Ciência de Dados Quântica 2021/22

Variational Quantum Circuits:
barren plateaus

LUÍS PAULO SANTOS

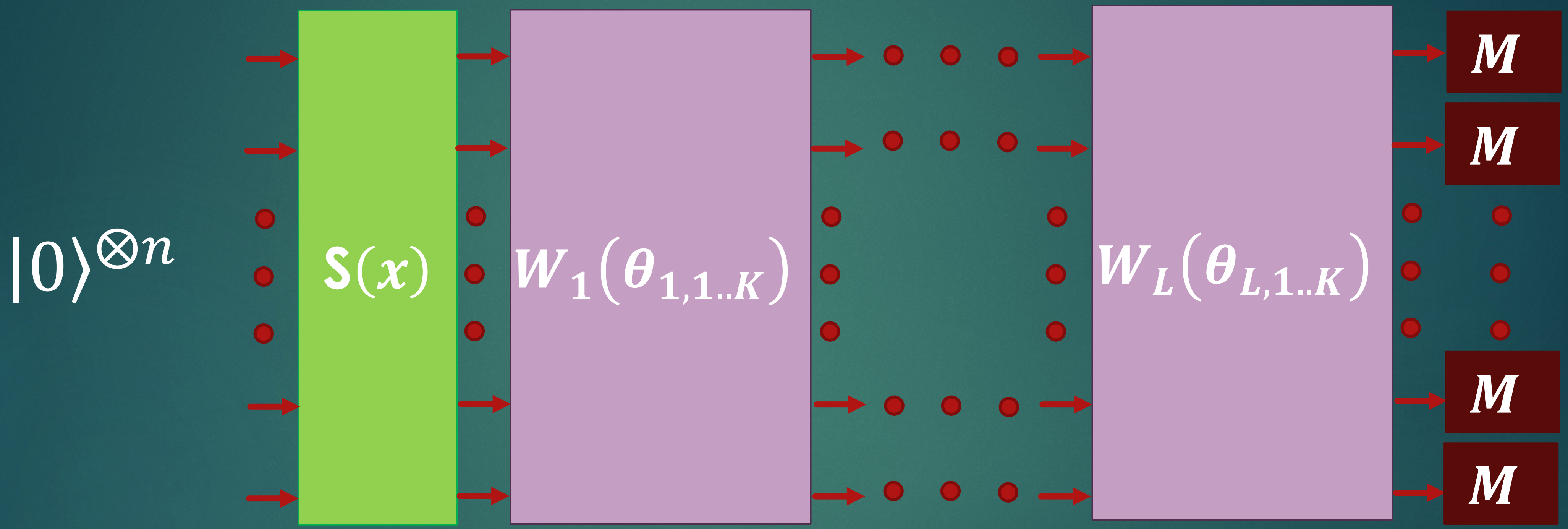
Material de Consulta

- [Schuld2021] – Sec. 5.3.3

- [Tacchino2021] – 2021 Qiskit Global Summer School on Quantum Machine Learning:
  Barren Plateaus, Trainability Issues and How to Avoid Them
  https://learn.qiskit.org/summer-school/2021/lec8-2-barren-plateaus-trainability-issues-how-avoid-them

- [Qiskit] –Qiskit Quantum Machine Learning Course:
  Training parameterized quantum circuits
  https://learn.qiskit.org/course/machine-learning/training-quantum-circuits

Parameterized quantum circuits



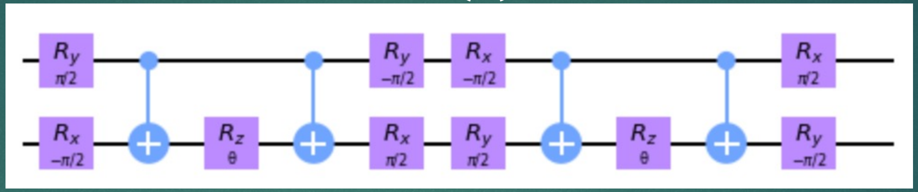
$$\hat{y}(x) = \langle 0 | U^{\dagger}(x, \vec{\theta})\, M\, U(x, \vec{\theta}) | 0 \rangle$$

$$U(x, \vec{\theta}) = \left[\prod_{l=1}^{L} W_l(\vec{\theta}_l) \right] S(x) = \mathcal{W}(\vec{\theta}) S(x)$$

The *ansatz*

PROBLEM

$\mathcal{W}(x)$

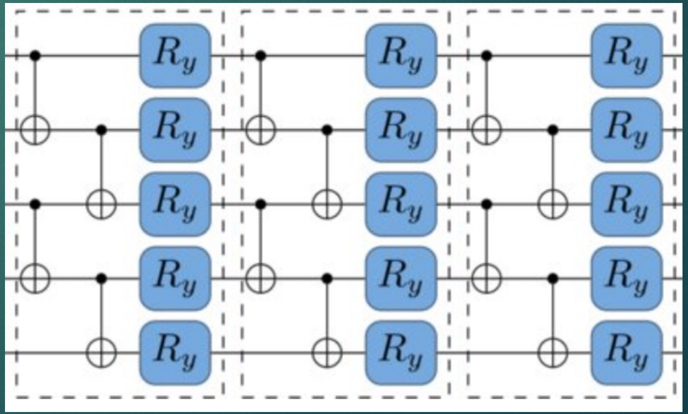


SOLUTION

Problem Specific

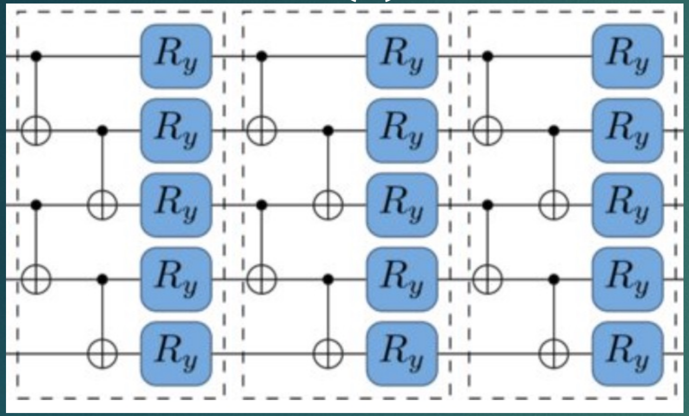
Example: Chemistry ground state problems

Problem Agnostic / Hardware Efficient



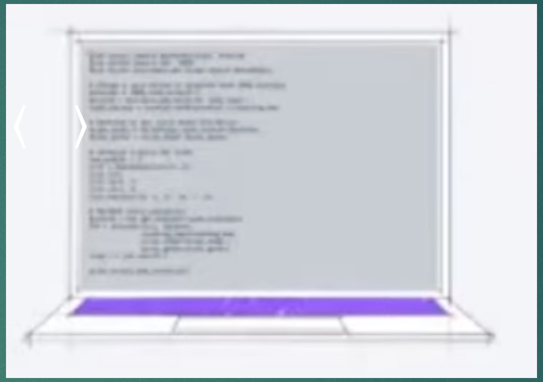
Vanishing Gradients / Barren plateaus

$\mathcal{W}(x)$



$\xrightarrow{\langle O \rangle_{\theta^t}}$

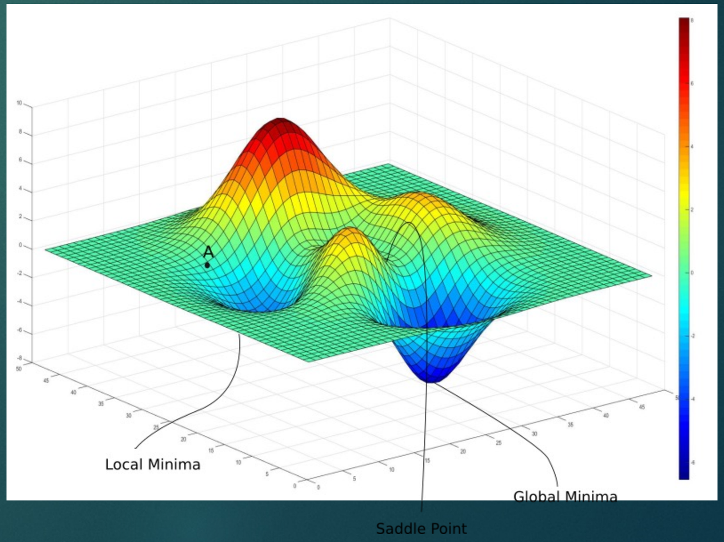
$\xleftarrow{\theta^{t+1}}$



Classical Optimization

$$\theta^{t+1} = f\left(C(\theta^t)\right)$$

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\left(C(\theta)\right)$$



Vanishing Gradients / Barren plateaus

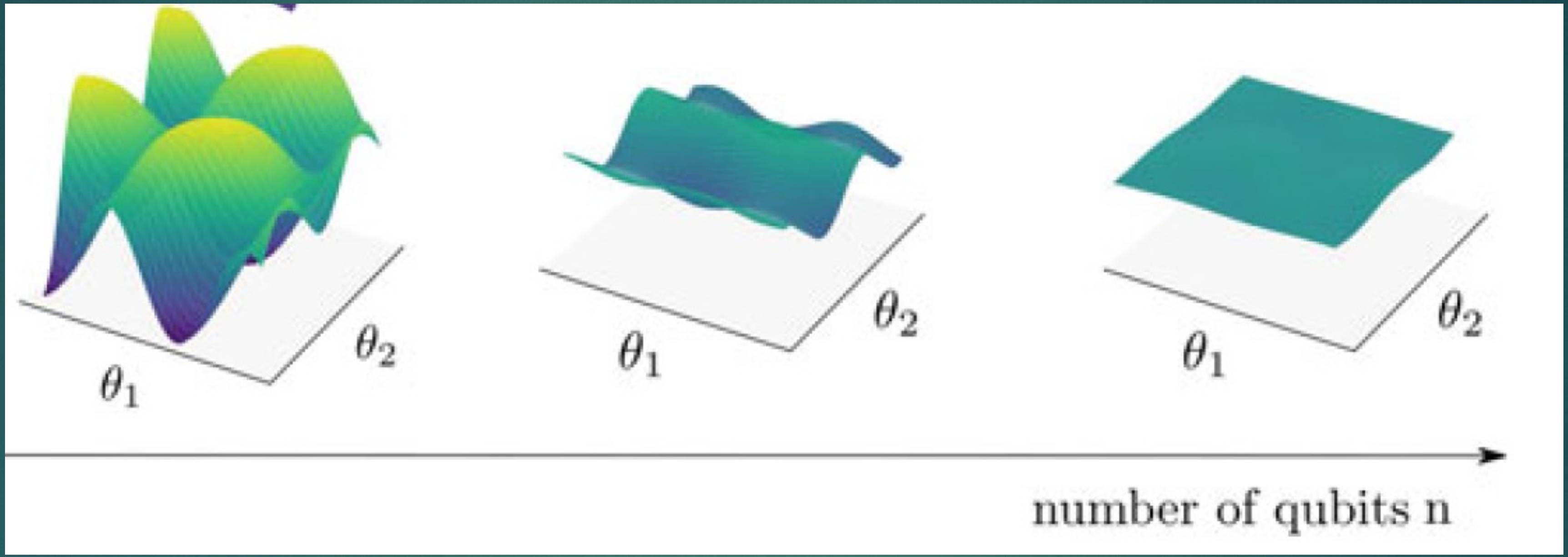
- Hardware efficient *ansatz*
  - "Deep" random parameterized circuits : $depth \sim \mathcal{O}(poly(n))$
  - Random initialization of parameters

$$\langle \partial_{\theta_k} C(\vec{\theta}) \rangle \sim 0$$

$$Var\left(\partial_{\theta_k} C(\vec{\theta})\right) \sim \mathcal{O}\left(\frac{1}{2^n}\right)$$

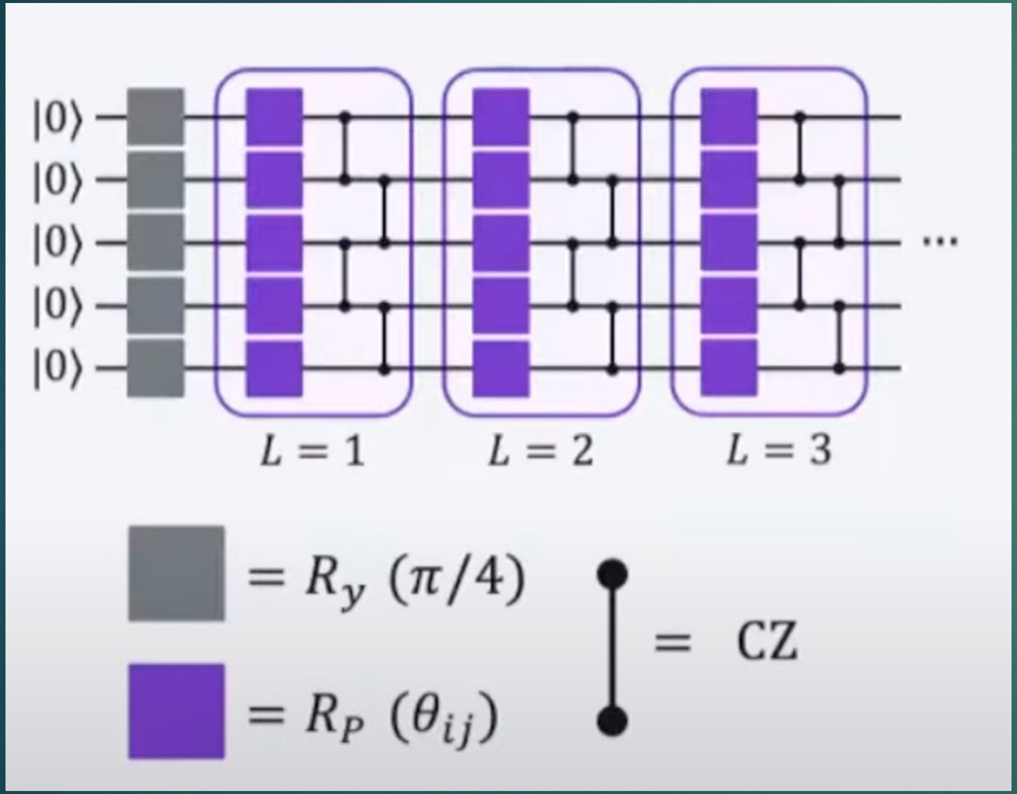
Gradient vanishes exponentially fast with the number of qubits, n

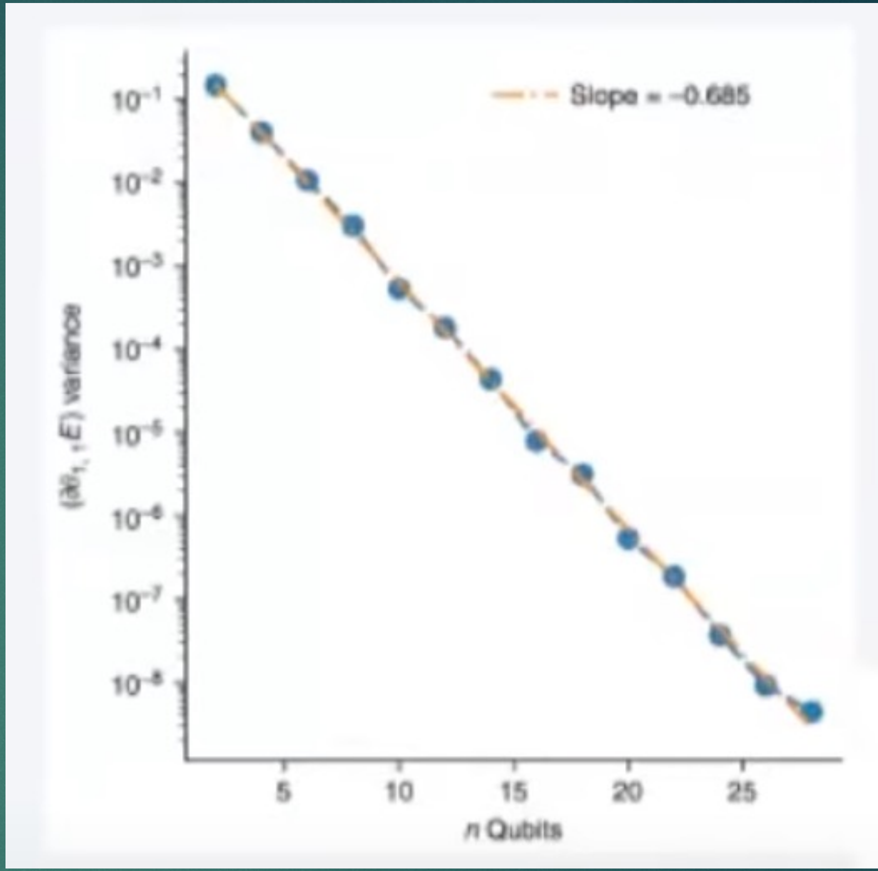
Barren plateaus



[Schuld2021]

Barren Plateaus





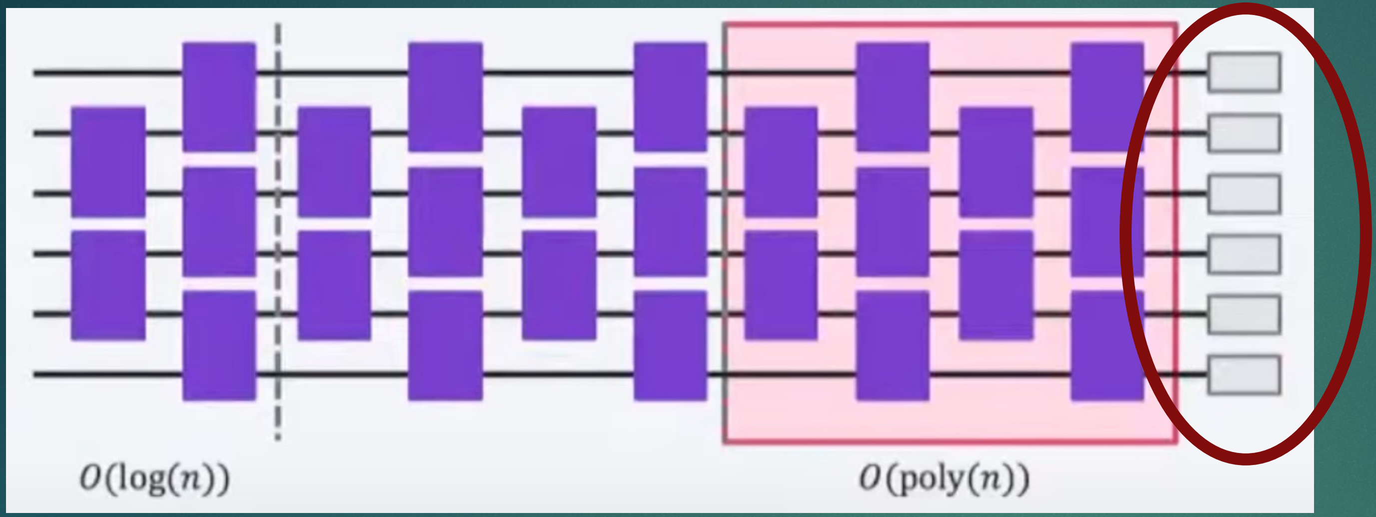
The variance of the gradients decreases exponentially with the number of qubits

$$L = 10n \sim \mathcal{O}(poly(n))$$

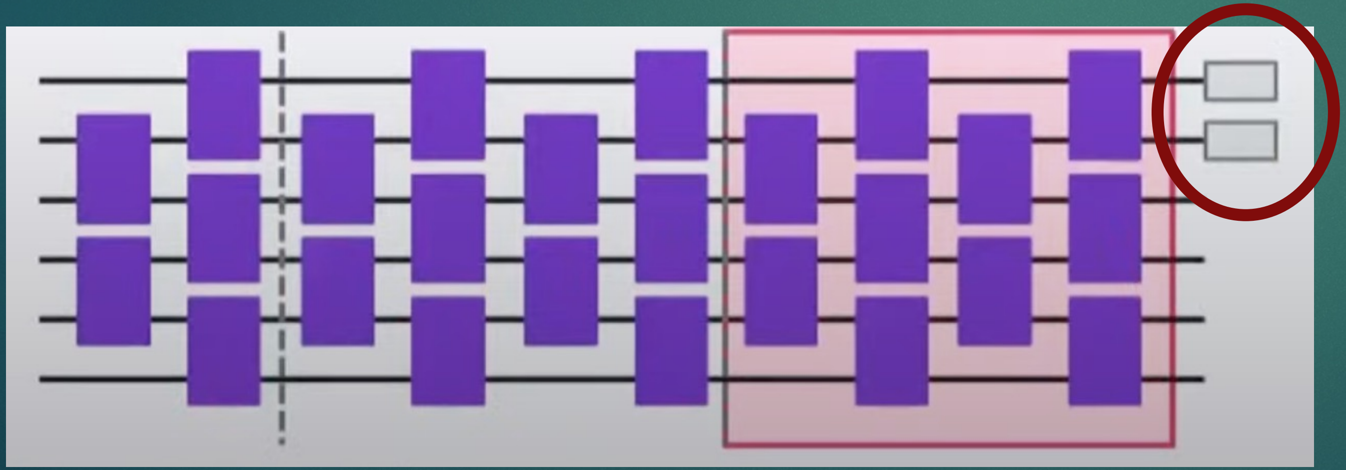
Barren plateaus

- The problem of barren plateaus becomes more acute as the number of qubits increases
- Trainability becomes exponentially more difficult with the number of qubits
- Interesting problems from the real world are expected to require a large number of qubits

Observables



Global Observable



Local Observable

Vanishing Gradients

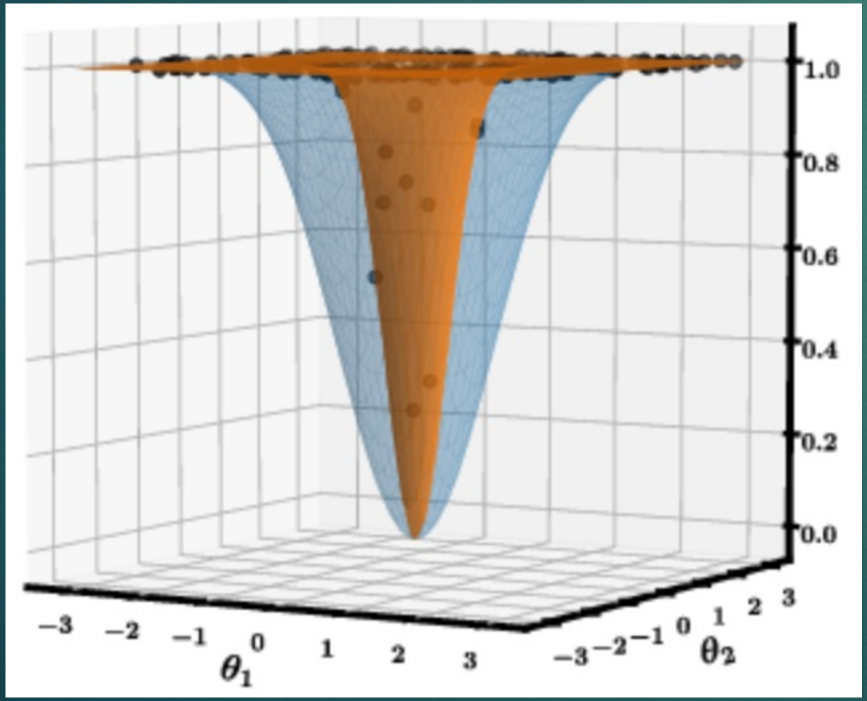
Gradients

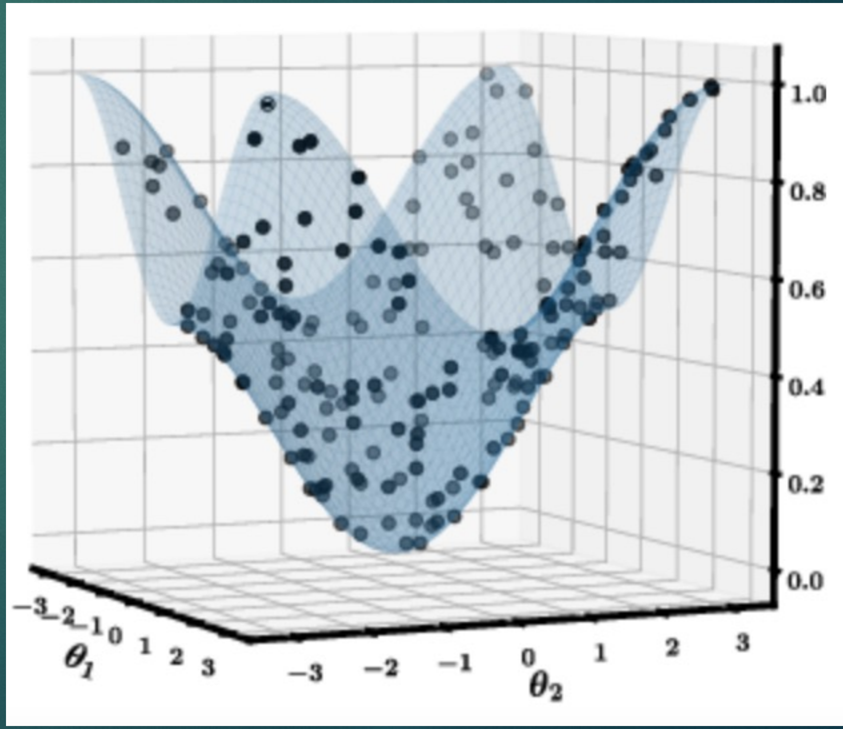
Global Observable

$$\mathrm{Var}[\partial_j C_G] = \frac{1}{8}\left(\frac{3}{8}\right)^{n-1}$$

Local Observable

$$\mathrm{Var}[\partial_j C_L] = \frac{1}{8n^2}$$

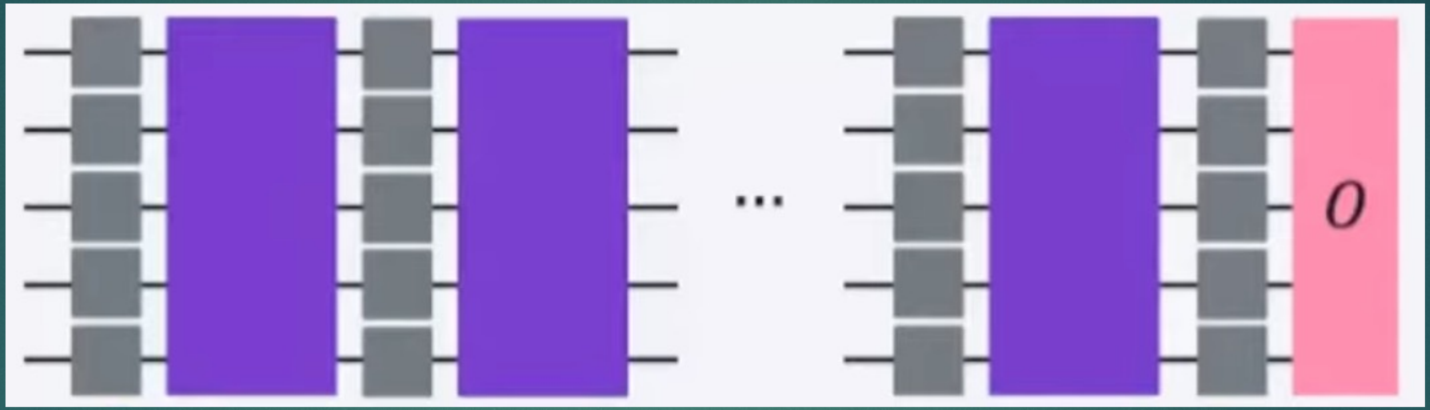




Vanishing Gradients

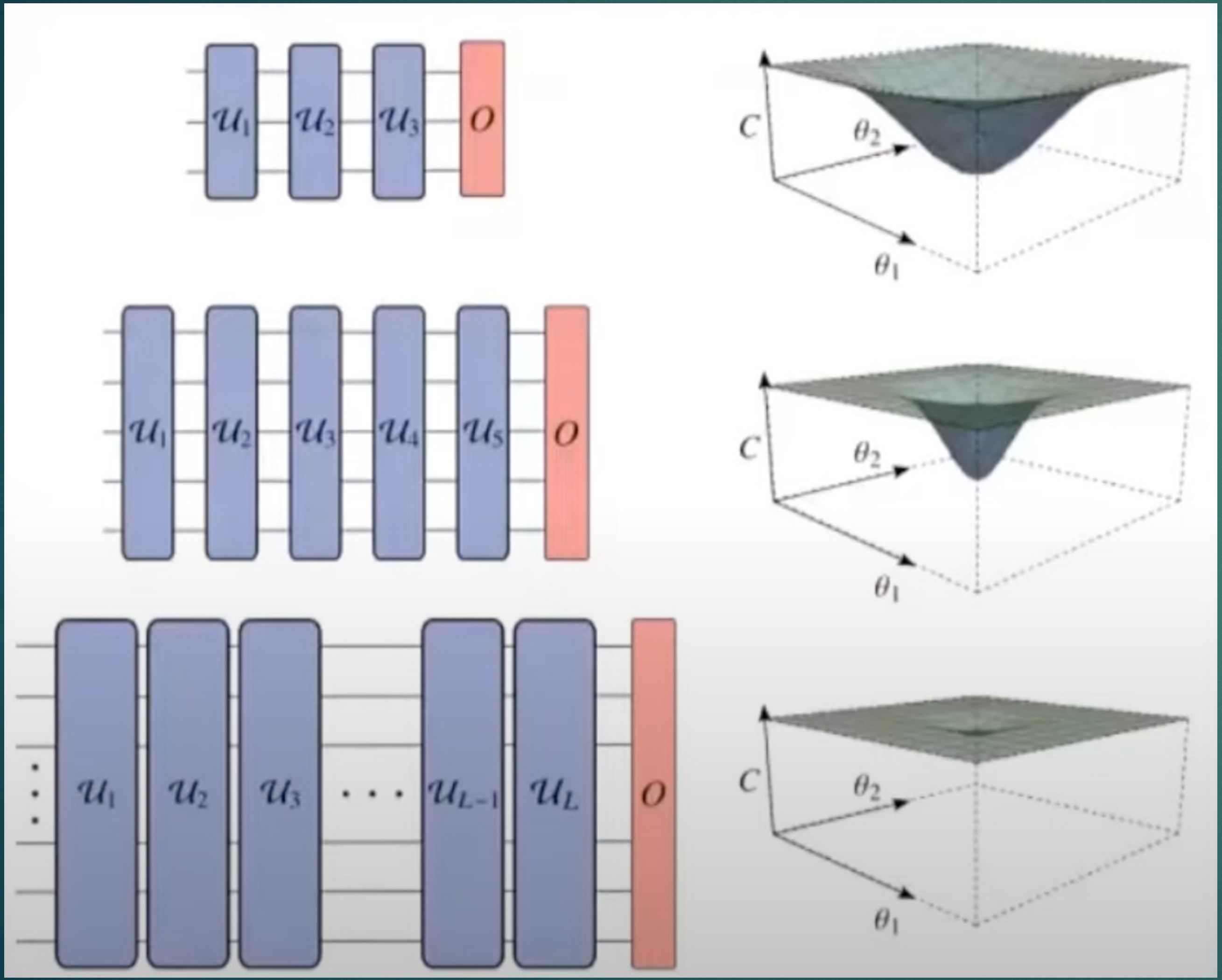
		Depth	
		$\mathcal{O}(\log(n))$	$\mathcal{O}(\mathrm{poly}(n))$
Observable	Global	👎	👎
	Local	👍	👎

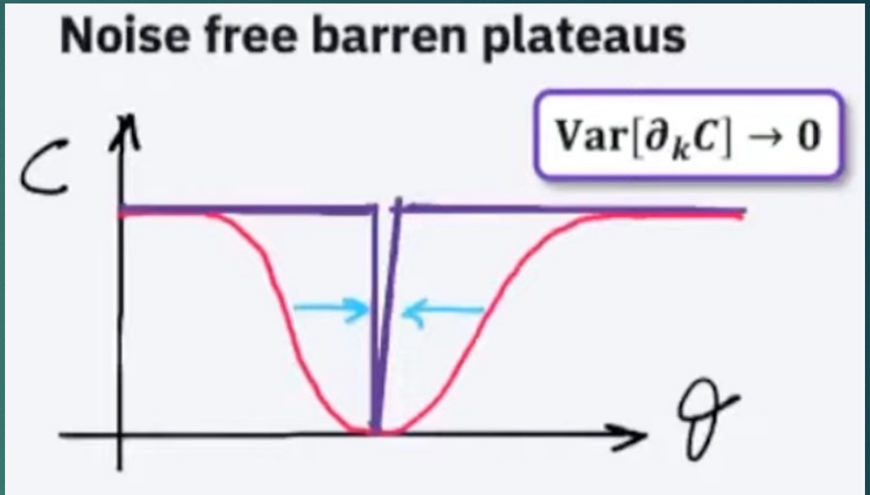
Noise induced barren plateaus

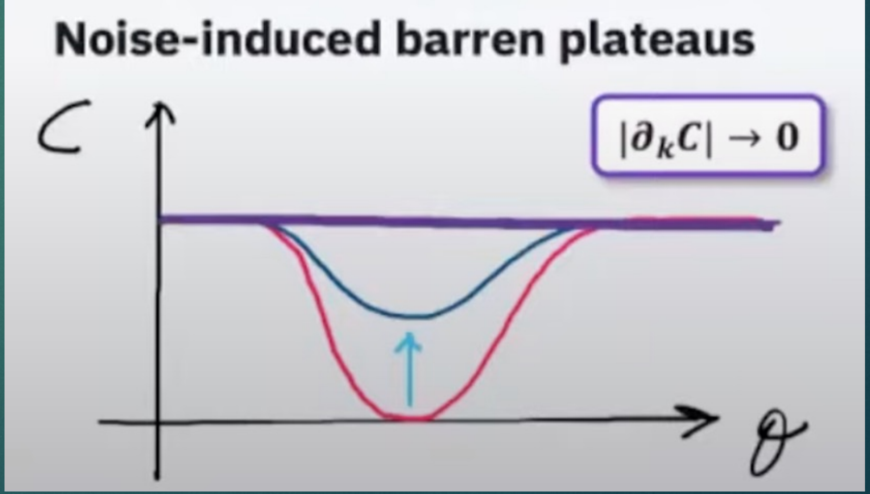


- Gradients do vanish **exponentially with depth**, **independently of** using global or local **measurements**

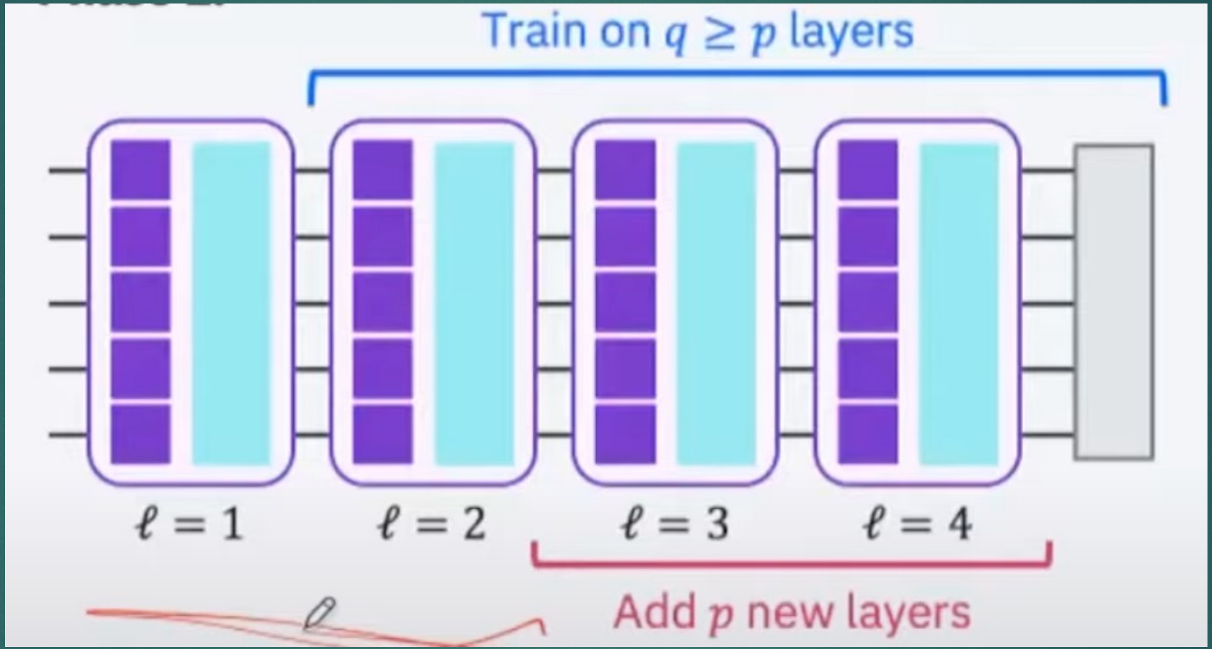
Noise induced barren plateaus





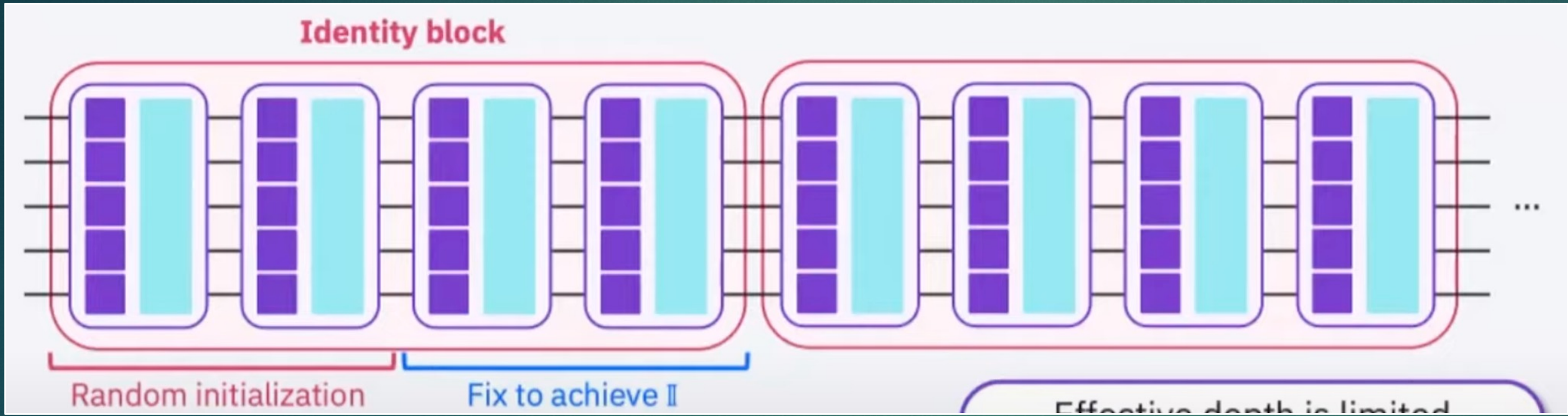


Mitigating barren plateaus: layerwise learning



- Learn parameters of the first set of layers
- Add layers, fixing the previous learnt parameters
- Initialize the new set of layers parameters such that it implements identity

Mitigating barren plateaus: initialization



Effective depth is limited
The gradients will not vanish exponentially fast
at least in the 1st training step

- Partition the circuit into shallow blocks
- Initialize some parameters with random values
- Fix the remaining such that the block implements identity