nature portfolio

Peer Review File

A Lie algebraic theory of Barren Plateaus for Deep Parameterized Quantum Circuits

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The paper provides a unifying framework for understanding the barren plateaus in parameterized quantum circuits. The paper is clearly written and easy to follow. The theoretical analysis is spelled out very carefully with good level of detail and appears correct.

The results are clearly of great interest to the community, as indicated by a paper showing very similar (albeit slightly more general) results appearing on arXiv a few days prior to this manuscript's release ("The adjoint is all you need: Characterizing barren plateaus in quantum ansätze", arXiv:2309.07902, Ref 55 in the manuscript). The similarities between the two independently produced papers further testify to the accuracy of the conclusions of the paper and to their significance.

Reviewer #2 (Remarks to the Author):

Barren plateaus (BP) have been an active area of research over the last few years and a simple conceptual tie together could be quite helpful in understanding how to best sidestep this phenomenon. Given the wide spread study of this topic and relevance to research, the topic area is potentially high impact enough to warrant publication in Nature Comms. That said, there are some lingering questions for the authors that I think would help me better determine the impact of this work in particular. At the moment, I think I remain uncertain about the exact extent to which this work changes the state of our knowledge on the topic and whether it offers new directions forward in combatting this phenomenon.

One concern I have is that while the mathematics contained in the work are quite general, and likely correct, the use of the dynamical lie algebra seems like it is a bit restrictive in its applicability. In particular, there seem to be an incredibly small number of algebras and hence circuit ansatz for which one would not immediately get the entirety of the space. The examples listed, the transverse field ising model (or equivalently free fermions) references in the work is one of maybe 2 or three non-trivially separable cases I am aware of that don't

just immediately result in filling the space.

Moreover, if one fills the space, this is an indictment of a circuit in that family that could be, in principle, quite long (or infinitely so, to cover the space). It says little about the convergence rate or practical matters relating to performance of finite length circuits. Different choices of ansatz in the same DLA could lead to extremely different convergence rates as a function of the circuit depth, initialization strategy etc, but that can't really be captured here.

Perhaps what I'm learning here is that the use of DLAs is an attempt to circumvent BPs for an infinite length circuit within a certain family, and this is very difficult to do, and hence I should turn to formulations that allow me to make design choices at finite length, and I'm not sure these formulas apply. So my question to the authors is what can these formulas tell us about the finite circuit length case?

Maybe a related question here is if it might be possible to create some formalism of approximate or effective DLAs that are captured by finite time horizons and lightcones within circuits.

As a general note as well, given the appearance of a work by the many of the same authors "Showcasing a BP theory beyond the dynamical lie algebra", I wonder if it might be prudent to include a more sizable open questions section in the work. The word unified in the title implies a bit a closing of the theory, while the subsequent paper suggests it's not really wrapped up. It perhaps wouldn't be appropriate to ask the authors to reference the new work since it appeared later, but it may be a decent idea to clearly frame what questions remain unresolved as is relevant to the new work so it is easier to understand how they relate in context and the extent to which the authors believe this work concludes what is known about BPs and what is left unanswered.

Reviewer #3 (Remarks to the Author):

The authors present a method to integrate symmetries into examining and quantifying the

variance of loss functions in variational algorithms. Previously, this so-called barren plateau (BP) phenomenon has proven to be a roadblock in the development of quantum variational algorithms. The central results in the paper are:
(1) Incorporation of symmetries into barren plateau calculations.
(2) Simple mathematical equations that can calculate barren plateaus (though converting an ansatz into this mathematical formulation may be challenging)
(3) Discussion of practical implications such as how to incorporate SPAM noise into calculation of barren plateaus.

Overall, the paper is nicely written and the results are neat. It is nice that this paper confirms the intuition that the dimension of the dynamical Lie algebra corresponds generally to the severity of the barren plateau. Additionally, the proofs are easy to follow.

However, upon a critical read, I have some concerns about publishing especially in Nature Communications. None of the results are really surprising and the “unification” is not as unified as one may hope (more on this later). My main criticisms upon reflection are:
- The main technical drawback is that the paper makes a crucial assumption that hurts how general their approach is: namely, they assume circuits converge to a two design (see comment below about approximate two design assumption as well). This essentially sidesteps the main question which is “when does a given circuit ansatz actually become an approximate two design?” The aim to provide a holistic approach necessitates a deeper dive into this particular question, and this paper still leaves much blank in answering that question.
- Separate from technical details, the study of barren plateaus is, in my opinion, a relatively small piece of the quantum machine learning challenge as it is a statement about initialization and says nothing about a host of other untrainability issues that arise with quantum machine learning (and are more interesting in my admittedly biased opinion).
- Given its quantum-specific focus particularly on improving results in barren plateaus, one wonders if the paper's findings might be more suited for a specialized venue rather than a broad-spectrum journal. This isn't a critique of the research quality but a suggestion for its optimal placement.

More detailed questions and concerns:

- The paper presents a quantity in eq. (15) quantifying the rate of convergence, which is associated to the largest eigenvalue of some Markov like mixing process. However, this quantity is incredibly hard to calculate in practice. The main question still remains: given a circuit, how fast does it converge to a BP? Eq. 15 and its associated formal statement may help in answering that, but that formula is still incredibly complicated to estimate or calculate — so complicated that it renders the formula potentially useless. See [1] for more relevant techniques and ideas on this front.
- Barren plateaus expressed in terms of dynamical Lie algebras give a nice mathematical picture. However, similar to the above point, these dynamical Lie algebras can be elusive mathematical objects. I expect in many cases, we may not know what the dimensions of them are let alone how a given circuit ansatz spans this algebra. The work of [4] seems to indicate understanding this problem may not be so easy and rife with other issues as well.
- How relevant are these results in the setting of approximate symmetry or equivariance which is the setting one typically finds in practice? E.g. for classical deep learning, convolutional networks are not truly symmetric but approximately symmetric to translations.
- Beyond BPs, variational algorithms face a number of untrainability issues that are of practical relevance. E.g., as mentioned before for geometric learning, see [4] which studies disconnectedness in the geometric manifold of unitaries. These untrainability issues are in some ways related to BPs but also need to be mentioned for a more holistic picture.

Proofs were briefly checked and no errors were found. The proofs were rather nice and relatively easy to follow for someone with a reasonable but non-expert background in representation theory.

More minor comments:

- A description (or at least a reference) is needed for the statement where the authors say "In the following, we assume that the parametrized circuit is deep enough so that it forms

an approximate 2-design". After all, readers may not know what this is.

- Related to the above, I would also prefer that the authors state specifically in theorems that the distribution forms a 2-design rather than assuming it in a sentence early on and subsequently leaving it out of formal statements.
- Can authors give a citation for Lemma 3 where they say the fact is well known? I understand a proof comes after, but it would be good to still cite where this is from.
- I do not agree with the language surrounding a "unified theory" of barren plateaus. This paper really only generalizes the phenomenon to circuits constrained under symmetry and where we know enough about its Lie algebraic properties to perform calculations as mentioned before. Furthermore, many other papers make similar claims of unification (e.g. [1-3]), some which the authors cite, so I would also be careful in using this language.

References:

[1] Napp, John. "Quantifying the barren plateau phenomenon for a model of unstructured variational ans\"{a} tze." arXiv preprint arXiv:2203.06174 (2022).

[2] Zhao, Chen, and Xiao-Shan Gao. "Analyzing the barren plateau phenomenon in training quantum neural networks with the ZX-calculus." Quantum 5 (2021): 466.

[3] Diaz, N. L., et al. "Showcasing a Barren Plateau Theory Beyond the Dynamical Lie Algebra." arXiv preprint arXiv:2310.11505 (2023).

[4] Marvian, Iman. "Restrictions on realizable unitary operations imposed by symmetry and locality." Nature Physics 18.3 (2022): 283-289.

A Unified Theory of Barren Plateaus for Deep Parametrized Quantum Circuits

LETTER TO THE REFEREES

Dear Referees,

We would first like to thank the Referees for their time and they valuable suggestions on how to improve our work. In particular, we would like to thank the kind words by the Referees. Referee 1 stated "*The paper is clearly written and easy to follow. The theoretical analysis is spelled out very carefully with good level of detail and appears correct. The results are clearly of great interest to the community*", while Referee 2 said ""*a simple conceptual tie together could be quite helpful in understanding how to best sidestep this phenomenon [...] potentially high impact enough to warrant publication in Nature Comms.*" In a similar spirit Referee 3 wrote " *Overall, the paper is nicely written and the results are neat. It is nice that this paper confirms the intuition that the dimension of the dynamical Lie algebra corresponds generally to the severity of the barren plateau. Additionally, the proofs are easy to follow.*"

In the following reply, we address the points raised by the Referees. In particular, we have argued that our work has significantly improved our understanding of the Barren Plateau phenomenon, by not only unifying previous results, but by providing further interpretation and operational meaning to the terms that appear in the loss function variance. Here, we stress that our present work has paved the way for to the latest advancements in the study of barren plateaus. We have also addressed the concerns by Referee 2 and 3 about how our results would hold for circuits that have finite depth and which are not 2-designs. In this context, we have added a new theorem (current Theorem 3) that quantifies how the variance will deviate from that of Theorem 1 for finite depth circuits.

Here, are the main changes to our manuscript:

- We have clarified that Theorem 1 holds for circuit forms a 2-design. As such, we now say that Theorem 2 quantifies how close a general finite-depth circuit's unitary distributions are from being a 2-design, with Theorem 3 studying how much the variances for an exact and for an approximate design deviate.

- We have added a new section in the Methods that provides additional intuition for the Dynamical Lie Algebra (DLA), arguing that in general the DLA will contain multiple ideals. Here, we also argue that while well-known DLA being semi-simple (e.g., either free-fermionic or universal) is an artifact from the the DLA study is still in its infancy.

- We have added a discussion in the outlook explaining how our results bring a novel perspective to the barren plateau study.

- We have added a new section on the Methods with Theorem 3, with the corresponding proof in the Supplemental Information.

- We have improved the clarity and readability of our proofs (e.g., by adding the new Lemma 7).

- We have also added in the Supplemental Information a use-case example of Theorem 2, where we exactly compute the spectral gap of the circuit's distribution second moment, thus illustrating that the bounds in Theorems 2 and 3 are not only useful, but also computable.

Overall, the helpful comments made by the Referees allowed us to substantially improve the quality of the manuscript, which we think meets the acceptance criteria for *Nature Communications*. Below, we provide a detailed response to the Referee comments, as well as attach a diff file highlighting the changes to our work.

Looking forward to your answer.

Michael Ragone, Bojko N. Bakalov, Frédéric Sauvage, Alexander F. Kemper, Carlos Ortiz Marrero, Martín Larocca, and M. Cerezo

REPLY TO REFEREE 1

First referee: The paper provides a unifying framework for understanding the barren plateaus in parameterized quantum circuits. The paper is clearly written and easy to follow. The theoretical analysis is spelled out very carefully with good level of detail and appears correct. The results are clearly of great interest to the community, as indicated by a paper showing very similar (albeit slightly more general) results appearing on arXiv a few days prior to this manuscript's release ("The adjoint is all you need: Characterizing barren plateaus in quantum ansätze," arXiv:2309.07902, Ref 55 in the manuscript). The similarities between the two independently produced papers further testify to the accuracy of the conclusions of the paper and to their significance.

Response: We thank the Referee for taking the time to review our manuscript and for their support and positive feedback.

REPLY TO REFEREE 2

Second referee: Barren plateaus (BP) have been an active area of research over the last few years and a simple conceptual tie together could be quite helpful in understanding how to best sidestep this phenomenon. Given the wide spread study of this topic and relevance to research, the topic area is potentially high impact enough to warrant publication in Nature Comms. That said, there are some lingering questions for the authors that I think would help me better determine the impact of this work in particular.

Response: First, we would like to thank the Referee for their feedback on our work. As the Referee points out, we do think that our work is extremely timely given the large amount of interest that variational quantum algorithms and quantum machine learning have accrued over the last few years. We hope that our comments to the Referee's questions will convince them that our work does indeed warrant publication in Nature Communications.

At the moment, I think I remain uncertain about the exact extent to which this work changes the state of our knowledge on the topic and whether it offers new directions forward in combating this phenomenon.

To understand how our work changes our knowledge about Barren Plateaus (BPs), allow us to make a (very brief) review of this topic's chronology. First, we recall that the first results on BPs were derived for unstructured hardware-efficient ansatzes [1, 2]. From here, several *exact* variance calculations were performed (for instance in permutation equivariant circuits [3], dissipative quantum neural networks [4], among others), but these were applicable only to specific circuit architectures. In [5], the authors numerically analyzed several architectures and conjectured that a general formalism for BPs would likely involve the Dynamical Lie Algebra (DLA). More importantly, in all the manuscripts reporting exact variance calculations it was obvious that while the size of the DLA $\dim(\mathfrak{g})$ indeed played a role in determining the scaling of the variance, so did the choice of the operator O and the choice of the initial state ρ. As such, we can safely claim that prior to our work the state of the field was such that:

1. Results were case-by-case dependent;
2. It was conjectured, but not proven, that the DLA $\mathfrak{g}$ was important;
3. It was known that the initial state ρ and measurement operator O played a role, but it wasn't obvious how they affected the variance.

As we argue in the main text of the manuscript, our work significantly refine our understanding of BPs as it:

1. Presents a general framework that encompasses and generalize known results;
2. Precisely characterizes the role of the DLA;
3. Quantifies the exact role of the initial state ρ and measurement operator O in terms of the DLA.

In particular, we argue that while our results hold for states or measurements in the DLA, this is the most standard case in the literature (e.g., many instances of Variational Quantum Eigensolver and Quantum Approximate Optimization Algorithm satisfy this assumption).

In addition, we would like to highlight the important fact that we don't just provide a mathematical equation for the variance, but also provide an operational interpretation of the contributing terms as generalized forms of entanglement and locality. Such a result thus significantly enhances how we understand BPs and allows us to show that many results taken as true in the literature are indeed false. For instance, it is widely believed that parametrized quantum circuits will always have barren plateaus if the initial state is "too entangled" or if the measurement operator is "too global." These misconceptions originate from extrapolating results that are valid for one specific architecture (e.g., [2, 6]), and assuming that they are generically true. Yet, as we argue in our discussion, we can readily see from our theorems that

BPs do not arise as a consequence of standard notions of entanglement or globality, but rather as a consequence of the *generalized* entanglement and the *generalized* globality, which are defined in terms of the DLA, and may greatly differ from their standard counterparts.

Taken together, we do believe that our results can give rise to new directions for combating the BP phenomenon, as we can pick initial states and measurements that are well aligned with the algebra, and not following misconceptions propagated in the community.

> One concern I have is that while the mathematics contained in the work are quite general, and likely correct, the use of the dynamical lie algebra seems like it is a bit restrictive in its applicability. In particular, there seem to be an incredibly small number of algebras and hence circuit ansatz for which one would not immediately get the entirety of the space. The examples listed, the transverse field ising model (or equivalently free fermions) references in the work is one of maybe 2 or three non-trivially separable cases I am aware of that don't just immediately result in filling the space.

We thank the Referee for their comments. While we agree that most commonly referenced algebras are either controllable/universal (i.e., $\mathfrak{g} = \mathfrak{su}(2^n)$) or somehow extremely structured (e.g., like the Free Fermionic one $\mathfrak{g} = \mathfrak{so}(2n)$); we must highlight that this is an artifact of the fact that very little effort has been put forward towards computing DLAs. In our opinion, the field of DLA study is still in its infancy, and more work needs to be done for determining simpler ways of computing the DLA. For instance, a systematic study of DLAs was initiated in the recent paper Ref. [7], which contains many new examples of DLAs.

In particular, we respectfully disagree with the statement that "*there seems to be an incredibly small number of algebras and hence circuit ansatz for which one would not immediately get the entirety of the space*". To see that this is the case let us note that for any ansatz containing symmetries, we know that $\mathfrak{g}$ cannot be equal to $\mathfrak{su}(2^n)$, but must be a proper subalgebra. Specifically, let us denote as $\mathrm{comm}(\mathcal{G})$ the commutant of the set of generators $\mathcal{G}$ (i.e., its symmetries), given by

$$\mathrm{comm}(\mathcal{G}) = \{A \in \mathbb{C}^{2^n \times 2^n} \;\mid\; [A, H] = 0 \;\forall H \in \mathcal{G}\}\,, \tag{1}$$

and as $\mathfrak{z}$ the center of $\mathrm{comm}(\mathcal{G})$, i.e.,

$$\mathfrak{z} = \{B \in \mathrm{comm}(\mathcal{G}) \;\mid\; [B, L] = 0 \;\forall L \in \mathrm{comm}(\mathcal{G})\}\,. \tag{2}$$

Then, we know that the DLA will decompose as

$$\mathfrak{g} = \bigoplus_{\lambda=1}^{\dim(\mathfrak{z})} \mathbb{1}_{m_\lambda} \otimes \mathfrak{g}_\lambda\,, \tag{3}$$

with m_λ denoting the multiplicity of the λ-th irrep. We may interpret such a direct sum decomposition as saying "there exists a basis of our Hilbert space which block-diagonalizes the matrices in $\mathfrak{g}$." Equation (3) explicitly shows that ansatzes with non-trivial commutants give rise to complex DLAs that are not universal.

For concreteness, let us consider the example of the Quantum Alternating Operator Ansatz (QAOA) [8] for a MaxCut problem. In this case, the DLA will likely neither be controllable nor be a Free Fermion one, as its generator symmetries can be non-trivial (this is especially true for realistic problems, which contain high levels of symmetry). Recall that the standard QAOA ansatz for a graph with edge set E can be expressed as

$$U(\boldsymbol{\theta}) = \prod_{l=1}^{L} e^{-iH_M\theta_{l,2}} e^{-iH_P\theta_{l,1}}\,, \tag{4}$$

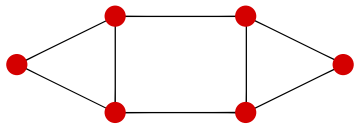

Figure 1. **Example Graph.** Example of a graph with six nodes.

where

$$H_P = \sum_{(i,j)\in E} Z_i Z_j \,, \qquad H_M = \sum_{i=1}^{n} X_i \,, \tag{5}$$

are respectively called problem and mixer Hamiltonians. Hence, we have $\mathcal{G} = \{H_M, H_P\}$ and we can see that the generators of the ansatz have three types of symmetries:

1. Parity symmetry $\mathbb{Z}_2$ (arising from the fact that the cut value is invariant under a bitwise negation of the cut value),
2. Automorphism symmetry (spanning from the node permutations that leave the graph, and thus the cut value unchanged),
3. Symmetries arising from the fact that we are using local gates [9–12].

For example, consider the graph in Fig. (1). Here we have that the automorphism group is $\mathbb{Z}_2 \times \mathbb{Z}_2$ (reflection through the vertical and horizonal axes) and a direct calculation using the Magma software package [13] reveals that

$$\mathfrak{g} = \mathfrak{su}(14) \oplus \mathfrak{su}(9) \oplus \mathfrak{su}(6) \oplus (\mathbb{1}_3 \otimes \mathfrak{su}(3)) \oplus \mathfrak{so}(10) \oplus (\mathbb{1}_2 \otimes \mathfrak{su}(1)) \,. \tag{6}$$

Hence, we can see that the algebra's dimension of 381, obtained from Eq. (6), is much smaller than the total dimension of 2046 and that its structure is non-universal, and very much non-trivial. [1]

With the previous being said, we can only predict that as our understanding and ability to calculate DLAs increases, our results will become more and more important to study the trainability of structured parametrized quantum circuits.

To clarify the fact that most DLAs are expected to be non-trivial, we have included a new section in the Methods entitled "*Symmetries and the DLA structure*", which we reproduce here for convenience of the reader:

In Eq. (6), we have argued that the DLA can be generically decomposed as a direct sum of commuting ideals. Here we provide additional intuition as to why this is the case. First, let us begin by denoting as $\mathrm{comm}(\mathcal{G})$ *the commutant of the set of generators* $\mathcal{G}$ *(i.e., the symmetries of the gate generators), that is,*

$$\mathrm{comm}(\mathcal{G}) = \{A \in \mathbb{C}^{2^n \times 2^n} \;\mid\; [A, H] = 0 \;\forall H \in \mathcal{G}\} \,, \tag{7}$$

and as $\mathfrak{z}$ *the center of* $\mathrm{comm}(\mathcal{G})$*, i.e.,*

$$\mathfrak{z} = \{B \in \mathrm{comm}(\mathcal{G}) \;\mid\; [B, L] = 0 \;\forall L \in \mathrm{comm}(\mathcal{G})\} \,. \tag{8}$$

The importance of the commutant $\mathrm{comm}(\mathcal{G})$ *and its center* $\mathfrak{z}$ *span from the fact that they determine the invariant subspaces of the Hilbert space under the action of* $G = e^{\mathfrak{g}}$*. Moreover, the dimension of the center is directly related to the DLA's decomposition as we can rewrite the decomposition of* $\mathfrak{g}$ *into commuting ideals as*

$$\mathfrak{g} = \bigoplus_{\lambda=1}^{\dim(\mathfrak{z})} \mathbb{1}_{m_\lambda} \otimes \mathfrak{g}_\lambda \,, \tag{9}$$

[1] We note that we have not added this example to the main text as it was borrowed from a different project that we will soon upload as a pre-print.

with m_λ denoting the multiplicity of the representation $\mathfrak{g}_\lambda$ of $\mathfrak{g}$. We may interpret such a direct sum decomposition as saying "there exists a basis of our Hilbert space that block-diagonalizes the matrices in $\mathfrak{g}$."

Equation (9) shows that a circuit with symmetries will have a rich and complex DLA structure, while universal, unstructured circuits will likely lead to controllable circuits with $\mathfrak{g} = \mathfrak{su}(2^n)$. In particular, we note that while there has been significant recent efforts for computing DLAs [3, 7, 9–12, 14], this field is still in its infancy, and we predict that as our understanding and ability to calculate the DLAs increases, the results in our present work will become more and more important to study the trainability of structured parametrized quantum circuits.

> Moreover, if one fills the space, this is an indictment of a circuit in that family that could be, in principle, quite long (or infinitely so, to cover the space). It says little about the convergence rate or practical matters relating to performance of finite length circuits. Different choices of ansatz in the same DLA could lead to extremely different convergence rates as a function of the circuit depth, initialization strategy etc, but that can't really be captured here. Perhaps what I'm learning here is that the use of DLAs is an attempt to circumvent BPs for an infinite length circuit within a certain family, and this is very difficult to do, and hence I should turn to formulations that allow me to make design choices at finite length, and I'm not sure these formulas apply. So my question to the authors is what can these formulas tell us about the finite circuit length case? Maybe a related question here is if it might be possible to create some formalism of approximate or effective DLAs that are captured by finite time horizons and lightcones within circuits.

We would like to again thank the Referee for their question. First, we would like to note that our results do not require circuits with infinite depth. As stated in the main text, for our theorems to hold, the circuit must be deep enough so that it *only* forms an approximate 2-design over the Lie group $G = e^{\mathfrak{g}}$. When we say deep enough, we do not mean that the number of layers L goes to infinity. In fact, we can bound the number of layers for the circuit to be an ϵ-approximate 2-design via Theorem 2, which says that $U(\boldsymbol{\theta})$ will form an ϵ-approximate 2-design over G when

$$L \geqslant \frac{\log(1/\epsilon)}{\log\left(1/\left\|\mathcal{A}^{(2)}_{\mathcal{E}_1}\right\|_\infty\right)}. \tag{10}$$

Let us recall the definition of this norm. The Schatten p-norms for $p = 1, \infty$ on the spaces of operators $\mathfrak{gl}(\mathcal{H}^{\otimes t})$ are spectral norms: given the singular values $s_i(A) \geqslant 0$ of A, we may express them as

$$\|A\|_1 = \sum_i s_i(A), \qquad \|A\|_\infty = \max_i s_i(A). \tag{11}$$

This induces corresponding operator norms for linear maps $\mathcal{M} : \mathfrak{gl}(\mathcal{H}^{\otimes t}) \to \mathfrak{gl}(\mathcal{H}^{\otimes t})$ by the usual

$$\|\mathcal{M}\|_p = \sup_{\substack{A \neq 0 \\ A \in \mathfrak{gl}(\mathcal{H}^{\otimes t})}} \frac{\|\mathcal{M}(A)\|_p}{\|A\|_p}. \tag{12}$$

While Eq. (10) shows that we need a finite number of layers to obtain an ϵ-approximate 2-design over G, we do agree that it does not tell us by how much the variance will deviate from that in Theorem 1 for finite depth. To account for this limitation in our manuscript, we have derived a new theorem that tells us about the variance for finite L-layered circuits that need not form a 2-design. To present this new theorem we have added a new section to the Methods entitled "*Variance for a circuit that does not form a 2-design over G*," which reads:

In the main text, we have assumed that the circuit is deep enough so that it forms a 2-design (or an approximate one) over G, and Theorem 2 provides bounds on the necessary number of layers for this to happen. Here, we instead ask the question: How different will the variance be from that in Theorem 1 if the circuit does not form an approximate 2-design over G?

For this purpose, let us consider an L-layered circuit. In what follows, we do not assume that the circuit forms an approximate 2-design over G. Just as before, we denote the ensemble of unitaries generated by the L-layered circuit $U(\boldsymbol{\theta})$

by $\mathcal{E}_L$, and we define its t-th moment superoperator $\mathcal{M}_{\mathcal{E}_L}$. Then, denoting as $\mathrm{Var}_{\mathcal{E}_L}$ and Var_G the variance obtained by sampling circuits over the distribution $\mathcal{E}_L$ and over the Haar measure over G, respectively, we find that the following theorem holds.

Theorem 3. *Let $\mathfrak{g} \subseteq \mathfrak{u}(2^n)$ be any dynamical Lie Algebra, and suppose either ρ or O are in $i\gamma$ with ρ a density matrix. The difference between the loss function variance of an L-layered circuit, with ensemble of unitaries $\mathcal{E}_L$, and the variance for a circuit that forms a 2-design over G can be bounded as*

$$|\mathrm{Var}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)] - \mathrm{Var}_G[\ell_{\boldsymbol{\theta}}(\rho, O)]| \leqslant 3\left(\left\|\mathcal{A}^{(2)}_{\mathcal{E}_1}\right\|_\infty\right)^L \|O\|_1^2, \tag{13}$$

where the usual Schatten p-norm and induced Schatten p-norm are given by Eq. (11) and Eq. (12).

Theorem 3 shows that the variance of an L-layered circuit will converge exponentially fast in L to the variance obtained by assuming that $U(\boldsymbol{\theta})$ forms a 2-design. The rate of convergence is again determined by the single layers expressiveness $\left\|\mathcal{A}^{(2)}_{\mathcal{E}_1}\right\|_\infty$. The proof of Theorem 3 is given in the Supplemental Information.

For completeness, we also present here the proof of Theorem 3, which has been added to the Supplemental Information:

Proof. Let us begin by observing that we may split

$$\begin{aligned} |\mathrm{Var}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)] - \mathrm{Var}_G[\ell_{\boldsymbol{\theta}}(\rho, O)]| &= \left|\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)]^2 - (\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)])^2 + \mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho, O)]^2 - (\mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho, O)])^2\right| \\ &\leqslant \left|\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)]^2 - \mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho, O)]^2\right| + \left|(\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)])^2 - (\mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho, O)])^2\right|. \end{aligned} \tag{14}$$

To estimate the first term of Eq. (14), we write the second moments in terms of the second moment operators of ensembles $\mathcal{E}$, which again are linear maps $\mathcal{M}^{(2)}_{\mathcal{E}} : \mathfrak{gl}(\mathcal{H})^{\otimes 2} \to \mathfrak{gl}(\mathcal{H})^{\otimes 2}$,

$$\begin{aligned} \left|\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)]^2 - \mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho, O)]^2\right| &= \left|\mathrm{Tr}\left[\mathcal{M}^{(2)}_{\mathcal{E}_L}(\rho^{\otimes 2})O^{\otimes 2}\right] - \mathrm{Tr}\left[\mathcal{M}^{(2)}_{G}(\rho^{\otimes 2})O^{\otimes 2}\right]\right| \\ &= \left|\mathrm{Tr}\left[(\mathcal{M}^{(2)}_{\mathcal{E}_L} - \mathcal{M}^{(2)}_{G})(\rho^{\otimes 2})O^{\otimes 2}\right]\right| \\ &\leqslant \left\|\rho^{\otimes 2}\right\|_1 \left\|O^{\otimes 2}\right\|_1 \left\|\mathcal{A}^{(2)}_{\mathcal{E}_L}\right\|_\infty \\ &\leqslant \left\|O^{\otimes 2}\right\|_1 \left\|\mathcal{A}^{(2)}_{\mathcal{E}_L}\right\|_\infty, \end{aligned} \tag{15}$$

where we have Hölder's inequality for matrices and that $\left\|\rho^{\otimes 2}\right\|_\infty \leqslant 1$ since ρ a density matrix and so has eigenvalue at most 1.

To estimate the second term of Equation (14), we use the difference of squares formula for two real numbers $a^2 - b^2 = (a+b)(a-b)$ to write

$$\left|(\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)])^2 - (\mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho, O)])^2\right| = |\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)] + \mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho, O)]| \cdot |\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)] - \mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho, O)]|. \tag{16}$$

Recalling that for any ensemble $\mathcal{E}$ the first moment operator $\mathcal{M}^{(1)}_{\mathcal{E}} : \mathfrak{gl}(\mathcal{H}) \to \mathfrak{gl}(\mathcal{H})$ has operator norm $\left\|\mathcal{M}^{(1)}_{\mathcal{E}}\right\|_\infty \leqslant 1$ by Lemma 2, we see that

$$|\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)] + \mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho, O)]| = \left|\mathrm{Tr}\left[(\mathcal{M}^{(1)}_{\mathcal{E}_L} + \mathcal{M}^{(1)}_{G})(\rho)O\right]\right| \leqslant \|O\|_1 \left\|\mathcal{M}^{(1)}_{\mathcal{E}_L} + \mathcal{M}^{(1)}_{G}\right\|_\infty \leqslant 2\|O\|_1. \tag{17}$$

again by Hölder and $\|\rho\|_\infty < 1$. Essentially the same estimate as for the variance shows that

$$|\mathbb{E}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)] - \mathbb{E}_G[\ell_{\boldsymbol{\theta}}(\rho, O)]| \leqslant \|O\|_1 \left\|\mathcal{A}^{(1)}_{\mathcal{E}_L}\right\|_\infty. \tag{18}$$

Putting it all together into Eq. (14) and using that $\|O\|_1^2 = \left\|O^{\otimes 2}\right\|_1$*, we have that*

$$\begin{aligned} |\mathrm{Var}_{\mathcal{E}_L}[\ell_{\boldsymbol{\theta}}(\rho, O)] - \mathrm{Var}_G[\ell_{\boldsymbol{\theta}}(\rho, O)]| &\leqslant \|O\|_1^2 \left(\left\|\mathcal{A}^{(2)}_{\mathcal{E}_L}\right\|_\infty + 2\left\|\mathcal{A}^{(1)}_{\mathcal{E}_L}\right\|_\infty \right) \\ &= \|O\|_1^2 \left(\left\|\mathcal{A}^{(2)}_{\mathcal{E}_1}\right\|_\infty^L + 2\left\|\mathcal{A}^{(1)}_{\mathcal{E}_1}\right\|_\infty^L \right) \\ &\leqslant 3\|O\|_1^2 \left\|\mathcal{A}^{(2)}_{\mathcal{E}_1}\right\|_\infty^L, \end{aligned} \tag{19}$$

where in the second line we have used Lemma 3, and in the third line we have used Lemma 1 to upper bound $\left\|\mathcal{A}^{(1)}_{\mathcal{E}_1}\right\|_\infty \leqslant \left\|\mathcal{A}^{(2)}_{\mathcal{E}_1}\right\|_\infty$*.* □

Lemma 1. *Let* $\mathcal{E}$ *be an ensemble of unitaries which forms an* ϵ*-approximate* G t*-design, i.e.* $\left\|\mathcal{M}^{(t)}_{\mathcal{E}} - \mathcal{M}^{(t)}_{G}\right\|_\infty < \epsilon$*. Then* $\mathcal{E}$ *forms an* ϵ*-approximate* s*-design for all* $1 \leqslant s \leqslant t$*.*

Proof. We show that $\mathcal{E}$ *forms an* ϵ*-approximate* $t-1$*-design, whence the other cases follow. Let* $\mathcal{M}^{(t)}_{\mathcal{E}}, \mathcal{M}^{(t)}_{G} : \mathfrak{gl}(\mathcal{H}^{\otimes t}) \to \mathfrak{gl}(\mathcal{H}^{\otimes t})$*. We may first note that for any* $\rho \in \mathfrak{gl}(\mathcal{H}^{\otimes(t-1)})$*,*

$$\mathcal{M}^{(t)}_{\mathcal{E}}(\rho \otimes \mathbb{1}_{\mathcal{H}}) = \int_{\mathcal{E}} U^{\otimes(t-1)} \rho (U^\dagger)^{\otimes(t-1)} \otimes UU^\dagger \, d\mu = \left(\int_{\mathcal{E}} U^{\otimes(t-1)} \rho (U^\dagger)^{\otimes(t-1)} \, d\mu \right) \otimes \mathbb{1}_{\mathcal{H}} = \mathcal{M}^{(t-1)}_{\mathcal{E}}(\rho) \otimes \mathbb{1}_{\mathcal{H}}, \tag{20}$$

and likewise for the Haar ensemble G*. Then computing the norm,*

$$\begin{aligned} \left\|\mathcal{M}^{(t-1)}_{\mathcal{E}} - \mathcal{M}^{(t-1)}_{G}\right\|_\infty = \sup_{\substack{\rho \in \mathfrak{gl}(\mathcal{H}^{\otimes(t-1)}) \\ \rho \neq 0}} \frac{\left\|(\mathcal{M}^{(t-1)}_{\mathcal{E}} - \mathcal{M}^{(t-1)}_{G})(\rho)\right\|_\infty}{\|\rho\|_\infty} &= \sup_{\substack{\rho \in \mathfrak{gl}(\mathcal{H}^{\otimes(t-1)}) \\ \rho \neq 0}} \frac{\left\|(\mathcal{M}^{(t)}_{\mathcal{E}} - \mathcal{M}^{(t)}_{G})(\rho \otimes \mathbb{1})\right\|_\infty}{\|\rho \otimes \mathbb{1}\|_\infty} \\ &\leqslant \left\|\mathcal{M}^{(t)}_{\mathcal{E}} - \mathcal{M}^{(t)}_{G}\right\|_\infty \\ &< \epsilon. \end{aligned} \tag{21}$$

□

Lemma 2. *Every eigenvalue* λ *of* $\mathcal{M}^{(t)}_{\mathcal{E}_L}$ *has* $|\lambda| \leqslant 1$*.*

Proof. Since $\mathcal{M}^{(t)}_{\mathcal{E}_L}$ *is self-adjoint, its singular values are equal to the absolute value of its eigenvalues and so every eigenvalue has* $|\lambda| \leqslant \left\|\mathcal{M}^{(t)}_{\mathcal{E}_L}\right\|_\infty$ *by Eq. (11). Let* $A \in \mathfrak{gl}(\mathcal{H}^{\otimes t})$ *have* $\|A\|_\infty = 1$*. Then we have*

$$\left\|\mathcal{M}^{(t)}_{\mathcal{E}_L}(A)\right\|_\infty \leqslant \int_{\mathcal{E}_L} d\mu(U) \left\|U^{\otimes t} A (U^\dagger)^{\otimes t}\right\|_\infty = \int_{\mathcal{E}_L} d\mu(U)\, 1 = 1,$$

where we used that the Schatten norms are unitarily invariant. □

Lemma 3. *Let* $\mathcal{E}_L \subseteq G$ *be the ensemble of unitaries generated by an* L*-layered circuit* $U(\boldsymbol{\theta})$*. Then*

$$\mathcal{A}^{(t)}_{\mathcal{E}_L} = \underbrace{\mathcal{A}^{(t)}_{\mathcal{E}_1} \circ \mathcal{A}^{(t)}_{\mathcal{E}_1} \circ \cdots \circ \mathcal{A}^{(t)}_{\mathcal{E}_1}}_{l \text{ times}}. \tag{22}$$

In particular, $\left\|\mathcal{A}^{(t)}_{\mathcal{E}_L}\right\|_\infty = \left\|\mathcal{A}^{(t)}_{\mathcal{E}_1}\right\|_\infty^L$*.*

Proof. Let us more explicitly write $\mathcal{A}^{(t)}_{\mathcal{E}_L}$:

$$\mathcal{A}^{(t)}_{\mathcal{E}_L}(\cdot) = \mathcal{M}^{(t)}_{\mathcal{E}_L}(\cdot) - \mathcal{M}^{(t)}_{G}(\cdot) = \int_{\mathcal{E}_L} d\mu(U(\boldsymbol{\theta}))\, U(\boldsymbol{\theta})^{\otimes t}(\cdot)(U(\boldsymbol{\theta})^{\dagger})^{\otimes t} - \int_{G} d\mu(U(\boldsymbol{\theta}))\, U(\boldsymbol{\theta})^{\otimes t}(\cdot)(U(\boldsymbol{\theta})^{\dagger})^{\otimes t}\,. \tag{23}$$

Taking the composition of two such linear maps leads to

$$\mathcal{A}^{(t)}_{\mathcal{E}_L} \circ \mathcal{A}^{(t)}_{\mathcal{E}_L}(\cdot) = \mathcal{M}^{(t)}_{\mathcal{E}_L} \circ \mathcal{M}_{\mathcal{E}_L} + \mathcal{M}^{(t)}_{G} \circ \mathcal{M}^{(t)}_{G}(\cdot) - \mathcal{M}^{(t)}_{G} \circ \mathcal{M}^{(t)}_{\mathcal{E}_L}(\cdot) - \mathcal{M}^{(t)}_{\mathcal{E}_L} \circ \mathcal{M}^{(t)}_{G}(\cdot)\,. \tag{24}$$

Then, note that that the following results hold

$$\mathcal{M}^{(t)}_{\mathcal{E}_L} \circ \mathcal{M}^{(t)}_{\mathcal{E}_L}(\cdot) = \mathcal{M}^{(t)}_{\mathcal{E}_{2L}}(\cdot)\,, \qquad \mathcal{M}^{(t)}_{G} \circ \mathcal{M}^{(t)}_{G}(\cdot) = \mathcal{M}^{(t)}_{G} \circ \mathcal{M}^{(t)}_{\mathcal{E}_L}(\cdot) = \mathcal{M}^{(t)}_{\mathcal{E}_L} \circ \mathcal{M}^{(t)}_{G}(\cdot) = \mathcal{M}^{(t)}_{G}\,. \tag{25}$$

The first equality simply states that the multiplication of two distribution from L layered circuits is equal to the distribution of a $2L$ layered circuit, while the latter ones follows from the left- and right-invariance of the Haar measure. Hence, we have

$$\mathcal{A}^{(t)}_{\mathcal{E}_L} \circ \mathcal{A}^{(t)}_{\mathcal{E}_L}(\cdot) = \mathcal{A}^{(t)}_{\mathcal{E}_{2L}}(\cdot)\,. \tag{26}$$

In particular, we obtain

$$\mathcal{A}^{(t)}_{\mathcal{E}_L} = \underbrace{\mathcal{A}^{(t)}_{\mathcal{E}_1} \circ \mathcal{A}^{(t)}_{\mathcal{E}_1} \circ \cdots \circ \mathcal{A}^{(t)}_{\mathcal{E}_1}}_{l \text{ times}}, \tag{27}$$

where $\mathcal{A}^{(t)}_{\mathcal{E}_1}$ corresponds to the single layer ensemble $\mathcal{E}_1$.

The final comment that $\left\|\mathcal{A}^{(t)}_{\mathcal{E}_L}\right\|_{\infty} = \left\|\mathcal{A}^{(t)}_{\mathcal{E}_1}\right\|^{L}_{\infty}$ follows by noting that $\mathcal{A}^{(t)}_{\mathcal{E}_1}$ is self-adjoint, and so its singular values are exactly the absolute values of its eigenvalues. Take the largest singular value $|\lambda|$ and a corresponding eigenvector A of $\mathcal{A}^{(t)}_{\mathcal{E}_1}$. Then by Eq. (27), A is an eigenvector of $\mathcal{A}^{(t)}_{\mathcal{E}_L}$ of largest singular value $|\lambda|^L$. □

To finish, we make two important remarks. First, we note that both Theorems 2 and 3 clearly capture the dependence on the initialization strategy and layer structure. In particular, different gate placements and different parameter initialization strategies lead to different operators $\mathcal{A}^{(2)}_{\mathcal{E}_1}$. To make this point clearer, we have added the following paragraph to the methods section:

Finally, we note that the expressiveness of a single layer will depend on several choices in the ansatz such as how the gates are distributed in a single layer, and how the parameters are sampled therein. We leave the study of the single layer expressiveness changes as a function of those factors for future work.

Second, we remark that our theorems give us the variance for a circuit that forms a 2-design (Theorem 1), or tells us how much we deviate from such variance when the circuit does not form a 2-design (Theorem 3). As such, we are not providing exact formula for the variance for circuits that do not form 2-designs. As stated by the Referee, such calculations would indeed require time horizons and lightcones within circuits. We are currently working on a formalism that will be able to capture this case, but this is still work in progress and beyond the scope of this work. To encourage the community to pursue this line of research, we have added the following sentence to the outlook:

Similarly, we encourage the community to develop tools to compute the exact variance for circuits that do not form approximate 2-designs. In this case, a Lie algebraic treatment such as the one presented here will not be available, meaning that new tools will have to be developed to study loss concentration.

As a general note as well, given the appearance of a work by the many of the same authors "Showcasing a BP theory beyond the dynamical Lie algebra," I wonder if it might be prudent to include a more sizable open questions section in the work. The word unified in the title implies a bit a closing of the theory, while the subsequent paper suggests it's not really wrapped up. It perhaps wouldn't be appropriate to ask the authors to reference the new work since it appeared later, but it may be a decent idea to clearly frame what questions remain unresolved as is relevant to the new work so it is easier to understand how they relate in context and the extent to which the authors believe this work concludes what is known about BPs and what is left unanswered.

We thank the Referee for their comments. We note that our title "unified" shows that the loss function variance terms can be entirely understood in terms of quantities related to the DLA. In that sense, the work "Showcasing a BP theory beyond the dynamical lie algebra" serves as a continuation (not a replacement) for the results and insights presented here. In fact, the present manuscript was foundational to derive the results in the "Showcasing" manuscript. Still, we agree that a mention is needed to clarify what the open questions are. As such, we have modified the outlook, which now reads:

We refer the readers to Ref. [15], where exact loss function variance calculations are presented for a parametrized matchgate circuit, which are valid for arbitrary measurements and initial states. We hope that the results therein will serve as a blueprint for a general theory for generic circuits.

REPLY TO REFEREE 3

> The authors present a method to integrate symmetries into examining and quantifying the variance of loss functions in variational algorithms. Previously, this so-called barren plateau (BP) phenomenon has proven to be a roadblock in the development of quantum variational algorithms. The central results in the paper are: (1) Incorporation of symmetries into barren plateau calculations. (2) Simple mathematical equations that can calculate barren plateaus (though converting an ansatz into this mathematical formulation may be challenging) (3) Discussion of practical implications such as how to incorporate SPAM noise into calculation of barren plateaus. Overall, the paper is nicely written and the results are neat. It is nice that this paper confirms the intuition that the dimension of the dynamical Lie algebra corresponds generally to the severity of the barren plateau. Additionally, the proofs are easy to follow. However, upon a critical read, I have some concerns about publishing especially in Nature Communications. None of the results are really surprising and the "unification" is not as unified as one may hope (more on this later).

We thank the Referee for carefully reviewing our work. We hope that our comments below address their concerns regarding our work, and that they would support publication in Nature Communications.

> My main criticisms upon reflection are:
>
> - The main technical drawback is that the paper makes a crucial assumption that hurts how general their approach is: namely, they assume circuits converge to a two design (see comment below about approximate two design assumption as well). This essentially sidesteps the main question which is "when does a given circuit ansatz actually become an approximate two design?" The aim to provide a holistic approach necessitates a deeper dive into this particular question, and this paper still leaves much blank in answering that question

We thank the reviewer for their comments. We first note that a similar question was asked by Referee 2, and we encourage the Referee to read the reply above. That being said, we would like to highlight the fact that the question "when does a given circuit ansatz actually become an approximate two design?" is fully answered by Theorem 2 in the Methods section. Namely, we provide a general bound that determines the number of layers needed for a circuit to become an approximate 2-design in terms of the one layer expressiveness. We recall here the Theorem 2 for convenience:

Theorem 2. *The ensemble of unitaries $\mathcal{E}_L \subseteq G$ generated by an L-layered circuit $U(\boldsymbol{\theta})$ will form an ϵ-approximate G 2-design, when the number of layers L is*

$$L \geqslant \frac{\log(1/\epsilon)}{\log\left(1/\left\|\mathcal{A}^{(2)}_{\mathcal{E}_1}\right\|\right)}, \tag{28}$$

where $\mathcal{E}_1$ is the ensemble generated by a single layer $U_1(\boldsymbol{\theta}_1)$.

As we can see, the question posed by the referee is fully answered by this theorem. Since this might not have been fully clear in our previous manuscript, we have modified the following paragraph in the Methods section:

As mentioned in the main text, we assume that if the circuit is deep enough, then it will form a 2-design on each of the simple or abelian components of the DLA. Here we provide theoretical justifications for this assumption, and we answer the question: "How many layers are needed for $U(\boldsymbol{\theta})$ to become an approximate 2-design over G?"

We would also like to highlight that, as described above in the reply to Referee 2, we have added a new theorem (Theorem 3 in the revised manuscript) that quantifies by how much the variance of a circuit that does not form a 2-design can deviate from the variance of a circuit that does form a 2-design. We believe that this result significantly enhances our work.

We would like to note that we are actively working on two follow-up works that study how the one layer superoperator $\mathcal{A}^{(2)}_{\mathcal{E}_1}$ depends on the initialization strategy (and hence how the convergence to a 2-design is affected), as well as provide exact variance calculations for circuits that do not form 2-designs. Such works are independent of the present manuscript, and will only strengthen the results found here.

- Separate from technical details, the study of barren plateaus is, in my opinion, a relatively small piece of the quantum machine learning challenge as it is a statement about initialization and says nothing about a host of other untrainability issues that arise with quantum machine learning (and are more interesting in my admittedly biased opinion).

We thank the Referee for their very useful comments. In our opinion, we have not yet fully understood the breath of insights that can be derived from studying barren plateaus. For instance, some of the authors of the present manuscript have recently uploaded to the arxiv a work entitled "Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing" (arXiv:2312.09121 [16]). Therein, it was shown that barren plateaus are intimately connected to the classical simulability of the corresponding quantum learning model, and thus, that understanding how to avoid barren plateaus can also improve our understanding of its classical simulability. More importantly, the work [15] shows that the generalized entanglement that appears in our variance calculation can be directly related to a fermionic entanglement measure (which is a resource for universal quantum computation). This again shows that a more detailed analysis of barren plateaus can unravel deep connections between quantum resources, barren plateaus, and classical simulability. Hence, in this context, we believe that our work has taken a pivotal step whose implications are still not fully fleshed out and understood. Finally, we would also like to note that the FS: concept of DLA is intrinsically related to the presence of local minima in variational quantum computing [17]. At this point, we have still not explored how the generalized entanglement and locality of the initial state and measurement, respectively, affect those other sources of untrainability, but we are certain that they will play a key role. To highlight this point, we have added the following sentence to the outlook:

Recent results have tied the DLA and the presence or absence of barren plateaus to other phenomena like the simulability of the quantum model [15, 16], to the presence of local minima in the optimization landscape [17], and to the reachability of solutions (through the disconnectedness in the geometric manifold of unitaries [10]). We expect that in the near future, the generalized notions of entanglement and locality discussed here will be connected to the resources needed to make a model more or less universal, more or less simulable, and more or less trainable. Such connections will shed additional light on the central role that the DLA plays in characterizing variational quantum computing models.

To wrap up, we would like to highlight that we see barren plateaus as a symptom of a deeper problem in how variational quantum computing models are built. As such, we firmly support the idea that their characterization and deep understanding can only help us to better understand when and where should quantum models be used, or avoided.

- Given its quantum-specific focus particularly on improving results in barren plateaus, one wonders if the paper's findings might be more suited for a specialized venue rather than a broad-spectrum journal. This isn't a critique of the research quality but a suggestion for its optimal placement.

We would like to remind the referee that according to Nature Communication guidelines, the journal's scope is "*Nature Communications is an open access, multidisciplinary journal dedicated to publishing high-quality research in all areas of the biological, physical, chemical and Earth sciences. Papers published by the journal represent important advances of significance to specialists within each field.*" As such, one can see that the journal is well suited for significant advances that are of interest to specialists in a given field (quantum computing in our case). Moreover, we would also like to recall that the original barren plateau paper by Jarrod McClean et al. [1] was published in Nature Communications. Finally, quantum computing is a field with growing interest from a wide variety of disciplines; our work sets important context for any practitioners of these disciplines that are considering the use of variational quantum optimization for their particular problem.

Taken together, we believe that given the significant advancement that our manuscript presents, it is indeed well suited for Nature Communications.

More detailed questions and concerns:

- The paper presents a quantity in eq. (15) quantifying the rate of convergence, which is associated to the largest eigenvalue of some Markov like mixing process. However, this quantity is incredibly hard to calculate in practice. The main question still remains: given a circuit, how fast does it converge to a BP? Eq. 15 and its associated formal statement may help in answering that, but that formula is still incredibly complicated to estimate or calculate — so complicated that it renders the formula potentially useless. See [1] for more relevant techniques and ideas on this front

We thank the Referee for their question. Indeed, we are very much aware that computing the eigenvalues of quantities such as $\mathcal{A}^{(2)}_{\mathcal{E}_1}$ can be a difficult task. As the Referee accurately points out, there are several approaches in the literature that can make theoretical predictions about quantities such as $\mathcal{A}^{(2)}_{\mathcal{E}_1}$ by considering Markov-like processes. In fact, we are currently working on efficient methods to compute quantities such as $\mathcal{A}^{(2)}_{\mathcal{E}_1}$ by following a procedure similar to those employed in Refs. [18–20]. The key insights of our current work is to realize that for circuits with local gates $\mathcal{A}^{(2)}_{\mathcal{E}_1}$ can be efficiently expressed as a matrix product state operator. Similarly, we have found that the eigenvector associated to its largest eigenvalue admits an efficient representation as a matrix product state of bond dimension $\chi = 4$. Such a realization allows us to obtain the largest eigenvalue of $\mathcal{A}^{(2)}_{\mathcal{E}_1}$ for a wide range of circuit ansatzes. This work is still in progress, and goes beyond the scope of our present manuscript. That being said, we do believe that the readers of our work should have an idea of how the norm of $\mathcal{A}^{(2)}_{\mathcal{E}_1}$ can be computed. As such, we have added a new section in the Supplemental Information entitled "*Example of a use-case of Theorem 2*", as well as the following sentence after Theorem 2:

More concretely, if we know what the expressiveness for one layer is, Theorem 2 allows us to compute exactly how many layers are needed for $\left\|\mathcal{A}^{(t)}_{\mathcal{E}_L}\right\|$ *to be smaller than a given* ϵ*. We present in the Supplemental Information an example where we can explicitly construct* $\mathcal{A}^{(2)}_{\mathcal{E}_1}$ *and obtain its largest eigenvalue.*

For convenience, we also present here the new section of the Supplemental Information.

In this section we present an example where we can use Theorem 2 to predict the number of layers needed for an L*-layered circuit to form an* ϵ*-approximate 2-design.*

Let S *be a subgroup of* $\mathbb{U}(d)$*, the unitary group of degree* $d = 2^n$ *acting on* $\mathcal{H}$*, and let* $\{P_j\}_{j=1}^{D}$ *be a basis for the commutant of the* t*-fold tensor representation* $\mathcal{H}^{\otimes t}$ *of* S*. That is,* $[P_j, U^{\otimes t}] = 0$ *for all* $U \in S$*. Then, we know that (see, e.g., [21]):*

$$\mathcal{M}_S^{(t)} = \int_S d\mu(U)\,(U \otimes U^*)^{\otimes t} = \sum_{i,j=1}^{D} W^{-1}_{ij} |P_i\rangle\rangle\langle\langle P_j|\,. \tag{29}$$

Here W^{-1} *is the so-called Weingarten matrix, the inverse of the Gram matrix* W *whose entries* $W_{ij} = \mathrm{Tr}\left[P_i^\dagger P_j\right]$ *are given by the Hilbert–Schmidt inner product of the commutant's basis elements. Moreover, we have denoted as* $|A\rangle\rangle$ *the standard vectorization of a linear operator* A*. Namely, given some* $A \in \mathcal{L}(\mathcal{H})$ *such that* $A = \sum_{\mu,\nu=1}^{2^n} A_{\mu\nu}|\mu\rangle\langle\nu|$*, then* $|A\rangle\rangle = \sum_{\mu,\nu=1}^{2^n} A_{\mu\nu}|\mu\rangle \otimes |\nu\rangle$*. Hence, if we can characterize the basis of the commutant of the* t*-fold tensor representation of* S*, then we can find a closed form expression for* $\mathcal{M}_S^{(t)}$*. Finally, note that* $\mathcal{M}_S^{(t)}$ *is an operator acting on* $2t$ *copies of the Hilbert space* $\mathcal{H}$*, and hence is of dimension* $2^{2nt} \times 2^{2nt}$*.*

Next, let us recall from the main text that we have factorized the parametrized quantum circuit into layers as $U(\boldsymbol{\theta}) = U_L(\boldsymbol{\theta}_L)U_{L-1}(\boldsymbol{\theta}_{L-1})\cdots U_1(\boldsymbol{\theta}_1)$*, and that each layer* $U_l(\boldsymbol{\theta}_l)$ *is given by* $U_l(\boldsymbol{\theta}_l) = \prod_{k=1}^{K} U_k(\theta_{l,k})$*. Hence, if the parameters in*

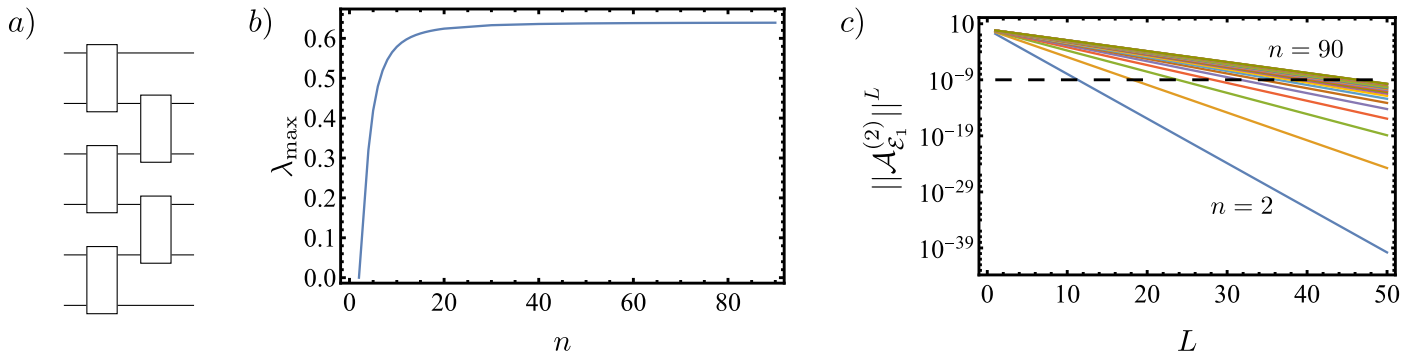

Figure 2. **Layers needed to become an ϵ-approximate 2-design.** (a) Hardware efficient ansatz where two-qubit gates act in a brick-like fashion on neighboring pairs of qubits. We assume that each gate is independently and randomly sampled from $\mathbb{SU}(4)$. (b) Largest eigenvalue of $\mathcal{A}^{(2)}_{\mathcal{E}_1}$ as a function of the number of qubits. (c) Each curve shows $(\mathcal{A}^{(2)}_{\mathcal{E}_1})^L$ for different values of n. The thick horizontal dashed line is plotted at 10^{-9}. Vertical axis is shown in a logarithmic scale.

each U_k are independently sampled, it is not hard to see that the t-th moment operator of a single layer can be expressed as the product of the t-th moment operators of each U_k. That is,

$$\mathcal{M}^{(t)}_{\mathcal{E}_1} = \prod_{k=1}^{K} \mathcal{M}^{(t)}_{\mathcal{E}_{1,k}}\,, \tag{30}$$

where we have denoted as $\mathcal{E}_{1,k}$ the distribution of unitaries obtained from each $U_k(\theta_{l,k})$. As such, if we assume that each $\mathcal{E}_{1,k}$ forms a subgroup, and that we sample it over its respective Haar measure, we will have via Eq. (29) that

$$\mathcal{M}^{(t)}_{\mathcal{E}_1} = \sum_{i_1,j_1=1}^{D_1} \cdots \sum_{i_K,j_K=1}^{D_K} \left((W_1^{-1})_{i_1 j_1} \cdots (W_K^{-1})_{i_K j_K} \langle\langle P_{j_1}|P_{i_2}\rangle\rangle \cdots \langle\langle P_{j_{K-1}}|P_{i_K}\rangle\rangle\right) |P_{i_1}\rangle\rangle\langle\langle P_{j_K}|\,. \tag{31}$$

Here, we denoted as $\{P_{j_k}\}_{j_k=1}^{D_k}$ be a basis for the commutant of the t-fold tensor representation of $\mathcal{E}_{1,k}$, and as W_k^{-1} its associated Weingarten matrix. Moreover, we recall that $\langle\langle A|B\rangle\rangle = \mathrm{Tr}\left[A^\dagger B\right]$. Equation (31) indicates that we can compute the single layer operator $\mathcal{M}^{(t)}_{\mathcal{E}_1}$ by decomposing the layer into unitaries whose ensembles form groups. In its lower form, such procedure can always be taken to the single-gate level where each $U_k(\theta_{l,k}) = e^{-iH_k\theta_{l,k}}$ with $H_k^2 = \mathbb{1}$, and hence where its associated distribution will be a representation of $\mathbb{U}(1)$.

For instance, let us consider a hardware efficient circuit composed of alternating layers of two-qubit gates acting on neighboring qubits (see Fig. 2(a)). Moreover, let us assume that each two-qubit gate forms an independent local 2-design over $\mathbb{SU}(4)$. If we consider a gate acting on qubits i and $i+1$, we will have [21]

$$\mathcal{M}^{(2)}_{\mathbb{SU}(4)} = \frac{1}{15}\left(|\mathbb{I}_i\mathbb{I}_{i+1}\rangle\rangle\langle\langle\mathbb{I}_i\mathbb{I}_{i+1}| + |\mathbb{S}_i\mathbb{S}_{i+1}\rangle\rangle\langle\langle\mathbb{S}_i\mathbb{S}_{i+1}| - \frac{1}{4}\left(\mathbb{S}_i\mathbb{S}_{i+1}| + |\mathbb{S}_i\mathbb{S}_{i+1}\rangle\rangle\langle\langle\mathbb{I}_i\mathbb{I}_{i+1}|\right)\right)\,. \tag{32}$$

where $\mathbb{I}_i$ and $\mathbb{S}_i$ denote the identity and SWAP operators acting on the two copies of the i-th qubit Hilbert spaces. By defining a basis for the operator space spanned from the non-orthogonal elements

$$\left\{|\mathbb{I}_i\mathbb{I}_{i+1}\rangle\rangle, |\mathbb{I}_i\mathbb{S}_{i+1}\rangle\rangle, |\mathbb{S}_i\mathbb{I}_{i+1}\rangle\rangle, |\mathbb{S}_i\mathbb{S}_{i+1}\rangle\rangle\right\}, \tag{33}$$

we can express $\mathcal{M}^{(2)}_{\mathbb{SU}(4)}$ as a 4×4 matrix

$$\mathcal{M}^{(2)}_{\mathbb{SU}(4)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2/5 & 0 & 0 & 2/5 \\ 2/5 & 0 & 0 & 2/5 \\ 2/5 & 0 & 0 & 2/5 \\ 0 & 0 & 0 & 1 \end{pmatrix}\,. \tag{34}$$

One can readily verify that $\mathcal{M}^{(2)}_{\mathbb{SU}(4)} \circ \mathcal{M}^{(2)}_{\mathbb{SU}(4)} = \mathcal{M}^{(2)}_{\mathbb{SU}(4)}$ *and that* $\mathrm{Tr}\left[\mathcal{M}^{(2)}_{\mathbb{SU}(4)}\right] = 2$*, as expected from the fact that it is a projector onto a subspace of dimension* 2*. Assuming for simplicity that* n *is even (the calculations below can be readily generalized to the case when* n *is odd), we find that the single layer moment operator is*

$$\mathcal{M}^{(t)}_{\mathcal{E}_1} = \left(\mathbb{1} \otimes \left(\mathcal{M}^{(2)}_{\mathbb{SU}(4)}\right)^{\otimes(n/2-1)} \otimes \mathbb{1}\right) \circ \left(\mathcal{M}^{(2)}_{\mathbb{SU}(4)}\right)^{\otimes(n/2)} , \tag{35}$$

where $\mathbb{1}$ *denotes the* 2×2 *identity matrix. Note that the previous has allowed us to effectively express* $\mathcal{M}^{(t)}_{\mathcal{E}_1}$ *as a* $2^n \times 2^n$ *matrix, which is much smaller than its vanilla form of* $16^n \times 16^n$*. Moreover, this approach follows very closely manuscripts in the literature where Haar average properties are computed by mapping objects such as* $\mathcal{M}^{(t)}_{\mathcal{E}_1}$ *to Markov-like mixing process matrices [18–20].*

To compute the norm of the operator $\mathcal{A}^{(2)}_{\mathcal{E}_1} = \mathcal{M}^{(2)}_{\mathcal{E}_1} - \mathcal{M}^{(2)}_{G}$ *we can use the fact that deep hardware efficient ansatz are universal [5], meaning that* $\mathfrak{g} = \mathfrak{su}(d)$*, and hence* $G = \mathbb{SU}(d)$*. From here, we can express*

$$\mathcal{M}^{(2)}_{\mathbb{SU}(d)} = \frac{1}{4^n - 1}\left(|\mathbb{I}\rangle\rangle\langle\langle\mathbb{I}| + |\mathbb{S}\rangle\rangle\langle\langle\mathbb{S}| - \frac{1}{2^n}\left(|\mathbb{I}\rangle\rangle\langle\langle\mathbb{S}| + |\mathbb{S}\rangle\rangle\langle\langle\mathbb{I}|\right)\right) , \tag{36}$$

where $|\mathbb{I}\rangle\rangle = \bigotimes_{i=1}^{n} |\mathbb{I}_i\rangle\rangle$ *and* $|\mathbb{S}\rangle\rangle = \bigotimes_{i=1}^{n} |\mathbb{S}_i\rangle\rangle$*. Following a similar procedure such as that used to build an efficient representation of* $\mathcal{M}^{(t)}_{\mathcal{E}_1}$ *in Eq. (35), we can express* $\mathcal{M}^{(2)}_{\mathbb{SU}(d)}$ *and thus* $\mathcal{A}^{(2)}_{\mathcal{E}_1}$ *as a* $2^n \times 2^n$ *matrix. Combining this with the fact that* $\mathcal{A}^{(2)}_{\mathcal{E}_1}$ *can be represented efficiently as a matrix product operator, and the fact that the eigenvector associated to its largest eigenvalue admits an efficient representation as a bond dimension 4 matrix product state, we can easily obtain the largest eigenvalue for* $\mathcal{A}^{(2)}_{\mathcal{E}_1}$*.*

For instance, in Fig. 2(b) we plot the largest eigenvalue $\lambda_{\max}$ *of* $\mathcal{A}^{(2)}_{\mathcal{E}_1}$ *as a function of the number of qubits* n *that* $U(\boldsymbol{\theta})$ *acts on. We have considered system sizes* $n = 2, \ldots, 90$*. Here we can see that for* $n = 2$*,* $\lambda_{\max} = 0$*, as expected. Moreover, as* n *increases, the value of* $\lambda_{\max}$ *appears to saturate at a value* ≈ 0.639*. Then, in Fig. 2(c) we plot* $\left\|\mathcal{A}^{(2)}_{\mathcal{E}_L}\right\| = \left(\left\|\mathcal{A}^{(2)}_{\mathcal{E}_1}\right\|\right)^L$ *for different values of* L*. Here, we find that as the number of layers increases, the norm of the expressiveness superoperator* $\left\|\mathcal{A}^{(2)}_{\mathcal{E}_L}\right\|$ *decreases exponentially. Indeed, we can now answer questions such as: "How many layers are needed for the circuit to become an* $\epsilon = 10^{-9}$ *approximate design?" For the number of qubits considered, we show that* $L = 47$ *will suffice.*

To finish, we note that here we have considered a circuit whose local gates are randomly taken from $\mathbb{SU}(4)$*. However, the methods presented can be extended and generalized to other circuit architectures where the gates are sampled from any subgroup. In particular, we can always assume that each gate in the layer is generated by a Pauli and hence samples from a representation of* $\mathbb{U}(1)$*. We leave the exploration of such cases for future work.*

- Barren plateaus expressed in terms of dynamical Lie algebras give a nice mathematical picture. However, similar to the above point, these dynamical Lie algebras can be elusive mathematical objects. I expect in many cases, we may not know what the dimensions of them are let alone how a given circuit ansatz spans this algebra. The work of [4] seems to indicate understanding this problem may not be so easy and rife with other issues as well.

We note that Referee 2 had a similar concern. As we replied above, we believe that the study of DLAs is still in its infancy. Already, recent results have made tremendous progress in characterizing entire families of DLAs (see for instance Ref. [7]; in a similar spirit we are currently preparing a manuscript where we classify the DLAs for the quantum alternating operator ansatz for Max-Cut problems). Still, more work needs to be done for determining simpler ways of computing the DLA. Indeed, while the seminal work of [10] indicates that using local gates will lead to additional elements of the commutant that will make the DLA more complex, this does not preclude their computation. We strongly believe that in the near future our understanding of the DLA (based on results such as that in [10]) will only improve and hence make our work more relevant.

- How relevant are these results in the setting of approximate symmetry or equivariance which is the setting one typically finds in practice? E.g. for classical deep learning, convolutional networks are not truly symmetric but approximately symmetric to translations.

We again thank the Referee for their comments. We would like to emphasize that results presented in this work do not require the circuit to respect a specific symmetry. In fact, they are equally valid for universal quantum circuits. Perhaps the Referee is assuming that our results are valid only for equivariant circuits such as those used in the geometric quantum machine learning literature [22–24]. However, such requirements are not necessary for our results to hold. In particular, we note that any circuit, whether is respects some symmetry, or only approximately respects it, will have a well-defined DLA, thus amendable to our theorems and results. We are sorry if our manuscript lead to this confusion. We have carefully reviewed our work and could not find any specific place where such confusion could arise.

- Beyond BPs, variational algorithms face a number of untrainability issues that are of practical relevance. E.g., as mentioned before for geometric learning, see [4] which studies disconnectedness in the geometric manifold of unitaries. These untrainability issues are in some ways related to BPs but also need to be mentioned for a more holistic picture.

We thank the reviewer for their comment. We agree that barren plateaus are by no means, the end of the story, and understanding how certain properties of the DLA can be related to other untrainability issues is an important line of research. To clarify this point, we have added the following sentence in the outlook:

Recent results have tied the DLA and the presence or absence of barren plateaus to other phenomena like the simulability of the quantum model [15, 16], to the presence of local minima in the optimization landscape [17], and to the reachability of solutions (through the disconnectedness in the geometric manifold of unitaries [10]). We expect that in the near future, the generalized notions of entanglement and locality discussed here will be connected to the resources needed to make a model more or less universal, more or less simulable, and more or less trainable. Such connections will shed additional light on the central role that the DLA plays in characterizing variational quantum computing models.

Proofs were briefly checked and no errors were found. The proofs were rather nice and relatively easy to follow for someone with a reasonable but non-expert background in representation theory.

We thank the Referee for checking our proofs and for their positive feedback.

More minor comments:

1. A description (or at least a reference) is needed for the statement where the authors say "In the following, we assume that the parametrized circuit is deep enough so that it forms an approximate 2-design". After all, readers may not know what this is.

2. Related to the above, I would also prefer that the authors state specifically in theorems that the distribution forms a 2-design rather than assuming it in a sentence early on and subsequently leaving it out of formal statements.

3. Can authors give a citation for Lemma 3 where they say the fact is well known? I understand a proof comes after, but it would be good to still cite where this is from.

4. I do not agree with the language surrounding a "unified theory" of barren plateaus. This paper really only generalizes the phenomenon to circuits constrained under symmetry and where we know enough about its Lie algebraic properties to perform calculations as mentioned before. Furthermore, many other papers make similar claims of unification (e.g. [1-3]), some which the authors cite, so I would also be careful in using this language.

References: [1] Napp, John. "Quantifying the barren plateau phenomenon for a model of unstructured variational ansätze." arXiv preprint arXiv:2203.06174 (2022). [2] Zhao, Chen, and Xiao-Shan Gao. "Analyzing the barren plateau phenomenon in training quantum neural networks with the ZX-calculus." Quantum 5 (2021): 466. [3] Diaz, N. L., et al. "Showcasing a Barren Plateau Theory Beyond the Dynamical Lie Algebra." arXiv preprint arXiv:2310.11505 (2023). [4] Marvian, Iman. "Restrictions on realizable unitary operations imposed by symmetry and locality." Nature Physics 18.3 (2022): 283-289.

1. We have added a footnote describing what a 2-design is. We have also added more information Section IIB, which both say more about the role of theorem 2 and point to where we define what a design is.

2. We thank the referee for this comment, as we agree with it. We have stated that Theorem 1 holds for a 2-design, while our new Theorem 3 quantifies how much the variances would change for a circuit that forms an approximate, rather than an exact, design.

3. A citation to a representation theory textbook has been added.

4. We respectfully disagree with the author. First we note that Refs. [1–2] is only valid for Hardware efficient ansatzes with locally $SU(2)$ random gates, whereas the results in Ref [3] hold only for matchgate circuits. Then, Ref [4] makes a statement about how large a given DLA can be when using local gates that respect a given symmetry, but makes no statement about loss function variances. As replied to Referee 2, we note that the "unification" terminology adopted in this work has two meaning. On one hand it showcases that our theory can be applied to any circuit, thus encapsulating architecture-specific results. On the other hand, the term highlights that all the contributions to the variance are *entirely* captured in terms of quantities relating to the DLA of the underlying circuit. As such, we deeply believe that such language is justified and would refrain to change it.

[1] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nature Communications **9**, 1 (2018).

[2] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nature Communications **12**, 1 (2021).

[3] L. Schatzki, M. Larocca, F. Sauvage, and M. Cerezo, Theoretical guarantees for permutation-equivariant quantum neural networks, arXiv preprint arXiv:2210.09974 (2022).

[4] K. Sharma, M. Cerezo, L. Cincio, and P. J. Coles, Trainability of dissipative perceptron-based quantum neural networks, Physical Review Letters **128**, 180505 (2022).

[5] M. Larocca, P. Czarnik, K. Sharma, G. Muraleedharan, P. J. Coles, and M. Cerezo, Diagnosing Barren Plateaus with Tools from Quantum Optimal Control, Quantum **6**, 824 (2022).

[6] S. Thanasilp, S. Wang, N. A. Nghiem, P. J. Coles, and M. Cerezo, Subtleties in the trainability of quantum machine learning models, arXiv preprint arXiv:2110.14753 (2021).

[7] R. Wiersema, E. Kökcü, A. F. Kemper, and B. N. Bakalov, Classification of dynamical lie algebras for translation-invariant 2-local spin systems in one dimension, arXiv preprint arXiv:2309.05690 (2023).

[8] S. Hadfield, Z. Wang, B. O'Gorman, E. G. Rieffel, D. Venturelli, and R. Biswas, From the quantum approximate optimization algorithm to a quantum alternating operator ansatz, Algorithms **12**, 34 (2019).

[9] Z. Zimborás, R. Zeier, T. Schulte-Herbrüggen, and D. Burgarth, Symmetry criteria for quantum simulability of effective interactions, Physical Review A **92**, 042309 (2015).

[10] I. Marvian, Restrictions on realizable unitary operations imposed by symmetry and locality, Nature Physics **18**, 283 (2022).

[11] I. Marvian, (non-)universality in symmetric quantum circuits: Why abelian symmetries are special, arXiv preprint arXiv:2302.12466 (2023).

[12] S. Kazi, M. Larocca, and M. Cerezo, On the universality of s_n-equivariant k-body gates, arXiv preprint arXiv:2303.00728 (2023).

[13] W. Bosma, J. Cannon, and C. Playoust, The Magma algebra system. I. The user language, J. Symbolic Comput. **24**, 235 (1997), computational algebra and number theory (London, 1993).

[14] E. Kökcü, T. Steckmann, Y. Wang, J. Freericks, E. F. Dumitrescu, and A. F. Kemper, Fixed depth hamiltonian simulation via cartan decomposition, Physical Review Letters **129**, 070501 (2022).

[15] N. L. Diaz, D. García-Martín, S. Kazi, M. Larocca, and M. Cerezo, Showcasing a barren plateau theory beyond the dynamical lie algebra, arXiv preprint arXiv:2310.11505 (2023).

[16] M. Cerezo, M. Larocca, D. García-Martín, N. L. Diaz, P. Braccia, E. Fontana, M. S. Rudolph, P. Bermejo, A. Ijaz, S. Thanasilp, *et al.*, Does provable absence of barren plateaus imply classical simulability? or, why we need to rethink variational quantum computing, arXiv preprint arXiv:2312.09121 https://doi.org/10.48550/arXiv.2312.09121 (2023).

[17] M. Larocca, N. Ju, D. García-Martín, P. J. Coles, and M. Cerezo, Theory of overparametrization in quantum neural networks, Nature Computational Science **3**, 542 (2023).

[18] A. Harrow and S. Mehraban, Approximate unitary t-designs by short random quantum circuits using nearest-neighbor and long-range gates, arXiv preprint arXiv:1809.06957 (2018).

[19] A. M. Dalzell, N. Hunter-Jones, and F. G. S. L. Brandão, Random quantum circuits anticoncentrate in log depth, PRX Quantum **3**, 010333 (2022).

[20] J. Napp, Quantifying the barren plateau phenomenon for a model of unstructured variational ansätze, arXiv preprint arXiv:2203.06174 (2022).

[21] A. A. Mele, Introduction to haar measure tools in quantum information: A beginner's tutorial, arXiv preprint arXiv:2307.08956 (2023).

[22] M. Larocca, F. Sauvage, F. M. Sbahi, G. Verdon, P. J. Coles, and M. Cerezo, Group-invariant quantum machine learning, PRX Quantum **3**, 030341 (2022).

[23] F. Sauvage, M. Larocca, P. J. Coles, and M. Cerezo, Building spatial symmetries into parameterized quantum circuits for faster training, arXiv preprint arXiv:2207.14413 https://doi.org/10.48550/arXiv.2207.14413 (2022).

[24] Q. T. Nguyen, L. Schatzki, P. Braccia, M. Ragone, M. Larocca, F. Sauvage, P. J. Coles, and M. Cerezo, A theory for equivariant quantum neural networks, arXiv preprint arXiv:2210.08566 (2022).

REVIEWER COMMENTS

Reviewer #2 (Remarks to the Author):

I want to thank the authors for their detailed and careful response to all the points that I raised. In summary, I think the additional theorem and analysis related to circuits at finite depth strengthen the paper to the point where I recommend acceptance.

As an afterthought -

I don't fully agree with the spirit of the points about how many non-trivial DLA's there are based on symmetries, as I think from a counting point of view "most" circuits would have none. That said, I do imagine there could be some value in the view of generalizing separable ansatz beyond the trivial case of just dividing up the qubits, and training within that space before destroying the symmetries to restore universality. Perhaps that is an interesting road towards more training and expressive ansatz at finite depths.

Reviewer #3 (Remarks to the Author):

I should qualify that in this round of reviews I did not look in detail through proofs again and only skimmed parts of the updated paper in the appendix. I did however read the authors' response in detail and thank the authors for their response. The response was very detailed and had some useful examples and new additions.

Before jumping into any detailed feedback, I would like to state that my opinion of the work remains largely unchanged. On the positive side, this paper has some nice extensions of barren plateau results to the symmetric setting along with examples and well-written proofs. It will be a useful reference for people in the quantum variational community. On the negative side, the results here are largely unsurprising and potentially challenging to use in applications (e.g. quantities in the main theorems can be hard to calculate for example).

Whether this paper meets the criteria of Nature Communications is up to the editor. I would be happy to support the editors' decision either way.

Points of further feedback and commentary:

- I thank the authors for adding the further details and examples about approximate convergence to 2-design.
- Regarding the question I asked earlier "How relevant are these results in the setting of approximate symmetry or equivariance which is the setting one typically finds in practice?", the authors responded believing this to be a criticism and arguing this fits within their framework already. Instead, I wanted simply to confirm my understand and urge the authors to discuss this further in their manuscript as I do not agree with their position. This is not a very major point, but I do believe this warrants some discussion about next steps and limitations. In practice, many symmetries are not so nice to fit into the setting here. For example, symmetries may sometimes only approximately hold within a problem, or other times, symmetries may not even be instances of Lie groups/algebras.
- Related to the above, the authors respond " In particular, we note that any circuit, whether is [sic] respects some symmetry, or only approximately respects it, will have a well-defined DLA.": Having a well defined DLA is great, but is it informative? What happens if I give you a symmetry which is not a group? E.g. some symmetry that is not invertible or some noisy symmetry that can only be approximate (e.g. a semi-group)? These show up all the time in classical machine learning (e.g. small pixel noise in images, translation edge effects, etc.) and even in physics (e.g. the answer to some problem may remain the same if only a few qubits are corrupted). I do not mean to solely criticize the authors' paper on this point, but simply ask the authors to consider this often more realistic setting and comment on it.
- I think the authors misunderstood my criticism regarding the approximate 2-design. I did not mean to imply that the paper did not have any formal theorems about when approximate 2-designs occur. Rather, given an architecture, how fast does that specific architecture approach a 2-design? I understand the authors have a description in terms of a variant of an eigenvalue gap reminiscent of similar convergence results in Markov chain algorithms. However, this quantity can be incredibly hard to calculate. In fact, the example in the appendix of the hardware efficient ansatz is nontrivial to estimate and appears to be one without any incorporation of symmetry. I do not agree with the authors that a better understanding of this "goes beyond the scope of our present manuscript". Instead, the authors should be honest about the limitations of the theorem in this regard.

- I am still against the use of the word “unification” to describe the results in this paper. As the authors state themselves in their response, the study of DLA is still in its infancy and there remains many open questions for barren plateaus. I also note reviewer 2 had similar concerns.

A Unified Theory of Barren Plateaus for Deep Parametrized Quantum Circuits

LETTER TO THE REFEREES

Dear Referees,

We would like to thank you again for your time invested in reviewing our work.

We would like to highlight that Referee 2 "*recommend[s] acceptance*" and that Referee 3 says "*this paper has some nice extensions of barren plateau results to the symmetric setting along with examples and well-written proofs. It will be a useful reference for people in the quantum variational community.*"

Below you can find our replies to the latest Referee comments. We believe that our work has significantly improved since its first version, and we hope that it now meets the criteria for publication. In fact, we have decided to add the following remark to our acknowledgments:

We also thank the anonymous referees for their thorough review of the manuscript and insightful comments, which motivated the addition of new results in the revised version.

Looking forward to your answer.

Michael Ragone, Bojko N. Bakalov, Frédéric Sauvage, Alexander F. Kemper, Carlos Ortiz Marrero, Martín Larocca, and M. Cerezo

REPLY TO REFEREE 2

I want to thank the authors for their detailed and careful response to all the points that I raised. In summary, I think the additional theorem and analysis related to circuits at finite depth strengthen the paper to the point where I recommend acceptance.

Response: We thank the Referee for their valuable time and for their positive feedback.

I don't fully agree with the spirit of the points about how many non-trivial DLA's there are based on symmetries, as I think from a counting point of view "most" circuits would have none. That said, I do imagine there could be some value in the view of generalizing separable ansatz beyond the trivial case of just dividing up the qubits, and training within that space before destroying the symmetries to restore universality. Perhaps that is an interesting road towards more training and expressive ansatz at finite depths.

Response: We thank the Referee for their comment. While it is true that "most" quantum circuits would have no symmetries, our argument is based on circuits built to tackle specific problems. Here, it is expected that these should contain symmetries arising from the problem at hand. Indeed, we also agree that the study of symmetry preserving and symmetry broken circuits is an interesting research direction that is beyond our current work.

REPLY TO REFEREE 3

I should qualify that in this round of reviews I did not look in detail through proofs again and only skimmed parts of the updated paper in the appendix. I did however read the authors' response in detail and thank the authors for their response. The response was very detailed and had some useful examples and new additions. Before jumping into any detailed feedback, I would like to state that my opinion of the work remains largely unchanged. On the positive side, this paper has some nice extensions of barren plateau results to the symmetric setting along with examples and well-written proofs. It will be a useful reference for people in the quantum variational community.

Response: We would like to thank the Referee for their positive feedback and for the time spent reviewing our work. Indeed, we believe that our results will be useful for the variational quantum computing community.

On the negative side, the results here are largely unsurprising and potentially challenging to use in applications (e.g. quantities in the main theorems can be hard to calculate for example).

Response: We would like to note here that, in our opinion, there are some very non-trivial results revealed from our work. For instance, we can highlight the intrinsic connection between barren plateaus and generalized notions of entanglement and locality. This connection was not previously explored nor noted in the community, and it appears to have deep implications regarding the connection between the resources that make something quantum "quantum" and how easy or hard it will be to train a variational quantum model. Similarly, we note that our results proved that there are many misconceptions in the literature (such as global measurement operators will always lead to barren plateaus) and that these can be fully understood under our framework — a task that was previously impossible. In addition, we would like to highlight that while the computation of certain quantities required in our main theorems can be hard, there is an ever-increasing body of literature on this topic. For instance, several recent papers have computed the dynamical Lie algebra of quantum circuits and the eigenvalues of moment operators. As such, we hope that our general results will become more applicable as more tools are developed, and we encourage the development of such methods.

Finally, we wish to address the 'surprisingness' of our results. It may well be true that to some, the fact that barren plateaus arise when the DLA is large is unsurprising. Indeed, this was conjectured in previous work. However, this view is not pervasive, a fact that can (again) be gleaned from the large body of literature on the topic, and the attempts to circumvent barren plateaus. We have provided a constructive proof of the relationship, which in principle shuts some particular avenues for avoiding BPs, and thus has a wide-ranging impact.

Points of further feedback and commentary: - I thank the authors for adding the further details and examples about approximate convergence to 2-design.

We thank the Referee for pointing us in this direction as we believe that our work is much stronger now.

- Regarding the question I asked earlier "How relevant are these results in the setting of approximate symmetry or equivariance which is the setting one typically finds in practice?", the authors responded believing this to be a criticism and arguing this fits within their framework already. Instead, I wanted simply to confirm my understand and urge the authors to discuss this further in their manuscript as I do not agree with their position. This is not a very major point, but I do believe this warrants some discussion about next steps and limitations. In practice, many symmetries are not so nice to fit into the setting here. For example, symmetries may sometimes only approximately hold within a problem, or other times, symmetries may not even be instances of Lie groups/algebras.
- Related to the above, the authors respond "In particular, we note that any circuit, whether is [sic] respects some symmetry, or only approximately respects it, will have a well-defined DLA.": Having a well defined DLA is great, but is it informative? What happens if I give you a symmetry which is not a group? E.g. some symmetry that is not invertible or some noisy symmetry that can only be approximate (e.g. a semi-group)? These show up all the time in classical machine learning (e.g. small pixel noise in images, translation edge effects, etc.) and even in physics (e.g. the answer to some problem may remain the same if only a few qubits are corrupted). I do not mean to solely criticize the authors' paper on this point, but simply ask the authors to consider this often more realistic setting and comment on it.

We thank the Referee for the clarification on this point. As we claimed in our previous response, if we take a parametrized quantum circuit as in Eq. (1), we can always define a dynamical Lie algebra as in Eq. (5). However, as the referee claims, this is of course an assumption and certain scenarios, including the case when some form of noise acts on the circuit, are not covered here. To address this issue, we have added the following sentence to the Outlook:

Moreover, one can also envision considering more realistic noise settings where noise channels are interleaved with the unitaries. Clearly, since a noisy parametrized quantum circuit no longer forms a group, our Lie algebraic formalism no longer applies.

- I think the authors misunderstood my criticism regarding the approximate 2-design. I did not mean to imply that the paper did not have any formal theorems about when approximate 2-designs occur. Rather, given an architecture, how fast does that specific architecture approach a 2-design? I understand the authors have a description in terms of a variant of an eigenvalue gap reminiscent of similar convergence results in Markov chain algorithms. However, this quantity can be incredibly hard to calculate. In fact, the example in the appendix of the hardware efficient ansatz is nontrivial to estimate and appears to be one without any incorporation of symmetry. I do not agree with the authors that a better understanding of this "goes beyond the scope of our present manuscript." Instead, the authors should be honest about the limitations of the theorem in this regard.

We agree that these cases are not covered by our theorems. This limitation is now acknowledged in our Outlook, where we have added the following sentence highlighting methods that can be used instead of the techniques leveraged in this work:

Similarly, we encourage the community to develop tools to compute the exact variance for circuits that do not form approximate 2-designs, as these cases are not covered by our main theorems. In this case, a Lie algebraic treatment such as the one presented here will not be available. Here, one will have to instead make additional assumptions such as local gates being sampled from a local group, thus mapping the variance evaluation to a Markov chain-like process which can be analyzed analytically [1–11], numerically via Monte Carlo [12] and tensor networks [13, 14], or studied via XZ-calculus [15].

Indeed, we are fully aware of the limitations of our theorems, which is one of the reasons why we are actively trying to develop tools that can be used in more general scenarios (see, e.g., our recent work [13]).

- I am still against the use of the word “unification” to describe the results in this paper. As the authors state themselves in their response, the study of DLA is still in its infancy and there remains many open questions for barren plateaus. I also note reviewer 2 had similar concerns.

We thank the Referee for their comment. We will defer here to the Editor’s opinion.

[1] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nature Communications **12**, 1 (2021).

[2] A. Uvarov and J. D. Biamonte, On barren plateaus and cost function locality in variational quantum algorithms, Journal of Physics A: Mathematical and Theoretical **54**, 245301 (2021).

[3] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, Absence of barren plateaus in quantum convolutional neural networks, Physical Review X **11**, 041011 (2021).

[4] V. Heyraud, Z. Li, K. Donatella, A. L. Boité, and C. Ciuti, Efficient estimation of trainability for variational quantum circuits, arXiv preprint arXiv:2302.04649 (2023).

[5] Z. Liu, L.-W. Yu, L.-M. Duan, and D.-L. Deng, The presence and absence of barren plateaus in tensor-network based machine learning, Physical Review Letters **129**, 270501 (2022).

[6] T. Barthel and Q. Miao, Absence of barren plateaus and scaling of gradients in the energy optimization of isometric tensor network states, arXiv preprint arXiv:2304.00161 (2023).

[7] Q. Miao and T. Barthel, Isometric tensor network optimization for extensive hamiltonians is free of barren plateaus, arXiv preprint arXiv:2304.14320 (2023).

[8] R. J. Garcia, C. Zhao, K. Bu, and A. Jaffe, Barren plateaus from learning scramblers with local cost functions, Journal of High Energy Physics **2023**, 1 (2023).

[9] A. Letcher, S. Woerner, and C. Zoufal, From tight gradient bounds for parameterized quantum circuits to the absence of barren plateaus in qgans, arXiv preprint arXiv:2309.12681 (2023).

[10] J. M. Arrazola, O. Di Matteo, N. Quesada, S. Jahangiri, A. Delgado, and N. Killoran, Universal quantum circuits for quantum chemistry, Quantum **6**, 742 (2022).

[11] H.-K. Zhang, S. Liu, and S.-X. Zhang, Absence of barren plateaus in finite local-depth circuits with long-range entanglement, arXiv preprint arXiv:2311.01393 (2023).

[12] J. Napp, Quantifying the barren plateau phenomenon for a model of unstructured variational ansätze, arXiv preprint arXiv:2203.06174 (2022).

[13] P. Braccia, P. Bermejo, L. Cincio, and M. Cerezo, Computing exact moments of local random quantum circuits via tensor networks, arXiv preprint arXiv:2403.01706 (2024).

[14] H.-Y. Hu, A. Gu, S. Majumder, H. Ren, Y. Zhang, D. S. Wang, Y.-Z. You, Z. Minev, S. F. Yelin, and A. Seif, Demonstration of robust and efficient quantum property learning with shallow shadows, arXiv preprint arXiv:2402.17911 (2024).

[15] C. Zhao and X.-S. Gao, Analyzing the barren plateau phenomenon in training quantum neural networks with the ZX-calculus, Quantum **5**, 466 (2021).