

# Does provable absence of barren plateaus imply classical simulability?

Received: 19 March 2024

Accepted: 8 August 2025

Published online: 25 August 2025

 Check for updates

M. Cerezo<sup>1,2</sup>✉, Martin Larocca<sup>3,4</sup>, Diego García-Martin<sup>1</sup>, N. L. Diaz<sup>1,5</sup>, Paolo Braccia<sup>3</sup>, Enrico Fontana<sup>6</sup>, Manuel S. Rudolph<sup>7</sup>, Pablo Bermejo<sup>1,8</sup>, Aroosa Ijaz<sup>3,9,10</sup>, Supanut Thanasilp<sup>7,11</sup>, Eric R. Anschuetz<sup>12,13</sup> & Zoë Holmes<sup>7</sup>

A large amount of effort has recently been put into understanding the barren plateau phenomenon. In this perspective article, we face the increasingly loud elephant in the room and ask a question that has been hinted at by many but not explicitly addressed: *Can the structure that allows one to avoid barren plateaus also be leveraged to efficiently simulate the loss classically?* We collect evidence—on a case-by-case basis—that many commonly used models whose loss landscapes avoid barren plateaus can also admit classical simulation, provided that one can collect some classical data from quantum devices during an initial data acquisition phase. This follows from the observation that barren plateaus result from a curse of dimensionality, and that current approaches for solving them end up encoding the problem into some small, classically simulable, subspaces. Thus, while stressing that quantum computers can be essential for collecting data, our analysis sheds doubt on the information processing capabilities of many parametrized quantum circuits with provably barren plateau-free landscapes. We end by discussing the (many) caveats in our arguments including the limitations of average case arguments, the role of smart initializations, models that fall outside our assumptions, the potential for provably superpolynomial advantages and the possibility that, once larger devices become available, parametrized quantum circuits could heuristically outperform our analytic expectations.

In recent years, the initial excitement attracted by variational quantum algorithms<sup>1–3</sup> and quantum machine learning<sup>4–9</sup> has been tempered by the barren plateau phenomenon<sup>10–59</sup>. Namely, there is a growing awareness that a large class of quantum learning architectures exhibit loss function landscapes that concentrate exponentially in system size towards their mean value. On such landscapes, exponential resources are required for training,

prohibiting the successful scaling of variational quantum algorithms. Hence, identifying architectures and training strategies that provably do not lead to barren plateaus has become a highly active area of research. Examples of such strategies include shallow circuits with local measurements<sup>15,24–28</sup>, dynamics with small Lie algebras<sup>22,50–54,58,60</sup>, identity initializations<sup>39,55,61</sup>, embedding symmetries into the circuit's architecture<sup>62–69</sup>, adding non-unital noise or intermediate

<sup>1</sup>Information Sciences, Los Alamos National Laboratory, Los Alamos, NM, USA. <sup>2</sup>Quantum Science Center, Oak Ridge, TN, USA. <sup>3</sup>Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA. <sup>4</sup>Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM, USA. <sup>5</sup>Departamento de Física-IFLP/CONICET, Universidad Nacional de La Plata, La Plata, Argentina. <sup>6</sup>University of Strathclyde, Glasgow, UK. <sup>7</sup>Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. <sup>8</sup>Donostia International Physics Center, San Sebastián, Spain. <sup>9</sup>Department of Physics and Astronomy, University of Waterloo, Waterloo, ON, Canada. <sup>10</sup>Vector Institute, MaRS Centre, Toronto, ON, Canada. <sup>11</sup>Chula Intelligent and Complex Systems, Department of Physics, Faculty of Science, Chulalongkorn University, Bangkok, Thailand. <sup>12</sup>Institute for Quantum Information and Matter Caltech, Pasadena, CA, USA. <sup>13</sup>Walter Burke Institute for Theoretical Physics Caltech, Pasadena, CA, USA. ✉e-mail: [cerezo@lanl.gov](mailto:cerezo@lanl.gov)

measurements<sup>56,70–73</sup>, and certain classes of quantum generative models<sup>29,32,33</sup>.

However, these strategies all, in some sense, make use of some simple underlying structure of the problem. This provokes the question: *Could the very same structure that allows one to provably avoid barren plateaus be leveraged to efficiently simulate the loss function classically?* Here we argue that the answer to this question is “Yes and No”. Specifically, we argue and present strong evidence that a wide class of loss landscapes which provably do not exhibit barren plateaus can be simulated using either a classical algorithm or what we call a “quantum-enhanced” classical algorithm that runs in polynomial time. In the latter case, this simulation still necessitates the use of a quantum computer during an initial data acquisition phase<sup>74–77</sup>, but it does not require hybrid quantum-classical optimization loops. These arguments can be understood as a soft form of dequantization of the information processing capabilities of many variational quantum circuits in barren plateau-free landscapes.

Core to our argument is the observation that any loss based on evolving an initial state  $\rho$  through a parameterized quantum circuit  $U(\theta)$  and then estimating the expectation value of a non-trivial observable  $O$  can be written as an inner product between the Heisenberg evolved observable  $U(\theta)^\dagger O U(\theta)$  and the state  $\rho$ . Given that both of these objects live in the exponentially large vector space of operators, one can generally expect this overlap to be –on average over  $\theta$ – exponentially small. This is the essence of the barren plateaus phenomenon—the curse of dimensionality. If, however, the evolved observable is confined to a polynomially large subspace, then the loss becomes the inner product between two objects in this reduced space and can therefore avoid barren plateaus. But, in this case one can also simulate the loss by representing the initial state, circuit, and measurement operator as polynomially large objects contained in, and acting on, the small subspace.

Our general argument is supported by an analysis of widely used schemes through which we show that all considered methods for avoiding barren plateaus can be efficiently classically simulated, as well as by recent works which explicitly perform the classical simulations<sup>70,78–84</sup>. Fundamentally, it is the very proof of absence of barren plateaus that allows us to identify the polynomially-sized subspaces in which the relevant part of the computation lives. Using this information, we can then determine the set of expectation values one needs to estimate (either classically or quantumly) to enable classical simulations. In the latter case the quantum computer is still essential, but is accessed non-adaptively to generate a *classical surrogate* rather than via a hybrid optimization loop.

Given the potential for misunderstanding, let us first state a few caveats to our claims. Firstly, our argument applies to widely used models and algorithms that employ a loss function formulated as the expectation of an observable for a state evolved under a parametrized quantum circuit, as well as variants using measurements of this form followed by classical post-processing. This encompasses the majority of popular quantum architectures including most standard variational quantum algorithms, many quantum machine learning models and certain families of quantum generative schemes. However, it does not cover all possible quantum learning protocols.

Secondly, while for all our case studies it is possible to identify the ingredients necessary for simulation, we do not prove that this will always be possible. Thus, in principle, there could be models for which the landscape is free of barren plateaus, and yet we do not know how to simulate it. This could arise for sub-regions of a landscape which could be explored via smart initialization strategies, when the small subspace is otherwise unknown, or even when the problem lives in the full exponential space but is highly structured. Indeed, we provide an explicit (but highly contrived) construction for the latter and raise the possibility that more such cases might be found heuristically once larger quantum devices become available.

Finally, having identified these caveats, we present new opportunities and research directions that follow from our results. In particular, we discuss the potential offered by warm starts and the fact that even if polynomial-time classical simulation is available, the computational cost might still be too large, thus enabling potential polynomial advantages when running the variational quantum computing scheme on a quantum computer. More exotically, we suggest that by exploiting the structure of conventional fault-tolerant quantum algorithms, it might yet be possible to construct highly structured variational architectures for which superpolynomial quantum advantages can be realized.

### Definitions for barren plateaus and simulability

Variational quantum computing algorithms encode a problem of interest into an optimization task. The standard approach is to train a parametrized quantum circuit to minimize a loss function that quantifies the quality of the solution<sup>1–7</sup>. These algorithms are hybrid computational models in the sense that they use quantum hardware to obtain an estimate of the loss and then leverage the power of classical optimizers to determine parameter updates for the next set of experiments.

In what follows, we will assume that the loss function takes the form

$$\ell_\theta(\rho, O) = \text{Tr}[U(\theta)\rho U^\dagger(\theta)O]. \tag{1}$$

Here,  $\rho \in \mathcal{B}$  is an  $n$ -qubit input state belonging to the set of bounded operators  $\mathcal{B}$  acting on a  $2^n$ -dimensional Hilbert space  $\mathcal{H}$ ,  $O \in \mathcal{B}$  is some non-trivial Hermitian operator (with  $\|O\|_\infty \leq 1$ ),  $U(\theta)$  is the parametrized quantum circuit, and  $\theta$  is a set of trainable parameters. Losses of this form can be used to tackle a wide range of problems through different choices of  $\rho$ ,  $O$  and  $U(\theta)$ . We note that while algorithms can employ more general loss functions that require computing multiple such quantities (e.g., by sending different states through the circuit, or by estimating the expectation value of several operators), we will focus on the fundamental case where the loss is given by Eq. (1), as the lessons derived here can be extrapolated to other scenarios.

To better understand and classify the problems we are focusing on, we find it convenient to define *problems* specified by *classes*  $\mathcal{C} = \{\mathcal{I}\}$  of *problem instances*. A problem instance  $\mathcal{I}$  is determined by an efficiently-sampleable parameter distribution  $\mathcal{P}$ , and some efficient *classical description* of  $\rho$ ,  $U(\theta)$ , and  $O$  that can be used to estimate  $\ell_\theta(\rho, O)$  on a quantum computer in polynomial time. These could be a quantum circuit that prepares  $\rho$  from some fiducial state, a dictionary of the gate types and placements in  $U(\theta)$ , and the Pauli decomposition of  $O$ . We assume that these descriptions can be encoded in a string of size in  $\mathcal{O}(\text{poly}(n))$ .

For concreteness, let us consider an example of a problem class. Let  $\mathcal{C}_{\text{shallowHEA}}$  be the class of all instances where the circuit is a one-dimensional hardware efficient ansatz<sup>24,85</sup> composed of  $L \in \mathcal{O}(\log(n))$  layers of two-qubit gates acting on neighboring qubits in a brick-layered structure. Moreover, we take  $\rho$  to be an  $n$ -qubit state preparable by a circuit with  $\mathcal{O}(\text{poly}(n))$  gates when acting on the all-zero state, and  $O$  some Pauli operator that is diagonal in the computational basis (e.g.,  $O = Z^{\otimes n}$  or  $O = Z_\mu$  for some  $\mu \in \{1, \dots, n\}$ ). Here, we further assume that all the gates in the circuit are parametrized, and every parameter is sampled uniformly at random.

Over the past few years, there has been a tremendous amount of work put forward to understand if the loss functions in a given problem class based on variational circuits are *trainable* (and we refer the reader to ref. 86 for a subtle discussion of what trainable, or even variational, even means). Several sources of untrainability have been detected<sup>7</sup> such as the presence of sub-optimal local minima<sup>22,87–90</sup> and expressivity limitations<sup>91</sup>. However, the vast majority of trainability analysis has been concentrated around the *barren plateau*

phenomenon<sup>11–59</sup> (see ref. 11 for a review on barren plateaus). When a problem exhibits a barren plateau, its loss function becomes—on average—exponentially concentrated with the system size<sup>10,24</sup>. We say a class  $\mathcal{C}$  of problems is *provably barren plateau-free*, i.e.,  $\mathcal{C} \in \overline{\text{BP}}$ , if one can show that

$$\text{Var}_{\theta \sim \mathcal{P}}[\ell_{\theta}(\rho, O)] \in \Omega\left(\frac{1}{\text{poly}(n)}\right), \quad (2)$$

for all loss functions  $\ell_{\theta}(\rho, O)$  in the class  $\mathcal{C}$  of parameterized quantum circuits. Note that one can define a more general class of barren plateau-free problems where an explicit proof of absence of barren plateaus is not needed, but for now we will consider this restricted class.

Proving that certain types of problems are in  $\overline{\text{BP}}$  has recently become an active area of research. While such studies are extremely important, it is worth noting that just because the loss functions for a given problem are barren plateau-free does not mean that they are practically useful or that they can achieve a quantum advantage. For instance, one should wonder if the quantum computer is being employed in a meaningful way. That is, one would ideally like to show that beyond absence of barren plateaus, there exists no classical algorithm that can also efficiently compute the loss.

To better tackle this question, we then first need to define what it means to *compute* a loss function, and also what we understand by *classical simulability*. In particular, since the notion of barren plateaus is an average statement over the landscape, in this context the natural notions for computing and simulating the loss will also be average ones. However, stronger notions of simulability, such as one where one guarantees that the loss can be computed for all points on the landscape, will also be discussed below.

First, let us define the task of *computing a loss function*, such as that in Eq. (1), for a given problem. We will say that an algorithm can compute the loss functions in an instance  $\mathcal{I}$  if, with high probability  $\theta \sim \mathcal{P}$ , it can implement a function  $\tilde{\ell}_{\theta}(\rho, O)$  approximating the loss up to error  $\epsilon$ , i.e.,

$$(\tilde{\ell}_{\theta}(\rho, O) - \ell_{\theta}(\rho, O))^2 \leq \epsilon. \quad (3)$$

An algorithm can compute the loss functions in the problem  $\mathcal{C}$  if it can compute them for all instances  $\mathcal{I}$  in  $\mathcal{C}$ . One could also define problems in terms of being able to compute the loss function  $\ell_{\theta}(\rho, O)$  and its gradients  $\nabla \ell_{\theta}(\rho, O)$ . Being able to access the derivative information can be useful during the parameter training process. While in some cases computing the loss allows us to access derivatives (e.g., via parameter-shift rule<sup>92,93</sup>), we will restrict ourselves to only estimating the loss function.

As previously mentioned, the above definition guarantees that we can compute the loss function with high probability when the parameters are sampled according to  $\mathcal{P}$ . While simulating—with high probability—random point in the landscape is typically easy (see for instance<sup>79</sup>) it may not be particularly useful as it does not necessarily allow us to train the circuit. One might also be interested in computing the loss given *any* parameter settings. Thus, we will also consider a stronger version where an algorithm can compute the loss function in an instance  $\mathcal{I}$  if *for all*  $\theta$ , it can implement a function  $\tilde{\ell}_{\theta}(\rho, O)$  approximating the loss up to error  $\epsilon$ . Of course, more pragmatic definitions exist where we do not care about approximating the loss function in average, or across all points, but where we instead focus on actually solving the optimization task at hand. As such, one may wish to obtain a  $\tilde{\ell}_{\theta}(\rho, O)$  such that the solution to the optimization problem  $\arg \min_{\theta} \tilde{\ell}_{\theta}(\rho, O)$  is as good (according to some metric) as that obtained

by solving  $\arg \min_{\theta} \ell_{\theta}(\rho, O)$ . We will not focus on this case here but we note that its exploration is an open area of research<sup>78,94</sup>.

Now that we have defined what it means to compute (with high probability or with certainty over  $\theta$ ) a loss, we can present different notions of what it means to simulate it. We begin with the most basic and intuitive definition for classical simulability, which we simply dub *Classical Simulation* (CSIM). A problem  $\mathcal{C}$  is in CSIM if a polynomial-time classical algorithm can compute every instance in  $\mathcal{C}$ . As schematically shown in Fig. 1, CSIM is performed entirely on a classical device with no need nor access to a quantum computer.

We also consider classical simulation algorithms enhanced by polynomial-size data obtained from quantum experiments, which we denote as *Quantum Enhanced Classical Simulation* (QESIM), which morally means that the quantum landscape can be *classically surrogated*. In this case, one is given a problem instance and is allowed to use a quantum computer for an initial data acquisition phase, which takes no more than polynomial time (and uses no more than polynomial-memory). During this phase, one can prepare copies of the initial quantum state, apply some operations, and independently measure them via some efficient tomographic or classical shadow techniques<sup>75,95,96</sup>. Such a procedure can be used to obtain an efficiently storable classical representation (i.e., storable in string of size in  $\mathcal{O}(\text{poly}(n))$ ) of the state, unitary or measurement operator. Once this initial phase is over, one cannot access the quantum device anymore, and the computation of  $\ell_{\theta}(\rho, O)$  has to be done purely classically. A problem  $\mathcal{C}$  is in QESIM if a polynomial-time classical algorithm, which can utilize data obtained from quantum devices in an initial data acquisition phase, can compute every instance in  $\mathcal{C}$ . Here we note that the problem of estimating a loss with data from a quantum computer can be cast into a decision problem. In this case, CSIM can be related to BPP, QESIM is closely connected to BPP/Samp<sup>74</sup> and even more closely connected to BPP/qgenpoly<sup>97</sup>, and QSIM is related to BQP. Speaking more informally, algorithms in QESIM are also often described as classical surrogate methods.

We stress that all problems in CSIM and QESIM do not require running a parametrized quantum circuit on the quantum computer. Quantum resources are either entirely unnecessary (CSIM), or used only in some initial data acquisition phase (QESIM).

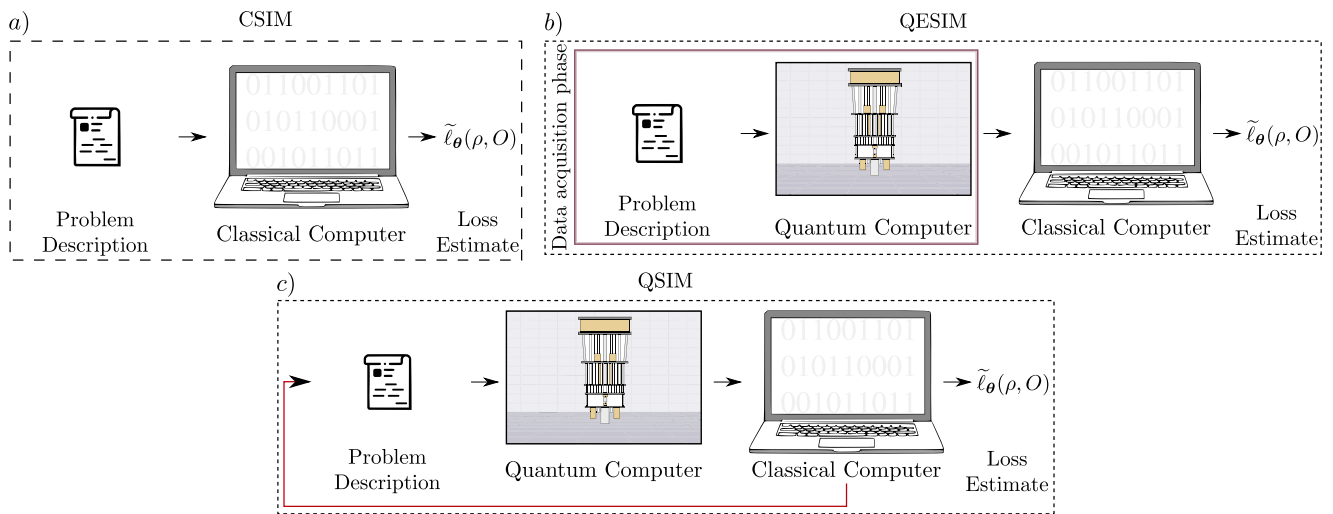
This leads us to define *Quantum Simulation* (QSIM). A problem class  $\mathcal{C}$  is in QSIM if a polynomial-time quantum algorithm can compute all instances in  $\mathcal{C}$ . As depicted in Fig. 1, models in QSIM allow for feedback between the classical and quantum computer. Moreover, the models in QSIM will usually require implementing the parametrized quantum circuit on quantum hardware.

This is a convenient point to make a few important remarks. First, as shown in Fig. 2, we note that, by definition, the following inclusions hold:

$$\text{CSIM} \subseteq \text{QESIM} \subseteq \text{QSIM}. \quad (4)$$

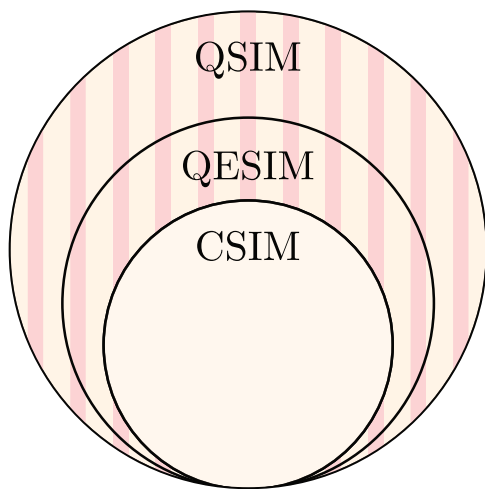
For instance,  $\text{CSIM} \subseteq \text{QESIM}$  because if the loss is fully classically simulable then it can be estimated by simply skipping the quantum computer. The rest of the inclusions follow similarly. Clearly, QSIM is *not* the largest possible set, and any loss that requires exponential time to estimate, even with a quantum computer, is beyond QSIM.

A quantum advantage is possible for problems where the loss can be simulated only if we have access to a quantum computer, i.e., for any problem in  $\text{QSIM} \cap \overline{\text{CSIM}}$ . In fact, any problem where the loss is in QESIM but not CSIM is already capable of a quantum advantage as it requires a quantum device. The problems that fall in  $\text{QESIM} \cap \overline{\text{CSIM}}$  may well be the most suited for implementation on a near-term quantum computer as the data acquisition phase could be less noisy than fully implementing the parametrized quantum circuit<sup>74,97–101</sup>.



**Fig. 1 | Schematic description of the simulation classes.** **a** Problems in CSIM are those for which there exists a fully classical algorithm that takes as input the problem description and estimates the loss function in polynomial time with a classical computer. Here, access to a quantum computer is not needed. **b** Problems in QESIM are those where one is allowed access to a quantum computer for an initial data acquisition phase that takes no more than polynomial time. At the end of this phase, access to the quantum computer is no longer allowed. A classical algorithm

then takes the problem's description, and the data obtained from the quantum device, to estimate the loss function in polynomial time. Note that one can compute problems in QESIM without needing to run a parametrized quantum circuit on the quantum hardware. **c** Problems in QSIM are those where one allows 'on demand' access to a quantum computer. Here, one usually implements the parametrized quantum circuit on the device.



**Fig. 2 | Class inclusions.** We show the inclusions between the classes CSIM, QESIM and QSIM. The region outside of CSIM (highlighted with stripes) corresponds to problems where a quantum advantage could potentially be achieved.

With the definitions above, we are ready to ask the main question that motivates this work: *Are all provably barren plateau-free loss functions also classically simulable (given polynomial-size data)?* That is, if:

$$\overline{\text{BP}} \stackrel{?}{\subset} \text{CSIM} \text{ or } \overline{\text{BP}} \stackrel{?}{\subset} \text{QESIM}. \tag{5}$$

### What leads to absence of barren plateaus?

To understand whether barren plateau-free losses are simulable, one must first understand the conditions leading to non-exponential concentration. While the study of barren plateaus was initially limited to case-by-case analyses, recent results have transformed our understanding of this phenomenon<sup>52–54</sup>. As such, what previously seemed

like a fragmented patchwork of special cases has started to coalesce into a cohesive unified theory of barren plateaus. In what follows, we will attempt to give an intuition for the sources of barren plateaus, and concomitantly how to avoid them.

To begin, let us note the simple, yet extremely important, fact that the loss function can be re-written as

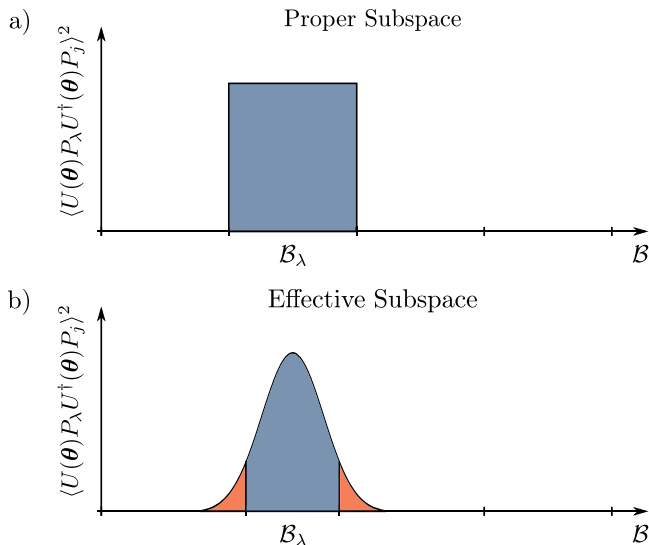
$$\ell_{\theta}(\rho, O) = \langle \rho(\theta), O \rangle = \langle \rho, O(\theta) \rangle, \tag{6}$$

where we have defined  $\rho(\theta) = U(\theta)\rho U^{\dagger}(\theta)$ ,  $O(\theta) = U^{\dagger}(\theta)OU(\theta)$ , and where  $\langle A, B \rangle = \text{Tr}[A^{\dagger}B]$  denotes the Hilbert–Schmidt inner product<sup>102</sup>. At a first glance, Eq. (6) indicates that the loss is expressed as the inner product—a similarity measure—between two (exponentially large) operators of  $\mathcal{B}$ . This fact should already raise some red flags as one can expect that, under quite general assumptions, the inner product between two exponentially large objects will be (on average) exponentially small and concentrated. For instance, we refer the reader to<sup>11,53</sup> for a formalization of such argument. As such, problems with loss functions such as those in Eq. (6) can generically be expected to have barren plateaus.

If, however, the unitary  $U(\theta)$  possesses additional structure, such as respecting some symmetry, then the loss can inherit this structure and potentially avoid the aforementioned issues. In particular, let us consider the adjoint action of  $U(\theta)$  over the operator space  $\mathcal{B}$ , and let us analyze if it leads to either subspaces, or, effective subspaces. That is, given some  $\theta$  and operator  $P$  in an appropriate orthogonal basis of  $\mathcal{B}$  (such as a Pauli operator), we want to know: *Where can  $U(\theta)PU^{\dagger}(\theta)$  go?* Note that one can, and should, also ask the question: *Where can  $U^{\dagger}(\theta)PU(\theta)$  go?* But for simplicity of notation we will consider the first question.

Mathematically, this question can be answered by computing the inner products  $\langle U(\theta)PU^{\dagger}(\theta), P_j \rangle$  for the rest of the basis elements  $P_j$ . For instance, if the inner product is non-zero for all  $P_j$ , then we know that the unitary can transform  $P$  into an operator that can spread out and reach operators across all of  $\mathcal{B}$ . On the other hand, it could happen that the adjoint action of the unitary can only reach certain operators in  $\mathcal{B}$ . For instance, as shown in Fig. 3a, the adjoint action of  $U(\theta)$  can lead to well-defined subspaces in  $\mathcal{B}$  such that  $\langle U(\theta)PU^{\dagger}(\theta), P_j \rangle$  is only





**Fig. 3 | Adjoint action subspaces.** **a** Given an operator  $P_\lambda$  and some parameters  $\theta$ , we say  $\mathcal{B}_\lambda \subset \mathcal{B}$  is a proper subspace if  $\langle U(\theta)P_\lambda U^\dagger(\theta), P_j \rangle^2$  is non-zero only for operators  $P_j$  in  $\mathcal{B}_\lambda$ . **b** We will instead say that  $\mathcal{B}_\lambda$  is an effective subspace if  $\langle U(\theta)P_\lambda U^\dagger(\theta), P_j \rangle^2$  is non-zero for many operators outside of  $\mathcal{B}_\lambda$ , but is only large for operators in  $\mathcal{B}_\lambda$ . Proper and effective subspaces can arise either for all  $\theta$ , or with high probability when sampling  $\theta$  from  $\mathcal{P}$ .

non-zero for operators in some  $\mathcal{B}_p \subset \mathcal{B}$ . This case can arise for instance in circuits with small Lie algebraic modules<sup>22,52–54,60,67,103,104</sup> or in shallow-depth hardware efficient ansätze<sup>24–27</sup>.

A second case of interest arises when we have that  $\langle U(\theta)P_\lambda U^\dagger(\theta), P_j \rangle$  is non-zero for a wide range of operators, but is large only for operators in a given subspace  $\mathcal{B}_p$ . This case is depicted schematically in Fig. 3(b) and we will say that given  $P$ , the adjoint action of  $U(\theta)$  leads to ‘effective’ subspaces, rather than proper ones. We note that effective subspaces can either arise for all  $\theta$  or with high probability for  $\theta \sim \mathcal{P}$ . In the latter case, for most  $\theta$  values for the  $\langle U(\theta)P_\lambda U^\dagger(\theta), P_j \rangle$  are large only within a small subspace, but for some low probability  $\theta$  values, large overlaps outside this subspace could be observed. Effective subspaces appear, with high probability for  $\theta \sim \mathcal{P}$ , for quantum convolutional neural networks<sup>15,105</sup>, in circuits with small angle initialization<sup>39,55,61</sup>, or even due to the effects of noise<sup>70</sup>.

To study whether or not a given problem lives within a subspace, it is convenient to express  $O$  in an orthogonal basis of  $\mathcal{B}$  as

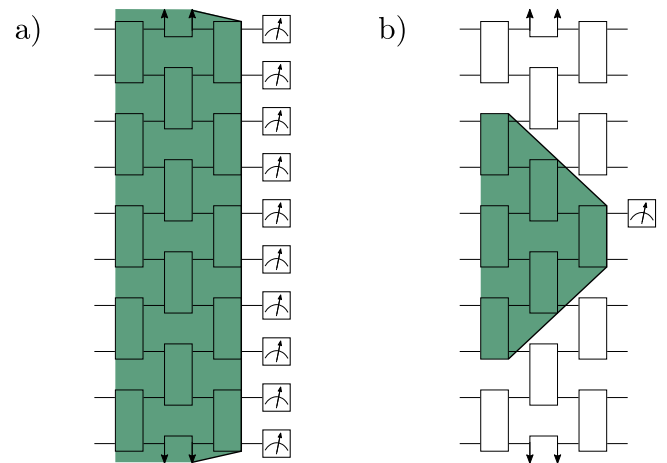
$$O = \sum_\lambda c_\lambda P_\lambda. \tag{7}$$

Then, denote as  $\mathcal{B}_\lambda \equiv \mathcal{B}_{P_\lambda}$  the subspace associated to each  $P_\lambda$  that is induced by the adjoint action of  $U(\theta)$ . Given a state  $\rho \in \mathcal{B}$ , we define its projection onto  $\mathcal{B}_\lambda$  as  $\rho_\lambda = \sum_{P_j \in \mathcal{B}_\lambda} a_j P_j$  with  $a_j = \langle \rho, P_j \rangle / \sqrt{\langle P_j, P_j \rangle}$ . In this notation, the loss function becomes:

$$\ell_\theta(\rho, O) = \sum_\lambda c_\lambda \langle \rho, P_\lambda(\theta) \rangle = \sum_\lambda c_\lambda \langle \rho_\lambda, P_\lambda(\theta) \rangle, \tag{8}$$

which reveals that  $\ell_\theta(\rho, O)$  is the sum of the inner products in each subspace. Note that in the previous equation we expanded the measurement operator and expressed the loss in terms of the subspaces obtained by Heisenberg evolving each basis element. However, one could also expand  $\rho$  and study the loss in terms of the subspaces that concomitantly arise.

If any of the subspaces appearing in Eq. (8) is only polynomially large, then we can see that some component of the loss arises from



**Fig. 4 | Subspaces for shallow hardware efficient ansätze.** We consider classes of problems where the unitary  $U(\theta)$  is an  $L$ -layered hardware efficient ansatz with two-qubit gates acting on alternating pairs of neighboring qubits in a brick-like fashion. We further assume that  $L \in \mathcal{O}(\log(n))$ . **a** For a global operator such as  $O = Z^{\otimes n}$ , the subspace obtained by adjoint action of  $U(\theta)$  is exponentially large  $\forall L$ . **b** Given a local operator, such as  $O = Z_\mu$ , one can see that for all  $\theta$  the ensuing subspace is proper and only contains Pauli operators acting on at most  $\mathcal{O}(\log(n))$  neighboring qubits. Hence, this subspace is only polynomially large. Colored regions depict the backwards light cone of the measurement operator when Heisenberg evolved<sup>42</sup>.

comparing objects (via their Hilbert-Schmidt inner product) in non-exponentially large spaces. In what follows, we will use  $\mathcal{C}_{\text{polySub}}$  to denote the class of problems where the action of  $U(\theta)$  on some term of the measurement operator is (either with high probability for  $\theta \sim \mathcal{P}$  or for all  $\theta$ ) contained in an identifiable polynomially scaling (proper or effective) subspace, i.e., such that  $\dim(\mathcal{B}_\lambda) \in \mathcal{O}(\text{poly}(n))$ , and where a basis for  $\mathcal{B}_\lambda$  can be classically obtained.

To make these ideas clearer, let us go back to the class of problems with shallow hardware efficient ansätze  $\mathcal{C}_{\text{shallowHEA}}$ . Consider first the case where the measurement operator is global, meaning that it acts on all qubits, such as  $O = Z^{\otimes n}$ . As schematically shown in Fig. 4a, by applying the circuit to this operator one can obtain exponentially many other Pauli operators acting non-trivially on all qubits, meaning that the associated subspace is exponentially large. On the other hand, as seen in Fig. 4b, if the measurement is a local operator such as  $O = Z_\mu$ , then due to the bounded light cone structure of the circuit, it can only be mapped to Paulis acting on at most  $2L \in \mathcal{O}(\log(n))$  neighboring qubits. Since there are  $\mathcal{O}(\text{poly}(n))$  such operators, the resulting subspace is a proper polynomial-sized subspace for any  $\theta$ . Hence, if  $O$  contains any local term, we will have that  $\mathcal{C}_{\text{shallowHEA}} \subset \mathcal{C}_{\text{polySub}}$ .

As mentioned above, showing that certain problem classes are in  $\mathcal{C}_{\text{polySub}}$  indicates that some part of the loss arises from comparing objects in polynomially large spaces. While this appears to be a necessary step towards non-concentrated loss functions, it is also clearly *not* a sufficient one. In particular, consider a subspace  $\mathcal{B}_\lambda$  (for some  $P_\lambda$  in Eq. (7)) such that  $\dim(\mathcal{B}_\lambda) \in \mathcal{O}(\text{poly}(n))$ , but  $\rho$  or  $O$  have almost no component when projected down into  $\mathcal{B}_\lambda$ . For instance, if  $|\langle \rho_\lambda, \rho_\lambda \rangle|$  or  $|c_\lambda^2 \langle P_\lambda, P_\lambda \rangle|$  are in  $\mathcal{O}(1/\exp(n))$ , then we will clearly have an issue since the signal in the loss coming from the polynomially-sized subspace can therefore still be exponentially weak. If the previous occurs for all small subspaces, then the loss will be obtained by comparing objects in exponentially large spaces, which we already know can lead to concentration issues.

To showcase the importance of having measurement operators and initial states which are ‘well-aligned’ with the polynomial subspace, let us consider again the class of problems with shallow hardware efficient ansätze. We already saw that if  $O$  is local (or contains

relevant local terms), then  $C_{\text{shallowHEA}} \subset C_{\text{polySub}}$ . Thus, to know whether the loss will concentrate, we need to study the Hilbert-Schmidt norm of the projected  $\rho$  into the subsets of  $2L \in \mathcal{O}(\log(n))$  adjacent qubits in the backwards light cone of the local measurement. For instance, it is not hard to check that if  $\rho$  is pure and follows a volume law of entanglement<sup>24,42</sup>, then  $|\langle \rho_\lambda, \rho_\lambda \rangle|$  will be exponentially close to zero, and the loss will be exponentially concentrated. On the other hand, if  $\rho$  is pure and satisfies an area law of entanglement<sup>42,106</sup>, then the previous inner product will be, at most, polynomially vanishing. Putting these realizations together we see that

$$C_{\text{shallowHEA}} \subset \overline{\text{BP}}, \text{ if } \rho \text{ follows area law and } O \text{ is local.}$$

Note that these results actually correspond to a reinterpretation of Theorems 1 and 2 in ref. 24, where it was shown that shallow-depth hardware efficient ansätze lead to barren plateaus for global measurements, but are barren plateau-free if the measurement operator is local and the initial state follows an area law of entanglement.

The previous example of  $C_{\text{shallowHEA}}$  illustrates an important point. We started with a problem class, identified the measurement operators which, when evolved under the adjoint of  $U(\theta)$ , remain in polynomially small subspaces, and then used this to determine the states for which the problem does not exhibit a barren plateau. This argument then leads to the question: *How general is the connection between absence of barren plateaus and polynomially small subspaces arising from the unitary's adjoint actions?* To address this, we have performed a detailed analysis of widely used barren plateau-free models and found that in all cases, it is *precisely* the existence of such subspaces which allows us to avoid exponential concentration. As such, we believe that the following claim is true for all widely used architectures and techniques:

**Claim 1.** [Standard provably barren plateau-free architectures live in classically identifiable polynomial subspaces.] For all standard problem classes  $\mathcal{C}$  such that  $\mathcal{C} \subset \overline{\text{BP}}$ , we have that

$$\mathcal{C} \subset C_{\text{polySub}}. \tag{9}$$

That is, take a problem class which provably avoids barren plateaus. Then, if one studies the parametrized unitary's adjoint action on the measurement operation and initial state, one will find that it generates operators that live either exactly, or approximately, in a polynomially sized subspace in operator space. Moreover, such subspaces can be identified classically.

By 'standard problems' we here refer to the conventional variational quantum architectures which have been proven to be in  $\overline{\text{BP}}$ . In Table 1 we present a non-exhaustive list of such architectures. Therein, we report the relevant polynomial subspace, indicating whether it is proper or effective, as well as the conditions on the measurement operator and initial state necessary for absence of barren plateaus. We emphasize that in many cases, it was the proof of absence of barren plateau itself which allowed us to determine the reported information. The majority of strategies studied, including shallow hardware efficient ansatz and highly symmetrized models, lead to proper subspaces for all  $\theta$ . The most important examples of models where one lives in effective subspaces (with high probability  $\theta \sim \mathcal{P}$ ) are quantum convolutional neural networks and small angle initialization strategies.

### Connection between absence of barren plateaus and simulability

In the previous section we claimed that barren plateaus arise as a curse of dimensionality for loss functions that compare objects in exponentially large spaces. We then argued that our attempts to fix this issue have ultimately led us to encode the problem in some polynomially small subspace, which we can classically identify. Here we

show that the existence of such a subspace can be exploited to classically simulate the loss.

To illustrate our arguments, let us again start by considering a shallow hardware efficient ansatz class where  $O = Z_\mu$ . Since the measurement is local the adjoint action of  $U(\theta)$  leads to a proper subspace  $\forall \theta$  of Pauli operators acting on at most  $2L \in \mathcal{O}(\log(n))$ -neighboring qubits. From Fig. 4(b), we can then see that if we 'drop' all the gates from  $U(\theta)$  that are outside of the measurement's backwards light cone, the loss function remains unchanged. By denoting as  $U_\lambda(\theta)$  the reduction of  $U(\theta)$  that acts only on the  $2L$  qubits in the backwards light cone, we find that

$$l_\theta(\rho, Z_\mu) = \langle \rho, U^\dagger(\theta) Z_\mu U(\theta) \rangle = \langle \rho_\lambda, U_\lambda^\dagger(\theta) Z_\mu U_\lambda(\theta) \rangle.$$

In the last equality, one computes the inner product between objects living, and acting on, the polynomially-sized subspace.

Indeed, we have taken an important step in the right direction as classically computing the loss now requires working with non-exponential operators. Still, we need to determine what  $\rho_\lambda$  is. If  $\rho$  is some product state, then one can classically find its projection onto the subspace by computing (e.g., via pen and paper) the expectation value for all operators  $P_j \in \mathcal{B}_\lambda$ . However, if  $\rho$  is some state obtained at the output of a given circuit, then there is no generic strategy which allows us to obtain these expectation values classically (even if we are promised that  $\rho$  is an area law state<sup>96</sup>). On the other hand, given access to a quantum computer, one can efficiently estimate  $\rho_\lambda$  during the initial data acquisition phase. In particular, here one simply prepares the quantum state, and measures the expectation value for all (polynomially many) operators in a basis of  $\mathcal{B}_\lambda$ . One could also simply perform standard classical shadow tomography<sup>75,95</sup>, and this information will suffice.

Hence, for local observables and area law initial states, we have that

$$C_{\text{shallowHEA}} \subset \text{CSIM}, \text{ or } C_{\text{shallowHEA}} \subset \text{QESIM},$$

and it is clear that the loss can be estimated without ever needing to use a quantum computer to run the parametrized quantum circuit. Hence, no problem based on shallow hardware efficient ansätze with a local measurement can be outside of QESIM. We note that this result was originally reported in ref. 30. It is worth additionally highlighting that, given the simplicity of the requisite measurement procedures here, one could lift the restriction that  $\rho$  is preparable by a poly depth quantum circuit and also consider quantum states that result from quantum experiments running in polynomial time, e.g., from analog or non-unitary processes. This makes the quantum-enhanced classical simulation method more flexible than directly implementing the parameterized quantum circuit on hardware<sup>107</sup>.

Here we argue that the process described above for simulating shallow HEA circuits can be applied to any problem in  $C_{\text{polySub}}$ . In all cases, the simulation follows the following three steps:

1. Identify the subspaces  $\mathcal{B}_\lambda$  of polynomial dimension.
2. Characterize the adjoint action of  $U(\theta)$  on the basis element  $P_\lambda$  in the decomposition of  $O$  (or  $\rho$ ) that lead to the  $\mathcal{B}_\lambda$ .
3. Compute or measure the component of  $\rho$  (or  $O$ ) that is in the relevant polynomial  $\mathcal{B}_\lambda$  subspaces.

Identifying the polynomial subspace  $\mathcal{B}_\lambda$  is the most important step and usually requires understanding the internal properties of  $U(\theta)$  and how these translate into the circuit's adjoint action. For instance, in the shallow hardware efficient ansatz example, we needed to realize that the relevant subspace is composed only of operators acting on the qubits in the backwards light cone. More generally, given a problem in  $\overline{\text{BP}}$ , one can obtain  $\mathcal{B}_\lambda$  by carefully analyzing the proof techniques used to show non-exponential concentration and reverse engineering what

**Table 1 | Problem classes in  $\overline{\text{BP}}$  and also in  $\mathcal{C}_{\text{polySub}}$**

Problem instance $\mathcal{C}$ based on	Refs.	Operators in the polynomial-sized $\mathcal{B}_\lambda$	$\mathcal{O}$ and $\rho$ leading to $\mathcal{C} \subset \overline{\text{BP}}$
Shallow hardware efficient ansatz	24–27,29–31	$\mathcal{O}(\log(n))$ neighboring qubits (P)	Local $\mathcal{O}$ , area law $\rho$
Generic shallow locally circuits	177,178 <sup>a</sup>	$\mathcal{O}(1)$ -weight qubits (E)	Local $\mathcal{O}$ , area law $\rho$
Quantum convolutional neural network	15	$\mathcal{O}(1)$ -weight qubits (E)	Local $\mathcal{O}$ , area law $\rho$
$U(1)$ -equivariant	22,50,51	Proj. with $\mathcal{O}(1)$ Hamming weight (P)	Equivariant $\mathcal{O}$ , $\rho \in \mathcal{B}_\lambda$
$S_n$ -equivariant	67	Permutation equivariant (P)	$\mathcal{O} \in \mathcal{B}_\lambda$ , $\text{Tr}[\rho_\lambda^2] \in \Omega(1/\text{poly}(n))$
Matchgate circuit	54	Product of $\mathcal{O}(1)$ Majoranas (P)	$\mathcal{O} \in \mathcal{B}_\lambda$ , $\text{Tr}[\rho_\lambda^2] \in \Omega(1/\text{poly}(n))$
Small angle initialization	39,55,61	$\mathcal{O}(\text{poly}(n))$ Pauli Operators (E)	Local $\mathcal{O}$ , area law $\rho$
Small Lie algebra $\mathfrak{g}$	52,58,65	Operators in $\mathfrak{g}$ (P)	$\mathcal{O} \in \mathfrak{g}$ , $\text{Tr}[\rho_\mathfrak{g}^2] \in \Omega(1/\text{poly}(n))$
Non-unital noisy-circuits	70	$\mathcal{O}(\log(n))$ qubits (E)	Local $\mathcal{O}$ , any $\rho$
Dynamic circuits	73	$\mathcal{O}(1)$ -weight qubits (E)	Local $\mathcal{O}$ , any $\rho$
Quantum generative modeling <sup>b</sup>	29,32	Tensor networks (e.g., MPS) (P)	$\mathcal{O}$ , $\rho$ computational basis proj.

In this table, we present the problems considered in our analysis. In all cases, we also provide the references where absence of barren plateaus was proven, as well as the relevant polynomially large subspace  $\mathcal{B}_\lambda$  (indicating whether it is proper (P) or effective (E)), and the states and measurement operators for which  $\mathcal{C}$  is in  $\overline{\text{BP}}$ . We refer the reader to Table 2 and the Supplemental Information for additional details on how the classical simulation is performed. Here ‘proj.’ is used as a short-hand for projector, while MPS stands for matrix product states<sup>79</sup>.

<sup>a</sup>See also the Supplemental Information for a detail proof of this claim.

<sup>b</sup>Here we refer to quantum circuit Born machines with the maximum mean discrepancy loss<sup>32</sup> and quantum generative adversarial networks with certain families of discriminators<sup>29</sup> (both using shallow hardware efficient circuits).

the exploited structure is. In particular, all proofs for the absence of barren plateaus come with fine print, in the sense that they only hold for specific choices of  $\rho$  and  $\mathcal{O}$ . Thus, most of the work has already been performed in proving absence of barren plateaus and one can take the initial states and measurement operators for which the proof holds and use these to infer  $\mathcal{B}_\lambda$ .

Once we have identified  $\mathcal{B}_\lambda$ , we can proceed to study the adjoint action of  $U(\theta)$  over the relevant  $P_\lambda$  operator. The key insight here is to note that  $U(\theta)P_\lambda U^\dagger(\theta)$  can only map to operators in  $\mathcal{B}_\lambda$ , and hence, can be somehow represented by its effective action on this subspace. The specific construction of this effective action is case dependent, but as detailed in the Supplemental Information (which contains additional details and proofs, which contains refs. 108–147.), we have been able to derive it for all considered barren plateau-free problems. In some cases, such as the aforementioned hardware efficient ansatz, one can trivially reduce  $U(\theta)$  to a smaller dimensional unitary acting on  $2L$  qubits by simply discarding all of the gates that are not in the light cone of the measurement (see Fig. 4(b)). In some other cases, one needs to employ classical simulation algorithms based on tensor networks, operator truncation, or Lie algebraic techniques<sup>104,148–150</sup>. We refer the reader to the Supplemental Information for additional details. It is worth noting that constructing the adjoint action of  $U(\theta)$  on  $\mathcal{B}_\lambda$  has a computational cost associated with it, which one needs to take into consideration to estimate the simulation cost. Importantly, while we have found algorithms whose cost is polynomial in the number of qubits, it can nevertheless have poor scaling.

The final task is to obtain the component of  $\rho$  and  $\mathcal{O}$  in  $\mathcal{B}_\lambda$ , which we respectively denote as  $\rho_\lambda$  and  $\mathcal{O}_\lambda$ . While it is possible to find the polynomially large subspaces and the effective adjoint of  $U(\theta)$  fully classically, this will not be generically true for determining  $\mathcal{O}_\lambda$  and  $\rho_\lambda$  (as we saw, for example, for the shallow HEA). While  $\mathcal{O}_\lambda$  is not much of a concern—since the classical description of  $\mathcal{O}$  could already contain this information—let us focus on obtaining  $\rho_\lambda$ . Recalling that, given an orthogonal basis  $\{|P_j\rangle\}$ , one can write  $\rho_\lambda = \sum_{P_j \in \mathcal{B}_\lambda} \langle P_j, \rho \rangle P_j / \sqrt{\langle P_j, P_j \rangle}$ , then we see that one needs to compute the expectation values  $\langle P_j, \rho \rangle$ . While for simple states such as the all-zero state this task can be performed fully classically, for general input states  $\rho$  this might not be the case.

With the previous arguments, we state the following claim.

**Claim 2.** [Problems in known polynomial subspaces are classically simulable (potentially requiring data from a quantum computer).]

Consider a problem class such that  $\mathcal{C} \subset \mathcal{C}_{\text{polySub}}$ . Then either

(i.)  $\mathcal{C} \subset \text{CSIM}$

if  $\mathcal{O}_\lambda, \rho_\lambda$  can be obtained classically, or

(ii.)  $\mathcal{C} \subset \text{QESIM}$

if  $\mathcal{O}_\lambda, \rho_\lambda$  need to be estimated on a quantum computer.

In particular, this opens the door towards the simulation of barren plateau-free loss functions  $\mathcal{C} \subset \overline{\text{BP}}$  living in a known polynomial subspace (as described in Claim 1) using a classical algorithm without the need to implement a parametrized quantum circuit on a quantum computer.

Again, for all barren plateau-free problems considered here, we have determined the relevant subspace, found efficient procedures to estimate the projections of the initial state and measurement operator, as well as for the action of the unitary. A summary of the resulting simulation algorithm and measurement protocols are reported in Table 2 and expanded upon in the Supplemental Information.

As mentioned above, and as indicated in Table 2, for problems that live in proper polynomial subspaces (which constitute the vast majority of barren plateau-free schemes considered in the literature), one can simulate the loss in the strong sense, meaning that there exists a classical algorithm that approximates the loss for any  $\theta$ . In the case of problems that live in effective subspaces, the classical simulability is available with high probability over  $\theta \sim \mathcal{P}$ . In particular, here we can simulate the parts of the loss that do not have barren plateaus. In the next section we will discuss the caveats resulting from this slightly weaker claim, but also opportunities to mitigate its disadvantages.

To wrap up, we highlight that combining Claim 1 and Claim 2 with our case-by-case analysis we can arrive at the conclusion: *For the families of problems and architectures with provable absence of barren plateaus we analyzed, it does not seem that the parametrized quantum circuit need to be implemented on a quantum computer in order to estimate the loss in polynomial time.* Note that our results do not imply a dequantization of variational quantum computing as a whole, since a quantum device might still be needed for the initial data acquisition phase. Viewed this way, our results can be viewed positively as highlighting the potential of a different learning paradigm where the quantum computer is used non-adaptively to create a classical

**Table 2 | Simulation algorithm for the problems in  $\overline{\text{BP}}$**

Problem instance $\mathcal{C}$ based on	Tomographic procedure for $\rho$	Simulation algorithm based on	Simulable
Shallow hardware efficient ansatz	Pauli classical shadows <sup>95</sup>	Light-cone sim. reduced $U(\theta)$	$\forall \theta$
Generic shallow locally circuits	Pauli classical shadows	Pauli Propagation <sup>79</sup>	$\theta \sim \mathcal{P}$
Quantum convolutional neural network <sup>a</sup>	Pauli classical shadows	Pauli Propagation	$\theta \sim \mathcal{P}$
$U(1)$ -equivariant	Computational basis measurement	Givens Rotations <sup>80</sup>	$\forall \theta$
$S_n$ -equivariant	Permutation invariant shadows <sup>181</sup>	$\mathfrak{g}$ -sim <sup>150</sup>	$\forall \theta$
Matchgate circuit	Expectation value of Pauli operators	$\mathfrak{g}$ -sim	$\forall \theta$
Small angle initialization	Pauli measurements	Tensor Networks <sup>82</sup> , Pauli Prop. <sup>80</sup>	$\theta \sim \mathcal{P}$
Small Lie algebra $\mathfrak{g}$	Expectation value of algebra elements	$\mathfrak{g}$ -sim	$\forall \theta$
Non-unital noisy-circuits	Not needed	Light-cone sim. <sup>70</sup> , Pauli Prop. <sup>151</sup>	$\theta \sim \mathcal{P}$
Dynamic circuits	Pauli classical shadows	Pauli Propagation <sup>79</sup>	$\theta \sim \mathcal{P}$
Quantum generative modeling <sup>b</sup>	Not needed	Tensor Networks	$\forall \theta$

In this table we present the problem instances considered in our analysis and an example algorithm that we can use to simulate them. Alternatives to tensor networks can also be used, e.g., refs. 148,149,182–184. We also provide the tomographic procedure that could be used for a generic initial state as well as indicating whether the simulation is available on average, or for all points in the landscape.

<sup>a</sup>In the Supplemental Information we also show that certain parameter restricted quantum convolutional neural networks can be simulated for all  $\theta$ . For a general simulation of vanilla quantum convolutional neural networks see<sup>78</sup>.

<sup>b</sup>Here we refer to quantum circuit Born machines with the maximum mean discrepancy loss<sup>32</sup> and quantum generative adversarial networks with certain families of discriminators<sup>39</sup> (both using shallow hardware efficient circuits).

surrogate of the loss landscape. Still, in a way, Claims 1 and 2 *do*, in a sense, dequantize the variational part of the model and shed some serious doubts on the non-classicality of the information processing being done by a barren plateau-free loss function.

### Caveats and future directions

In this section, we present several caveats to our arguments, as well as interesting new research directions.

#### Caveats

First and foremost, we would like to highlight the fact that our general arguments are based on intuition gathered from a case-by-case study of widely used circuit architectures and techniques (although we do refer the reader to<sup>70,78–83,151</sup> for explicit simulation algorithms). It was not possible to analyze every work claiming absence of barren plateaus –we studied only those in Table 1—but we highly encourage the community to check if our results are as widely applicable as we believe them to be. Moreover, while we have argued that absence of barren plateaus is linked to the presence of relevant polynomially small subspaces, we do not close the door to the existence of other non-exponential concentration mechanisms. For instance, Claim 1 states *for all cases studied* that, if  $\mathcal{C} \subset \overline{\text{BP}}$  then  $\mathcal{C} \subset \mathcal{C}_{\text{polySub}}$ . We do *not* claim that  $\overline{\text{BP}} \subset \mathcal{C}_{\text{polySub}}$ . Thus we are *not* claiming that *any* non-concentrated loss can always be classically simulated.

In fact, one can construct examples of non-concentrated loss functions that are not classically simulable (see the Supplemental Information for one such example based on cryptographic hardness). Crucially, these examples do not resemble current mainstream variational quantum algorithms but instead draw inspiration from conventional fault-tolerant quantum algorithms for which we expect a superpolynomial quantum advantage to be achievable. In doing so, they break our initial assumption that comparing objects living in exponentially large spaces generically leads to concentrated expectation values, as the circuits used therein are not generic but rather purposely constructed. In fact, textbook quantum algorithms do not suffer from the curse of dimensionality as the exponentially large quantum states are manipulated in a well-thought and orderly manner, rather than a variational one. Nonetheless, as evidenced by our rather contrived examples in the Supplemental Information, it is not obvious how to use these techniques in variational quantum computing as textbook quantum algorithms are only built for very specific tasks.

Another caveat is that the amount of quantum resources required by problems that fall into QESIM varies on a case-by-case basis. Many cases, as shown in Table 2, require only Pauli measurements and thus the load on the quantum device is very light. Indeed, a universal quantum computer may not be required and rather a more bespoke analogue simulator or quantum experiment may suffice. At the other extreme one could consider a deep quantum circuit with only a single trainable parameter  $\phi$ <sup>94</sup>. Such a circuit could be constructed so that it does not have a barren plateau and yet also fall into QESIM, as one can trivially construct a classical surrogate for the landscape by running the circuit in ‘a data collection phase’ at different  $\phi$  values. The absence of barren plateau results for small *patches* of quantum landscapes close to minima<sup>152,153</sup> which can also be surrogated by running the parametrized quantum circuit and making appropriate measurements<sup>80</sup>. These cases thus strictly align with our claims that problems that can provably avoid barren plateaus fall into QESIM but stretch the original spirit of QESIM by requiring substantially more quantum resources from the quantum device.

It is worth noting that our general arguments against conventional variational quantum algorithms require provable absence of barren plateaus, as the proof itself is used in our derivation of the classical simulability techniques. As such, we cannot comment on situations where one can heuristically find large gradients (e.g., via numerics<sup>41,154,155</sup>) but where the relevant subspace is not known. Such situations could arise for very special parts of the landscape, such as warm-starts or other smart initialization strategies. Although we believe that a careful analysis of many such cases will highlight that the problem is essentially contained in a polynomially-sized subspace. For instance, we can refer to the specific case of ADAPT-VQE<sup>41</sup>, which has no formal proof of absence of barren plateaus but has been heuristically shown to have large gradients for close to identity initializations. Here, it has been recently shown that Majorana Propagation<sup>84</sup> could efficiently simulate the action of the ADAPT-VQE circuit. Of course, these are preliminary results and there is as of yet no efficient protocol for characterizing the adjoint action of  $U(\theta)$  throughout training and guaranteeing that it will remain in the simulable region. Indeed, as suggested by our counterexamples, it might be that the clever initializations explore an exponentially large space in some structured—but unknown—manner.

Relatedly, let us highlight that, for problems that live in effective but not proper subspaces, classical simulation is possible for  $\theta \sim \mathcal{P}$



only with high probability. While this can enable for the simulation of randomly chosen points in certain regions of the landscape<sup>80,156</sup>, this need not be necessarily useful for training, as one may be in a non-interesting part of the landscape, or the training might take us towards regions that not classically simulable. More concretely, if the subspace is effective (as in Fig. 3(b)), our methods will provide simulation techniques that will only be faithful in the vicinity of the initialization. However, if the relevant sector  $B_1$  shifts during training the simulation might cease to be reliable. The extent to which this is possible is yet to be explored, but recent results have shown that this phenomenon does not seem to occur in quantum convolutional neural networks<sup>78</sup> when benchmarked in standard tests, as there one initializes –and remains during training– in the same classically simulable region of the landscape.

Finally, we would like to again highlight that our analysis is oriented towards loss functions based on quantities such as that in Eq. (1). Hence, our claims are not directly applicable to problems such as quantum Boltzmann machines<sup>157</sup> where one employs different types of loss functions. In fact, quantum Boltzmann machines are another example of a variational model built upon a quantum primitive (thermal state preparation) which one expects is hard to simulate classically. Hence this is an interesting avenue for balancing quantum advantage and trainability considerations.

### New opportunities

Perhaps the main goal of this perspective article is to provide new and exciting research directions that take into account the connection between barren plateaus and simulability. As such, we here discuss several new opportunities that we think could be fruitful to pursue.

First and foremost, we would like to note that just because some loss function can be classically simulated in polynomial time, this does not mean that it is still practical to do so. In particular, some of the simulation algorithms we have found (see the Supplemental Information) could still be prohibitively expensive to implement. Thus, by embracing the fact that some barren plateau-free loss functions are classically simulable, one could compare the computational complexity of the simulation versus that of estimating the loss function in a quantum computer, and thus potentially find provable polynomial speed-ups or at least more favorable constant factors with the same polynomial scaling. We refer the reader to refs. 81,158,159 for some examples on this research direction.

Second, we highlight the fact that being able to classically simulate and train a model without the need for expensive computational resources could be extremely useful. In particular, it is well known that there exist problems where one can classically train a variational state by minimizing some expectation value, but sampling from such a state on a classical computer is prohibitively expensive (such as in certain optimization problems or generative modeling tasks)<sup>160–165</sup> or may not be supported by the particular simulation algorithm. Similarly, one can envision tasks where the goal is to variationally prepare some state for a quantum sensing task<sup>166–168</sup>. Here, the ultimate goal is to obtain a metrological advantage rather than a computational one, meaning that classical simulations could be beneficial to save precious quantum resources. This paves the way for finding tasks where one ultimately cares about being able to prepare or sample from a state, as here one can train the parameters on a classical computer, transfer them to a quantum device (e.g., via refs. 169,170), and perform additional operations and measurements therein.

Third, and as discussed in the Supplemental Information, there appears to be no ‘one simulation algorithm’ to rule them all, as we have found that different loss functions require different simulation algorithms. Hence, given some barren plateau-free model, we envision that asking the question ‘How can we simulate this loss?’ will lead us to new data-driven and quantum-inspired simulation algorithms. In the same line of thought, we think that our quest to better understand what

makes an initial state well-aligned with the relevant subspaces, and thus a barren plateau-free loss, can allow us to uncover what are the true resources that make a loss function more difficult to simulate, but also more concentrated. For instance, the recent work of<sup>171</sup> studied why certain quantum experiments<sup>172</sup> could be efficiently simulated via tensor networks?. Therein, the authors showed that the dynamics generated by the quantum circuit ultimately live in a small (identifiable) subspace. In parallel, as most classical algorithms will have their limits where they particularly excel or perform poorly, it may also turn out that quantum methods where one runs the circuit on the device could still provide the most flexible ‘catch all’ simulation method.

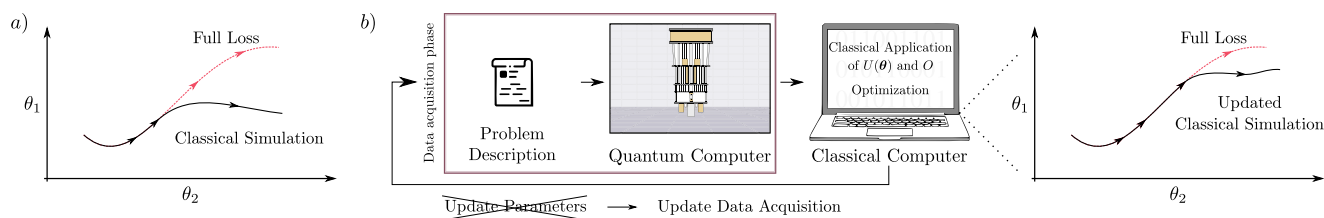
As an example, we refer the reader to ref. 54, where a connection between computational resources in fermionic linear optic circuits, absence of barren plateaus, and simulability is unraveled. Therein, it was shown that when the quantum resources (measured by fermionic entanglement) are increased, the loss function becomes more concentrated. Simultaneously, even if the state has a large component in the polynomial subspace one still needs to estimate it, and whether such computation can be classically performed will likely determine if the loss is in CSIM or QESIM (see Claim 2). We thus leave for future work a detailed study of the initial states and measurement operators for which the problems are actually in CSIM. Note that this analysis could promote simulations that are available on average, to faithful simulations for all  $\theta$ . For instance, a quantum convolutional neural network with product initial states can be efficiently simulated via tensor networks<sup>105</sup>, and hence is actually in CSIM  $\forall \theta$ . We believe that such pursuits will lead to deep insights into architecture-dependent resource theories and the limits of classical simulation algorithms.

Going further, we have noted that problems which are in QESIM but not in CSIM could be precisely those that are more amenable for implementation in the near term. This is due to the fact that the data acquisition might require shallower circuits than the implementation of  $U(\theta)$ , and thus be less impacted by hardware noise. Moreover, optimizing directly on the classical device will generally be much more efficient given the relative speed of classical devices and the possibility of using tools such as automatic differentiation through gradient back-propagation. Indeed, while the loss is defined as a unitary evolution of the state’s projection onto the polynomial subspace, once the data is stored classically, one can potentially analyze it with more general procedures such as sending it through an appropriate classical neural network. This then begs the question of whether the best way to analyze the tomographic data is via a unitary map, or some more general function.

More fundamentally, the very fact that a study of barren plateaus has led us to tomographic techniques<sup>75,95,96</sup> is in itself extremely interesting. Here, one can wonder about the limits of measure-first algorithms where initial data acquisition phases are allowed. While it is known that generically using the same shadows protocol for every problem has its limitations<sup>101</sup>, our results indicate that for every barren plateau-free problem there is an associated tomographic procedure, and that these protocols can vary widely from one task to the other. The extent to which these data acquisition methods are connected, and what their limitations are, is therefore an open question.

Regarding the utility of variational quantum computing and machine learning for classical data<sup>4–8</sup>, our results should help us better understand if the quantum computer is being used as a meaningful form of enhanced feature-space. Namely, if the problem’s loss function is barren plateau-free, and we can identify how to simulate it, one should ask whether the loss is in CSIM or in QESIM. Answering these questions could rule out situations where the use of a quantum computer is not necessary. One could further explore the parallels here with recent research into the possibility of classical surrogates for quantum models using classical data<sup>107,173–176</sup>.

As discussed in the previous section, we know that if the problem has effective subspaces, then one could potentially simulate and train



**Fig. 5 | Absence of barren plateaus, simulability, and faithful training.** **a** We consider a problem in  $\mathcal{BP}$  such that, with high probability for some  $\theta \sim \mathcal{P}$ , the circuit’s adjoint action leads to a polynomial effective subspace as in Fig. 3b. Training on the classically estimated loss can be faithful for the first few optimization steps. However, as the optimization continues, a vital contribution to the loss could arise from operators not in  $\mathcal{B}_1$ . If this occurs, training on the classically estimated loss can be unfaithful and one can converge towards a parameter region

that does not correspond to the minima of the true ‘full’ loss function. **b** To mitigate the aforementioned issues, one could use an alternative form of hybrid variational quantum computing where multiple, iterative, data acquisition steps are used. This information is then used by a classical computer to update a classical simulation of the loss and make it more faithful as the optimization progresses. As schematically shown, this scheme could make training the loss on a classical computer more faithful.

the loss given some initial set of parameters, but that the loss might not be faithful for the full optimization procedure (see Fig. 5(a)). To mitigate this issue, we envision a novel form of hybrid quantum computation as shown in Fig. 5(b). Here, there is still a feedback loop between the classical and quantum device (similar to that of standard variational methods), but instead of updating parameters in a quantum circuit, one updates the data acquired from the initial state. As such, as the optimization progresses, one uses the quantum computer to take new measurements on the initial state, and then uses this information for a more faithful, updated classical simulation. While we have not tested the performance of this protocol, we believe it might be useful and encourage the community to try it out.

To finish, we note that many schemes are not covered by our results. These include certain clever initialization strategies and many quantum learning models (such as some schemes in generative modeling). Moreover, we have shown that by drawing inspiration from conventional quantum algorithms one can construct unorthodox variational settings where manipulation of exponentially large objects is enabled, barren plateaus can be avoided, and a quantum advantage is still possible. In this regard, we note that under widely held cryptographic assumptions it is believed that  $\text{BQP} \not\subseteq \text{BPP/poly}$ . Ref. 77 uses this fact in a PAC-learning setting to argue that many-body physics problems that are BQP-complete could be used to construct learning problems that are not classically learnable. There are many differences between their setting and ours and so whether or not similar arguments could be applied here requires careful work. Nonetheless, this cautiously hints towards the exciting possibility that there could still be useful situations where the implementation of parameterized circuits on quantum hardware can be used to achieve an exponential advantage.

## Conclusions

The arrival of variational quantum algorithms effectively democratized the world of quantum computing. Whereas coming up with new conventional quantum algorithms requires a careful consideration of how best to manipulate and extract quantum information, proposing a new variational quantum algorithm is relatively straightforward. One can simply identify a potential loss function and a circuit ansatz to use. The hard work of minimizing that cost is offloaded to a classical optimizer, reducing the burden both on the quantum computer and on the quantum researcher, or so it was hoped.

While the hopes for this approach were largely fueled by the overwhelming success of neural networks in classical machine learning, it did not take long for the community to realize that variational quantum computing can be significantly more challenging than its classical counterpart. In fact, as any practitioner can testify, actually performing the optimization for moderate-sized problems is frustrating. This opened the door to studying causes of untrainability, leading to the discovery of several sources of barren plateaus and how to avoid

them. While this quest was in many ways fruitful, with numerous approaches identified, here we have argued that all of these approaches are at heart rather simple. More concretely, we have argued that the strategies for avoiding barren plateaus considered by the community so far lead to algorithms that effectively live in polynomially sized subspaces. From here, one can do away with parameterized quantum circuits and instead simulate the algorithms classically (potentially after an initial data acquisition stage on quantum hardware).

Our argument here at times borders on trivial, and may be—in some form, or another—already known to many. However, the connection between the absence of barren plateaus and simulability still has not permeated the field’s Zeitgeist, and it is not uncommon to find claims where the absence of barren plateaus is equated with practical usefulness. In this manner, our conclusions push against current practice in the community and the net result is that variational quantum algorithms need a rethink.

*So, where does this leave us?* One path forward is to embrace our case-by-case argument as something positive. While some classical simulations have been performed<sup>70,78–84</sup>, much work remains to be done in this field of “quantum machine learning dequantization”. Moreover, the fact that a quantum computer is generally not required to implement the parameterized quantum circuit, but might be needed for an initial data acquisition stage, makes such algorithms much easier to implement. To take advantage of this opportunity will require research into how to best perform the data acquisition stage and how to best perform the classical simulation needed to optimize the loss. Moreover, careful analytic work will be required to understand the scaling of the classical simulation algorithms. It may well turn out that simulating the parametrized quantum circuit on quantum hardware enjoys polynomial advantages over our best classical algorithms.

Alternatively, one could rally against our arguments here and strive harder to identify potential avenues for achieving a provable exponential quantum advantage with parameterized quantum circuits. Of course, there are numerous classically hard quantum circuits and so it’s natural to ask: *Can’t this be used to show that there are variational quantum algorithms that are barren plateau-free but not classically simulable even with data from quantum experiments?* Indeed, we showed that this is possible for a few contrived examples. The key question is whether the non-simulability of quantum circuits can be used more generally to construct useful, trainable, and non-classically simulable variational quantum algorithms. To achieve this we should seek to draw inspiration from conventional fault-tolerant quantum algorithms, and carefully consider how best to manipulate information in the exponentially large spaces in which quantum operators live. Variational quantum computing may still be a viable line of research in this regard but it will require a principled approach and a healthy dose of imagination.

Finally, our focus here has been on analytically studying the scaling of variational quantum algorithms because we simply do not currently have the hardware to study how these algorithms perform for problem sizes we are actually interested in. However, there might be cases where, although we cannot prove that an algorithm does not have a barren plateau, training is possible. Indeed, this is the case for classical machine learning, whose heuristic success goes well beyond what can be guaranteed analytically. As such, there could be architectures which we do not know how to simulate, and where we cannot prove they are barren plateau-free, and yet seem to work in practice.

We hope this perspective encourages the community to pause, reflect and take a more principled approach to variational quantum computing.

## References

- Cerezo, M. et al. Variational quantum algorithms. *Nat. Rev. Phys.* **3**, 625–644 (2021).
- Bharti, K. et al. Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.* **94**, 015004 (2022).
- Endo, S., Cai, Z., Benjamin, S. C. & Yuan, X. Hybrid quantum-classical algorithms and quantum error mitigation. *J. Phys. Soc. Jpn.* **90**, 032001 (2021).
- Wiebe, N., Kapoor, A. and Svore, K.M., Quantum deep learning, <https://arxiv.org/abs/1412.3489> arXiv preprint arXiv:1412.3489 (2014).
- Schuld, M., Sinayskiy, I. & Petruccione, F. An introduction to quantum machine learning. *Contemp. Phys.* **56**, 172 (2015).
- Biamonte, J. et al. Quantum machine learning. *Nature* **549**, 195 (2017).
- Cerezo, M., Verdon, G., Huang, H.-Y., Cincio, L. & Coles, P. J. Challenges and opportunities in quantum machine learning. *Nat. Computational Sci.* **2**, 567–576 (2022).
- Di Meglio, A. et al. Quantum computing for high-energy physics: state of the art and challenges. summary of the qc4hep working group. *Prx quantum* **5**, 037001 (2024).
- Abbas, A. et al. On quantum backpropagation, information reuse, and cheating measurement collapse, In *Advances in Neural Information Processing Systems* 36 [https://papers.nips.cc/paper\\_files/paper/2023/hash/8c3caae2f725c8e2a55ecd600563d172-Abstract-Conference.html](https://papers.nips.cc/paper_files/paper/2023/hash/8c3caae2f725c8e2a55ecd600563d172-Abstract-Conference.html) (2024).
- McClellan, J. R., Boixo, S., Smelyanskiy, V. N., Babbush, R. & Neven, H. Barren plateaus in quantum neural network training landscapes. *Nat. Commun.* **9**, 1 (2018).
- Larocca, M. et al. A review of barren plateaus in variational quantum computing. *Nat. Rev. Phys.* **3**, 625–644 (2025).
- Marrero, C. O., Kieferová, M. & Wiebe, N. Entanglement-induced barren plateaus. *PRX Quantum* **2**, 040316 (2021).
- Sharma, K., Cerezo, M., Cincio, L. & Coles, P. J. Trainability of dissipative perceptron-based quantum neural networks. *Phys. Rev. Lett.* **128**, 180505 (2022).
- Patti, T. L., Najafi, K., Gao, X. & Yelin, S. F. Entanglement devised barren plateau mitigation. *Phys. Rev. Res.* **3**, 033090 (2021).
- Pesah, A. et al. Absence of barren plateaus in quantum convolutional neural networks. *Phys. Rev. X* **11**, 041011 (2021).
- Uvarov, A. & Biamonte, J. D. On barren plateaus and cost function locality in variational quantum algorithms. *J. Phys. A: Math. Theor.* **54**, 245301 (2021).
- Cerezo, M. & Coles, P. J. Higher order derivatives of quantum neural networks with barren plateaus. *Quantum Sci. Technol.* **6**, 035006 (2021).
- Uvarov, A., Biamonte, J. D. & Yudin, D. Variational quantum eigensolver for frustrated quantum systems. *Phys. Rev. B* **102**, 075104 (2020).
- Wang, S. et al. Noise-induced barren plateaus in variational quantum algorithms. *Nat. Commun.* **12**, 1 (2021).
- Abbas, A. et al. The power of quantum neural networks. *Nat. Computational Sci.* **1**, 403 (2021).
- Arrasmith, A., Holmes, Z., Cerezo, M. & Coles, P. J. Equivalence of quantum barren plateaus to cost concentration and narrow gorges. *Quantum Sci. Technol.* **7**, 045015 (2022).
- Larocca, M. et al. Diagnosing Barren Plateaus with Tools from Quantum Optimal Control. *Quantum* **6**, 824 (2022).
- Holmes, Z., Sharma, K., Cerezo, M. & Coles, P. J. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. *PRX Quantum* **3**, 010313 (2022).
- Cerezo, M., Sone, A., Volkoff, T., Cincio, L. & Coles, P. J. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nat. Commun.* **12**, 1 (2021).
- Khatri, S. et al. Quantum-assisted quantum compiling. *Quantum* **3**, 140 (2019).
- Zhao, C. & Gao, X.-S. Analyzing the barren plateau phenomenon in training quantum neural networks with the ZX-calculus. *Quantum* **5**, 466 (2021).
- Liu, Z., Yu, L.-W., Duan, L.-M. & Deng, D.-L. The presence and absence of barren plateaus in tensor-network based machine learning. *Phys. Rev. Lett.* **129**, 270501 (2022).
- Miao, Q. & Barthel, T. Isometric tensor network optimization for extensive hamiltonians is free of barren plateaus. *Phys. Rev. A* **109**, L050402 (2024).
- Letcher, A., Woerner, S. & Zoufal, C. Tight and efficient gradient bounds for parameterized quantum circuits. *Quantum* **8**, 1484 (2024).
- Basheer, A., Feng, Y., Ferrie, C. & Li, S. Alternating layered variational quantum circuits can be classically optimized efficiently using classical shadows, <https://arxiv.org/abs/2208.11623> arXiv preprint arXiv:2208.11623 (2022).
- Suzuki, Y. & Li, M. Effect of alternating layered ansatzes on trainability of projected quantum kernel <https://journals.aps.org/prabstract/10.1103/PhysRevA.110.012409> (2023).
- Rudolph, M. S. et al. Trainability barriers and opportunities in quantum generative modeling. *npj Quantum Inf.* **10**, 116 (2024).
- Kieferova, M., Carlos, O.M. and Wiebe, N. Quantum generative training using rényi divergences, <https://arxiv.org/abs/2106.09567> arXiv preprint arXiv:2106.09567 (2021).
- Thanaslip, S., Wang, S., Nghiem, N. A., Coles, P. J. & Cerezo, M. Subtleties in the trainability of quantum machine learning models. *Quantum Mach. Intell.* **5**, 21 (2023).
- Lee, J., Magann, A. B., Rabitz, H. A. & Arenz, C. Progress toward favorable landscapes in quantum combinatorial optimization. *Phys. Rev. A* **104**, 032401 (2021).
- Shaydulin, R. & Wild, S. M. Importance of kernel bandwidth in quantum machine learning. *Phys. Rev. A* **106**, 042407 (2022).
- Holmes, Z. et al. Barren plateaus preclude learning scramblers. *Phys. Rev. Lett.* **126**, 190501 (2021).
- Leadbeater, C., Sharrock, L., Coyle, B. & Benedetti, M. F-divergences and cost function locality in generative modelling with quantum circuits. *Entropy* **23**, 1281 (2021).
- Zhang, K., Liu, L., Hsieh, M.H. & Tao, D. Escaping from the barren plateau via Gaussian initializations in deep variational quantum circuits, in <https://openreview.net/forum?id=jXgbJdQ2Yly> *Advances in Neural Information Processing Systems* (2022).
- Martín, E. C., Plekhanov, K. & Lubasch, M. Barren plateaus in quantum tensor network optimization. *Quantum* **7**, 974 (2023).
- Grimsley, H. R., Mayhall, N. J., Barron, G. S., Barnes, E. & Economou, S. E. Adaptive, problem-tailored variational quantum eigensolver mitigates rough parameter landscapes and barren plateaus. *npj Quantum Inf.* **9**, 19 (2023).
- Leone, L., Oliviero, S. F., Cincio, L. & Cerezo, M. On the practical usefulness of the hardware efficient ansatz. *Quantum* **8**, 1395 (2024).



43. Sack, S. H., Medina, R. A., Michailidis, A. A., Kueng, R. & Serbyn, M. Avoiding barren plateaus using classical shadows. *PRX Quantum* **3**, 020365 (2022).
44. Kashif, M. & Al-Kuwari, S. The impact of cost function globality and locality in hybrid quantum neural networks on nisq devices. *Mach. Learn.: Sci. Technol.* **4**, 015004 (2023).
45. Friedrich, L. & Maziero, J. Quantum neural network cost function concentration dependency on the parametrization expressivity. *Sci. Rep.* **13**, 9978 (2023).
46. García-Martín, D., Larocca, M. & Cerezo, M. Deep quantum neural networks form gaussian processes, <https://doi.org/10.1038/s41567-025-02883-z> *Nature Physics*, 1 (2025).
47. Kulshrestha, A. & Saffro, I. Beinit: Avoiding barren plateaus in variational quantum algorithms, in <https://doi.org/10.1109/QCE53715.2022.00039> 2022 IEEE International Conference on Quantum Computing and Engineering (QCE) (IEEE, 2022) pp. 197–203.
48. Volkoff, T. J. Efficient trainability of linear optical modules in quantum optical neural networks. *J. Russian Laser Res.* **42**, 250 (2021).
49. Kashif, M. & Al-Kuwari, S. The unified effect of data encoding, ansatz expressibility and entanglement on the trainability of hqnn. *Int. J. Parallel, Emergent Distrib. Syst.* **38**, 362 (2023).
50. Monbroussou, L., Landman, J., Grilo, A. B., Kukla, R. & Kashefi, E. Trainability and expressivity of hamming-weight preserving quantum circuits for machine learning. *Quantum* **9**, 1745 (2025).
51. Raj, S. et al. Quantum deep hedging. *Quantum* **7**, 1191 (2023).
52. Fontana, E. et al. Characterizing barren plateaus in quantum ansätze with the adjoint representation. *Nat. Commun.* **15**, 7171 (2024).
53. Ragone, M. et al. A lie algebraic theory of barren plateaus for deep parameterized quantum circuits. *Nat. Commun.* **15**, 7172 (2024).
54. Diaz, N.L., García-Martín, D., Kazi, S., Larocca, M. & Cerezo, M. Showcasing a barren plateau theory beyond the dynamical lie algebra, *arXiv preprint arXiv:2310.11505* (2023).
55. Park, C.-Y. & Killoran, N. Hamiltonian variational ansatz without barren plateaus. *Quantum* **8**, 1239 (2024).
56. Sannia, A., Tacchino, F., Tavernelli, I., Giorgi, G.L. & Zambrini, R. Engineered dissipation to mitigate barren plateaus. *npj Quantum Inf.* **10**, 81 (2024).
57. Thanasilp, S., Wang, S., Cerezo, M. & Holmes, Z. Exponential concentration in quantum kernel methods. *Nat. Commun.* **15**, 5200 (2024).
58. West, M. T., Heredge, J., Sevir, M. & Usman, M. Provably trainable rotationally equivariant quantum machine learning. *PRX Quantum* **5**, 030320 (2024).
59. Mao, R., Tian, G. & Sun, X. Towards determining the presence of barren plateaus in some chemically inspired variational quantum algorithms. *Commun. Phys.* **7**, 342 (2024).
60. Heredge, J. et al. Prospects of privacy advantage in quantum machine learning, <https://arxiv.org/abs/2405.08801> *arXiv preprint arXiv:2405.08801* (2024).
61. Wang, Y., Qi, B., Ferrie, C. & Dong, D. Trainability enhancement of parameterized quantum circuits via reduced-domain parameter initialization. *Phys. Rev. Appl.* **22**, 054005 (2024).
62. Larocca, M. et al. Group-invariant quantum machine learning. *PRX Quantum* **3**, 030341 (2022).
63. Meyer, J. J. et al. Exploiting symmetry in variational quantum machine learning. *PRX Quantum* **4**, 010328 (2023).
64. Skolik, A., Cattelan, M., Yarkoni, S., Bäck, T. & Dunjko, V. Equivariant quantum circuits for learning on weighted graphs. *npj Quantum Inf.* **9**, 47 (2023).
65. Ragone, M. et al. Representation theory for geometric quantum machine learning, <https://arxiv.org/abs/2210.07980> *arXiv preprint arXiv:2210.07980* (2022).
66. Nguyen, Q. T. et al. Theory for equivariant quantum neural networks. *PRX Quantum* **5**, 020328 (2024).
67. Schatzki, L., Larocca, M., Nguyen, Q. T., Sauvage, F. & Cerezo, M. Theoretical guarantees for permutation-equivariant quantum neural networks. *npj Quantum Inf.* **10**, 12 (2024).
68. Zheng, H., Li, Z., Liu, J., Strelchuk, S. & Kondor, R. Speeding up learning quantum states through group equivariant convolutional quantum ansätze. *PRX Quantum* **4**, 020327 (2023).
69. East, R.D., Alonso-Linaje, G. & Park, C.Y. All you need is spin: Su (2) equivariant variational quantum circuits based on spin networks, <https://arxiv.org/abs/2309.07250> *arXiv preprint arXiv:2309.07250* (2023).
70. Mele, A.A. et al. Noise-induced shallow circuits and absence of barren plateaus, <https://arxiv.org/abs/2403.13927> *arXiv preprint arXiv:2403.13927* (2024).
71. Fefferman, B., Ghosh, S., Gullans, M., Kuroiwa, K. & Sharma, K. Effect of Nonunital Noise on Random-Circuit Sampling. *PRX Quantum* **5**, 030317 (2024).
72. Crognalotti, G., Grossi, M. & Bassi, A. Estimates of loss function concentration in noisy parametrized quantum circuits, <https://arxiv.org/abs/2410.01893> *arXiv preprint arXiv:2410.01893* (2024).
73. Deshpande, A. et al. Dynamic parameterized quantum circuits: expressive and barren-plateau free, *arXiv preprint arXiv:2411.05760* <https://doi.org/10.48550/arXiv.2411.05760> (2024).
74. Huang, H.-Y. et al. Power of data in quantum machine learning. *Nat. Commun.* **12**, 1 (2021).
75. Elben, A. et al. The randomized measurement toolbox, *Nature Review Physics* <https://doi.org/10.1038/s42254-022-00535-2> (2022).
76. Huang, H.-Y., Kueng, R. & Preskill, J. Information-theoretic bounds on quantum advantage in machine learning. *Phys. Rev. Lett.* **126**, 190505 (2021).
77. Gyurik, C. & Dunjko, V. Exponential separations between classical and quantum learners, <https://arxiv.org/abs/2306.16028> *arXiv preprint arXiv:2306.16028* (2023).
78. Bermejo, P. et al. Quantum convolutional neural networks are (effectively) classically simulable, <https://arxiv.org/abs/2408.12739> *arXiv preprint arXiv:2408.12739* (2024).
79. Angrisani, A. et al. Classically estimating observables of noiseless quantum circuits, <https://arxiv.org/abs/2409.01706> *arXiv preprint arXiv:2409.01706* (2024).
80. Lerch, S. et al. Efficient quantum-enhanced classical simulation for patches of quantum landscapes, *arXiv preprint arXiv:2411.19896* <https://doi.org/10.48550/arXiv.2411.19896> (2024).
81. Anschuetz, E.R. & Gao, X. Arbitrary polynomial separations in trainable quantum machine learning, *arXiv preprint arXiv:2402.08606* <https://doi.org/10.48550/arXiv.2402.08606> (2024).
82. Shin, S., Teo, Y. S. & Jeong, H. Dequantizing quantum machine learning models using tensor networks. *Phys. Rev. Res.* **6**, 023218 (2024).
83. Ermakov, I., Lychkovskiy, O. & Byrnes, T. Unified framework for efficiently computable quantum circuits, *arXiv preprint arXiv:2401.08187* <https://doi.org/10.48550/arXiv.2401.08187> (2024).
84. Miller, A. et al. Simulation of fermionic circuits using majorana propagation, <https://doi.org/10.48550/arXiv.2503.18939> *arXiv preprint arXiv:2503.18939* (2025).
85. Kandala, A. et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242 (2017).
86. Gil-Fuster, E., Gyurik, C., Pérez-Salinas, A. & Dunjko, V. On the relation between trainability and dequantization of variational quantum learning models, <https://arxiv.org/abs/2406.07072> *arXiv preprint arXiv:2406.07072* (2024).



87. Bittel, L. & Kliesch, M. Training variational quantum algorithms is NP-hard. *Phys. Rev. Lett.* **127**, 120502 (2021).
88. Fontana, E., Cerezo, M., Arrasmith, A., Rungger, I. & Coles, P. J. Non-trivial symmetries in quantum landscapes and their resilience to quantum noise. *Quantum* **6**, 804 (2022).
89. Anschuetz, E. R. & Kiani, B. T. Beyond barren plateaus: Quantum variational algorithms are swamped with traps. *Nat. Commun.* **13**, 7760 (2022).
90. Anschuetz, E. R. Critical points in quantum generative models, <https://openreview.net/forum?id=2flz55GVQN> *International Conference on Learning Representations* (2022).
91. Tikku, A. & Kim, I.H. Circuit depth versus energy in topologically ordered systems, <https://arxiv.org/abs/2210.06796> *arXiv preprint arXiv:2210.06796* (2022).
92. Mitarai, K., Negoro, M., Kitagawa, M. & Fujii, K. Quantum circuit learning. *Phys. Rev. A* **98**, 032309 (2018).
93. Schuld, M., Bergholm, V., Gogolin, C., Izaac, J. & Killoran, N. Evaluating analytic gradients on quantum hardware. *Phys. Rev. A* **99**, 032331 (2019).
94. Gil-Fuster, E., Eisert, J. & Bravo-Prieto, C. Understanding quantum machine learning also requires rethinking generalization. *Nat. Commun.* **15**, 2277 (2024).
95. Huang, H.-Y., Kueng, R. & Preskill, J. Predicting many properties of a quantum system from very few measurements. *Nat. Phys.* **16**, 1050 (2020).
96. Anshu, A. & Arunachalam, S. A survey on the complexity of learning quantum states. *Nat. Rev. Phys.* **6**, 59 (2024).
97. Jerbi, S., Gyurik, C., Marshall, S. C., Molteni, R. & Dunjko, V. Shadows of quantum machine learning. *Nat. Commun.* **15**, 5676 (2024).
98. McClean, J. R., Kimchi-Schwartz, M. E., Carter, J. & De Jong, W. A. Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states. *Phys. Rev. A* **95**, 042308 (2017).
99. Parrish, R. M., Hohenstein, E. G., McMahan, P. L. & Martinez, T. J. Quantum computation of electronic transitions using a variational quantum eigensolver. *Phys. Rev. Lett.* **122**, 230401 (2019).
100. Bharti, K. & Haug, T. Quantum-assisted simulator. *Phys. Rev. A* **104**, 042418 (2021).
101. Gyurik, C., Molteni, R. & Dunjko, V. Limitations of measure-first protocols in quantum machine learning, <https://arxiv.org/abs/2311.12618> *arXiv preprint arXiv:2311.12618* (2023).
102. Knapp, A.W. [https://doi.org/10.1007/978-1-4757-2453-0\\_1](https://doi.org/10.1007/978-1-4757-2453-0_1) *Lie Groups Beyond an Introduction*, Vol. 140 (Springer Science & Business Media, 2013).
103. Kazi, S. et al. Analyzing the quantum approximate optimization algorithm: ansätze, symmetries, and lie algebras, <https://arxiv.org/abs/2410.05187> *arXiv preprint arXiv:2410.05187* (2024).
104. Anschuetz, E. R., Bauer, A., Kiani, B. T. & Lloyd, S. Efficient classical algorithms for simulating symmetric quantum systems. *Quantum* **7**, 1189 (2023).
105. Cong, I., Choi, S. & Lukin, M. D. Quantum convolutional neural networks. *Nat. Phys.* **15**, 1273 (2019).
106. Eisert, J., Cramer, M. & Plenio, M. B. Colloquium: Area laws for the entanglement entropy. *Rev. Mod. Phys.* **82**, 277 (2010).
107. Jerbi, S. et al. The power and limitations of learning quantum dynamics incoherently, <https://arxiv.org/abs/2303.12834> *arXiv preprint arXiv:2303.12834* (2023).
108. Wiersema, R. et al. Exploring entanglement and optimization within the Hamiltonian variational ansatz. *PRX Quantum* **1**, 020319 (2020).
109. Jozsa, R. & Miyake, A. Matchgates and classical simulation of quantum circuits. *Proc. R. Soc. A: Math., Phys. Eng. Sci.* **464**, 3089 (2008).
110. Wan, K., Huggins, W. J., Lee, J. & Babbush, R. Matchgate shadows for fermionic quantum simulation. *Commun. Math. Phys.* **404**, 629 (2023).
111. De Melo, F., Ćwikliński, P. & Terhal, B. M. The power of noisy fermionic quantum computation. *N. J. Phys.* **15**, 013015 (2013).
112. Diaz, N.L. et al. Parallel-in-time quantum simulation via page and wootters quantum time, <https://arxiv.org/abs/2308.12944> *arXiv preprint arXiv:2308.12944* (2023).
113. Bravyi, S. Lagrangian representation for fermionic linear optics. *Quantum Info Comput.* **5**, 216–238 (2005).
114. Dias, B. & Koenig, R. Classical simulation of non-gaussian fermionic circuits. *Quantum* **8**, 1350 (2024).
115. Cudby, J. & Strelchuk, S. Gaussian decomposition of magic states for matchgate computations, <https://arxiv.org/abs/2307.12654> *arXiv preprint arXiv:2307.12654* (2023).
116. Gigena, N. & Rossignoli, R. Entanglement in fermion systems. *Phys. Rev. A* **92**, 042326 (2015).
117. Arrazola, J. M. et al. Universal quantum circuits for quantum chemistry. *Quantum* **6**, 742 (2022).
118. Johri, S. et al. Nearest centroid classification on a trapped ion quantum computer. *npj Quantum Inf.* **7**, 122 (2021).
119. Lopez-Piqueres, J., Chen, J. & Perdomo-Ortiz, A. Symmetric tensor networks for generative modeling and constrained combinatorial optimization, *Machine Learning: Science and Technology* <https://doi.org/10.1088/2632-2153/ace0f5> (2022).
120. Somma, R., Barnum, H., Ortiz, G. & Knill, E. Efficient solvability of Hamiltonians and limits on the power of some quantum computational models. *Phys. Rev. Lett.* **97**, 190501 (2006).
121. Zeier, R. & Schulte-Herbrüggen, T. Symmetry principles in quantum systems theory. *J. Math. Phys.* **52**, 113510 (2011).
122. Wiersema, R., Kökcü, E., Kemper, A. F. & Bakalov, B. N. Classification of dynamical lie algebras of 2-local spin systems on linear, circular and fully connected topologies. *npj Quantum Inf.* **10**, 110 (2024).
123. Kökcü, E., Wiersema, R., Kemper, A.F. & Bakalov, B.N. Classification of dynamical lie algebras generated by spin interactions on undirected graphs, *arXiv preprint arXiv:2409.19797* <https://doi.org/10.48550/arXiv.2409.19797> (2024).
124. Aguilar, G., Cichy, S., Eisert, J. & Bittel, L. Full classification of pauli lie algebras, <https://arxiv.org/abs/2408.00081> *arXiv preprint arXiv:2408.00081* (2024).
125. Žnidarič, M. Solvable non-hermitian skin effect in many-body unitary dynamics. *Phys. Rev. Res.* **4**, 033041 (2022).
126. Deneris, A.E., Bermejo, P., Braccia, P., Cincio, L. & Cerezo, M. Exact spectral gaps of random one-dimensional quantum circuits, *arXiv preprint arXiv:2408.11201* <https://doi.org/10.48550/arXiv.2408.11201> (2024).
127. Braccia, P., Bermejo, P., Cincio, L. & Cerezo, M. Computing exact moments of local random quantum circuits via tensor networks. *Quantum Mach. Intell.* **6**, 54 (2024).
128. Belkin, D. et al. Approximate t-designs in generic circuit architectures, *PRX Quantum* **5**, 040344 (2024).
129. Mittal, S. & Hunter-Jones, N. Local random quantum circuits form approximate designs on arbitrary architectures, <https://arxiv.org/abs/2310.19355> *arXiv preprint arXiv:2310.19355* (2023).
130. Benedetti, M. et al. A generative modeling approach for benchmarking and training shallow quantum circuits. *npj Quantum Inf.* **5**, 45 (2019).
131. Coyle, B., Mills, D., Danos, V. & Kashefi, E. The born supremacy: quantum advantage and training of an ising born machine. *npj Quantum Inf.* **6**, 60 (2020).
132. Alcazar, J., Leyton-Ortega, V. & Perdomo-Ortiz, A. Classical versus quantum models in machine learning: insights from a finance application. *Mach. Learn.: Sci. Technol.* **1**, 035003 (2020).

133. Benedetti, M., Grant, E., Wossnig, L. & Severini, S. Adversarial quantum circuit learning for pure state approximation. *N. J. Phys.* **21**, 043023 (2019).
134. Perdomo-Ortiz, A., Benedetti, M., Realpe-Gómez, J. & Biswas, R. Opportunities and challenges for quantum-assisted machine learning in near-term quantum computers. *Quantum Sci. Technol.* **3**, 030502 (2018).
135. C., Zoufal, Generative quantum machine learning, <https://arxiv.org/abs/2111.12738> arXiv preprint arXiv:2111.12738 (2021).
136. Ferris, A. J. & Vidal, G. Perfect sampling with unitary tensor networks. *Phys. Rev. B* **85**, 165146 (2012).
137. Stoudenmire, E. & White, S. R. Minimally entangled typical thermal state algorithms. *N. J. Phys.* **12**, 055026 (2010).
138. Markov, I. L. & Shi, Y. Simulating quantum computation by contracting tensor networks. *SIAM J. Comput.* **38**, 963 (2008).
139. Verstraete, F. & Cirac, J. I. Matrix product states represent ground states faithfully. *Phys. Rev. B* **73**, 094423 (2006).
140. Evenbly, G. & Vidal, G. Tensor network states and geometry. *J. Stat. Phys.* **145**, 891 (2011).
141. Begušić, T., Gray, J. & Chan, G. K.-L. Fast and converged classical simulations of evidence for the utility of quantum computing before fault tolerance. *Sci. Adv.* **10**, eadk4321 (2024).
142. Wecker, D., Hastings, M. B. & Troyer, M. Progress towards practical quantum variational algorithms. *Phys. Rev. A* **92**, 042303 (2015).
143. Caro, M. C. et al. Generalization in quantum machine learning from few training data. *Nat. Commun.* **13**, 4919 (2022).
144. Liu, Y.-J., Smith, A., Knap, M. & Pollmann, F. Model-independent learning of quantum phases of matter with quantum convolutional neural networks. *Phys. Rev. Lett.* **130**, 220603 (2023).
145. Hur, T., Kim, L. & Park, D. K. Quantum convolutional neural network for classical data classification. *Quantum Mach. Intell.* **4**, 3 (2022).
146. Umeano, C., Paine, A. E., Elfving, V. E. & Kyriienko, O. What can we learn from quantum convolutional neural networks?, <https://arxiv.org/abs/2308.16664> arXiv preprint arXiv:2308.16664 (2023).
147. Mele, A. A. Introduction to haar measure tools in quantum information: A beginner's tutorial. *Quantum* **8**, 1340 (2024).
148. Fontana, E., Rudolph, M. S., Duncan, R., Rungger, I. & Cirstoiu, C. Classical simulations of noisy variational quantum circuits. *npj Quantum Inf.* **11**, 1 (2025).
149. Rudolph, M. S., Fontana, E., Holmes, Z. & Cincio, L. Classical surrogate simulation of quantum systems with LOWESA, <https://arxiv.org/abs/2308.09109> arXiv preprint arXiv:2308.09109 (2023).
150. Goh, M. L., Larocca, M., Cincio, L., Cerezo, M. & Sauvage, F. Lie-algebraic classical simulations for quantum computing, <https://arxiv.org/abs/2308.01432> arXiv preprint arXiv:2308.01432 (2023).
151. Schuster, T., Yin, C., Gao, X. & Yao, N. Y. A Polynomial-Time Classical Algorithm for Noisy Random Circuit Sampling. *Proceedings of the 55th Annual ACM Symposium on Theory of Computing* 945–957 <https://doi.org/10.1145/3564246.3585234> (2023).
152. Haug, T. & Kim, M. Optimal training of variational quantum algorithms without barren plateaus, <https://arxiv.org/abs/2104.14543> arXiv preprint arXiv:2104.14543 (2021).
153. Puig, R., Drudis, M., Thanasilp, S. & Holmes, Z. Variational quantum simulation: A case study for understanding warm starts. *PRX Quantum* **6**, 010317 (2025).
154. Dborin, J., Barratt, F., Wimalaweera, V., Wright, L. & Green, A. G. Matrix product state pre-training for quantum machine learning. *Quantum Sci. Technol.* **7**, 035014 (2022).
155. Rudolph, M. S. et al. Synergistic pretraining of parametrized quantum circuits via tensor networks. *Nat. Commun.* **14**, 8367 (2023).
156. Mhiri, H. et al. A unifying account of warm start guarantees for patches of quantum landscapes, arXiv preprint arXiv:2502.07889 <https://doi.org/10.48550/arXiv.2502.07889> (2025).
157. Coopmans, L. & Benedetti, M. On the sample complexity of quantum boltzmann machine learning. *Commun. Phys.* **7**, 274 (2024).
158. Anschuetz, E. R., Hu, H.-Y., Huang, J.-L. & Gao, X. Interpretable quantum advantage in neural sequence learning. *PRX Quantum* **4**, 020338 (2023).
159. Miao, Q. & Barthel, T. Convergence and quantum advantage of trotterized mera for strongly-correlated systems. *Quantum* **9**, 1631 (2025).
160. Medvidović, M. & Carleo, G. Classical variational simulation of the quantum approximate optimization algorithm. *npj Quantum Inf.* **7**, 101 (2021).
161. Díez-Valle, P., Porras, D. & García-Ripoll, J. J. Quantum approximate optimization algorithm pseudo-boltzmann states. *Phys. Rev. Lett.* **130**, 050601 (2023).
162. Hadfield, S. A. *Quantum algorithms for scientific computing and approximate optimization* (Columbia University, 2018).
163. Sreedhar, R. et al. The quantum approximate optimization algorithm performance with low entanglement and high circuit depth, <https://arxiv.org/abs/2207.03404> arXiv preprint arXiv:2207.03404 (2022).
164. Krovi, H. Average-case hardness of estimating probabilities of random quantum circuits with a linear scaling in the error exponent, <https://arxiv.org/abs/2206.05642> arXiv preprint arXiv:2206.05642 (2022).
165. Farhi, E. & Harrow, A. W. Quantum supremacy through the quantum approximate optimization algorithm, <https://arxiv.org/abs/1602.07674> arXiv preprint arXiv:1602.07674 (2016).
166. Endo, S., Sun, J., Li, Y., Benjamin, S. C. & Yuan, X. Variational quantum simulation of general processes. *Phys. Rev. Lett.* **125**, 010501 (2020).
167. Beckey, J. L., Cerezo, M., Sone, A. & Coles, P. J. Variational quantum algorithm for estimating the quantum Fisher information. *Phys. Rev. Res.* **4**, 013083 (2022).
168. Huerta Alderete, C. et al. Inference-based quantum sensing. *Phys. Rev. Lett.* **129**, 190501 (2022).
169. Ran, S.-J. Encoding of matrix product states into quantum circuits of one-and two-qubit gates. *Phys. Rev. A* **101**, 032310 (2020).
170. Rudolph, M. S., Chen, J., Miller, J., Acharya, A. & Perdomo-Ortiz, A. Decomposition of matrix product states into shallow quantum circuits. *Quantum Sci. Technol.* **9**, 015012 (2023).
171. Tindall, J. & Sels, D. Confinement in the transverse field ising model on the heavy hex lattice. *Phys. Rev. Lett.* **133**, 180402 (2024).
172. Kim, Y. et al. Evidence for the utility of quantum computing before fault tolerance. *Nature* **618**, 500 (2023).
173. Schuld, M. and Petruccione, F. Quantum models as kernel methods, *Machine Learning with Quantum Computers*, 217 (2021).
174. Schreiber, F. J., Eisert, J. & Meyer, J. J. Classical surrogates for quantum learning models. *Phys. Rev. Lett.* **131**, 100803 (2023).
175. Landman, J., Thabet, S., Dalyac, C., Mhiri, H. & Kashefi, E. Classically approximating variational quantum machine learning with random fourier features, arXiv preprint arXiv:2210.13200 <https://doi.org/10.48550/arXiv.2210.13200> (2022).
176. Sweke, R. et al. Potential and limitations of random fourier features for dequantizing quantum machine learning. *Quantum* **9**, 1640 (2025).
177. Napp, J. Quantifying the barren plateau phenomenon for a model of unstructured variational ansätze, <https://arxiv.org/abs/2203.06174> arXiv preprint arXiv:2203.06174 (2022).
178. Zhang, H.-K., Liu, S. & Zhang, S.-X. Absence of barren plateaus in finite local-depth circuits with long-range entanglement. *Phys. Rev. Lett.* **132**, 150603 (2024).
179. Verstraete, F., Murg, V. & Cirac, J. I. Matrix product states, projected entangled pair states, and variational renormalization

- group methods for quantum spin systems. *Adv. Phys.* **57**, 143 (2008).
180. Kerenidis, I., Landman, J. & Mathur, N. Classical and quantum algorithms for orthogonal neural networks. <https://arxiv.org/abs/2106.07198> *arXiv preprint arXiv:2106.07198* (2021).
181. Tóth, G. et al. Permutationally invariant quantum tomography. *Phys. Rev. Lett.* **105**, 250403 (2010).
182. Nemkov, N. A., Kiktenko, E. O. & Fedorov, A. K. Fourier expansion in variational quantum algorithms. *Phys. Rev. A* **108**, 032406 (2023).
183. Shao, Y., Wei, F., Cheng, S. & Liu, Z. Simulating noisy variational quantum algorithms: A polynomial approach. *Phys. Rev. Lett.* **133**, 120603 (2024).
184. Begušić, T., Hejazi, K. & Chan, G.K. Simulating quantum circuit expectation values by Clifford perturbation theory. *J. Chem. Phys.* **162**, 154110 (2025).

## Acknowledgements

We are extremely grateful to Hsin-Yuan Huang for his invaluable contributions to this work. We thank Andrew Sornborger, Lukasz Cincio, Nathan Wiebe, Chae-Yeun Park, Nathan Killoran, Maria Schuld, Xanadu's Toronto office staff, and the QTML 2023 community for thoughtful and insightful conversations. M.C. acknowledges support from Los Alamos National Laboratory (LANL) ASC Beyond Moore's Law project. M.L. was supported by the Center for Nonlinear Studies at LANL. M.C., D.G.M., N.L.D. and P.B. were supported by Laboratory Directed Research and Development (LDRD) program of LANL under project numbers 20230527ECR and 20230049DR. Also, N.L.D. acknowledges support from CONICET Argentina, and P.B acknowledges support of DIPC. A.I. acknowledges support by the U.S. Department of Energy (DOE) through a quantum computing program sponsored by the LANL Information Science & Technology Institute and by the U.S. DOE, Office of Science, Office of Advanced Scientific Computing Research, under Computational Partnerships program. E.F. acknowledges the support of the UK department for Business, Energy and Industrial Strategy through the National Quantum Technologies Programme, and the support of an industrial CASE (iCASE) studentship, funded by the Engineering and Physical Sciences Research Council (grant EP/T517665/1), in collaboration with the University of Strathclyde, the National Physical Laboratory, and Quantinuum. E.R.A. acknowledges support from the Walter Burke Institute for Theoretical Physics at Caltech. S.T. and Z.H. acknowledge support from the Sandoz Family Foundation-Monique de Meuron program for Academic Promotion. ST further acknowledges the grants for development of new faculty staff, Ratchadaphiseksomphot Fund, Chulalongkorn University [grant number 3230120336 DNS\_68\_052\_2300\_012], as well as funding from National Research Council of Thailand (NRCT) [grant number N42A680126]. M.C., Z.H., and E.R.A. thank the organizers of the PennyLane Research Retreat, where part of this work was undertaken, for their hospitality. This material is based upon work supported by the U.S. Department of Energy, Office of

Science, National Quantum Information Science Research Centers, Quantum Science Center (LC). This work was also supported by the Quantum Science Center (QSC), a National Quantum Information Science Research Center of the U.S. DOE.

## Author contributions

The project was conceived by M.C., M.L., and Z.H.; Theoretical results were derived by M.C., N.L.D., and E.R.A.; All authors contributed to the manuscript review process.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-63099-6>.

**Correspondence** and requests for materials should be addressed to M. Cerezo.

**Peer review information** *Nature Communications* thanks Qiang Miao and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025