



# FROM SMOOTH VALLEYS TO BARREN PLATEAUS:

THE HIDDEN GEOMETRY OF QUANTUM OPTIMIZATION

Dr. Van Binh Henri VU

Quantum Computing Lab, IT4Innovations National Supercomputing Center

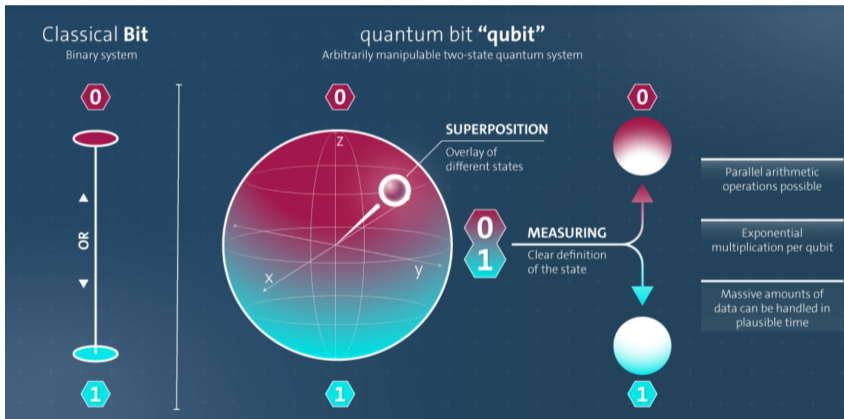
05/19/2026



# Motivation

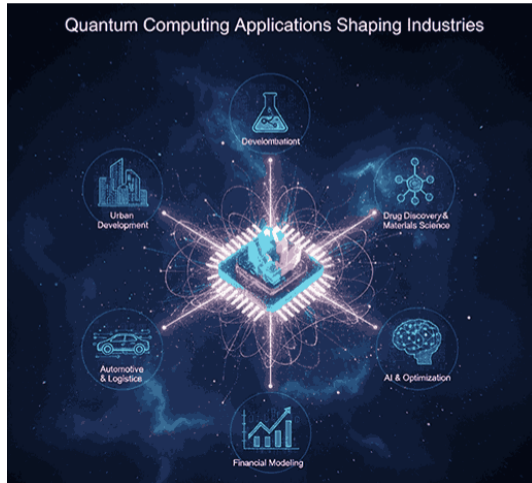
# Why do we need quantum computing ?

## Motivation



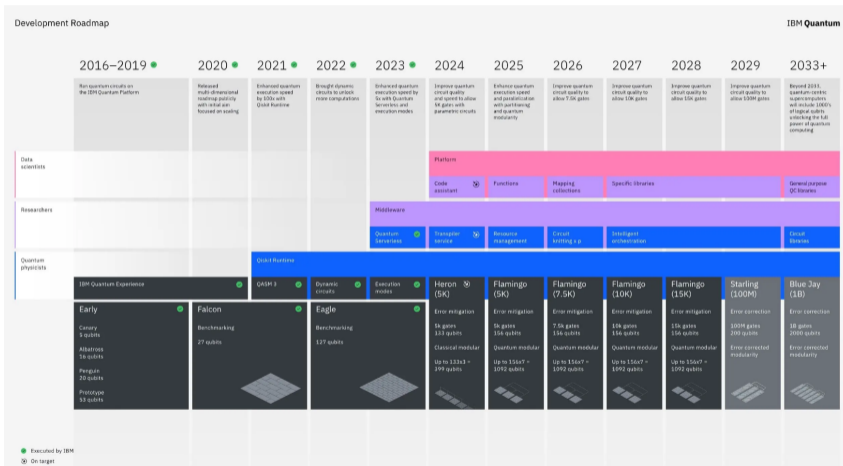
# Quantum computing will not replace entirely classical computing

Motivation



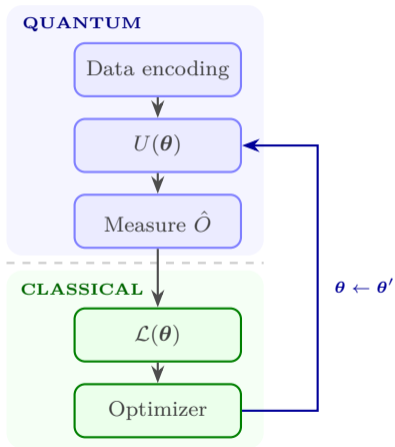
# Today, Quantum Hardware is noisy

## Motivation



# Hybrid Quantum Classical Computing

Motivation



- Using Variational Quantum Algorithms (VQA)
- Parameterized circuit  $U(\theta)$  prepares quantum state [5]
- Cost evaluated via measurement
- Classical optimizer updates  $\theta$
- Leading paradigm for near-term quantum devices

# The Trainability Crisis

## Motivation

VQA success depends on the structure of the cost-function landscape, but current practice offers limited tools to assess it:

- **Trial and error:** practitioners run optimization and hope for convergence, with no guarantee beforehand
- **Gradient variance:** the standard diagnostic [14] detects barren plateaus, but says nothing about local minima, saddle points, or rugged curvature
- **No dynamical insight:** knowing that gradients exist does not tell us whether the optimizer can actually follow them to a good solution

**Gap:** No unified methodology exists to characterize *both* the geometry and the dynamics of VQA landscapes *before* optimization

## Research Question

Can the intrinsic geometric and dynamical properties of a VQA landscape predict trainability *a priori*?

In the other word, *Can we look at the mathematical shape and chaos of a quantum circuit's landscape before training it, and know immediately if the algorithm is going to be able to learn?*



IT4I

# Approach

# Spectral Hessian Analysis

## Approach

The Hessian encodes local curvature of  $\mathcal{L}(\theta)$ :

$$H_{ij} = \frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_j}$$

Its eigenvalues  $\{\lambda_k^{\text{Hess}}\}$  describe principal directions:

- $\forall k, \lambda_k^{\text{Hess}} > 0$ : Local minimum
- $\forall k, \lambda_k^{\text{Hess}} < 0$ : Local maximum
- $\exists i, j : \lambda_i^{\text{Hess}} < 0 < \lambda_j^{\text{Hess}}$ : Saddle point
- $\forall k, \lambda_k^{\text{Hess}} \approx 0$ : flat, plateau direction

### Key statistics:

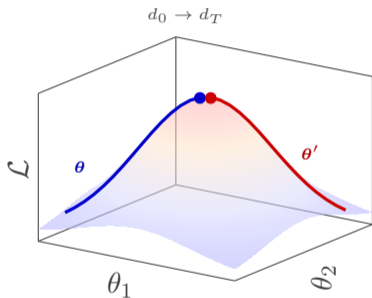
- $\sigma_{\lambda^{\text{Hess}}}^2$ : spectral variance – anisotropy of the landscape
- $\lambda_{\max}^{\text{Hess}}$ : steepest direction; bounds stable learning rate  $\eta < 2/\lambda_{\max}^{\text{Hess}}$  [7]
- Fraction  $\lambda_k^{\text{Hess}} < 0$ : saddle-point density

### Dimensions:

- $p \times p$  where  $p$  is number parameters of ansatz

# Lyapunov Exponents

## Approach



Two nearby trajectories diverge at rate [8]:

$$\lambda^{\text{Lyap}} \approx \frac{1}{T} \sum_{t=0}^{T-1} \ln \frac{d_{t+1}}{d_t}$$

where  $d_t = \|\theta_t - \theta'_t\|$ .

- $\lambda^{\text{Lyap}} > 0$ : **chaotic** divergence
- $\lambda^{\text{Lyap}} \approx 0$ : marginally stable
- $\lambda^{\text{Lyap}} < 0$ : **stable** convergence

**Dimensions:**

- $N_L$  : number of paths

# The Geometry-Dynamics Correspondence

## Approach

Near a critical point, linearization of gradient descent reveals that the Hessian spectrum directly governs trajectory stability [18]:

$$\lambda_{\max}^{\text{Lyap}} \approx -\eta \lambda_{\min}^{\text{Hess}}$$

A saddle direction with  $\lambda_k^{\text{Hess}} < 0$  produces exponential instability at rate  $|\lambda_k^{\text{Hess}}|$ , so  $\lambda^{\text{Lyap}} \approx \eta |\lambda_k^{\text{Hess}}|$  for small learning rate  $\eta$ .

### Why this matters:

- Studied in classical neural networks [18], but *never tested in quantum optimization landscapes*
- If the correspondence holds for VQAs, a single Hessian computation can predict dynamical trainability without running optimization

# How to detect global phenomena?

Approach

**Local  $\rightarrow$  Global:**

- We calculate at 20 different locations

- 

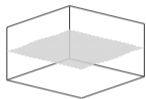
$\lambda \rightarrow \sigma^2$  (Spectral variance) &  $\bar{\lambda}$  (Spectral mean)



# Our Target

# Four Landscape Regimes

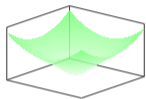
Our Target



**Barren Plateau**

$$\sigma^2 \approx 0, \bar{\lambda}^{\text{Lyap}} \approx 0$$

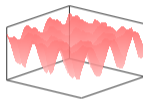
gradient descent fails



**Sharp Convergent**

$$\sigma^2 \uparrow, \bar{\lambda}^{\text{Lyap}} < 0$$

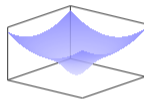
optimal for GD



**Rugged & Chaotic**

$$\sigma^2 \uparrow, \bar{\lambda}^{\text{Lyap}} > 0$$

GD unreliable



**Smooth & Convex**

$$\sigma^2 \downarrow, \bar{\lambda}^{\text{Lyap}} > 0$$

trainable, slow GD

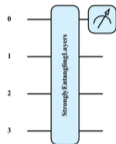
Central hypothesis: **Hessian spectrum predicts  $\lambda^{\text{Lyap}}$  sign and magnitude *a priori***



IT4I

# Results

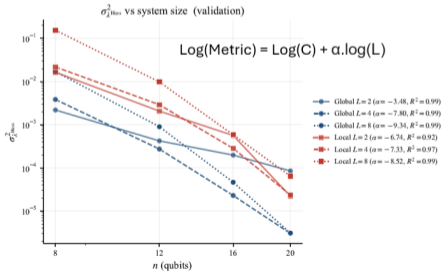
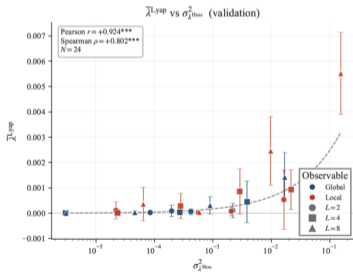
# Validation Experiment Results



Local observable



Global observable



# Real cases

## Results

*Can  $\lambda^{\text{Hess}}$  spectra predict  $\lambda^{\text{Lyap}}$  without running optimization?*

Experiment	Pearson $r$	Sig.	Dominant predictor
Validation (local obs.)	+0.92	***	$\sigma_{\lambda^{\text{Hess}}}^2$ and $\lambda_{\text{max}}^{\text{Hess}}$ both reliable
Quantum Chemistry	+0.90	***	molecule identity dominates; $\lambda_{\text{max}}^{\text{Hess}}$ most reliable
Graph Coloring	+0.81	***	$\lambda_{\text{max}}^{\text{Hess}}$ ; all configs divergent
QML (Moons)	+0.64	**	$\lambda_{\text{max}}^{\text{Hess}}$ ; regime driven by $n$
MaxCut QAOA	+0.53	***	$\lambda_{\text{max}}^{\text{Hess}}$ stronger than $\sigma^2$
QML (Iris)	+0.37	*	saddle density additionally predictive

**Answer: yes, in 6 of 7 cases.**

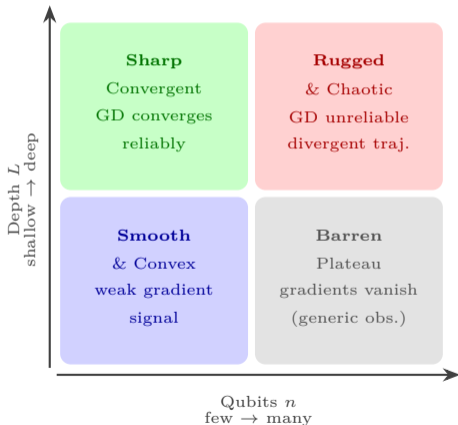
$\lambda_{\text{max}}^{\text{Hess}}$  is the most consistent single predictor of  $\lambda^{\text{Lyap}}$  across all domains.



# Summary

# Design Intuition: Navigating the Landscape

## Summary



## Three rules of thumb:

1. **More qubits**  $\rightarrow$  flatter landscape  
*unless* your Hamiltonian energy scale grows with  $n$  (e.g. MaxCut shifts the boundary right)
2. **More depth**  $\rightarrow$  sharper curvature  
universal, but too deep pushes into Rugged & Chaotic
3. **Hamiltonian**  $>$  **ansatz**  
same circuit on different cost functions  
= completely different regime

## The payoff:

a single Hessian measurement at a random point classifies the regime *before* any optimization is run.

# References

## References

- [1] E. R. Anschuetz and B. T. Kiani. “Quantum variational algorithms are swamped with traps”. In: *Nature Communications* 13.1 (2022), p. 7760. DOI: 10.1038/s41467-022-35364-5.
- [2] A. Asadi et al. “Hybrid quantum programming with PennyLane Lightning on HPC platforms”. In: *arXiv preprint arXiv:2403.02512* (2024).
- [3] V. Bergholm et al. “PennyLane: Automatic differentiation of hybrid quantum-classical computations”. In: *arXiv preprint arXiv:1811.04968* (2018).
- [4] M. Cerezo et al. “Cost function dependent barren plateaus in shallow parametrized quantum circuits”. In: *Nature Communications* 12.1 (2021), p. 1791. DOI: 10.1038/s41467-021-21728-w.
- [5] M. Cerezo et al. “Variational quantum algorithms”. In: *Nature Reviews Physics* 3.9 (2021), pp. 625–644. DOI: 10.1038/s42254-021-00348-9.
- [6] D. Chemnitz and M. Engel. “Characterizing Dynamical Stability of Stochastic Gradient Descent in Overparameterized Learning”. In: *arXiv preprint arXiv:2407.20209* (2025).
- [7] J. M. Cohen et al. “Gradient descent on neural networks typically occurs at the edge of stability”. In: *arXiv preprint arXiv:2103.00065* (2021).
- [8] J.-P. Eckmann and D. Ruelle. “Ergodic theory of chaos and strange attractors”. In: *Reviews of Modern Physics* 57.3 (1985), pp. 617–656. DOI: 10.1103/RevModPhys.57.617.
- [9] E. Farhi, J. Goldstone, and S. Gutmann. “A quantum approximate optimization algorithm”. In: *arXiv preprint arXiv:1411.4028* (2014).

- [10] E. Fontana et al. “Beyond unital noise in variational quantum algorithms: noise-induced barren plateaus and limit sets”. In: *Quantum* 9 (2025), p. 1617. DOI: 10.22331/q-2025-01-30-1617.
- [11] E. Grant et al. “An initialization strategy for addressing barren plateaus in parametrized quantum circuits”. In: *Quantum* 3 (2019), p. 214. DOI: 10.22331/q-2019-12-09-214.
- [12] M. Larocca et al. “Barren plateaus in variational quantum computing”. In: *Nature Reviews Physics* 7.4 (2025), pp. 174–189. DOI: 10.1038/s42254-025-00813-9.
- [13] C. Liu, L. Zhu, and M. Belkin. “Loss landscapes and optimization in over-parameterized non-linear systems and neural networks”. In: *Applied and Computational Harmonic Analysis* 59 (2022), pp. 85–116. DOI: 10.1016/j.acha.2021.12.009.
- [14] J. R. McClean et al. “Barren plateaus in quantum neural network training landscapes”. In: *Nature Communications* 9.1 (2018), p. 4812. DOI: 10.1038/s41467-018-07090-4.
- [15] A. Mürwald. “Integrating curvature into the cost landscape of variational quantum algorithms”. PhD thesis. University of Stuttgart, 2025.
- [16] A. Pérez-Salinas et al. “Data re-uploading for a universal quantum classifier”. In: *Quantum* 4 (2020), p. 226. DOI: 10.22331/q-2020-02-06-226.
- [17] A. Peruzzo et al. “A variational eigenvalue solver on a photonic quantum processor”. In: *Nature Communications* 5.1 (2014), p. 4213. DOI: 10.1038/ncomms5213.
- [18] M. Sasdelli et al. “A Chaos Theory Approach to Understand Neural Network Optimization”. In: (2020).



# References

## References

- [19] A. Uvarov and J. Biamonte. “On barren plateaus and cost function locality in variational quantum algorithms”. In: *Journal of Physics A: Mathematical and Theoretical* 54.24 (2021), p. 245301. DOI: [10.1088/1751-8121/abfac7](https://doi.org/10.1088/1751-8121/abfac7).
- [20] S. Wang et al. “Noise-induced barren plateaus in variational quantum algorithms”. In: *Nature Communications* 12.1 (2021), p. 6961. DOI: [10.1038/s41467-021-27045-6](https://doi.org/10.1038/s41467-021-27045-6).



Any Question?

# What this thesis establishes

## Summary

- A unified diagnostic framework combining  $\lambda^{\text{Hess}}$  spectra and  $\lambda^{\text{Lyap}}$  for VQA landscapes across six problem domains
- The H–L correspondence [18] extends to quantum optimization in 6 of 7 tested settings ( $r$  from +0.37 to +0.92)
- $\lambda_{\text{max}}^{\text{Hess}}$  is the most reliable single predictor of dynamical trainability *a priori*
- Four landscape regimes (Barren Plateau, Sharp Convergent, Smooth/Convex, Rugged & Chaotic) are empirically distinguishable from a single Hessian measurement
- **Cost function structure, not circuit architecture, is the dominant landscape determinant**

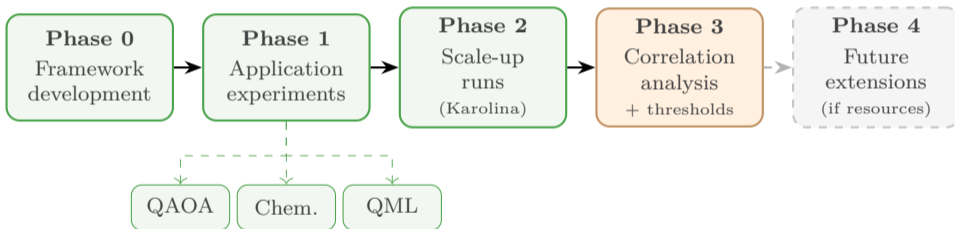
# Principled limits identified

## Summary

- The H–L link fails when Hamiltonian structure (attractive basins, hard constraints) dominates local curvature – community detection is the principled exception
- Curvature *magnitude* predicts dynamics; eigenvalue *sign* (saddle density) does not
- Depth drives chaoticity universally; qubit scaling depends on cost function – growing for MaxCut, decaying for generic circuits and graph coloring
- Molecule identity in VQE dominates over circuit depth – problem encoding, not ansatz depth, sets the chemistry landscape
- Spectral saturation at large  $n$ , deep  $L$ : Hessian and Lyapunov decouple in a *nonlinearity-driven* regime

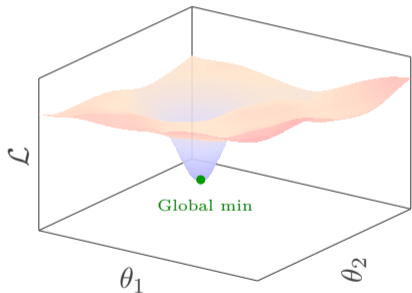
# Roadmap

## Summary



# Landscape Structure: The Trainable Regime

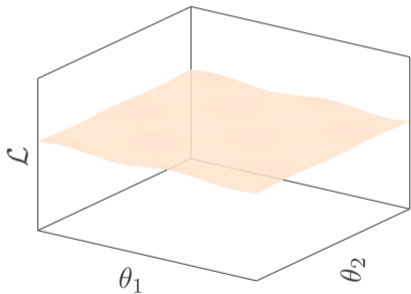
Summary



- Gradient variance vanishes as  $\mathcal{O}(2^{-n})$  for global observables [14, 19]
- Ansatz approaches unitary 2-design
- Local observables: polynomial scaling [4]

# Barren Plateaus: The Flat Regime

Summary

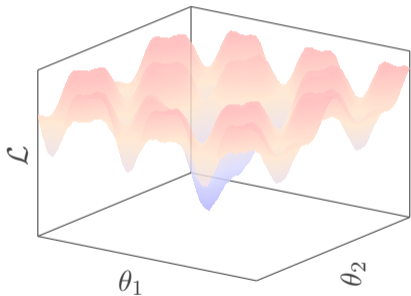


As system size  $n$  increases:

- Landscape flattens exponentially
- Optimizer receives no directional information
- Exponentially many evaluations needed to find structure

# Local Minima Proliferation

## Summary



- Shallow circuits avoid barren plateaus [1]
- But minima concentrate far from the global optimum
- Trainable regime: narrow corridor between flatness and traps



# Infrastructure

## Software:

- PennyLane + Lightning-GPU [3]
- JAX automatic differentiation
- Finite-difference Hessian (GPU)
- Lyapunov tracking via gradient descent

## Hardware (Karolina IT4I):

- NVIDIA A100 (40 GB HBM2)
- NVLink multi-GPU pooling



# MPI State-Vector Partitioning

## Infrastructure

### The memory bottleneck:

At  $n$  qubits, the state vector holds  $2^n$  complex amplitudes:

$$\text{Memory} = 2^n \times 16 \text{ B} \quad (\text{complex128})$$

### Solution: distributed partitioning

Lightning-GPU compiled with MPI support partitions the  $2^n$  amplitudes evenly across  $R$  GPU ranks via NCCL/CUDA-aware MPI collectives.

#### Execution difference from single-GPU:

JAX transforms (`jax.grad`, `jax.jit`) bypass MPI lockstep; all circuit calls use `qml.execute(..., diff_method=None)` directly, with gradients reconstructed via central finite differences

## Q2: Identify Landscape Regimes *A Priori*

Infrastructure

*Do four landscape regimes appear, and can they be predicted from the Hessian alone?*

### All four regimes observed:

- **Barren Plateau:** MaxCut  $L=1$ ; global obs.  
 $n=20$ ,  $L=8$   
 $\sigma^2 \rightarrow 0$ ,  $\lambda^{\text{Lyap}} \leq 0$
- **Sharp Convergent:** MaxCut  $L=2-3$ ;  
 chemistry (LiH, H<sub>2</sub>O)  
 high  $\sigma^2$ ,  $\lambda^{\text{Lyap}} < 0$ ; optimal for gradient descent
- **Smooth/Convex:** validation (global obs., shallow  $L$ ); QML at small  $n$   
 low  $\sigma^2$ ,  $\lambda^{\text{Lyap}} > 0$
- **Rugged/Chaotic:** graph coloring (all configs); HF; MaxCut  $L=4$   
 high  $\sigma^2$ ,  $\lambda^{\text{Lyap}} > 0$ ; gradient descent unreliable

### Predictive thresholds:

- Regime correctly classified in 6 of 7 experiments via  $\sigma^2$  median threshold and  $\lambda^{\text{Lyap}}$  sign
- $\sigma_{\lambda^{\text{Hess}}}^2 \propto n^{-\alpha}$  decay identifies BP onset:  $\alpha$  steepens from 3.5 ( $L=2$ ) to 9.3 ( $L=8$ ) for global obs.
- QML recovers  $\sim n^{-1}$  scaling – at the boundary between exponential and polynomial suppression

**Answer: yes.** Static Hessian computation suffices to classify the regime before optimization begins in the majority of settings.

# Q3: Depth and Qubit Count Drive Transitions

Infrastructure

*Where are the phase boundaries, and what controls them?*

## Depth ( $L$ ) effects – universal sharpening:

- $\sigma_{\lambda_{\text{Hess}}}^2$  grows as power law in  $L$  across *all* domains; exponents from +0.31 (H<sub>2</sub>O) to +1.83 (MaxCut  $n=6$ )
- **Convergent** → **chaotic transition in MaxCut**:  $\lambda^{\text{Lyap}}$  crosses from  $-0.47$  ( $L=1$ ) to  $+0.07$  ( $L=4$ ) at  $n=16$ ; transition is depth-driven and approximately  $n$ -independent
- Chemistry and community detection: depth modulates intensity, but *never causes regime change* – problem structure locks the regime

**Answer:** depth drives chaoticity universally; qubit scaling depends on cost function structure. Problem Hamiltonians that grow with  $n$  invert the barren plateau scaling.

## Qubit ( $n$ ) effects – two opposing regimes:

- **Decaying**  $\sigma^2$ : generic circuits (validation  $\alpha$  up to  $-9.3$ ), graph coloring (one-hot dilution,  $\alpha \approx -1.2$  to  $-7$ ), QML ( $\sim n^{-1}$ )
- **Growing**  $\sigma^2$ : MaxCut ( $\sigma^2 \propto n^{+1.9}$  at  $L=1$ , near-quadratic from Hamiltonian energy scale  $|E| \propto n^2$ ); community detection similar
- MaxCut  $L=1$ : larger graphs are *more* convergent – shallow QAOA on large random graphs has stronger basin attraction

## Q4: H–L Universal or Problem-Specific Infrastructure

*Does the geometry-dynamics correspondence extend beyond synthetic validation?*

### Where it generalizes:

- **Combinatorial optimization** (MaxCut, graph coloring): strong H–L link; both  $\sigma^2$  and  $\lambda_{\max}^{\text{Hess}}$  predictive
- **Quantum chemistry** ( $\text{H}_2$ , LiH,  $\text{H}_2\text{O}$ , HF): very strong; molecule identity is the dominant factor, not circuit depth
- **QML** (Moons, Iris): significant even for data-driven cost functions;  $\lambda_{\max}^{\text{Hess}}$  most reliable

### Where it fails:

- **Community detection**: H–L correlation  $r = -0.13$  (ns)  
Attractive basin structure in the modularity landscape overrides local curvature geometry; 83% of configurations remain convergent even at  $L=4$
- **Saddle density** (fraction  $\lambda_k^{\text{Hess}} < 0$ ): non-predictive in most experiments  
Curvature *magnitude* ( $\sigma^2$ ,  $\lambda_{\max}^{\text{Hess}}$ ) matters; eigenvalue *sign* does not

**Answer:** broadly universal, with one principled exception.

The link breaks when the Hamiltonian imposes strong attractive structure that dominates local geometry.

# Spectral formulas

Infrastructure

$$\bar{\lambda}^{Hess} = \frac{1}{p} \sum^p \lambda_k^{Hess}$$

$$\bar{\lambda}^{Lyap} = \frac{1}{N_L} \sum^{N_L} \lambda_k^{Lyap}$$

$$\sigma_{\lambda^{Hess}}^2 = \frac{1}{p} \sum^p (\lambda_k^{Hess} - \bar{\lambda}^{Hess})^2$$

$$\sigma_{\lambda^{Lyap}}^2 = \frac{1}{N_L} \sum^p (\lambda_k^{Lyap} - \bar{\lambda}^{Lyap})^2$$

# Key parameters

Infrastructure

- Finite-Difference:
  - $T = 100$
  - $\delta = 10^{-5}$
- Lyapunov :  $\eta = 0.01$

### Our goals:

1. Combine Hessian spectral analysis (curvature) with Lyapunov exponents (dynamics) into a single diagnostic framework
2. Apply it systematically across VQA domains: combinatorial optimization, quantum chemistry, and QML
3. Identify quantitative thresholds that separate trainable from untrainable configurations *before* running optimization