# 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*" this week on arxiv.
N. L. Diaz, Diego García-Martín, Sujay Kazi, Martin Larocca, and **MC**

10/17/2023 @IPAM Workshop

# 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

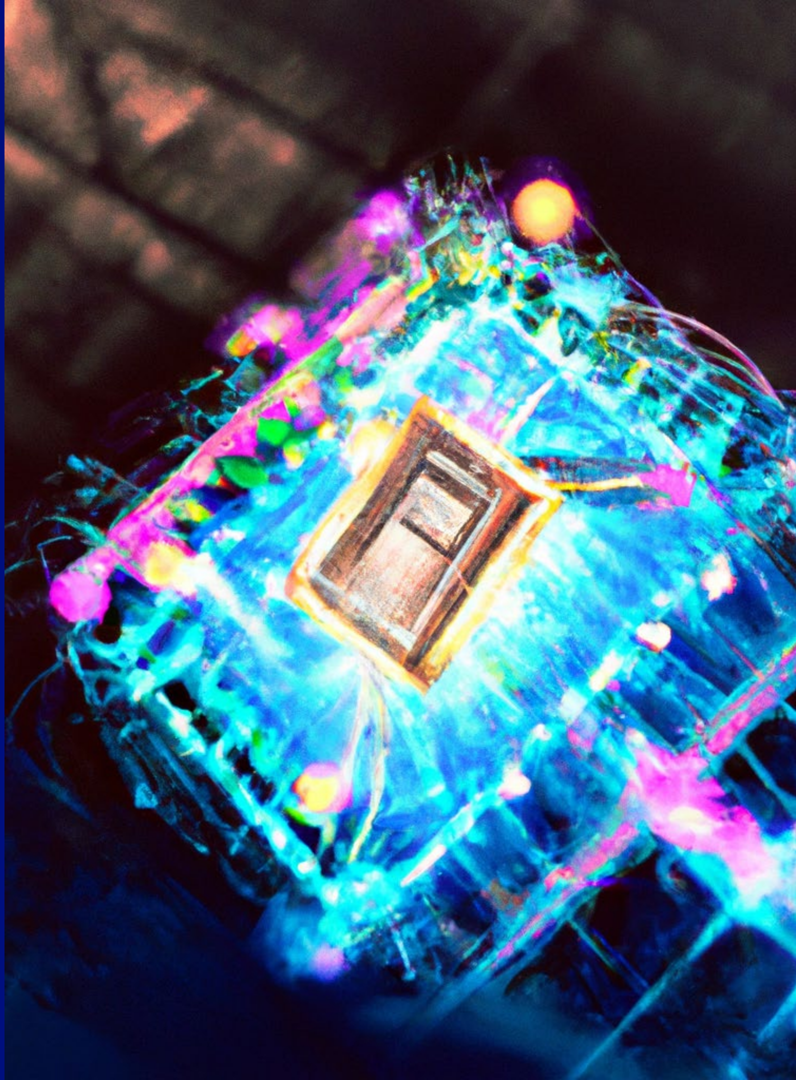
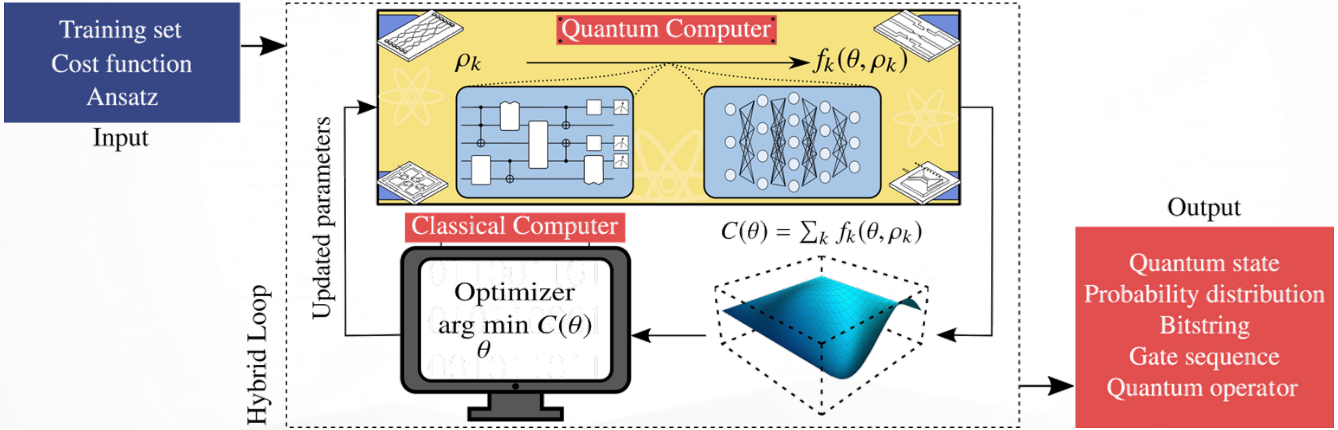


Image credits: Our AI overlord Dall-E 2

# 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:

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

We will assume (standard) $\|O\|_2^2 \leq 2^n$

Review: *Variational Quantum Algorithms,* **MC** et al, Nature Reviews Physics 3, 625-644 (2021)
Review: *Noisy intermediate-scale quantum algorithms,* K. Bharti et al, Rev. Mod. Phys. 94, 015004 (2022)

# How and when can this computational approach fail?

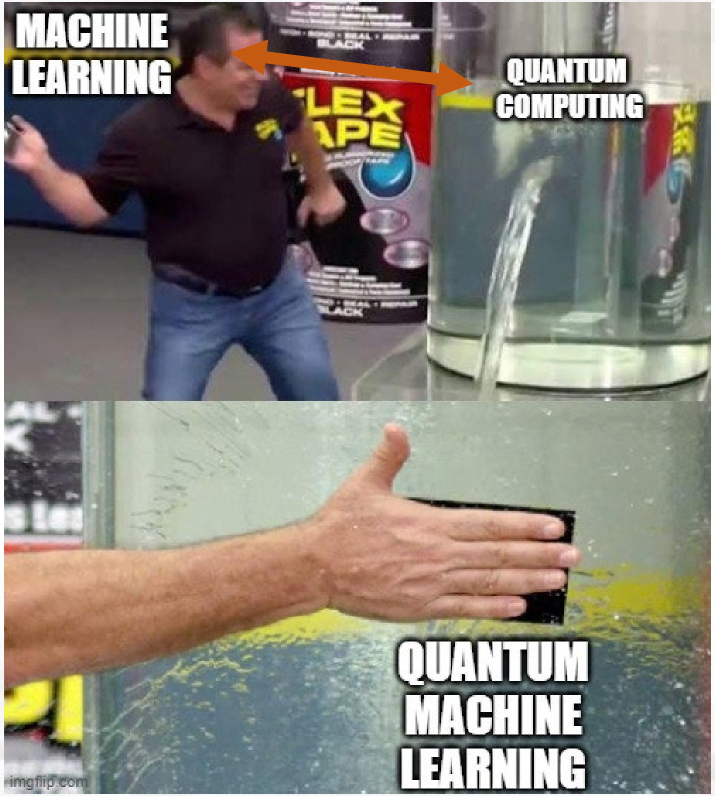
While NN are widely used today, their historical development saw periods of great stagnation (or winters).


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.



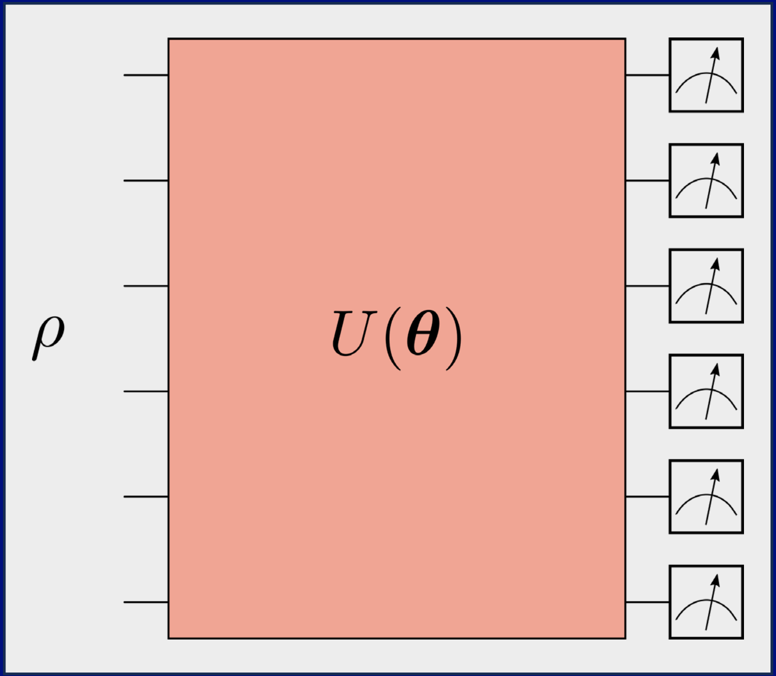
*Challenges and opportunities in quantum machine learning,* MC et al, Nature Computational Science 2, 567–576 (2022)

# 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 <3


- Global c… ially on all q…
- Too mu… Ps.
- Deep ci…

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

*They can be shown to be equivalent.*

*Equivalence of quantum barren plateaus to cost concentration and narrow gorges, A. Arrasmith, et al, Quantum Sci. Technol. (2021)*

# Barren Plateaus

No Barren Plateau

Barren Plateau

Image Credits: Samson Wang

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

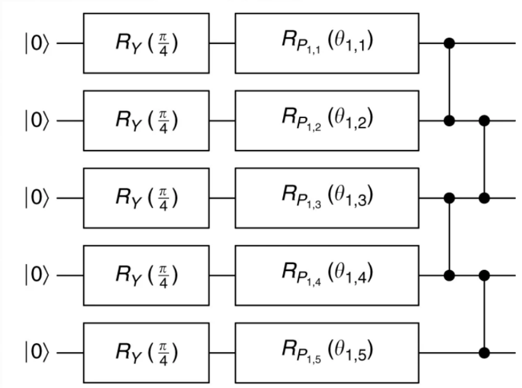
$\mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] \in \mathcal{O}(1/b^n)$ with $b > 1$.

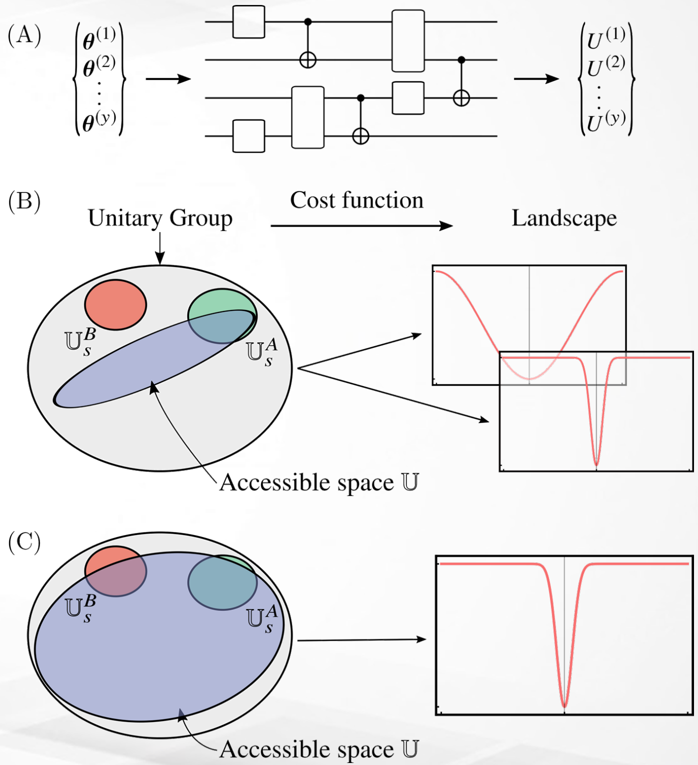
# Expressiveness in the circuit,

## or, a general form of No-Free-Lunch theorem

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

Seminal Google paper: $U(\boldsymbol{\theta})$ forms a 2-design




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

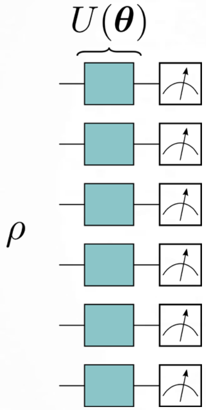
We have a BP for any $\rho$, and $O$.

*Barren Plateaus in Quantum Neural Networks,* McClean et al, Nat. Commun. (2018)
*Connecting ansatz expressibility to gradient magnitudes and barren plateaus, Z. Holmes, et al,PRX Quantum (2022).*

# Locality of measurement operator

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

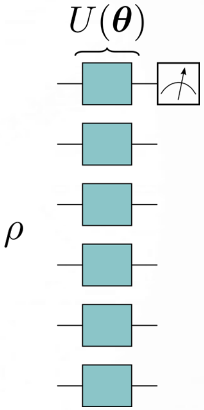
Let’s take the most inexpressive imaginable circuit and a global measurement $O = Z^{\otimes n}$




Always has a BP:

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

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




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

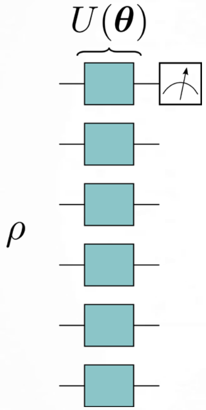
If input is $\rho = |0\rangle\langle 0|^{\otimes n}$, then no BP.. But…

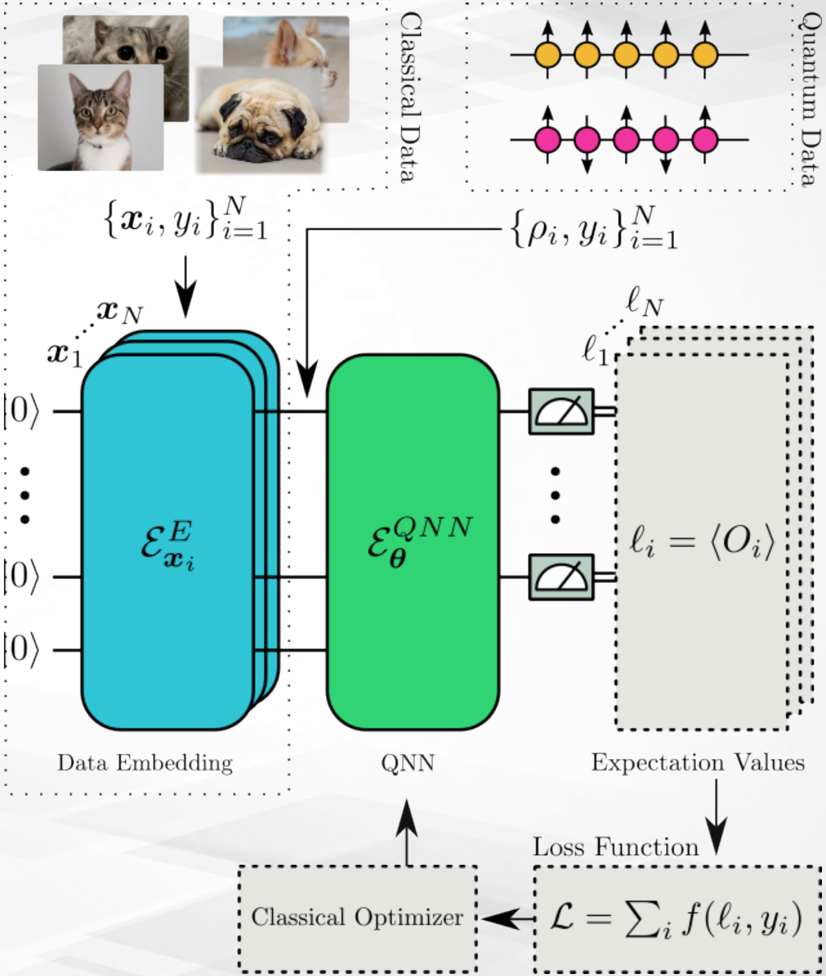
*Cost Function Dependent Barren Plateaus in Shallow Parametrized Quantum Circuits, MC et al, Nat. Commun. (2020)*

# Entanglement in the initial state

## or, why untamed entanglement can be bad

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



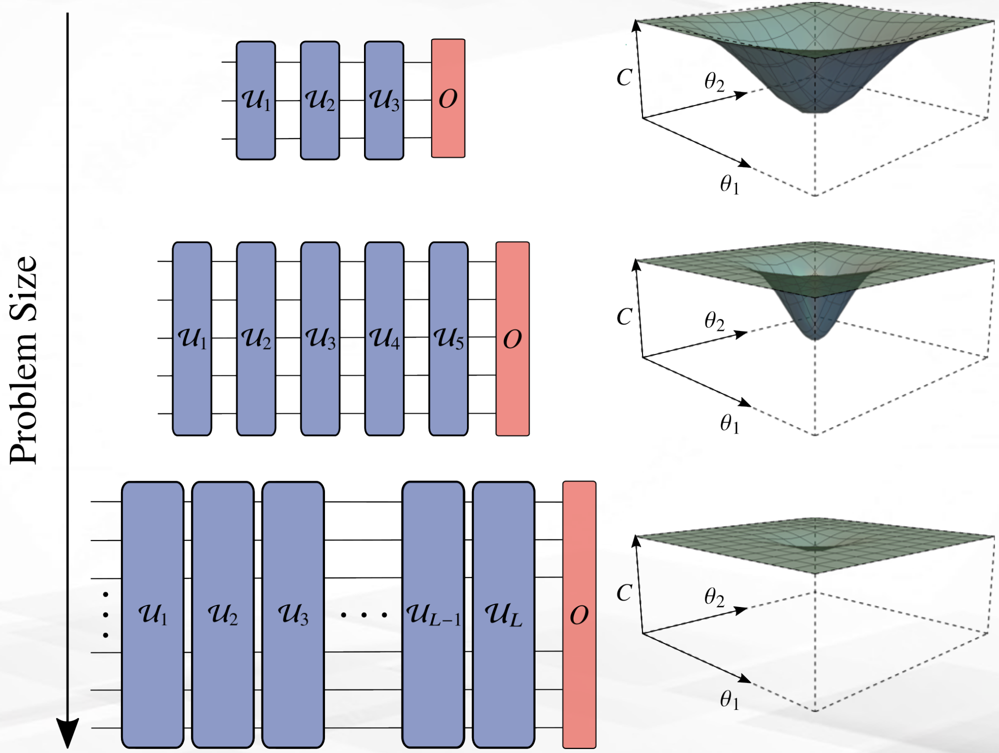

Has a BP: $\mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] \in \mathcal{O}(1/b^n)$

Despite no expressiveness and local measurements

*Subtleties in the trainability of quantum machine learning models, S. Thanasilp et al, arxiv (2021).*

# And of course, Noise

## or, why existence is pain




*Noise Induced-barren plateaus, S. Wang et al, Nat. Commun. (2021).*

# Previous understanding of BPs
## or, yes this slide looks ugly, but it will get better soon.

**Architecture Specific Analysis**

Specific $U(\boldsymbol{\theta})$

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 $\boldsymbol{\theta}$, $\boldsymbol{L}$.

*2) Quantifies the ultimate expressiveness* $U(\boldsymbol{\theta})$.

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

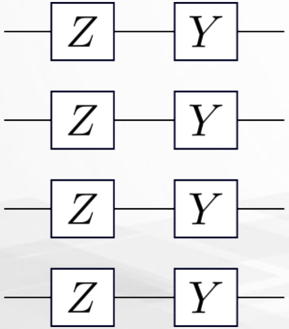
$C = A + B + \frac{1}{2}[A,B] + \frac{1}{12}[A,[A,B]] - \frac{1}{12}[B,[A,B]])$

# Examples of DLA

Single qubit rotations:

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

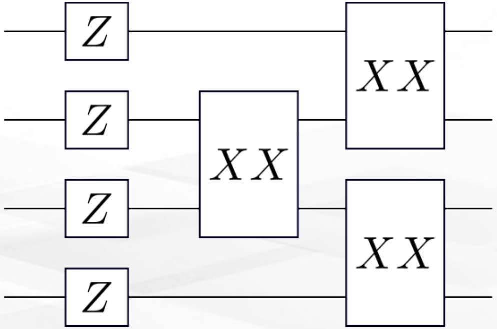
$$\mathfrak{g} = su(2)^{\oplus n}$$




Matchgate circuit:

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

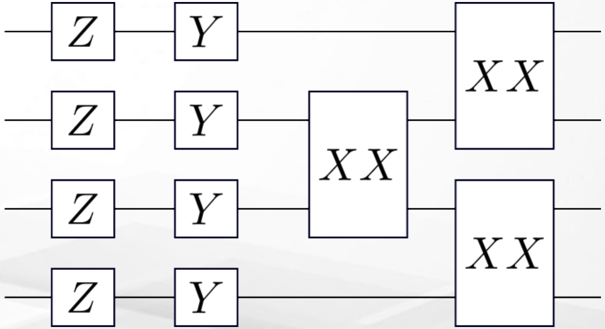
$$\mathfrak{g} = so(2n)$$




Universal circuit:

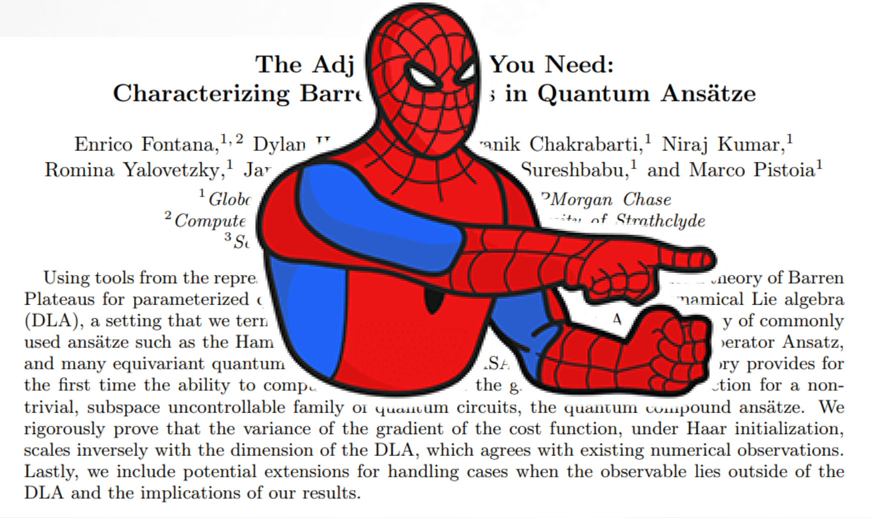
$$\mathcal{G} = \{Z_i, Y_i\}_{i=1}^{n} \cup \{X_i X_{i+1}\}_{i=1}^{n-1}$$

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

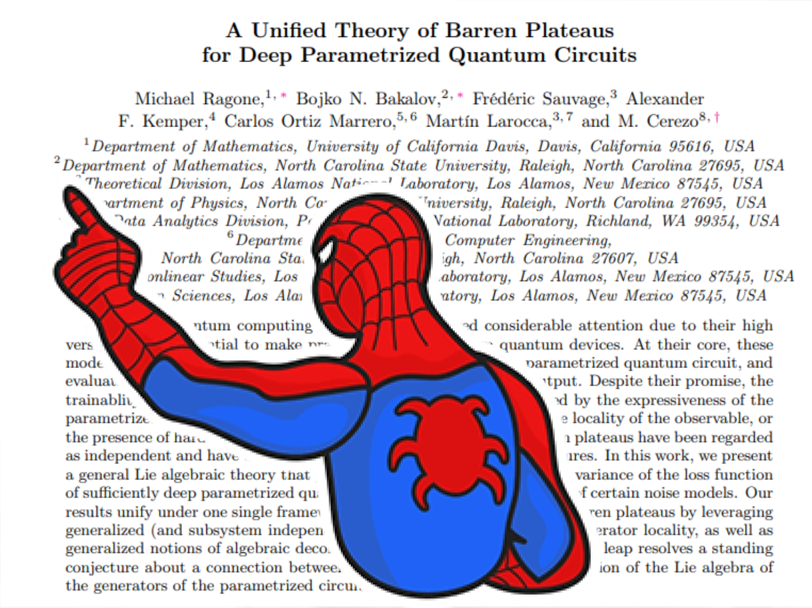


# Enough math. Why is the DLA important?

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.

[1] *Diagnosing Barren Plateaus with tools from quantum optimal control, M. Larocca, et al, Quantum. (2022).*

# 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

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

Then, recalling that any DLA

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

We have bounds for circuit depth

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

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

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

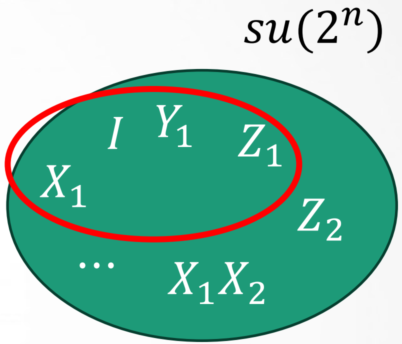
EXACT Formulae

*A Unified Theory of Barren Plateaus for Deep Parametrized Quantum Circuits,* M. Ragone et al, arxiv (2023)

# What can we learn from here?

Assume $\mathfrak{g}$ is simple

$$\mathrm{E}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] = 0$$
$$\mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] = \frac{\mathcal{P}_g(\rho)\mathcal{P}_g(O)}{\dim(\mathfrak{g})}$$

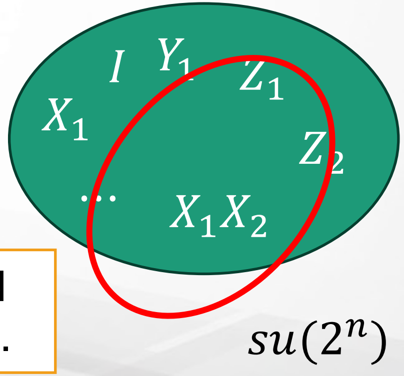



If either $\dim(\mathfrak{g}), 1/\mathcal{P}_g(\rho), 1/\mathcal{P}_g(O) \in \Omega(b^n)$ with $b > 2$, one has a BP!

There are 3 sources of BPs here!

Proves a conjecture in [1]!

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




Maximized for O in DLA

Maximized for a HWS.

[1] *Diagnosing Barren Plateaus with tools from quantum optimal control,* M. Larocca, et al, Quantum *(2022)*
*Generalizations of entanglement based on coherent states and convex sets*, H. Barnum et al, Phys. Rev. A (2003)
*A subsystem-independent generalization of entanglement,* H. Barnum et al, Phys. Rev. Lett. (2004)

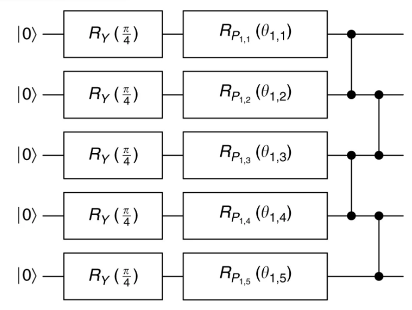
# Revisiting previous examples

Google's result:

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

$$\mathcal{P}_g(\rho) = \mathrm{Tr}[\rho^2] - 1/2^n \; (\rho = \rho_{\mathfrak{g}} - I/2^n)$$

$$\mathcal{P}_g(O) = \mathrm{Tr}[O^2] \; (O = O_{\mathfrak{g}} - \mathrm{Tr}[O]\, I/2^n)$$



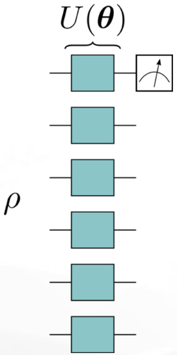

$$\mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] = \frac{1}{4^n - 1}(\mathrm{Tr}[\rho^2] - 1/2^n)(\mathrm{Tr}[O^2] - \mathrm{Tr}[O]^2/2^n)$$

Local measurement $O = Z_1$

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

$$\mathcal{P}_{g_1}(\rho) = \mathrm{Tr}[\rho_1^2] - 1/2^n \; (\rho_1 = \mathrm{Tr}_{\bar{1}}[\rho])$$

$$\mathcal{P}_g(O) = 2$$




Generalized entanglement = standard entanglement

Algebra IS local

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

*A Unified Theory of Barren Plateaus for Deep Parametrized Quantum Circuits,* M. Ragone et al, arxiv (2023)

# An important lesson learnt

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

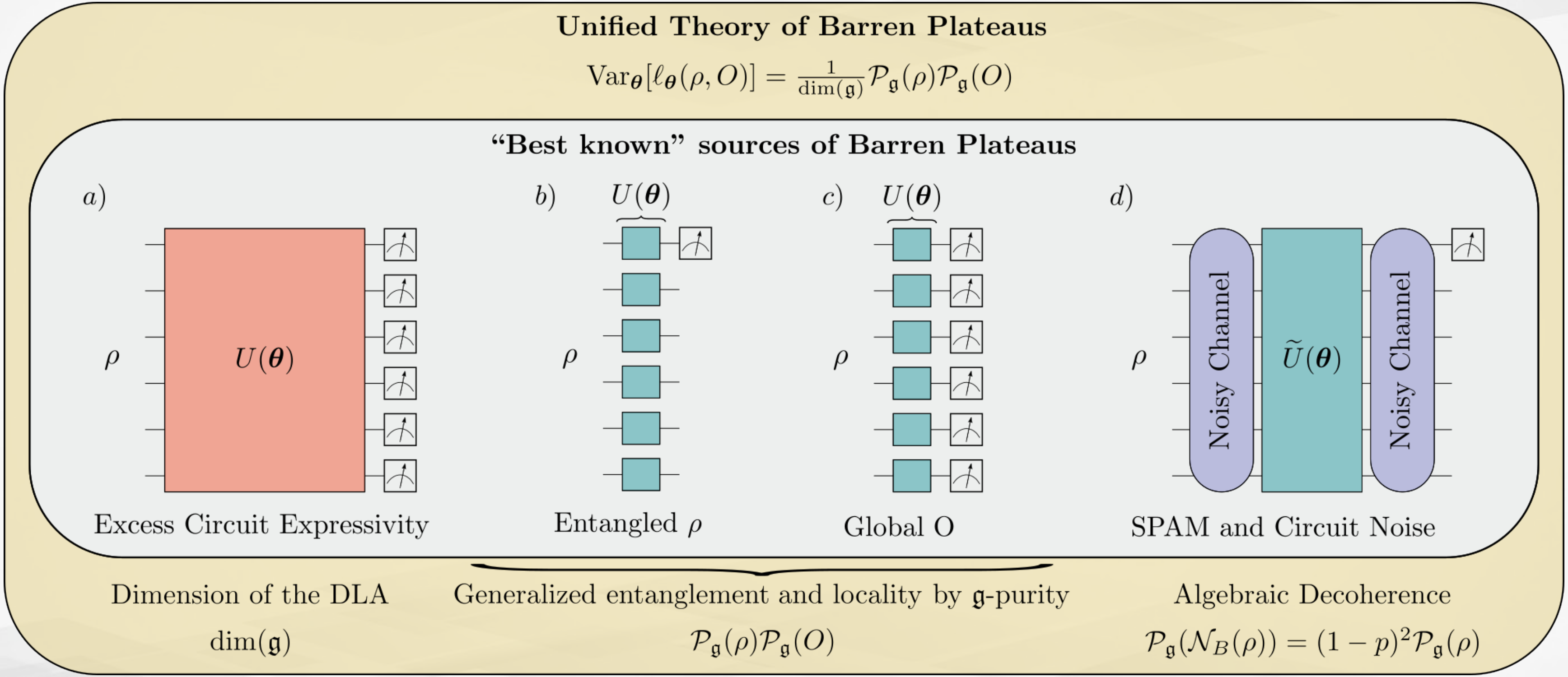
- Global observables (acting untrainable
- Too much entanglement le
- Deep circuits are have BPs


We show that generalized entanglement in $\rho$ 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.*

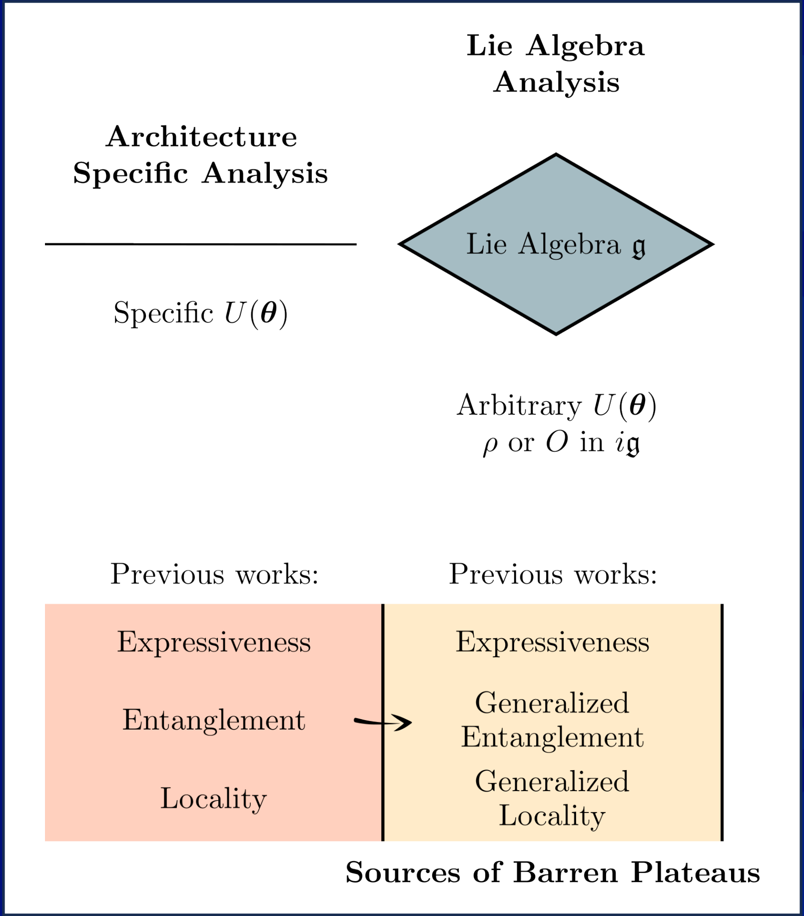
*A Unified Theory of Barren Plateaus for Deep Parametrized Quantum Circuits,* M. Ragone et al, arxiv (2023)

# Summary




*A Unified Theory of Barren Plateaus for Deep Parametrized Quantum Circuits,* M. Ragone et al, arxiv (2023)

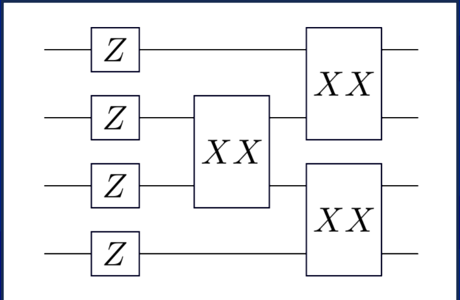

THAT'S A WRAP
FOLKS
memegenerator.net


Lie Algebra Analysis
Architecture Specific Analysis
Lie Algebra $\mathfrak{g}$
Specific $U(\boldsymbol{\theta})$
Arbitrary $U(\boldsymbol{\theta})$
$\rho$ or $O$ in $i\mathfrak{g}$
Previous works:
Previous works:
Expressiveness
Entanglement
Locality
Expressiveness
Generalized Entanglement
Generalized Locality
Sources of Barren Plateaus

# 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 $\rho$ or $O$ in $i\mathfrak{g}$. The DLA was the central element. Now we need to work with group modules.

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

$\forall M_\kappa \in \mathcal{B}_\kappa$, then $U M_\kappa U^\dagger \in \mathcal{B}_\kappa$ for any $U \in e^{\mathfrak{g}}$

Use Jordan-Wigner, and define Majorana operators

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

$$c_2 = YI \cdots I,\ c_3 = ZYI \cdots I,\ c_5 = Z \cdots ZY$$

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

# Moar exact variance formulae,

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

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

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

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

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

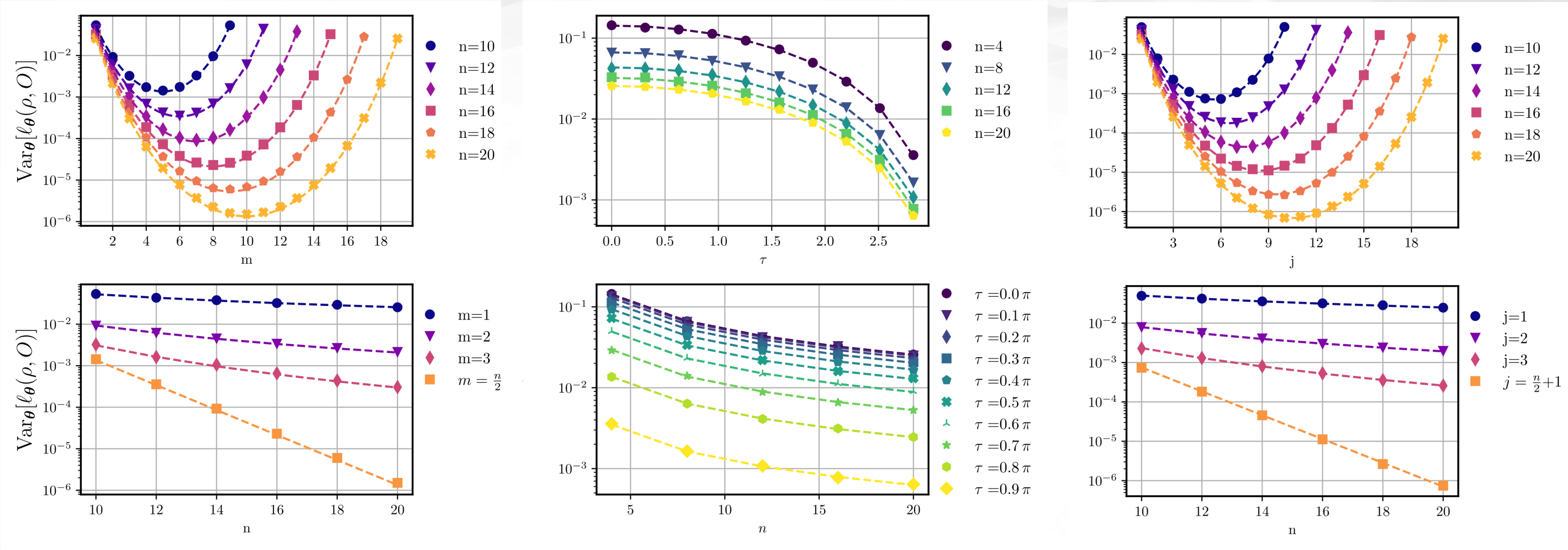
$\mathcal{P}_\kappa(M) = \langle M_\kappa, M_\kappa \rangle_I$, $\mathcal{C}_\kappa(M) = i^{\kappa \bmod 2}\langle M_\kappa, M_\kappa \rangle_P$, and $\langle M_1, M_2 \rangle_\Gamma = \mathrm{Tr}[\Gamma M_1^\dagger M_2]$

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!

*Showcasing a Barren Plateau Theory Beyond the Dynamical Lie Algebra,* NL Diaz et al, to appear on arxiv this week.
Entanglement in fermion systems, N. Gigena et al, Physical Review A (2015).

# 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)

$$|\Psi(\tau)\rangle = \left[\frac{|0000\rangle + |0011\rangle + |1100\rangle + e^{i\tau}\,|111\rangle}{2}\right]^{\otimes n/4}$$

Measure in the algebra (small dimension)

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

Non FermionicState

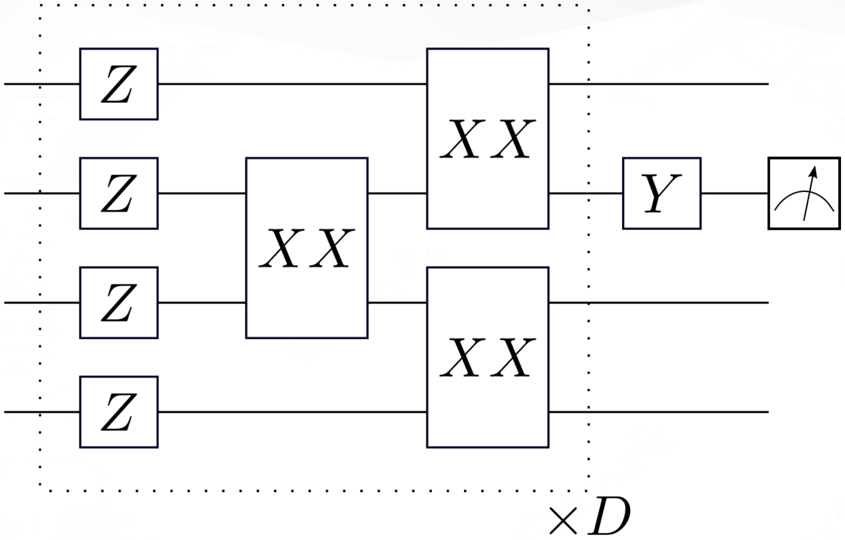
$$|\psi\,\rangle = \alpha|0\rangle^{\otimes n} + \beta|1\rangle|0\rangle^{\otimes n}$$

Increase generalized globality:

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

*Showcasing a Barren Plateau Theory Beyond the Dynamical Lie Algebra,* NL Diaz et al, to appear on arxiv this week.
*Classical simulation of non-gaussian fermionic circuits*, B. Dias arXiv (2023).

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

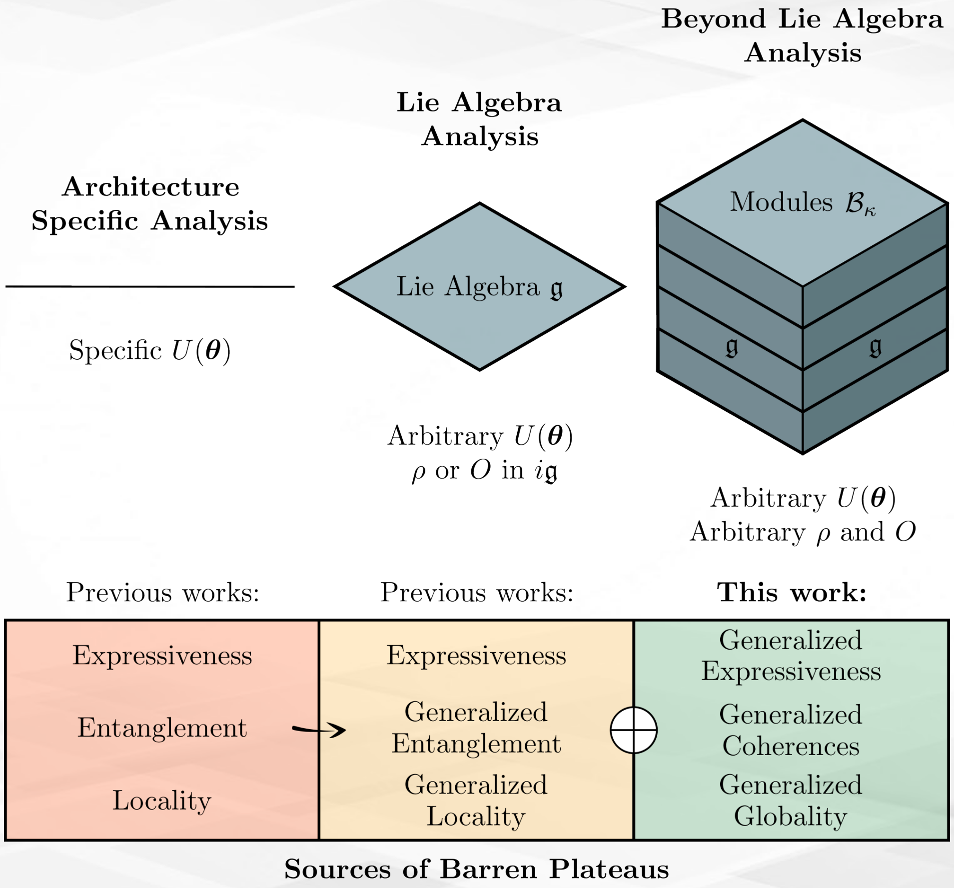



$$Z_{n/2} \to X_{n/2}$$

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

*Showcasing a Barren Plateau Theory Beyond the Dynamical Lie Algebra,* NL Diaz et al, to appear on arxiv this week.
*Classical simulation of non-gaussian fermionic circuits*, B. Dias arXiv (2023).

# Towards a super duper, unified mega complete BP theory




*Showcasing a Barren Plateau Theory Beyond the Dynamical Lie Algebra,* NL Diaz et al, to appear on arxiv this week.

# 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.***

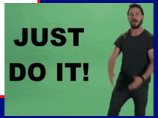
We even propose a simulation method for exponential magic states that are non-simulable via Wick theorem simulation-based techniques!)

Is this connection more general?

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

# Thanks for your attention!

This guy is amazing and looking for a job. HIRE HIM. Like NOW. No really, DO IT. DO IT NOW! He's in the audience rn!




*Michael Ragone*
*UC Davis*

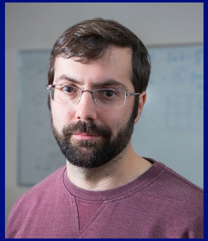
*Frederic Sauvage*
*LANL*

*Martin Larocca*
*LANL*

*Bojko Bakalov*
*NCU*

*Lex Kemper*
*NCU*

*Carlos Ortiz Marrero*
*PNNL*



*Diego Garcia Martin*
*LANL*

*Nahuel Diaz*
*LANL*


LANL's 2024 Quantum Computing Summer School
Applications opening soon!
cerezo@lanl.gov