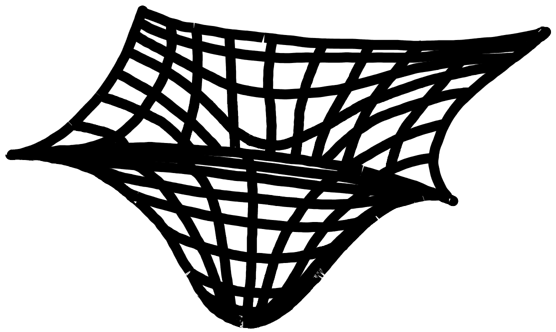
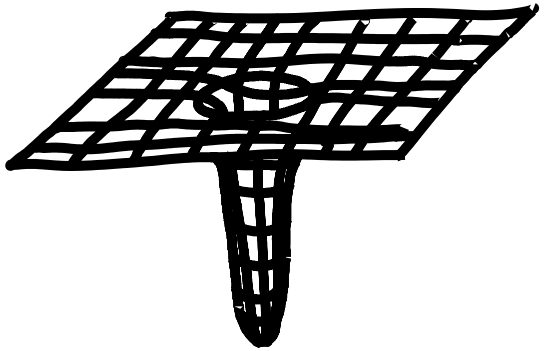
Training landscapes for parameterized quantum circuits

Patrick Coles
Theoretical Division
Los Alamos National Laboratory



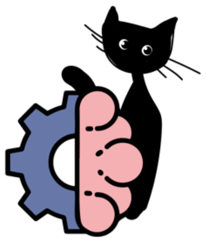
Trainable landscape



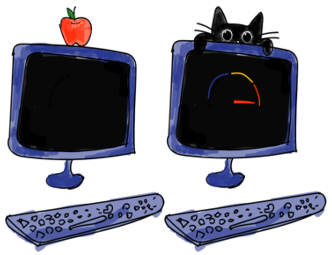
Barren plateau landscape

A commonality between fields: **Parametrized Quantum Evolution**

Quantum Machine Learning (QML)



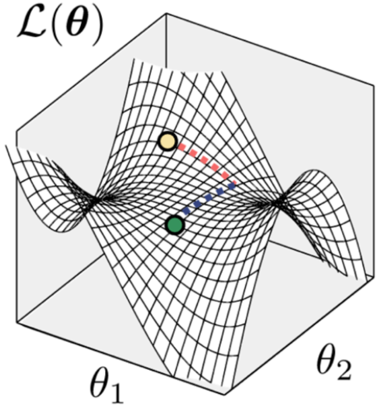
Variational Quantum Algorithms (VQAs)



Quantum Optimal Control (QOC)

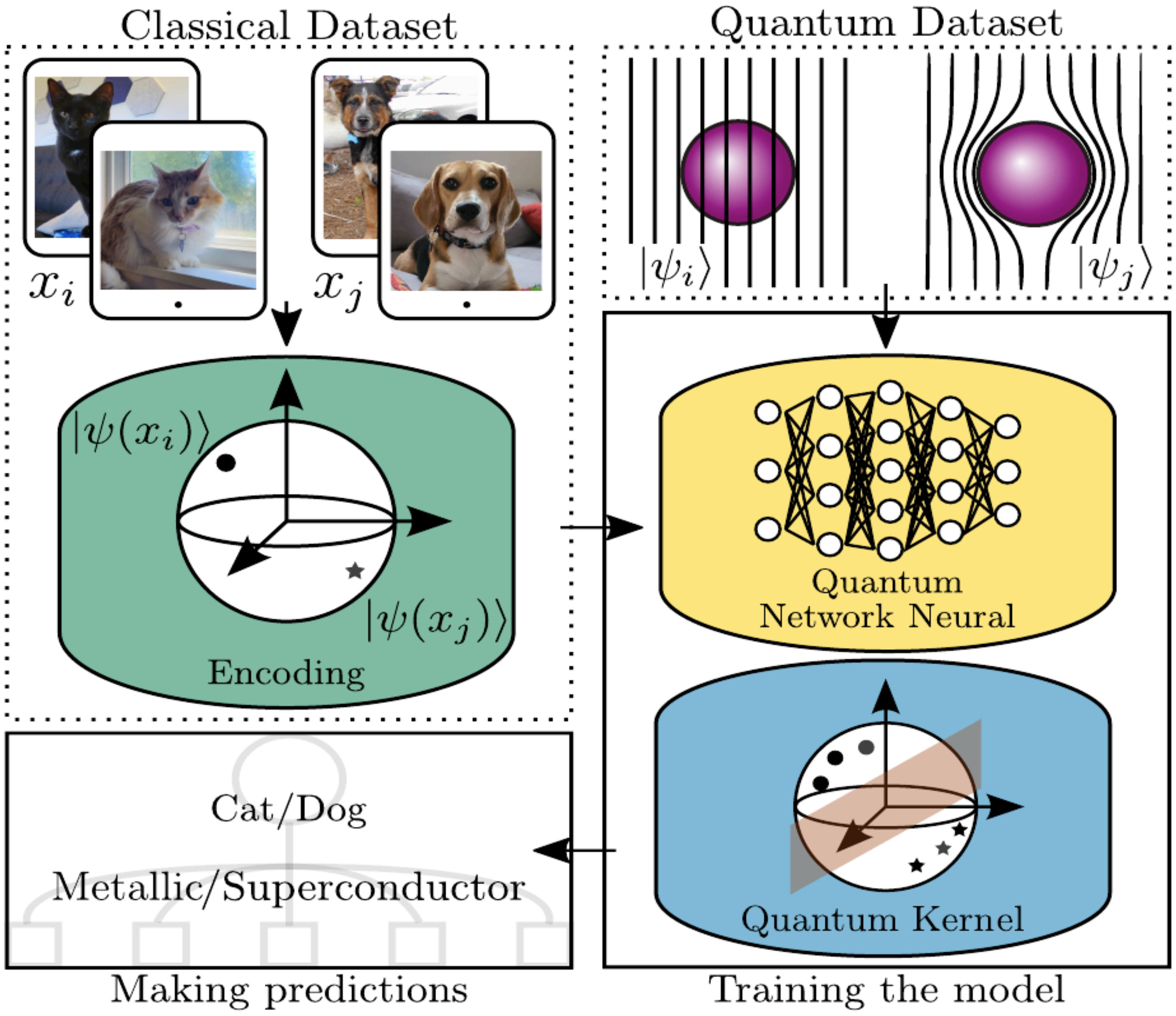
$$|\psi\rangle \rightarrow \boxed{U(\boldsymbol{\theta})} \rightarrow |\psi(\boldsymbol{\theta})\rangle$$

+ task



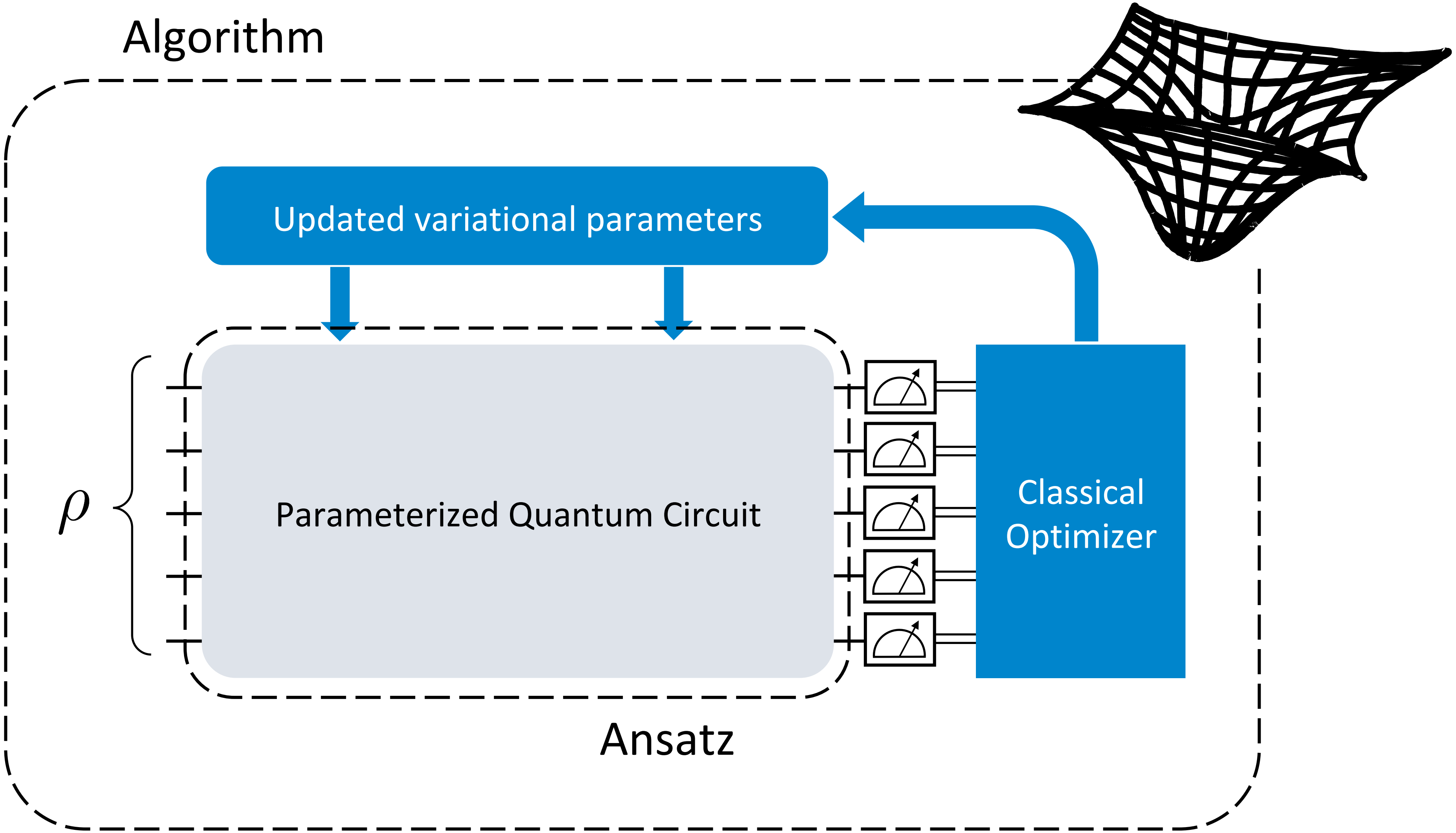
Loss Function Landscape

Quantum Machine Learning



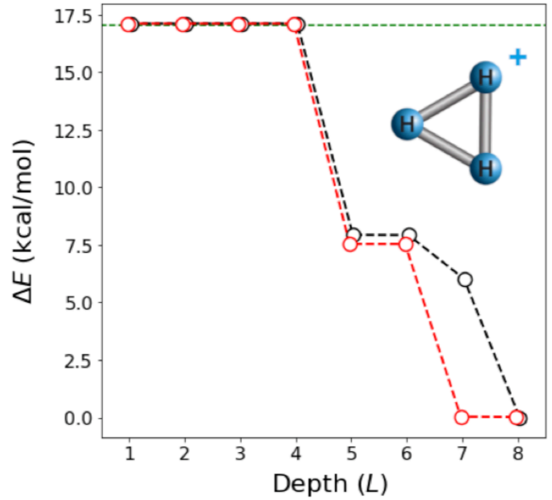
Variational Quantum Algorithms (VQAs)

→automated programming of quantum computers

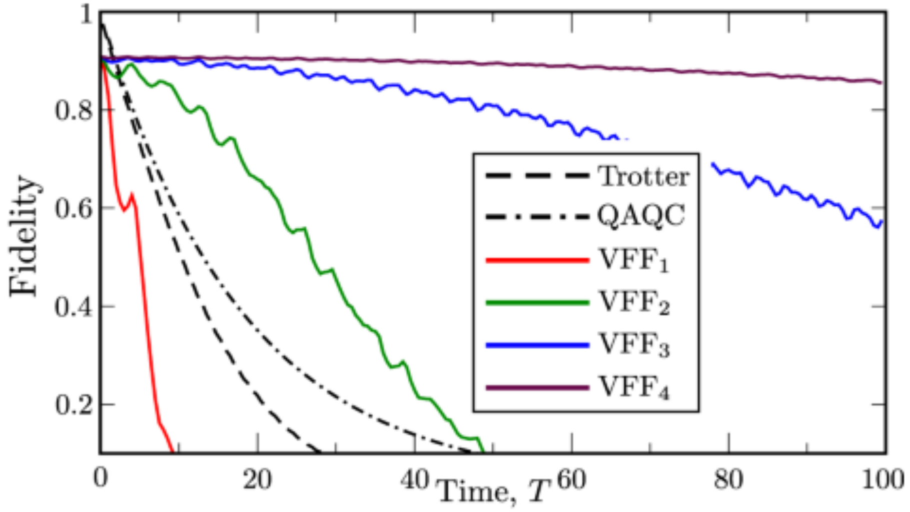


Applications

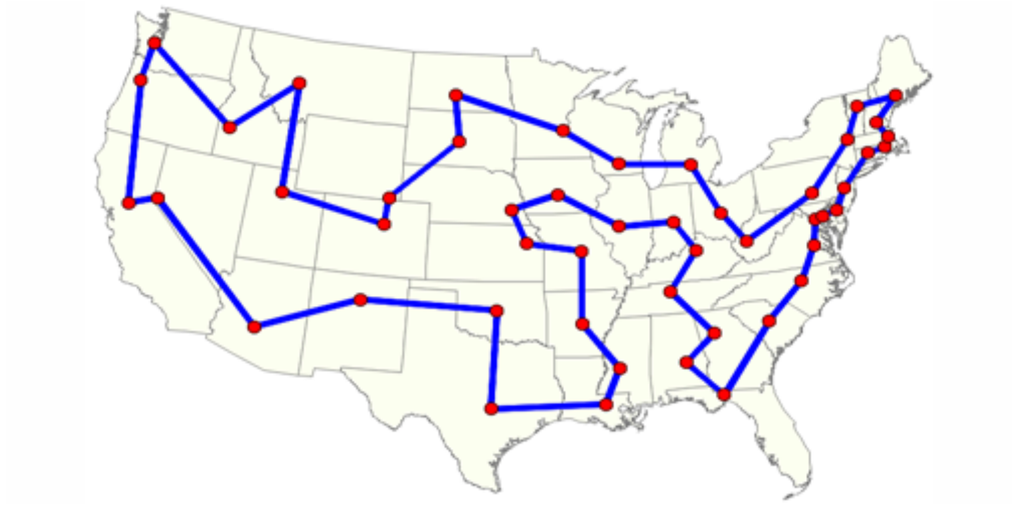
Electronic and Nuclear Structure



Dynamical Simulation



Combinatorial Optimization

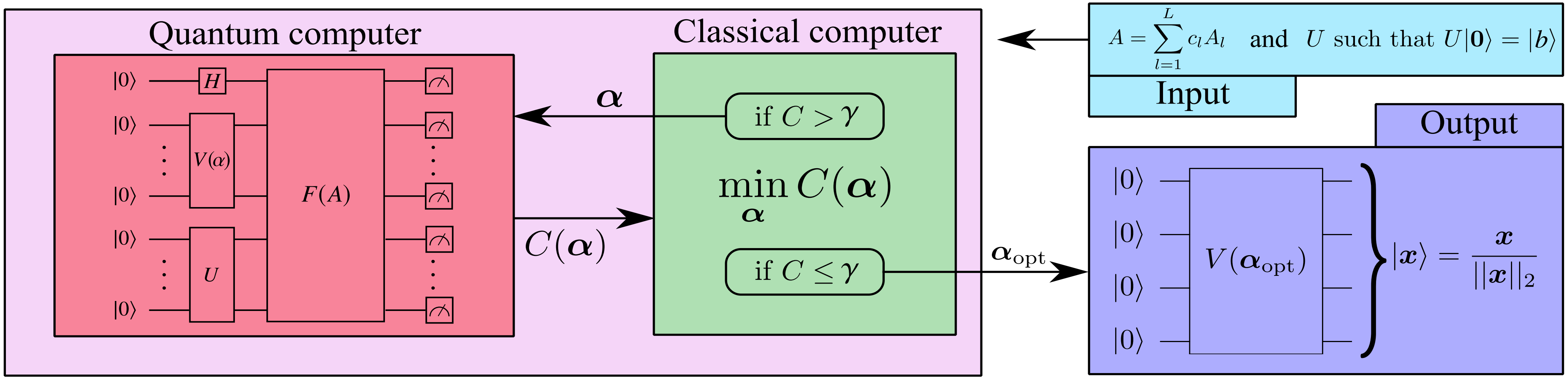


Linear Systems of Equations

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Example: Variational Quantum Linear Solver

Goal: Prepare |*x*> such that: $A|x\rangle \propto |b\rangle$



Input

Linear combination of unitaries

$$A = \sum_{l=1}^{L} c_l A_l$$

Unitary *U* that prepares |*b*>

Output

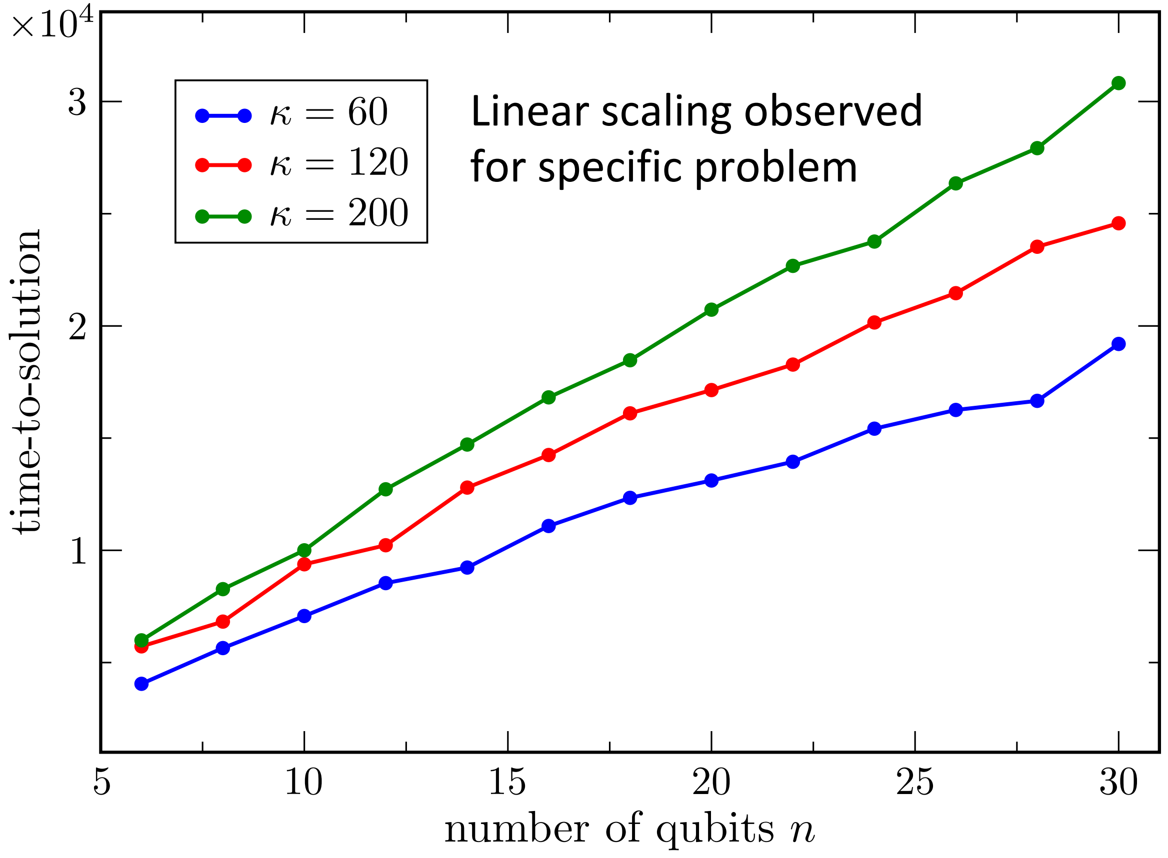
Gate sequence $V(\alpha)$

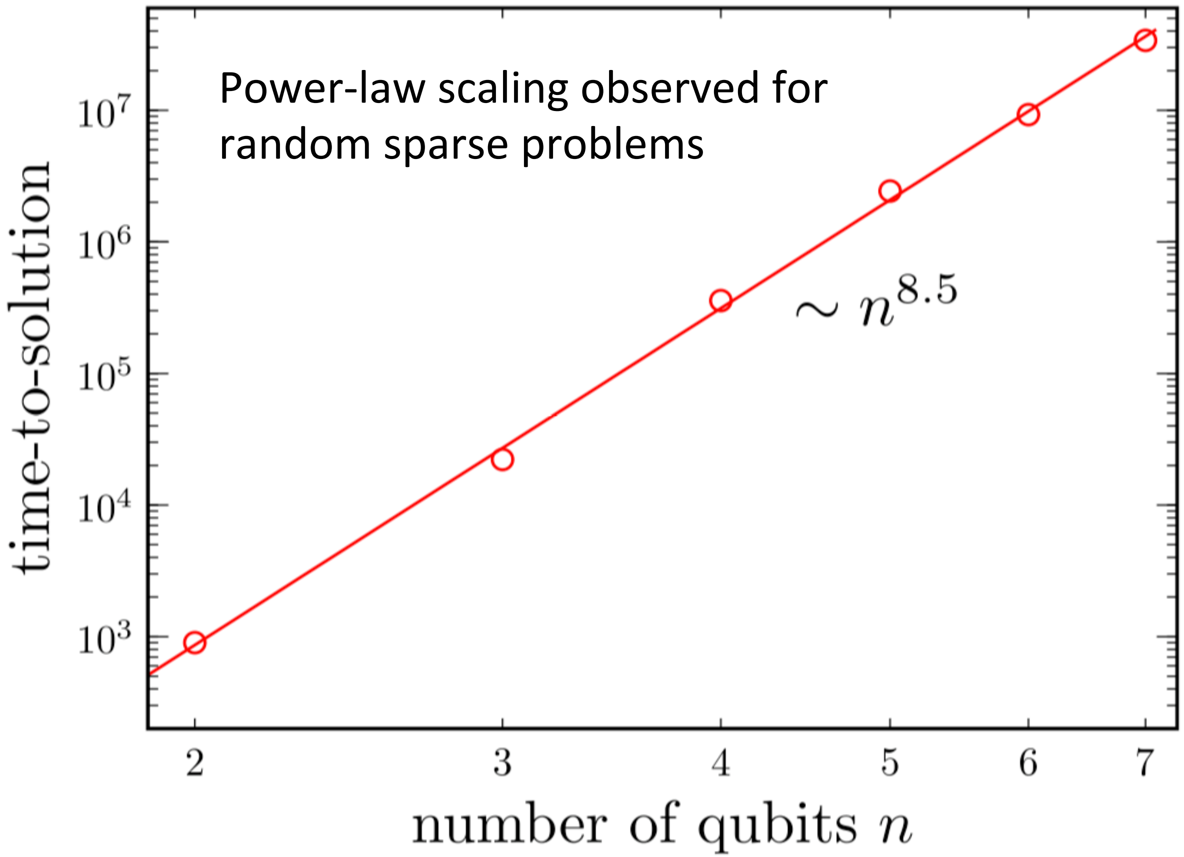
$$|x(\boldsymbol{\alpha})\rangle = V(\boldsymbol{\alpha})|\mathbf{0}\rangle$$

Key Question: Will VQAs scale well?

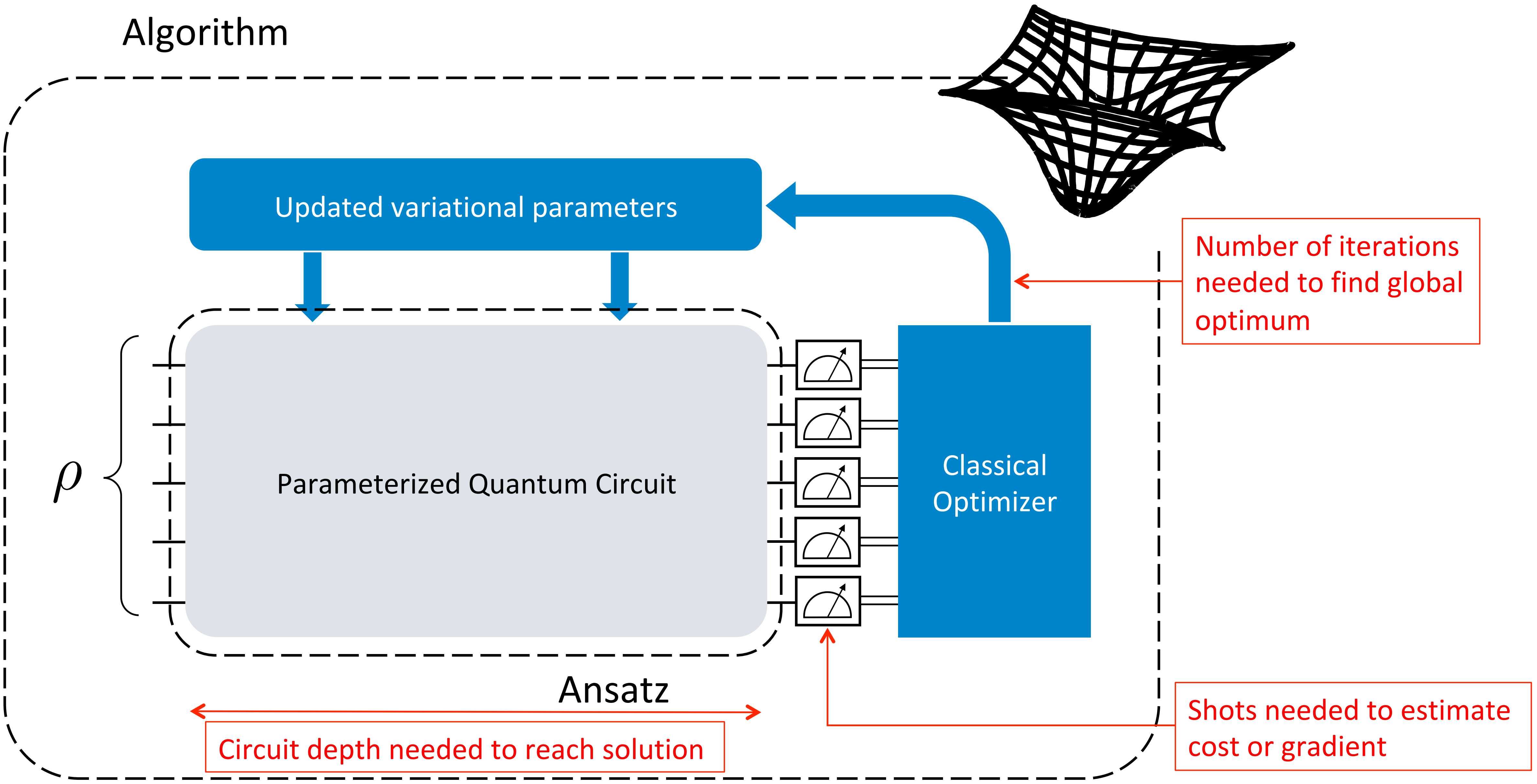
Heuristics play crucial role in answering this

Example: Variational Quantum Linear Solver (VQLS) for solving linear systems





Sources of scaling inefficiency



Sources of scaling inefficiency

Circuit depth needed to reach solution

→ *Obtstacles to variational quantum optimization from symmetry protection*. Bravyi et al. PRL (2020)

Number of iterations needed to find global optimum

→ *Training variational quantum algorithms is NP-hard*. Bittel, Kliesch. arXiv:2101:07267 (2021)

Shots needed to estimate cost or gradient

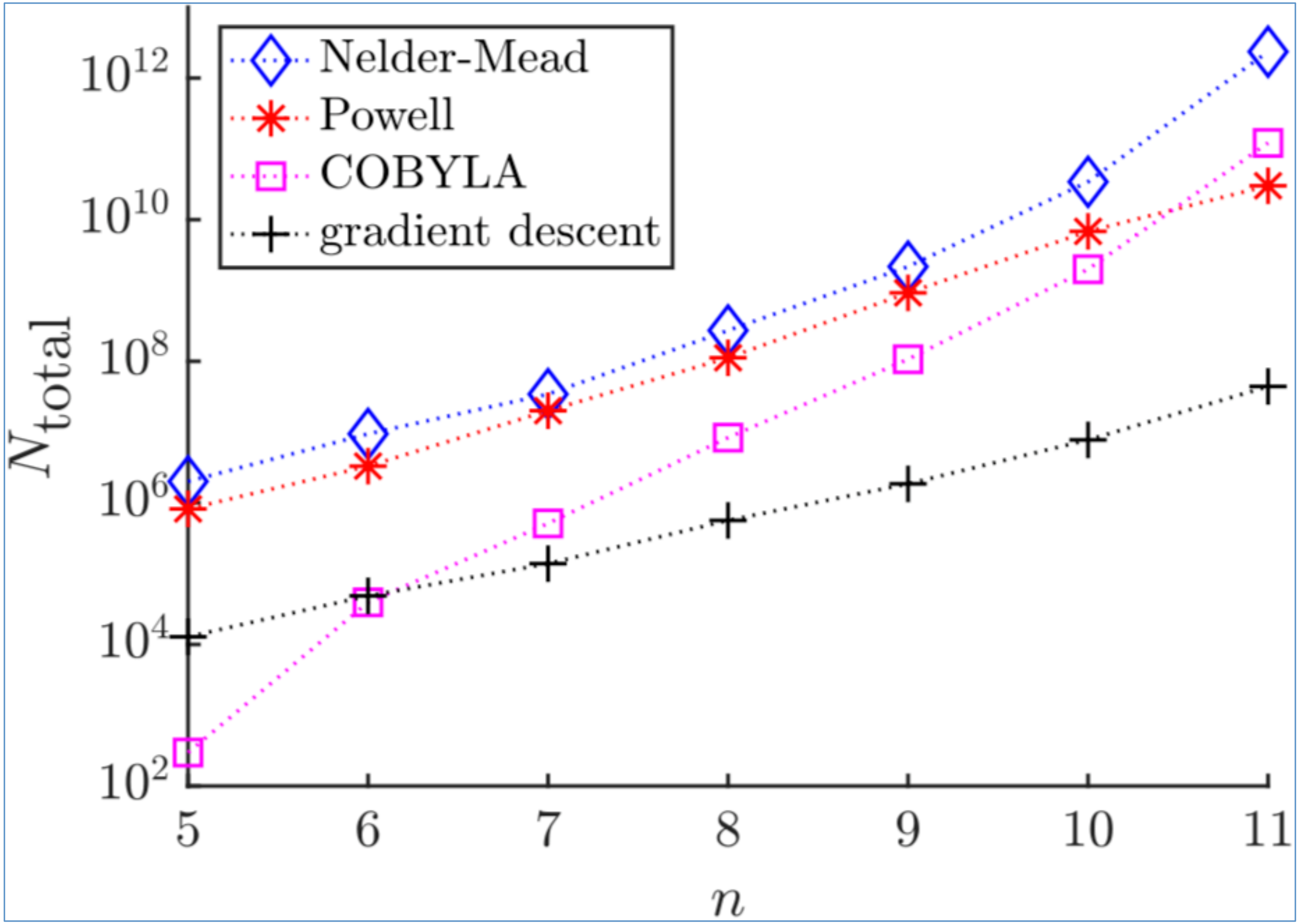
→ *Our work and other groups' work on barren plateaus*

Focus of this talk

A new phenomenon:

Shot cost to start training grows exponentially

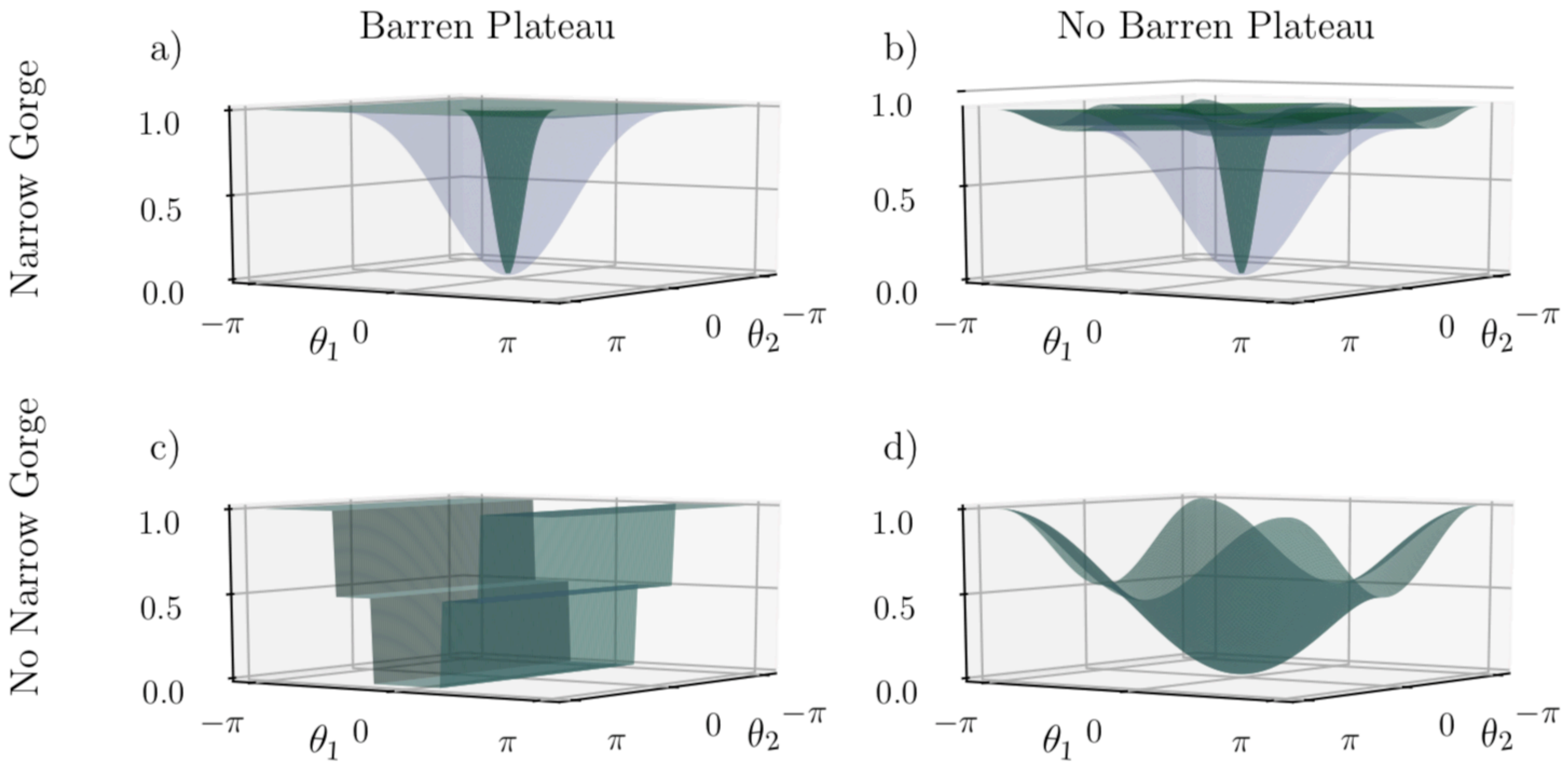
Arxiv: 2011.12245



Why does this happen and how can we prevent it?

Need to understand VQA landscapes

Understanding landscape could allow us to predict the shot cost



Barren plateaus in quantum neural network training landscapes

Jarrod R. McClean,[1,*] Sergio Boixo,[1,†] Vadim N. Smelyanskiy,[1,‡] Ryan Babbush,[1] and Hartmut Neven[1]
[1]*Google Inc., 340 Main Street, Venice, CA 90291, USA*
(Dated: March 30, 2018)

Many experimental proposals for noisy intermediate scale quantum devices involve training a parameterized quantum circuit with a classical optimization loop. Such hybrid quantum-classical algorithms are popular for applications in quantum simulation, optimization, and machine learning. Due to its simplicity and hardware efficiency, random circuits are often proposed as initial guesses for exploring the space of quantum states. We show that the exponential dimension of Hilbert space and the gradient estimation complexity make this choice unsuitable for hybrid quantum-classical algorithms run on more than a few qubits. Specifically, we show that for a wide class of reasonable parameterized quantum circuits, the probability that the gradient along any reasonable direction is non-zero to some fixed precision is exponentially small as a function of the number of qubits. We argue that this is related to the 2-design characteristic of random circuits, and that solutions to this problem must be studied.

Rapid developments in quantum hardware have motivated advances in algorithms to run in the so-called noisy intermediate scale quantum (NISQ) regime [1]. Many of the most promising application-oriented approaches are hybrid quantum-classical algorithms that rely on optimization of a parameterized quantum circuit [2–8]. The resilience of these approaches to certain types of errors and high flexibility with respect to coherence time and gate requirements make them especially attractive for NISQ implementations [3, 9–11].

The first implementation of such algorithms was developed in the context of quantum simulation with the variational quantum eigensolver [2, 3]. This algorithm has been successfully demonstrated on a number of experimental setups with extensions to excited states and

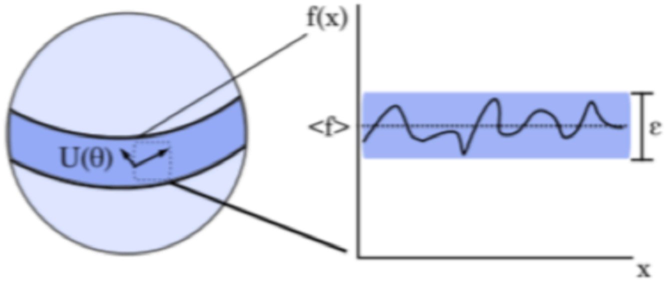
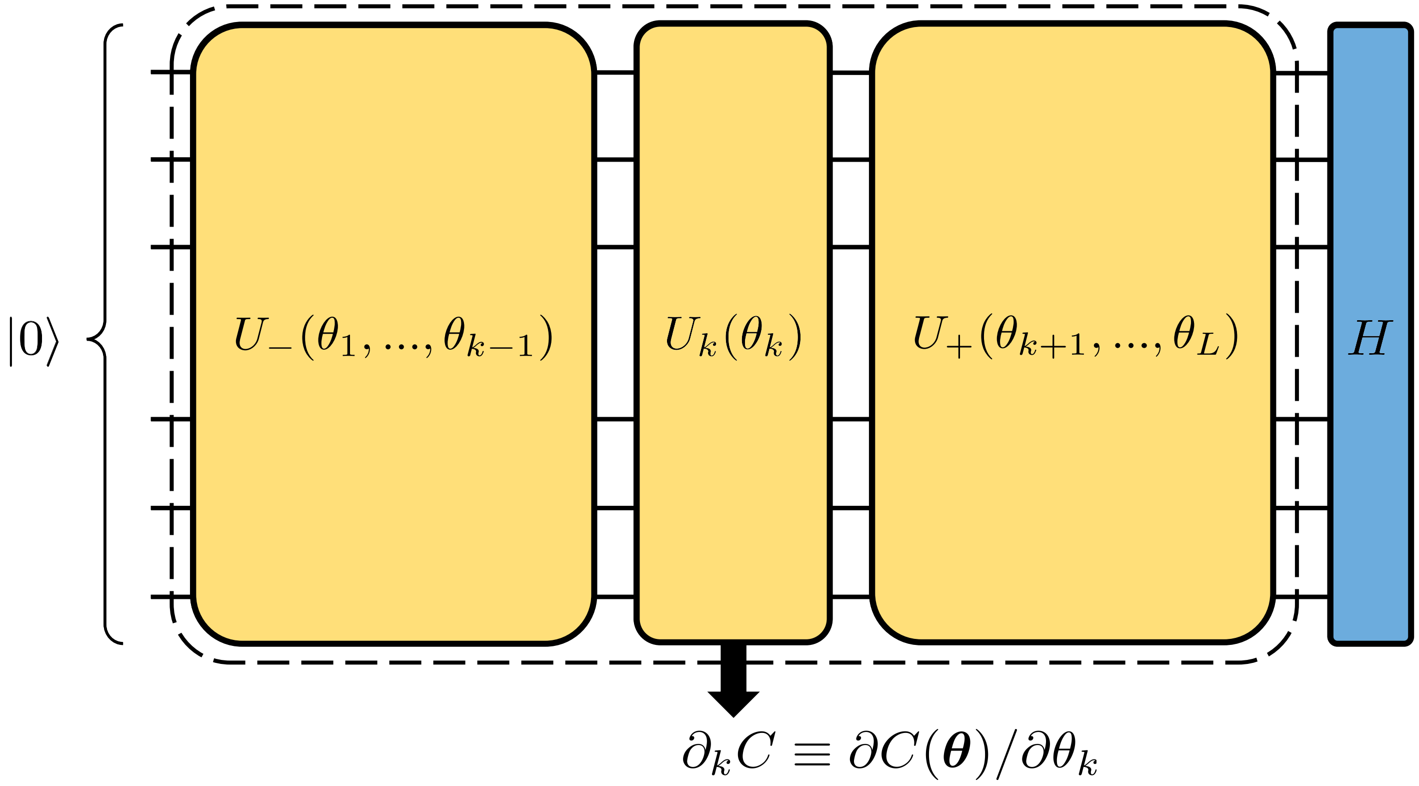


FIG. 1. A cartoon of the general geometric results from this work. The sphere depicts the phenomenon of concentration of measure in quantum space: the fraction of states that fall outside a fixed angular distance from zero along any coordinate decreases exponentially in the number of qubits [37]. This implies a flat plateau where observables concentrate on their average over Hilbert space and the gradient is exponentially

Barren plateaus in deep circuits

[McClean et al., 2016]

Setting:



Parameterized gates:

$$U_k(\theta_k) = \exp(-i\theta_k V_k)$$

Cost function:

$$C(\boldsymbol{\theta}) = \langle 0| U(\boldsymbol{\theta})^\dagger H U(\boldsymbol{\theta}) |0\rangle$$

(Or energy *E*)

Barren plateaus in deep circuits

[McClean et al., 2016]

Result:

Forms 2-design

$$\mathrm{Var}\left[\partial_k E\right] = \begin{cases} -\frac{\mathrm{Tr}\left(\rho^2\right)}{2^{2n}} \mathrm{Tr}\left\langle \left[V, u^\dagger H u\right]^2 \right\rangle_{U_+} & U_- \\ -\frac{\mathrm{Tr}\left(H^2\right)}{2^{2n}} \mathrm{Tr}\left\langle \left[V, u \rho u^\dagger\right]^2 \right\rangle_{U_-} & U_+ \\ 2\, \mathrm{Tr}\left(H^2\right) \mathrm{Tr}\left(\rho^2\right) \left(\frac{\mathrm{Tr}\left(V^2\right)}{2^{3n}} - \frac{\mathrm{Tr}(V)^2}{2^{4n}} \right) & U_- \,\&\, U_+ \end{cases}$$

Recall, t-design:

$$\frac{1}{|X|} \sum_{U \in X} U^{\otimes t} \otimes (U^*)^{\otimes t} = \int_{U(d)} U^{\otimes t} \otimes (U^*)^{\otimes t} dU$$

Barren plateaus in deep circuits

[McClean et al., 2016]

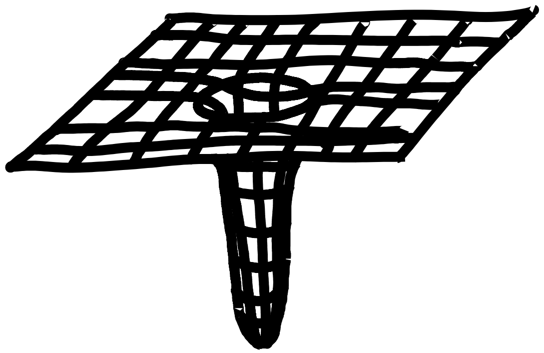
Result:

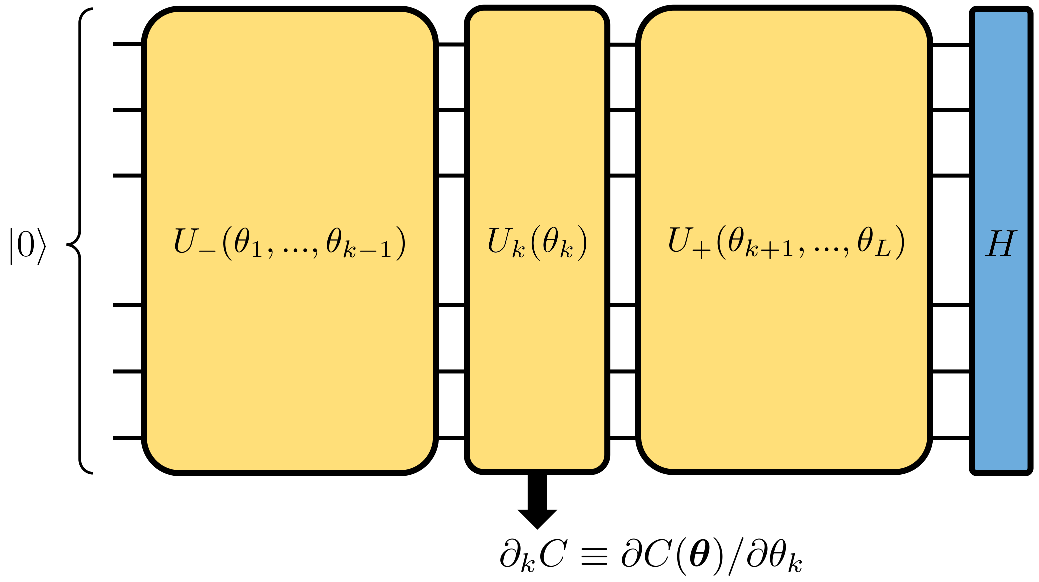
If U_+ or U_- form a 2-design then:

$$\mathrm{Var}\left[\partial_k C\right] = G(n)$$

where

$$G(n) \in \mathcal{O}(1/2^{2n})$$





Parameterized gates:

$$U_k\left(\theta_k\right) = \exp\left(-i\theta_k V_k\right)$$

Cost function:

$$C(\boldsymbol{\theta}) = \langle 0| U(\boldsymbol{\theta})^\dagger H U(\boldsymbol{\theta}) |0\rangle$$

Barren plateaus in deep circuits

[McClean et al., 2016]

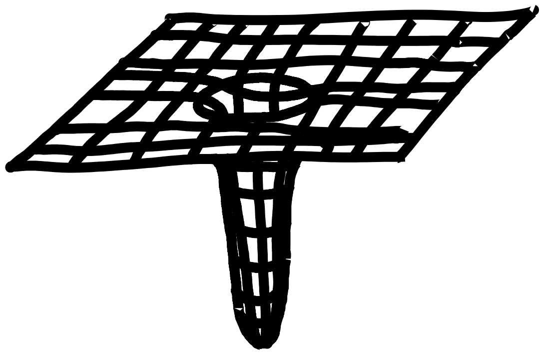
Result:

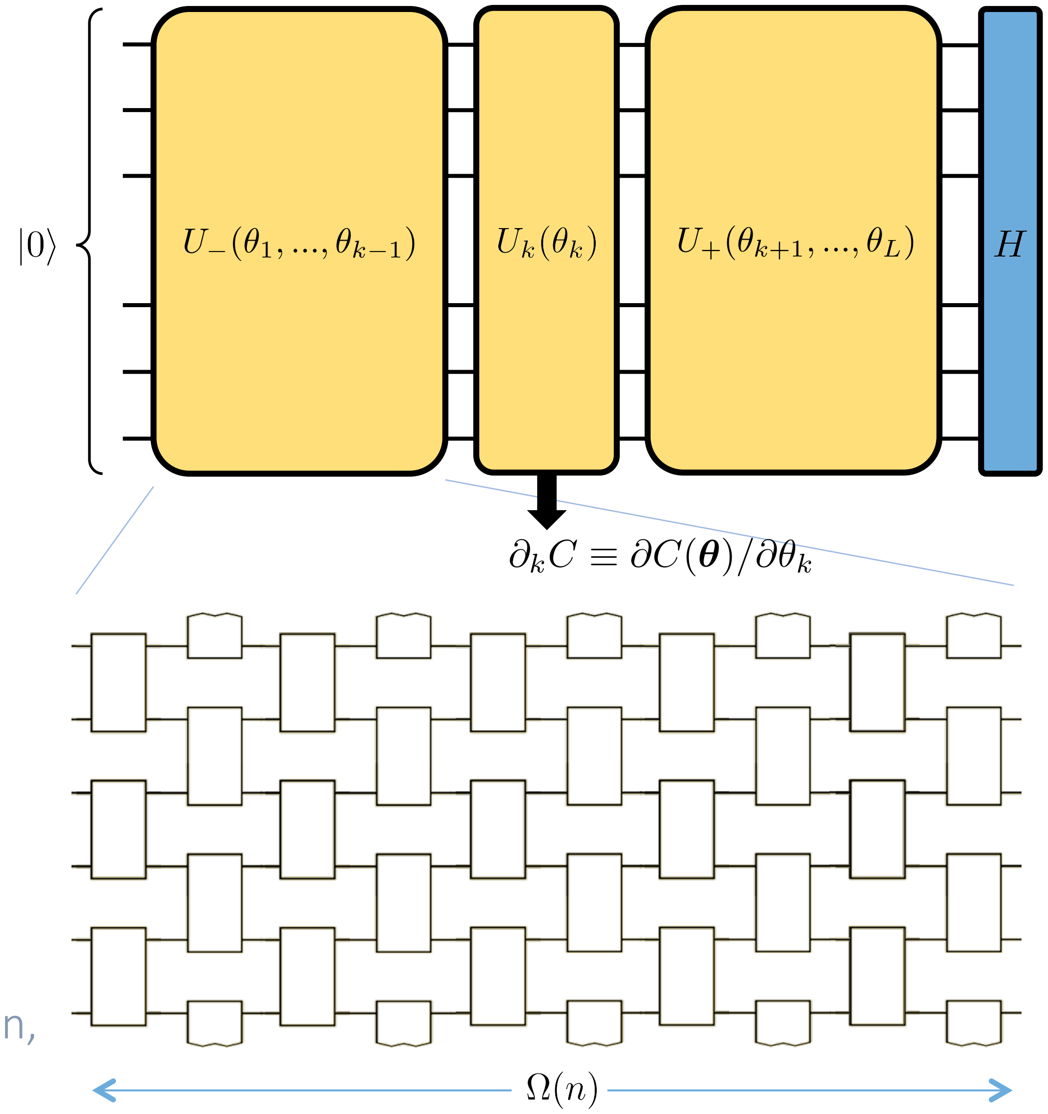
If U_+ or U_- form a 2-design then:

$$\mathrm{Var}\left[\partial_k C\right] = G(n)$$

where

$$G(n) \in \mathcal{O}(1/2^{2n})$$



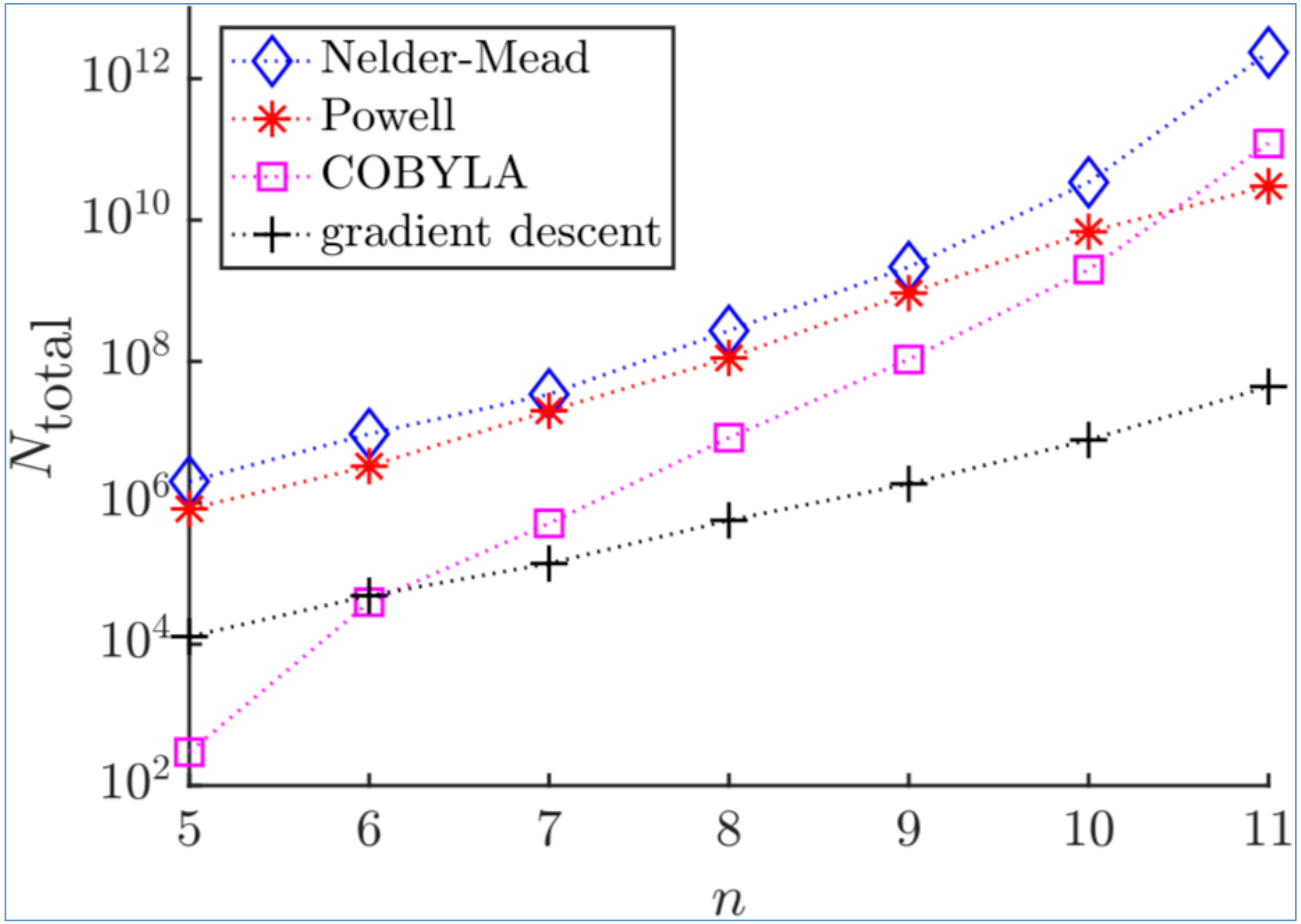


[Harrow & Mehraban, 2018]

Impact on Gradient-based and Gradient-free optimizers

(Arrasmith, Cerezo, Czarnik, Cincio, Coles. Arxiv: 2011.12245)

Now we can understand why exponential scaling occurs:
Exponential precision required to escape a barren plateau



Proposition 1. *Consider the cost function of Eq. (1). Let $\boldsymbol{\theta}_A$ be a randomly chosen point in parameter space. Let $\boldsymbol{\theta}_B = \boldsymbol{\theta}_A + L\hat{\boldsymbol{\ell}}$ be a point at a distance $L = \|\boldsymbol{\theta}_B - \boldsymbol{\theta}_A\|$ from $\boldsymbol{\theta}_A$ in parameter space, for some unit vector $\hat{\boldsymbol{\ell}}$. If the cost exhibits a barren plateau according to Definition 1, then the expectation value of the difference $\Delta C = C(\boldsymbol{\theta}_B) - C(\boldsymbol{\theta}_A)$ is*

$$E_{\boldsymbol{\theta}_A}[\Delta C] = 0\,, \tag{6}$$

and the variance is exponentially vanishing with n as

$$\mathrm{Var}_{\boldsymbol{\theta}_A}[\Delta C] \leqslant G(n)\,, \tag{7}$$

with

$$G(n) = m^2 L^2 F(n)\,, \quad \textit{and} \quad G(n) \in \widetilde{\mathcal{O}}\left(\frac{1}{b^n}\right)\,, \tag{8}$$

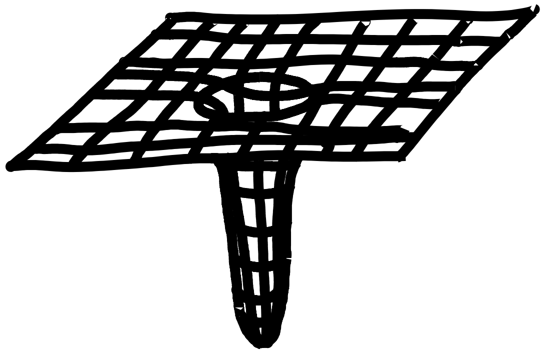
for some $b > 1$. Here m is the dimension of the parameter space, and $F(n)$ was defined in (4).

Extension to Hessian and Higher Order Derivatives

(Cerezo, Coles. Arxiv: 2008.07454)

Proposition 1. *Consider a cost function of the form (1), for which the parameter shift rule of (2) holds. Let H_{ij} be the matrix elements of the Hessian as defined in (5). Then, assuming that $\langle \partial_i C \rangle_{\boldsymbol{\theta}} = 0$, the following inequality holds for any $c > 0$,*

$$\Pr(|H_{ij}| \geqslant c) \leqslant \frac{2\mathrm{Var}_{\boldsymbol{\theta}}[\partial_i C]}{c^2} . \tag{7}$$



Proposition 2. *Consider a cost function of the form (1), for which the parameter shift rule of (2) holds. Let $D^{i,\boldsymbol{\alpha}}C(\boldsymbol{\theta})$ be a higher order partial derivative of the cost as defined in (16). Then, assuming that $\langle \partial_i C \rangle_{\boldsymbol{\theta}} = 0$, the following inequality holds for any $c > 0$,*

$$\Pr\left(\left|D^{i,\boldsymbol{\alpha}}C(\boldsymbol{\theta})\right| \geqslant c\right) \leqslant \frac{2^{|\boldsymbol{\alpha}|}\mathrm{Var}_{\boldsymbol{\theta}}[\partial_i C]}{c^2} . \tag{21}$$

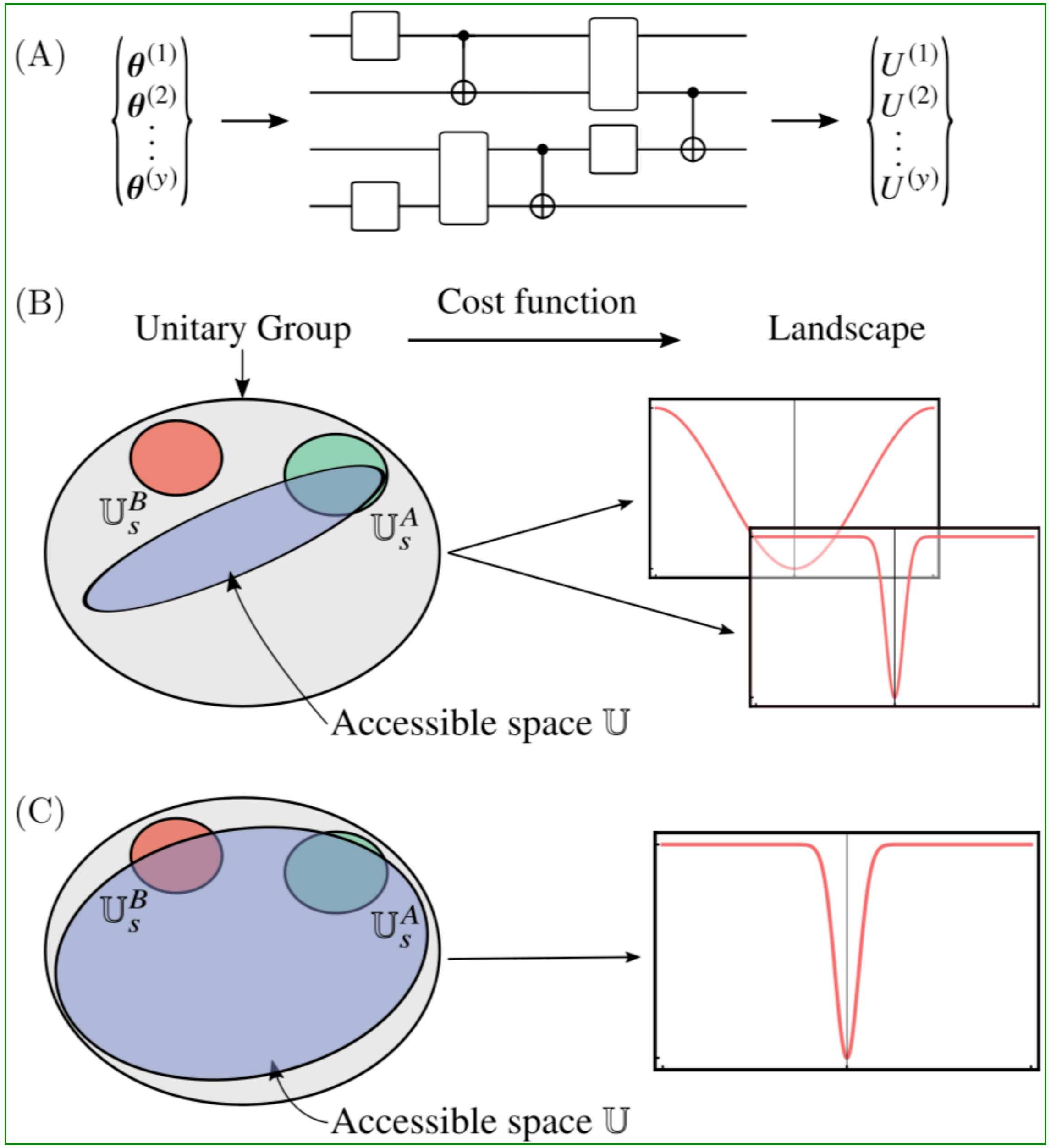
Changing the optimizer does not seem to solve the exponential scaling!

Q: Where do barren plateaus come from?

A: Highly expressive ansatzes (i.e., ignorance)

Connecting ansatz expressibility to gradient magnitudes and barren plateaus

(Holmes, Sharma, Cerezo, Coles. Arxiv: 2101.02138)



Ansatz Expressibility:

(Sim, Johnson, Aspuru-Guzik. AQT (2019))

$$\varepsilon_{\mathbb{U}}^{\rho} := ||\mathcal{A}_{\mathbb{U}}(\rho^{\otimes 2})||_2$$

$$\varepsilon_{\mathbb{U}}^{H} := ||\mathcal{A}_{\mathbb{U}}(H^{\otimes 2})||_2$$

$$\mathcal{A}_{\mathbb{U}}^{(t)}(\cdot) := \int_{\mathcal{U}(d)} d\mu(V) V^{\otimes t}(\cdot)(V^{\dagger})^{\otimes t} - \int_{\mathbb{U}} dU\, U^{\otimes t}(\cdot)(U^{\dagger})^{\otimes t}$$

Cost Function:

$$C_{\rho,H}(\boldsymbol{\theta}) = \mathrm{Tr}[HU(\boldsymbol{\theta})\rho U(\boldsymbol{\theta})^\dagger]$$

Ansatz Expressibility:

$$\varepsilon_{\mathbb{U}}^{\rho} := ||\mathcal{A}_{\mathbb{U}}(\rho^{\otimes 2})||_2$$
$$\varepsilon_{\mathbb{U}}^{H} := ||\mathcal{A}_{\mathbb{U}}(H^{\otimes 2})||_2$$

$$\mathcal{A}_{\mathbb{U}}^{(t)}(\cdot) := \int_{\mathcal{U}(d)} d\mu(V) V^{\otimes t}(\,\cdot\,)(V^\dagger)^{\otimes t} - \int_{\mathbb{U}} dU\, U^{\otimes t}(\,\cdot\,)(U^\dagger)^{\otimes t}$$

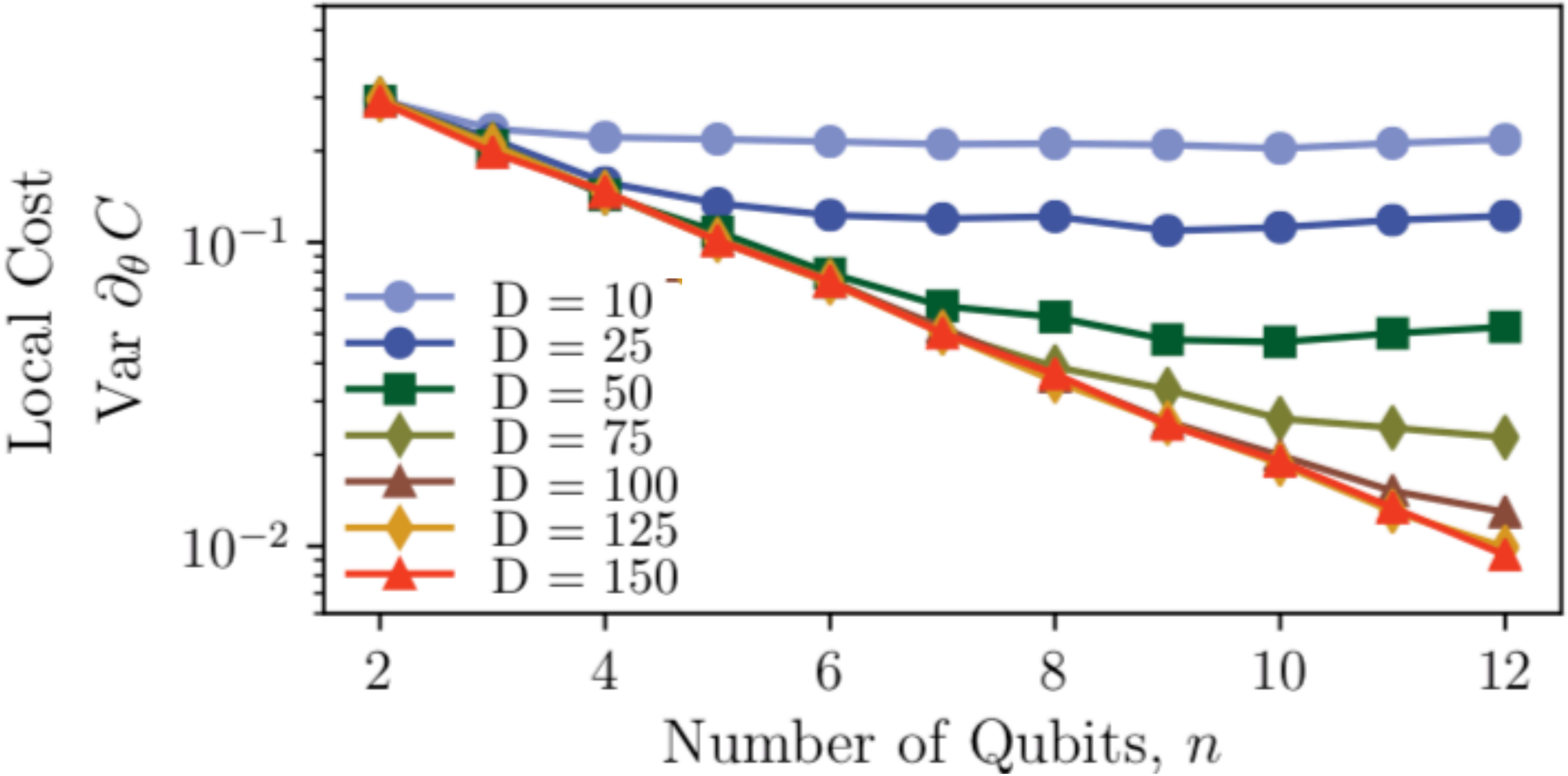
Main Result:

$$\mathrm{Var}\,\partial_k C \leqslant \frac{g_{L,R}(\rho, H, U)}{2^{2n}-1} + f(\varepsilon_L^H, \varepsilon_R^\rho)$$

Generalizes the Google result to approximate 2-designs

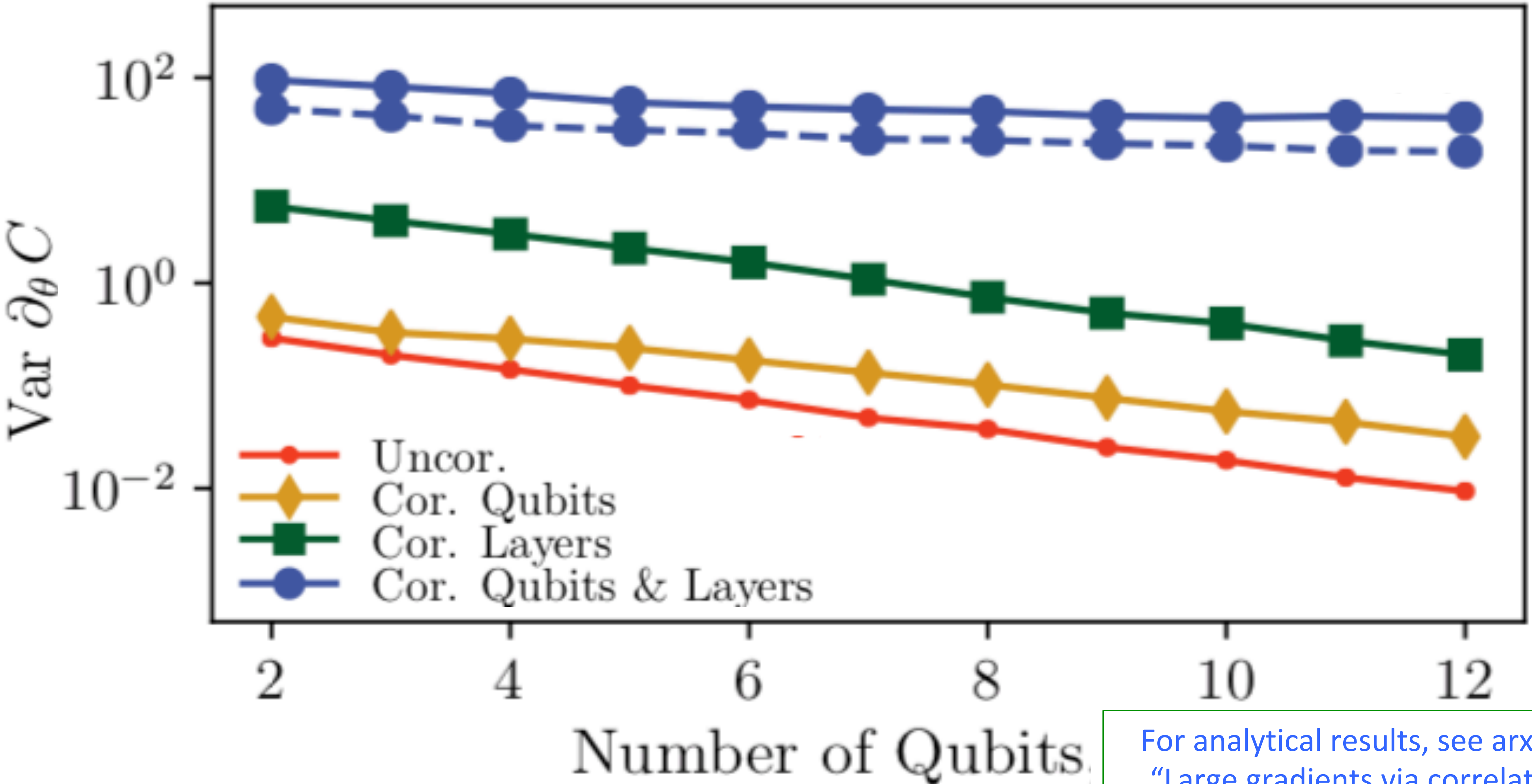
Reducing expressibility is necessary to avoid BPs

Reducing circuit depth helps



Reducing expressibility is necessary to avoid BPs

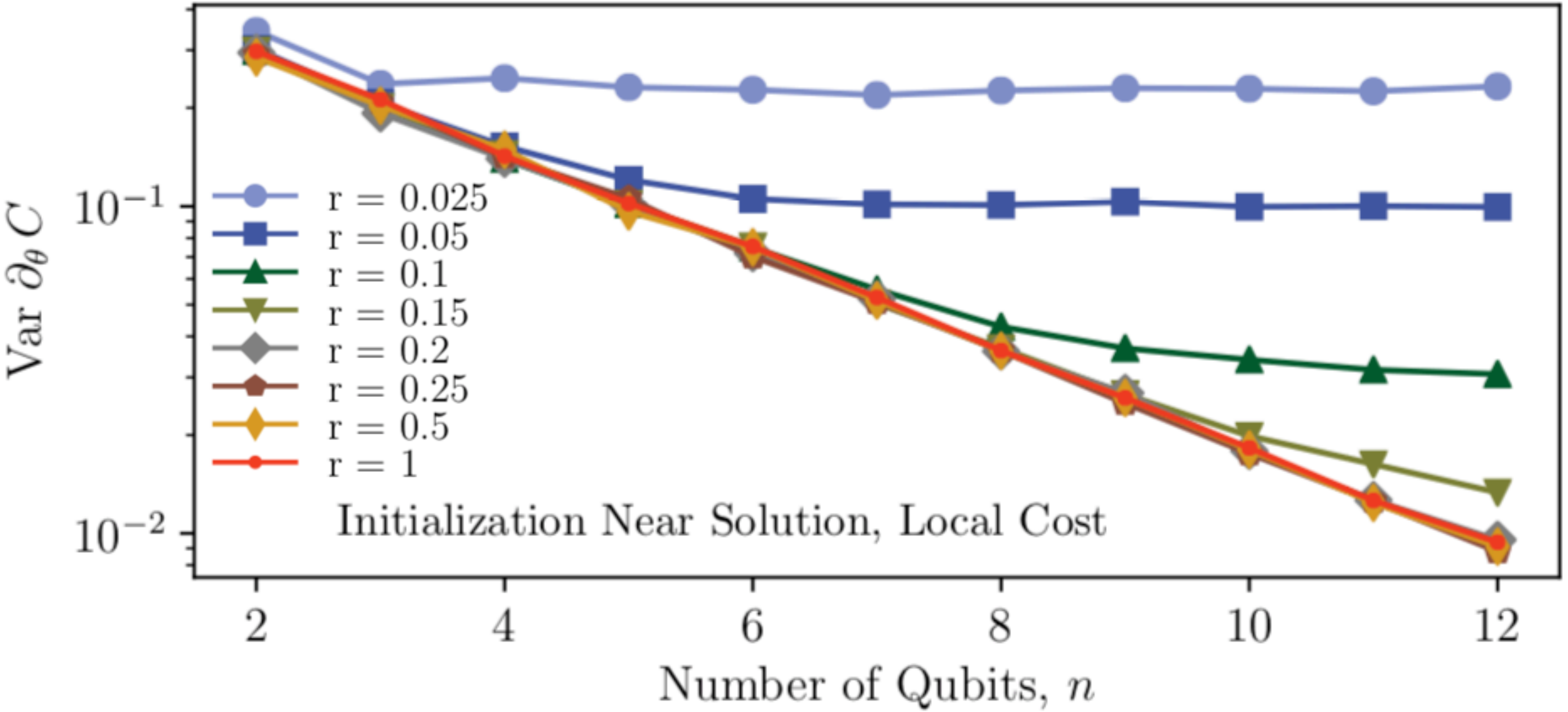
Correlating parameters helps



For analytical results, see arxiv: 2005.12200
"Large gradients via correlation in random parameterized quantum circuits"

Reducing expressibility is necessary to avoid BPs

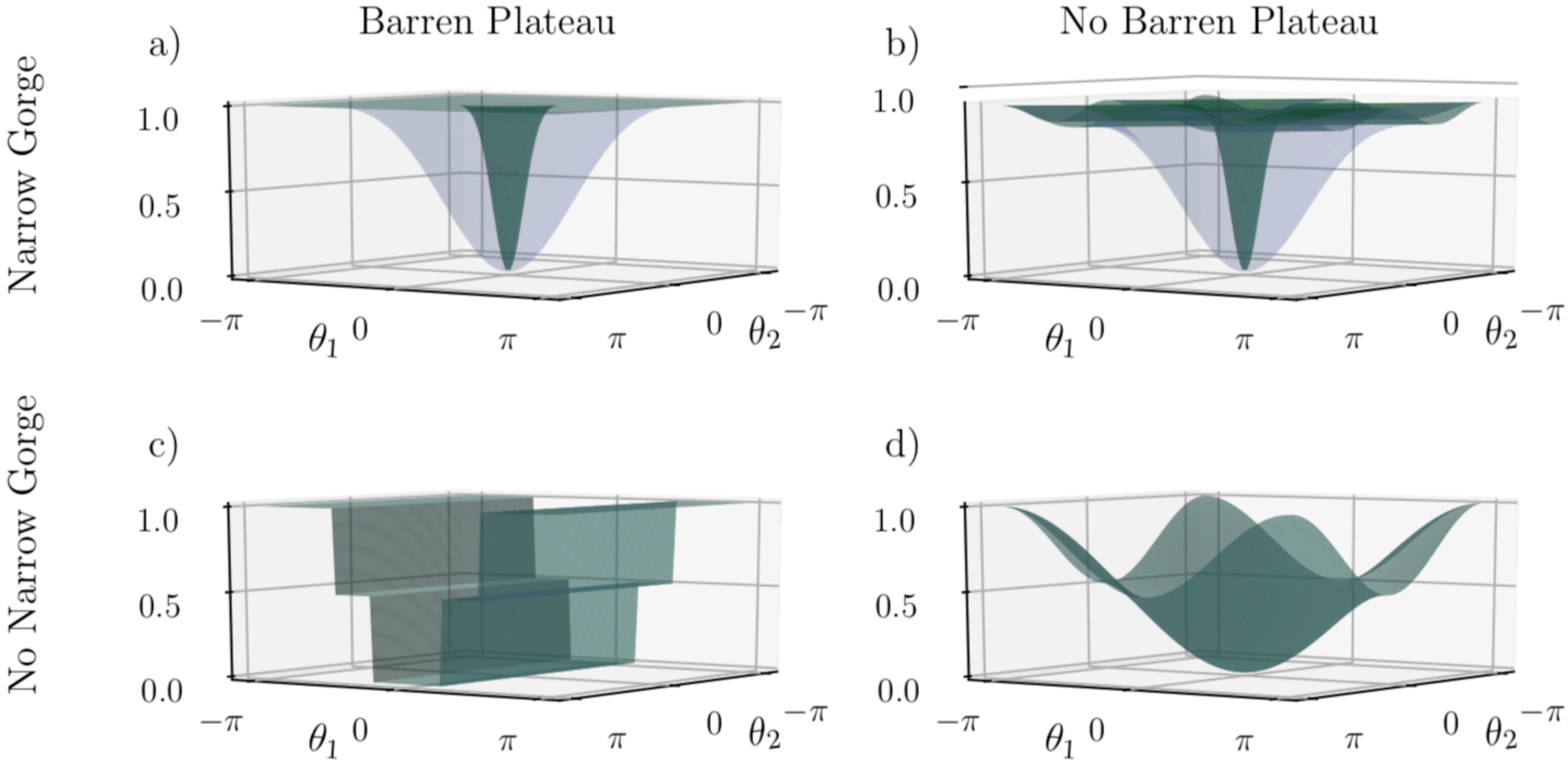
Restricting rotation angle range helps, when near solution



How to diagnose barren plateaus

Equivalence of barren plateaus to cost concentration and narrow gorges

(arxiv:2104.05868)

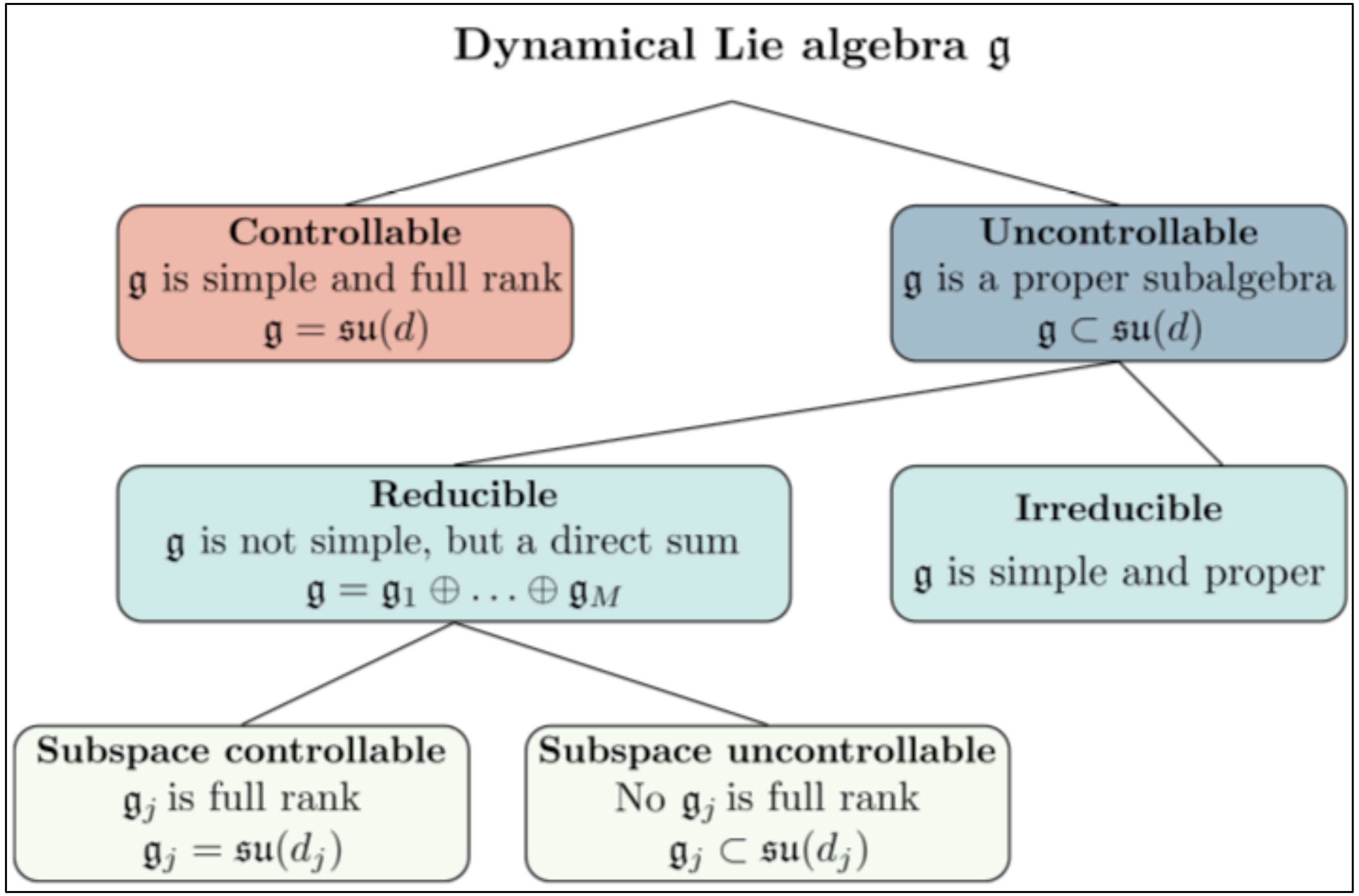


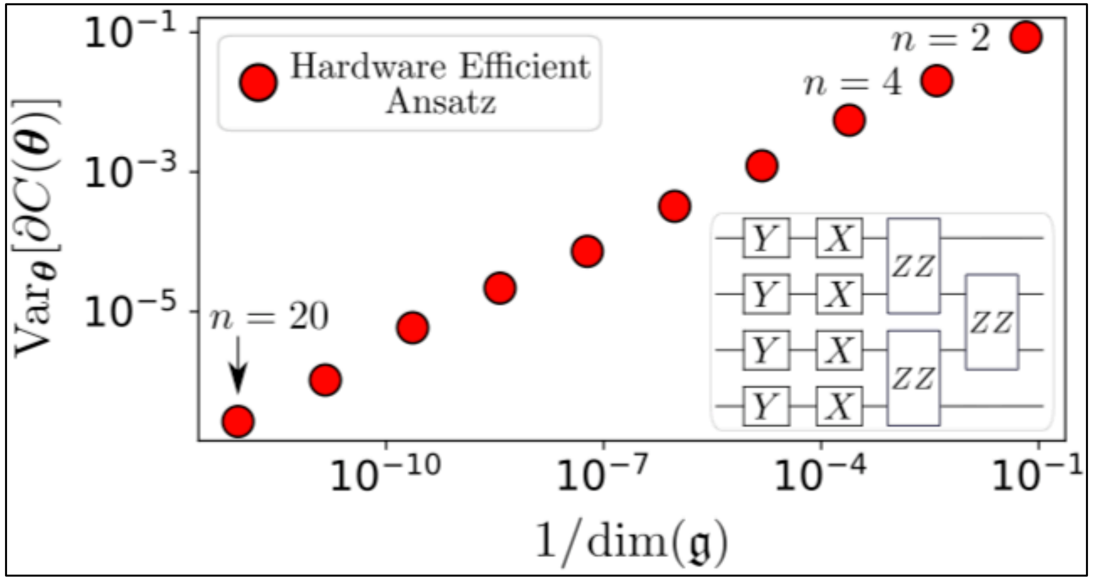
Sampling cost function differences (instead of gradients) can diagnose presence/absence of BPs

Predicting barren plateaus with tools from quantum optimal control

(arxiv:2105.14377)

Periodic ansatz: $$U(\boldsymbol{\theta}) = \prod_{l=1}^{L} U_l(\boldsymbol{\theta}_l), \quad U_l(\boldsymbol{\theta}_l) = \prod_{k=0}^{K} e^{-iH_k\theta_{lk}}$$





Proposition 1 (Controllable). *There exists a scaling of the depth for which controllable systems form ε-approximate 2-designs with $\varepsilon \in \mathcal{O}(1/2^n)$, and hence the system exhibits a barren plateau according to Definition 1.*

Observation 1. *Let the state ρ belong to a subspace $\mathcal{H}_k$ associated with a DLA $\mathfrak{g}_k$. Then, the scaling of the variance of the cost function partial derivative is inversely proportional to the scaling of the dimension of the DLA as*

$$\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})] \in \mathcal{O}\left(\frac{1}{\mathrm{poly}(\dim(\mathfrak{g}_k))}\right). \quad (24)$$

Q: Can we extend the barren plateau result to shallow depth?

A: Yes, if we allow it to depend on the cost function

Cost-Function-Dependent Barren Plateaus in Shallow Quantum Neural Networks

M. Cerezo,[1,2] Akira Sone,[1,2] Tyler Volkoff,[1] Lukasz Cincio,[1] and Patrick J. Coles[1]

[1] *Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA.*

[2] *Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM, USA*

Variational quantum algorithms (VQAs) optimize the parameters $\boldsymbol{\theta}$ of a quantum neural network $V(\boldsymbol{\theta})$ to minimize a cost function C. While VQAs may enable practical applications of noisy quantum computers, they are nevertheless heuristic methods with unproven scaling. Here, we rigorously prove two results, assuming $V(\boldsymbol{\theta})$ is a hardware-efficient ansatz composed of blocks forming local 2-designs. Our first result states that defining C in terms of global observables leads to an exponentially vanishing gradient (i.e., a barren plateau) even when $V(\boldsymbol{\theta})$ is shallow. This implies that several VQAs in the literature must revise their proposed cost functions. On the other hand, our second result states that defining C with local observables leads to at worst a polynomially vanishing gradient, so long as the depth of $V(\boldsymbol{\theta})$ is $\mathcal{O}(\log n)$. Taken together, our results establish a connection between locality and trainability. Finally, we illustrate these ideas with large-scale simulations, up to 100 qubits, of a particular VQA known as quantum autoencoders.

I. Introduction

One of the most important technological questions is whether Noisy Intermediate-Scale Quantum (NISQ) computers will have practical applications [1]. NISQ devices are limited both in qubit count and in gate fidelity, hence preventing the use of quantum error correction.

The leading strategy to make use of these devices are variational quantum algorithms (VQAs) [2]. VQAs employ a quantum computer to efficiently evaluate a cost function C, while a classical optimizer trains the parameters $\boldsymbol{\theta}$ of a parameterized quantum circuit $V(\boldsymbol{\theta})$. The latter represents a quantum generalization of a neural network, i.e., a quantum neural network (QNN). The benefits of VQAs are three-fold. First, VQAs al-

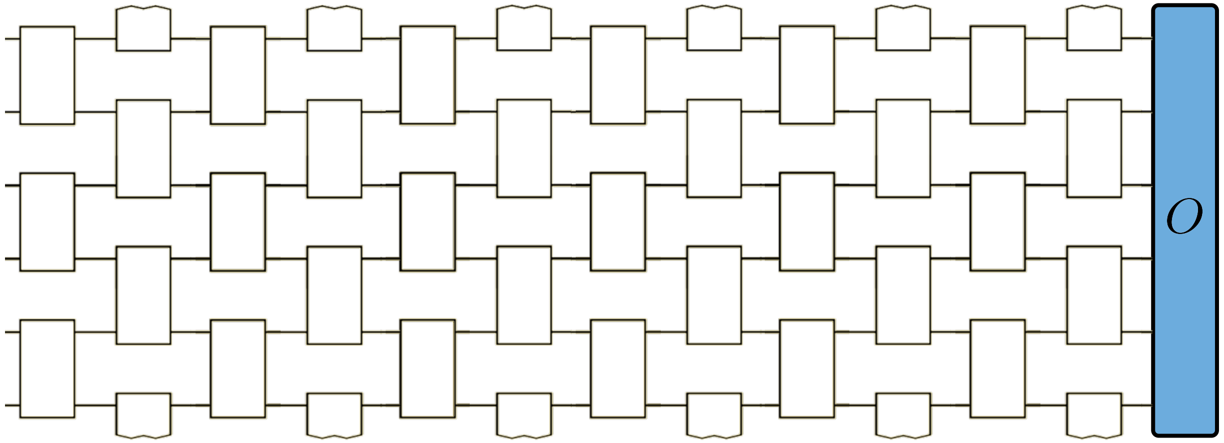
obviously the energy $C = \langle\psi|H|\psi\rangle$ of the trial state $|\psi\rangle$. However, VQAs have been proposed for other applications, like quantum data compression [9], quantum error correction [10], quantum metrology [11], quantum compiling [12–15], quantum state diagonalization [16, 17], quantum simulation [18–21], fidelity estimation [22], unsampling [23], consistent histories [24], and linear systems [25–27]. For these applications, the choice of C is less obvious. Put another way, if one reformulates these VQAs as ground-state problems (which can be done in many cases), the choice of Hamiltonian H is less intuitive. This is because many of these applications are abstract, rather than associated with a physical Hamiltonian.

Here we connect the trainability of VQAs to the choice of C. For the abstract applications in Refs. [9–27], it is important for C to be operational, so that small values of

Barren Plateaus: Locality of the cost function [Cerezo et al., 2020]

Setting:

Blocks of local 2-designs



Global cost function:

$$O = c_0 \mathbb{1} + \sum_{i=1}^{N} c_i \widehat{O}_{i1} \otimes \widehat{O}_{i2} \otimes \cdots \otimes \widehat{O}_{i\xi}$$

Local cost function:

$$O = c_0 \mathbb{1} + \sum_{i=1}^{N} c_i \widehat{O}_i^{\mu_i} \otimes \widehat{O}_i^{\mu_i'}$$

Barren Plateaus: Locality of the cost function [Cerezo et al., 2020]

Results:

Global cost function:

$$O = c_0 \mathbb{1} + \sum_{i=1}^{N} c_i \widehat{O}_{i1} \otimes \widehat{O}_{i2} \otimes \cdots \otimes \widehat{O}_{i\xi}$$

$\mathrm{Var}\left[\partial_\nu C\right]$ exponentially vanishes for

$$L \in \mathcal{O}(\mathrm{poly}(\log(n)))$$

Local cost function:

$$O = c_0 \mathbb{1} + \sum_{i=1}^{N} c_i \widehat{O}_i^{\mu_i} \otimes \widehat{O}_i^{\mu_i'}$$

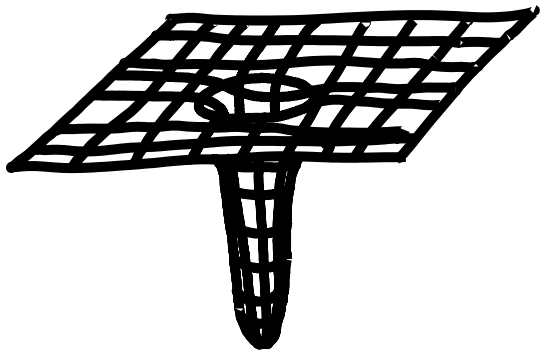
$\mathrm{Var}\left[\partial_\nu C\right]$ at worst polynomially decreases for

$$L \in \mathcal{O}(\log(n))$$

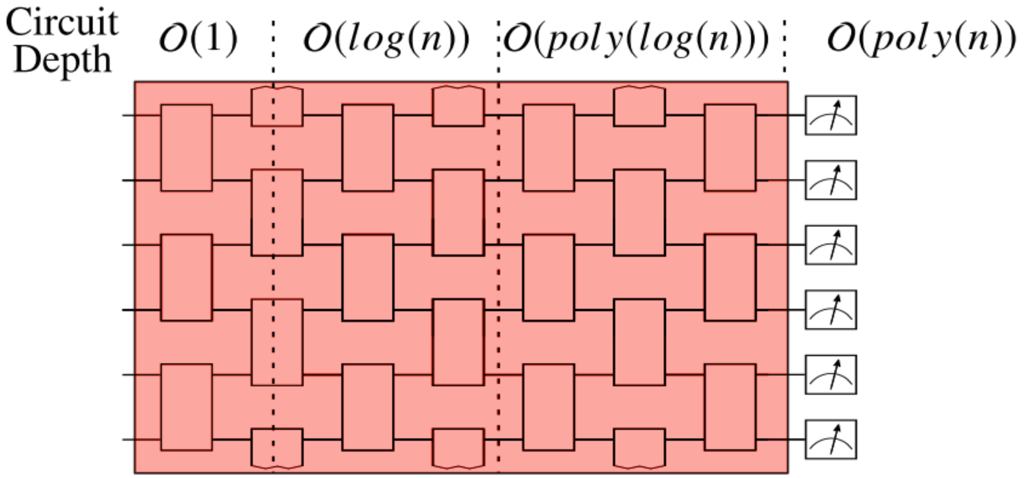
Barren Plateaus: Summary

$\mathrm{Var}\left[\partial_\nu C\right]$ exponentially vanishes for

$$L \in \Omega(n)$$



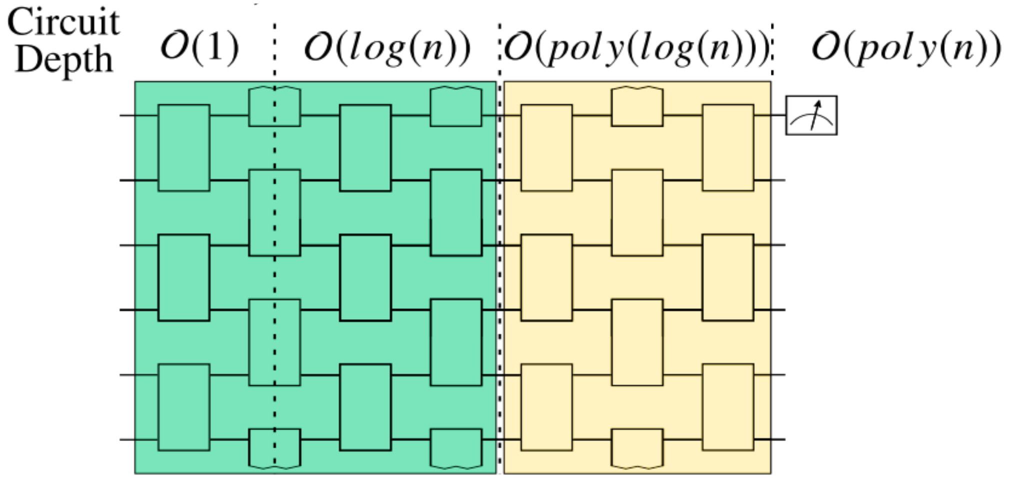
Global cost function:



$\mathrm{Var}\left[\partial_\nu C\right]$ exponentially vanishes for

$$L \in \mathcal{O}(\mathrm{poly}(\log(n)))$$

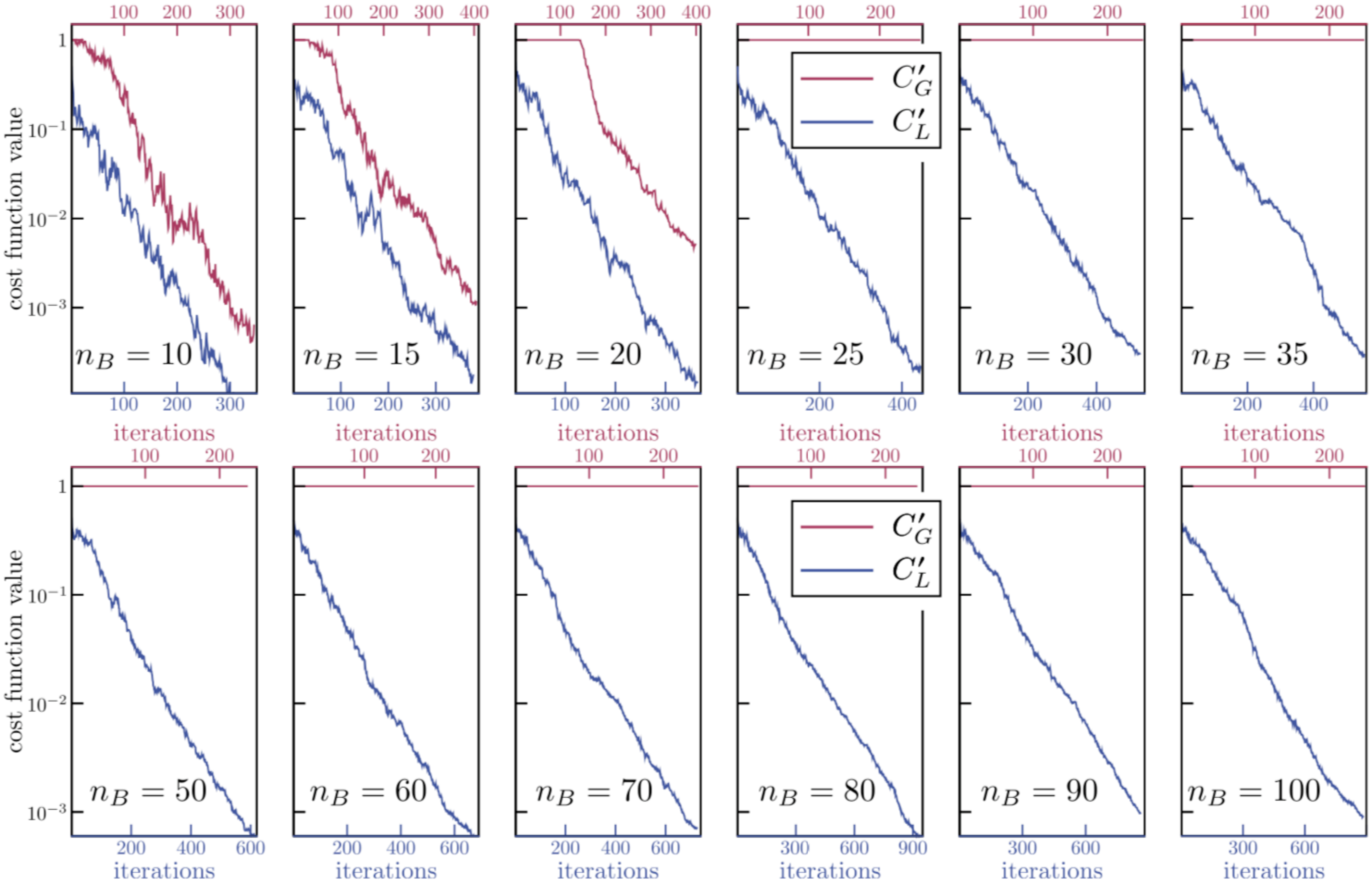
Local cost function:



$\mathrm{Var}\left[\partial_\nu C\right]$ at worst polynomially decreases for

$$L \in \mathcal{O}(\log(n))$$

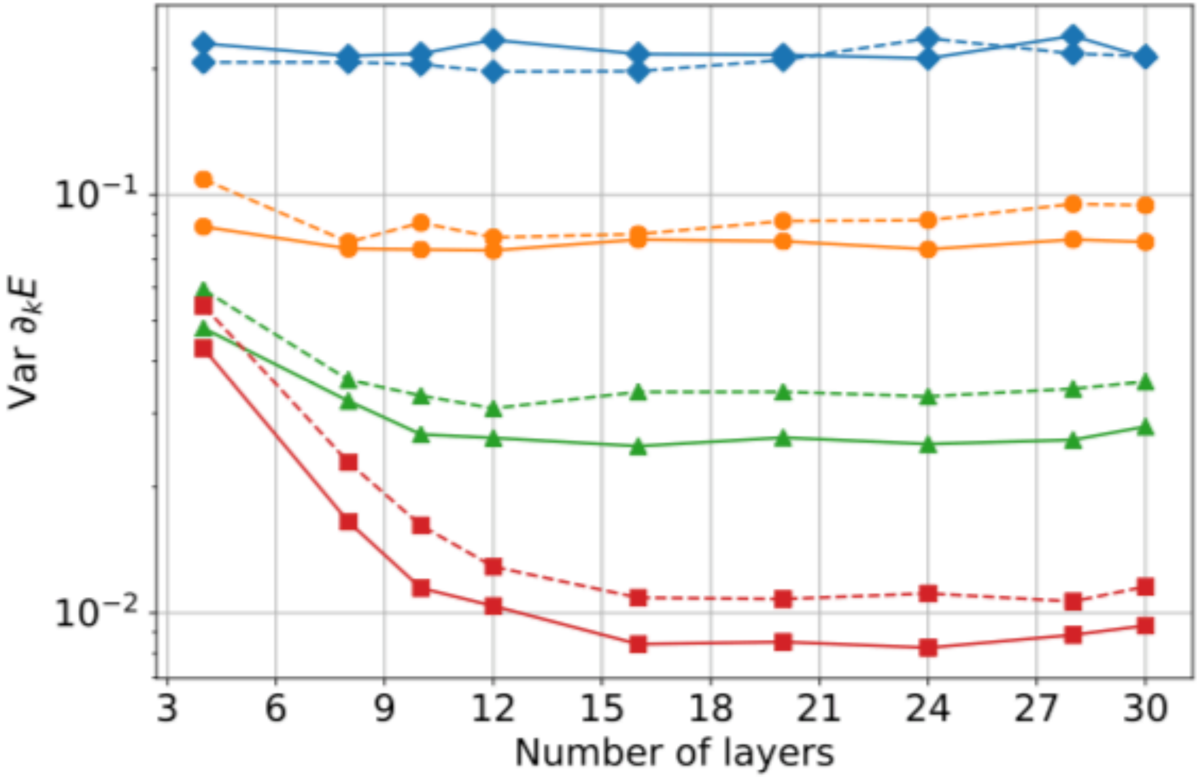
Example: Quantum Autoencoders (Romero et al. Quantum Sci. Tech. (2017))



Implication: fermion-to-qubit mapping in VQE

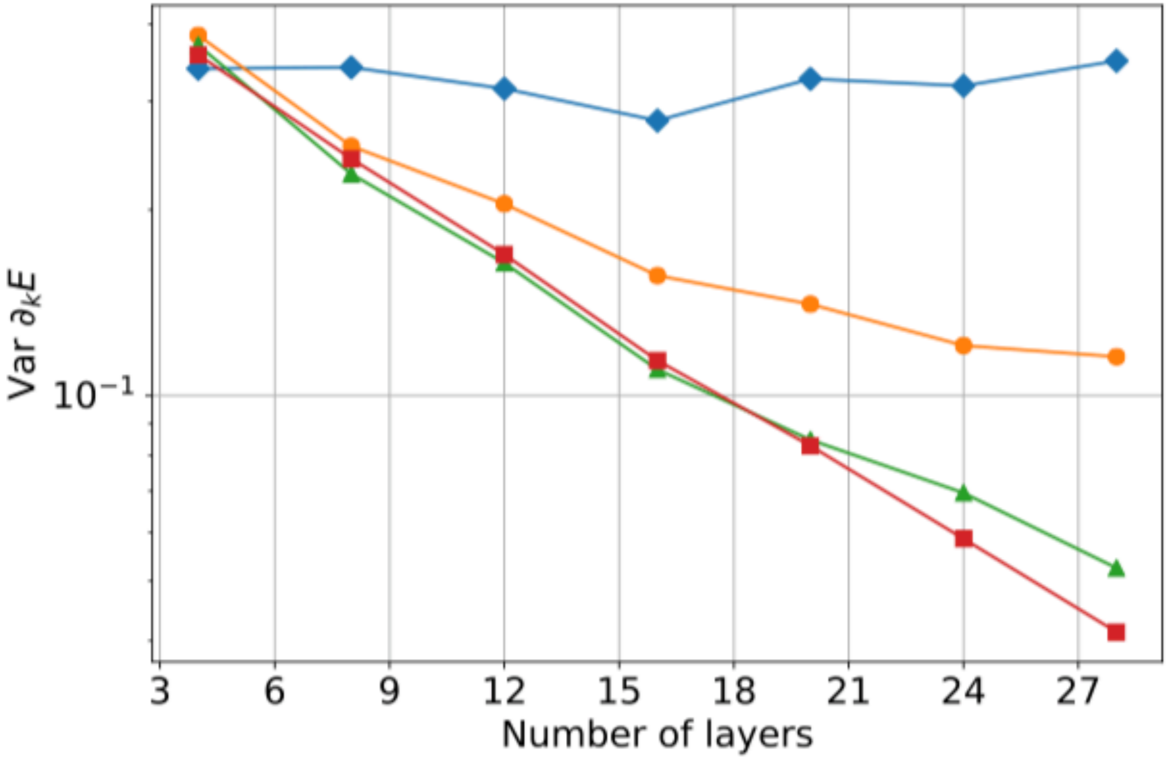
Confirmed by numerics of Moscow group:
"Variational Quantum Eigensolver for Frustrated Quantum Systems"
Uvarov, Biamonte, Yudin (2020)

JW Transformation



BP for Shallow Depth

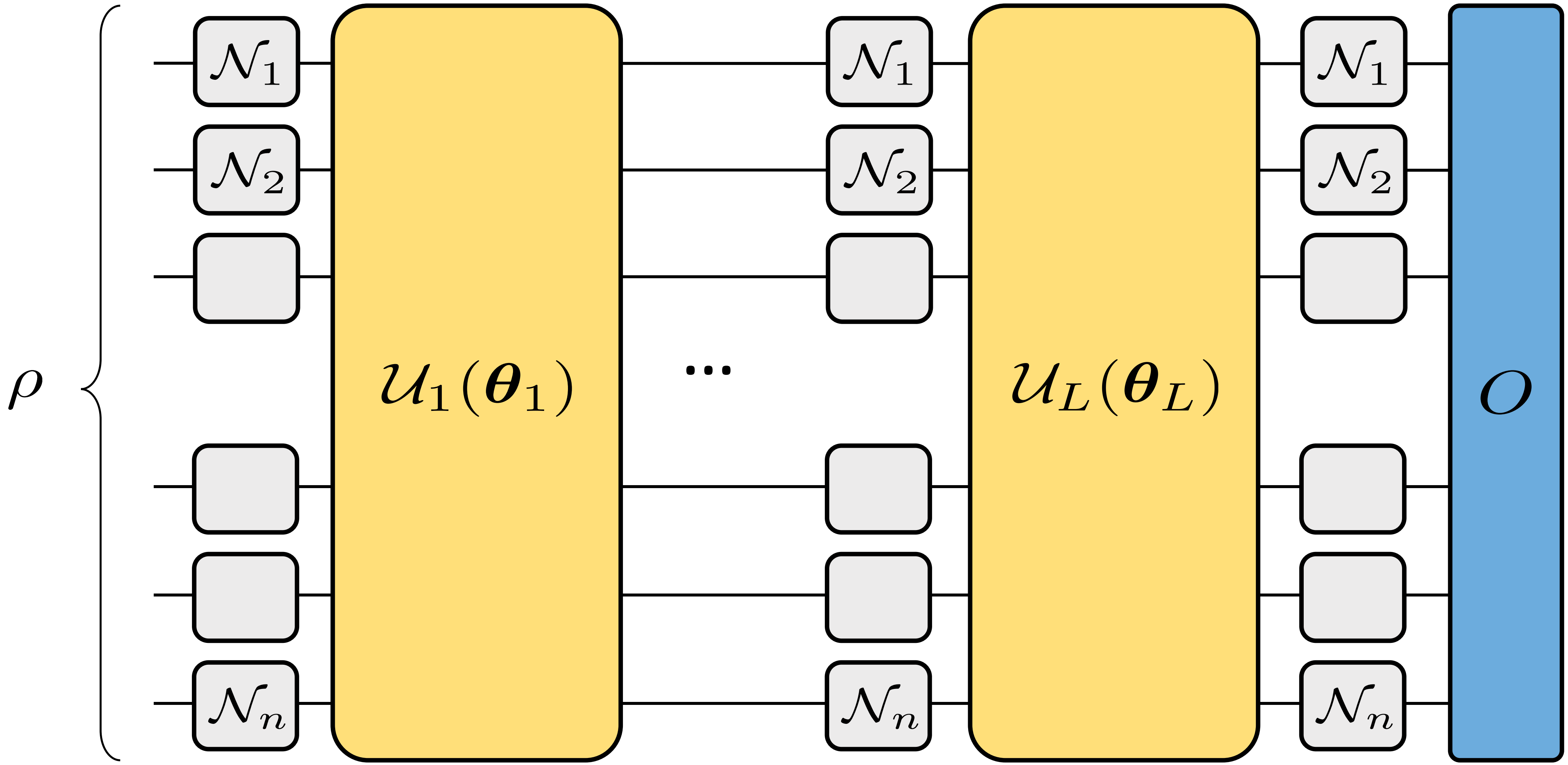
BK Transformation



No BP for Shallow Depth

Noise-induced barren plateaus

Wang, Fontana, Cerezo, Sharma, Sone, Cincio, Coles. *Nature Comm.* (2021)



Noise-induced barren plateaus

Wang, Fontana, Cerezo, Sharma, Sone, Cincio, Coles. *Nature Comm.* (2021)

Main result:

We have

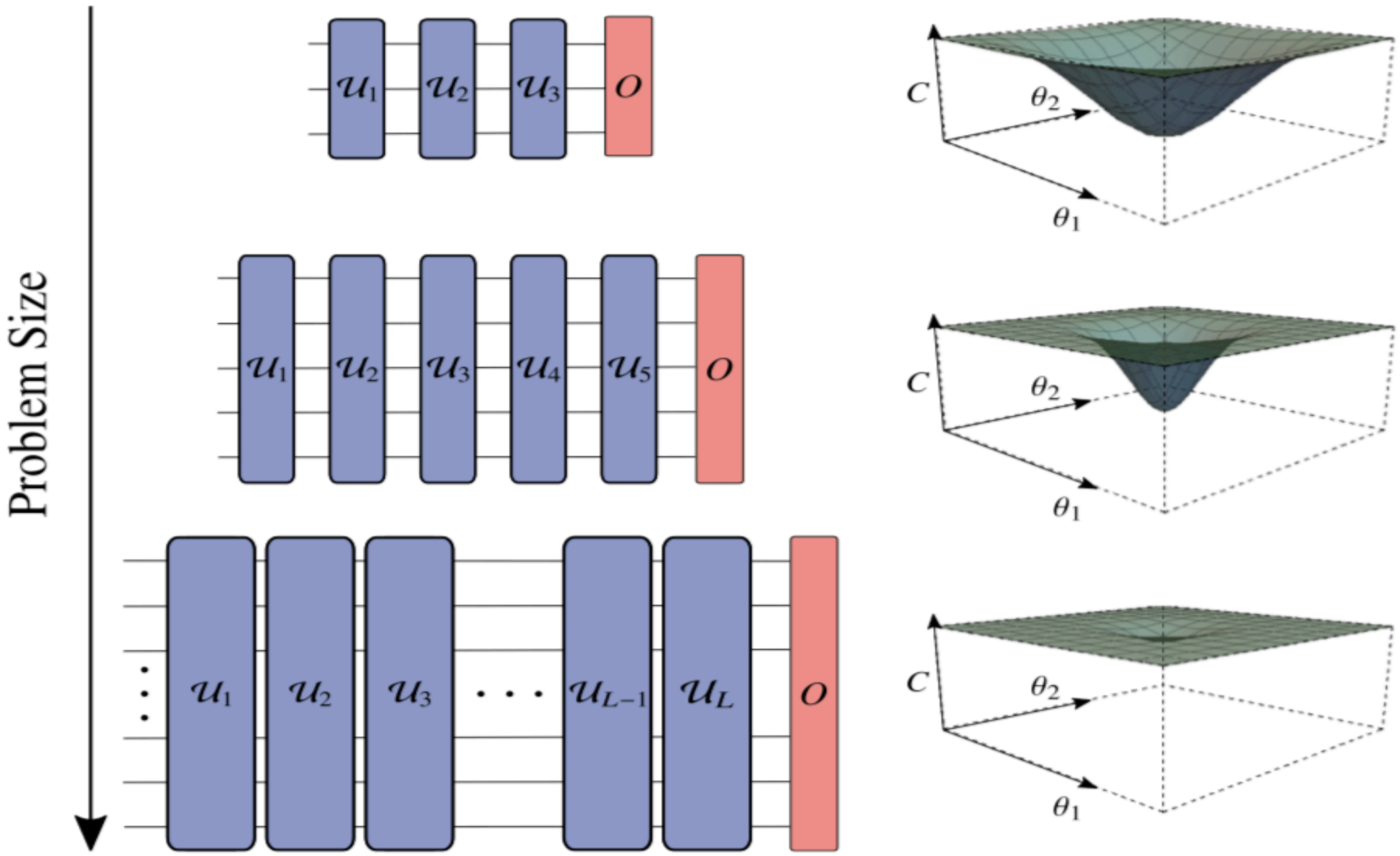
$$|\partial_{lm}\widetilde{C}| \leqslant F(n)$$

where

$$F(n) \in \mathcal{O}\left(2^{-\alpha n}\right)$$

if

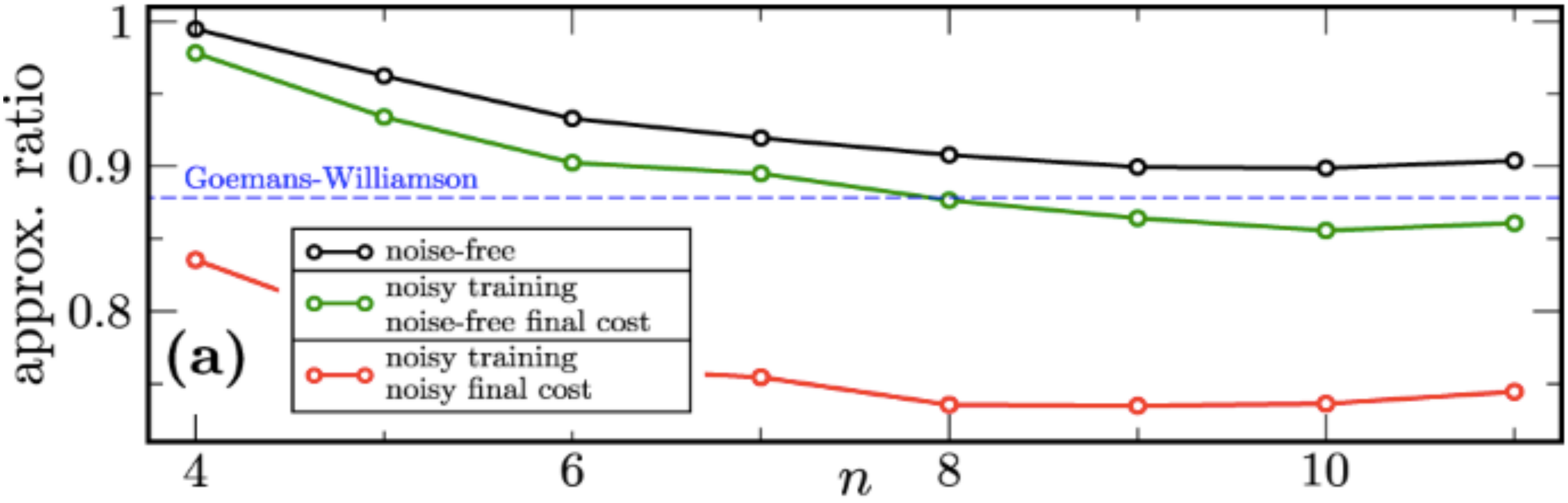
$$L \in \Omega(n)$$



Impacts problem-inspired ansatzes for VQE and QAOA!

NIBPs can destroy quantum advantage

QAOA numerics for MaxCut:

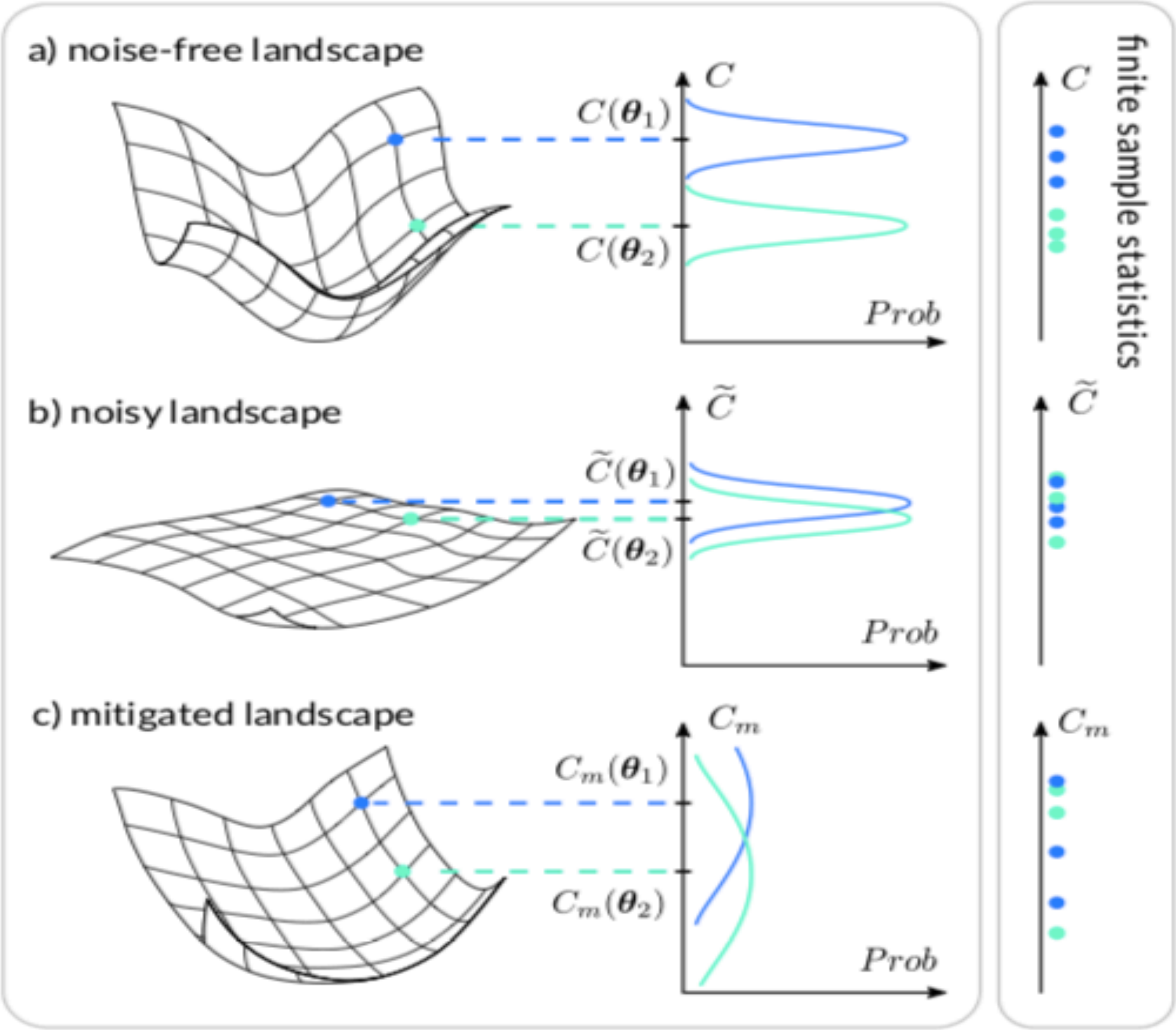


Can error mitigation solve NIBPs?

Wang, Czarnik, Arrasmith, Cerezo, Cincio, Coles. *Arxiv:2109.01051*

Answer: No!

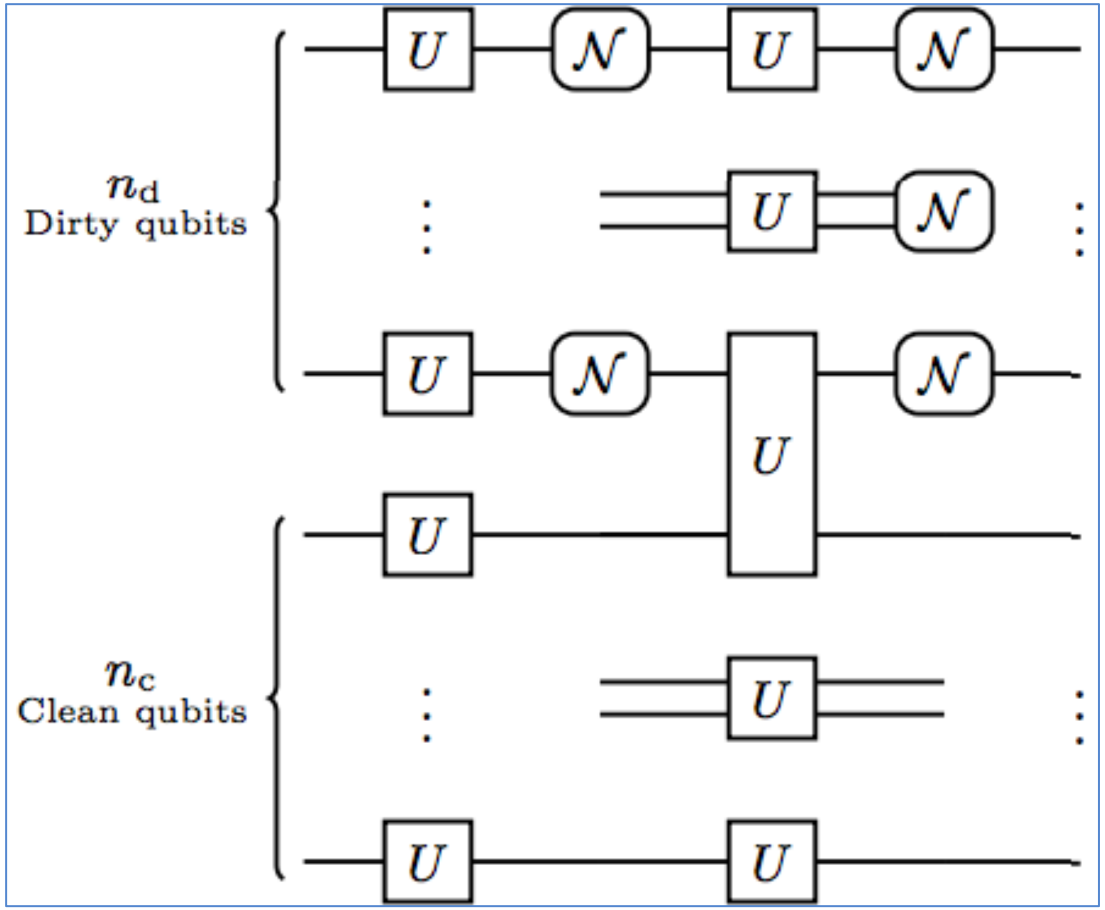
Exponential scaling of resources still occurs

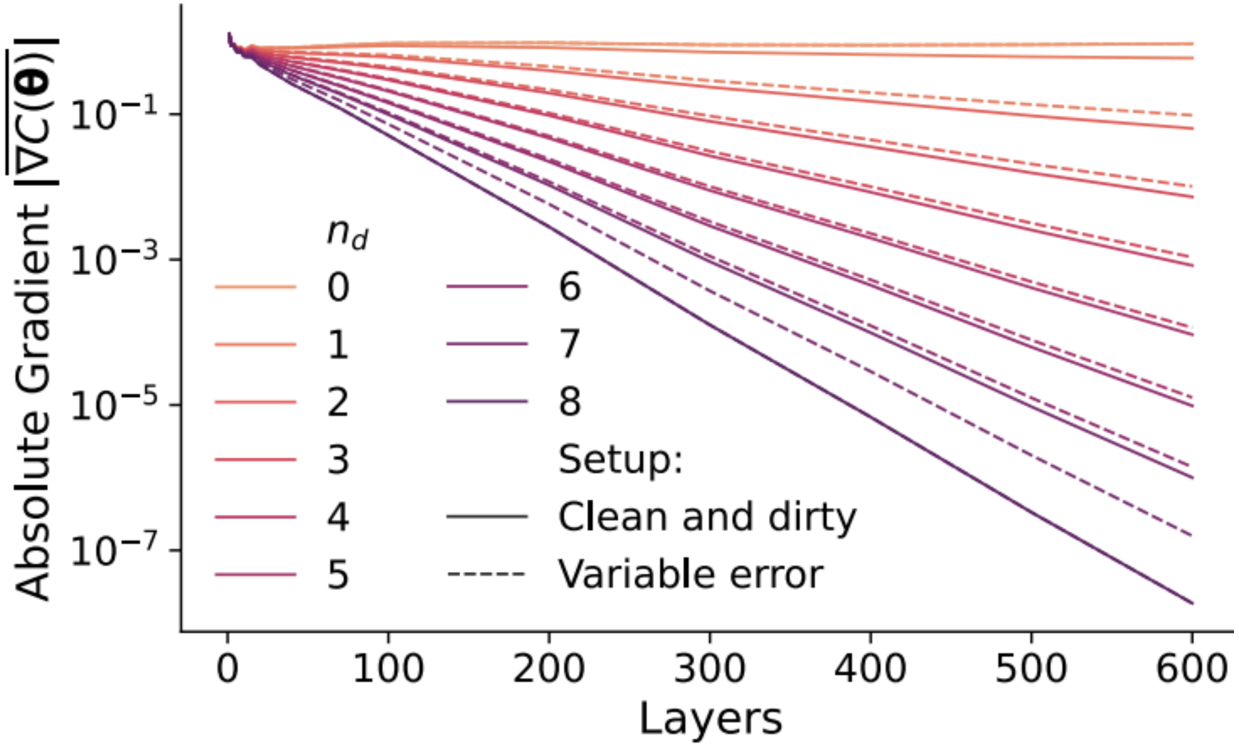


Can error correction solve NIBPs?

Bultrini, Wang, Czarnik, Hunter Gordon, Cerezo, Coles, Cincio. *Arxiv:2205.13454*

Clean and dirty model

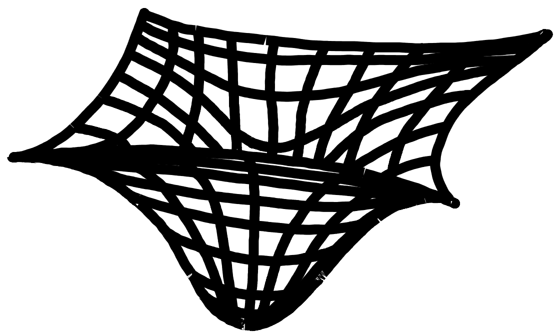




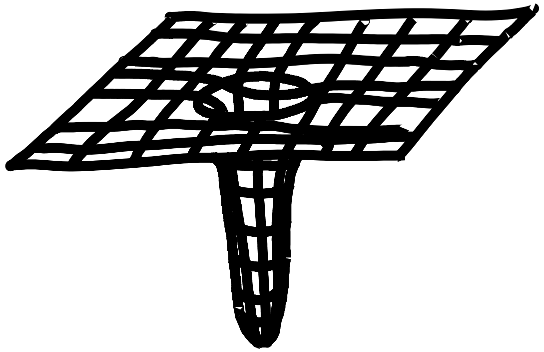
Each clean qubit exponentially increases gradient sizes

Conclusions

- Barren plateaus are one of the few complexity theoretic results we have for VQAs
- Barren plateaus (BPs) can determine whether or not quantum speedup is possible
- Trainability is connected to (1) expressibility, (2) controllability, (3) locality, (4) noise
- Strategies to avoid barren plateaus is a crucial area of research



Trainable landscape



Barren plateau landscape