Investigating and mitigating barren plateaus in variational quantum circuits: a survey

Jack Cunningham[1] · Jun Zhuang[1]

Abstract

In recent years, variational quantum circuits (VQCs) have been widely explored to advance quantum circuits against classic models on various domains, such as quantum chemistry and quantum machine learning. Similar to classic machine-learning models, VQCs can be trained through various optimization approaches, such as gradient-based or gradient-free methods. However, when employing gradient-based methods, the gradient variance of VQCs may dramatically vanish as the number of qubits or layers increases. This issue, a.k.a. barren plateaus (BPs), seriously hinders the scaling of VQCs on large datasets. To mitigate the barren plateaus, extensive efforts have been devoted to tackling this issue through diverse strategies. In this survey, we conduct a systematic literature review of recent works from both investigation and mitigation perspectives. Furthermore, we propose a new taxonomy to categorize most existing mitigation strategies into five groups and introduce them in detail. Also, we compare the concurrent survey papers about BPs. Finally, we provide insightful discussion on future directions for BPs.
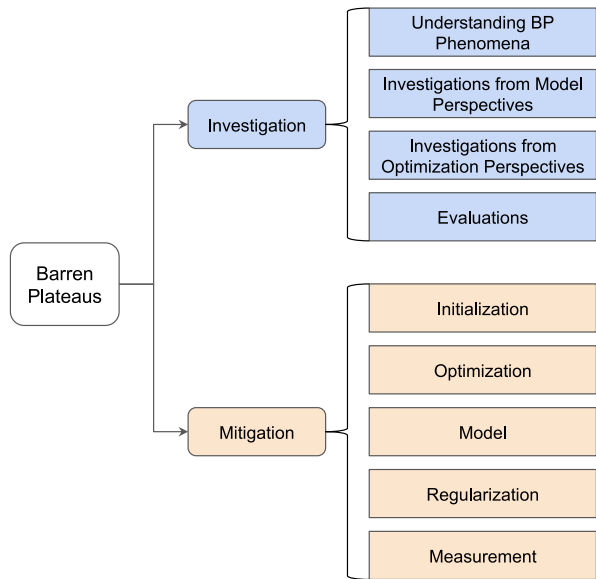
1 Introduction

In the era of noisy intermediate-scale quantum (NISQ), quantum computing has achieved significant advancement in various domains [1], such as quantum chemistry [2], quantum machine learning [3–5], and quantum architecture [6, 7]. Despite these advancements, the superiority of quantum computing over classical computing in the foreseeable future remains uncertain. One uncertainty arises during the optimization of variational quantum circuits (VQCs), which have emerged as pivotal models for advancing quantum computing. Similar to classical models, VQCs

✉ Jun Zhuang
junzhuang@boisestate.edu

[1] Computer Science, Boise State University, Boise, ID 83725, USA

Fig. 1 Our proposed taxonomy. We categorize the most existing works about barren plateaus into two aspects: investigation and mitigation

can be optimized by gradient-based approaches [8, 9]. During optimization, however, the training of VQCs may be initially trapped in a flat landscape as the model size increases. McClean et al. [10] first categorize this issue as barren plateaus (BPs) in VQCs and ascertain that the variance of the gradient will exponentially decrease as the model size increases when the VQCs match the 2-design Haar distribution. Under this occurrence, most gradient-based techniques would fail. Thus, it is essential to mitigate the barren plateau issues in VQCs.

Extensive efforts have been devoted to tackling the challenges of barren plateaus through various approaches [11]. To categorize these studies, we conduct comprehensive literature reviews on recent works about BPs from two perspectives. On the one hand, some studies aim to investigate the barren plateau phenomenon. In this group, several works conduct experiments to understand and evaluate the barren plateaus. On the other hand, extensive strategies are proposed to mitigate the barren plateaus. To better categorize these works, we propose a new taxonomy presented in Fig. 1. Overall, our contributions[1] to this survey can be summarized as follows:

- We conduct comprehensive literature reviews on recent works about barren plateaus.
- We propose a new taxonomy to help researchers better understand the background and existing works about BPs from two aspects: investigation and mitigation.
- We provide insightful discussions in this survey, including a summary of most existing mitigation methods, a comparison of concurrent survey papers, a contrast between gradient issues and barren plateaus, and a discussion of future directions.

---

[1] We maintain a repo at https://github.com/junzhuang-code/Barren_Plateaus_Survey/

In the rest of this survey, we first introduce the preliminary background of BPs in Sect. 2. Besides, we present the investigation of BPs in Sect. 3 and the mitigation strategies in Sect. 4. Last, we provide insightful discussion in Sect. 5.

2 Preliminary of barren plateaus

The term, "Barren Plateaus (BPs)", typically denotes an issue where gradient-based approaches may initially struggle to optimize variational quantum circuits (VQCs). Generally, VQCs contain a finite sequence of unitary gates $U(\theta)$ parameterized by θ. The unitary gates $U(\theta)$ can be formulated as:

$$U(\theta) = U(\theta_1, \ldots, \theta_L) = \prod_{l=1}^{L} U_l(\theta_l), \tag{1}$$

where $U_l(\theta_l) = e^{-i\theta_l V_l}$, V_l is a Hermitian operator, and L denotes the number of layers in VQCs.

Similar to classical models, VQCs' parameters can be optimized by gradient-based methods. The cost function for optimization can be denoted as $C(\theta)$. Thus, $C(\theta)$ can be formulated as the expectation over Hermitian operator H:

$$C(\theta) = \langle 0|U(\theta)^{\dagger} H U(\theta)|0\rangle. \tag{2}$$

Based on the above loss function $C(\theta)$, we can further denote its gradient as $\frac{\partial C(\theta)}{\partial \theta_l}$, a.k.a. ∂C. Note that, in this work, we use the term "optimization" to describe a process that searches optimal parameters for VQCs via numerical methods.

In 2018, McClean et al. [10] first investigated barren plateaus and revealed that under the assumption of the two-design Haar distribution, the variance of the gradient $Var[\partial C]$ in VQCs would exponentially decrease to zero during optimization as the model size, such as the number of qubits or layers, increases. This phenomenon is rooted in the assumption of the Haar distribution, under which $U(\theta)$ is considered sufficiently random in Hilbert space, making the gradient landscape increasingly flat as the size of VQCs grows. This randomness, a.k.a. Haar randomness, is characterized by the left- and right invariance of the Haar measure $\mu(U)$ on the unitary group $U(N)$, which ensures that the distribution of any function $f(U)$ remains unchanged under transformations by V as follows.

$$\int_{U(N)} d\mu(U) f(U) = \int d\mu(U) f(VU) = \int d\mu(U) f(UV). \tag{3}$$

The invariance presented in Eq. 3 plays a crucial role in ensuring the uniformity of random unitaries. To achieve such invariance, however, VQCs require large computational resources, which is impractical in both simulation and real-world scenarios. To mitigate this drawback, researchers often leverage unitary t-designs [12] for restricted

classes of $f(U)$, which significantly reduce computational demands while approximating the properties of the Haar measure. Formally, suppose that we sample unitary V_i with probability p_i, then a unitary ensemble $\{p_i, V_i\}$ is a t-design if it satisfies the following condition:

$$\sum_i p_i V_i^{\otimes t} \rho (V_i^\dagger)^{\otimes t} = \int \mathrm{d}\mu(U) U^{\otimes t} \rho (U^\dagger)^{\otimes t}, \qquad (4)$$

where ρ denotes the density matrix, and $\otimes$ denotes the tensor product.

Equation 4 ensures that a t-design can replicate the statistical properties of Haar randomness for functions $f(U)$ of degree t or lower. In practice, this means that if $f(U)$ is a polynomial with a maximum degree of t in unitary U, averaging $f(U)$ over the t-design $\{p_i, V_i\}$ approximates the result of averaging it over the unitary group w.r.t. the Haar measure [10].

In brief, while the Haar measure represents the theoretical upper bound of VQCs' expressivity, it is often unattainable in practice due to excessive computational resource demands. t-designs provide a practical alternative to enhance VQCs' expressivity by approximating Haar randomness with significantly reduced costs.

While t-designs mitigate the computational challenges associated with Haar randomness, they can still result in highly random unitary operators, particularly as the degree increases. This introduces a critical issue in VQCs' training known as barren plateaus (BPs). This issue poses significant challenges to the practical deployment of VQCs. To generally define the BPs, Qi et al. [11] provide a simple but generic definition as follows:

$$\mathrm{Var}[\partial C] \leq F(N), \qquad (5)$$

where $\mathrm{Var}[\partial C]$ denotes the variance of gradient, $F(N) \in o\left(\frac{1}{b^N}\right)$, for some $b > 1$ and, N denotes the number of qubits in VQCs.

According to Eq. 5, $\mathrm{Var}[\partial C]$ will exponentially decrease to zero when the number of qubits N increases. In this case, most gradient-based approaches will fail to train VQCs. Therefore, it is critical to mitigate barren plateau issues to train VQCs robustly. Notably, in this survey, we aim to investigate the mitigation methods designed to avoid the occurrence of barren plateaus at the beginning of the training process. Most mitigation methods theoretically or empirically verify the effectiveness under certain conditions, such as specific ranges of qubits or layers. The results indicate that the proposed methods can avoid BPs under these circumstances but do not necessarily guarantee the elimination of BPs for all cases. Thus, we use the term "mitigate" to describe these efforts.

3 Investigation of barren plateaus

In this section, we will introduce several works that aim to investigate barren plateaus (BPs). First, we introduce some works that explore and understand the barren plateau

phenomena in Sect. 3.1. Furthermore, we present the related works that investigate BPs from the perspective of model architectures in Sect. 3.2 and optimization in Sect. 3.3. Last, we discuss several works that provide tools to evaluate the strategies for mitigating BPs in Sect. 3.4.

3.1 Understanding barren plateau phenomena

Since the BP phenomena are categorized, several studies continue to delve into whether any other conditions or reasons may trigger the same issue. In this section, we overview related works that offer an understanding of the barren plateaus.

To dive deeper into BPs, Wang et al. [13] further prove that the gradient will vanish exponentially under the consideration of local Pauli noise, which is quite different from the noise-free setting in [10]. Similarly, Singkanipa et al. [14] explore the effects of noise on VQCs, focusing on how non-unital noise contributes to the formation of barren plateaus and fixed points in the optimization landscape. Marrero et al. [15] show that excessive entanglement between the visible and hidden units in VQCs can hinder learning capacity. Holmes et al. [16] first investigate the scrambling processes and show that any variational ansatz is highly probable to have a barren plateau landscape. Later, Holmes et al. [17] study the expressibility of VQCs and reveal that highly expressive VQCs may exhibit flatter optimization landscapes. Similarly, Ragone et al. [18] provide a unified framework to study the expressibility of VQCs. Liu et al. [19] theoretically provide an understanding of the distinction between BPs and "laziness". In this study, laziness refers to the exponential suppression of variational angle updates during gradient descent. By studying laziness, researchers have a deeper understanding of the ansatz structure as a trade-off between flat landscapes and trainability, i.e., trainability in the presence of BPs via laziness). Friedrich et al. [20] examine how the dimensionality of qudits influences the occurrence of BPs in VQCs. Fontana et al. [21] introduce the concept of adjoint differentiation to analyze BPs and provide theoretical insights for designing quantum ansätze to avoid barren plateaus. Comparably, Diaz et al. [22] extend the theoretical understanding of BPs beyond the conventional dynamical Lie algebra (DLA) framework and propose new criteria to identify barren plateaus. Last, Anschuetz et al. [23] investigate other untrainable scenarios beyond barren plateaus.

3.2 Investigations from model perspectives

A portion of the research works aims to investigate BPs from the perspective of existing quantum circuits. Wiersema et al. [24] first examine a family of VQCs, named the Hamiltonian variational Ansatz (HVA), and reveal that HVA is absent from BPs due to its favorable structural properties. Mao et al. [25] theoretically investigate the existence of barren plateaus for alternated disentangled UCC (dUCC) ansatz. Besides, Liu et al. [26] provide critical insights into the trainability of tensor network models and reveal that barren plateaus are absent near the local minimum. This discovery has led to many investigations in new VQC architectures. For example, Martin et al. [27] investigate the BP issues in several randomly chosen parameterized circuits and reveal that some

model architectures like tree tensor network (qTTN) and multi-scale entanglement renormalization ansatz (qMERA) could avoid BPs. Cybulski et al. [28] investigate four scenarios about barren plateaus by examining the model depth, layer-by-layer pretraining, circuit block structure, and model creation without any constraints. Beyond evaluating VQCs, several works validate that BPs do not exhibit in classical-quantum hybrid models. Abbas et al. [29] empirically demonstrate that a class of quantum neural networks (QNNs) can outperform comparable feedforward networks and prove that certain QNNs can avoid BPs because of the unique optimization landscapes. Similarly, Pesah et al. [30] analyze the cost function of QCNNs and show that BPs do not exhibit in QCNNs, indicating the potential of QCNNs for advancing NISQ computing. Coelho et al. [31] explore the trainability and performance of VQCs on deep Q-learning models in classical environments and indicate that increasing the number of qubits does not lead to BPs for these models.

3.3 Investigations from optimization perspectives

Another portion of the research works aims to investigate BPs from the perspective of the optimization process. Zhao et al. [32] propose a scheme to analyze the barren plateau phenomenon in training VQCs with the ZX-calculus. Cerezo et al. [8] study the connections between locality and trainability of different cost functions in shallow VQCs and reveal that optimizing VQCs via cost functions with global observables exhibit BPs, whereas the gradient vanishes polynomially when using local observables. Concurrently, Cerezo et al. [33] verify that Hessian-based approaches do not circumvent the exponential scaling associated with BPs. Later, Cerezo et al. [34] also reveal the cost function can be classically simulated when the gradient variance vanishes. This result arises from a curse of dimensionality. Current methods essentially encode this problem into small and classically simulatable subspaces. Arrasmith et al. [35] empirically verify that the gradient-free optimizer does not solve the barren plateau problem. Furthermore, Arrasmith et al. [36] investigate the interdependent connections among three optimization landscapes, barren plateaus, exponential cost concentration around the gradient mean, and narrow gorges, during the training phase of VQCs. This work indicates that in high-dimensional space, BPs will cause cost concentration and thus reduce the training efficiency. Similarly, Miao et al. [37] evaluate BP issues and show that BPs would be addressed in the Riemannian formulation of such optimization problems. Furthermore, Miao et al. [38, 39] provide an in-depth analysis of isometric tensor network optimization methods applied to extensive Hamiltonians, i.e., VQCs with extensive size, and demonstrate that these approaches are inherently free of barren plateaus. Also, these works evaluate the gradient variance of tensor-network-based approaches and their scaling property. Pérez-Salinas et al. [40] investigate landscapes that contain relevant information causing optimization hardness, contributing connections between information content and the scale of the gradient. Nemkov et al. [41] investigate BPs in the presence of local minima and reveal that enlarging gradients cannot sufficiently address BPs under the above circumstance. From the theoretical perspectives in optimization, Uvarov et al. [9] provide a lower bound for training VQCs and argue that the magnitude of gradient variance can be heavily influenced by

Table 1 Summary of keywords in the toolbox for mitigating BPs

Authors	Keywords
Patti et al. [43]	Entanglement metrics, Meta-learning, Solution Factorization
Larocca et al. [44]	Quantum control, dynamical lie algebra, BP Diagnosis
Park et al. [45]	Code optimization, gradient visualization, QNN IoT systems
Kashif et al. [46]	Evaluate initialization-based strategies

the Hamiltonian and circuit structure. Letcher et al. [42] study the bounds of gradient descent in VQCs, providing tight upper and lower bounds during VQCs' training for different amounts of observables.

3.4 Evaluations

Several works provide frameworks or tools to evaluate the strategies that can mitigate BPs. In this section, we discuss several tools as follows: Patti et al. [43] provide a toolbox that integrates several techniques, such as initialization or regularization strategies, to mitigate BPs. This toolbox enables developers to experiment and tinker around with additional entanglement and noise in the circuit. Similarly, Larocca et al. [44] introduce a theoretical framework to diagnose the presence of BPs for problem-based ansatz such as QAOA and HVA. This research focuses on utilizing concepts from quantum optimal control (i.e., dynamical lie algebra) to predict whether BPs will arise in the training landscape of problem-based ansatz. Park et al. [45] introduce a new software engineering tool that aims at assisting developers in efficiently writing, debugging, and testing code related to BP experiments. This tool is a run-time testing, analysis, and code optimization (TACO) for QNNs in advanced IoT systems. Kashif et al. [46] analyze the impact of initialization-based strategies from classical machine learning in random VQCs from the aspect of the BP phenomenon. In brief, we summarize the keywords for these toolboxes, which can help researchers write code and conduct barren plateau experiments, in Table 1.

4 Mitigation of barren plateaus

In this section, we categorize most existing mitigation strategies into five groups from the perspective of initialization, optimization, model architecture, regularization, and measurement. Besides, we summarize and discuss the strategies for each group.

4.1 Initialization-based strategies

In this section, we explore various initialization-based strategies applied to VQCs. Within this category, most strategies aim to employ different initialization methods to help VQCs get rid of barren plateaus in the initialization stage. To better illustrate these strategies, we present a general idea in Fig. 2.
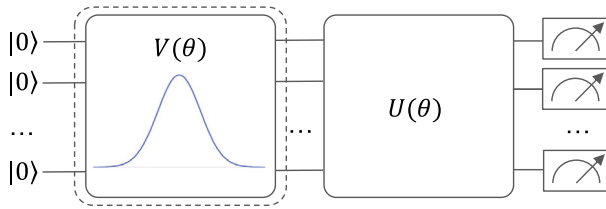
Fig. 2 A general structure of a VQC where the initialization-based strategy is circled in a dotted box. Generally speaking, a VQC will take the encoded quantum-state data as inputs and learn the data representation through a sequence of unitary operations. For each unitary gate, VQC has a corresponding parameter. In the beginning, initialization-based strategies $V(\theta)$ are applied to initialize parameters θ of the unitary quantum circuits $U(\theta)$ before training. In the end, a measurement layer is applied to measure the output of the VQC

In the following paragraph, we will briefly introduce each work within this category. Grant et al. [47] propose a strategy that first randomly initializes some gates and further sets the rest gates to produce the identity matrix, ensuring initial gradients non-vanishing. Sauvage et al. [48] introduce a flexible initialization strategy that can adapt suitable initial parameters to various sizes of quantum circuits. Sack et al. [49] design a method to circumvent weak barren plateaus by using classical shadows protocol at the initialization stage and during the training process. Rad et al. [50] introduce a fast-and-slow algorithm that employs the Bayesian Learning approach to initialize parameters for VQCs by initializing a fast local optimizer to find the global optimum point efficiently. In this algorithm, the Bayesian approach can help predict effective starting points, thereby reducing the risk of barren plateaus. Kulshrestha et al. [51] first initialize the model parameters using a beta distribution, which can better model the data distribution in binary classification, and further add noise into the parameters during training to avoid the optimization being trapped in saddle points. They empirically verify that utilizing beta distribution can effectively mitigate the rapid decrease of gradient variance in initialization as the model size increases. Similarly, Zhang et al. [52] employ a Gaussian distribution as initial model parameters and demonstrate that such a strategy can effectively help VQCs avoid BPs. Friedrich et al. [53] utilize a classical neural network, namely a convolutional neural network (CNN), to randomly generate the model parameters, indicating the effectiveness of this method for mitigating BPs. Mele et al. [54] offer an iterative search scheme to transfer smooth solutions from smaller systems as a warm start for larger systems. Such a scheme has been proven to help optimization avoid the barren plateau issue. Grimsley et al. [55] first examine how the Adaptive Problem-Tailored Variational Quantum Eiegensolver (ADAPT-VQE) ansätze is affected by local minima and further propose a strategy that can initialize parameters with over an order of magnitude smaller error than random initialization. Liu et al. [56] present a transfer-learning approach where model parameters learned from small-scale tasks are transferred to larger tasks. This method leverages the smooth solutions from simpler problems to provide a good starting point for more complex VQCs, therefore avoiding barren plateaus and enhancing training efficiency. Park et al. [57] first show that a time-evolution operator generated by a k-local Hamiltonian can help VQCs avoid BPs and then propose an initialization scheme using this operator. The k-local Hamiltonian can perform efficiently as it acts

	Authors	Methods
Table 2 Summary of initialization-based strategies	Grant et al. [47]	Identity block strategy
	Sauvage et al. [48]	Flexible strategy
	Sack et al. [49]	Classical shadows protocol
	Rad et al. [50]	Fast-and-slow algorithm
	Kulshrestha et al. [51]	Beta initialization
	Zhang et al. [52]	Gaussian initialization
	Friedrich et al. [53]	Initialization via CNNs
	Mele et al. [54]	Transfer from small to large
	Grimsley et al. [55]	ADAPT-VQE
	Liu et al. [56]	Transfer-learning method
	Park et al. [57]	A time-evolution operator

non-trivially on at most k qubits [58, 59], such as addressing the contains with finite correlation length in quantum spin lattice systems [60]. Some studies discussed in this survey also utilize the concept of k-local Hamiltonian [49, 61].

Overall, the above initialization-based strategies can be summarized in Table 2.

4.2 Optimization-based strategies

In this section, we go over the optimization-based strategies that mainly claim to address BPs in optimization, which is an essential step for optimizing VQCs [62]. To better visualize the optimization-based strategies, we present a general layout in Fig. 3. Note that many optimization-based strategies can enhance the trainability of VQCs but some of them only solve the gradient vanishing problem during training rather than focusing on BPs. To follow the topic of this survey, we don't discuss these papers in this section.

Ostazewski et al. [63] propose a gradient-free algorithm, *Rotoselect*, to enhance the optimization process of VQCs by sequentially adjusting angles of rotation and circuit structure to efficiently minimize the energy. This algorithm can significantly enhance the performance of shallow circuits compared to those relying solely on parameter updates, rendering this method particularly advantageous for noisy intermediate-scale quantum computers. On the contrary, most works alleviate BPs via gradient-based approaches. Skolik et al. [64] propose a layer-wise training strategy that can incrementally grow the depth of VQCs during optimization and only update a subset of model parameters in each training epoch. This work also verifies the effectiveness of addressing BPs. Similarly, Gharibyan et al. [65] introduce a new hierarchical framework that can incrementally learn at different scales of VQCs. This framework starts from a block with a small number of qubits and incrementally expands the blocks with a larger number of qubits. Such a hierarchical design can not only ensure the performance of VQCs but also avoid BPs. The above two methods commonly grow the model size during training, but the main difference between them is that the layer-wise learning method [64] stacks single-rotation gates and two-qubit gates to form all-to-all
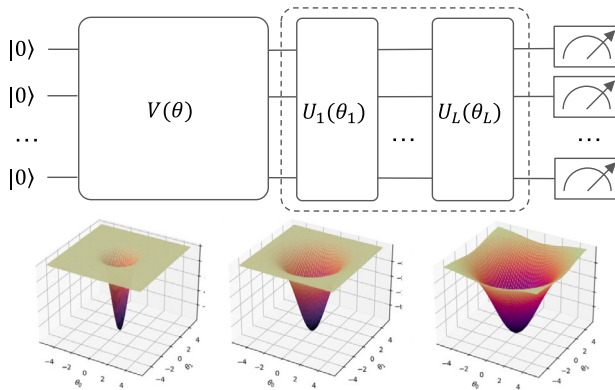
Fig. 3 A general structure of VQCs where the optimization-based strategies are usually applied in the training of $U(\theta)$ (dotted box), which is decomposed into L layer from $U_1(\theta_1)$ to $U_L(\theta_L)$ (upper side). On the lower side, we present the process for addressing BPs. The left-most cost-function landscape represents BPs, a flat landscape with no discernible slope toward the minimum. In this case, many optimization approaches may be initially trapped in the flat landscape, leading to a failure of training. By employing optimization-based strategies, the cost-function landscape could be gradually recovered, as shown in the middle and right-most landscapes

connected layers successively, whereas the hierarchical learning method [65] expands the blocks with a larger number of qubits. From the perspective of parameters reduction, compared to layer-wise learning [64], which updates a subset of parameters iteratively, Liu et al. [66] propose a state-efficient ansatz (SEA) to accurately estimate the ground state with much fewer parameters than common VQCs while efficiently mitigating barren plateaus and improving the trainability of VQCs.

Furthermore, several approaches attempt to mitigate BPs via new proposed mechanisms in optimization. Tobias Haug and Kim [67] propose methods to enhance the training process of VQCs by leveraging Gaussian kernels to derive adaptive learning rates during optimization and by introducing the generalized quantum natural gradient. The proposed methods ensure the stability of training and reduce the risk of barren plateaus. Kieferova et al. [68] examine the assumptions of BPs and indicate that the issues can be circumvented by an unbounded loss function. Based on this examination, they propose a training algorithm that minimizes the maximal Rényi divergence of order two. They further verify the effectiveness of the proposed algorithm on quantum neural networks (QNNs) and quantum Boltzmann machines (QBM) across two cases, thermal state learning and Hamiltonian learning. Heyraud et al. [69] provide an efficient method to compute the gradient of the cost function and its variance for a wide class of VQCs by mapping from randomly initialized circuits to a set of Clifford circuits, which can be simulated on classical devices. The method can not only mitigate BPs but also increase the scalability of VQCs. Sannia et al. [70] aim to mitigate barren plateaus and enhance the trainability of VQCs via engineered Markovian dissipation processes. To achieve this goal, the proposed method artificially manufactures additional noise after each gate in the VQC. The experiments show that the added noise can help restore the trainability of a VQC by mitigating the exponential scaling of the gradient's variance. Mele et al. [71] examine the role of noise in creating shallow

Table 3 Summary of optimization-based strategies

Authors	Methods
Ostaszewski et al. [63]	A gradient-free algorithm, "Rotoselect"
Skolik et al. [64]	Layer-wise learning method
Gharibyan et al. [65]	Hierarchical variational circuit framework
Liu et al. [66]	State-efficient ansatz
Haug et al. [67]	Adaptively adjust learning rates
Kieferova et al. [68]	Rényi 2-divergence
Heyraud et al. [69]	Map randomly circuits to Clifford circuits
Sannia et al. [70]	Engineered Markovian dissipation processes
Mele et al. [71]	Noise-induced shallow circuits
Broers et al. [72]	Introduce a Fourier-based ansatz
Zambrano et al. [73]	Evaluate geometric entanglement
Sciorilli et al. [74]	Pauli-correlations encoding
Falla et al. [75]	Graph embedding

quantum circuits that do not exhibit barren plateaus for cost functions composed of local observables. Broers et al. [72] propose an optimization approach based on trainable Fourier coefficients of Hamiltonian system parameters and demonstrate that this approach can mitigate BPs. Zambrano et al. [73] address BP issues in the variational quantum computation of geometric entanglement and present an approach to maintain gradients throughout the optimization process of VQCs.

Besides improving the trainability of VQCs, another group of optimization-based methods addresses BPs from the perspective of data encoding. We categorize them in this section as optimization is the main contribution of these methods while their proposed encoding mechanisms can address BPs. Sciorilli et al. [74] introduce a hybrid quantum-classic solver for combinatorial optimizations and further analyze that employing specific qubit-efficient encoding methods can mitigate barren plateaus. Also, this work verifies the effectiveness of the proposed methods on max-cut problems. Falla et al. [75] first investigate several graph embedding techniques for parameter transferability, such as the transferability between different classes of max-cut instances and further effectively mitigate BPs during variational optimization. Although this is a naive approach, it effectively reduces the number of iterations in optimization, obtaining a speedup for parameter convergence.

Overall, the above optimization-based strategies are summarized in Table 3.

4.3 Model-based strategies

In this section, we explore several model-based strategies that propose new model architectures to address BP issues over the past few years [61, 76–78]. These methods encompass a wide variety of new circuit structures and new gradient-based or gradient-free frameworks. An overall idea of model-based strategies is presented in Fig. 4.
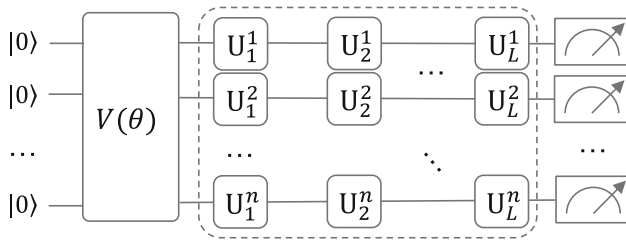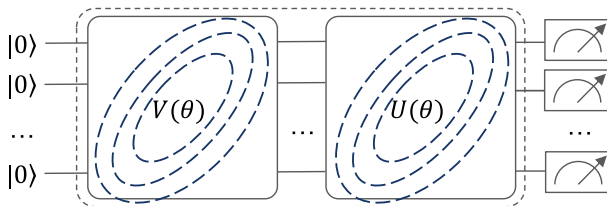
Fig. 4 An overall idea of model-based strategies that are usually applied to the decomposed unitary circuit $U(\theta) \in \mathbf{C}^{n \cdot L}$ (dotted box), where U_L^n denotes that $U(\theta)$ consists of n qubits and L layers

Within this category, Li et al. [79] first propose a new hybrid quantum-classical framework, Variational Shadow Quantum Learning (VSQL), that utilizes the classical shadows of quantum data, a kind of side information of quantum data, in supervised learning. This framework can not only reduce noise but also significantly avoid BP issues. Bharti et al. [80] pioneered a groundbreaking approach to quantum simulation by introducing a novel quantum-classical algorithm. Unlike most existing methods, their algorithm operates without relying on any classical-quantum feedback loop during training. Notably, their work demonstrates to be free from BP problems, marking a significant advancement in simulating the dynamics of quantum systems. Du et al. [81] introduce a new search scheme called quantum architecture search (QAS), aimed at identifying approximately optimal ansatz structures by assessing the expressivity of VQCs. This innovative approach enables QAS to dynamically generate ansatz structures tailored to specific tasks, offering a flexible solution to address challenges such as mitigating the negative effects of quantum noise and bypassing BP issues. Zhang et al. [82] introduce a gradient-free framework, Quark, that optimizes the models via quantum optimization. This framework can avoid BPs and frequent classical-quantum interactions because it does not compute the gradient during model optimization. Selvarajan et al. [83] design a variational-autoencoder-based circuit to reduce states represented in large Hilbert spaces with half the number of qubits while retaining the features of the starting state. This reduction can potentially help improve training efficiency on higher-dimensional Hilbert spaces and thus overcome BP issues. Tuysuz et al. [84] propose a classical splitting ansatz that decomposes an n-qubit circuit into circuits of size $O(\log N)$. This ansatz has been shown to mitigate BPs with the demonstration of their absence in tasks such as binary classification. Inspired by classical residual neural networks, Kashif et al. [85] introduce residual connections in VQC. This proposed model, a.k.a., resQNets, breaks down a VQC structure into smaller nodes and introduces residual connections between these nodes. The experimental results show that resQNets can outperform conventional VQCs and offer a new model architecture to solve BP issues. Shin et al. [86] propose a unified methodology, layer-wise quantum convolutional neural networks (LQCNNs), to avoid BPs and to provide a solution for large qubit states. This study rigorously proves that LQCNNs preserve fundamental quantum properties, such as different entropies and fidelity of quantum states, despite their deep structure. Zhang et al. [87] theoretically analyze lower bounds on gradient variances of VQCs and propose a unified framework, finite local-depth

Table 4 Summary of model-based strategies

Authors	Methods
Li et al. [79]	Utilizes side information of quantum data
Bharti et al. [80]	Training without a feedback loop
Du et al. [81]	Dynamically generate ansatz structures
Zhang et al. [82]	Introduce a gradient-free framework
Selvarajan et al. [83]	Design a variational-autoencoder-based circuit
Tuysuz et al. [84]	Propose a splitting ansatz to decompose circuit
Kashif et al. [85]	Introduce residual connections in VQC
Shin et al. [86]	Propose a unified layer-wise QNN
Zhang et al. [87]	Propose a unified framework, FLDC

Fig. 5 An overall idea of regularization-based strategies. This strategy could be applied to the model parameters during initialization or training. We use dotted circles to represent the regularization in this figure

circuits (FLDC), to mitigate BPs. In brief, we summarize the above model-based strategies in Table 4.

4.4 Regularization-based strategies

Regularization is a machine-learning technique to constrain the search space for boosting optimization performance. In other words, regularization achieves its goal by penalizing complex model parameters θ, effectively generalizing the model to unseen data. In general, regularization can be formulated as follows.

$$\min_{\theta}[C(\theta) + \lambda Reg(\theta)], \tag{6}$$

where $Reg(\cdot)$ denotes the regularization method and λ denotes its coefficient.

In this survey, regularization-based strategies aim to apply regularization to VQCs. The regularization could be applied to the model parameters in the initialization stage or the optimization stage. We present the overall idea of this strategy in Fig. 5.

Patti et al. [43] utilize regularization to penalize entanglement and show that this method could ameliorate BPs and further decrease both training time and error. Zhuang et al. [88] propose regularization methods that can be applied during initialization and optimization. Specifically, the proposed methods regularize the initial distribution of
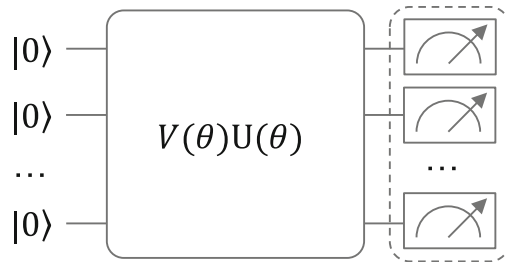
Fig. 6 An example of measurement-based strategy, which is applied to the measurement of the VQC $V(\theta)U(\theta)$, where $V(\theta)$ denotes the initialization module and $U(\theta)$ denotes the unitary circuit. The measurement step (shown within the dotted box) typically focuses on measuring the output of the VQCs

model parameters with prior knowledge of the train data and iteratively add diffused Gaussian noise to the parameters during training. This is the first work that leverages regularization strategies to simultaneously address barren plateaus and saddle points in optimizing VQCs.

4.5 Measurement-based strategies

Measurement techniques aim to provide more informative metrics about the qubits in the system during and after the optimization process. By obtaining richer information about the quantum data, researchers may gain a deeper understanding of what their data is and structure the VQC with the data accordingly. This allows for more efficient training and higher performance in VQCs. In Fig. 6, we show an example of a measurement-based strategy.

One measurement-based strategy introduced by Rappaport et al. [89] demonstrates that entanglement-induced BPs could be suppressed by adding measurements with post-selection. This work empirically studies two measures, measurement-induced landscape transition (MILT) and measurement-induced phase transition (MIPT), and further reveals the benefits of using MILT for optimization.

5 Discussion

In this section, we first present the summary of mitigation strategies in a table. Second, we compare our work with the concurrent survey papers. Third, we highlight the similarity and distinction between gradient issues and barren plateaus. Last, we provide insightful discussion on the future directions.

5.1 Summary of mitigation strategies

In Table 6, we summarize several types of mitigation methods for barren plateaus. The category is proposed from the perspective of initialization, optimization, model architecture, regularization, and measurement. In the column title of the table, the

attributes "$\|N\|$" and "$\|L\|$" denote the maximum number of qubits and layers in the corresponding paper. "Problem Type" denotes the problem that the work mainly focuses on. Note that in most works, authors will examine different numbers of qubits or layers, whereas we only present the maximum numbers mentioned in the paper.

In the table, we have several observations as follows. First, most existing methods aim to mitigate barren plateaus from the perspective of initialization, optimization, and model architecture, whereas regularization and measurement are two relatively new directions for mitigation. Second, in the initialization-base strategies, many works examine new data encoding methods, indicating that data encoding is a crucial component for mitigation. Third, ground state is a vital problem in physics so many researchers are interested in it. Fourth, classification is an important task for the application of VQCs, whereas most works use phase encoding in this task.

5.2 Comparison with concurrent survey papers about BPs

In this subsection, we compare three concurrent survey papers about BPs and discuss their pros and cons. Also, we summarize their key contributions in Table 5.

Qi et al. [11] provide an analysis of the barren plateau phenomenon, which significantly affects the training efficiency of quantum neural networks (QNNs). The review categorizes various causes of barren plateaus, including cost function design, circuit depth, and parameter initialization. It also discusses current strategies to mitigate this issue, such as circuit architecture modifications, and highlights emerging trends aimed at overcoming these challenges to improve the scalability and performance of QNNs. This survey is the first publication that introduces related works about BP issues in QNNs but it only mentions limited related work.

Larocca et al. [90] conduct an in-depth review of barren plateaus in variational quantum algorithms (VQAs). This review discusses how various factors such as ansatz design, initial states, measurement observables, loss functions, and hardware noise contribute to the occurrence of barren plateaus. It also highlights current theoretical and heuristic strategies to mitigate these issues. The study of BPs is essential for improving the trainability of VQAs and has implications for related fields like quantum optimal control, tensor networks, and learning theory. This work provides a comprehensive review and larger scope of BPs, including the trainability of VQAs and other gradient issues. However, it doesn't introduce the mitigation strategies in detail.

Gelman et al. [91] explore various quantum circuits and types of BPs that hinder the optimization of parameterized quantum circuits (PQCs). Specifically, this survey covers the basic concept of PQCs and further describes BPs. It also introduces various types of quantum ansatz and metrics. Furthermore, this paper discusses several reasons that may cause BPs and corresponding mitigation methods. Basically, this work focuses more on the reasons and types of BPs rather than the mitigation methods.

Compared with the above three concurrent survey papers, our work provides a new taxonomy to categorize the existing mitigation strategies for BPs and introduce them in detail. We summarize the above three works in Table 5 to highlight their key contributions.

Table 5 Comparison with concurrent survey papers about BPs

Authors	Key contributions
Qi et al. [11]	Introduce the origin, solutions, and trends of BPs in QNNs
Larocca et al. [90]	Introduce the frameworks of VQAs and broader topics about BPs
Gelman et al. [91]	Introduce various types of PQCs and discuss the types of BPs with corresponding mitigation methods

5.3 Gradient issues vs barren plateaus

Gradient issues are commonly seen during the optimization process of utilizing gradient-based methods. In classical models, common gradient issues include gradient vanishing, local minima, saddle points, etc. Gradient vanishing occurs when the gradient diminishes as it backpropagates through the layers of a neural network, leading to negligible updates in the early layers during training. Local minima refer to points in the loss landscape where the model is stuck in a state that is not the global minimum but cannot escape due to the nature of the gradient descent. Besides local minima, training sometimes gets stuck in a saddle point where the gradient is zero along certain directions, making it difficult for the optimization to proceed efficiently.

In contrast, barren plateaus (BPs) represent a specific type of gradient issue particularly associated with VQCs. Unlike classical gradient issues, BPs occur when the variance of the gradients diminishes rapidly as the number of qubits or layers increases. This rapid decay results in a flat loss landscape, where the gradients are so small that the optimization process becomes effectively stalled right from the beginning of training. Thus, BPs can significantly hinder the training of VQCs.

In this review, we focus on the BP issues occurring in VQCs, examining how this issue arises and its mitigation strategies for the training of VQCs.

5.4 Future directions

Based on the above discussion, in this survey, we present several promising research directions to mitigate barren plateaus in VQCs as follows.

- *AI-driven Initialization.* Most initialization-base strategies utilize classical neural networks, or machine-learning algorithms to generate initial distribution. In the future, AI-driven methods, such as LLMs, could be a promising tool to adaptively initialize VQC's parameters.
- *New Optimization Methods.* Several directions in optimization are worth exploring, such as employing local cost functions to manipulate a smaller subset of qubits to avoid a flat landscape and improving layer-wise training methods to incrementally avoid BPs in deep VQCs.

Table 6 Summary of mitigation strategies for barren plateau issues

Authors	Data encoding	$\|N\|$	$\|L\|$	Problem type
Initialization-based strategies				
Grant et al. [47]	Amplitude	12	120	Ground State
Sauvage et al. [48]	FLIP encode-decode	16	16	MaxCut, Ground State
Sack et al. [49]	×	16	100	Ground state
Rad et al. [50]	×	6	2	Algorithm Evaluation
Kulshrestha et al. [51]	Phase	10	30	Classification
Zhang et al. [52]	Basis	15	72	Ground State
Friedrich et al. [53]	Qubit encode	10	30	Model Assessment
Mele et al. [54]	Symmetry encode	20	100	Ground State
Grimsley et al. [55]	×	14	×	Ground State
Liu et al. [56]	Block identity encode	4	4	Ground State
Park et al. [57]	Local Hamiltonian	28	×	Thermalization
Optimization-based strategies				
Ostaszewski et al. [63]	×	7	15	Ground State
Skolik et al. [64]	Phase	18	200	Classification
Haug et al. [67]	×	25	20	Ground State
Kieferova et al. [68]	×	×	×	Training Assessment
Heyraud et al. [69]	×	100	10	Ground State
Gharibyan et al. [65]	Amplitude	27	9	Model Assessment
Sannia et al. [70]	×	10	20	Ground State

Table 6 continued

Authors	Data encoding	$\|N\|$	$\|L\|$	Problem type
Mele et al. [71]	×	6	10	×
Broers et al. [72]	×	8	×	Ground State
Zambrano et al. [73]	×	18	×	Ground State
Liu et al. [66]	×	18	2	Ground state
Sciorilli et al. [74]	Pauli-correlations	17	11	MaxCut
Falla et al. [75]	Graph embedding	27	3	MaxCut
Model-based strategies				
Li et al. [79]	Phase	10	5	Classification
Bharti et al. [80]	×	12	200	Ground State
Du et al. [81]	Phase	10	3	Ground State, Classification
Zhang et al. [82]	Basis	4	2	Edge Detection, Classification
Selvarajan et al. [83]	Amplitude	8	5	Classification
Tuysuz et al. [84]	Phase	24	200	Classification
Kashif et al. [85]	×	20	6	Analyze Loss Landscapes
Shin et al. [86]	×	×	×	×
Zhang et al. [87]	×	12	16	Ground State
Regularization-based strategies				
Patti et al. [43]	×	9	200	Ground State
Zhuang et al. [88]	Phase	16	50	Classification
Measurement-based Strategies				
Rappaport et al. [89]	×	18	100	Ground State

"$\|N\|$" and "$\|L\|$" denote the maximum number of qubits and layers in the corresponding paper. "Problem Type" denotes the problem that the work mainly focuses on

- *Novel Circuit Architecture.* It is promising to explore new quantum circuit architectures that integrate neural networks to address BPs and further improve the trainability of VQCs.
- *Noise-based Techniques.* Noise could induce BPs but also be used for mitigation. For example, noise can be injected into model parameters to avoid a flat landscape. In the meanwhile, these techniques bring us an open question: Will noise-based techniques weaken the optimization of VQCs?
- *Graph Embedding.* Besides classic data encoding techniques, embedding data with graph representation could be a potential solution to mitigate BPs.

In brief, the above future directions offer researchers new perspectives to address the challenge of barren plateaus in VQCs. These directions pave the effective way for mitigating BPs and further enhance trainability of VQCs.

6 Conclusion

In this survey, we conduct a comprehensive overview of related literature about the barren plateaus (BPs) in variational quantum circuits (VQCs) and propose a new taxonomy from the perspective of investigation and mitigation. Especially, we categorize most existing mitigation strategies into five groups from the point of initialization, optimization, model architecture, regularization, and measurement. We introduce these strategies in detail and summarize them in a table. Furthermore, in the discussion section, we compare our work with concurrent survey papers and highlight the similarity and distinction between gradient issues and barren plateaus. Last, we provide insightful discussion on the future directions. We hope that our survey paper can provide more inspiration for researchers who are interested in BP issues.

Declarations

References

1. Preskill, J.: Quantum computing in the NISQ era and beyond. Quantum **2**, 79 (2018)
2. Bauer, B., Bravyi, S., Motta, M., Chan, G.K.-L.: Quantum algorithms for quantum chemistry and quantum materials science. Chem. Rev. **120**(22), 12685–12717 (2020)
3. Liang, Z., Wang, H., Cheng, J., Ding, Y., Ren, H., Gao, Z., Hu, Z., Boning, D.S., Qian, X., Han, S., : Variational quantum pulse learning. In: 2022 IEEE International Conference on Quantum Computing and Engineering (QCE), pp. 556–565, IEEE (2022)
4. Thanasilp, S., Wang, S., Nghiem, N.A., Coles, P., Cerezo, M.: Subtleties in the trainability of quantum machine learning models. Quantum Mach. Intell. **5**(1), 21 (2023)
5. Zhang, B., Xu, P., Chen, X., Zhuang, Q.: Generative quantum machine learning via denoising diffusion probabilistic models. Phys. Rev. Lett. **132**(10), 100602 (2024)
6. Chiribella, G., D'Ariano, G.M., Perinotti, P.: Quantum circuit architecture. Phys. Rev. Lett. **101**(6), 060401 (2008)
7. Liang, Z., Cheng, J., Yang, R., Ren, H., Song, Z., Wu, D., Qian, X., Li, T., Shi, Y.: Unleashing the potential of LLMs for quantum computing: A study in quantum architecture design. arXiv preprint arXiv:2307.08191 (2023)
8. 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), 1791 (2021)
9. Uvarov, A., Biamonte, J.D.: On barren plateaus and cost function locality in variational quantum algorithms. J. Phys. A: Math. Theor. **54**(24), 245301 (2021)
10. McClean, J.R., Boixo, S., Smelyanskiy, V.N., Babbush, R., Neven, H.: Barren plateaus in quantum neural network training landscapes. Nat. Commun. **9**(1), 4812 (2018)
11. Qi, H., Wang, L., Zhu, H., Gani, A., Gong, C.: The barren plateaus of quantum neural networks: review, taxonomy and trends. Quantum Inf. Process. **22**(12), 435 (2023)
12. Renes, J.M., Blume-Kohout, R., Scott, A.J., Caves, C.M.: Symmetric informationally complete quantum measurements. J. Math. Phys. **45**(6), 2171–2180 (2004)
13. Wang, S., Fontana, E., Sharma, K., Sone, A., Cincio, L., Coles, P.J.: Noise-induced barren plateaus in variational quantum algorithms. Nat. Commun. **12**(1), 6961 (2021)
14. Singkanipa, P., Lidar, D.A.: Beyond unital noise in variational quantum algorithms: noise-induced barren plateaus and fixed points. arXiv preprint arXiv:2402.08721 (2024)
15. Marrero, C.O., Kieferová, M., Wiebe, N.: Entanglement-induced barren plateaus. PRX Quantum **2**(4), 040316 (2021)
16. Holmes, Z., Arrasmith, A., Yan, B., Coles, P.J., Albrecht, A., Sornborger, A.T.: Barren plateaus preclude learning scramblers. Phys. Rev. Lett. **126**(19), 190501 (2021)
17. Holmes, Z., Sharma, K., Cerezo, M., Coles, P.J.: Connecting ansatz expressibility to gradient magnitudes and barren plateaus. PRX Quantum **3**(1), 010313 (2022)
18. Ragone, M., Bakalov, B.N., Sauvage, F., Kemper, A.F., Marrero, C.O., Larocca, M., Cerezo, M.: A unified theory of barren plateaus for deep parametrized quantum circuits. arXiv preprint arXiv:2309.09342 (2023)
19. Liu, J., Lin, Z., Jiang, L.: Laziness, barren plateau, and noises in machine learning. Mach. Learn.: Sci. Technol. **5**(1), 015058 (2024)
20. Friedrich, L., Farias, T.d.S., Maziero, J.: Barren plateaus induced by the dimension of qudits. arXiv preprint arXiv:2405.08190 (2024)
21. Fontana, E., Herman, D., Chakrabarti, S., Kumar, N., Yalovetzky, R., Heredge, J., Sureshbabu, S.H., Pistoia, M.: The adjoint is all you need: characterizing barren plateaus in quantum ansätze. Bulletin of the American Physical Society (2024)
22. Diaz, N., 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)
23. Anschuetz, E.R., Kiani, B.T.: Quantum variational algorithms are swamped with traps. Nat. Commun. **13**(1), 7760 (2022)
24. Wiersema, R., Zhou, C., Sereville, Y., Carrasquilla, J.F., Kim, Y.B., Yuen, H.: Exploring entanglement and optimization within the Hamiltonian variational ansatz. PRX Quantum **1**(2), 020319 (2020)
25. Mao, R., Tian, G., Sun, X.: Barren plateaus of alternated disentangled UCC ansatzs. arXiv preprint arXiv:2312.08105 (2023)
26. Liu, Z., Yu, L.-W., Duan, L.-M., Deng, D.-L.: Presence and absence of barren plateaus in tensor-network based machine learning. Phys. Rev. Lett. **129**(27), 270501 (2022)

27. Martín, E.C., Plekhanov, K., Lubasch, M.: Barren plateaus in quantum tensor network optimization. Quantum **7**, 974 (2023)
28. Cybulski, J.L., Nguyen, T.: Impact of barren plateaus countermeasures on the quantum neural network capacity to learn. Quantum Inf. Process. **22**(12), 442 (2023)
29. Abbas, A., Sutter, D., Zoufal, C., Lucchi, A., Figalli, A., Woerner, S.: The power of quantum neural networks. Nat. Comput. Sci. **1**(6), 403–409 (2021)
30. Pesah, A., Cerezo, M., Wang, S., Volkoff, T., Sornborger, A.T., Coles, P.J.: Absence of barren plateaus in quantum convolutional neural networks. Phys. Rev. X **11**(4), 041011 (2021)
31. Coelho, R., Sequeira, A., Santos, L.P.: Vqc-based reinforcement learning with data re-uploading: Performance and trainability. arXiv preprint arXiv:2401.11555 (2024)
32. Zhao, C., Gao, X.-S.: Analyzing the barren plateau phenomenon in training quantum neural networks with the zx-calculus. Quantum **5**, 466 (2021)
33. Cerezo, M., Coles, P.J.: Higher order derivatives of quantum neural networks with barren plateaus. Quantum Sci. Technol. **6**(3), 035006 (2021)
34. Cerezo, M., Larocca, M., García-Martín, D., Diaz, N., Braccia, P., Fontana, E., Rudolph, M.S., Bermejo, P., Ijaz, A., Thanasilp, S., et al.: Does provable absence of barren plateaus imply classical simulability? or, why we need to rethink variational quantum computing. arXiv preprint arXiv:2312.09121 (2023)
35. Arrasmith, A., Cerezo, M., Czarnik, P., Cincio, L., Coles, P.J.: Effect of barren plateaus on gradient-free optimization. Quantum **5**, 558 (2021)
36. Arrasmith, A., Holmes, Z., Cerezo, M., Coles, P.J.: Equivalence of quantum barren plateaus to cost concentration and narrow gorges. Quantum Sci. Technol. **7**(4), 045015 (2022)
37. Miao, Q., Barthel, T.: Equivalence of cost concentration and gradient vanishing for quantum circuits: an elementary proof in the Riemannian formulation. arXiv preprint arXiv:2402.07883 (2024)
38. Barthel, T., Miao, Q.: Absence of barren plateaus and scaling of gradients in the energy optimization of isometric tensor network states. arXiv preprint arXiv:2304.00161 (2023)
39. Miao, Q., Barthel, T.: Isometric tensor network optimization for extensive Hamiltonians is free of barren plateaus. Phys. Rev. A **109**(5), 050402 (2024)
40. Pérez-Salinas, A., Wang, H., Bonet-Monroig, X.: Analyzing variational quantum landscapes with information content. NPJ Quantum Inf. **10**(1), 27 (2024)
41. Nemkov, N.A., Kiktenko, E.O., Fedorov, A.K.: Barren plateaus are swamped with traps. arXiv preprint arXiv:2405.05332 (2024)
42. Letcher, A., Woerner, S., Zoufal, C.: From tight gradient bounds for parameterized quantum circuits to the absence of barren plateaus in QGANs. arXiv preprint arXiv:2309.12681 (2023)
43. Patti, T.L., Najafi, K., Gao, X., Yelin, S.F.: Entanglement devised barren plateau mitigation. Phys. Rev. Res. **3**(3), 033090 (2021)
44. Larocca, M., Czarnik, P., Sharma, K., Muraleedharan, G., Coles, P.J., Cerezo, M.: Diagnosing barren plateaus with tools from quantum optimal control. Quantum **6**, 824 (2022)
45. Park, S., Kim, J.: Quantum neural network software testing, analysis, and code optimization for advanced IoT systems: Design, implementation, and visualization. arXiv preprint arXiv:2401.10914 (2024)
46. Kashif, M., Rashid, M., Al-Kuwari, S., Shafique, M.: Alleviating barren plateaus in parameterized quantum machine learning circuits: investigating advanced parameter initialization strategies. In: 2024 Design, Automation & Test in Europe Conference & Exhibition (DATE), pp. 1–6, IEEE (2024)
47. Grant, E., Wossnig, L., Ostaszewski, M., Benedetti, M.: An initialization strategy for addressing barren plateaus in parametrized quantum circuits. Quantum **3**, 214 (2019)
48. Sauvage, F., Sim, S., Kunitsa, A.A., Simon, W.A., Mauri, M., Perdomo-Ortiz, A.: Flip: a flexible initializer for arbitrarily-sized parametrized quantum circuits. arXiv preprint arXiv:2103.08572 (2021)
49. Sack, S.H., Medina, R.A., Michailidis, A.A., Kueng, R., Serbyn, M.: Avoiding barren plateaus using classical shadows. PRX Quantum **3**(2), 020365 (2022)
50. Rad, A., Seif, A., Linke, N.M.: Surviving the barren plateau in variational quantum circuits with bayesian learning initialization. arXiv preprint arXiv:2203.02464 (2022)
51. Kulshrestha, A., Safro, I.: Beinit: Avoiding barren plateaus in variational quantum algorithms. In: 2022 IEEE International Conference on Quantum Computing and Engineering (QCE), pp. 197–203, IEEE (2022)
52. Zhang, K., Liu, L., Hsieh, M.-H., Tao, D.: Escaping from the barren plateau via gaussian initializations in deep variational quantum circuits. Adv. Neural. Inf. Process. Syst. **35**, 18612–18627 (2022)

53. Friedrich, L., Maziero, J.: Avoiding barren plateaus with classical deep neural networks. Phys. Rev. A **106**(4), 042433 (2022)
54. Mele, A.A., Mbeng, G.B., Santoro, G.E., Collura, M., Torta, P.: Avoiding barren plateaus via transferability of smooth solutions in a Hamiltonian variational ansatz. Phys. Rev. A **106**(6), 060401 (2022)
55. Grimsley, H.R., Barron, G.S., Barnes, E., Economou, S.E., Mayhall, N.J.: Adaptive, problem-tailored variational quantum eigensolver mitigates rough parameter landscapes and barren plateaus. NPJ Quantum Inf. **9**(1), 19 (2023)
56. Liu, H.-Y., Sun, T.-P., Wu, Y.-C., Han, Y.-J., Guo, G.-P.: Mitigating barren plateaus with transfer-learning-inspired parameter initializations. New J. Phys. **25**(1), 013039 (2023)
57. Park, C.-Y., Killoran, N.: Hamiltonian variational ansatz without barren plateaus. Quantum **8**, 1239 (2024)
58. Kempe, J., Kitaev, A., Regev, O.: The complexity of the local Hamiltonian problem. SIAM J. Comput. **35**(5), 1070–1097 (2006)
59. Cubitt, T., Montanaro, A.: Complexity classification of local Hamiltonian problems. SIAM J. Comput. **45**(2), 268–316 (2016)
60. Anshu, A.: Concentration bounds for quantum states with finite correlation length on quantum spin lattice systems. New J. Phys. **18**(8), 083011 (2016)
61. Park, C.-Y., Kang, M., Huh, J.: Hardware-efficient ansatz without barren plateaus in any depth. arXiv preprint arXiv:2403.04844 (2024)
62. Nádori, J., Morse, G., Majnay-Takács, Z., Zimborás, Z., Rakyta, P.: The promising path of evolutionary optimization to avoid barren plateaus. arXiv preprint arXiv:2402.05227 (2024)
63. Ostaszewski, M., Grant, E., Benedetti, M.: Structure optimization for parameterized quantum circuits. Quantum **5**, 391 (2021)
64. Skolik, A., McClean, J.R., Mohseni, M., Van Der Smagt, P., Leib, M.: Layerwise learning for quantum neural networks. Quantum Mach. Intell. **3**, 1–11 (2021)
65. Gharibyan, H., Su, V., Tepanyan, H.: Hierarchical learning for quantum ml: novel training technique for large-scale variational quantum circuits. arXiv preprint arXiv:2311.12929 (2023)
66. Liu, X., Liu, G., Zhang, H.-K., Huang, J., Wang, X.: Mitigating barren plateaus of variational quantum eigensolvers. IEEE Trans. Quantum Eng. (2024). https://doi.org/10.1109/TQE.2024.3383050
67. Haug, T., Kim, M.: Optimal training of variational quantum algorithms without barren plateaus. arXiv preprint arXiv:2104.14543 (2021)
68. Kieferova, M., Carlos, O.M., Wiebe, N.: Quantum generative training using r\'enyi divergences. arXiv preprint arXiv:2106.09567 (2021)
69. Heyraud, V., Li, Z., Donatella, K., Le Boité, A., Ciuti, C.: Efficient estimation of trainability for variational quantum circuits. PRX Quantum **4**(4), 040335 (2023)
70. Sannia, A., Tacchino, F., Tavernelli, I., Giorgi, G.L., Zambrini, R.: Engineered dissipation to mitigate barren plateaus. arXiv preprint arXiv:2310.15037 (2023)
71. Mele, A.A., Angrisani, A., Ghosh, S., Khatri, S., Eisert, J., França, D.S., Quek, Y.: Noise-induced shallow circuits and absence of barren plateaus. arXiv preprint arXiv:2403.13927 (2024)
72. Broers, L., Mathey, L.: Mitigated barren plateaus in the time-nonlocal optimization of analog quantum-algorithm protocols. Phys. Rev. Res. **6**(1), 013076 (2024)
73. Zambrano, L., Muñoz-Moller, A.D., Muñoz, M., Pereira, L., Delgado, A.: Avoiding barren plateaus in the variational determination of geometric entanglement. Quantum Sci. Technol. **9**(2), 025016 (2024)
74. Sciorilli, M., Borges, L., Patti, T.L., García-Martín, D., Camilo, G., Anandkumar, A., Aolita, L.: Towards large-scale quantum optimization solvers with few qubits. arXiv preprint arXiv:2401.09421 (2024)
75. Falla, J., Langfitt, Q., Alexeev, Y., Safro, I.: Graph representation learning for parameter transferability in quantum approximate optimization algorithm. arXiv preprint arXiv:2401.06655 (2024)
76. Yao, Y., Hasegawa, Y.: Avoiding barren plateaus with entanglement. arXiv preprint arXiv:2406.03748 (2024)
77. Friedrich, L., Maziero, J.: Quantum neural network with ensemble learning to mitigate barren plateaus and cost function concentration. arXiv preprint arXiv:2402.06026 (2024)
78. Shi, X., Shang, Y.: Avoiding barren plateaus via gaussian mixture model. arXiv preprint arXiv:2402.13501 (2024)
79. Li, G., Song, Z., Wang, X.: Vsql: Variational shadow quantum learning for classification. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 8357–8365 (2021)
80. Bharti, K., Haug, T.: Quantum-assisted simulator. Phys. Rev. A **104**(4), 042418 (2021)

81. Du, Y., Huang, T., You, S., Hsieh, M.-H., Tao, D.: Quantum circuit architecture search for variational quantum algorithms. NPJ Quantum Inf. **8**(1), 62 (2022)
82. Zhang, Z., Chen, Z., Huang, H., Jia, Z.: Quark: a gradient-free quantum learning framework for classification tasks. arXiv preprint arXiv:2210.01311 (2022)
83. Selvarajan, R., Sajjan, M., Humble, T.S., Kais, S.: Dimensionality reduction with variational encoders based on subsystem purification. Mathematics **11**(22), 4678 (2023)
84. Tüysüz, C., Clemente, G., Crippa, A., Hartung, T., Kühn, S., Jansen, K.: Classical splitting of parametrized quantum circuits. Quantum Mach. Intell. **5**(2), 34 (2023)
85. Kashif, M., Al-Kuwari, S.: Resqnets: a residual approach for mitigating barren plateaus in quantum neural networks. EPJ Quantum Technol. **11**(1), 4 (2024)
86. Shin, M., Lee, S., Lee, M., Ji, D., Yeo, H., Lee, H.J., Jeong, K.: Layerwise quantum convolutional neural networks provide a unified way for estimating fundamental properties of quantum information theory. arXiv preprint arXiv:2401.07716 (2024)
87. 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**(15), 150603 (2024)
88. Zhuang, J., Cunningham, J., Guan, C.: Improving trainability of variational quantum circuits via regularization strategies. arXiv preprint arXiv:2405.01606 (2024)
89. Rappaport, S., Gyawali, G., Sereno, T., Lawler, M.J.: Measurement-induced landscape transitions in hybrid variational quantum circuits. arXiv preprint arXiv:2312.09135 (2023)
90. Larocca, M., Thanasilp, S., Wang, S., Sharma, K., Biamonte, J., Coles, P.J., Cincio, L., McClean, J.R., Holmes, Z., Cerezo, M.: A review of barren plateaus in variational quantum computing. arXiv preprint arXiv:2405.00781 (2024)
91. Gelman, M.: A survey of methods for mitigating barren plateaus for parameterized quantum circuits. arXiv preprint arXiv:2406.14285 (2024)

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