Pacific
Northwest
NATIONAL LABORATORY

Quantum Neural Networks: Issues, Training, and Applications

September 2023

Carlos Ortiz Marrero
Nathan Wiebe
James Furches
Michael Ragone

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights**. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fox: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

Quantum Neural Networks: Issues, Training, and Applications

September 2023

Carlos Ortiz Marrero
Nathan Wiebe
James Furches
Michael Ragone

Pacific Northwest National Laboratory
Richland, Washington 99354

Table of Contents

Abstract

Our work in the field aims at explaining the limitations and expressive power of Quantum Machine Learning models, as well as finding feasible training algorithms that could be implemented in near-term Quantum Computers. The promise of Quantum Machine Learning is that by incorporating quantum effects, such as entanglement, into machine learning models researchers can improve model performance and understand more complex datasets. This pledge is particularly pronounced in the design of Quantum neural networks (QNNs), a promising framework for creating quantum algorithms, that promise to outperform classical models by combining the speedups of quantum computation with the widespread successes of deep learning. We show that applying this approach alone to quantum deep learning is problematic given that an excess of entanglement between the hidden and visible layers can destroy the predictive power of our QNN models. We address the barren plateau problem by suggesting the use of a generative, unbounded, nonlinear loss function with simple gradients. The loss function quantifies how much the quantum states generated by the QNNs differ from the data and the goal during training is to minimize it. Finally, we showcase how to use generative training to construct a "classical-quantum" neural network to accurately interpolate between the ground states of a Molecular Hamiltonian, a central question in Quantum Chemistry.

Summary

Our work aims at explaining the limitations and expressive power of Quantum Machine Learning models, as well as finding feasible training algorithms, that support scientific and mission-critical applications. The promise of Quantum Machine Learning is that by incorporating quantum effects, such as entanglement, into machine learning models researchers can improve model performance and understand more complex datasets. This pledge is particularly pronounced in the design of Quantum neural networks (QNNs), a promising framework for creating quantum algorithms, that promise to outperform classical models by combining the speedups of quantum computation with the widespread successes of deep learning.

A reason why QNNs can outperform existing deep learning models is the presence of entanglement between the visible and hidden layers of a model. In joint work with Nathan Wiebe (University of Toronto) and Mária Kieferová (UT Sydney), we show that applying this approach alone to quantum deep learning is problematic given that an excess of entanglement between the hidden and visible layers can destroy the predictive power of QNN models. Our key insight is that barren plateaus i.e., vanishing gradients as a model scale in the number of units, can occur because of an excess of entanglement between visible and hidden units in deep quantum neural networks. This surplus of entanglement to some extent defeats the purpose of deep learning by causing information to be non-locally stored in the correlations between the layers rather than in the layers themselves. As a result, when one tries to remove the hidden units, as is customary in deep learning, we find that the resulting state is close to the maximally mixed state i.e., no better than random guessing. Indeed, we show that such situations are generic and gradient descent methods are unlikely to allow the user to escape from such barren plateaus at a low cost.

To address the barren plateau problem we proposed the use of a generative, unbounded, nonlinear loss function with simple gradients. The loss function quantifies how much the quantum states generated by the QNNs differ from the data and the goal during training is to minimize it. Existing algorithms almost exclusively utilize a linear bounded operator as a loss function and linearity of the loss function is a central assumption behind all barren plateau results, including our discovered Entanglement Induced barren plateau. Our proposed training algorithm minimizes a maximal Quantum Rényi divergence, akin to the KL-divergence and an upper bound of the quantum relative entropy, between two quantum states, that is, the output of our QNN and the quantum data. This choice implies that the standard arguments for barren plateau theorems do not apply, in part because Rényi-divergences experience a logarithmic divergence when the two states are nearly orthogonal. This causes the gradients of the divergence between nearly orthogonal quantum states to be large and thereby provides a workaround for all known barren plateau results.

Our previously mentioned work allowed us to apply our algorithms to problems in Quantum Chemistry and Quantum Sensing. In collaboration with domain-experts, we showcase how to use generative training algorithms to construct a "classical-quantum" neural network to accurately interpolate between the ground states of a Molecular Hamiltonian. Quantum Computers were originally proposed for the exact task of simulating chemical processes, which are fundamentally quantum mechanical. Finding the lowest energy state of a chemical Hamiltonian is a central problem in Quantum Chemistry, but it is an expensive computational task even for near-term Quantum Computers. In our work, we used a generative, hybrid "Classical-Quantum" neural network to interpolate between, and generate, the actual ground states of a molecular Hamiltonian. After training, you can then use your model to produce new

ground states to extract any observable, including the ground state energy, without the need to employ expensive Quantum Algorithms like Quantum Phase Estimation each time.
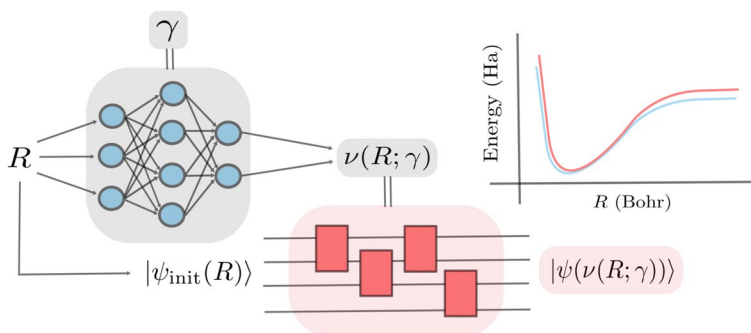

Figure 1: Hybrid Classical-Quantum Generative model.

Our goal in Quantum Sensing is to construct a measurement scheme that can successfully extract quantities from a source system under the influence of competing forces or environmental factors. Current examples include atomic clocks and magnetic resonance imagers, while developing applications include dark-matter detection and atomic microscopy. The goal is to isolate and measure a specific quantity of interest of the evolving quantum sensor in the presence of other competing forces or the environment. We hope to improve on this task by recasting our parameter isolation task as a code-space learning problem in a stabilizer error-correcting code. This entails the development of Quantum Machine Learning algorithms that can map desired parameter regions to regions in code-space that can be distinguished by stabilizer measurements. Undesired parameter regions may then be "corrected", only leaving the desired quantities that are to be estimated during the extraction phase.

Acknowledgments

1.0   Introduction

Despite the optimism, training QNNs has proved to be computationally difficult because of a concentration of measure phenomenon known as a *barren plateau* that leads to exponentially small gradients as we scale the number of neurons. Barren plateaus are known to emerge in many classes of QNNs [1, 4–7], leading to performance that is no better than random guessing. Additionally, recent work showed that under mild assumptions barren plateaus arise whenever QNNs are highly expressible [8] and the issue cannot be alleviated using gradient-free methods [9] or higher-derivatives [10]. These results paint a bleak picture for the future of quantum machine learning and finding a scalable approach to train generic QNNs has become a central problem in the field. Existing approaches that overcome barren plateaus are either based on empirical evidence [11, 12] or are constrained to a specific architecture [5, 13, 14] but none provide a generic way to train quantum models that is guaranteed to avoid these no-go results.

Let us start by defining some terminology. We define a *pure* quantum state, as a unit vector $|\psi\rangle \in (\mathbb{C}^2)^{\otimes n}$. When $n = 1$, it is customary to write $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, where $|0\rangle = (1,0)^t$ and $|1\rangle = (0,1)^t$. Not all pure quantum states describe the complex landscape of behavior that we observe in the lab. It turns out that probabilistic mixtures of pure states are also valid quantum states (these are referred to as *mixed* quantum states). In 1932, John von Neumann presented the mathematical formalism that resolved the issues when describing all possible quantum states [15]. General probabilistic mixtures of pure states cannot be described by a state vector, but von Neumann pointed out that all the measurable information you can extract from a state is encoded in its density matrix. He showed that to every quantum state you can associate a positive semi-definite matrix ρ with $Tr(\rho) = 1$ (called the *density matrix* of the state) and from this matrix you can extract information from all measurable quantities of interest. In the case of a pure state $|\psi\rangle$, it associated density matrix is simply $|\psi\rangle\langle\psi|$, where $\langle\psi| = |\psi\rangle^t$. A more geometric way to understand the set of quantum states is to notice that the set of density matrices is a convex set, and pure states are the extremal points of that set.

Now let us define two popular types of quantum neural networks (QNNs): the unitary Quantum neural network and a quantum Boltzmann machine, see Fig. A. A *Unitary quantum neural networks* are characterized by a sequence of parameterized Unitary matrices acting on an initial state. In general, we can write Unitary QNNs as $U(\theta_1,...,\theta_n) := e^{-iH_n\theta_n}...e^{-iH_1\theta_1}$, where θ_i are the parameters we aim to learn and H_i are Hermitian matrices that specify the QNN. At inference time, $U(\theta_1,...,\theta_n)$ can act on some initial state to get our model output. *Quantum Boltzmann machines* model the data by taking the matrix exponential of a parameterized Hamiltonian and normalizing the resulting matrix to get a density matrix of the quantum state for our model. In both cases, the goal is to find a suitable assignment of parameters given a target quantum state, i.e. "training data". The QNNs make use of two quantum layers; the *visible units* correspond to qubits that we use to construct the output, and the remaining qubits correspond to *hidden units* and provide additional expressive power to the model. The inclusion of hidden units is essential in traditional machine learning and a necessary ingredient of any practical deep neural network. In our work, we show that indiscriminately adding more hidden units to quantum deep learning is problematic given that an excess of entanglement between hidden and visible layers can destroy the predictive power of these models. In effect, we showed the presence of a new type of barren plateau due to entanglement.

The most common approach to training is to find a "good" assignment of parameters via gradient descent. We quantify "goodness" of the parameters in terms of an objective function O $= Tr((O_{obj} \otimes 1)\rho(\theta_1,...,\theta_n))$ where $\rho(\theta_1,...,\theta_n)$ is the density matrix from the quantum state outputted by the model, and $\{\theta_1,...,\theta_n\} \in \mathbb{R}$ are the parameters (O_{obj} is an operator that acts on the visible units and 1 acts on the hidden units). Common choices for the operator O_{obj} include a Hamiltonian or a projector on a given state. The goal of the training phase is to minimize O,

often through stochastic gradient descent. The neural network is used to estimate the gradients $\frac{\partial \mathcal{O}}{\partial \theta_i}$ and parameters θ_i are then systematically adjusted to follow the gradient. The requirement for efficient training is that the gradients be bounded away from zero, otherwise we experience a barren plateau in the loss landscape.

In the seminal work by McClean et al. [4], the authors showed that, with high probability, the derivatives of the expectation value of a random QNN over an observable will be exponentially small for all but an exponentially small fraction of the values the parameters. This result is referred to as the *barren plateau* problem and we restate it below as follows,

Definition 1 (Barren Plateau [16]). *The cost function* O = *Tr*$((O_{obj} \otimes 1)\rho(\theta_1,...,\theta_n))$ *exhibits a barren plateau with respect to parameter θ_j for a quantum model with density ρ, if $\frac{\partial M}{\partial \theta_j}$ exist and is continuous on a compact subset $A \subset \Omega$ and for every $\epsilon > 0$, there exists $0 < b < 1$ such that* $P_A\left(\left|\frac{\partial M}{\partial \theta_j}\right| \geq \epsilon\right) \in O\left(b^{n_v + n_h}\right)$, *where P_A is the probability measure on A induced by P, Ω is a compact set of parameters equipped with a probability density P, O_{obj} is an observable, and n_v, n_h are the numbers of visible and hidden units, respectively, of the quantum model.*
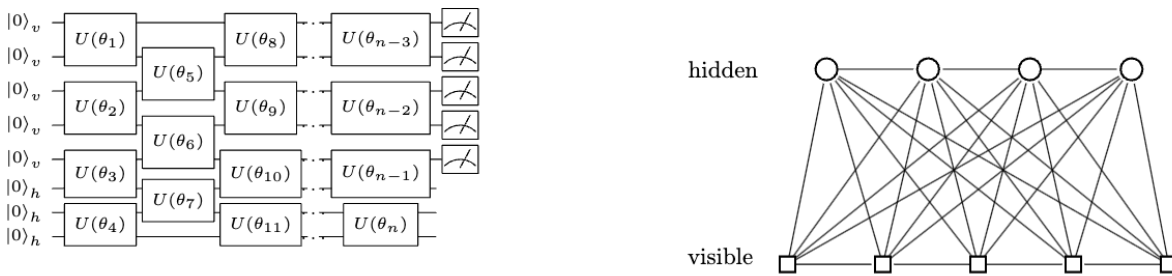

Figure 2: Two types of QNNs: (a) A Unitary QNN with 5 visible units and 3 hidden units. We refer to qubits that are measured at the end as visible units and the remaining qubits as hidden using deep learning notation as an analogy. We denote these qubits $|0\rangle_v$ and $|0\rangle_h$ respectively. Each Unitary (box) acts on the qubit subspace denoted by the wires that go into the box in sequence to produce a final output state. (b) Quantum Boltzmann machine defined on a graph. Each edge and each vertex correspond to a weight on a local Hamiltonian. The top layer (circles) corresponds to visible units and the bottom layer (rectangles) are the hidden units.

Quantum Boltzmann machines model the output as a thermal state $\frac{e^{-H(\theta)}}{Z(\theta)} := \frac{e^{-\sum_i \theta_i H_i}}{\text{Tr}\left(e^{-\sum_i \theta_i H_i}\right)}$. The aim when training a quantum Boltzmann machine is to learn a vector θ such that for a training objective function that acts on the visible subsystem is minimized.

2.0 Entanglement-Induced Barren Plateaus

To appreciate our result and the role of hidden units in our models, let us consider our neural networks as a bipartite quantum system. A classic thermalization result [17] shows that for a random initial state, the state on the visible units is with high probability exponentially close to a maximally mixed state. However, the assumption of a random state is too strong. Instead, we assume that the states are chosen from a 2-design [18]. This is a reasonable assumption for outputs of a randomly initialized, not yet trained, QNN of sufficient depth. Similarly, for a quantum Boltzmann machine we assume that the eigenvalues of the Hamiltonian are chosen at random and then the diagonal matrix is conjugated with a unitary drawn from a distribution that is a unitary 2-design. Now we are ready to state the two main theorems of our work:

Theorem 2 (Gradient in unitary networks). *Assume that $\rho(\theta)$ is drawn from a unitary 2-design where $\rho(\theta)$ is generated by a unitary QNN of the form*

$$\rho(\theta) = \left(\prod_{j=1}^{N} e^{-iH_j\theta_j} \right) |0\rangle\langle 0| \left(\prod_{j=N}^{1} e^{iH_j\theta_j} \right) \tag{1}$$

that acts on a Hilbert space that is the product of a hidden and visible space of dimensions D_h and

D_v respectively. Further, let $H_k(\theta) = \prod_{j=1}^{k} e^{-iH_j\theta_j} H_k \prod_{j=k}^{1} e^{iH_j\theta_j}$ for each k obeys $E(H_k(\theta)\rho(\theta)) = E(H_k(\theta))E(\rho(\theta))$. We then have that

$$E(|\mathrm{Tr}_v(O_{\mathrm{obj}}\mathrm{Tr}_h(\rho(\theta)))|)$$

*is a Lipshitz continuous function of θ with constant Λ obeying*

$$\Lambda \in O\left(\|O_{\mathrm{obj}}\|_\infty \|H_k\|_\infty \sqrt{\frac{D_v}{D_h}} \right). \tag{2}$$

Simply put, under some independence conditions and an assumption that $\|O_{\mathrm{obj}}\|_\infty, \|H_k\|_\infty \in O(1)$, the gradients for QNNs with more hidden units than visible units are exponentially small. It is worth noting that the gradients will be small with high probability as a consequence of the Markov inequality. We proved a similar Theorem for Quantum Boltzmann Machines but we needed to impose some mild assumptions on the structure of the Hamiltonian. If we assumed that Haar-randomness rather than 2-design, we would be able to derive similar results using Levy's lemma [19].

While the most of our contributions consists of theoretic asymptotical results, I complemented our results by providing numerical examples (See Figure 2). While it is impossible to draw a clear conclusion from a small-scale numerical study, our numerical examples demonstrate the vanishing gradients are not only an asymptotic effect and they do occur even for small instances.

3.0 Training Quantum Neural Networks using Rényi Divergences

In the previous subsection we discussed how barren plateaus prevent QNNs from being trained efficiently. In this subsection we discuss a way to prevent the barren plateau problem by using an unbounded, nonlinear loss function with simple gradients instead of a linear objective function. Existing algorithms almost exclusively utilize a linear bounded operator as a loss function and linearity of the loss function is an assumption behind all barren plateau results. We saw that the loss function is typically estimated by measuring the expectation values of Hermitian operators but it is not at all necessary. Our proposed training algorithm minimizes a maximal Quantum Rényi divergence which upper bounds the quantum analog of the KL-divergence between two quantum
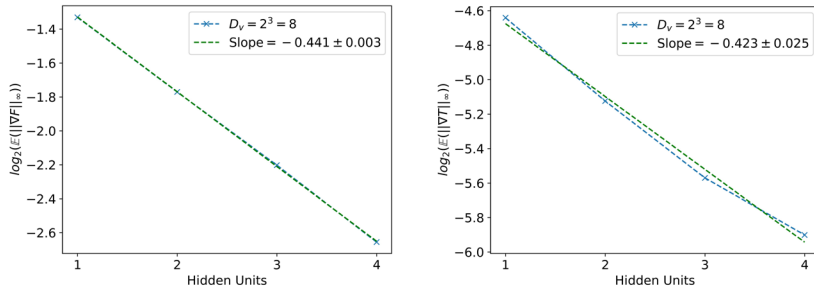
(a) (b)

FIG. 2: Vanishing gradients. (a-b) Semi-log plot highlighting the decay in the variance of the ∞-norm of the gradient vector over an ensemble of initialized models. The dashed blue represent the average over 1000 model instances. The dimension of the visible subpsace is fixed (3 qubits)
while we increase the dimension of hidden subspace. The dashed green line represents is the best fit we obtain from least squares. (a) Gradient estimates for the Unitary Model. (b) Gradient estimates for a normalized Quantum Boltzmann Machine.

states, that is, the output of our QNN and the quantum data. We show how this choice implies that the standard arguments for barren plateau theorems do not apply because the Rényi-divergences experience a logarithmic divergence when the two states are nearly orthogonal. This causes the gradients of the divergence between nearly orthogonal quantum states to be large and thereby provides a workaround for all known barren plateau results.

In our work [2], we derive a closed-form expression for the gradients and provide sufficient conditions for when this loss function avoids barren plateaus. We showcase our method by learning thermal states, a class of mixed states that are hard to generate computationally, but easy to generate experimentally. We provide two quantum algorithms for computing the gradients, a near-term and a fault-tolerant algorithm. I complemented our theoretical work with small numerical experiments and the code available at https://github.com/pnnl/renyiqnets. We observed an absence of gradient decay or barren plateaus in all the learning tasks we performed.

The exact form of the loss function that we consider is a generalization of quantum relative entropy known as quantum Rényi divergence or "sandwiched" Rényi relative entropy [20, 21]. For two quantum states ρ and σ, the quantum Rényi

$$\frac{1}{\alpha-1} \log \left[\mathrm{Tr} \left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}} \right) \right]$$

divergence D_α takes the form $D_\alpha(\rho\|\sigma) = \alpha$ for $\alpha \in [0,\infty)\backslash\{1\}$. The Quantum Rényi divergence inherits many

of the mathematical properties of the Rényi divergence and in the case where $\alpha \to 1$ it reduces to the quantum relative entropy (Quantum analog of the KL-Divergence). One can additionally define an upper-bound on the quantum Rényi divergence $D_\alpha(\rho\|\sigma) \leq D_{e\alpha}(\rho\|\sigma) = \frac{1}{\alpha-1} \log \mathrm{Tr}$ $\left(\sigma^{\frac{1}{2}} \left(\sigma^{\frac{-1}{2}} \rho \sigma^{\frac{-1}{2}}\right)^\alpha \sigma^{\frac{1}{2}}\right)$ where $\widetilde{D}_\alpha$ defines the maximal Rényi divergence, which, until now, has not been widely used for quantum computing. We focused on the case where $\alpha = 2$, first proposed by Petz [22],

$$D_{e2}(\rho\|\sigma(\theta)) = \log \mathrm{Tr}\left(\rho^2 \sigma^{-1}\right). \qquad (3)$$

Here ρ is the training data state and $\sigma(\theta)$ corresponds to the output of the QNN as a function of the parameters θ. The main reason for using $D_{e2}(\rho\|\sigma)$ as a loss function is that it upper-bounds the quantum relative entropy and its gradients are considerably simpler than that of the ordinary Rényi divergence and quantum relative entropy.

We also considered the divergence with reversed arguments,

$$D_{e2}(\sigma(\theta)\|\rho) = \log \mathrm{Tr}\left(\sigma^2 \rho^{-1}\right). \qquad (4)$$

Note that in general, $D_{e2}(\rho\|\sigma(\theta)) \models D_{e2}(\sigma(\theta)\|\rho)$. However, if both ρ and $\sigma(\theta)$ are full rank,

$D_{e2}(\rho\|\sigma(\theta)) = D_{e2}(\sigma(\theta)\|\rho) = 0$ if and only if $\rho = \sigma$ [23]. In this case, $D_{e2}(\sigma(\theta)\|\rho)$ becomes a reasonable loss function to consider for training.

Furthermore, these divergences have closed-form gradients that can be in many cases efficiently computable on a Quantum Computer. For a unitary quantum neural network, we define $\sigma_v :=$

$\mathrm{Tr}^h \left[\Pi_{j=1}^N e^{-iH_j\theta_j} |0\rangle\langle 0| \Pi_{j=N}^1 e^{iH_j\theta_j}\right]$ to be the measured output density of the network, where Tr_h is the partial trace over the hidden subspace generate by the hidden units. Then the gradients of the maximal Rényi divergence between ρ and σ_v takes the form

$$\partial_\theta \widetilde{D}_2(\rho\|\sigma_v(\theta)) = \qquad \mathrm{Tr}\left(\rho^2\sigma_v^{-1}\right)\frac{i\mathrm{Tr}\left(\rho^2\sigma_v^{-1}\mathrm{Tr}_h([\widetilde{H}_k,\sigma])\sigma_v^{-1}\right)}{} \qquad (5)$$

where $\widetilde{H}_k = \Pi_{j=1}^{k-1} e^{-iH_j\theta_j} H_k \Pi_{j=k-1}^1 e^{iH_j\theta_j}$. Similarly, the gradient of the reverse divergence $D_{e2}(\sigma_v\|\rho)$ gives

$$\partial_{\theta_k} D_{e2}(\sigma_v\|\rho) = \frac{-i\mathrm{Tr}\left(\left\{\mathrm{Tr}_h([\widetilde{H}_k,\sigma]), \sigma_v\right\}\rho^{-1}\right)}{\mathrm{Tr}(\sigma_2\rho_{-1})}. \qquad (6)_v$$

We derive gradients for a quantum Boltzmann machine in a similar fashion. It is worth emphasizing that these are considerably simpler than those of quantum relative entropy or quantum negative log-likelihood.

The key advantage of our training method is that it indeed avoids barren plateaus. We proved, under some mild assumptions, that our algorithm does not experience gradient decay for unitary QNNs and Quantum Boltzmann Machines. Furthermore, our numerical results indicate that our mild assumptions might not be necessary in practice, see Fig. 3.

We showcase our training routine by learning thermal states, i.e. $\rho = e^{-H}/\mathrm{Tr}(e^{-H})$ for some Hamiltonian H. For this case, the gradient of the reverse Rényi divergence (6) simplifies to

$$\partial_{\theta_k} D_{e2}(\sigma_v \| \rho) = \frac{-i\mathrm{Tr}\left(\left\{\mathrm{Tr_h}([\widetilde{H}_k, \sigma]), \sigma_v\right\} e^H\right)}{\mathrm{Tr}(\sigma_2 e_H)} . \qquad (7)_v$$

We propose two quantum algorithms for estimating the gradient (7). Our first algorithm prioritizes shorter, noisy computation and could be implemented with near-term Quantum hardware. The second algorithm builds on our first algorithm but uses more sophisticated techniques like amplitude estimation [24].

In addition to our theoretical results, I included a small series of small scale numerical experiments to showcase our ability to learn a thermal state with the Unitary model using the analytical gradients calculated in equation (7). In all our experiments we saw no evidence of gradient decay or barren plateaus during training. Moreover, our model fidelity continued to increase with an increasing number of hidden units, also suggesting an absence of entanglement-induced barren plateaus [1].

Our work focused strictly on generative training but could be applied in a broader Quantum Machine Learning context as a pre-training step. In this approach, one would train a quantum model to perform a generative task first and then train it to perform a specific task in a second phase. This was a common strategy in the early days of Deep Learning. Pre-training would ensure that the model at the start of a second phase is not random and thus will not suffer from gradient decay due to barren plateaus.
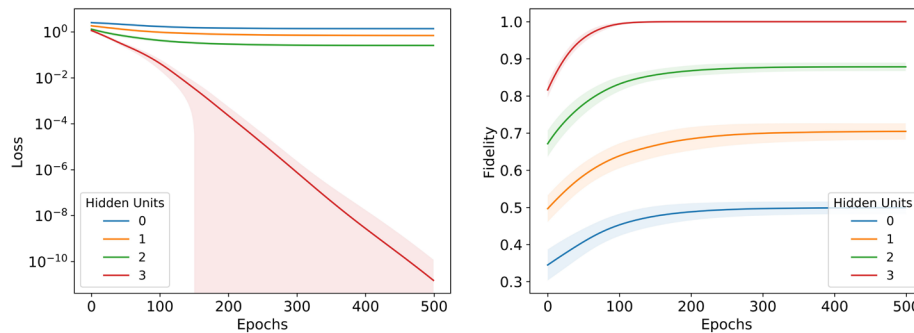
(a) Loss   (b) Fidelity i.e. Accuracy

FIG. 3: We trained the Unitary model with three visible units and an increasing number of hidden units. The target state is a random thermal state. No entanglement-induced barren plateau is observed. The solid lines represent the average epoch value and the width of the shaded area two standard deviations over 50 runs. (a) Training loss (i.e. Rényi Divergence) of our model. (b) Fidelity between the target state and our model.

4.0 Quantum Chemistry, Quantum Sensing, and Nonlocal games

Equipped with our new toolbox, I am now moving into applying them to problems in Quantum Chemistry, Quantum Sensing, and Non-Local Games. We recently published a preprint outlining how we can use generative training to interpolate between the ground states of a parameterized molecular Hamiltonian [3]. Quantum Computers were originally proposed for the exact task of simulating chemical processes, which are fundamentally quantum mechanical [25]. Finding the lowest energy state of a chemical Hamiltonian is a central problem in Quantum Chemistry, but it is an expensive computational task even for near-term Quantum Computers. In our work, we used a generative, hybrid "Classical-Quantum" neural network to interpolate between, and generate, the actual ground states of a molecular Hamiltonian. After training, you can then use your model to produce new ground states to extract any observable, including the ground state energy, without the need to employ expensive Quantum Algorithms like Quantum Phase Estimation [26] each time.

In Quantum Sensing the goal is to construct a measurement scheme that can successfully extract quantities from a source system under the influence of competing forces or environmental factors. Current examples include atomic clocks and magnetic resonance imagers, while developing applications include dark-matter detection and atomic microscopy [27]. The goal is to isolate and measure a specific quantity of interest of the evolving quantum sensor in the presence of other competing forces or the environment. We hope to improve on this task by recasting our parameter isolation task as a code-space learning problem in a stabilizer error-correcting code [28]. This entails the development of Quantum Machine Learning algorithms that can map desired parameter regions to regions in code-space that can be distinguished by stabilizer measurements. Undesired parameter regions may then be "corrected", only leaving the desired quantities that are to be estimated during the extraction phase.

In 1964, John Stewart Bell effectively showed that for some simple games you can achieve higher than expected win rates, if you allow players to share quantum resources when designing strategies [29, 30]. This breakthrough solidified quantum theory as foundational and redefined the term "optimal strategy" in game theory. Recent developments in quantum information theory have highlighted the efficiency of quantum strategies over classical ones [31]. In particular, for some combinatorial games players can convince a referee, with perfect certainty, that they possess an "exponentially" more optimal strategy that wins the game due to their ability to craft a quantum strategy using Quantum resources [32]. I studied and mathematically characterized the strategies of these combinatorial games as part of my thesis work [33]. Most of the existing work on strategies is theoretical in nature [33–35] and little work has been done on algorithm development to find such strategies [36]. This game-theoretic framework provides us with the perfect playground to develop new novel training algorithms that can handle domain constraints: it contains non-trivial but manageable problem instances (e.g. on the order of a dozen qubits [35]), all constraints are known, and enforceable during training. With James Furches (PNNL Intern), We developed a variational algorithm for computing strategies of nonlocal games and show that it can yield optimal strategies for small examples of both convex and non-convex games. We showed that our algorithm returns an optimal quantum strategy for a graph coloring game; whereas no optimal quantum strategy was previously known for this problem. Moreover, we describe how this technique can be run on quantum computers to discover shallow-depth circuits that yield optimal quantum strategies. We argue that such circuits will be useful for benchmarking quantum computers because of the ability to verify the solutions at scale and the experiment's sensitivity to 2-qubit gate noise.

Finally, we demonstrate the use of nonlocal games as a benchmarking strategy experimentally on 11 IBM quantum computers [37].

5.0 A Unified Theory of Barren Plateaus for Deep Parametrized Quantum Circuits

Due to the tremendous limitations that BPs place on the potential to scale variational schemes to large problem sizes, a significant amount of effort has been put forward towards understanding why and when BPs arise. It was found that BPs result from several seemingly disparate aspects, including the expressiveness of the Parametrized Quantum Circuit (PQC) (i.e., the breadth of unitaries that the PQC can express), the locality of the measurement observable, and the entanglement in the input state to the circuit. Hardware noise further exacerbates these issues.
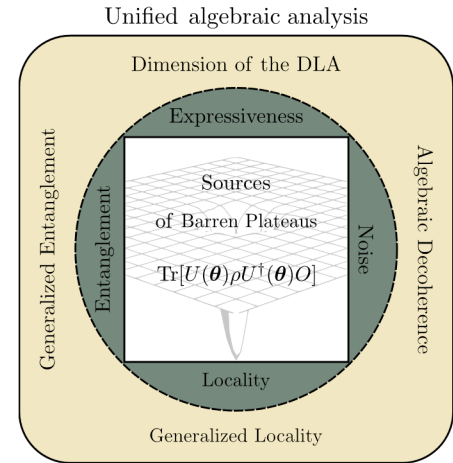

FIG. 1. It is known that BPs can arise due to the expressiveness of the PQC, the locality of the measurement operator, the entanglement in the initial state, or hardware noise. We propose a unified view of BPs through the lens of Lie algebras.

Yet, most of the results in the literature have been derived, and can only be applied, for specific circuit architectures or scenarios. In other words, we do not have a holistic unifying theory that captures the interplay of the various aspects that give rise to BPs.

In order to unify this fragmented understanding, we present a general Lie algebraic theory for BPs. Our theory is based on the study of the Lie group and the associated Lie algebra $\mathfrak{g}$, also known as the *dynamical Lie algebra*, generated by a PQC. Our results can be applied to any deep parametrized quantum circuit architecture, in the presence of state preparation and measurement (SPAM) errors and also *coherent* errors, provided that the measurement operator or the input state belongs to $\mathfrak{g}$ [38].

Our results have extremely powerful implications. First, we can see that the variance depends on only three quantities: the dimension of g and the g-purities of *O* and *ρ*. *Hence, if any of these is exponentially small compared to the rest, the cost will exhibit a BP.* In what follows we analyze each of these three potential causes of barren plateaus and relate them to the expressiveness, generalized entanglement, and locality.

Expressiveness – The expressiveness of the circuit is often advanced as a main cause of BPs. Our result cements this understanding, as the variance is inversely proportional to dim(g), which directly quantifies the expressiveness of the circuit: more expressive circuits (larger dim(g)) lead to more concentrated loss functions. We find that PQCs with exponential DLA are generally untrainable, *regardless of* the initial state or the measurement operator. On the other hand, if $\dim(\mathfrak{g}) \in O(\mathrm{poly}(n))$, the expressiveness of the circuit will not induce BPs by itself. However, this does not exclude the possibility that the initial state or the measurement operator can still lead to an exponential concentration.

Entanglement – It has been realized that if the input state to certain PQCs is too entangled, then the loss has a BP. We can directly relate this observation to $P_{\mathfrak{g}}(\rho)$, the g-purity of the state.

Indeed, the g-purity has been established as a measure of generalized state entanglement. This measure is relative to a subspace of operators (here g) rather than relying on locality concepts based on subsystem decomposition. From this point of view, $P_g(\rho)$ effectively quantifies the amount of mixedness of ρ with respect to the DLA, and is maximized (i.e., ρ is non-g-entangled) if ρ belongs to the orbit of the highest weight state of g. A smaller g-purity implies larger generalized entanglement and, in turn, smaller variances. Therefore, if $P_g(\rho) \in O(1/b^n)$ for $b > 2$ (highly g-entangled state), the loss concentrates exponentially regardless of the expressiveness of the circuit, that is, even if dim(g) is polynomial. Common intuitions about standard entanglement fail to explain variances scalings for general g but are fully captured. This includes situations where local unitary transformations reduce the gpurity (or variance), and also the fact that highly entangled states can still lead to maximal g-purity (or variance).

Locality – It has been shown that for certain PQCs, measuring local operators (acting on a single qubit) leads to a trainable loss, whereas measuring global operators induces BPs. Turning our focus on the g-purity of the measurement operator O, we can carry over the notion of generalized entanglement to define a generalized notion of locality: We call an operator generalized local if it belongs to the preferred subspace of observables given by g. In this case, $\Pi_g(O) = O$. On the other hand, we will call it (fully) generalized-nonlocal if $\Pi_g(O) = 0$. With these definitions, one can readily see that generalized local operators maximize the variance. On the other hand, when $P_g(O) \in O(1/b^n)$ (highly generalized nonlocal measurements), the loss exhibits a BP regardless of the DLA dimension. As was the case for the entanglement, this Lie algebraic perspective allows us to capture cases where standard notions of locality fail, meaning that we can prove the absence of BPs even when measuring operators that act on all qubits.

Noise – While most BPs are discussed in a noiseless setting, it is also known that noise can induce forms of BPs. Consider the case of state preparation with $O \in i$g. The effect of the noise is to map the initial state from ρ to $N_B(\rho)$. We can see that the only way noise can affect the scaling of the variance is through changes in the g-purity of the state. Any channel decreasing this g-purity will reduce its variance. These include incoherent noise, as studied in, such as global depolarization, and also unitary noise (see earlier discussion). These are distinct forms of *algebraic decoherence*, whereby the state becomes entangled with the environment (e.g., for depolarizing noise) or with an effective "algebraic environment" within the system. Similarly, measurement errors can be understood through the action of N_A^{-1} (the potentially unphysical inverse of N_A) on the observable O, and the induced changes in its g-purity. Finally, for circuit errors occurring as uncontrolled unitaries during the circuit, these will increase the set of generators to $G_e = G \cup \{iK_l\}$ with effective DLA $_e$g $\supseteq$ g. Interestingly, this means that coherent noise can increase the expressiveness of the circuit at the cost of decreasing the variance.

Discussions – Finding ways to avoid BPs has been one of the central topics of research in variational quantum computing. This has led the community to develop good-practice guidelines such as: "global observables are untrainable" or "too much entanglement leads to BPs." Although widely regarded as universally true, these are obtained by extrapolating results pertaining to specific scenarios and assuming that they will hold in others. With our work, we suggest a more rigorous and unifying look at BPs through the lens of Lie algebras.

Despite the simplicity of our main result, its implications are far reaching, and, unlike previous results, it provides *exact* variance calculations rather than bounds, allowing us to precisely determine the presence or absence of a BP. Conceptually, our results show that in order to explain the BP phenomenon, one needs to understand the interplay of generalized flavors of entanglement, locality, and also problem size. Finally, it is also worth noting that the formalism adopted here can explain how common noise models, such as SPAM and coherent errors, can lead to BP through the way they increase the generalized entanglement of the initial state, reduce the generalized locality of the measurement operator, or increase the expressive power of the circuit.

Looking ahead, we see different ways to extend our results. For example, our theorems are derived for the case where ρ or O are in g. Although this case encompasses most algorithms in the literature, it would be interesting to generalize our results to operators and measurements not in the DLA. Moreover, one can also envision considering more general noise settings where noise channels are interleaved with the unitaries. Clearly, since a noisy parametrized quantum circuit no longer forms a group, our Lie algebraic formalism no longer applies. As such, new tools will have to be developed to study loss concentration in this scenario.

6.0   References

[1]   Carlos Ortiz Marrero, Mária Kieferová, and Nathan Wiebe. Entanglement-induced barren plateaus. *PRX Quantum*, 2(4):040316, 2021.

[2]   Maria Kieferova, Ortiz Marrero Carlos, and Nathan Wiebe. Quantum generative training using r\'enyi divergences. *arXiv preprint arXiv:2106.09567*, 2021.

[3]   Jack Ceroni, Torin F Stetina, Maria Kieferova, Carlos Ortiz Marrero, Juan Miguel Arrazola, and Nathan Wiebe. Generating approximate ground states of molecules using quantum machine learning. *arXiv preprint arXiv:2210.05489*, 2022.

[4]   Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature communications*, 9(1):1–6, 2018.

[5]   M Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles. Cost-function-dependent barren plateaus in shallow quantum neural networks. *arXiv preprint arXiv:2001.00550*, 2020.

[6]   Samson Wang, Enrico Fontana, Marco Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J Coles. Noise-induced barren plateaus in variational quantum algorithms. *arXiv preprint arXiv:2007.14384*, 2020.

[7]   Zoë Holmes, Andrew Arrasmith, Bin Yan, Patrick J Coles, Andreas Albrecht, and Andrew T Sornborger.
Barren plateaus preclude learning scramblers. *arXiv preprint arXiv:2009.14808*, 2020.

[8]   Zoë Holmes, Kunal Sharma, M Cerezo, and Patrick J Coles. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. *arXiv preprint arXiv:2101.02138*, 2021.

[9]   Andrew Arrasmith, M Cerezo, Piotr Czarnik, Lukasz Cincio, and Patrick J Coles. Effect of barren plateaus on gradient-free optimization. *arXiv preprint arXiv:2011.12245*, 2020.

[10] Marco Cerezo and Patrick J Coles. Higher order derivatives of quantum neural networks with barren plateaus. *Quantum Science and Technology*, 2021.

[11] Edward Grant, Leonard Wossnig, Mateusz Ostaszewski, and Marcello Benedetti. An initialization strategy for addressing barren plateaus in parametrized quantum circuits. *Quantum*, 3:214, 2019.

[12] Andrea Skolik, Jarrod R McClean, Masoud Mohseni, Patrick van der Smagt, and Martin Leib. Layerwise learning for quantum neural networks. *Quantum Machine Intelligence*, 3(1):1–11, 2021.

[13] Arthur Pesah, M Cerezo, Samson Wang, Tyler Volkoff, Andrew T Sornborger, and Patrick J Coles. Absence of barren plateaus in quantum convolutional neural networks. *arXiv preprint arXiv:2011.02966*, 2020.

[14] Kunal Sharma, M Cerezo, Lukasz Cincio, and Patrick J Coles. Trainability of dissipative perceptron-based quantum neural networks. *arXiv preprint arXiv:2005.12458*, 2020.

[15] John Von Neumann. *Mathematical foundations of quantum mechanics: New edition*. Princeton university press, 2018.

[16] Tyler Volkoff and Patrick J Coles. Large gradients via correlation in random parameterized quantum circuits. *Quantum Science and Technology*, 6(2):025008, 2021.

[17] Sandu Popescu, Anthony J Short, and Andreas Winter. Entanglement and the foundations of statistical mechanics. *Nature Physics*, 2(11):754–758, 2006.

[18] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Physical Review A*, 80(1):012304, 2009.

[19] Patrick Hayden, Debbie W Leung, and Andreas Winter. Aspects of generic entanglement. *Communications in mathematical physics*, 265(1):95–117, 2006.

[20] Mark M Wilde, Andreas Winter, and Dong Yang. Strong converse for the classical capacity of entanglement-breaking and hadamard channels via a sandwiched rényi relative entropy. *Communications in Mathematical Physics*, 331(2):593–622, 2014.

[21] Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. On quantum rényi entropies: A new generalization and some properties. *Journal of Mathematical Physics*, 54(12):122203, 2013.

[22] Dénes Petz. Quasi-entropies for finite quantum systems. *Reports on mathematical physics*, 23(1):57–65, 1986.

[23] Tim Van Erven and Peter Harremos. Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.

[24] Gilles Brassard, Peter Hoyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, 305:53–74, 2002.

[25] Richard Phillips Feynman and FL Vernon Jr. The theory of a general quantum system interacting with a linear dissipative system. *Annals of physics*, 281(1-2):547–607, 2000.

[26] A Yu Kitaev. Quantum measurements and the abelian stabilizer problem. *arXiv preprint quantph/9511026*, 1995.

[27] Quantum Information Science Committee On Science. Bringing Quantum Sensors to Fruition. Technical report, National Science and Technology Council, March 2022.

[28] Daniel Gottesman. *Stabilizer codes and quantum error correction*. California Institute of Technology, 1997.

[29] John S Bell. On the einstein podolsky rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964.

[30] John F Clauser, Michael A Horne, Abner Shimony, and Richard A Holt. Proposed experiment to test local hidden-variable theories. *Physical review letters*, 23(15):880, 1969.

[31] Carlos A Melo-Luna, Cristian E Susa, Andrés F Ducuara, Astrid Barreiro, and John H Reina. Quantum locality in game strategy. *Scientific Reports*, 7(1):1–11, 2017.

[32] David Avis, Jun Hasegawa, Yosuke Kikuchi, and Yuuya Sasaki. A quantum protocol to win the graph colouring game on all hadamard graphs. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 89(5):1378–1381, 2006.

[33] Carlos M Ortiz and Vern I Paulsen. Quantum graph homomorphisms via operator systems. *Linear Algebra and its Applications*, 497:23–43, 2016.

[34] Martino Lupini, Laura Mančinska, Vern I Paulsen, David E Roberson, Giannicola Scarpa, Simone Severini, Ivan G Todorov, and Andreas Winter. Perfect strategies for non-local games. *Mathematical Physics, Analysis and Geometry*, 23(1):1–31, 2020.

[35] Laura Mančinska and David E Roberson. Oddities of quantum colorings. *arXiv preprint arXiv:1801.03542*, 2018.

[36] Kishor Bharti, Tobias Haug, Vlatko Vedral, and Leong-Chuan Kwek. How to teach ai to play bell non-local games: Reinforcement learning. *arXiv preprint arXiv:1912.10783*, 2019.

[37] J. Furches, N. Wiebe, and C. O. Marrero, "Variational Methods for Computing Non-Local Quantum Strategies." arXiv, Nov. 02, 2023. Accessed: Nov. 09, 2023. [Online]. Available: http://arxiv.org/abs/2311.01363.

[38] M. Ragone et al., "A Unified Theory of Barren Plateaus for Deep Parametrized Quantum Circuits." arXiv, Sep. 20, 2023. doi: 10.48550/arXiv.2309.09342.

Pacific Northwest
National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov