

Challenges and Opportunities in Quantum Machine Learning

M. Cerezo,^{1, 2, 3} Guillaume Verdon,^{4, 5, 6} Hsin-Yuan Huang,^{7, 8} Lukasz Cincio,^{9, 3} and Patrick J. Coles^{10, 9, 3}

¹*Information Sciences, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

²*Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

³*Quantum Science Center, Oak Ridge, TN 37931, USA*

⁴*X, Mountain View, CA, USA*

⁵*Institute for Quantum Computing, University of Waterloo, ON, Canada*

⁶*Department of Applied Mathematics, University of Waterloo, ON, Canada*

⁷*Institute for Quantum Information and Matter, California Institute of Technology, USA*

⁸*Department of Computing and Mathematical Sciences, California Institute of Technology, USA*

⁹*Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

¹⁰*Normal Computing Corporation, New York, New York, USA*

At the intersection of machine learning and quantum computing, Quantum Machine Learning (QML) has the potential of accelerating data analysis, especially for quantum data, with applications for quantum materials, biochemistry, and high-energy physics. Nevertheless, challenges remain regarding the trainability of QML models. Here we review current methods and applications for QML. We highlight differences between quantum and classical machine learning, with a focus on quantum neural networks and quantum deep learning. Finally, we discuss opportunities for quantum advantage with QML.

I. INTRODUCTION

The recognition that the world is quantum mechanical has allowed researchers to embed well-established, but classical, theories into the framework of quantum Hilbert spaces. Shannon’s information theory, which is the basis of communication technology, has been generalized to quantum Shannon theory (or quantum information theory), opening up the possibility that quantum effects could make information transmission more efficient [1]. The field of biology has been extended to quantum biology to allow for a deeper understanding of biological processes like photosynthesis, smell, and enzyme catalysis [2]. Turing’s theory of universal computation has been extended to universal quantum computation [3], potentially leading to exponentially faster simulations of physical systems.

One of the most successful technologies of this century is machine learning (ML), which aims to classify, cluster, and recognize patterns for large data sets. Learning theory has been simultaneously developed alongside of ML technology in order to understand and improve upon its success. Concepts like support vector machines, neural networks, and generative adversarial networks have impacted science and technology in profound ways. ML is now ingrained into society to such a degree that any fundamental improvement to ML leads to tremendous economic benefit.

Like other classical theories, ML and learning theory can in fact be embedded into the quantum mechanical formalism. Formally speaking, this embedding leads to the field known as Quantum Machine Learning (QML) [4–6], which aims to understand the ultimate limits of data analysis allowed by the laws of physics. Practically speaking, the advent of quantum computers, with the hope of achieving a so-called quantum advantage (as defined below) for data analysis, is what has made

QML so exciting. Quantum computing exploits entanglement, superposition, and interference to perform certain tasks with significant speedups over classical computing, sometimes even exponentially faster. Indeed while such speedup has already been observed for a contrived problem [7], reaching it for data science is still uncertain even at the theoretical level, but this is one of the main goals for QML.

In practice, QML is a broad term that encompasses all of the tasks shown in Fig. 1. For example, one can apply machine learning to quantum applications like discovering quantum algorithms [8] or optimizing quantum experiments [9, 10], or one can use a quantum neural network to process either classical or quantum information [11]. Even classical tasks can be viewed as QML when they are quantum inspired [12]. We note that the focus of this article will be on quantum neural networks, quantum deep learning, and quantum kernels, even though the field of QML is quite broad and goes beyond these topics.

After the invention of the laser, it was called a solution in search of a problem. To some degree, the situation with QML is similar. The complete list of applications of QML is not fully known. Nevertheless, it is possible to speculate that all the areas shown in Fig. 2 will be impacted by QML. For example, QML will likely benefit chemistry, materials science, sensing and metrology, classical data analysis, quantum error correction, and quantum algorithm design. Some of these applications produce data that is inherently quantum mechanical, and hence it is natural to apply QML (rather than classical ML) to them.

While there are similarities between classical and quantum ML, there are also some differences. Because QML employs quantum computers, noise from these computers can be a major issue. This includes hardware noise like decoherence as well as statistical noise (i.e., shot noise) that arises from measurements on quantum states. Both

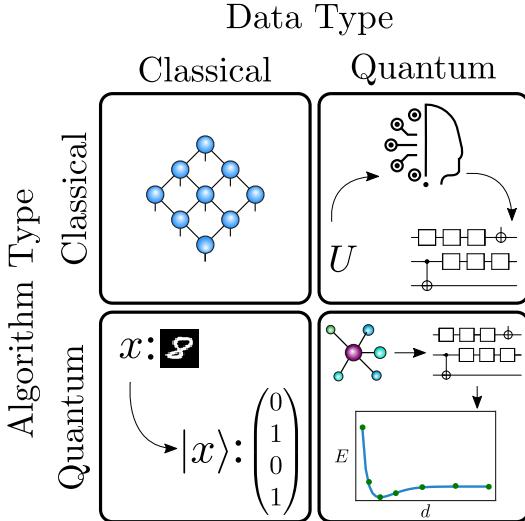


FIG. 1. Quantum Machine Learning (QML) tasks. Quantum machine learning is usually considered for four main tasks. These include tasks where the data is either classical or quantum, and where the algorithm is either classical or quantum. Top left: tensor networks are quantum-inspired classical methods that can analyze classical data. Top right: unitary time-evolution data U from a quantum system can be classically compiled into a quantum circuit. Bottom left: handwritten digits can be mapped to quantum states for classification on a quantum computer. Bottom right: molecular ground state data can be classified directly on a quantum computer. The figure shows ground state energy E dependence on the distance d between the atoms.

of these noise sources can complicate the QML training process. Moreover, non-linear operations (e.g., neural activation functions) that are natural in classical ML require more careful design of QML models due to the linearity of quantum transformations.

For the field of QML, the immediate goal for the near-future is demonstrating quantum advantage, i.e., outperforming classical methods, in a data science application. Achieving this goal will require keeping an open mind about which applications will benefit most from QML (e.g., it may be an application that is inherently quantum mechanical). Understanding how QML methods scale to large problem sizes will also be required, including analysis of trainability (gradient scaling) and prediction error. The availability of high quality quantum hardware [13, 14] will also be crucial.

Finally, we note that QML provides a new way of thinking about established fields, like quantum information theory, quantum error correction, and quantum foundations. Viewing such applications from a data science perspective will likely lead to new breakthroughs.

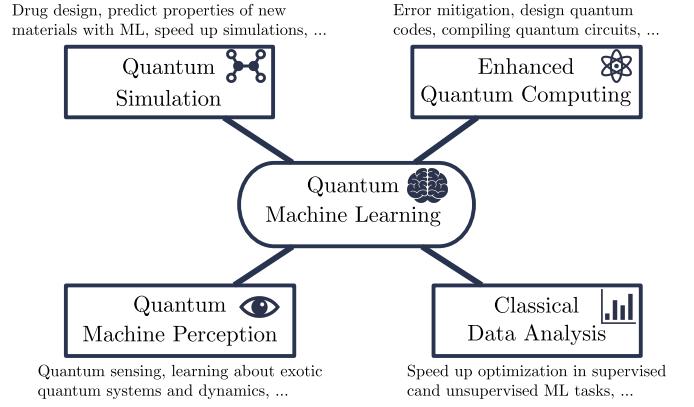


FIG. 2. Key Applications for QML. QML has been envisioned to bring a computational advantage in many applications. QML can enhance quantum simulation for chemistry (e.g., molecular ground states [112], equilibrium states [47], and time evolution [114]) and materials science (e.g., quantum phase recognition [11] and generative design with a target property in mind [132]). QML can enhance quantum computing by learning quantum error correction codes [11, 111] and syndrome decoders, performing quantum control, learning to mitigate errors, and compiling and optimizing quantum circuits. QML can enhance sensing and metrology [46, 106–109] and extract hidden parameters from quantum systems. Finally, QML may speed up classical data analysis, including clustering and classification.

II. FRAMEWORK

A. Data

As shown in Fig. 3, QML can be used to learn from either classical or quantum data, and thus we begin by contrasting these two types of data. Classical data is ultimately encoded in bits, each of which can be in a 0 or 1 state. This includes images, texts, graphs, medical records, stock prices, properties of molecules, outcomes from biological experiments, and collision traces from high energy physics experiments. Quantum data is encoded in quantum bits, called qubits, or higher-dimensional analogs. A qubit can be represented by the states $|0\rangle$, $|1\rangle$, or any normalized complex linear superposition of these two. Here, the states contain information obtained from some physical process such as quantum sensing [15], quantum metrology [16], quantum networks [17], quantum control [18], or even quantum analog-digital transduction [19]. Moreover, quantum data can also be the solution to problems obtained on a quantum computer, e.g., the preparation of various Hamiltonians' ground states.

In principle, all classical data can be efficiently encoded in systems of qubits: a classical bitstring of length n can be easily encoded onto n qubits. However, the same cannot be said for the converse, since one cannot efficiently encode quantum data in bit systems, i.e., the state of a general n qubit system requires $(2^n - 1)$ complex num-

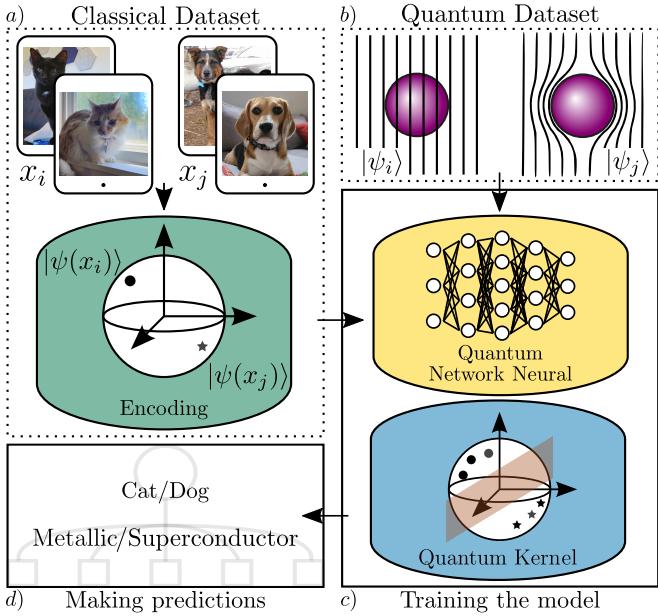


FIG. 3. Classification with QML. a) The classical data x , i.e., images of cats and images of dogs, is encoded into a Hilbert space via some map $x \rightarrow |\psi(x)\rangle$. Ideally, data from different classes (here represented by dots and stars) is mapped to different regions of the Hilbert space . b) Quantum data $|\psi\rangle$ can be directly analyzed on a quantum device. Here the dataset is composed of states representing metallic or superconducting systems. c) The dataset is used to train a QML model. Two common paradigms in QML are quantum neural networks and quantum kernels, both of which allow for classification of either classical or quantum data. In Kernel methods one fits a decision hyperplane that separates the classes. d) Once the model is trained, it can be used to make predictions.

bers to be specified. Hence, systems of qubits (and more generally the quantum Hilbert space) constitute the ultimate data representation medium, as they can encode not only classical information but also quantum information obtained from physical processes.

In a QML setting, the term quantum data refers to data that is naturally already embedded in a Hilbert space \mathcal{H} . When the data is quantum, it is already in the form of a set of quantum states $\{|\psi_j\rangle\}$ or a set of unitaries $\{U_j\}$ that could prepare these states on a quantum device (via the relation $|\psi_j\rangle = U_j |\mathbf{0}\rangle$). On the other hand, when the data x is classical, it first needs to be encoded in a quantum system through some embedding mapping $x_j \rightarrow |\psi(x_j)\rangle$, with $|\psi(x_j)\rangle$ in \mathcal{H} . In this case, the hope is that the QML model can solve the learning task by accessing the exponentially large dimension of the Hilbert space [20–23].

One of the most important and reasonable conjectures to make is that the availability of quantum data will significantly increase in the near future. The mere fact that people will use the quantum computers that are available will logically lead to more quantum problems being

solved and quantum simulations being performed. These computations will produce quantum data sets, and hence it is reasonable to expect the rapid rise of quantum data. Note that, in the near term, this quantum data will be stored on classical devices in the form of efficient descriptions of quantum circuits that prepare the datasets.

Finally, as our level of control over quantum technologies progress, coherent transduction of quantum information from the physical world to digital quantum computing platforms may be achieved [19]. This would quantum mechanically mimic the main information acquisition mechanism for classical data from the physical world, that being analog-digital conversion. Moreover, we can expect that the eventual advent of practical quantum error correction [24] and quantum memories [25] will allow us to store quantum data on quantum computers themselves.

B. Models

Analyzing and learning from data requires a parameterized model, and many different models have been proposed for QML applications. Classical models like neural networks and tensor networks (as shown in Fig. 1) are often useful for analyzing data coming from quantum experiments. However, due to their novelty, we will focus our discussion on quantum models using quantum algorithms, where one applies the learning methodology directly at the quantum level.

Similar to classical ML, there exists several different QML paradigms: supervised learning (task-based) [26–28], unsupervised learning (data-based) [29, 30] and reinforced learning (reward-based) [31, 32]. While each of these fields is exciting and thriving on its own, supervised learning has recently received considerable attention for its potential to achieve quantum advantage [26, 33], resilience to noise [34], and good generalization properties [35–37], which makes it a strong candidate for near-term applications. In what follows we discuss two popular QML models: quantum neural networks (QNNs) and quantum kernels, shown in Fig. 3, with a particular emphasis on QNNs as these are the primary ingredient of several supervised, unsupervised, and reinforced learning schemes.

1. Quantum neural networks

The most basic and key ingredient in QML models are Parameterized Quantum Circuits (PQCs). These involve a sequence of unitary gates acting on the quantum data states $|\psi_j\rangle$, some of which have free parameters θ that will be trained to solve the problem at hand [38]. PQCs are conceptually analogous to neural networks, and indeed this analogy can be made precise, i.e., classical neural networks can be formally embedded into PQCs [39].

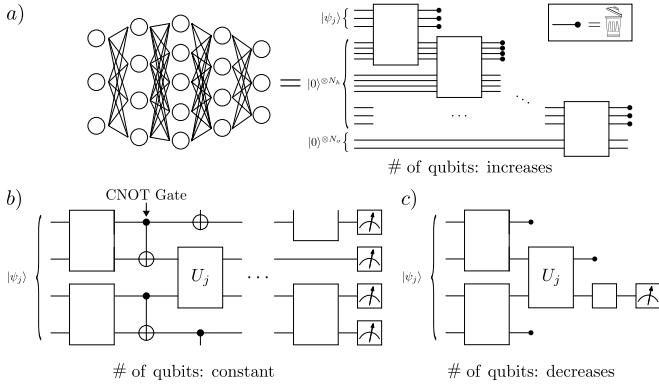


FIG. 4. Examples of QNN architectures. a) A classical feed-forward neural network has input, hidden, and output layers. This can be generalized to the quantum setting with a dissipative QNN, where some qubits are discarded and replaced by new qubits during the algorithm. Here we show a quantum circuit representation for the dissipative QNN. In a circuit diagram each horizontal line represents a qubit, and the logical operations, or quantum gates, are represented by boxes connecting the qubit lines. Circuits are read from left to right. For instance, here the circuit is initialized in a product state $|\psi_j\rangle \otimes |0\rangle^{\otimes(N_h+N_o)}$, where $|\psi_j\rangle$ encodes the input data and N_h (N_o) is the number of qubits in the hidden (output) layer. As one performs logical operations, the information forward propagates through the circuit. b) Another possible QNN strategy is to keep the qubits fixed, without discarding or replacing them. The circuit represents consecutive application of two-qubit gates U_j and controlled-NOT (denoted by CNOT) gates. c) Quantum convolutional neural networks (QCNNs) measure and discard qubits during the algorithm. The QCNN circuit considered here is built with two-qubit quantum gates U_j and is initialized in $|\psi_j\rangle$.

This has led researchers to refer to certain kinds of PQCs as Quantum Neural Networks (QNNs). In practice, the term QNN is used whenever a PQC is employed for a data science application, and hence we will use the term QNN in what follows. QNNs are employed in all three QML paradigms mentioned above. For instance, in a supervised classification task, the goal of the QNN is to map the states in different classes to distinguishable regions of the Hilbert space [26]. Moreover, in the unsupervised learning scenario of [29] a clustering task is mapped onto a MAXCUT problem and solved by training a QNN to maximize distance between classes. Finally, in the reinforced learning task of [32], a QNN can be used as the Q-function approximator, which can be used to determine the best action for a learning agent given its current state.

Figure 4 gives examples of three distinct QNN architectures where at each layer the number of qubits in the model is increased, preserved, or decreased. In Fig. 4(a) we show a dissipative QNN [40] which generalizes the classical feed-forward network. Here, each node corresponds to a qubit, while lines connecting qubits are unitary operations. The term dissipative arises from the fact that qubits in a layer are discarded after the information

forward-propagates to the (new) qubits in the next layer. Figure 4(b) shows a standard QNN where quantum data states are sent through a quantum circuit, at the end of which some or all of the qubits are measured. Here, no qubits are discarded or added as one goes deeper into the QNN. Finally, Fig. 4(c) depicts a convolutional QNN [11], where at each layer qubits are measured to reduce the dimension of the data while preserving its relevant features. Many other QNNs have been proposed [41–45], and constructing QNN architectures is currently an active area of research.

To further accommodate for the limitation of near-term quantum computers, one can also employ a hybrid approach with models that have both classical and quantum neural networks [46]. Here, QNNs act coherently on quantum states while deep classical neural networks alleviate the need for higher-complexity quantum processing. Such hybridization distributes the representational capacity and computational complexity across both quantum and classical computers. Moreover, since quantum states generally have a mixture of classical correlations and quantum correlations, hybrid quantum-classical models allow for the use of quantum computers as an additive resource to increase the ability for classical models to represent quantum-correlated distributions. Applications of hybrid models include generating [47] or learning and distilling information [46] from multipartite-entangled distributions.

2. Quantum kernels

As an alternative to QNNs, researchers have proposed quantum versions of kernel methods [26, 28]. A kernel method maps each input to a vector in a high-dimensional vector space, known as the reproducing kernel Hilbert space. Then, a kernel method learns a linear function in the reproducing kernel Hilbert space. The dimension of the reproducing kernel Hilbert space could be infinite, which enables the kernel method to be very powerful in terms of the expressiveness. To learn a linear function in a potentially infinite-dimensional space, the kernel trick [48] is employed, which only requires efficient computation of the inner product between these high-dimensional vectors. The inner product is also known as the kernel [48]. Quantum kernel methods consider the computation of kernel functions using quantum computers. There are many possible implementations. For example, [26, 28] considered a reproducing kernel Hilbert space equal to the quantum state space, which is finite dimensional. Another approach [13] is to study an infinite-dimensional reproducing kernel Hilbert space that is equivalent to transforming classical vector using a quantum computer. It then maps the transformed classical vectors to infinite-dimensional vectors.

3. Inductive bias

For both QNNs and quantum kernels, an important design criterion is their inductive bias. This bias refers to the fact that any model represents only a subset of functions and is naturally biased towards certain types of functions (i.e., functions relating the input features to the output prediction). One aspect of achieving quantum advantage with QML is to aim for QML models that have an inductive bias that is inefficient to simulate with a classical model. Indeed, it was recently shown [49] that quantum kernels with this property can be constructed, albeit with some subtleties regarding their trainability.

Generally speaking, inductive bias encompasses any assumptions made in the design of the model or the optimization method which bias the search of the potential models to a subset in the set of all possible models. In the language of Bayesian probabilistic theory, we usually call these assumptions our prior. Having a certain parameterization of potential models, like QNNs, or choosing a particular embedding for quantum kernel methods [13, 14, 26] is itself a restriction of the search space, and hence a prior. Adding a regularization term to the optimizer or modulating the learning rate to keep searches geometrically local also adds inherently a prior and focuses the search, and thus provides inductive bias.

Ultimately, inductive biases from the design of the ML model, combined with a choice of training process, are what make or break an ML model. The main advantage of QML will then be to have the ability to sample from and learn models that are (at least partially) natively quantum mechanical. As such, they have inductive biases that classical models do not have. This discussion assumes that the dataset to be represented is quantum mechanical in nature, and is one of the reasons why researchers typically believe that QML has greater promise from quantum rather than classical data.

C. Training and Generalization

The ultimate goal of ML (classical or quantum) is to train a model to solve a given task. Thus, understanding the training process of QML models is fundamental for their success.

Consider the training process, whereby one aims to find the set of parameters θ that lead to the best performance. The latter can be accomplished, for instance, by minimizing a loss function $\mathcal{L}(\theta)$ encoding the task at hand. Some methods for training QML models are leveraged from classical ML, like stochastic gradient descent. However, shot noise, hardware noise, and unique landscape features often make off-the-shelf classical optimization methods perform poorly for QML training [50]. This realization led to development of quantum-aware optimizers, which account for the quantum idiosyncrasies of the QML training process. For example, shot-frugal optimizers [51–54] can employ stochastic gradient descent

while adapting the number of shots (or measurements) needed at each iteration, so as not to waste too many shots during the optimization. Quantum natural gradient [55, 56] adjusts the step size according to the local geometry of the landscape (based on the quantum Fisher information metric). These and other quantum-aware optimizers often outperform standard classical optimization methods in QML training tasks.

For the case of supervised learning, one is not only interested in learning from a training data set but also in making accurate predictions on (generalize to) previously unseen data. This translates into achieving small training and prediction errors, with the second usually hinging on the first. Thus, let us now consider prediction error, also known as generalization error, which has been studied only very recently for QML [13, 14, 35, 37, 57, 58]. Formally speaking, this error measures the extent to which a trained QML model performs well on unseen data. Prediction error depends on both the training error as well as the complexity of the trained model. If the training error is large, the prediction error is also typically large. If the training error is small but the complexity of the trained model is large, then the prediction error is likely still large. The prediction error is small only if training error is itself small and the complexity of the trained model is moderate (i.e., sufficiently smaller than training data size) [14, 35]. The notion of complexity depends on the QML model. We have a good understanding of the complexity of quantum kernel methods [13, 14], while more research is needed on QNN complexity. Recent theoretical analysis of QNNs shows that their prediction performance is closely linked to the number of independent parameters in the QNN, with good generalization obtained when the amount of training data is roughly equal to the number of parameters [35]. This gives the exciting prospect of using only a small amount of training data to obtain good generalization.

III. CHALLENGES IN QML

Heuristic fields can face periods of stagnation (or “winters”) due to unforeseen technical challenges. Indeed in classical ML, there was a gap between introducing a single perceptron [59] and the multi-layer perceptron [60] (i.e., neural network), and there was also a gap between attempts to train multiple layers and the introduction of the backpropagation method [61].

Naturally we would like to avoid these stagnations or winters for QML. The obvious strategy is to try to determine all of the challenges as quickly possible, and focus research effort on addressing them. Fortunately, QML researchers have taken this strategy. Figure 5 showcases some of the different elements of QML models, as well as the challenges associated with them. In this section we detail various QML challenges and how one could potentially avoid them.

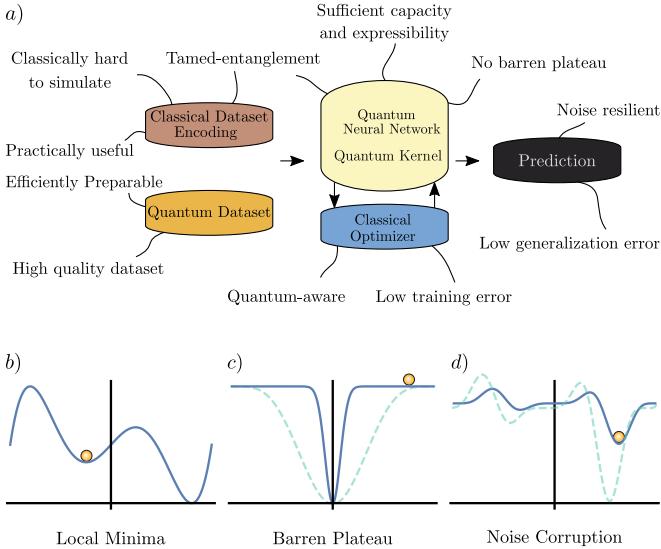


FIG. 5. Challenges for QML. a) There are several ingredients and priors needed to build a QML model: a dataset (and an encoding scheme for classical data), the choice of parameterized model, loss function, and classical optimizer. In this diagram, we show some of the challenges of the different components of the model. b-d) The success of the QML model hinges on an accurate and efficient training of the parameters. However, there are certain phenomena that can hinder the QML trainability. These include the abundance of low-quality local minima solutions shown in b), as well as the barren plateau phenomenon in c). When a QML architecture exhibits a barren plateau, the landscape becomes exponentially flat (on average) as the number of qubits increases (seen as a transition from dashed to solid line). The presence of hardware noise has been shown to erase the features in the landscape as well as potentially shift the position of the minima. Here, the dashed (solid) line corresponds to the noiseless (noisy) landscape shown in d).

A. Embedding schemes and quantum datasets

The access to high-quality, standardized datasets has played a key role in advancing classical ML. Hence, one could conjecture that such datasets will be crucial for QML as well.

Currently, most QML architectures are benchmarked using classical datasets (such as MNIST, Dogs vs Cats, and Iris). While using classical datasets is natural due to their accessibility, it is still unclear how to best encode classical information onto quantum states. Several embedding schemes have been proposed [22, 26, 62], and there are some desirable properties they must possess. One such property is that the inner product between output states of the embedding is classically hard to simulate (otherwise the quantum kernel would be classically simulable). In addition, the embedding should be practically useful, i.e., in a classification task, the states should be in distinguishable regions of the Hilbert space. Unfortunately, embeddings that satisfy one of these properties do not necessarily satisfy the other [63]. Thus, developing

encoding schemes is an active area of research, especially those that are equipped with an inductive bias containing information about the dataset [49].

Furthermore, some recent results suggest that achieving a quantum advantage with classical data might not be straightforward [49]. On the other hand, QML models with quantum data have a more promising route towards a quantum advantage [64–67]. Despite this fact, there is still a dearth of truly quantum datasets for QML, which just a few recently proposed [68, 69]. Hence, the field needs standardized quantum datasets with easily preparable quantum states, as these can be used to benchmark QML models on true quantum data.

B. Quantum landscapes

Training the parameters of the QML model corresponds in a wide array of cases to minimizing a loss function and navigating through a (usually non-convex) loss function landscape in search for its global minimum [70]. Quantum landscape theory [71] aims to understand QML landscape properties and how to engineer them. Local minima and barren plateaus have received significant attention in quantum landscape theory.

1. Local minima in quantum landscapes

As schematically shown in Fig. 5(b), similar to classical ML, the quantum loss landscape can have many local minima. Ultimately, this can lead to the overall non-convex optimization being NP-hard [72], which is again similar to the classical case. There have been some methods proposed to address local minima. For example, variable structure QNNs [73, 74], which grow and contract throughout the optimization, adaptively change the model’s prior and allow some local minima to be turned into saddle points. Moreover, evidence of the overparametrization phenomenon has been seen for QML [75, 76]. Here, the optimization undergoes a computational phase transition, due to spurious local minima disappearing, whenever the number of parameters exceeds a critical value.

2. Overview of barren plateaus

Local minima are not the only issue facing QML, as it has been shown that quantum landscapes can exhibit a fascinating property known as a *barren plateau* [58, 77–89]. As depicted in Fig. 5(c), in a barren plateau the loss landscape becomes, on average, exponentially flat with the problem size. When this occurs, the valley containing the global minimum also shrinks exponentially with problem size, leading to a so-called *narrow gorge* [71]. As a consequence, one requires exponential resources (e.g., numbers of shots) to navigate through the

landscape. The latter impacts the complexity of one’s QML algorithm and can even destroy quantum speedup, since quantum algorithms typically aim to avoid the exponential complexity normally associated with classical algorithms.

3. Barren plateaus from ignorance or insufficient inductive bias

The barren plateau phenomenon was first studied in deep hardware-efficient QNNs [77], where they arise due to the high *expressivity* of the model [81]. By making no assumptions about the underlying data, deep hardware-efficient architectures aims to solve a problem by being able to prepare a wide range of unitary evolutions. In other words, the prior over hypothesis space is relatively uninformed. Barren plateaus in this unsharp prior are caused by ignorance or the lack of sufficient inductive bias, and therefore a means to avoid them is to input knowledge into the construction of the QNN - making the design of QNNs with good inductive biases for the problem at hand a key solution.

Fortunately various strategies have been developed to address these barren plateaus, such as clever initialization [90], pre-training, and parameter correlation [82, 83]. These are all examples of adding a sharper prior to one’s search over the over-expressive parameterizations of hardware efficient QNNs. Below we further discuss how QNN architectures can be designed to further introduce inductive bias.

4. Barren plateaus from global observables

Other mechanisms have been linked to barren plateaus. Simply defining a loss function based on a global observable (i.e., observables measuring all qubits) leads to barren plateaus even for shallow circuits with sharp priors [78], while local observables (those comparing quantum states at the single-qubit level) avoid this issue [78, 87]. The latter is not due to bad inductive biases but rather to the fact that comparing objects in exponentially large Hilbert spaces requires an exponential precision, as their overlap is usually exponentially small.

5. Barren plateaus from entanglement

While entanglement is one of the most important quantum resources for information processing tasks in quantum computers, it can also be detrimental for QML models. QNNs (or embedding schemes) that generate too much entanglement also lead to barren plateaus [84, 86, 88]. Here, the issue arises when one entangles the visible qubits of the QNN (those that one measures at the QNN’s output) with a large number of qubits in the hidden layers. Due to entanglement, the information of the

state is stored in non-local correlations across all qubits, and hence the reduced state of the visible qubits concentrates around the maximally mixed state. This type of barren plateau can be solved by taming the entanglement generated across the QNN.

C. QNN architecture design

One of the most active areas is developing QNN architectures that have sharp priors. Since QNNs are a fundamental ingredient in supervised learning (deep learning, kernel methods), but also in unsupervised learning and reinforced learning, developing good QNN architectures is crucial for the field.

For instance, it has been shown that QNNs with sharp priors can avoid issues such as barren plateaus altogether. One such example are Quantum Convolutional Neural Networks (QCNNs) [11]. QCNNs possess an inductive bias from having a prior over the space of architectures that is much sharper than that of deep hardware-efficient architectures, as QCNNs are restricted to be hierarchically structured and translationally invariant. The significant reduction in the expressivity and parameter space dimension from this translational invariance assumption yields the greater trainability [82].

The idea of embedding knowledge about the problem and dataset into our models (to achieve helpful inductive bias) will be key to improve the trainability of QML models. Recent proposals use Quantum Graph Neural Networks [91] for scenarios where quantum subsystems live on a graph, and potentially have further symmetries. For instance, the underlying graph-permutation symmetries of a quantum communication dataset were taken into account by a quantum graph convolutional network. Similarly, a quantum recurrent neural network has been used in scenarios where temporal recurrence of parameters occurs, e.g., as in the quantum dynamics of a stationary (time-dependent) quantum dynamical process.

To better understand how to go beyond the aforementioned inductive biases from temporal and/or translational invariance in grids and graphs, we can take inspiration from recent advances in the theory of classical deep learning. In classical ML, the study of the group theory behind graph neural networks, namely the concepts of invariance and equivariance to various group actions on the input space, has led to a unifying theory of deep learning architectures based on group theory, called Geometric Deep Learning theory [92].

In order to have a prescription to create arbitrary architectures and inductive biases suitable for a given set of quantum physical data, a theory of quantum geometric deep learning could be key to design architectures with the right prior over the transformation space and inductive biases to ensure trainability and generalization. As the study of physics is often about the identification of inherent or emergent symmetries in particular systems, there is great potential for a future unifying theory of

quantum geometric deep learning to provide consistent methods to create QML model architectures with inductive biases encoding knowledge of the basic symmetries and principles of the quantum physical system underlying given quantum datasets. This approach has been recently explored in [93–95]. Moreover, the works of [76, 96] have also shown that the Lie algebra obtained from the generators of the QNN can be linked to properties of the QML landscape such as the presence of barren plateaus or the overparametrization phenomenon.

D. Effect of quantum noise

The presence of hardware noise during quantum computations is one of the defining characteristics of Noisy Intermediate-Scale Quantum (NISQ) computing. Despite this fact, most QML research neglects noise in the analytical calculations and numerical simulations while still promising that the methods are near-term compatible. Accounting for the effects of hardware noise should be a crucial aspect of QML analysis if one wishes to pursue a quantum advantage with currently available hardware.

Noise corrupts the information as it forward propagates in a quantum circuit, meaning that deeper circuits with longer run-times will be particularly affected. As such, noise affects all aspects of the model that make use of quantum computers. This includes the dataset preparation scheme as well as circuits used to compute quantum kernels. Moreover, when using QNNs, noise can hinder their trainability as it leads to noise-induced barren plateaus [89, 97]. Here, the relevant features of the landscape get exponentially suppressed by noise as the depth of the circuit increases (see Fig. 5(d)). Ultimately, the effects of noise translate into a deformation of the inductive bias of the model from its original one, and an effective reduction of the dimension of the quantum feature space. Despite the critical impact of quantum noise, its effects are still largely unexplored, particularly on its impact on the classical simulability of the QML model [98, 99].

Addressing noise-induced issues will likely require either: (1) reduction in hardware error rates, (2) partial quantum error correction [100], or (3) employing QNNs that are relatively shallow (i.e., whose depth grows sub-linearly in the problem size) [89], such as QCNNs. Error mitigation techniques [101–103] can also improve performance of QML models in the presence of noise, although they may not solve noise-induced trainability issues [97]. A different approach to dealing with noise is to engineer QML models with noise-resilient properties [34, 104, 105] (such as the position of the minima not changing due to noise).

IV. OUTLOOK

A. Potential for Quantum advantage

The first quantum advantages in QML will likely come from hidden parameter extraction from quantum data. This can be for quantum sensing or quantum state classification/regression. Fundamentally, we know from the theory of optimal measurement that non-local quantum measurements can extract hidden parameters using less samples. Using QML, one can form and search over a parameterization of hypotheses for such measurements.

This is particularly useful when such optimal measurements are not known *a priori*, for example, identifying the measurement that extracts an order parameter or identifies a particular phase of matter. As the information about this classical parameter is embedded in the structure of quantum correlations between subsystems, it is natural that a trained QML model with good inductive biases can exhibit an advantage over local measurements and classical representations.

Another area of application where classical parameter extraction may yield an advantage is in quantum machine perception [46, 64, 106–109], i.e. quantum sensing, metrology, and beyond. Here, leveraging the variational search over multipartite-entangled states for input to exposure to a quantum signal along with the optimization for optimal control and/or over post-processing schemes can find optimal measurements for the estimation of hidden parameters in the incoming signal. In particular, the variational approach may be able to find the optimal entanglement, exposure, and measurement scheme which filters signal from the noise [110], akin to variationally learning the quantum error correcting code which filters signal from noise, instead applied to quantum metrology.

Beyond classical parameter extraction embedded in quantum data, there may be an advantage for the discovery of quantum error correcting codes (QECCs) [111]. QECC's fundamentally encode data (typically) non-locally into a subsystem or subspace of the Hilbert space. As deep learning is fundamentally about the discovery of submanifolds of data space, identifying and decoding subspaces/subsystems from a Hilbert space which correspond to a quantum error correction subspace/subsystem is a natural place where differentiable quantum computing may yield an advantage. This is a barely explored area, mainly due to the difficulty of gaining insights with small-scale numerical simulations. Fundamentally, it is akin to a quantum data version of classical parameter embedding/extraction advantage.

Finally, a quantum advantage for generative modelling may be achieved when one can generate ground states [112], equilibrium states [47, 113], or quantum dynamics [114], using generative models incorporating QNNs, in a way where the distribution cannot be sampled classically, and yields more accurate predictions or more extensively generalization compared to classical ML approaches. The nearest-term possibility for demonstrating such an ad-

vantage would likely be from variational optimization at the continuous time optimal control level on analogue quantum simulators.

B. What will quantum advantage look like?

When the data originates from quantum-mechanical processes, such as from experiments in chemistry, material science, biology, and physics, it is more likely to see exponential quantum advantage in ML. The quantum advantage could be in *sample complexity* or *time complexity*. An exponential advantage in sample complexity always implies an exponential advantage in time complexity, but the reverse is not generally true. It was recently shown [64, 66, 115, 116] that there is an exponential quantum advantage in sample complexity when we can use a quantum sensor, quantum memory, and quantum computer to retrieve, store, and process quantum information from experiments. Such a sample complexity advantage can be proven rigorously without the possibility of being *dequantized* [12, 65, 117] in the future, i.e., it is impossible to find improved classical algorithms such that there is no exponential advantage. This significant quantum advantage has recently been demonstrated on the Sycamore processor [64] raising the hope for achieving quantum advantage using NISQ devices [118].

The situation for advantage in *time complexity* is more subtle. Classical simulation of quantum process is intractable in many cases, hence one would expect exponential advantage in time complexity to be prevalent. However, one should be cautious about the availability of data in ML tasks, which makes classical ML algorithms computationally more powerful [13, 119]. For instance, Ref. [119] shows that *in the worst case*, there is no exponential quantum advantage in predicting ground state properties in geometrically local gapped Hamiltonians. Furthermore, the emergence of effective classical theory in quantum-mechanical processes could enable classical machines to provide accurate predictions. For example, density functional theory [120, 121] allows accurate prediction of molecular properties when we have an accurate approximation to the exchange-correlation functionals by conducting real-world experiments. It is still likely that an exponential advantage is possible in physical systems of practical interest, but there are no rigorous proofs yet.

When the data is of a purely classical origin, such as in applications for recommending products to customers [12], performing portfolio optimization [122, 123], and processing human languages [124] and everyday images [125], there is no known exponential advantage [117]. However, it is still reasonable to expect polynomial advantage. Furthermore, a quadratic advantage can be rigorously proven [126, 127] for purely classical problems. So we likely have a potential impact in the long-term when we have fault-tolerant quantum computers, albeit with the speedup significantly damped by the overheads of quantum error correction [128] for currently

known fault-tolerant quantum computing schemes.

C. Transition to the fault-tolerant era and beyond

While QML has been proposed as a candidate to achieve a quantum advantage in the near-term using NISQ devices, one can still pose the question about its usability in the future. Here, researchers envision two different chronological eras post-NISQ. In the first, which we can refer to as “partial error corrected”, quantum computers will have enough physical qubits (a couple of hundred of them), and sufficiently small error rates, to allow for a small number of fully error corrected logical qubits. Since one logical qubit is comprised of multiple physical qubits, in this era one will have the freedom to trade off and split the qubits in the device onto a subset of error corrected qubits, along with a subset of non-error corrected qubits. The next era, i.e., the “fault-tolerant era” will arise when the quantum hardware has a large number of error corrected qubits.

Indeed, one can easily envision QML being useful in both of these post-NISQ eras. First, in the partial error corrected era, QML models will be able to execute high-fidelity circuits and thus have an improved performance. This will naturally enhance the trainability of the models by mitigating noise-induced barren plateaus, and also reduce noise-induced classification errors in QML models. Most importantly, QML will likely see its most widespread and critical use during the fault-tolerant era. Here, quantum algorithms such as those for quantum simulation [129, 130] will be able to accurately prepare quantum data, and to faithfully store it in quantum memories [131]. Therefore QML will be the natural model to learn, infer, and make predictions from quantum data, as here the quantum computer will learn from the data itself directly.

On the further-term horizon, we anticipate it will be possible to capture quantum data from nature directly via transduction from its natural analog form to one that is quantum digital (e.g., via quantum analog-digital interconversion [19]). This data will then be able to be shuttled around quantum networks for distributed and/or centralized processing with quantum machine learning models, using fault-tolerant quantum computation and error-corrected quantum communication. At this point, quantum machine learning will have reached a stage similar to where machine learning is today, where edge sensors capture data, the data is relayed to a central cloud, and machine learning models are trained on the aggregated data. As the modern advent of widespread classical machine learning arose at this point of abundant data, one could anticipate that ubiquitous access to quantum data in the fault-tolerant era could similarly propel quantum machine learning to even greater widespread use.

ACKNOWLEDGEMENTS

MC acknowledges support from the Los Alamos National Laboratory (LANL) LDRD program under project number 20210116DR. MC was also supported by the Center for Nonlinear Studies at LANL. LC and PJC were supported by the U.S. Department of Energy (DOE), Office of Science, Office of Advanced Scientific Computing Research, under the Accelerated Research in Quantum Computing (ARQC) program. LC also acknowledges support from U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Science Center. PJC was also supported by the NNSA’s Advanced Simulation and Computing Beyond Moore’s Law Program at LANL. GV would like to thank Faris Sbahi, Antonio J. Martinez,

and Petar Velickovic for useful discussions. X, formerly known as Google[x], is part of the Alphabet family of companies, which includes Google, Verily, Waymo, and others (www.x.company). HH is supported by a Google PhD Fellowship.

AUTHOR CONTRIBUTIONS

PJC drafted the manuscript structure. The manuscript was written and revised by MC, GV, HYH, LC and PJC.

COMPETING INTERESTS

The authors declare no competing interests.

- [1] Michael A. Nielsen and Isaac L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).
- [2] Jennifer C Brookes, “Quantum effects in biology: golden rule in enzymes, olfaction, photosynthesis and magnetodetection,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **473**, 20160822 (2017).
- [3] David Deutsch, “Quantum theory, the church-turing principle and the universal quantum computer,” *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **400**, 97–117 (1985).
- [4] Nathan Wiebe, Ashish Kapoor, and Krysta M Svore, “Quantum deep learning,” *arXiv preprint arXiv:1412.3489* (2014).
- [5] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione, “An introduction to quantum machine learning,” *Contemporary Physics* **56**, 172–185 (2015).
- [6] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd, “Quantum machine learning,” *Nature* **549**, 195–202 (2017).
- [7] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G. S. L. Brandao, David A. Buell, Brian Burkett, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Andrew Dunsworth, Edward Farhi, Brooks Foxen, Austin Fowler, Craig Gidney, Marissa Giustina, Rob Graff, Keith Guerin, Steve Habegger, Matthew P. Harrigan, Michael J. Hartmann, Alan Ho, Markus Hoffmann, Trent Huang, Travis S. Humble, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Paul V. Klimov, Sergey Knysh, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Mike Lindmark, Erik Lucero, Dmitry Lyakh, Salvatore Mandrà, Jarrod R. McClean, Matthew McEwen, Anthony Megrant, Xiao Mi, Kristel Michelsen, Masoud Mohseni, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Murphy Yuezen Niu, Eric Ostby, Andre Petukhov, John C. Platt, Chris Quintana, Eleanor G. Rieffel, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Kevin J. Sung, Matthew D. Trevithick, Amit Vainsencher, Benjamin Villalonga, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven, and John M. Martinis, “Quantum supremacy using a programmable superconducting processor,” *Nature* **574**, 505–510 (2019).
- [8] Lukasz Cincio, Yiğit Subaşı, Andrew T Sornborger, and Patrick J Coles, “Learning the quantum algorithm for state overlap,” *New Journal of Physics* **20**, 113022 (2018).
- [9] Aaron D Tranter, Harry J Slatyer, Michael R Hush, Anthony C Leung, Jesse L Everett, Karun V Paul, Pierre Vernaz-Gris, Ping Koy Lam, Ben C Buchler, and Geoff T Campbell, “Multiparameter optimisation of a magneto-optical trap using deep learning,” *Nature communications* **9**, 1–8 (2018).
- [10] Raphael Kaubruegger, Denis V Vasilyev, Marius Schulte, Clemens Hammerer, and Peter Zoller, “Quantum variational optimization of ramsey interferometry and atomic clocks,” *Physical Review X* **11**, 041045 (2021).
- [11] Iris Cong, Soonwon Choi, and Mikhail D Lukin, “Quantum convolutional neural networks,” *Nature Physics* **15**, 1273–1278 (2019).
- [12] Ewin Tang, “A quantum-inspired classical algorithm for recommendation systems,” in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (2019) pp. 217–228.
- [13] Hsin-Yuan Huang, Michael Broughton, Masoud Mohseni, Ryan Babbush, Sergio Boixo, Hartmut Neven, and Jarrod R McClean, “Power of data in quantum machine learning,” *Nature Communications* **12**, 1–9 (2021).
- [14] Leonardo Banchi, Jason Pereira, and Stefano Pirandola, “Generalization in quantum machine learning: A quantum information standpoint,” *PRX Quantum* **2**, 040321 (2021).

- [15] C. L. Degen, F. Reinhard, and P. Cappellaro, “Quantum sensing,” *Rev. Mod. Phys.* **89**, 035002 (2017).
- [16] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone, “Advances in quantum metrology,” *Nat. Photonics* **5**, 222–229 (2011).
- [17] Giulio Chiribella, Giacomo Mauro D’Ariano, and Paolo Perinotti, “Theoretical framework for quantum networks,” *Physical Review A* **80**, 022339 (2009).
- [18] D. D’Alessandro, *Introduction to Quantum Control and Dynamics*, Chapman & Hall/CRC Applied Mathematics & Nonlinear Science (Taylor & Francis, 2007).
- [19] Guillaume Verdon-Akzam, “Quantum Analog-digital Interconversion For Encoding And Decoding Quantum Signals,” (2020), United States Patent Application No. 17063595.
- [20] Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd, “Quantum support vector machine for big data classification,” *Physical review letters* **113**, 130503 (2014).
- [21] Maria Schuld and Nathan Killoran, “Quantum machine learning in feature hilbert spaces,” *Physical review letters* **122**, 040504 (2019).
- [22] Seth Lloyd, Maria Schuld, Aroosa Ijaz, Josh Izaac, and Nathan Killoran, “Quantum embeddings for machine learning,” *arXiv preprint arXiv:2001.03622* (2020).
- [23] Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer, “Effect of data encoding on the expressive power of variational quantum-machine-learning models,” *Physical Review A* **103**, 032430 (2021).
- [24] Joschka Roffe, “Quantum error correction: an introductory guide,” *Contemporary Physics* **60**, 226–245 (2019).
- [25] Peter W Shor, “Scheme for reducing decoherence in quantum computer memory,” *Physical review A* **52**, R2493 (1995).
- [26] Vojtěch Havlíček, Antonio D Cárcoles, Kristan Temme, Aram W Harrow, Abhinav Kandala, Jerry M Chow, and Jay M Gambetta, “Supervised learning with quantum-enhanced feature spaces,” *Nature* **567**, 209–212 (2019).
- [27] Yunchao Liu, Srinivasan Arunachalam, and Kristan Temme, “A rigorous and robust quantum speed-up in supervised machine learning,” *Nature Physics* , 1–5 (2021).
- [28] Maria Schuld, “Quantum machine learning models are kernel methods,” *arXiv preprint arXiv:2101.11020* (2021).
- [29] J. S. Otterbach, R. Manenti, N. Alidoust, A. Bestwick, M. Block, B. Bloom, S. Caldwell, N. Didier, E. Schuyler Fried, S. Hong, P. Karalekas, C. B. Osborn, A. Papa-george, E. C. Peterson, G. Prawiroatmodjo, N. Rubin, Colm A. Ryan, D. Scarabelli, M. Scheer, E. A. Sete, P. Sivarajah, Robert S. Smith, A. Staley, N. Tezak, W. J. Zeng, A. Hudson, Blake R. Johnson, M. Reagor, M. P. da Silva, and C. Rigetti, “Unsupervised machine learning on a hybrid quantum computer,” *arXiv preprint arXiv:1712.05771* (2017).
- [30] Iordanis Kerenidis, Jonas Landman, Alessandro Luongo, and Anupam Prakash, “q-means: A quantum algorithm for unsupervised machine learning,” *Advances in Neural Information Processing Systems* **32** (2019).
- [31] V. Saggio, B. E. Asenbeck, A. Hamann, T. Strömbärg, P. Schiansky, V. Dunjko, N. Friis, N. C. Harris, M. Hochberg, D. Englund, S. Wölk, H. J. Briegel, and P. Walther, “Experimental quantum speed-up in reinforcement learning agents,” *Nature* **591**, 229–233 (2021).
- [32] Andrea Skolik, Sofiene Jerbi, and Vedran Dunjko, “Quantum agents in the gym: a variational quantum algorithm for deep q-learning,” *arXiv preprint arXiv:2103.15084* (2021).
- [33] Hsin-Yuan Huang, Michael Broughton, Jordan Cotler, Sitan Chen, Jerry Li, Masoud Mohseni, Hartmut Neven, Ryan Babbush, Richard Kueng, John Preskill, and Jarrod R. McClean, “Quantum advantage in learning from experiments,” *arXiv preprint arXiv:2112.00778* (2021).
- [34] Ryan LaRose and Brian Coyle, “Robust data encodings for quantum classifiers,” *Physical Review A* **102**, 032420 (2020).
- [35] Matthias C. Caro, Hsin-Yuan Huang, M. Cerezo, Kunal Sharma, Andrew Sornborger, Lukasz Cincio, and Patrick J. Coles, “Generalization in quantum machine learning from few training data,” *arXiv preprint arXiv:2111.05292* (2021).
- [36] Matthias C. Caro, Hsin-Yuan Huang, Nicholas Ezzell, Joe Gibbs, Andrew T. Sornborger, Lukasz Cincio, Patrick J. Coles, and Zoe Holmes, “Out-of-distribution generalization for learning quantum dynamics,” *arXiv preprint arXiv:2204.10268*.
- [37] Matthias C. Caro, Elies Gil-Fuster, Johannes Jakob Meyer, Jens Eisert, and Ryan Sweke, “Encoding-dependent generalization bounds for parametrized quantum circuits,” *Quantum* **5**, 582 (2021).
- [38] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles, “Variational quantum algorithms,” *Nature Reviews Physics* **3**, 625–644 (2021).
- [39] Kwok Ho Wan, Oscar Dahlsten, Hlér Kristjánsson, Robert Gardner, and MS Kim, “Quantum generalisation of feedforward neural networks,” *npj Quantum information* **3**, 1–8 (2017).
- [40] Kerstin Beer, Dmytro Bondarenko, Terry Farrelly, Tobias J. Osborne, Robert Salzmann, Daniel Scheiermann, and Ramona Wolf, “Training deep quantum neural networks,” *Nature Communications* **11**, 808 (2020).
- [41] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione, “The quest for a quantum neural network,” *Quantum Information Processing* **13**, 2567–2586 (2014).
- [42] Pierre-Luc Dallaire-Demers and Nathan Killoran, “Quantum generative adversarial networks,” *Physical Review A* **98**, 012324 (2018).
- [43] Edward Farhi and Hartmut Neven, “Classification with quantum neural networks on near term processors,” *arXiv preprint arXiv:1802.06002* (2018).
- [44] Nathan Killoran, Thomas R Bromley, Juan Miguel Arrazola, Maria Schuld, Nicolás Quesada, and Seth Lloyd, “Continuous-variable quantum neural networks,” *Physical Review Research* **1**, 033063 (2019).
- [45] Johannes Bausch, “Recurrent quantum neural networks,” *arXiv preprint arXiv:2006.14619* (2020).
- [46] Michael Broughton, Guillaume Verdon, Trevor McCourt, Antonio J. Martinez, Jae Hyeon Yoo, Sergei V. Isakov, Philip Massey, Ramin Halavati, Murphy Yuezhen Niu, Alexander Zlokapa, Evan Peters, Owen Lockwood, Andrea Skolik, Sofiene Jerbi, Vedran Dunjko, Martin Leib, Michael Streif, David Von Dollen, Hongxiang Chen, Shuxiang Cao, Roeland Wiersema, Hsin-Yuan Huang, Jarrod R. McClean, Ryan Babbush, Sergio Boixo, Dave Bacon, Alan K. Ho, Hartmut Neven,

- and Masoud Mohseni, “Tensorflow quantum: A software framework for quantum machine learning,” *arXiv preprint arXiv:2003.02989* (2020).
- [47] Guillaume Verdon, Jacob Marks, Sasha Nanda, Stefan Leichenauer, and Jack Hidary, “Quantum Hamiltonian-based models and the variational quantum thermalizer algorithm,” *arXiv preprint arXiv:1910.02071* (2019).
- [48] Corinna Cortes and Vladimir Vapnik, “Support-vector networks,” *Mach. Learn.* **20**, 273–297 (1995).
- [49] Jonas M Kübler, Simon Buchholz, and Bernhard Schölkopf, “The inductive bias of quantum kernels,” *arXiv preprint arXiv:2106.03747* (2021).
- [50] This is due to the fact that extracting information from a quantum state requires computing the expectation values of some observable, which in practice need to be estimated via measurements on a noisy quantum computer. Hence, given a finite number of shots (measurement repetitions), these can only be resolved up to some additive errors. Moreover, such expectation values will be subject to corruption due to hardware noise.
- [51] Jonas M Kübler, Andrew Arrasmith, Lukasz Cincio, and Patrick J Coles, “An adaptive optimizer for measurement-frugal variational algorithms,” *Quantum* **4**, 263 (2020).
- [52] Andrew Arrasmith, Lukasz Cincio, Rolando D Somma, and Patrick J Coles, “Operator sampling for shot-frugal optimization in variational algorithms,” *arXiv preprint arXiv:2004.06252* (2020).
- [53] Andi Gu, Angus Lowe, Pavel A Dub, Patrick J Coles, and Andrew Arrasmith, “Adaptive shot allocation for fast convergence in variational quantum algorithms,” *arXiv preprint arXiv:2108.10434* (2021).
- [54] Ryan Sweke, Frederik Wilde, Johannes Jakob Meyer, Maria Schuld, Paul K Fährmann, Barthélémy Meynard-Piganeau, and Jens Eisert, “Stochastic gradient descent for hybrid quantum-classical optimization,” *Quantum* **4**, 314 (2020).
- [55] James Stokes, Josh Izaac, Nathan Killoran, and Giuseppe Carleo, “Quantum natural gradient,” *Quantum* **4**, 269 (2020).
- [56] Bálint Koczor and Simon C Benjamin, “Quantum natural gradient generalised to non-unitary circuits,” *arXiv preprint arXiv:1912.08660* (2019).
- [57] Kunal Sharma, M. Cerezo, Zoë Holmes, Lukasz Cincio, Andrew Sornborger, and Patrick J Coles, “Reformulation of the no-free-lunch theorem for entangled data sets,” *arXiv preprint arXiv:2007.04900* (2020).
- [58] Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner, “The power of quantum neural networks,” *Nature Computational Science* **1**, 403–409 (2021).
- [59] Frank Rosenblatt, *The perceptron, a perceiving and recognizing automaton Project Para* (Cornell Aeronautical Laboratory, 1957).
- [60] Simon Haykin, *Neural networks: a comprehensive foundation* (Prentice Hall PTR, 1994).
- [61] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams, “Learning representations by backpropagating errors,” *nature* **323**, 533–536 (1986).
- [62] Thomas Hubregtsen, David Wierichs, Elies Gil-Fuster, Peter-Jan HS Derkx, Paul K Faehrmann, and Johannes Jakob Meyer, “Training quantum embedding kernels on near-term quantum computers,” *arXiv preprint arXiv:2105.02276* (2021).
- [63] Supanut Thanasilp, Samson Wang, Nhat A Nghiem, Patrick J Coles, and M Cerezo, “Subtleties in the trainability of quantum machine learning models,” *arXiv preprint arXiv:2110.14753* (2021).
- [64] Hsin-Yuan Huang, Michael Broughton, Jordan Cotler, Sitam Chen, Jerry Li, Masoud Mohseni, Hartmut Neven, Ryan Babbush, Richard Kueng, John Preskill, and Jarrod R. McClean, “Quantum advantage in learning from experiments,” *Science* **376**, 1182–1186 (2022).
- [65] Jordan Cotler, Hsin-Yuan Huang, and Jarrod R McClean, “Revisiting dequantization and quantum advantage in learning tasks,” *arXiv preprint arXiv:2112.00811* (2021).
- [66] Sitam Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li, “A hierarchy for replica quantum advantage,” *arXiv preprint arXiv:2111.05874* (2021).
- [67] Sitam Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li, “Exponential separations between learning with and without quantum memory,” *arXiv preprint arXiv:2111.05881* (2021).
- [68] Elija Perrier, Akram Youssry, and Chris Ferrie, “Qdataset: Quantum datasets for machine learning,” *arXiv preprint arXiv:2108.06661* (2021).
- [69] Louis Schatzki, Andrew Arrasmith, Patrick J. Coles, and M. Cerezo, “Entangled datasets for quantum machine learning,” *arXiv preprint arXiv:2109.03400* (2021).
- [70] Technically speaking, the loss function defines a map from the model’s parameter space to the real values. The loss function value can quantify, for instance, the model’s error in solving a given so that our goal is to find the set of parameters that minimizes such error.
- [71] Andrew Arrasmith, Zoë Holmes, M Cerezo, and Patrick J Coles, “Equivalence of quantum barren plateaus to cost concentration and narrow gorges,” *arXiv preprint arXiv:2104.05868* (2021).
- [72] Lennart Bittel and Martin Kliesch, “Training variational quantum algorithms is np-hard,” *Phys. Rev. Lett.* **127**, 120502 (2021).
- [73] M Bilakis, M Cerezo, Guillaume Verdon, Patrick J Coles, and Lukasz Cincio, “A semi-agnostic ansatz with variable structure for quantum machine learning,” *arXiv preprint arXiv:2103.06712* (2021).
- [74] Ryan LaRose, Arkin Tikku, Étude O’Neil-Judy, Lukasz Cincio, and Patrick J Coles, “Variational quantum state diagonalization,” *npj Quantum Information* **5**, 1–10 (2019).
- [75] Bobak Toussi Kiani, Seth Lloyd, and Reevu Maity, “Learning unitaries by gradient descent,” *arXiv preprint arXiv:2001.11897* (2020).
- [76] Martin Larocca, Nathan Ju, Diego García-Martín, Patrick J. Coles, and M. Cerezo, “Theory of overparametrization in quantum neural networks,” *arXiv preprint arXiv:2109.11676* (2021).
- [77] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven, “Barren plateaus in quantum neural network training landscapes,” *Nature Communications* **9**, 1–6 (2018).
- [78] M Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles, “Cost function dependent barren plateaus in shallow parametrized quantum circuits,” *Nature Communications* **12**, 1–12 (2021).
- [79] M. Cerezo and Patrick J Coles, “Higher order derivatives of quantum neural networks with barren plateaus,”

- Quantum Science and Technology** **6**, 035006 (2021).
- [80] Andrew Arrasmith, M. Cerezo, Piotr Czarnik, Lukasz Cincio, and Patrick J Coles, “Effect of barren plateaus on gradient-free optimization,” *Quantum* **5**, 558 (2021).
 - [81] Zoë Holmes, Kunal Sharma, M. Cerezo, and Patrick J Coles, “Connecting ansatz expressibility to gradient magnitudes and barren plateaus,” *PRX Quantum* **3**, 010313 (2022).
 - [82] Arthur Pesah, M. Cerezo, Samson Wang, Tyler Volkoff, Andrew T Sornborger, and Patrick J Coles, “Absence of barren plateaus in quantum convolutional neural networks,” *Physical Review X* **11**, 041011 (2021).
 - [83] Tyler Volkoff and Patrick J Coles, “Large gradients via correlation in random parameterized quantum circuits,” *Quantum Science and Technology* **6**, 025008 (2021).
 - [84] Kunal Sharma, Marco Cerezo, Lukasz Cincio, and Patrick J Coles, “Trainability of dissipative perceptron-based quantum neural networks,” *Physical Review Letters* **128**, 180505 (2022).
 - [85] Zoë Holmes, Andrew Arrasmith, Bin Yan, Patrick J Coles, Andreas Albrecht, and Andrew T Sornborger, “Barren plateaus preclude learning scramblers,” *Physical Review Letters* **126**, 190501 (2021).
 - [86] Carlos Ortiz Marrero, Maria Kieferova, and Nathan Wiebe, “Entanglement induced barren plateaus,” *arXiv preprint arXiv:2010.15968* (2020).
 - [87] AV Uvarov and Jacob D Biamonte, “On barren plateaus and cost function locality in variational quantum algorithms,” *Journal of Physics A: Mathematical and Theoretical* **54**, 245301 (2021).
 - [88] Taylor L Patti, Khadijeh Najafi, Xun Gao, and Susanne F Yelin, “Entanglement devised barren plateau mitigation,” *Physical Review Research* **3**, 033090 (2021).
 - [89] Samson Wang, Enrico Fontana, Marco Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J Coles, “Noise-induced barren plateaus in variational quantum algorithms,” *Nature Communications* **12**, 1–11 (2021).
 - [90] Guillaume Verdon, Michael Broughton, Jarrod R McClean, Kevin J Sung, Ryan Babbush, Zhang Jiang, Hartmut Neven, and Masoud Mohseni, “Learning to learn with quantum neural networks via classical neural networks,” *arXiv preprint arXiv:1907.05415* (2019).
 - [91] Guillaume Verdon, Trevor McCourt, Enxhell Luzhnica, Vikash Singh, Stefan Leichenauer, and Jack Hidary, “Quantum graph neural networks,” *arXiv preprint arXiv:1909.12264* (2019).
 - [92] Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković, “Geometric deep learning: Grids, groups, graphs, geodesics, and gauges,” *arXiv preprint arXiv:2104.13478* (2021).
 - [93] Martin Larocca, Frederic Sauvage, Faris M Sbahi, Guillaume Verdon, Patrick J Coles, and M Cerezo, “Group-invariant quantum machine learning,” *arXiv preprint arXiv:2205.02261* (2022).
 - [94] Andrea Skolik, Michele Cattelan, Sheir Yarkoni, Thomas Bäck, and Vedran Dunjko, “Equivariant quantum circuits for learning on weighted graphs,” *arXiv preprint arXiv:2205.06109* (2022).
 - [95] Johannes Jakob Meyer, Marian Mularski, Elies Gil-Fuster, Antonio Anna Mele, Francesco Arzani, Alissa Wilms, and Jens Eisert, “Exploiting symmetry in variational quantum machine learning,” *arXiv preprint arXiv:2205.06217* (2022).
 - [96] Martin Larocca, Piotr Czarnik, Kunal Sharma, Gopikrishnan Muraleedharan, Patrick J. Coles, and M. Cerezo, “Diagnosing barren plateaus with tools from quantum optimal control,” *arXiv preprint arXiv:2105.14377* (2021).
 - [97] Samson Wang, Piotr Czarnik, Andrew Arrasmith, M Cerezo, Lukasz Cincio, and Patrick J Coles, “Can error mitigation improve trainability of noisy variational quantum algorithms?” *arXiv preprint arXiv:2109.01051* (2021).
 - [98] Abhinav Deshpande, Bill Fefferman, Alexey V Gorshkov, Michael J Gullans, Pradeep Niroula, and Oles Shtanko, “Tight bounds on the convergence of noisy random circuits to uniform,” *arXiv preprint arXiv:2112.00716* (2021).
 - [99] Shigeo Hakkaku, Yuichiro Tashima, Kosuke Mitarai, Wataru Mizukami, and Keisuke Fujii, “Quantifying fermionic nonlinearity of quantum circuits,” *arXiv preprint arXiv:2111.14599* (2021).
 - [100] Daniel Bultrini, Samson Wang, Piotr Czarnik, Max Hunter Gordon, M Cerezo, Patrick J Coles, and Lukasz Cincio, “The battle of clean and dirty qubits in the era of partial error correction,” *arXiv preprint arXiv:2205.13454* (2022).
 - [101] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta, “Error mitigation for short-depth quantum circuits,” *Phys. Rev. Lett.* **119**, 180509 (2017).
 - [102] Piotr Czarnik, Andrew Arrasmith, Patrick J. Coles, and Lukasz Cincio, “Error mitigation with Clifford quantum-circuit data,” *Quantum* **5**, 592 (2021).
 - [103] Suguru Endo, Zhenyu Cai, Simon C Benjamin, and Xiao Yuan, “Hybrid quantum-classical algorithms and quantum error mitigation,” *Journal of the Physical Society of Japan* **90**, 032001 (2021).
 - [104] Kunal Sharma, Sumeet Khatri, M. Cerezo, and Patrick J Coles, “Noise resilience of variational quantum compiling,” *New Journal of Physics* **22**, 043006 (2020).
 - [105] Lukasz Cincio, Kenneth Rudinger, Mohan Sarovar, and Patrick J. Coles, “Machine learning of noise-resilient quantum circuits,” *PRX Quantum* **2**, 010324 (2021).
 - [106] Alan Ho, Guillaume Verdon, and Masoud Mohseni, “Quantum Machine Perception,” (2020), United States Patent Application No. 17019564.
 - [107] Johannes Jakob Meyer, Johannes Borregaard, and Jens Eisert, “A variational toolbox for quantum multi-parameter estimation,” *NPJ Quantum Information* **7**, 1–5 (2021).
 - [108] Jacob L Beckey, M Cerezo, Akira Sone, and Patrick J Coles, “Variational quantum algorithm for estimating the quantum Fisher information,” *Physical Review Research* **4**, 013083 (2022).
 - [109] Jianwei Wang, Stefano Paesani, Raffaele Santagati, Sebastian Knauer, Antonio A. Gentile, Nathan Wiebe, Maurangelo Petruzzella, Jeremy L. O’Brien, John G. Rarity, Anthony Laing, and Mark G. Thompson, “Experimental quantum Hamiltonian learning,” *Nature Physics* **13**, 551–555 (2017).
 - [110] David Layden and Paola Cappellaro, “Spatial noise filtering through error correction for quantum sensing,” *npj Quantum Information* **4**, 1–6 (2018).
 - [111] Peter D Johnson, Jonathan Romero, Jonathan Olson, Yudong Cao, and Alán Aspuru-Guzik, “Qvector: an algorithm for device-tailored quantum error correction,” *arXiv preprint arXiv:1711.02249* (2017).

- [112] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O'brien, “A variational eigenvalue solver on a photonic quantum processor,” *Nature communications* **5**, 1–7 (2014).
- [113] Sam McArdle, Tyson Jones, Suguru Endo, Ying Li, Simon C Benjamin, and Xiao Yuan, “Variational ansatz-based quantum simulation of imaginary time evolution,” *npj Quantum Information* **5**, 1–6 (2019).
- [114] Cristina Cirstoiu, Zoe Holmes, Joseph Iosue, Lukasz Cincio, Patrick J Coles, and Andrew Sornborger, “Variational fast forwarding for quantum simulation beyond the coherence time,” *npj Quantum Information* **6**, 1–10 (2020).
- [115] Hsin-Yuan Huang, Richard Kueng, and John Preskill, “Information-theoretic bounds on quantum advantage in machine learning,” *Phys. Rev. Lett.* **126**, 190505 (2021).
- [116] Dorit Aharonov, Jordan Cotler, and Xiao-Liang Qi, “Quantum algorithmic measurement,” *arXiv preprint arXiv:2101.04634* (2021).
- [117] Nai-Hui Chia, András Gilyén, Tongyang Li, Han-Hsuan Lin, Ewin Tang, and Chunhao Wang, “Sampling-based sublinear low-rank matrix arithmetic framework for dequantizing quantum machine learning,” in *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing* (2020) pp. 387–400.
- [118] John Preskill, “Quantum computing in the nisq era and beyond,” *Quantum* **2**, 79 (2018).
- [119] Hsin-Yuan Huang, Richard Kueng, Giacomo Torlai, Victor V. Albert, and John Preskill, “Provably efficient machine learning for quantum many-body problems,” *arXiv preprint arXiv:2106.12627* (2021).
- [120] P. Hohenberg and W. Kohn, “Inhomogeneous electron gas,” *Phys. Rev.* **136**, B864–B871 (1964).
- [121] W. Kohn, “Nobel lecture: Electronic structure of matter—wave functions and density functionals,” *Rev. Mod. Phys.* **71**, 1253–1266 (1999).
- [122] Javier Alcazar, Vicente Leyton-Ortega, and Alejandro Perdomo-Ortiz, “Classical versus quantum models in machine learning: insights from a finance application,” *Machine Learning: Science and Technology* **1**, 035003 (2020).
- [123] Adam Bouland, Wim van Dam, Hamed Joorati, Iordanis Kerenidis, and Anupam Prakash, “Prospects and challenges of quantum finance,” *arXiv preprint arXiv:2011.06492* (2020).
- [124] Christopher Manning and Hinrich Schütze, *Foundations of statistical natural language processing* (MIT press, 1999).
- [125] John C Russ, *The image processing handbook* (CRC press, 2006).
- [126] Lov K Grover, “A fast quantum mechanical algorithm for database search,” in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing* (1996) pp. 212–219.
- [127] Ethan Bernstein and Umesh Vazirani, “Quantum complexity theory,” *SIAM Journal on computing* **26**, 1411–1473 (1997).
- [128] Ryan Babbush, Jarrod R McClean, Michael Newman, Craig Gidney, Sergio Boixo, and Hartmut Neven, “Focus beyond quadratic speedups for error-corrected quantum advantage,” *PRX Quantum* **2**, 010103 (2021).
- [129] Iulia M Georgescu, Sahel Ashhab, and Franco Nori, “Quantum simulation,” *Reviews of Modern Physics* **86**, 153 (2014).
- [130] Dominic W Berry, Andrew M Childs, Richard Cleve, Robin Kothari, and Rolando D Somma, “Simulating hamiltonian dynamics with a truncated taylor series,” *Physical Review Letters* **114**, 090502 (2015).
- [131] Alexander I Lvovsky, Barry C Sanders, and Wolfgang Tittel, “Optical quantum memory,” *Nature photonics* **3**, 706–714 (2009).
- [132] Benjamin Sanchez-Lengeling and Alán Aspuru-Guzik, “Inverse molecular design using machine learning: Generative models for matter engineering,” *Science* **361**, 360–365 (2018).