimension D, the gradient variance is at most $1/D$ and the dimension of the reachable manifold is at most D.

So **expressivity grows linearly with D, trainability shrinks linearly with $1/D$**, and the product (expressivity × trainability) is bounded.

This is the formal version of every intuition Module 4 and Module 8 have been building toward. This node states it precisely, walks through its implications, and discusses how to navigate the tradeoff in practice.

The Tradeoff In One Equation

For a parameterized circuit with dynamical Lie algebra $\mathfrak{g}$ of dimension D:

$$\underbrace{\dim \mathcal{R}(U) \leq D}_{\text{expressivity bound}} \quad \text{and} \quad \underbrace{\mathrm{Var}[\partial_j C] \leq \frac{P(\mathfrak{g}, H)}{D}}_{\text{trainability bound}}$$

where $P(\mathfrak{g}, H)$ is the cost-purity factor (a quantity that depends on how the cost Hamiltonian aligns with the algebra; typically $O(1)$).

Multiplying:

$$\dim \mathcal{R}(U) \cdot \mathrm{Var}[\partial_j C] \leq P(\mathfrak{g}, H).$$

This is the tradeoff. The product of expressivity (in the dimension sense) and trainability (in the gradient-variance sense) is bounded by a constant. You cannot have both large.

What The Tradeoff Says

A few interpretations:

1. There is no free lunch. Every gain in expressivity comes at a proportional cost in trainability. If you want to reach a larger family of states, you pay with smaller gradients.

2. The boundary is a curve. Plotting (expressivity, trainability) as a 2D point, all achievable ansätze lie below a hyperbola. The hyperbola is the **Pareto frontier** — the boundary where you can't improve one without sacrificing the other.

3. Most ansätze are far from the frontier. The bound is an inequality, not an equality. Many ansätze are strictly worse than the frontier — they have low expressivity *and* low trainability. Only a small subset are *Pareto optimal*.

4. Choosing where to land is the design decision. Given the frontier, the practitioner picks a point. High expressivity, low trainability (deep HEA on small problems where you can afford the optimization cost). Low expressivity, high trainability (UCCSD on chemistry where the symmetries do the work). Or somewhere in the middle.

The Ragone Curve, Explicitly

The Ragone et al. theorem provides an explicit curve.

For an ansatz indexed by a "structure parameter" s (e.g., depth, number of generators, allowed symmetries), as s increases:

- The DLA dimension $D(s)$ increases.
- The reachable manifold dimension grows: $\dim \mathcal{R}(s) \leq D(s)$.
- The gradient variance shrinks: $\mathrm{Var}(s) \leq P/D(s)$.

The achievable region in (expressivity, trainability) space is bounded by:

$$\text{trainability} \cdot \text{expressivity} \leq \text{const}.$$

For specific ansatz families, the curve has been computed. Examples:

Deep HEA family, parameterized by depth L: - $D(L)$ saturates at $4^n - 1$ (full unitary algebra) for $L = O(n)$. - Expressivity grows from polynomial to exponential. - Trainability drops from $O(1)$ to $O(2^{-n})$.

UCCSD family, parameterized by excitation order k: - $D(k)$ grows polynomially in k: $D \sim n^{2k}$. - Expressivity grows polynomially. - Trainability drops polynomially. - *Stays in the polynomial regime* — never hits the exponential plateau.

Symmetry-restricted ansätze, parameterized by the symmetry group G: - D is fixed by the dimension of the G-invariant operator algebra. - Both expressivity and trainability are determined by G, not directly tunable. - The "right" G is problem-specific.

Pareto Frontier And Optimal Ansätze

A Pareto-optimal ansatz is one where you cannot improve expressivity without losing trainability, or vice versa. The Pareto frontier is the curve of such ansätze.

Properties of the frontier:

1. **It is monotone:** moving along the frontier, gaining expressivity always loses trainability and vice versa. No "free improvements."

2. **It is convex** (in the appropriate scaling): the tradeoff has diminishing returns — going from low to medium expressivity costs little trainability, but going from medium to high expressivity costs proportionally more.
3. **Different problems have different optimal points:** the right point on the frontier depends on the cost function (specifically, how much expressivity is needed to represent the optimum).
4. **Most published ansätze are sub-optimal**, but the gap to the frontier is small in the cases studied so far.

Practical takeaway: when designing an ansatz, ask: "is this on the frontier, or below it?" If below, can it be improved (more efficient parameterization) to move it up?

Concrete Tradeoff Examples

The bounds in node 8.4 give numerical predictions you can check against this tradeoff.

Example 1: 6-qubit HEA, depth 1. - Expressivity (DLA dim): ~ 100. - Gradient variance (Ragone): ~ 0.01. - Verdict: low expressivity, moderate trainability. Probably below the frontier.

Example 2: 6-qubit HEA, depth 6. - Expressivity (DLA dim): ~ 4000 (full unitary algebra). - Gradient variance (Ragone): ~ 0.0003. - Verdict: maximum expressivity, minimum trainability. On the (exponential) frontier — barren plateau.

Example 3: 6-qubit UCCSD with 2 electrons. - Expressivity (DLA dim): ~ 35 (single + double excitations subspace). - Gradient variance (Ragone): ~ 0.03. - Verdict: structured, low expressivity, high trainability. On the (polynomial) frontier.

Example 4: 6-qubit Hamiltonian Variational Ansatz, 3 layers. - Expressivity (DLA dim): ~ 200 (problem-specific algebra). - Gradient variance: ~ 0.005. - Verdict: moderate expressivity, moderate trainability. Close to the frontier.

The pattern: **structured (problem-aligned) ansätze are on the frontier; generic ansätze are typically below it**.

Why The Tradeoff Is Tight

A natural question: is the tradeoff *fundamental* (a true theorem about quantum mechanics) or *artifact-of-our-bounds* (only as tight as the bounds, which could be loosened by smarter analysis)?

The honest answer: **partially fundamental, partially artifactual**.

Fundamental part: the relationship between Lie algebra dimension and reachable set is exact (it's a theorem of differential geometry). The relationship between Lie algebra dimension and gradient variance has the right scaling (also a theorem). The product is bounded by a constant that depends only on the cost Hamiltonian's purity.

Artifactual part: the constant $P(\mathfrak{g}, H)$ might be tightenable by smarter analysis. The gradient variance bound is an upper bound — empirically the variance can be smaller, which would make the tradeoff *worse* than predicted (more pessimistic). Lower bounds would be needed to show the tradeoff is tight in both directions.

The 2026 state of the art is that the tradeoff is **definitely valid** as a hyperbolic upper bound, and **probably tight up to constants** for generic ansätze. For special structured ansätze, there may be more room.

Implications For Ansatz Design

The tradeoff turns ansatz design into a constrained optimization problem:

$$\text{maximize trainability}(U) \quad \text{subject to expressivity}(U) \geq E_{\text{required}},$$

where E_{required} is the minimum expressivity needed to contain the answer to the problem at hand.

This formulation makes the design problem clean:

Step 1: Estimate E_{required} from the problem. For chemistry, this is "enough excitations to capture electron correlations." For optimization, "enough expressivity to express the optimal mixing operator." For machine learning, "enough Fourier modes for the data."

Step 2: Find the ansatz family with smallest possible expressivity that still meets E_{required}.

Step 3: Among that family, pick the variant with the largest trainability.

Step 4: Verify empirically.

This is what UCCSD does for chemistry, what QAOA does for optimization, what re-uploading does for machine learning. They are all different *operational embodiments* of the same constrained optimization principle.

Connection To The Thesis

The thesis frames the tradeoff in terms of **regularization**. The argument:

The expressivity-trainability tradeoff says: an unrestricted (highly expressive) ansatz is untrainable. To make it trainable, you must *restrict* it. There are two ways to restrict:

1. Restrict the ansatz — fewer parameters, smaller DLA, more symmetries. This is the *static* restriction, baked into the ansatz design.

2. Restrict the optimizer dynamics — population-based methods, regularized updates, momentum, trust regions. This is the *dynamic* restriction, baked into the optimizer.

The thesis claims **both** are forms of regularization, and the dynamic version is underexploited. Module 5 (regularization) and Module 6 (optimization dynamics) cover the dynamic version; this module covers the static version.

The key insight: **both forms of restriction move you to a different point on the expressivity-trainability tradeoff curve**. Static restriction trades expressivity at design time; dynamic restriction trades effective expressivity at optimization time. They are complementary.

What The Tradeoff Doesn't Say

A few things the tradeoff curve does *not* tell you:

1. It doesn't tell you where on the curve you should be. That depends on the problem.

2. It doesn't tell you which structured ansatz is best. Many ansätze can be at the same point on the curve, but with different fitness for different problems.

3. It doesn't tell you about local minima. The tradeoff is about *global* trainability properties (gradient variance). It doesn't address whether the landscape has many local minima.

4. It doesn't tell you about noise. Hardware noise modifies the effective tradeoff (Wang's bound) but the noise correction is on top of the unitary tradeoff.

5. It doesn't predict generalization. For VQA-as-machine-learning, there's a separate question about how well the ansatz generalizes from training data, which the tradeoff doesn't address.

These are all separate questions that nodes 8.6 (data encoding) and 8.7 (quantum advantage) take up.

Structural Role

This node is the **central theorem** of Module 8. It combines the expressivity machinery (8.1-8.3) with the trainability bounds (8.4) to produce the explicit tradeoff curve, and it provides the practical workflow for designing ansätze in light of the tradeoff.

It is the conceptual climax of Module 8, and the main reference point for Module 9 (Hardware Noise) and Module 10 (Adaptive Control), which discuss how the tradeoff plays out in real deployment.

Relations to Other Concepts

- **Ragone et al. (2024)** — the primary reference for the explicit tradeoff.
- **Holmes et al. (2022)** — earlier connection between expressibility and trainability.
- **Sim-Johnson-Killoran (2019)** — the empirical version of the same intuition.
- **No-free-lunch theorems (Wolpert-Macready 1997)** — classical analogue from optimization.
- **Module 4 (Barren plateaus)** — the qualitative version.

Worked Example — The Tradeoff For A Family Of Ansätze

Consider 6-qubit ansätze with single and double Pauli rotations, parameterized by the number of layers L.

L	DLA dimension	Expressivity (manifold dim)	Variance bound (P/D)	Where on curve
1	24	24	0.04	Below frontier
2	100	100	0.01	Approaching frontier
3	400	400	0.0025	On frontier
4	1500	~1500	0.0007	On frontier
5	3500	3500	0.0003	Saturated
6+	4096	4096 (max)	0.00024	Plateau regime

The product (expressivity × variance) hovers around 1 for layers 3-6, confirming the tradeoff. Below $L = 3$, the product is below 1 — the ansatz is sub-optimally compressing.

To use this ansatz for a problem requiring expressivity ~400 (manifold dimension), $L = 3$ is the right choice. Going to $L = 5$ or $L = 6$ gains nothing (the ansatz can already represent the answer) but loses an order of magnitude in trainability.

This is the kind of analysis the tradeoff curve enables.

Visualization

Pending. A 2D plot of trainability (y-axis, log scale) vs expressivity (x-axis, log scale) with the Pareto frontier curve, several ansatz points (UCCSD, HEA-1, HEA-3, HEA-deep, problem-aligned), and the "below frontier"

region shaded. Caption: "The tradeoff is real. Most ansätze are on the curve. The right point depends on your problem."

Exercises

1. For an ansatz with DLA dimension 50 and cost-purity 1, what is the maximum gradient variance? Is this trainable at 1000 shots?
2. The tradeoff says (expr $\times$ var) $\leq$ const. For two different cost Hamiltonians (one global, one local), the constants differ. Which has the larger constant — global or local? Explain.
3. (VQA) Design a 4-qubit ansatz with target expressivity (manifold dimension) ~20 and predict its gradient variance from the tradeoff. Verify by computing the DLA dimension explicitly.

Research Extensions

- **Tighter constants in the tradeoff** — pinning down $P(\mathfrak{g}, H)$ for various ansatz-Hamiltonian combinations.
- **Tradeoff under noise** — combining Ragone with Wang to get the effective tradeoff under hardware noise.
- **Tradeoff for non-standard cost functions** — non-Hermitian observables, multi-objective costs.
- **Tradeoff and Pareto-optimal ansatz search** — automated discovery of ansätze on the frontier.

Cross-links to Other Courses

- **Multi-objective optimization** — Pareto frontiers in classical optimization.
- **Information theory** — capacity-rate tradeoffs (Shannon's noisy channel theorem is the classical analogue).
- **Module 4 (Barren plateaus)** — the qualitative root of the tradeoff.
- **Module 5 (Regularization)** — the dynamics-side complement.

Variational Quantum Algorithms · Module 8 — Expressivity Vs Trainability · Node 8.5
Concepts: barren-plateaus, dimensionality, parameterized-circuits, optimization-landscape
Prerequisites: 8.1, 8.2, 8.4
Enables: 8.6, 8.7

hbar.university — own.your.cognition