

A Unified Theory of Barren Plateaus for Deep Parametrized Quantum Circuits

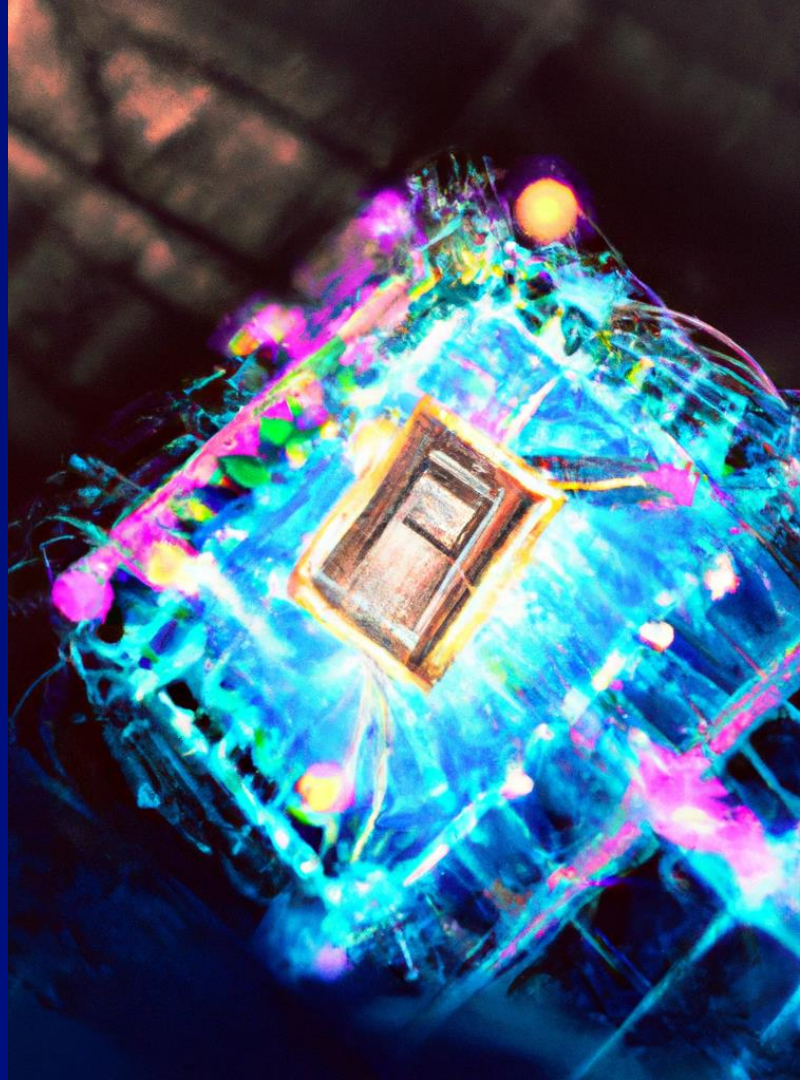
Marco Cerezo
CCS-3

“A Unified Theory of Barren Plateaus for Deep Parametrized Quantum Circuits” arxiv 2309.09342 (2023)
Michael Ragone, Bojko N. Bakalov, Frédéric Sauvage, Alexander F. Kemper, Carlos Ortiz Marrero, Martin Larocca, **MC**
“Showcasing a Barren Plateau Theory Beyond the Dynamical Lie Algebra” arXiv:2310.11505 (2023).
N. L. Diaz, Diego García-Martín, Sujay Kazi, Martin Larocca, and **MC**

11/21/2023 @QTML, CERN

Outline

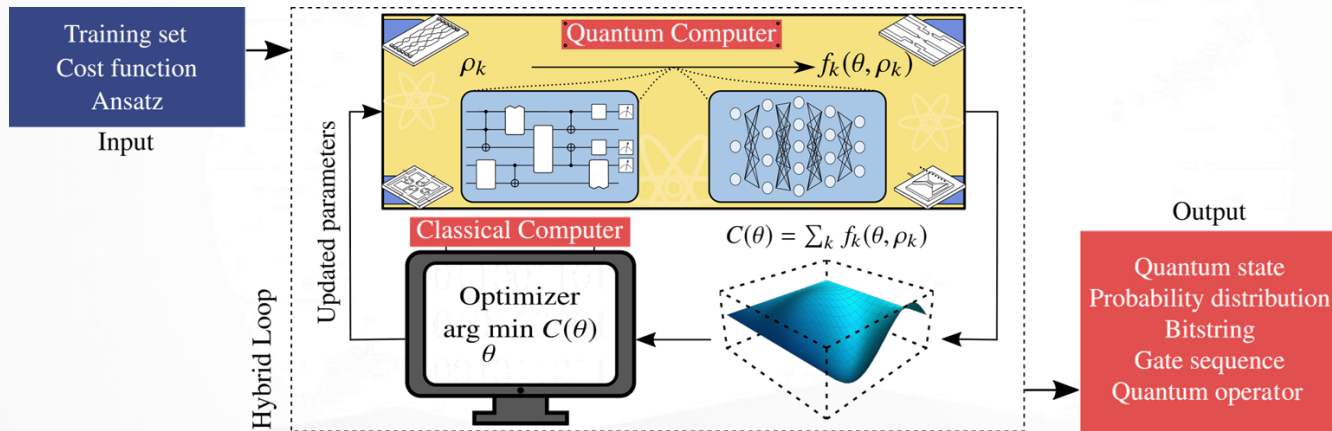
- (Very brief) Introduction to variational quantum computing
- A casual stroll through the history of Barren Plateaus (BPs)
- A unified Dynamical Lie algebraic perspective to BPs
- Beyond the Dynamical Lie algebra BP Theory
- Outlook



Variational Quantum Computing

Solve a problem of interest by encoding it as an optimization task.

These include **Variational Quantum Algorithms** (VQE, QAOA, etc), but also **Quantum Machine Learning** schemes. We will consider models passed on Parametrized Quantum Circuits, or, Quantum Neural Networks.



The loss function will take the form:

$$\ell_{\theta}(\rho, O) = \text{Tr}[U(\theta)\rho U^{\dagger}(\theta)O]$$

We will assume (standard) $|O\rangle_2 \leq 2^n$

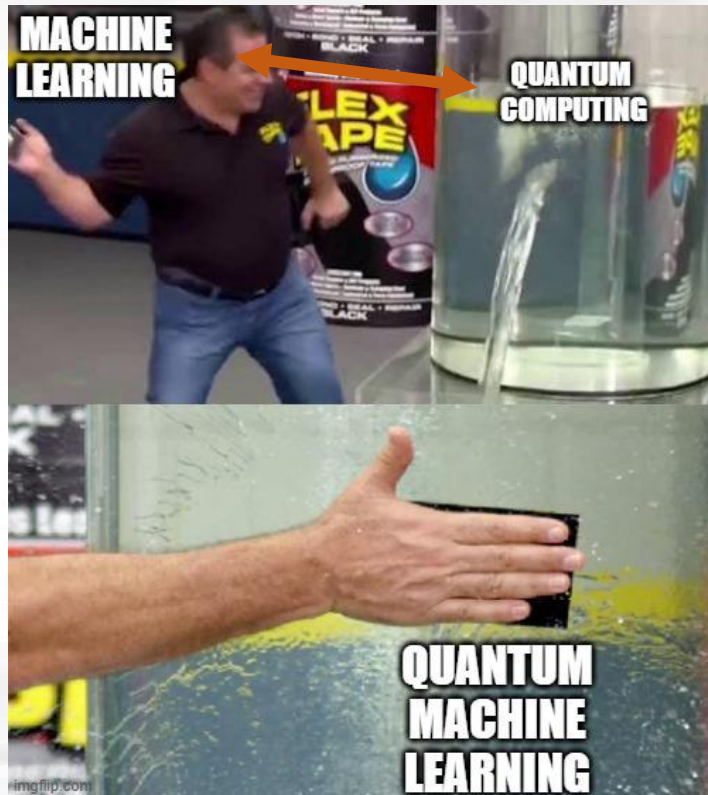
How and when can this computational approach fail?

While NN are widely used today, their historical development saw periods of great **stagnation** (or winters).



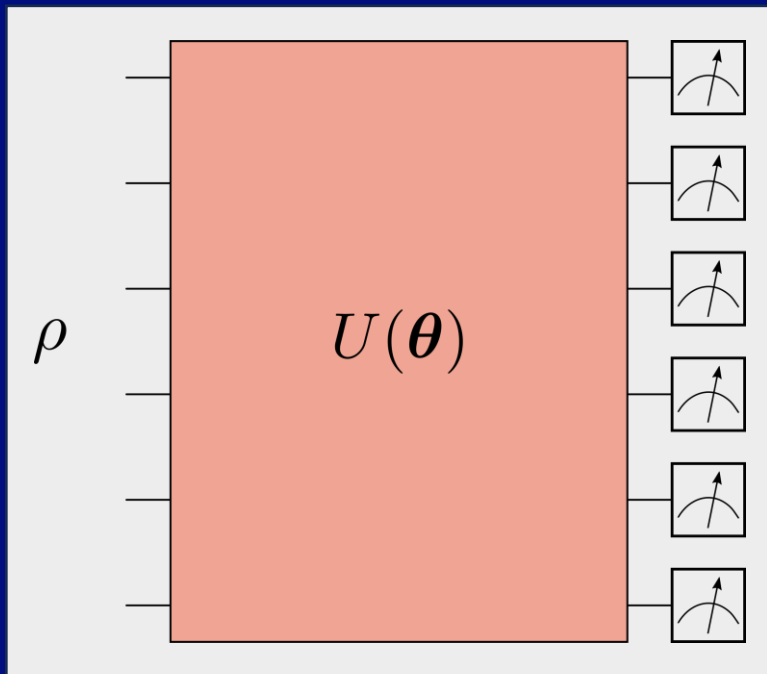
Variational Quantum Computing holds tremendous the tremendous promise to achieve a quantum advantage – a computational speedup – we need to **guarantee** that our architectures will work when scaled up and applied to large, realistic problems.

Two memes in a row? Oh, the humanity!



Z_2 symmetric
Meme

A review of Barren Plateaus and their sources



Good practices that we have come to learn and love $\langle \dots \rangle$

- Global cost function
- on all qubits
- Too many parameters
- Deep circuits



We will study *loss concentration*, not *partial derivative concentration*.

They can be shown to be *equivalent*.

Barren Plateaus

No Barren Plateau



Image Credits: Samson Wang

$$\text{Var}_{\theta}[\ell_{\theta}(\rho, O)] \in \Omega(1/\text{poly}(n))$$

Barren Plateau

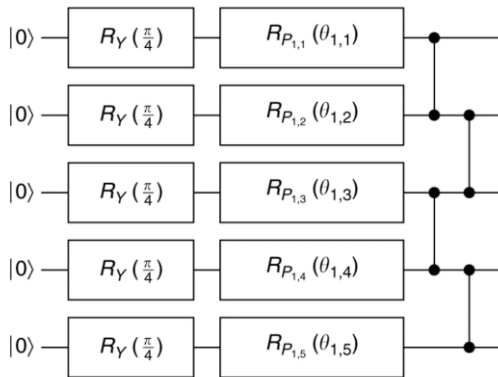


$$\text{Var}_{\theta}[\ell_{\theta}(\rho, O)] \in \mathcal{O}(1/b^n) \text{ with } b > 1.$$

Expressiveness in the circuit, or, a general form of **No-Free-Lunch theorem**

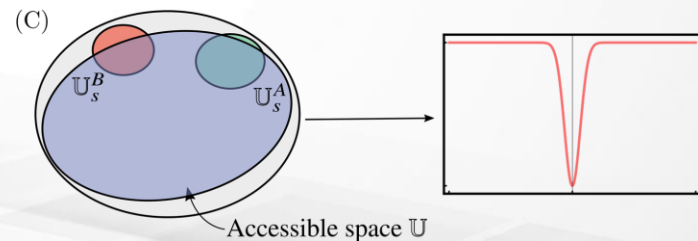
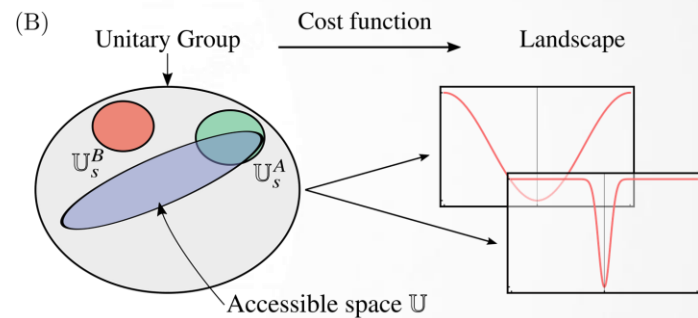
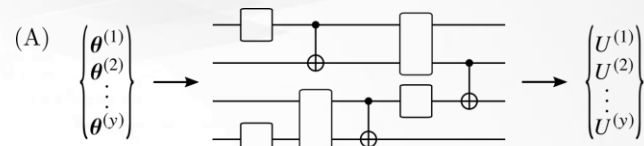
The first, and perhaps, most studied source of BPs:

Seminal Google paper: $U(\theta)$ forms a **2-design**



$$\text{Var}_{\theta}[\ell_{\theta}(\rho, O)] = \frac{1}{4^n - 1} (\text{Tr}[\rho^2] - 1/2^n)(\text{Tr}[O^2] - \text{Tr}[O]^2/2^n)$$

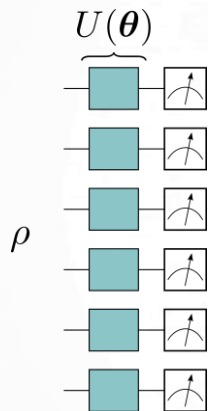
We have a **BP** for any ρ , and O .



Locality of measurement operator

or, why comparing **exponentially large objects** is bound to fail

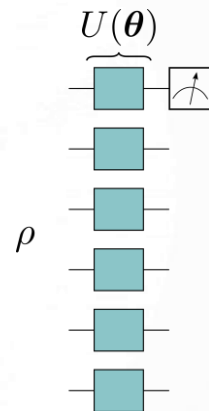
Let's take the most **inexpressive** imaginable circuit and a **global measurement** $O = Z^{\otimes n}$



Always has a BP:

$\text{Var}_{\theta}[\ell_{\theta}(\rho, O)] \in \mathcal{O}(1/b^n)$ with $b > 1$ for any ρ .

What if we take a **local measurement** $O = Z_1$



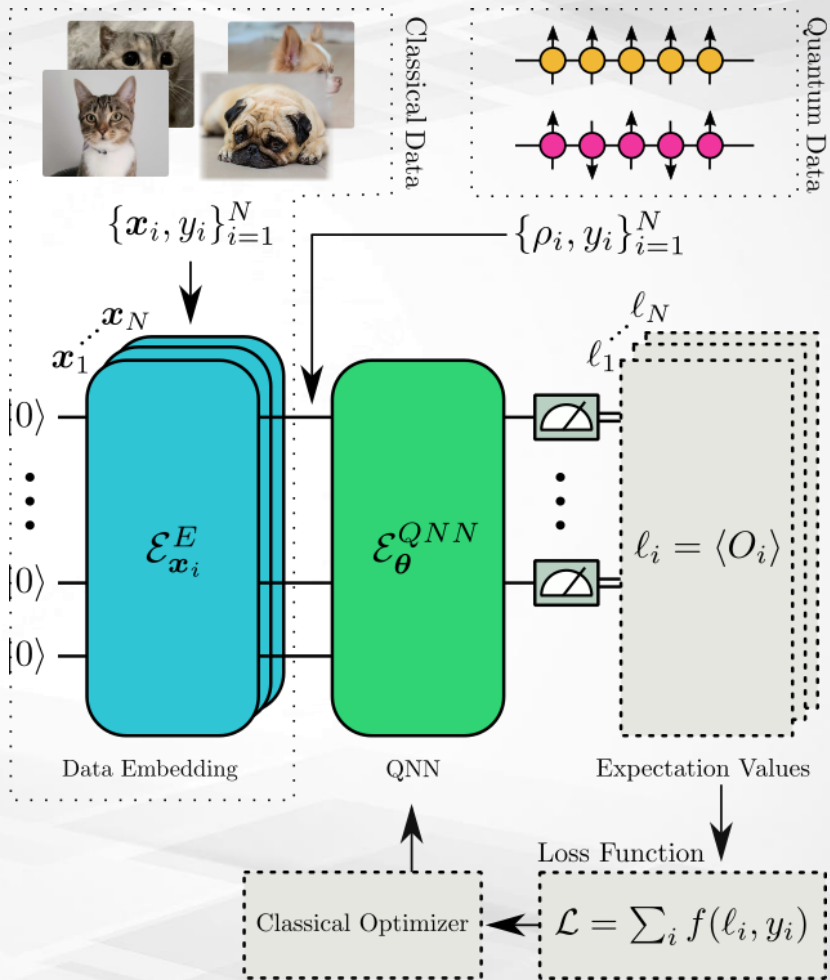
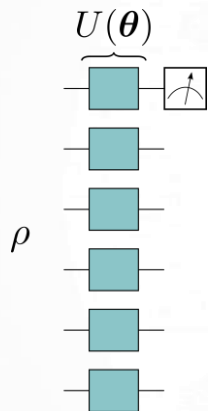
$$\text{Var}_{\theta}[\ell_{\theta}(\rho, O)] = \frac{2}{3} (\text{Tr}[(\rho_1)^2] - 1/2)$$

If input is $\rho = |0\rangle\langle 0|^{\otimes n}$, then no BP.. But...

Entanglement in the initial state

or, why **untamed** entanglement can be bad

What if we take a **local measurement** $O = Z_1$, but ρ follows a **volume law of entanglement**?

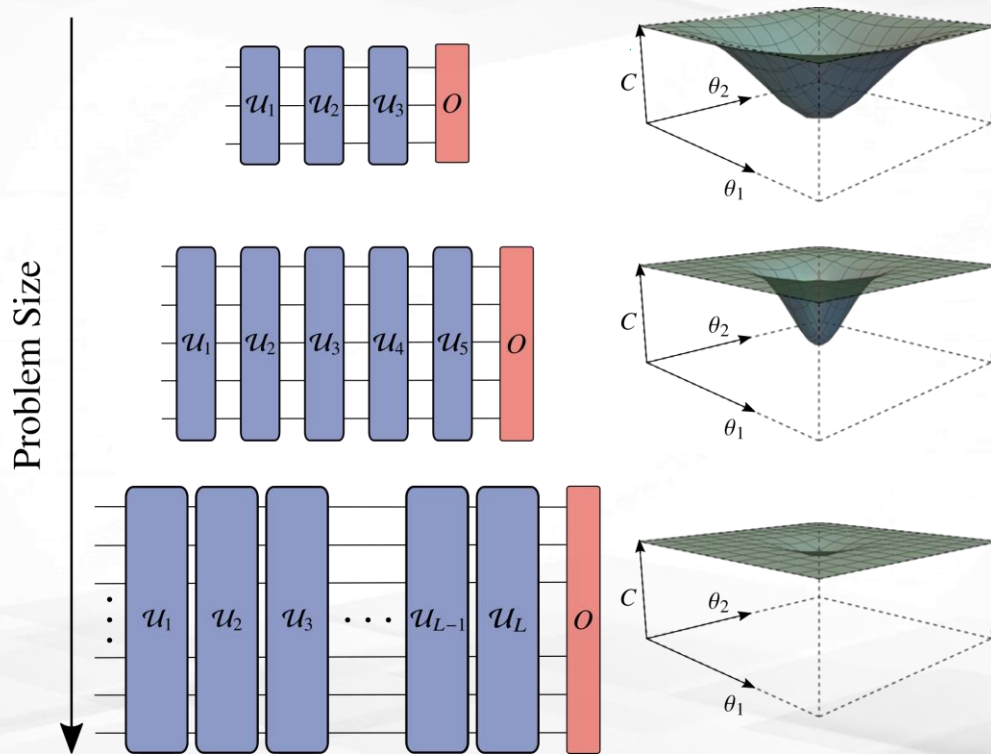


Has a BP: $\text{Var}_{\theta}[\ell_{\theta}(\rho, O)] \in \mathcal{O}(1/b^n)$

Despite no expressiveness and local measurements

And of course, Noise

or, why **existence is pain**



Previous understanding of BPs

or, yes this slide looks ugly, but it will get better soon.

Architecture
Specific Analysis

Specific $U(\theta)$

Previous works:

Expressiveness

Entanglement

Locality

Sources of Barren Plateaus

Can we understand and unify all these sources of BPs?

Yes! We need to take a **Lie algebraic** perspective. Assume

$$U(\boldsymbol{\theta}) = \prod_{l=1}^L e^{-i \theta_l H_l},$$

Where $H_l \in \mathcal{G}$ (set of generators).

Then we need to construct the circuit's **Dynamical Lie algebra (DLA)**:

$$\mathfrak{g} = \langle i\mathcal{G} \rangle_{Lie} \subseteq su(d)$$

DLA = subspace spanned by the nested commutators of the generators

The DLA is important for 2 reasons:

- 1) $U(\boldsymbol{\theta}) \in e^{\mathfrak{g}}$ for any $\boldsymbol{\theta}$, L .
- 2) *Quantifies the ultimate expressiveness $U(\boldsymbol{\theta})$.*

BCH Formula $e^A e^B = e^C$ with

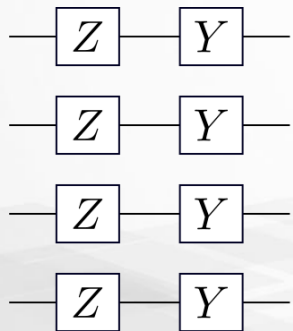
$$C = A + B + \frac{1}{2}[A, B] + \frac{1}{12}[A, [A, B]] - \frac{1}{12}[B, [A, B]] + \dots$$

Examples of DLA

Single qubit rotations:

$$\mathcal{G} = \{Z_i, Y_i\}_{i=1}^n$$

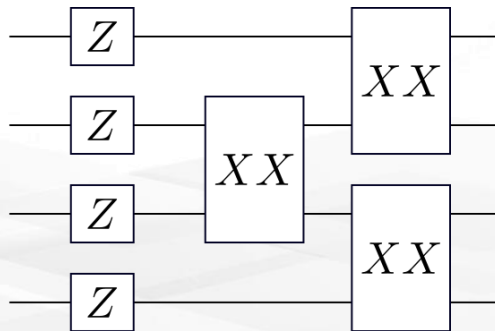
$$\mathfrak{g} = \mathfrak{su}(2)^{\oplus n}$$



Matchgate circuit:

$$\mathcal{G} = \{Z_i\}_{i=1}^n \cup \{X_i X_{i+1}\}_{i=1}^{n-1}$$

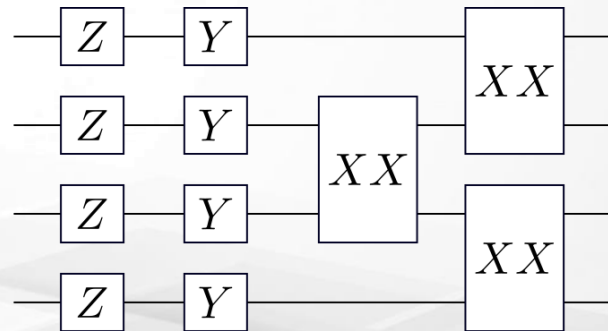
$$\mathfrak{g} = \mathfrak{so}(2n)$$



Universal circuit:

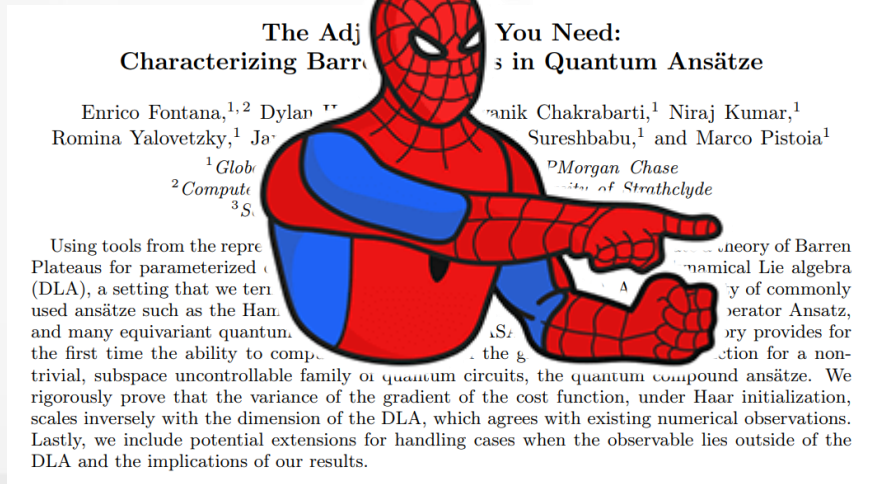
$$\mathcal{G} = \{Z_i, Y_i\}_{i=1}^n \cup \{X_i X_{i+1}\}_{i=1}^{n-1}$$

$$\mathfrak{g} = \mathfrak{su}(2^n)$$

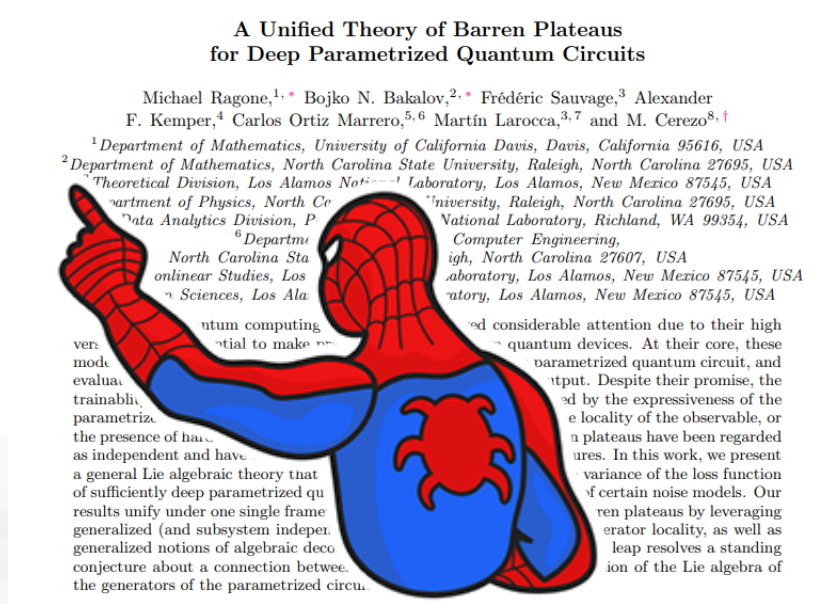


Enough math. Why is the DLA important?

The importance of the DLA was **conjectured** in our work [1], and recently **proven** to be central in the BP study. A classic QIP deadline induced rush:



arXiv:2309.07902



arXiv:2309.09342

Image Credits: Stan Lee, ABC's Spiderman Cartoon: "Double Identity" Season 1, Episode 19 (1968). Yeah, I googled it.

Exact variance formula for states or observables in the DLA

Given an **arbitrary subalgebra** $\mathfrak{k} \subseteq su(2^n)$ with Hermitian basis $\{B_j\}_{j=1}^{\dim(\mathfrak{k})}$, we define the **\mathfrak{k} -purity** of an operator H as

$$\mathcal{P}_{\mathfrak{k}}(H) = \text{Tr}[H_{\mathfrak{k}}^2] = \sum_{j=1}^{\dim(\mathfrak{k})} \text{Tr}[B_j^\dagger H]^2.$$

Then, recalling that any DLA

$$\mathfrak{g} = \underbrace{\mathfrak{g}_1 \oplus \mathfrak{g}_2 \oplus \cdots \oplus \mathfrak{g}_k}_{\text{simple}} \underbrace{\quad}_{\text{abelian}}$$

We have bounds for circuit depth

and **assuming that either ρ , or O are in \mathfrak{ig}** (QAOA, VEQ) and the circuit is a 2-design over $\mathfrak{e}\mathfrak{g}$:

$$\mathbb{E}_{\theta}[\ell_{\theta}(\rho, O)] = \text{Tr}[\rho_{\mathfrak{g}_k} O_{\mathfrak{g}_k}]$$

$$\text{Var}_{\theta}[\ell_{\theta}(\rho, O)] = \sum_{j=1}^{k-1} \frac{\mathcal{P}_{\mathfrak{g}_j}(\rho) \mathcal{P}_{\mathfrak{g}_j}(O)}{\dim(\mathfrak{g}_j)}$$

EXACT Formulae

What can we learn from here?

Assume \mathfrak{g} is simple

$$E_{\theta}[\ell_{\theta}(\rho, O)] = 0$$

$$\text{Var}_{\theta}[\ell_{\theta}(\rho, O)] = \frac{\mathcal{P}_g(\rho)\mathcal{P}_g(O)}{\dim(\mathfrak{g})}$$

If either $\dim(\mathfrak{g}), 1/\mathcal{P}_g(\rho), 1/\mathcal{P}_g(O) \in \Omega(b^n)$ with $b > 2$, one has a BP!

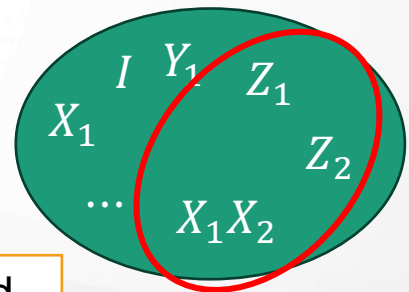
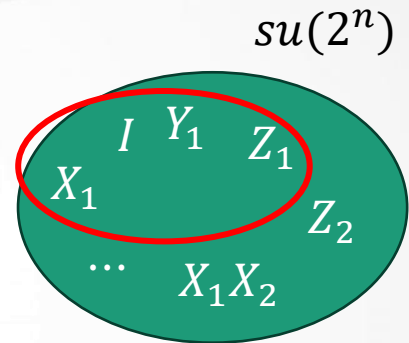
There are 3 sources of BPs here!

1. The circuit **expressiveness**, measured through $\dim(\mathfrak{g})$.
2. The **generalized entanglement** in ρ , measured through $\mathcal{P}_g(\rho)$.
3. The **generalized locality** of O , measured through $\mathcal{P}_g(O)$.

Proves a conjecture in [1]!

Maximized for O in DLA

Maximized for a HWS.



$su(2^n)$

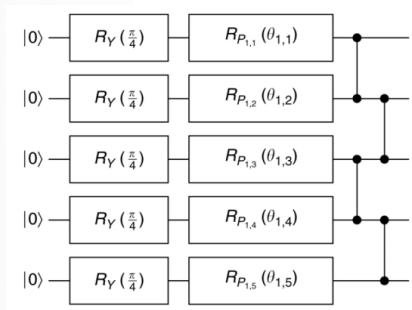
Revisiting previous examples

Google's result:

$$\mathfrak{g} = su(2^n); \quad \dim(\mathfrak{g}) = 4^n - 1$$

$$\mathcal{P}_{\mathfrak{g}}(\rho) = \text{Tr}[\rho^2] - 1/2^n \quad (\rho = \rho_{\mathfrak{g}} - I/2^n)$$

$$\mathcal{P}_{\mathfrak{g}}(O) = \text{Tr}[O^2] \quad (O = O_{\mathfrak{g}} - \text{Tr}[O]I/2^n)$$



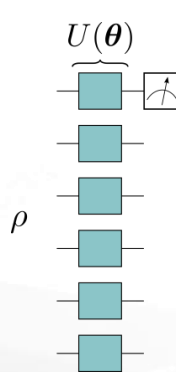
$$\text{Var}_{\theta}[\ell_{\theta}(\rho, O)] = \frac{1}{4^n - 1} (\text{Tr}[\rho^2] - 1/2^n)(\text{Tr}[O^2] - \text{Tr}[O]^2/2^n)$$

Local measurement $O = Z_1$

$$\mathfrak{g} = su(2)^{\oplus n} \quad (Z_1 \in \mathfrak{g}_1); \quad \dim(\mathfrak{g}_1) = 3$$

$$\mathcal{P}_{\mathfrak{g}_1}(\rho) = \text{Tr}[\rho_1^2] - 1/2^n \quad (\rho_1 = \text{Tr}_{\bar{1}}[\rho])$$

$$\mathcal{P}_{\mathfrak{g}_1}(O) = 2$$



Generalized
entanglement =
standard
entanglement

Algebra IS local

$$\text{Var}_{\theta}[\ell_{\theta}(\rho, O)] = \frac{2}{3} (\text{Tr}[(\rho_1)^2] - 1/2^n)$$

An important lesson learnt

The previous examples led the community to developing a series of good practice guidelines such as:

- Global observables (acting on all qubits) are intrainable
- Too much entanglement leads to barren plateaus
- Deep circuits are have BPs



We show that **generalized entanglement** in ρ is relative to the DLA, **generalized locality** of O , is relative to the DLA.

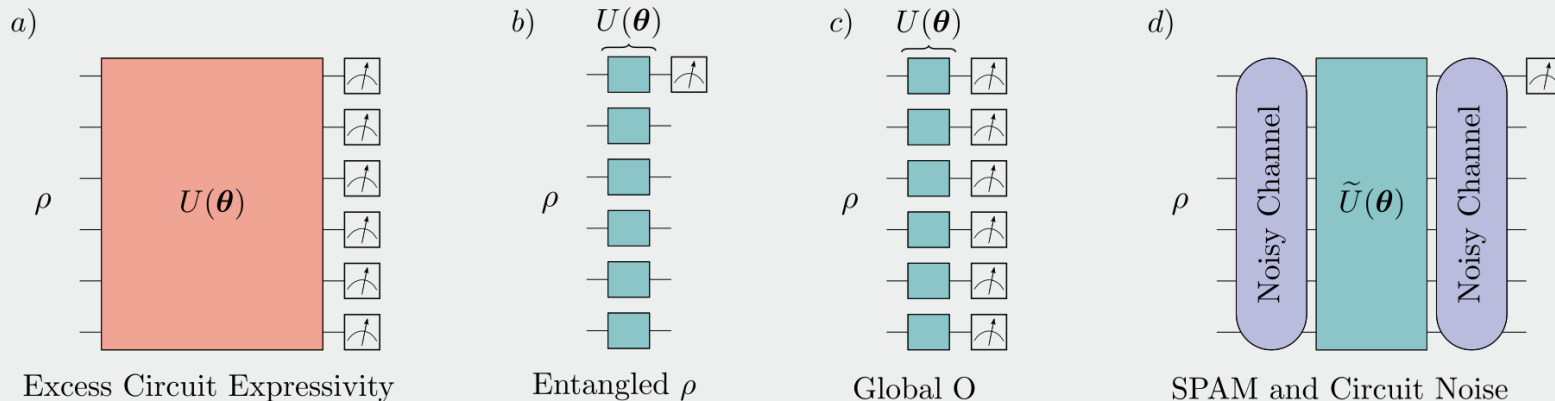
*Hence, it is entirely possible for a circuit to be trained on **highly entangled initial states** using **highly nonlocal measurements** (i.e., acting on all qubits) as long as they are well aligned with the underlying DLA of the circuit.*

Summary

Unified Theory of Barren Plateaus

$$\text{Var}_{\theta}[\ell_{\theta}(\rho, O)] = \frac{1}{\dim(\mathfrak{g})} \mathcal{P}_{\mathfrak{g}}(\rho) \mathcal{P}_{\mathfrak{g}}(O)$$

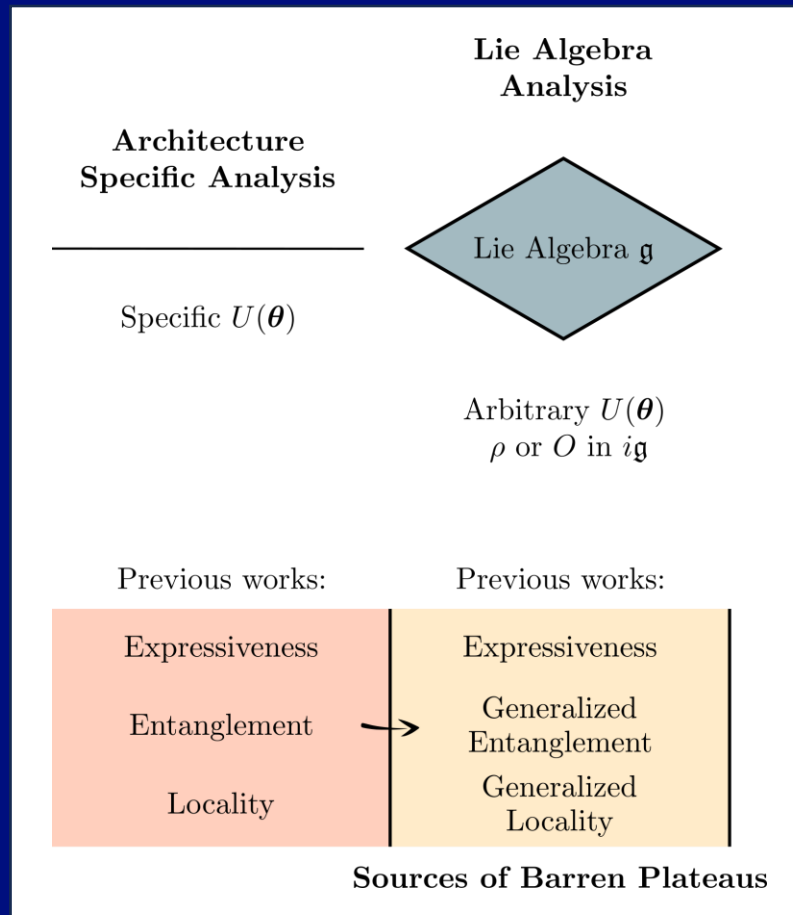
“Best known” sources of Barren Plateaus



Dimension of the DLA
 $\dim(\mathfrak{g})$

Generalized entanglement and locality by \mathfrak{g} -purity
 $\mathcal{P}_{\mathfrak{g}}(\rho) \mathcal{P}_{\mathfrak{g}}(O)$

Algebraic Decoherence
 $\mathcal{P}_{\mathfrak{g}}(\mathcal{N}_B(\rho)) = (1 - p)^2 \mathcal{P}_{\mathfrak{g}}(\rho)$



Showcasing a Barren Plateau Theory Beyond the Dynamical Lie Algebra

N. L. Diaz,^{1,2} Diego García-Martín,¹ Sujay Kazi,³ Martin Larocca,^{4,5} and M. Cerezo¹

¹*Information Sciences, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

²*Departamento de Física-IFLP/CONICET, Universidad Nacional de La Plata, C.C. 67, La Plata 1900, Argentina*

³*Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA*

⁴*Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

⁵*Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

Barren plateaus have emerged as a pivotal challenge for variational quantum computing. Our understanding of this phenomenon underwent a transformative shift with the recent introduction of a Lie algebraic theory capable of explaining most sources of barren plateaus. However, this theory requires either initial states or observables that lie in the circuit's Lie algebra. Focusing on parametrized matchgate circuits, in this work we are able to go beyond this assumption and provide an exact formula for the loss function variance that is valid for arbitrary input states and measurements. Our results reveal that new phenomena emerge when the Lie algebra constraint is relaxed. For instance, we find that the variance does not necessarily vanish inversely with the Lie algebra's dimension. Instead, this measure of expressiveness is replaced by a generalized expressiveness quantity: The dimension of the Lie group modules. By characterizing the operators in these modules as products of Majorana operators, we can introduce a precise notion of generalized globality and show that measuring generalized-global operators leads to barren plateaus. Our work also provides operational meaning to the generalized entanglement as we connect it with known fermionic entanglement measures, and show that it satisfies a monogamy relation. Finally, while parameterized matchgate circuits are not efficiently simulable in general, our results suggest that the structure allowing for trainability may also lead to classical simulability.

Can we go beyond the DLA?

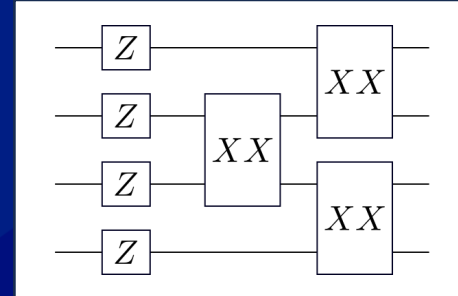
Yes! We need to take (at least for now) an **architecture specific analysis**.

Parametrized Matchgate circuits: $\mathcal{G} = \{Z_i\}_{i=1}^n \cup \{X_i X_{i+1}\}_{i=1}^{n-1}$

Free fermion DLA:

$$\mathfrak{g} = \text{span}_R \{Z_i, \widehat{X_i X_j}, \widehat{X_i Y_j}, \widehat{Y_i X_j}, \widehat{Y_i Y_j}\}_{1 \leq i < j \leq n} \simeq so(2n)$$

Where $\widehat{A_i B_j} = A_i Z_{i+1} \cdots Z_{j-1} B_j$



Before we needed ρ or O in ig. The DLA was the central element. Now we need to work with **group modules**.

$$\mathcal{B} = \bigoplus_{k=0}^{2n} \mathcal{B}_k; \dim(\mathcal{B}_k) = \binom{2n}{k}$$

$\forall M_k \in \mathcal{B}_k$, then $U M_k U^\dagger \in \mathcal{B}_k$ for any $U \in e^{\mathfrak{g}}$

Use Jordan-Wigner, and define **Majorana operators**

$$c_1 = XI \cdots I, c_3 = ZXI \cdots I, c_5 = Z \cdots ZX$$

$$c_2 = YI \cdots I, c_4 = ZYI \cdots I, c_6 = Z \cdots ZY$$

$$P = Z^{\otimes n} = (-i)^n c_1 \cdots c_{2n}$$

Moar exact variance formulae,

/môar/ - adj - A way of asking for more of something you want very badly


$$E_{\theta}[\ell_{\theta}(\rho, O)] = \sum_{\kappa=0,2n} \langle \rho_{\kappa}, O_{\kappa} \rangle_{I+P}$$

$$\text{Var}_{\theta}[\ell_{\theta}(\rho, O)] = \sum_{j=1}^{k-1} \frac{\mathcal{P}_{\kappa}(\rho)\mathcal{P}_{\kappa}(O) + \mathcal{C}_{\kappa}(\rho)\mathcal{C}_{\kappa}(O)}{\dim(\mathcal{B}_{\kappa})}$$

$O \in \mathcal{B}_{\kappa}$, then $E_{\theta}[\ell_{\theta}(\rho, O)] = 0$, and

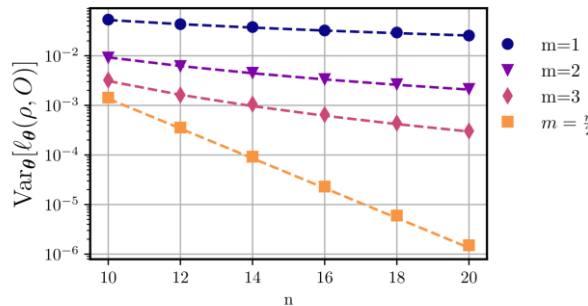
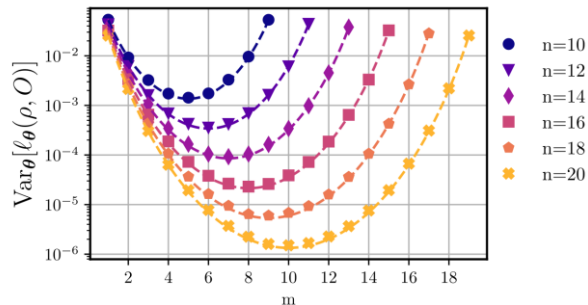
$$\text{Var}_{\theta}[\ell_{\theta}(\rho, O)] = \frac{\mathcal{P}_{\kappa}(\rho)\mathcal{P}_{\kappa}(O)}{\dim(\mathcal{B}_{\kappa})}$$

$\mathcal{P}_{\kappa}(M) = \langle M_{\kappa}, M_{\kappa} \rangle_I$, $\mathcal{C}_{\kappa}(M) = i^{\kappa \bmod 2} \langle M_{\kappa}, M_{\kappa} \rangle_P$, and $\langle M_1, M_2 \rangle_{\Gamma} = \text{Tr}[\Gamma M_1^{\dagger} M_2]$

1. The (local) **expressiveness** measure $\dim(\mathfrak{g})$ is replaced by the (global) **expressiveness** $\dim(\mathcal{B}_{\kappa})$. The conjecture in [1] is then not generally true. **Trainability is module dependent.**
2. We can define a notion of **generalized globality**, product of Majoranas: $\kappa \bmod n \in \Theta(1)$ no BP but $\kappa \bmod n \in \Theta(n)$ then BP!
3. For $\kappa = 2$ ($\mathcal{B}_2 = \mathfrak{g}$), **operational meaning** to $\mathcal{P}_{\kappa}(\rho)$ as a **Fermionic entanglement measure**. 
4. The covariances are forms of generalized-coherences **between isomorphic modules** (related to the parity sectors).

Remember, we have qubits!

I do theory, but... here are some cool numerics:

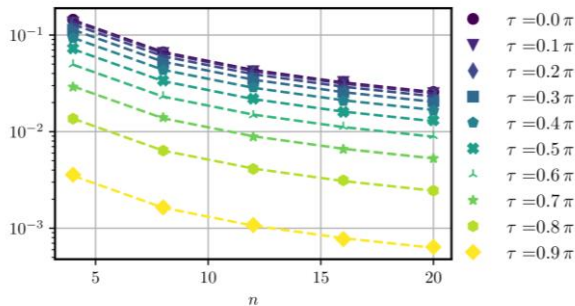
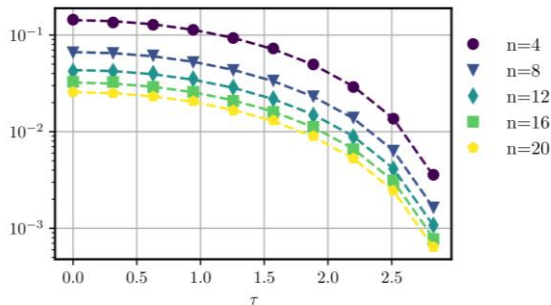


Gaussian State

$$\rho = |0\rangle\langle 0|^{\otimes n}$$

Increase standard globality:

$$O = Z^{\otimes m} \in \mathcal{B}_{2m}$$

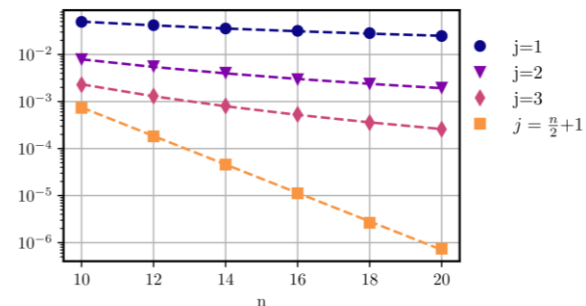
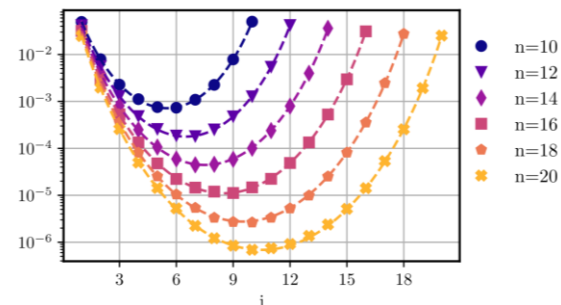


Magic State (exponential extent)

$$|\Psi(\tau)\rangle = \left[\frac{|0000\rangle + |0011\rangle + |1100\rangle + e^{i\tau} |1111\rangle}{2} \right]^{\otimes n/4}$$

Measure in the algebra (small dimension)

$$O = Z_1 \in \mathcal{B}_2$$



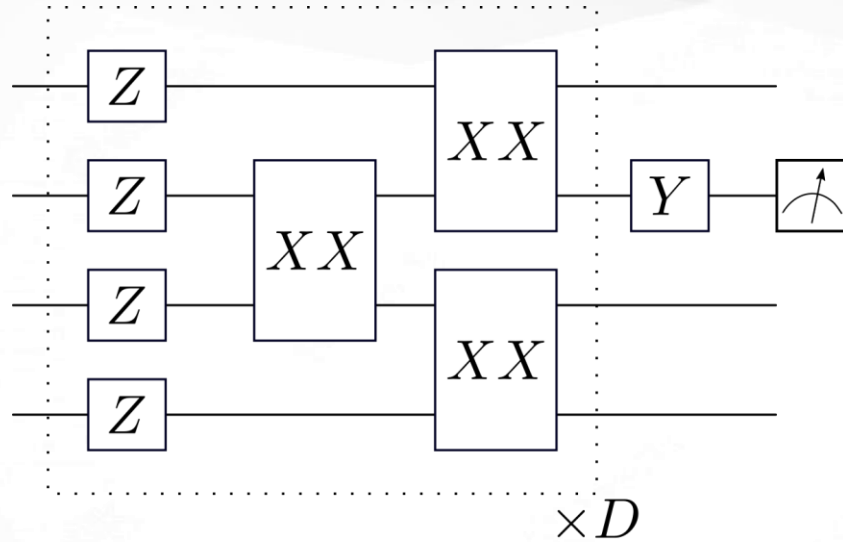
Non Fermionic State

$$|\psi\rangle = \alpha|0\rangle^{\otimes n} + \beta|1\rangle|0\rangle^{\otimes n-1}$$

Increase generalized globality:

$$O = X_j \in \mathcal{B}_{2j-1}$$

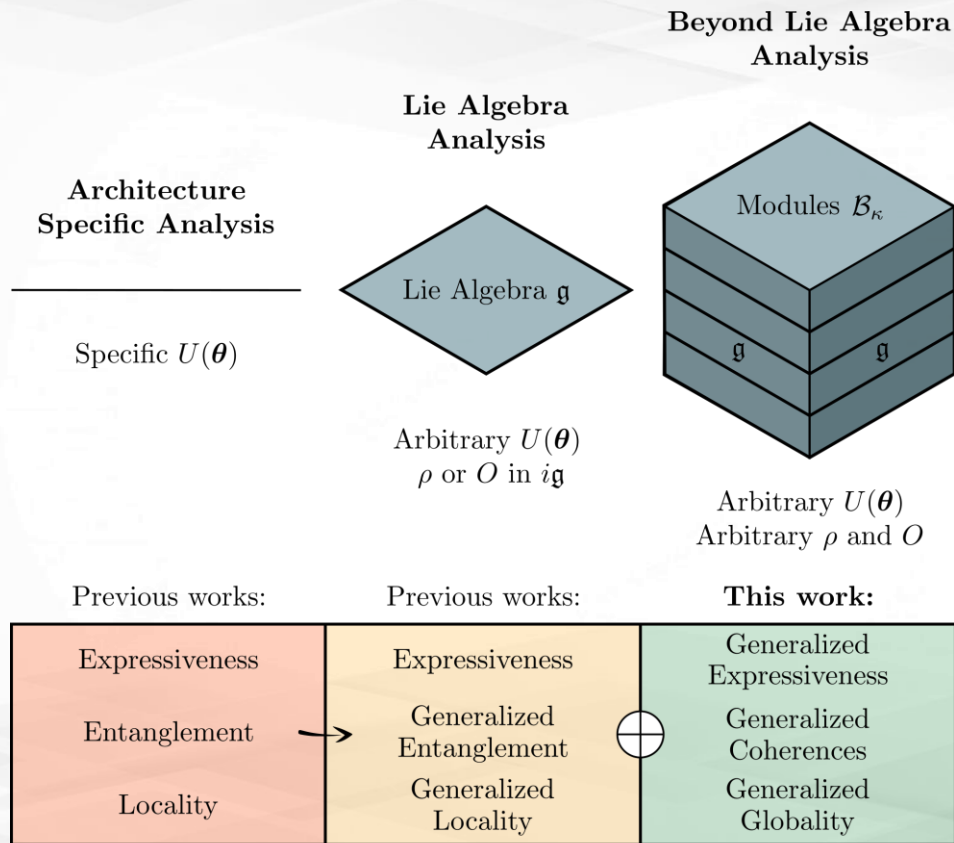
Enough is enough Marco! Ok, just one more cool result



$$Z_{n/2} \rightarrow X_{n/2}$$

$$\in \mathcal{B}_2 \quad \in \mathcal{B}_{n-1}$$

Towards a super duper, unified mega complete BP theory





Phew.. That was a lot... What did we learn?

IMHO: Our understanding of BPs underwent a recent **transformative shift**. We can use these results to **guide our model's design**.

Here are some **open questions**:

- In the Matchgate case, we can relate **trainability** to fermionic entanglement (a **computational resource** that can promote matchgates to universal circuits). We see that **e-v-e-r-y trainable model, is also (somehow*) classically simulable.**

We even propose a simulation method for exponential magic states that are non-simulable via Wick theorem simulation-based techniques!

Is this connection more general?

- What about noise?
- What about shallow depth?
- Non-expectation value-based models (generative)?

Thanks for your attention!



Michael Ragone
UC Davis



Frederic Sauvage
Ex-LANL
(Quantinuum)



Martin Larocca
LANL



Bojko Bakalov
NCU



Sujay Kazi
Duke



Lex Kemper
NCU



Carlos Ortiz Marrero
PNNL



Diego Garcia Martin
LANL



Nahuel Diaz
LANL



LANL's 2024 Quantum Computing Summer School
Applications now open!
cerezo@lanl.gov

