Classical Splitting of Parametrized Quantum Circuits

Cenk Tüysüz

DESY CQTA, Humboldt-Universität zu Berlin

in collaboration with;

Giuseppe Clemente, Arianna Crippa, Tobias Hartung, Stefan Kühn, Karl Jansen

For more details:

arXiv: 2206.09641

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

DESY.

Variational Quantum Algorithms

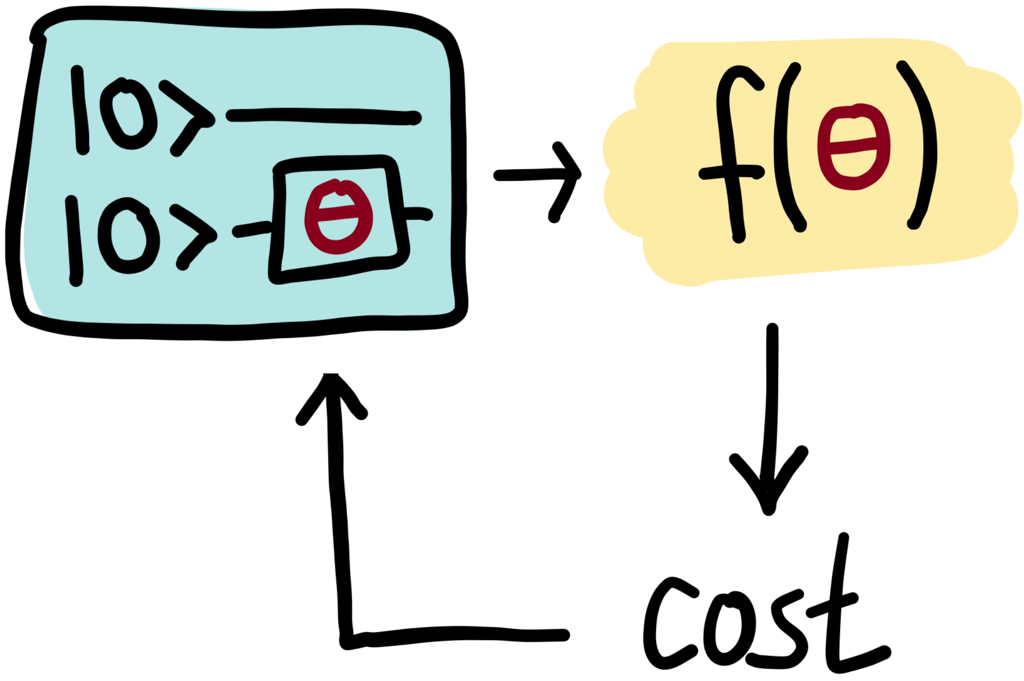


Image credit: https://pennylane.ai/qml/glossary/variational_circuit.html

Measuring Expectation values

Hoeffding's inequality defines the number of shots required to estimate $E[\hat{X}]$ to ϵ-precision with $1\text{-}\delta$ probability as;

$$N_{\text{hff}}^{\epsilon,\delta}(\hat{X}) = \frac{\ln(2/\delta)}{2\epsilon^2}\mathrm{R}[\hat{X}]^2$$

This means that we need to define a precision. How can we do this?

Measuring gradients

Parameter shift rule

→ The optimizer needs to obtain gradients to update the circuit parameters.

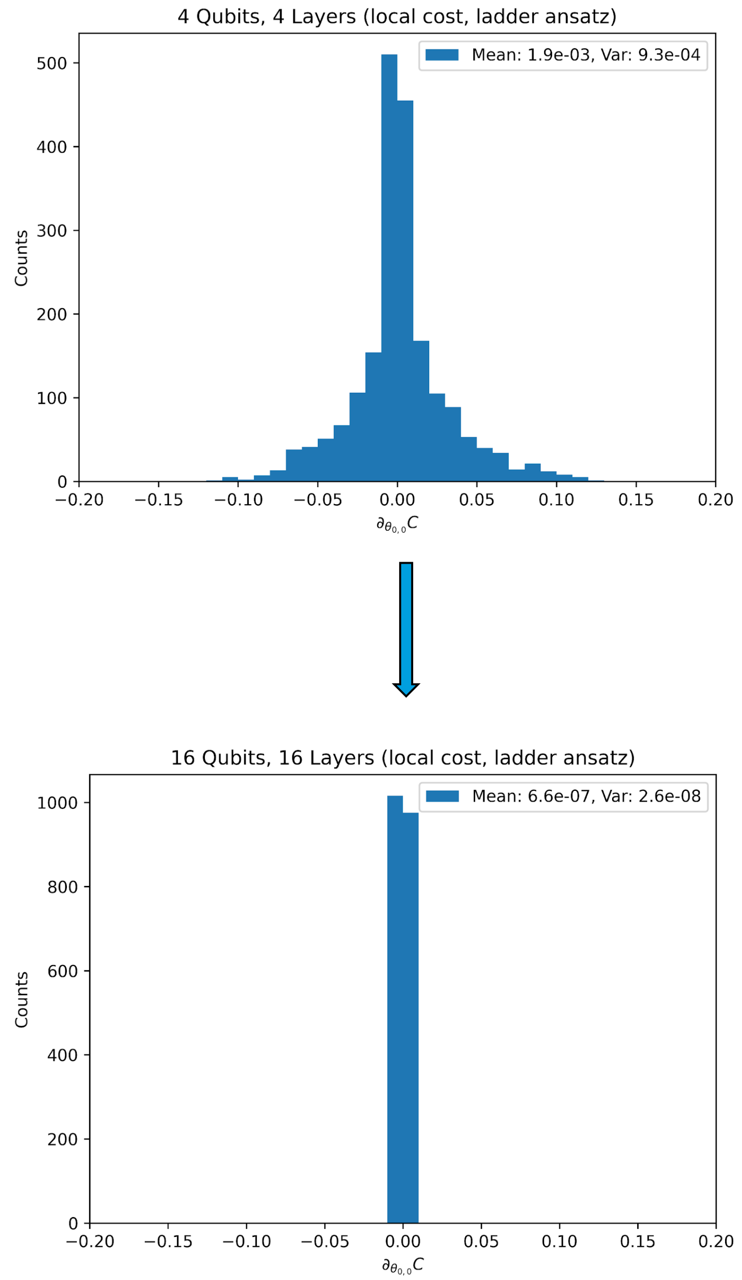
The parameter shift rule allows us to measure gradients of circuit parameters with respect to the expectation values.

$$\frac{d}{d\theta} f(\theta) = r\left[f(\theta + \tfrac{\pi}{4r}) - f(\theta - \tfrac{\pi}{4r})\right]$$

r=1/2 for simple Pauli rotation gates.

More details on this: Crooks 2019 arXiv: 1905.13311

Histogram of gradients of a single parameter

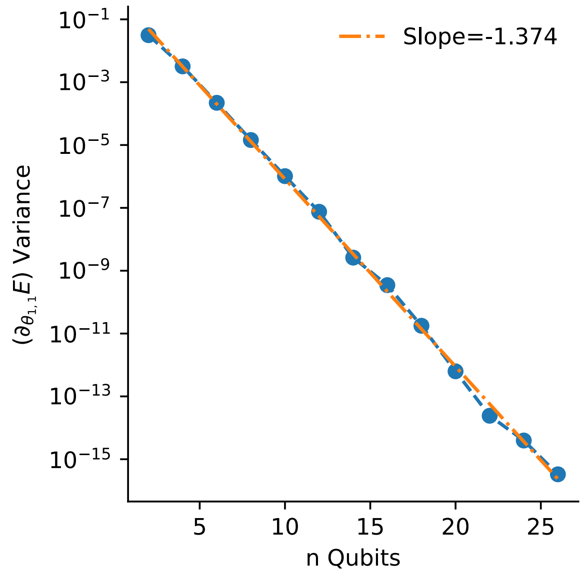


Barren Plateaus

What are they?

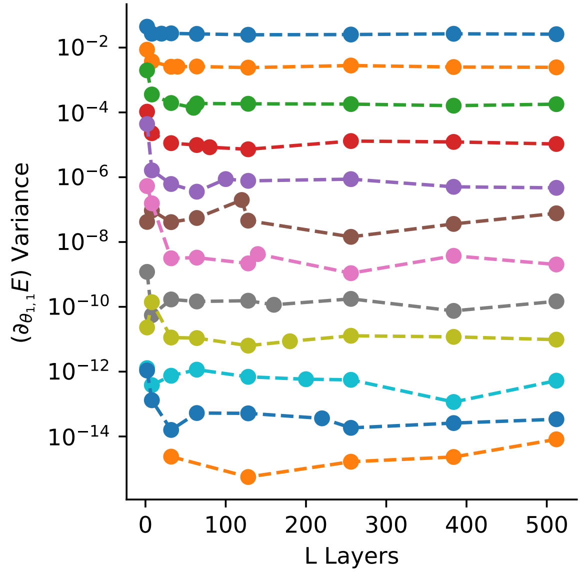
Barren Plateaus (BPs) are the exponential decay of gradient sizes with respect to a hyper-parameter of the model (e.g. number of qubits, number of layers, etc.).

McClean et al. 2018, Nature Comm.



→ Gradients of a VQA are centered around zero

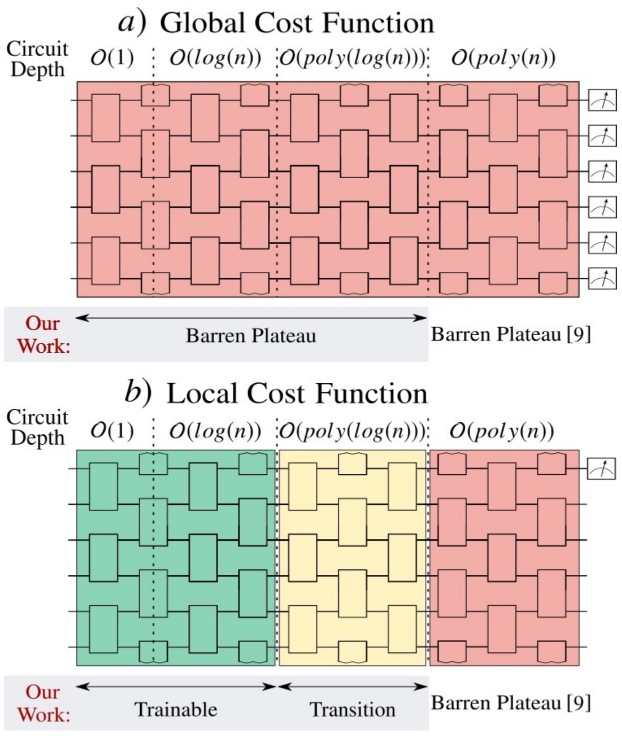
→ Variance of the gradients inform us about the flatness of the landscape.



Barren Plateaus

What causes them?

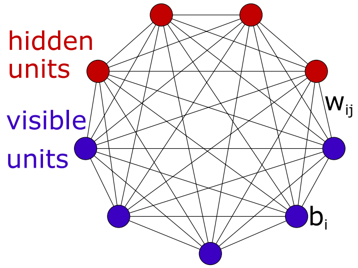
Global cost functions



Cerezo et al. 2021, Nature Comm.

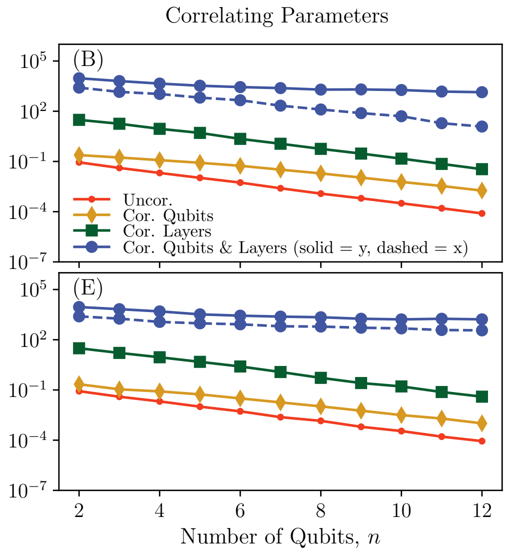
Entanglement

Volume law entanglement between hidden and visible units cause BPs



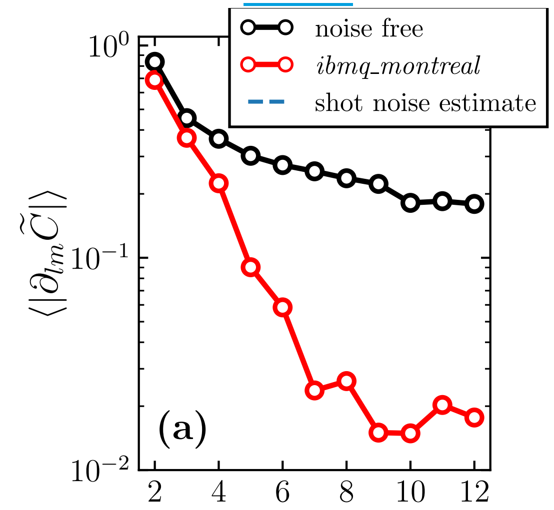
Ortiz Marrero et al. 2021, PRX Quantum

Expressivity



Holmes et al. 2022, PRX Quantum

Noise



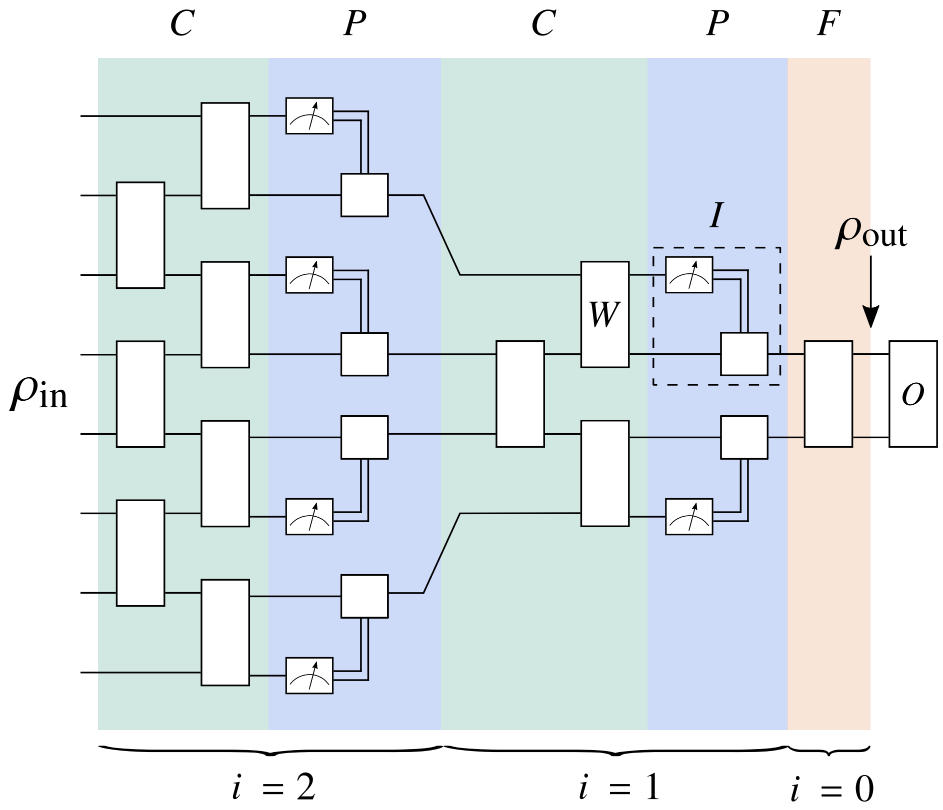
Wang et al. 2021, Nature Comm.

Barren Plateaus

How to avoid/mitigate them?

Some circuits do not exhibit BPs (e.g. Quantum Convolutional Neural Networks (QCNNs))

Pesah et al. 2021, Phys. Rev. X

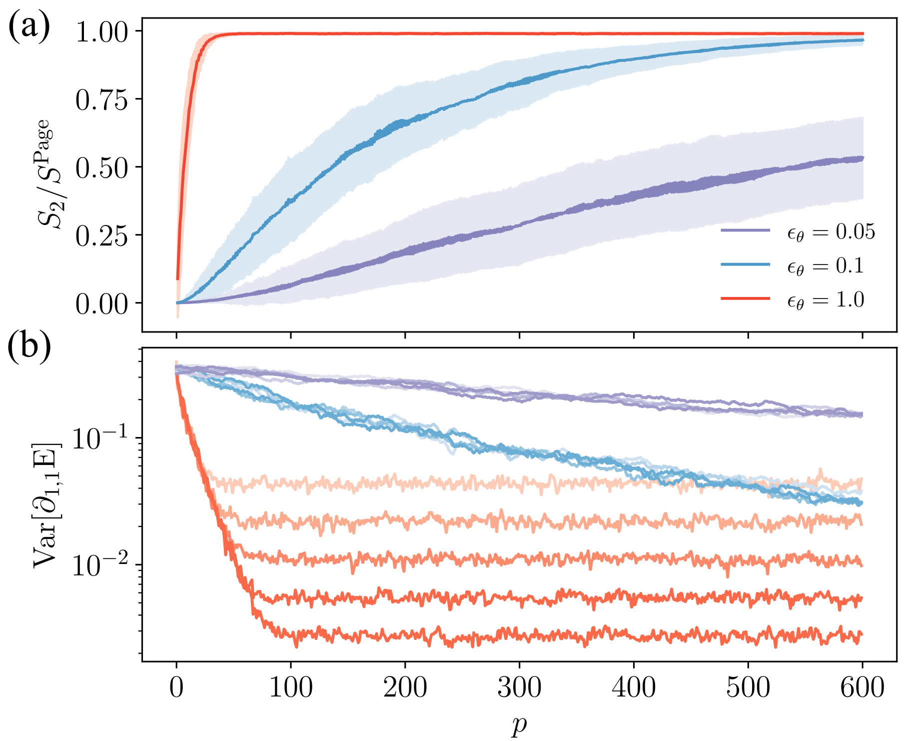


There are many methods suggested to avoid BPs by certain initialization methods. *Grant et al. 2019, Quantum*

→ These methods, in general, do not ensure trainability after the initialization.

Weak BPs were suggested as another measure that can be used to track if the model falls in to BPs.

→ Expensive (requires state tomography) but effective



Sack et al. 2022, PRX Quantum

Classical Splitting

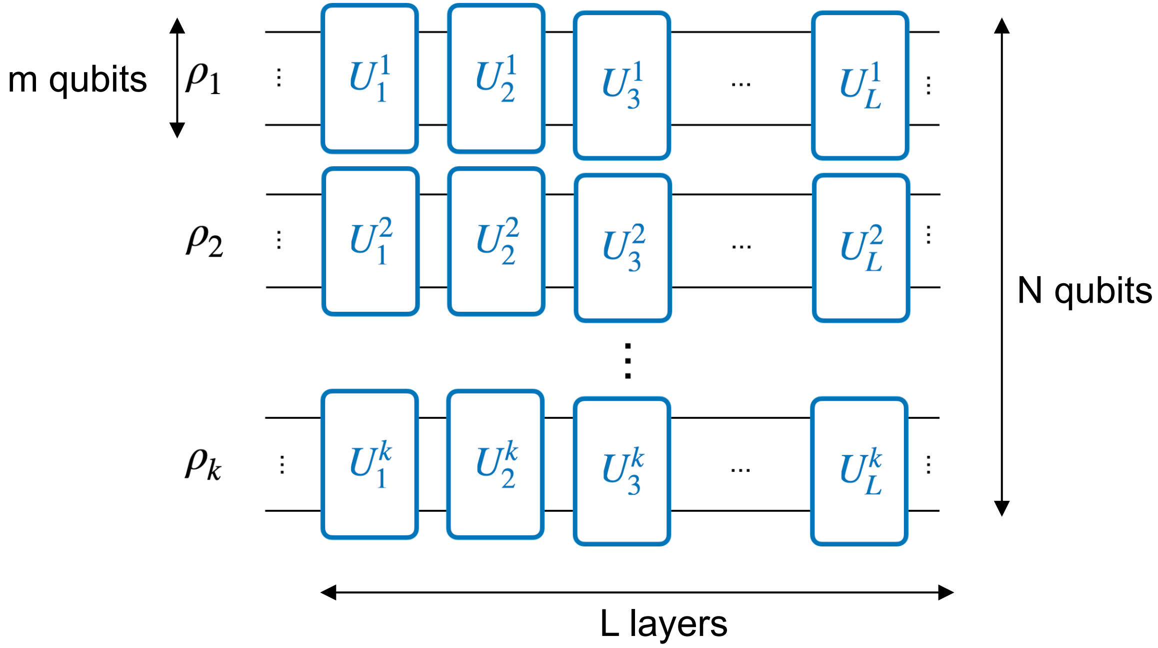
Split an N qubit ansatz to k many m qubit circuit.

Classical Splitting (CS) Ansatz:

k: number of circuits

m: number of local qubits

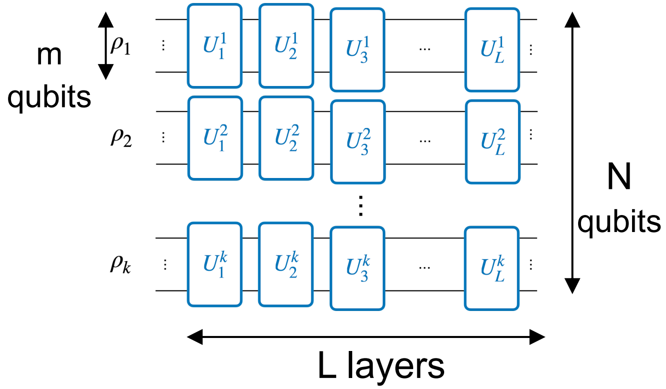
N: total number of qubits



Cost function becomes a sum of each subcircuit;

$$
\begin{aligned}
C(\boldsymbol{\theta}) &= \sum_{i=1}^{k} \mathrm{Tr}[\bigotimes_{j=1}^{k} ((O_i - \mathbb{1})\, \delta_{i,j} + \mathbb{1})\, U^j(\boldsymbol{\theta}^j)\rho_j U^{j\dagger}(\boldsymbol{\theta}^j)] \\
&= \sum_{i=1}^{k} \prod_{j=1}^{k} \mathrm{Tr}[((O_i - \mathbb{1})\, \delta_{i,j} + \mathbb{1})\, U^j(\boldsymbol{\theta}^j)\rho_j U^{j\dagger}(\boldsymbol{\theta}^j)] \\
&= \sum_{i=1}^{k} \mathrm{Tr}[O_i U^i(\boldsymbol{\theta}^i)\rho_i U^{i\dagger}(\boldsymbol{\theta}^i)].
\end{aligned}
$$

Classical Splitting

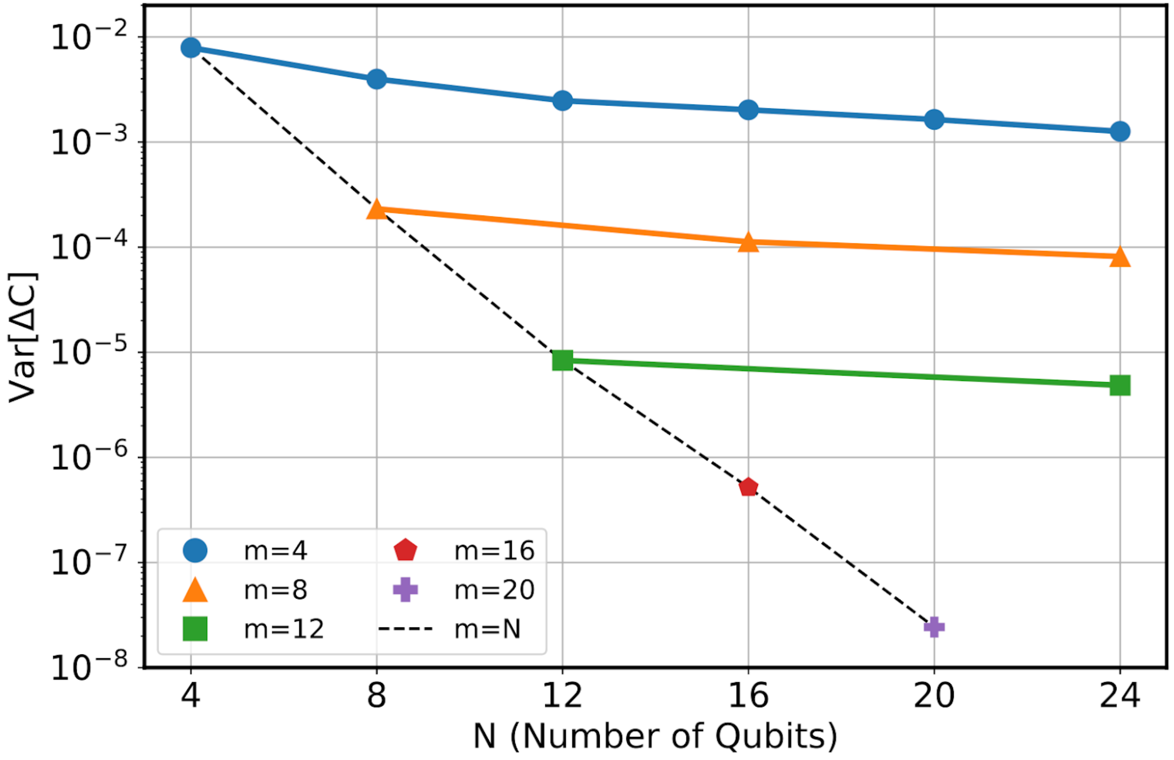


An ansatz with a local cost function has BPs for depth > O(logN). This means the variance of gradients decays <u>exponentially</u> with N;

$$\mathrm{Var}[\partial_j C(\boldsymbol{\theta})] \approx \mathcal{O}\left(\frac{1}{2^{6N}}\right)$$

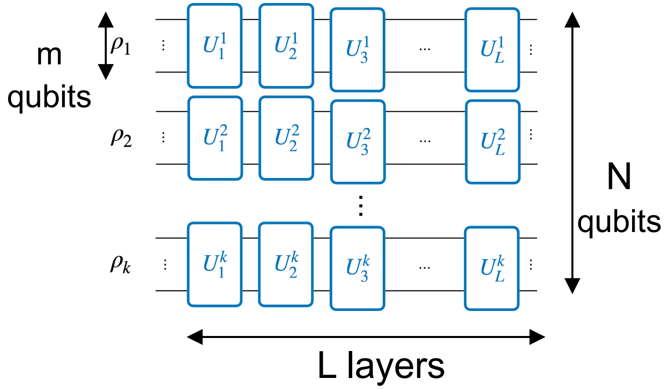
If we have classical splitting with $\mathrm{m} = \beta log_{\gamma} N$, this becomes <u>polynomial</u>;

$$\mathrm{Var}[\partial_j C(\boldsymbol{\theta})] \approx \mathcal{O}\left(\frac{1}{2^{(6m)}}\right) = \mathcal{O}\left(\frac{1}{N^{6\beta \log_{\gamma} 2}}\right)$$



Performance of the CS Ansatz

Binary Classification on Classical Data



We perform binary classification on a synthetic dataset with increasing number of features (number of features = number of qubits). We choose an ansatz with linear depth (L=N) and present validation accuracy results after training.

<u>No classical splitting (m=N):</u>

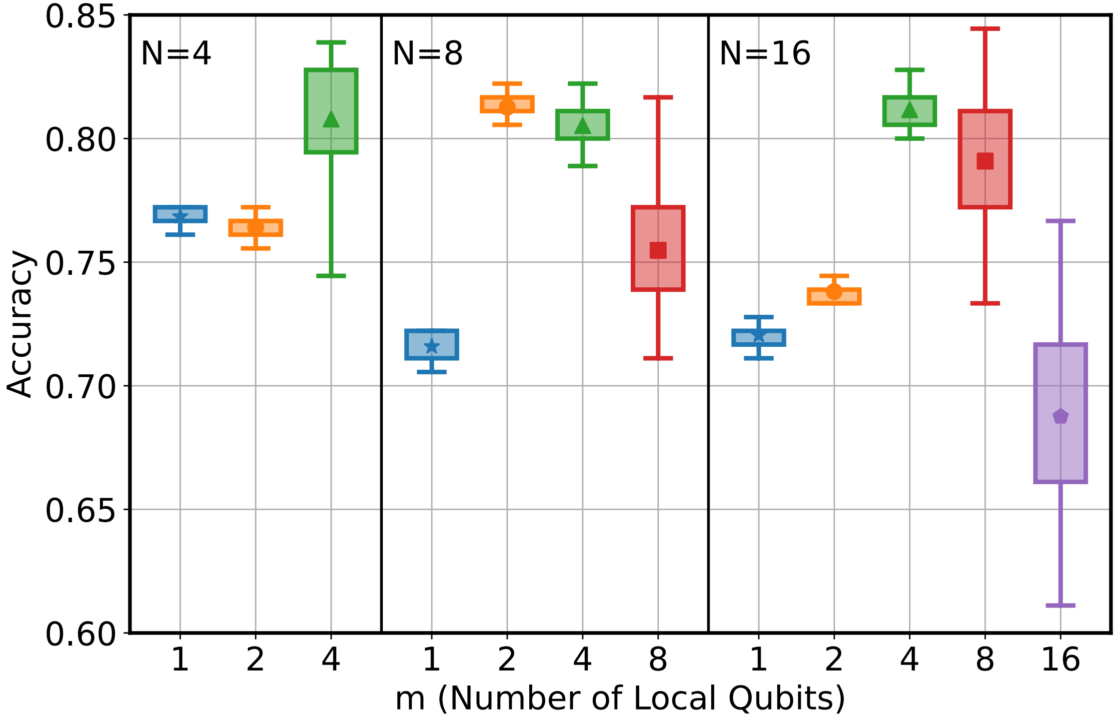
→ Accuracy drops rapidly with increasing N.

<u>With classical splitting (m<N):</u>

→ Accuracy stays stable for increasing N for $\boldsymbol{m = log_2 N}$.

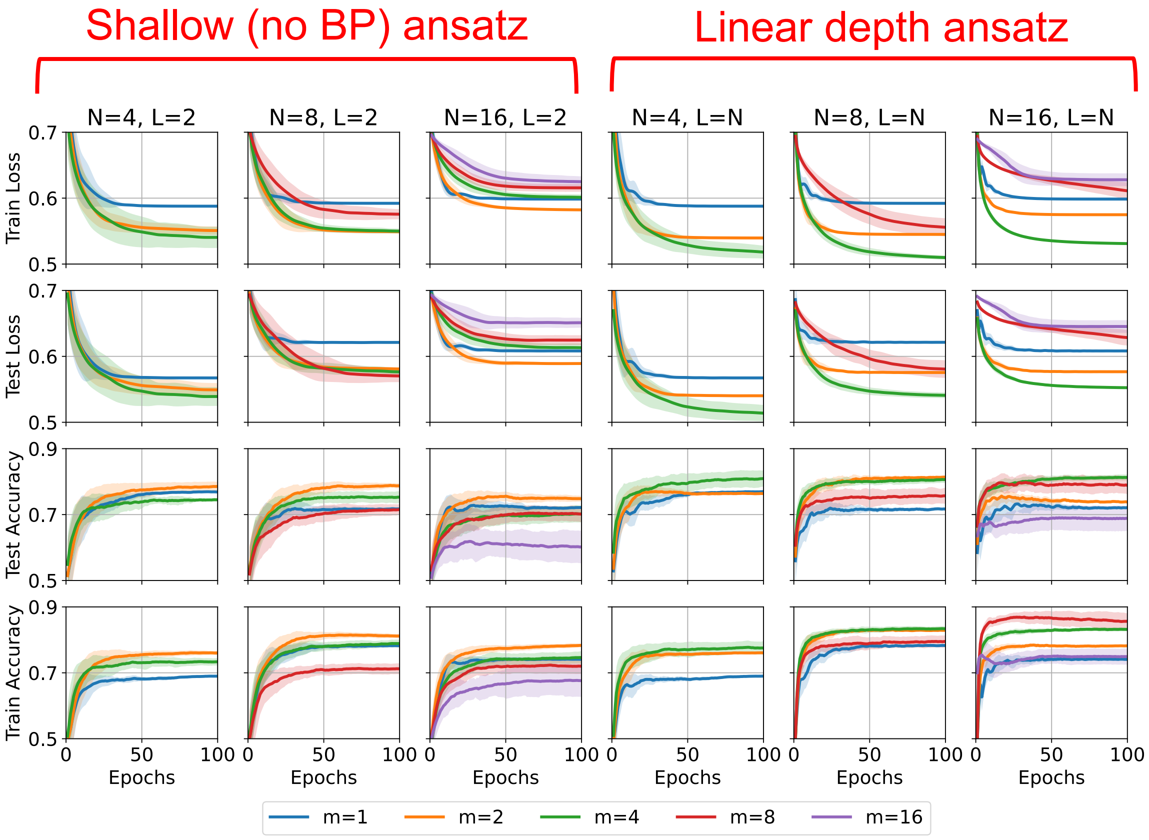
→ Classical splitting shows benefit even for $m = N/2$.

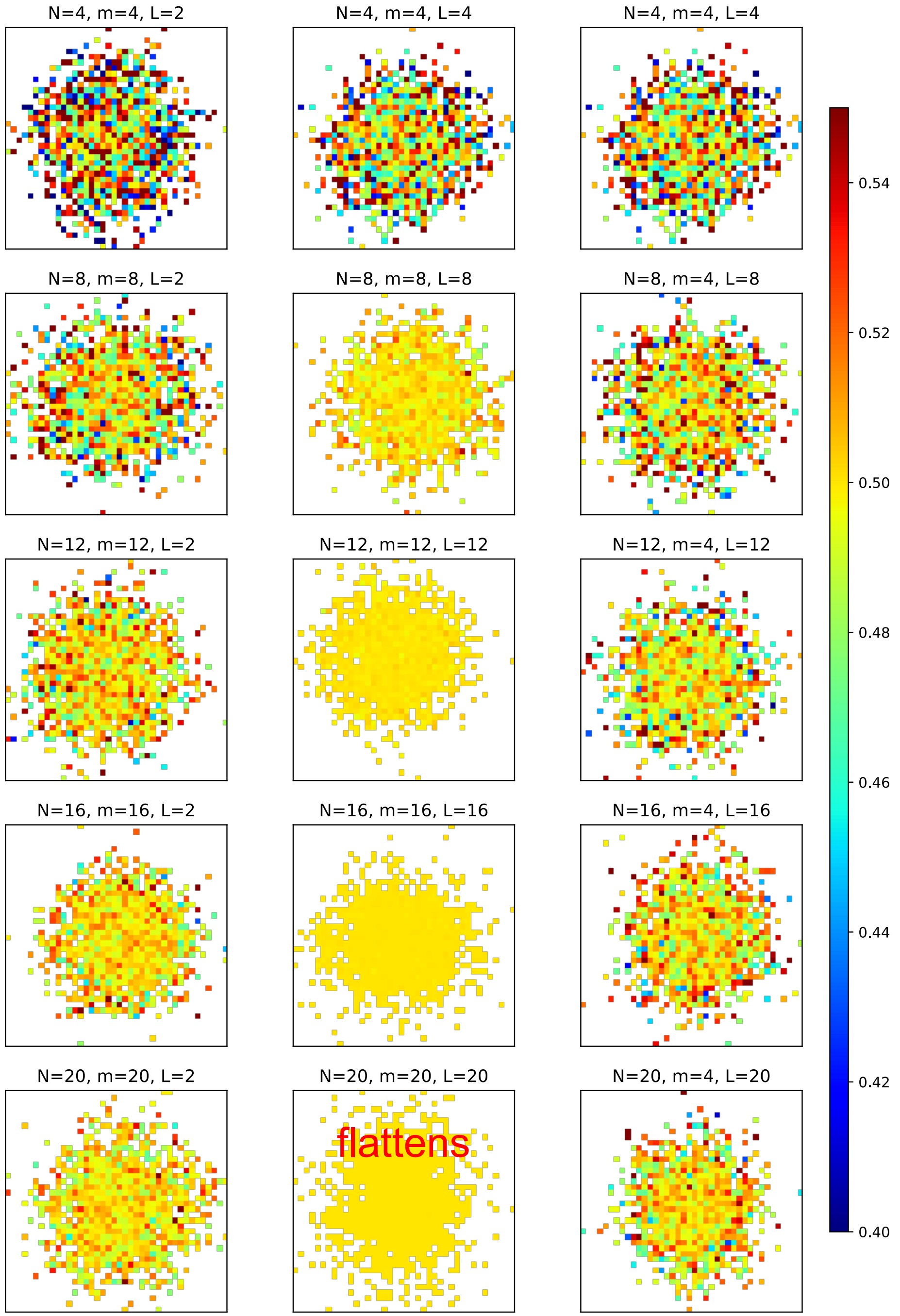
→ Accuracy benefits from CX layers

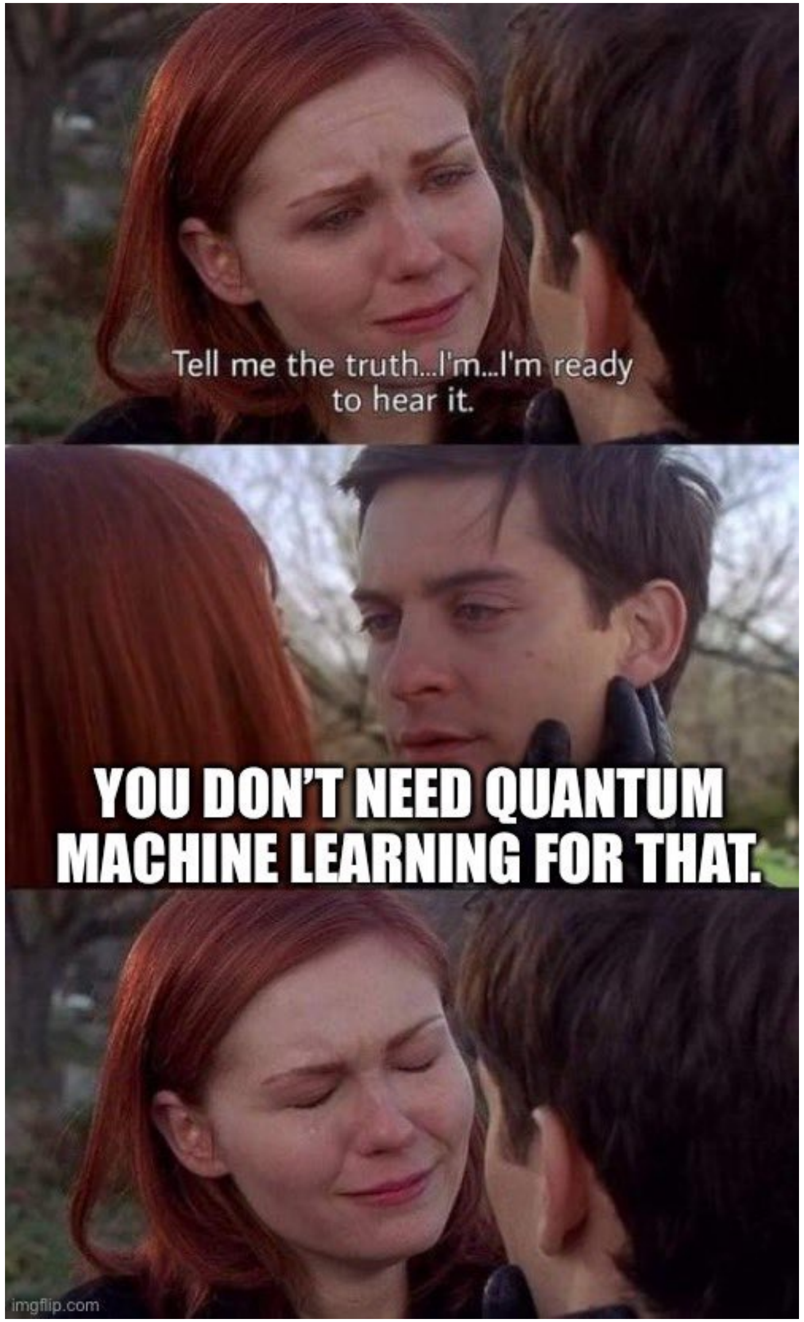


Performance of the CS Ansatz

Cont'd







Credit: @lpptb

QML for Quantum Problems

NTangled dataset

Concentrable Entanglement (CE) is an entanglement measure defined as;

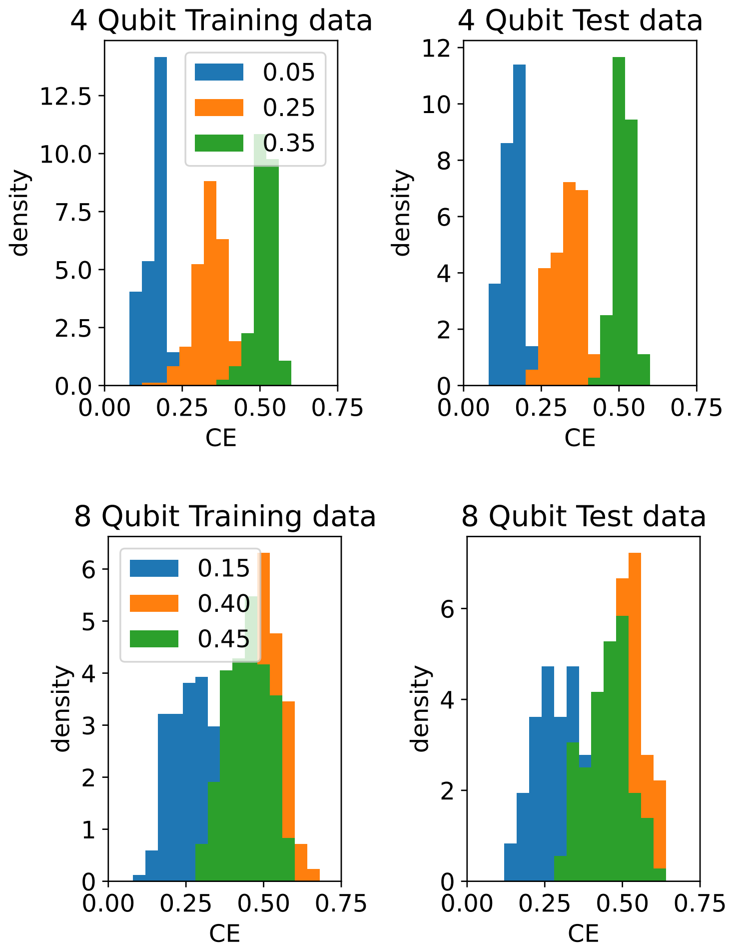
$$\mathrm{CE}(|\Psi\rangle) = 1 - \frac{1}{2^N} \sum_{\alpha \in Q} \mathrm{Tr}[\rho_\alpha^2]$$

where Q is the power set of the set {1,2,...,N}, and ρ_α is the reduced state of subsystems labeled by the elements of α associated to |Ψ⟩

The Ntangled dataset is one of the few publicly available "quantum" datasets!

Schatzki et al. 2021 arXiv: 2109.03400

Publicly available: https://github.com/LSchatzki/NTangled_Datasets



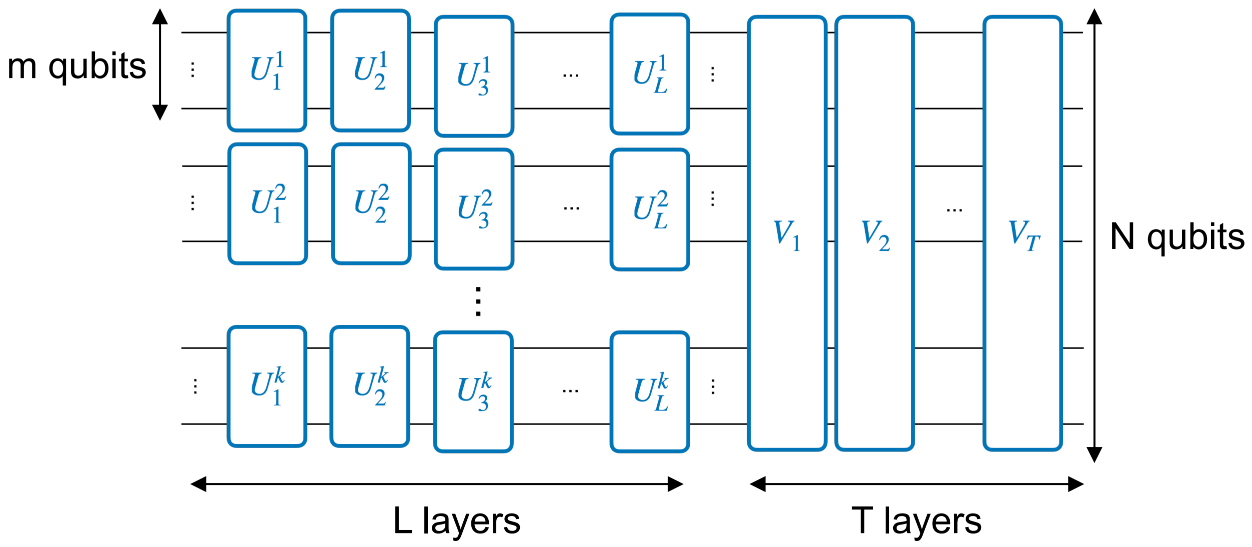
Results on the Ntangled dataset

Classification performance on the Ntangled dataset.

N	Task [CE Values]	L	m	Train Accuracy (%)	Avg. epochs to reach 90% Train Accuracy	Avg. epochs to reach 100% Train Accuracy	Test Accuracy (%)	Avg. epochs to reach 90% Test Accuracy	Avg. epochs to reach 100% Test Accuracy
4	0.05 vs. 0.35	4	1	$94.6^{+2.5}_{-1.7}$	$6.7^{+11.3}_{-5.7}$	N/A	$94.6^{+3.8}_{-1.8}$	$6.1^{+11.9}_{-5.1}$	N/A
			2	$\mathbf{100.0}^{+0.0}_{-0.5}$	$\mathbf{4.9}^{+12.1}_{-3.9}$	N/A	$100.0^{+0.0}_{-0.0}$	$\mathbf{3.9}^{+11.1}_{-2.9}$	$10.8^{+26.2}_{-9.8}$
			4	$99.9^{+0.1}_{-1.6}$	$5.4^{+6.6}_{-4.4}$	N/A	$100.0^{+0.0}_{-1.1}$	$4.1^{+8.9}_{-3.1}$	N/A
4	0.25 vs. 0.35	4	1	$90.4^{+4.1}_{-3.5}$	N/A	N/A	$86.4^{+6.9}_{-5.9}$	N/A	N/A
			2	$98.2^{+1.5}_{-1.3}$	$7.7^{+25.3}_{-5.7}$	N/A	$97.1^{+2.3}_{-1.6}$	$7.9^{+27.1}_{-6.9}$	N/A
			4	$\mathbf{100.0}^{+0.0}_{-0.4}$	$\mathbf{5.1}^{+9.9}_{-4.1}$	N/A	$\mathbf{100.0}^{+0.0}_{-1.1}$	$\mathbf{4.5}^{+11.5}_{-3.5}$	N/A
8	0.15 vs. 0.45	8	1	$99.9^{+0.1}_{-0.2}$	$3.3^{+3.7}_{-2.3}$	N/A	$100.0^{+0.0}_{-0.0}$	$2.4^{+2.6}_{-1.4}$	$6.1^{+10.9}_{-5.1}$
			2	$100.0^{+0.0}_{-0.0}$	$2.5^{+2.5}_{-1.5}$	$7.1^{+11.9}_{-6.1}$	$100.0^{+0.0}_{-0.0}$	$1.5^{+2.5}_{-0.5}$	$3.2^{+6.8}_{-2.2}$
			4	$100.0^{+0.0}_{-0.0}$	$\mathbf{2.4}^{+1.6}_{-1.4}$	$\mathbf{4.6}^{+4.4}_{-3.6}$	$100.0^{+0.0}_{-0.0}$	$\mathbf{1.4}^{+1.6}_{-0.4}$	$\mathbf{2.9}^{+7.1}_{-1.9}$
			8	$100.0^{+0.0}_{-0.0}$	$2.8^{+2.2}_{-0.8}$	$7.8^{+11.2}_{-5.8}$	$100.0^{+0.0}_{-0.0}$	$1.8^{+2.2}_{-0.8}$	$4.8^{+8.2}_{-3.8}$
8	0.40 vs. 0.45	8	1	$99.9^{+0.1}_{-0.4}$	$3.1^{+2.9}_{-2.1}$	N/A	$99.6^{+0.4}_{-0.7}$	$2.2^{+2.8}_{-1.2}$	N/A
			2	$100.0^{+0.0}_{-0.0}$	$2.8^{+4.2}_{-1.8}$	$9.2^{+11.8}_{-8.2}$	$100.0^{+0.0}_{-0.0}$	$1.9^{+4.1}_{-0.9}$	$5.2^{+5.8}_{-4.2}$
			4	$100.0^{+0.0}_{-0.0}$	$\mathbf{2.4}^{+1.6}_{-1.4}$	$\mathbf{5.3}^{+12.7}_{-4.3}$	$100.0^{+0.0}_{-0.0}$	$\mathbf{1.5}^{+1.5}_{-0.5}$	$\mathbf{3.2}^{+9.8}_{-2.2}$
			8	$100.0^{+0.0}_{-0.0}$	$2.9^{+3.1}_{-0.9}$	$8.2^{+9.8}_{-6.2}$	$100.0^{+0.0}_{-0.0}$	$1.9^{+3.1}_{-0.9}$	$5.7^{+5.3}_{-4.7}$

What about simulating physical systems?

The Extended CS Ansatz

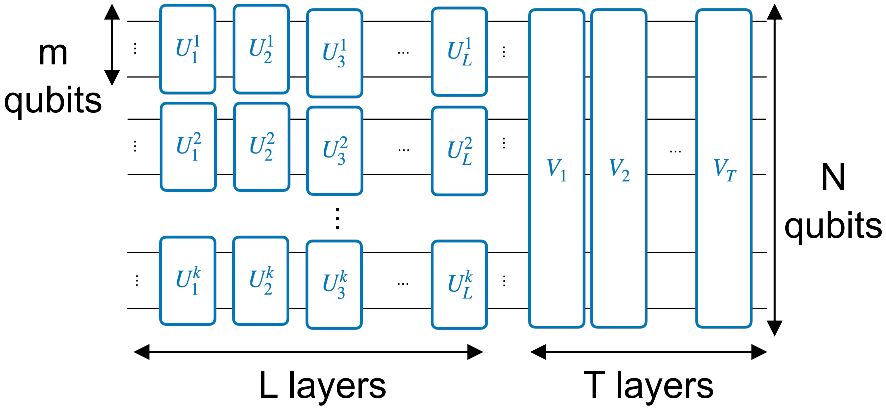


Allows us to reach to higher entanglement entropy states.

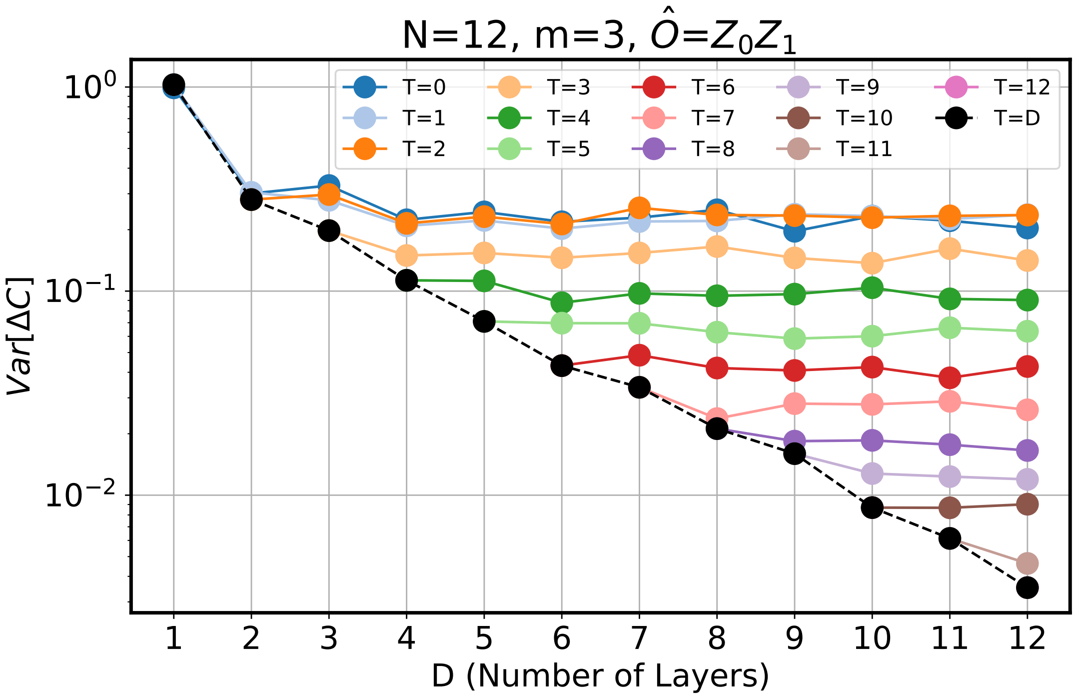
Classically split layers

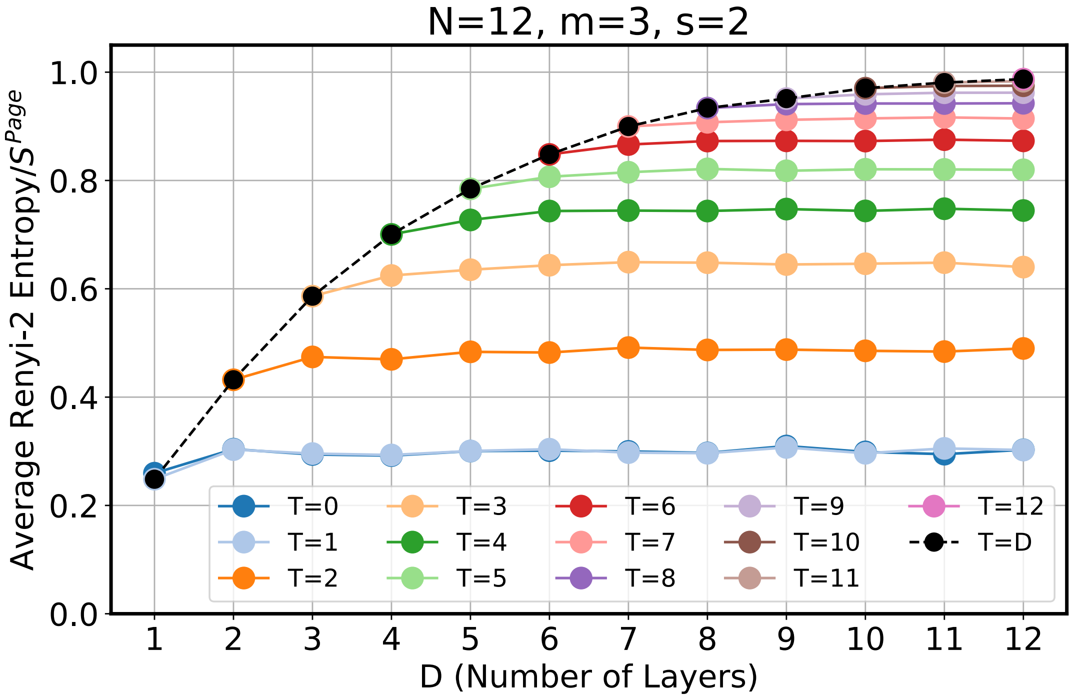
Standard layers

The Extended CS Ansatz

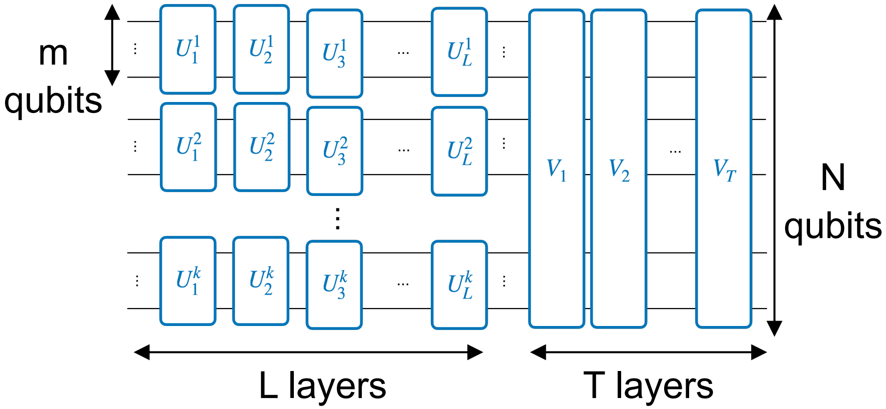


The ECS ansatz gives us the flexibility to choose certain levels of entanglement entropy.



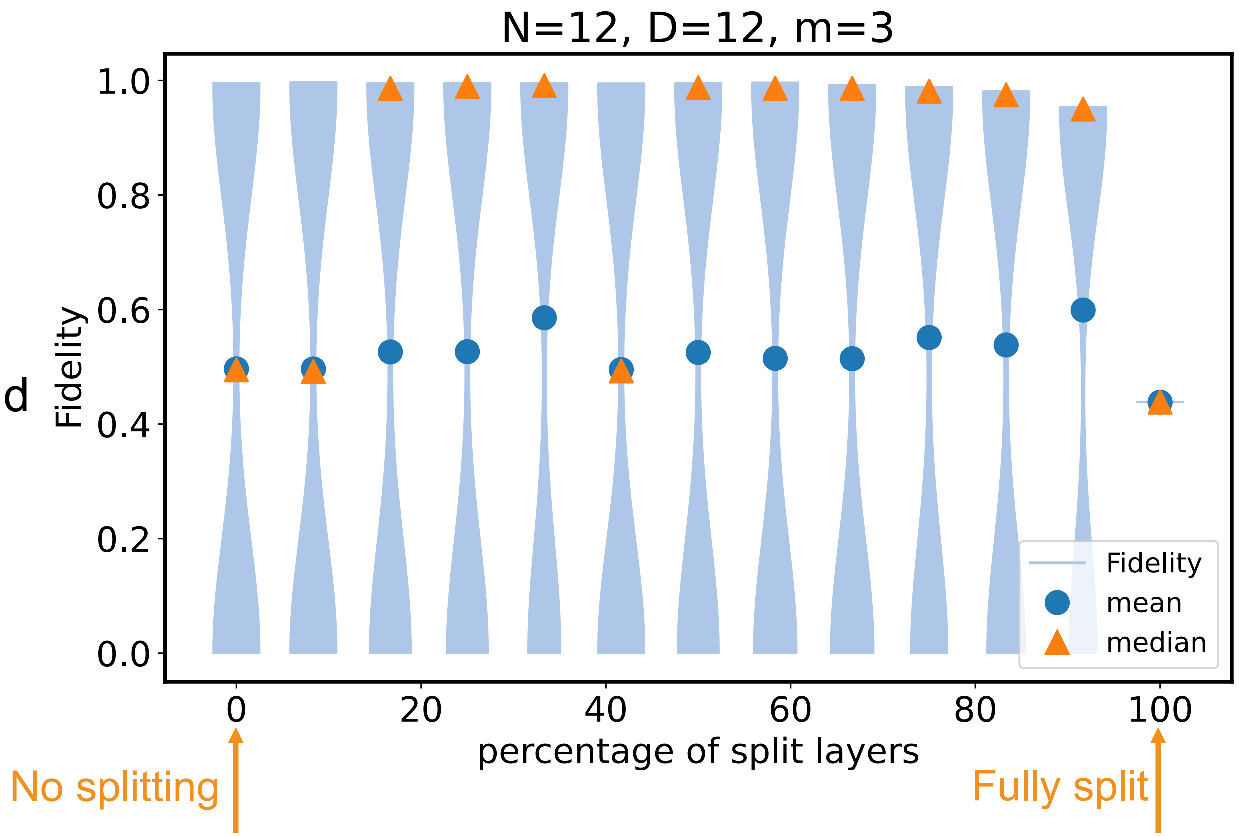


The Extended CS Ansatz



We simulate the transversal-field Ising Hamiltonian for J=h=1 on a 1D line for N=12, D=12, m=3 and N=16, D=16, m=4.

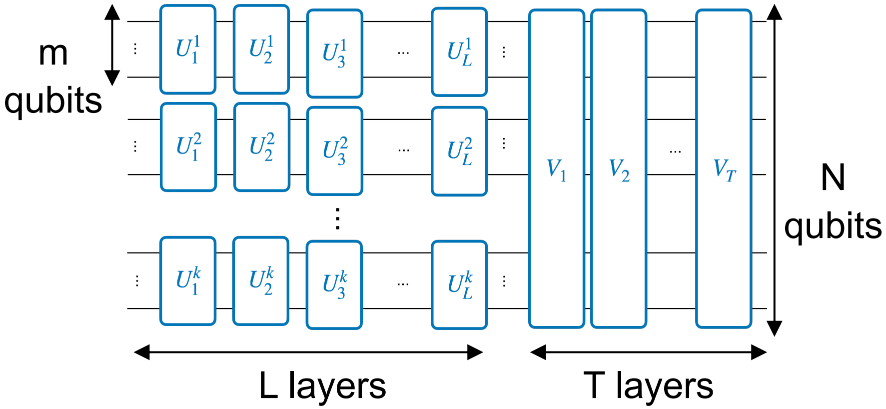
$$H = -J \sum_{i=1}^{N} Z_i Z_{i+1} - h \sum_{i=1}^{N} X_i$$



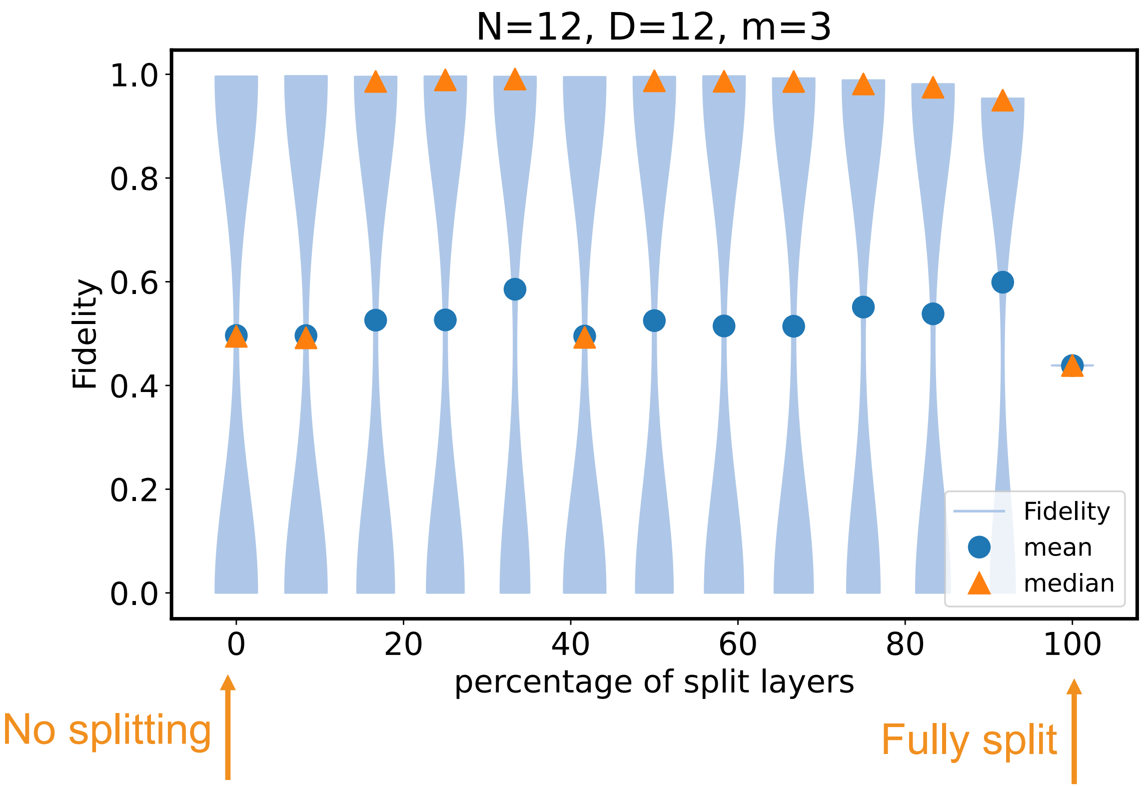
We present results from 100 runs with different L and T values.

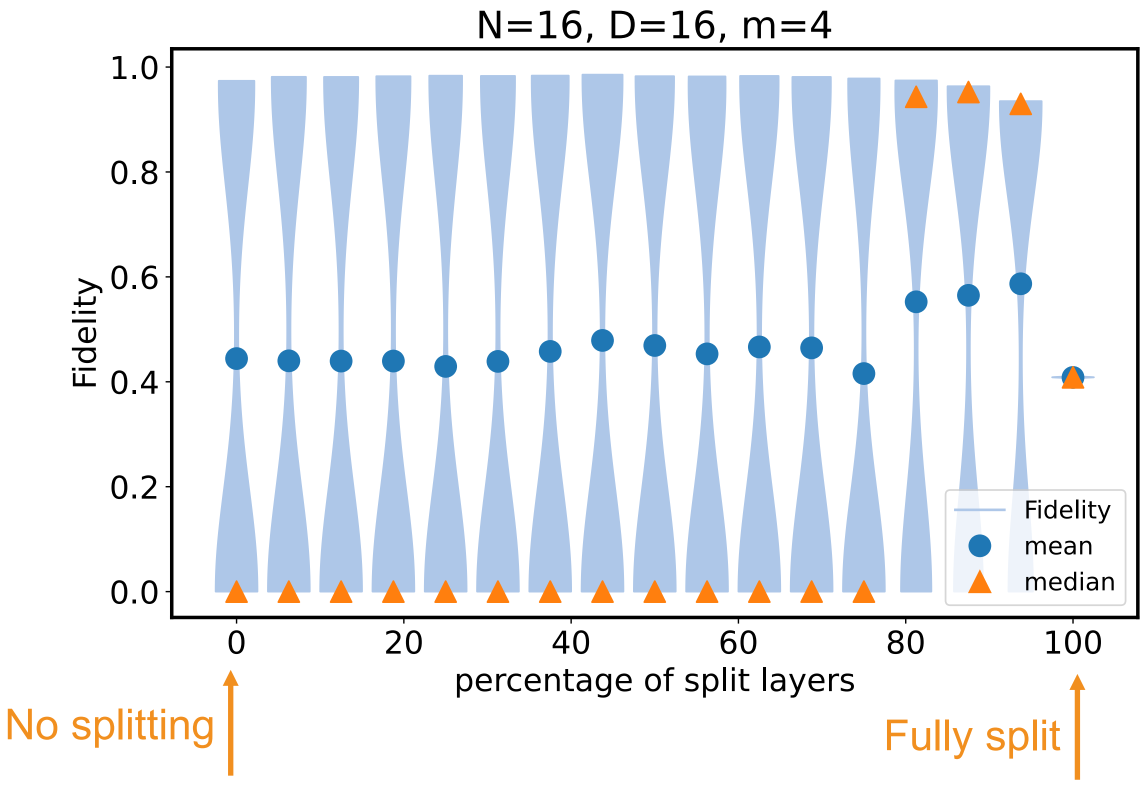
→ Success rate improves with classical splitting

Performance of the ECS Ansatz



→ Success rate improves with classical splitting
It is hard to pinpoint the transition point!





Summary

Classical splitting allows,

- Mitigating Barren Plateaus
  - better trainability
  - faster convergence
  - Doesn’t ensure better results out of the box!
- Extended CS ansatz:
  - Empirical evidence suggests a performance increase.
  - Hard to find the hyperparameters that gives the best results.

For more details:

arXiv: 2206.09641

For more details:

arXiv: 2206.09641

Contact

DESY. Deutsches Elektronen-Synchrotron

www.desy.de

Cenk Tüysüz
DESY CQTA

cenk.tueysuez@desy.de

@cenk_tuysuz

Credit: @itheodonis