Parameterized quantum circuits as machine learning models

Marcello Benedetti,[1,2,*] Erika Lloyd,[1] and Stefan Sack[1]

[1] *Cambridge Quantum Computing Limited, CB2 1UB Cambridge, United Kingdom*
[2] *Department of Computer Science, University College London, WC1E 6BT London, United Kingdom*

(Dated: Jun 18, 2019)

Hybrid quantum-classical systems make it possible to utilize existing quantum computers to their fullest extent. Within this framework, parameterized quantum circuits can be thought of as machine learning models with remarkable expressive power. This Review presents components of these models and discusses their application to a variety of data-driven tasks such as supervised learning and generative modeling. With experimental demonstrations carried out on actual quantum hardware, and with software actively being developed, this rapidly growing field could become one of the first instances of quantum computing that addresses real world problems.

I. INTRODUCTION

Developments in material science, hardware manufacturing, and disciplines such as error-correction and compilation, have brought us one step closer to the era of large-scale, fault-tolerant, universal quantum computation. However, the process is incremental and may take years; existing quantum hardware implements few physical qubits and can perform short sequences of gates before sources of noise take over. In such a setting, much anticipated algorithms such as Shor's are way out of reach. Yet it has been argued that noisy intermediate-scale quantum (NISQ) devices may find useful applications and commercialization in the next few years [1, 2]. As prototypes of quantum computers are made available to researchers for experimentation, algorithmic development is indeed adapting to the pace at which quantum hardware is developed.

Parameterized quantum circuits (PQCs) offer a concrete way to implement algorithms and demonstrate quantum supremacy in the NISQ era. PQCs are typically composed of fixed gates, e.g., controlled NOTs, and adjustable gates, e.g., qubit rotations. Even at low circuit depth, some classes of PQCs are capable of generating highly non-trivial outputs. For example, under well-believed complexity-theoretic assumptions, the class of PQCs called instantaneous quantum polynomial-time (IQP) cannot be efficiently simulated by classical resources (see Lund *et al.* [3] and Harrow and Montanaro [4] for accessible Reviews of quantum supremacy proposals). The demonstration of quantum supremacy is an important milestone in the development of quantum computers. In practice, however, we would like to obtain a quantum advantage over classical computing while attacking real-world problems.

The main approach taken by the community consists of formalizing problems of interest as variational optimization problems and use a combination of quantum and classical hardware to find approximate solutions. The intuition is that by outsourcing parts of the algorithm to classical hardware, we significantly reduce the burden on the quantum hardware. In particular, we reduce the required coherence time, circuit depth and number of qubits, hence allowing NISQ hardware to focus entirely on the computationally hard part of the problem.

This hybrid algorithmic approach turned out to be successful in attacking scaled-down problems in chemistry, combinatorial optimization and machine learning. For example, the variational quantum eigensolver (VQE) [5] has been used for searching the ground state of the electronic Hamiltonian of molecules [6, 7]. Similarly, the quantum approximate optimization algorithm (QAOA) [8] has been used to find approximate solutions of classical Ising models [9] and clustering problems formulated as MaxCut [10].

The focus of this Review is on hybrid approaches for machine learning (ML). Our aim is to present the ideas in a pedagogical and non-technical way, although we assume familiarity with basic ML definitions and methods (see Mehta *et al.* [11] for a physics-oriented introduction to ML), and basic working knowledge on quantum computing (see Nielsen and Chuang [12], Chapter 2, for an introduction). In this field, quantum circuits are seen as components of a *model* for some data-driven task. *Learning* is the process of iteratively updating the set of parameters in the model towards the goal. In practice, one obtains a pool of models and selects the one attaining low error on a test data. The selected model is then used in a production environment such as an automated forecasting and decision-making system.

The approach is illustrated in Fig. 1 and is made of three main components: the human, the classical computer, and the quantum computer. The human interprets the problem information and selects an initial model to represent it. The available data is pre-processed on a classical computer to determine a set of parameters for the PQC. The quantum hardware prepares a quantum state as prescribed by a PQC and performs measurements. Measurement outcomes are post-processed by the classical computer to generate a forecast. To improve the forecast, the classical computer implements a learning algorithm that updates the set of parameters. The overall algorithm is run in a closed loop between the classical and

* marcello.benedetti@cambridgequantum.com

quantum devices which comprise the hybrid system. The human supervises the process and uses forecasts towards the goal.

To the best of our knowledge, the earliest hybrid systems were proposed in the context of quantum algorithm learning. Bang *et al.* [13] describe a method where a classical computer controls the unitary operation implemented by a quantum device. Each execution of the quantum device is deemed as either a 'success' or 'failure', and the classical algorithm adjusts the unitary operation towards its target. Starting from a dataset of input-output pairs their simulated system learns an equivalent of Deutsch's algorithm for finding whether a function is constant or balanced. Gammelmark and Mølmer [14] took a more general approach where the parameters of the quantum system are quantized as well. In their simulations they successfully learn Grover's and Shor's algorithms.

These early proposals attacked problems that are well known within the quantum computing community, but much less known among ML researchers. More recently though, the hybrid approach based on PQCs has been shown to perform well on ML tasks such as classification, regression, and generative modeling, and has been applied to both classical and quantum data of small scale. The success is in part due to some similarities between PQCs and celebrated classical models such as kernel methods and artificial neural networks (ANNs). In this Review we provide an introduction to many of these bridging ideas, and we direct Readers interested to in-depth exploration towards the relevant literature. We make use of several acronyms when referring to models and algorithms; to help the Reader we summarize all the acronyms in Table I.

This novel field has not been restricted to theory and simulation. A series of experimental demonstrations on scaled-down problems have been performed in the past two years. In Table II we summarize the relevant demonstrations, and the Reader interested in experimental setups is invited to delve into the references therein.

The software side is also moving at fast pace (see Fingerhuth *et al.* [15] for a Review of general quantum computing software). There now exist several platforms for hybrid quantum-classical computation which are specifically dedicated to ML. They provide PQC models, automatic differentiation techniques, and interfaces to both simulators and existing quantum computers. This enables experimentation at a much higher rate than previously possible, a scenario reminiscent of the deep learning developments a decade ago. We summarize the relevant open-source software in Table III; we do not claim to be comprehensive since software packages evolve at a very fast pace.

The structure of the Review is as follows: in Section II we describe the components of ML models based on PQCs and their learning algorithms; in Section III we describe their applications to classical and quantum tasks; and in Section IV we summarize the advantages of this approach and give an outlook of the field.

Adam	adaptive moment estimation
ANN	artificial neural network
BO	Bayesian optimization
GA	genetic algorithm
GD	gradient descent
IQP	instantaneous quantum polynomial-time
MERA	multi-scale entanglement renormalization ansatz
ML	machine learning
MPS	matrix product state
NISQ	noisy intermediate-scale quantum
PAC	probably approximately correct
PQC	parameterized quantum circuit
PSO	particle swarm optimization
QAE	quantum autoencoder
QAOA	quantum approximate optimization algorithm
QCBM	quantum circuit Born machine
QKE	quantum kernel estimator
QGAN	quantum generative adversarial network
SHC	stochastic hill-climbing
SPSA	simultaneous perturbation stochastic approximation
SVM	support vector machine
TTN	tree tensor network
VQM	variational quantum model
VQE	variational quantum eigensolver
ZOOpt	zeroth-order optimization package

TABLE I. Acronyms used in this Review.

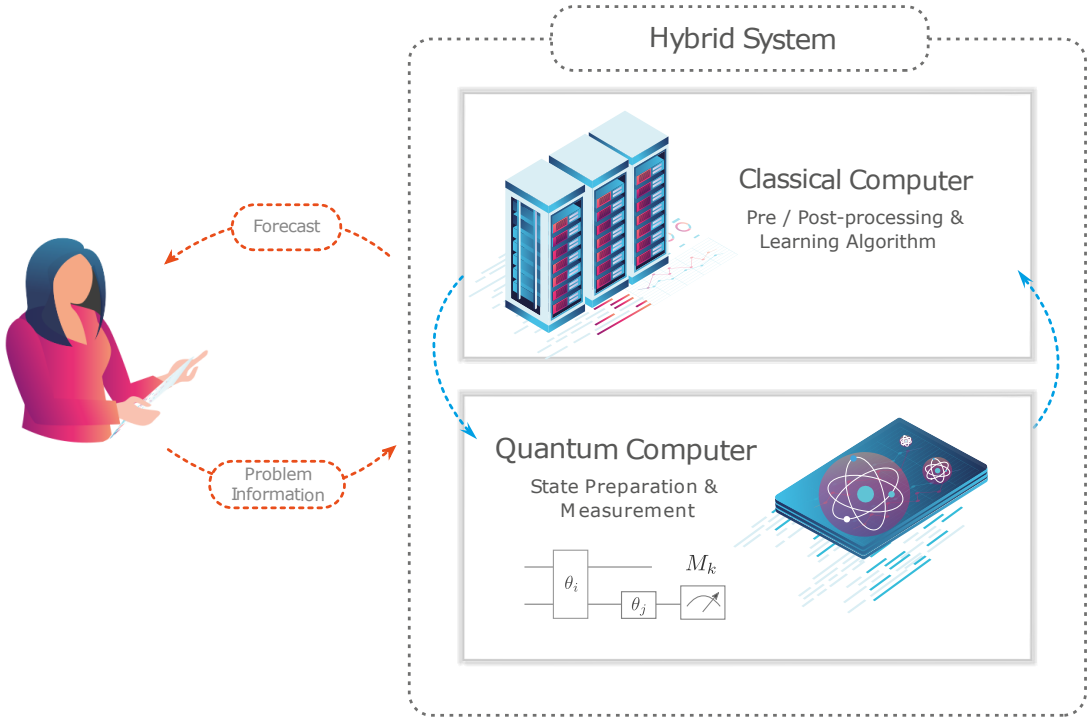

FIG. 1. High-level depiction of hybrid algorithms used for machine learning. The role of the human is to use the problem information to setup the model, assess the learning process, and use the forecasts. Within the hybrid system the quantum computer prepares quantum states according to a set of parameters. Using the measurement outcomes, the classical learning algorithm adjusts the set of parameters in order to minimize an objective function. The parameters, now defining a new quantum circuit, are fed back to the quantum hardware in a closed loop.

Reference	**Task**	**Model**	**Learning**	**Qubits**	**Computer**
Schuld *et al.* [16]	Classification	QKE	N/A	4	IBM Q5 Yorktown (S)
Grant *et al.* [17]	Classification	VQM	N/A	4	IBM Q5 Tenerife (S)
Havlíček *et al.* [18]	Classification	QKE, VQM	SPSA	2	IBM Q5 Yorktown (S)
Tacchino *et al.* [19]	Classification	Perceptron	GD	3	IBM Q5 Tenerife (S)
Benedetti *et al.* [20]	Generative modeling	QCBM	N/A	4	Custom (T)
Hamilton *et al.* [21]	Generative modeling	QCBM	GD w/Adam	4	IBM Q20 Tokyo (S)
Zhu *et al.* [22]	Generative modeling	QCBM	PSO, BO	4	Custom (T)
Leyton-Ortega *et al.* [23]	Generative modeling	QCBM	ZOOpt, SHC, GD w/Adam	4	Rigetti 16Q-Aspen (S)
Hu *et al.* [24]	Quantum state learning	QGAN	GD w/FD	1	Custom (S)
Rocchetto *et al.* [25]	Quantum state learning	PAC	N/A	6	Custom (P)
Otterbach *et al.* [10]	Clustering	QAOA	BO	19	Rigetti 19Q-Acorn (S)
Ding *et al.* [26]	Compression	QAE	GA	3	Rigetti 8Q-Agave (S)
Ristè *et al.* [27]	Learning parity with noise	Oracle	N/A	5	IBM Q5 Yorktown (S)

TABLE II. Overview of parameterized quantum circuit models that have been demonstrated experimentally on superconducting (S), trapped ion (T), and photonic (P) hardware. N/A labels the cases where a learning algorithm was either not required or not used, e.g., when learning is simulated classically and the model is deployed on quantum hardware.

Reference	Name	Developer	Features	Language	Backend
Aleksandrowicz *et al.* [28]	Qiskit Aqua	IBM Research	VQE, QAOA, VQM, QKE	Python	Superconducting, Simulator
Bergholm *et al.* [29]	Pennylane	Xanadu	VQE, VQM, QGAN	Python	Superconducting, Simulator
Luo *et al.* [30]	Yao	QuantumBFS	VQE, QAOA, QCBM	Julia	Simulator

TABLE III. Open-source software for developing machine learning models based on parameterized quantum circuits and, in some cases, for experimenting on existing quantum computers.

II. FRAMEWORK

We assume our computer to be a closed quantum system. With n qubits, its state can be described as a unit vector living in a complex inner product vector space $\mathbb{C}^{2^n}$. The computation always starts with an easy to prepare state of the computational basis, for example the product state $|0\rangle^{\otimes n}$. A unitary operator U is applied to the initial state producing a new state $U|0\rangle^{\otimes n}$. Here, the value of an observable quantity can be measured. Physical observables are associated with Hermitian operators. Let $M = \sum_i \lambda_i P_i$ be the Hermitian operator of interest, where λ_i is the i-th eigenvalue and P_i is the projector on the corresponding eigenspace. The *Born rule* states that the outcome of the measurement corresponds to one of the eigenvalues and follows probability distribution $p(\lambda_i) = \mathrm{tr}\big(P_i U|0\rangle\langle 0|U^\dagger\big)$. Plugging this in the definition of expectation values we obtain

$$\langle M\rangle = \sum_i \lambda_i p(\lambda_i) = \mathrm{tr}\big(MU|0\rangle\langle 0|U^\dagger\big). \qquad (1)$$

As we will see, one can exploit the probabilistic nature of quantum measurements to define a variety of machine learning models, and PQCs offer a concrete way to implement adjustable unitary operators U.

Figure 2 shows the components of a supervised learning model based on a PQC. First, a data vector is sampled from the training set and transformed by classical pre-processing, for example with de-correlation or standardization functions. Second, the transformed data point is mapped to the parameters of an *encoder circuit* $U_{\phi(\boldsymbol{x})}$. Third, a *variational circuit* $U_{\boldsymbol{\theta}}$, which possibly acts on an extended qubit register, implements the core operation of the model. This is followed by the estimation of a set of expectation values $\{\langle M_k\rangle_{\boldsymbol{x},\boldsymbol{\theta}}\}_{k=1}^K$ from measurements [1].

A post-processing function f is then applied to this set in order to provide a suitable output for the task. As an example, if we were to perform regression, f could be a linear combination of the kind $\sum_k w_k \langle M_k\rangle_{\boldsymbol{x},\boldsymbol{\theta}}$, with additional parameters w_k. Note that it is possible to parameterize and train all the components of the model, including pre- and post-processing functions.

Many of the proposals found in the literature fit within this framework with very small adaptation. For example, in generative modeling the encoder circuit may be used for a slightly different purpose, such as setting the initial state $|0\rangle$ to random computational basis states, in order to provide entropy to the model.

We now describe the encoder and variational circuits in detail and explain their links to other well-known machine learning models.

A. The encoder circuit $U_{\phi(\boldsymbol{x})}$

There are several ways to encode data into qubits and each one provides different expressive power. This choice of encoding is related to kernel methods, a well-established field whose goal is to embed data into a higher dimensional feature space where a specific problem may be easier to solve. For example, non-linear feature maps change the relative position between data points such that a dataset may become easier to classify in feature space. In a similar way, the process of encoding classical data into a quantum state in a high-dimensional Hilbert space can be interpreted as a feature map $\boldsymbol{x} \to U_{\phi(\boldsymbol{x})}|0\rangle$. Here, ϕ is a user-defined pre-processing function which transforms the data vector into circuit parameters.

The inner product of two data points in this Hilbert space defines a similarity function, or kernel, $k(\boldsymbol{x},\boldsymbol{x}') = \left|\langle 0|U^\dagger_{\phi(\boldsymbol{x}')}U_{\phi(\boldsymbol{x})}|0\rangle\right|^2$. Kernels can be estimated, for example, using the SWAP test shown in Fig. 3. This naturally leads to quantum versions of kernel based models such as the support vector machine (SVM), Gaussian processes, and principal component analysis.

Let us now discuss some examples. Stoudenmire and Schwab [31] encode data as products of local kernels, one for each component of the input vector, which results in a product quantum state (i.e., disentangled). This approach is often referred to as *qubit encoding*. The highly non-linear nature of the resulting kernel can be used in quantum-inspired classifiers. Mitarai *et al.* [32] use a similar approach, but encode each component of the data

[1] The number of repetitions required for the estimation of each term is determined by the desired precision as well as by the variance $\mathrm{Var}(M_k) = \langle M_k^2\rangle - \langle M_k\rangle^2$. In this Review we won't discuss estimation methods.

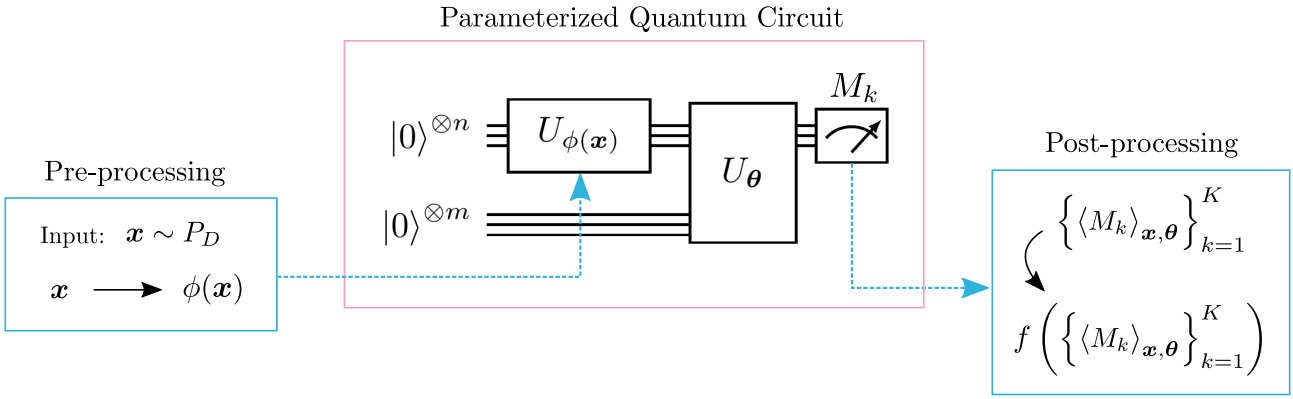

FIG. 2. A machine learning model comprised of classical pre/post-processing and parameterized quantum circuit. A data vector is sampled from the dataset distribution, $\boldsymbol{x} \sim P_D$. The pre-processing scheme maps it to the vector $\phi(\boldsymbol{x})$ that parameterizes the encoder circuit $U_{\phi(\boldsymbol{x})}$. A variational circuit $U_{\boldsymbol{\theta}}$, parameterized by a vector $\boldsymbol{\theta}$, acts on the state prepared by the encoder circuit and possibly on an additional register of ancilla qubits, producing the state $U_{\boldsymbol{\theta}}U_{\phi(\boldsymbol{x})}|0\rangle$. A set of observable quantities $\{\langle M_k\rangle_{\boldsymbol{x},\boldsymbol{\theta}}\}_{k=1}^{K}$ is estimated from the measurements. These estimates are then mapped to the output space through classical post-processing function f. For a supervised model, this output is the forecast associated to input $\boldsymbol{x}$. Generative models can be expressed in this framework with small adaptations.

vector into multiple qubits. This redundancy populates the wave function with higher-order terms that can be exploited to fit non-linear functions of the data. Vidal and Theis [33] investigate how this redundancy helps the task of data fitting. Using a linear-algebraic complexity measure, they found lower bounds of the required input redundancy to be logarithmic in the complexity of the function to be fit.

A different approach was taken by Wilson *et al.* [34]; the authors use a random linear map from data vectors to encoder circuit parameters to create a quantum version of the random kitchen sink [35]. They show that in the limit of many realizations of random linear maps, this approach implicitly implements a kernel. Interestingly, the form of the kernel depends on the circuit layout of the PQC, and not all layouts are capable of implementing useful kernels.

The examples discussed so far require low-depth encoder circuits and may therefore be robust to noise of some extent. A different approach is the so called *amplitude encoding*, a feature map encoding 2^n-dimensional data vectors into the wave function of merely n qubits. Assuming unit data vectors, this is a linear kernel $\boldsymbol{x} \to |\boldsymbol{x}\rangle$ with an exponential advantage in terms of memory. Preparing copies of this feature map one could even implement arbitrary polynomial kernels [36]. Unfortunately, the corresponding encoder circuit is expected to have exponential depth for generic inputs. Therefore, algorithms based on amplitude encoding could be impeded by our inability to coherently load data into quantum states. In fact, even if we were able to coherently prepare the required states, data loading and readout operations may dominate the computational complexity of the algorithm [37].

Havlíček *et al.* [18] argue that a feature map can be constructed so that the kernel is hard to estimate using classical resources, and that this is a form of quantum supremacy. They consider, for example, $U_{\phi(\boldsymbol{x})} = \exp(i\sum_{j,k}^{n}\phi_{j,k}(\boldsymbol{x})Z_jZ_k)H^{\otimes n}$ where Z_j is the Pauli operator for the j-th qubit, $\phi_{j,k}$ are real functions, and H is the Hadamard gate. They conjecture that two layers of such an encoder circuit make the estimation of the kernel $k(\boldsymbol{x},\boldsymbol{x'}) = |\langle 0|\, U^{\dagger}_{\phi(\boldsymbol{x'})}U^{\dagger}_{\phi(\boldsymbol{x'})}U_{\phi(\boldsymbol{x})}U_{\phi(\boldsymbol{x})}\,|0\rangle|^2$ classically intractable. This is due to its similarity to the circuits used in the hidden shift problem of Boolean bent functions, which are known to be classically hard to simulate [38].

Finally, it would be interesting to design feature maps that track closely other quantum supremacy proposals such as IQP circuits. Whether this leads to an advantage in practical ML tasks is an open question and should be tested empirically on existing computers. Ultimately, the form of the kernel and its parameters can be learned from data; this is a largely unexplored area in PQCs and has the potential to reduce the bias in kernel selection, and to automatically discover unknown feature maps that exhibit quantum supremacy.

B. The variational circuit $U_{\boldsymbol{\theta}}$

The variational circuit is the core of the model and consists of a sequence of operations applied to the encoded input state. Its purpose is to implement a function that is optimal for the task at hand. Similar to the universal approximation theorem in ANN [39], there always exists a circuit that can represent a target function exactly. The caveat is that such a circuit may be exponentially deep and therefore impractical. Lin *et al.* [40] argue that since real datasets arise from physical systems, they have characteristics such as symmetry and

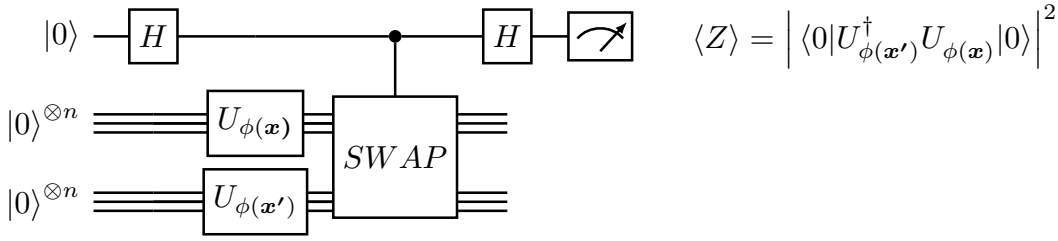

FIG. 3. The SWAP test can be used to estimate the implicit kernel implemented by an encoder circuit. Measurements of the Z Pauli observable on the ancilla qubit yield the absolute value squared of the inner product between $U_{\phi(\boldsymbol{x})}|0\rangle$ and $U_{\phi(\boldsymbol{x'})}|0\rangle$, respectively encoding data points $\boldsymbol{x}$ and $\boldsymbol{x'}$. The SWAP test finds several applications in machine learning and is a ubiquitous routine in quantum computing in general.

locality. These features makes it possible to use 'cheap' models, rather than exponentially costly ones, and still obtain remarkable accuracy in ML problems. Similarly, the variational circuit aims at extracting relevant quantum information from the input while remaining scalable in its number of parameters and depth. With a Hilbert space growing exponentially with the number of qubits, this is a daunting task and requires significant effort. Indeed, variational circuits are often assumed to have a fixed structure; while this limits their cost from the start, low-depth fixed circuits may suffer from a lack of expressivity unless sufficient non-linearity is introduced.

The first strategy to circuit design aims to comply with NISQ hardware which is usually equipped with few qubits and operates on a sparse qubit-to-qubit connectivity graph with rather simple gates. The so-called *hardware-efficient* circuits alternate layers of fixed entangling gates and parameterized single-qubit rotations [7]. Two examples of these layers are shown in Fig. 4, where (a) and (b) are designed around the connectivity and gate set of superconducting and trapped ion computers, respectively. The objective of these constructions is to maximize expressive power while keeping the number of parameters to a minimum – usually polynomial in the number of qubits. Heuristics can be used to strategically reduce the number of costly entangling gates. For example, Liu and Wang [41] use the Chow-Liu tree graph [42] to setup the entangling layers. First, the mutual information between all pairs of variables is estimated form the dataset. Then, entangling gates are placed between qubits so that most of the mutual information is represented.

Another principled approach to circuit design is inspired by quantum many-body physics. *Tensor networks* are methods to efficiently represent quantum states in terms of smaller interconnected tensors. This enables the numerical treatment of systems through layers of abstraction, reminiscent of deep ANNs. Importantly, tensor networks describe only the physically relevant portion of the Hilbert space containing states whose entanglement is constrained by local interactions. By looking only at this small corner of the Hilbert space, one lowers the computational cost to a polynomial dependence on system size. This language has been explored for generative and classification tasks using hierarchical representations such as matrix product states (MPS), tree tensor networks (TTN), and the multi-scale entanglement renormalization ansatz (MERA)[17, 43, 44].

These tensor network structures effectively implement a renormalization group flow, which is a technique to extract relevant degrees of freedom through coarse graining. At a theoretical level, the hierarchical representation used in deep learning for feature extraction can be mapped to the renormalization group [45], and entanglement has shown to be a relevant quantity to measure how deep networks express correlations in data [46]. This connection has allowed for the design of ANNs from a quantum entanglement perspective – and for ANNs to represent quantum states [47]. In light of these connections, and considering how classical models have already benefited from tensor networks [31], it is natural to propose variational circuits inspired by tensor networks.

Figure 5 (a) shows an example of a TTN for supervised learning. After the application of each unitary, half of the qubits are traced out, while the other half continues to the next layer. Huggins *et al.* [44] suggest a *qubit-efficient* version where the traced qubits are reinitialized and used as the inputs of another unitary, as shown in Fig. 5 (b). Qubit-efficient schemes could significantly reduce the required number of qubits, a favorable condition to some NISQ hardware.

ANN and deep learning have proven to be very successful and therefore provide a further source of inspiration for the design of PQCs. Both PQCs and ANNs can be thought of as layers of connected computational units controlled by adjustable parameters. This has led some authors to refer to variational circuits as 'quantum neural networks'. Here we shall briefly discuss the key differences that make this approach to circuit design rather difficult.

First, quantum circuit operations are unitary and therefore linear. This should be contrasted with the highly non-linear character of ANNs, which is a key component to their success and universality [49]. There are several ways to construct non-linear operations in quantum circuits, both coherently (i.e., exploiting entangle-

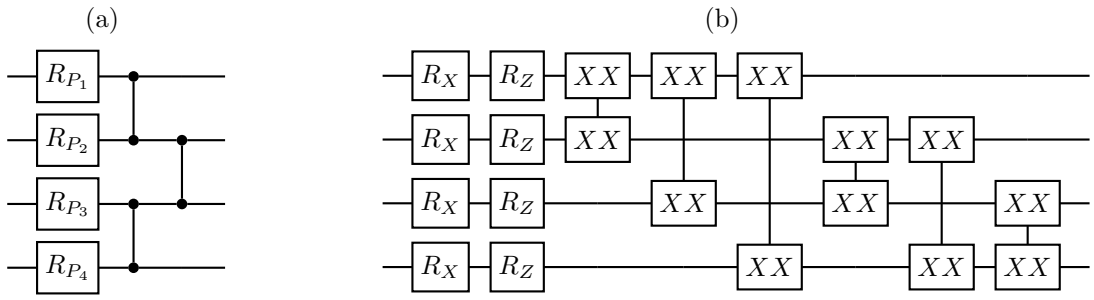

FIG. 4. Examples of hardware-efficient layers that can be used for encoder and variational circuits. Hardware-efficient constructions use entangling interactions that are naturally available on hardware and do not require compilation. Layers are repeated a number of times which is compatible with the hardware coherence time. (a) The construction in Ref. [48] uses single-qubit rotations $R_{P_j} = \exp\left(-\frac{i}{2}\theta_j P_j\right)$ about randomly sampled directions $P_j \in \{X, Y, Z\}$, and a ladder of control-Z entangling gates. Both the gate set and the connectivity are naturally implemented by many superconducting computers. (b) The construction in Ref. [20] uses single-qubit rotations about X and Y, and a fully-connected pattern of XX entangling gates. Both the gate set and the connectivity are naturally implemented by trapped ions computers.

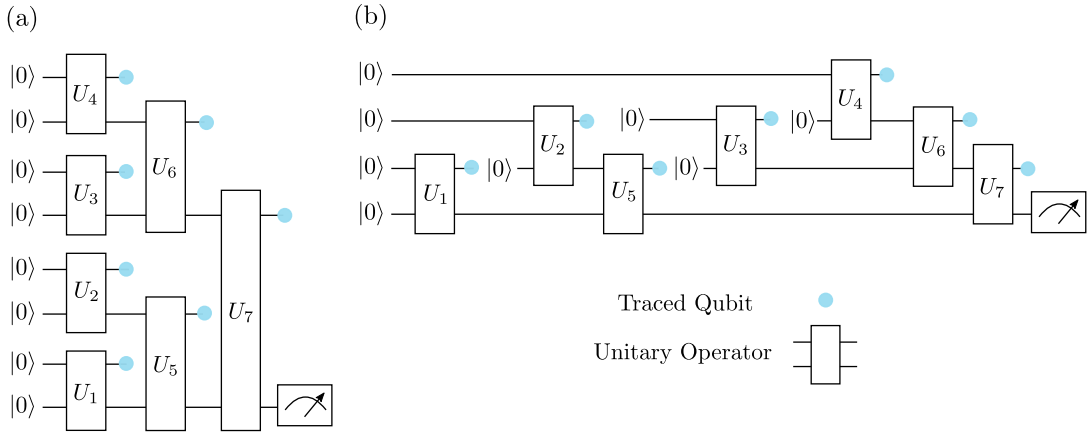

FIG. 5. Discriminative binary tree tensor network and its qubit-efficient version – adapted from Ref. [44]. (a) The binary TTN implements a coarse graining procedure by tracing over half of the qubits after the application of each unitary. (b) A qubit-efficient version re-initializes the discarded qubits to be used in parallel operations. This scheme implements the same operation in (a) but requires fewer qubits on the device. It may however result in a deeper circuit.

ment) or non-coherently (e.g., exploiting the natural coupling of the system to the environment). These can in turn be used to implement classical artificial neurons in quantum circuits [19, 50, 51].

The second key difference is that it is impossible to access the state of the PQC at intermediate points in the computation. Although measurement of ancillary quantum variables can be used to extract limited information, any attempt to observe the full state of the system would disrupt its quantum character. This implies that executing the PQC cannot be trivially seen as performing the forward pass of an ANN; it also makes it hard to design a PQC learning algorithm that truly resembles *backprop*, the gold standard algorithm for ANNs [52].

The questions of how to generalize a quantum artificial neuron and design a quantum backprop algorithm have been open for quite some time [53]. Some recent work goes towards this direction. Verdon *et al.* [54] quantize the parameters of the PQC model which are then prepared in superposition in a dedicated register. This enables a backprop-like algorithm which exploits quantum effects such as phase kickback and tunneling. Beer *et al.* [55] use separate qubit registers for PQC input and output, and define the quantum neuron as a completely positive map between the two. The resulting network is universal for quantum computation and can be trained by an efficient process resembling backprop.

C. Circuit learning

Just like classical models, PQCs are trained to perform data-driven tasks. Their learning algorithms can be categorized as either gradient-based or gradient-free. We discuss these two types of algorithms and how they can be applied to optimize the parameters of a variational

circuit $U_{\boldsymbol{\theta}}$.

The task of learning an arbitrary function from data is mathematically expressed as the minimization of a loss function $L(\boldsymbol{\theta})$, also known as the objective function, with respect to the parameter vector $\boldsymbol{\theta}$. One way to achieve this is by performing an iterative method called gradient descent (GD). In GD, parameters are updated towards the direction of steepest descent of the loss function

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \eta \nabla_{\boldsymbol{\theta}} L, \tag{2}$$

where $\nabla_{\boldsymbol{\theta}} L$ is the gradient vector and η is the learning rate – a hyperparameter controlling the size of the update. This procedure is iterated and, assuming suitable conditions, converges to a local minimum of the loss function.

The required partial derivatives can be calculated numerically using a finite difference scheme

$$\frac{\partial L}{\partial \theta_j} \approx \frac{L(\boldsymbol{\theta} + \Delta \boldsymbol{e}_j) - L(\boldsymbol{\theta} - \Delta \boldsymbol{e}_j)}{2\Delta}, \tag{3}$$

where Δ is a (small) hyperparameter and $\boldsymbol{e}_j$ is the Cartesian unit vector in the j direction. Note that in order to estimate the gradient vector $\nabla_{\boldsymbol{\theta}} L$, this approach evaluates the loss function twice for each parameter.

Alternatively, Spall's simultaneous perturbation stochastic approximation (SPSA) [56, 57] computes an approximate gradient vector with just two evaluations of the loss function as

$$\frac{\partial L}{\partial \theta_j} \approx \frac{L(\boldsymbol{\theta} + c\,\boldsymbol{\Delta}) - L(\boldsymbol{\theta} - c\,\boldsymbol{\Delta})}{2\,c\,\Delta_j}, \tag{4}$$

where $\boldsymbol{\Delta}$ is a random perturbation vector and c is a (small) hyperparameter.

There are cases when finite difference methods are ill-conditioned and unstable due to truncation and round-off errors. This is one of the reasons why ML relies on the analytical gradient when possible, and it is often calculated with automatic differentiation schemes [58]. The analytical gradient can also be estimated for PQC models, although the equations depend on the choice of parameterization for the gates. For our discussion, we consider circuits $U_{J:1} = U_J \cdots U_1$, where trainable gates are of the from $U_j = \exp\left(-\frac{i}{2}\theta_j P_j\right)$, and where $P_j \in \{I, Z, X, Y\}^{\otimes n}$ is a tensor product of n Pauli matrices. Arguably, this is the most common parameterization found in the literature.

Using this, Li *et al.* [59] propose a way to efficiently compute analytical gradients in the context of quantum optimal control. Mitarai *et al.* [32] bring this method to the context of supervised learning. Recall that the model's output is a function of expectation values $\langle M_k \rangle_{\boldsymbol{\theta}}$. Using the chain rule we can write the derivative $\frac{\partial L}{\partial \theta_j}$ as a function of the derivatives of the expectation values $\frac{\partial \langle M_k \rangle_{\boldsymbol{\theta}}}{\partial \theta_j}$. Each of these quantities can be estimated on quantum hardware using the so called 'parameter shift rule'

$$\frac{\partial \langle M_k \rangle_{\boldsymbol{\theta}}}{\partial \theta_j} = \frac{\langle M_k \rangle_{\boldsymbol{\theta} + \frac{\pi}{2}\boldsymbol{e}_j} - \langle M_k \rangle_{\boldsymbol{\theta} - \frac{\pi}{2}\boldsymbol{e}_j}}{2}, \tag{5}$$

where subscripts $\boldsymbol{\theta} \pm \frac{\pi}{2}\boldsymbol{e}_j$ indicate the shifted parameter vector to use for the evaluation (see Schuld *et al.* [60] for a detailed derivation). Note that this estimation can be performed by executing two circuits.

An alternative method can estimate the partial derivative with a single circuit, but at the cost of adding an ancilla qubit. A simple derivation using the gate parameterization introduced above (e.g., see Farhi and Neven [61]) shows that the partial derivative can be written as

$$\frac{\partial \langle M_k \rangle_{\boldsymbol{\theta}}}{\partial \theta_j} = \mathrm{Im}\left(\mathrm{tr}\Big(M_k U_{J:j+1} P_j U_{j:1} \left|0\right\rangle\!\left\langle 0\right| U_{J:1}^{\dagger}\Big)\right). \tag{6}$$

This can be thought of as an *indirect measurement* and can be evaluated using the Hadamard test shown in Fig. 6. This method can be generalized to compute higher order derivatives, as presented for example by Dallaire-Demers and Killoran [62], and with alternative gate parameterizations, as done for example by Schuld *et al.* [63].

We shall note that despite the apparent simplicity of the circuit in Fig. 6, the actual implementation of Hadamard tests may be challenging due to non-trivial controlled gates. Coherence must be guaranteed in order for quantum interference to produce the desired result. Mitarai and Fujii [64] propose a method for replacing a class of indirect measurements with direct ones. Instead of an interference circuit one can execute, in some cases, multiple simpler circuits that are suitable for implementations on NISQ computers. The 'parameters shift rule' in Eq. (5) is nothing but the direct version of the measurement in Eq. (6).

Compared to finite difference and SPSA, the analytical gradient has the advantage of providing an unbiased estimator. Additionally, Harrow and Napp [65] find evidence that training PQCs using the analytical gradient outperforms any finite difference method. This is done by showing that for n qubits and precision ϵ, the query cost of an oracle for convex optimization in the vicinity of the optimum scales as $\mathcal{O}(\frac{n^2}{\epsilon})$ for the analytical gradient, whereas finite difference needs at least $\Omega(\frac{n^3}{\epsilon^2})$ calls to the oracle. In practice though, it is found that SPSA performs well in small-scale noisy experimental settings (e.g. see Kandala *et al.* [7] and Havlíček *et al.* [18]).

Particular attention should be given to the problems of exploding and vanishing gradients which are well-known to the machine learning community. Classical models, in particular recurrent ANN, are often constrained to perform unitary operations so that their gradients cannot explode (see Wisdom *et al.* [66] for an example). PQCs naturally implement unitary operations and therefore avoid the exploding gradient problem altogether. On the other hand, McClean *et al.* [48] show that random circuits of

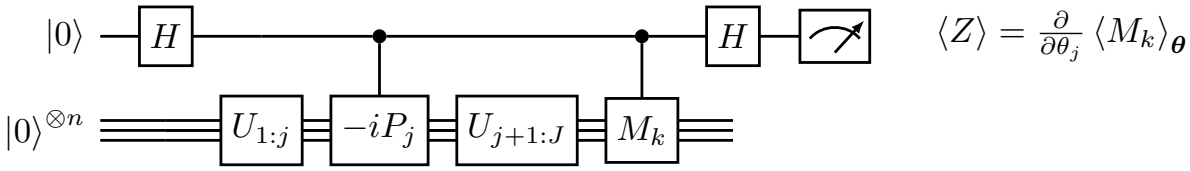

FIG. 6. The Hadamard test can be used to estimate the partial derivative of an expectation $\langle M_k\rangle_{\boldsymbol{\theta}}$ with respect to the parameter θ_j. Here we show a simple case where gates are of the form $U_j = \exp\left(-\frac{i}{2}\theta_j P_j\right)$ and where both P_j and M_k are tensor products of Pauli matrices. It can be shown that measurements of the Z Pauli observable on the ancilla qubit yield Eq. (6), the desired partial derivative. Hadamard tests can be designed to estimate higher order derivatives and to work with different measurements and gate parameterizations.

reasonable depth lead to an optimization landscape with exponentially large plateaus of vanishing gradients with an exponentially decaying variance. This can be understood as a consequence of Levy's lemma [67] which states that a random variable that depends on many independent variables is essentially constant. The learning algorithm is thus unable to estimate the gradient and may perform a random walk in parameter space. While this limits the effectiveness of PQCs initialized at random, the use of highly structured circuits could alleviate the problem (e.g., see Grant *et al.* [68] for a structured initialization strategy).

We shall stress here that in hybrid systems parameter updates are performed classically. This implies that some of the most successful deep learning algorithms can be readily used for training PQC models. For the case of gradient-based optimization, heuristics such as stochastic gradient descent [69], resilient backpropagation [70], and adaptive momentum estimation (Adam) [71], have already been applied with success. These were designed to deal with issues of practical importance such as large datasets, large noise in gradient estimates, and the need to find adaptive learning rates in Eq. (2). In practice, these choices can reduce the time for successful training from days to hours.

There are cases where gradient-based optimization may be challenging. For example, in a noisy experimental setting the loss function may be highly non-smooth and not suitable for GD. As another example, the objective function may be itself unknown and therefore should be treated as a black-box. In these cases, circuit learning can be carried out by gradient-free methods. A well-known method of this type is particle swarm optimization (PSO) [72]. Here the system is initialized with a number of random solutions called particles, each one moving through solution space with a certain velocity. The trajectory of each particle is adjusted according to its own experience and that of other particles so that they converge to a local minima. Another popular method is Bayesian optimization (BO) [73]. BO uses evaluations of the objective function to construct a model of the function itself. Subsequent evaluations can be chosen either to improve the model or to find a minima.

Zhu *et al.* [22] compare BO and PSO for training a generative model on a trapped ion quantum computer. While BO outperforms PSO in their setting, they found that the large number of parameters challenges both optimizers. They show that an ideal simulated system is not significantly faster than the experimental system, indicating that the actual bottleneck is the classical optimizer. Leyton-Ortega *et al.* [23] train a generative model on a superconducting quantum computer and compare the gradient-free methods of zeroth-order optimization package (ZOOpt) [74] and stochastic hill-climbing (SHC), with GD using Adam. They find that on average ZOOpt achieves the lowest loss on their hardware. They argue that the main optimization challenge is to overcome the variance of the loss function which is due to random parameter initialization, hardware noise, and finite number of measurements.

Genetic algorithms (GAs) [75] are another large class of gradient-free optimization algorithms. At each step, candidate solutions are evolved using biology-inspired operations such as recombination, mutation, and natural selection. When used to train PQCs, GAs define a set of allowed gates and the maximum number to be employed. Lamata *et al.* [76] suggest the use of GAs to train a PQC model for compression using a universal set of single- and two-qubit gates. Ding *et al.* [26] validate the idea experimentally by deploying a pre-trained PQC model on a superconducting computer and find that using a subsequent GA improves its fidelity.

To conclude, we note that optimization algorithms should be tailored for PQCs if we want to achieve better scalability. Very recent work has been approaching circuit learning from this perspective (e.g., see Ostaszewski *et al.* [77] and Nakanishi *et al.* [78]).

III. APPLICATIONS

In this Section we look at ML applications using PQC models where the goal is to obtain an advantage over classical models. For supervised learning with classical data we give a general overview of how PQCs can be applied to classification and regression. For unsupervised learning with classical data we focus on generative modeling since this comprises most of the literature.

PQCs can also handle inputs and outputs that are inherently quantum mechanical, i.e., already in superposition. These are often referred to as *quantum data* [79]. Quantum input data could originate remotely, for example, from other quantum computers transmitting over a quantum internet. Otherwise, if a preparation recipe is available, one could prepare the input data locally using a suitable encoder circuit. Assuming this data preparation is efficient, one can extend supervised and unsupervised learning to quantum states and perform other interesting processing of quantum information.

Figure 7 shows examples for all these cases. Intuitively each application is a specification of the components outlined in Fig. 2, which the Reader is encouraged to refer to throughout the Section for clarity.

In many practical decision-making scenarios there is no available data. In this case, the model needs to interact with its environment to obtain information and learn how to perform a task from its own experience. This is known as reinforcement learning. An example would be a video game character that learns a successful strategy by repeatedly playing the game, analyzing results, and improving. Although quantum generalizations and algorithms for reinforcement learning have been proposed, to the best of our knowledge, none of them are based on hybrid systems and PQCs.

A. Supervised learning

Let us first consider supervised learning tasks, e.g., classification and regression, on classical data. Given a dataset $\mathcal{D} = \{(\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)})\}_{i=1}^{N}$, the goal is to learn a model function $f : \mathcal{X} \to \mathcal{Y}$ that maps each $\boldsymbol{x} \in \mathcal{X}$ to its corresponding target $\boldsymbol{y} \in \mathcal{Y}$. A standard approach is to minimize a suitable regularized loss function, that is,

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} L\left(f(\boldsymbol{x}^{(i)}, \boldsymbol{\theta}), \boldsymbol{y}^{(i)}\right) + R(\boldsymbol{\theta}), \quad (7)$$

where $\boldsymbol{\theta}$ is the set of parameters defining the model function, L quantifies the error of a forecast, and R is a regularization function penalizing undesired values for the parameters. The latter is used to prevent overfitting; indeed, if the training set is not sufficiently large, the model could simply memorize the dataset and not generalize to unseen data.

In the PQC framework, we first setup an encoder circuit $U_{\phi(\boldsymbol{x})}$ to implement a feature map on the input data. From there we have two options: the quantum kernel estimator (QKE), and the variational quantum model (VQM). We now briefly discuss both, and refer the Reader to Schuld and Killoran [80] for a more in-depth theoretical exposition.

The QKE does not use a variational circuit $U_{\boldsymbol{\theta}}$ to process the data; instead, it uses the SWAP test (e.g., see Fig. 3) to estimate the possibly intractable kernel $k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})$. Then, resorting to the representer theorem [81], the model function is expressed as an expansion over kernel functions $f(\boldsymbol{x}, \boldsymbol{w}) = \sum_{d=1}^{D} w_d k(\boldsymbol{x}, \boldsymbol{x}^{(d)})$. The learning task is to find parameters $\boldsymbol{w}$ so that the model outputs correct forecasts. Note that these parameters define the classical post-processing function, as opposed to an operation of the PQC. In other words, this approach completely off-loads the modeling part to the classical computer. A potential caveat is that QKE relies on a coherent SWAP test which may be non-trivial to implement on NISQ computers.

The VQM uses a variational circuit $U_{\boldsymbol{\theta}}$ to process data directly in the Hilbert space. A set of expectation values $\{\langle M_k \rangle_{\boldsymbol{x},\boldsymbol{\theta}}\}_{k=1}^{K}$ is estimated and post-processed by the function f in order to construct a model output (see Fig. 2). For example, f could add a scaling factor for regression, or apply a threshold for classification. Note that in contrast to QKE, VQM parameters define the operations carried out by the quantum computer. This model requires a training algorithm of the kind discussed in Section II C.

Havlíček *et al.* [18] experimentally demonstrate QKE and VQM classifiers on two superconducting qubits of the IBM Q5 Yorktown. Their QKE estimates a classically intractable feature map (see Section II A for details) which is then fed into a SVM to find the separating hyperplane. Their VQM classifier uses a hardware-efficient circuit made of alternating layers of single qubit rotations and entangling gates. By employing a suitable error mitigation protocol, they find an increase in classification success with increasing circuit depth.

Other proposals with no variational circuit rely on random encoder circuits followed by classical models such as linear regression or ANNs (e.g., the quantum kitchen sink by Wilson *et al.* [34] and the quanvolutional network by Henderson *et al.* [82]).

We now focus our discussion on VQM proposals. Farhi and Neven [61] propose a VQM binary classifier for bitstrings. The encoder circuit simply maps bitstrings to computational basis states by applying identity and NOT gates at almost no cost. The variational circuit acts on the input register and one ancilla qubit which is measured to yield a class forecast. With n-bit data strings as the input, there are 2^{2^n} possible binary functions that could generate the class labels. The authors show that for any of the possible label functions there exists a variational circuit that achieves zero classification error. For some of these functions, the circuit is exponentially deep and therefore impractical. This result parallels the well known universal approximation theorem [39] which states that ANNs with an exponentially large hidden layer of non-linear neurons are able to represent any Boolean function.

Mitarai *et al.* [32] propose VQMs for classification and regression of real-valued data using a highly non-linear qubit-encoding. The variational circuit must then highly entangle the qubits such that a local observable can extract the relevant non-linear features. As discussed in Section II B one possible way to strategically construct

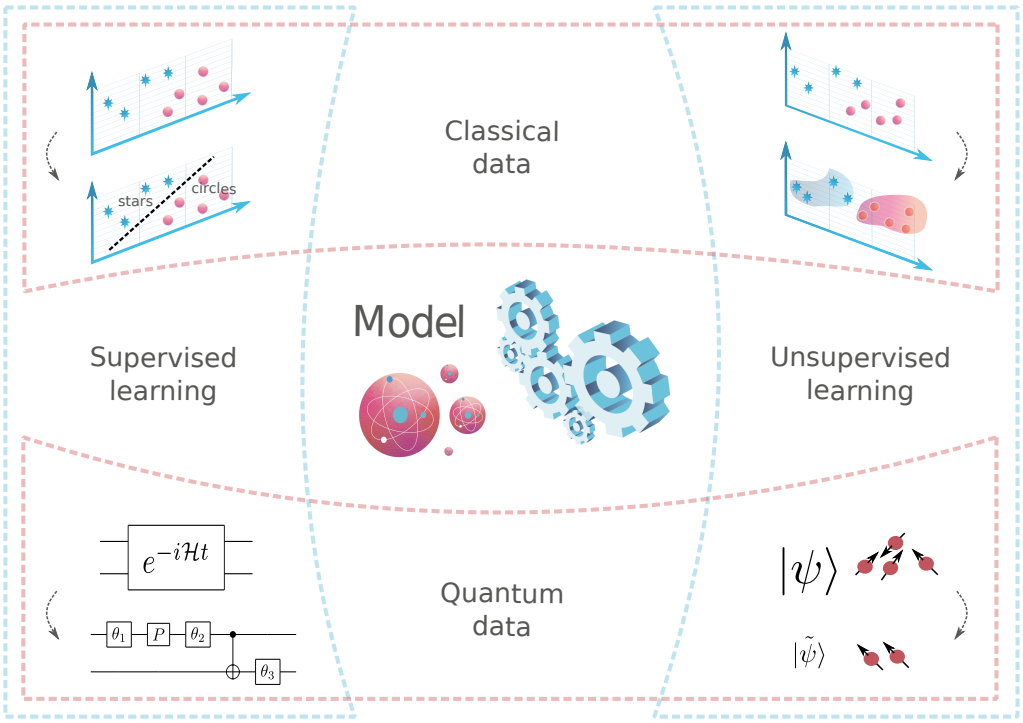

FIG. 7. Parameterized quantum circuit models can be trained for a variety of machine learning tasks, such as supervised and unsupervised learning, on both classical and quantum data. This figure shows examples from each category. In the top-left panel, the model learns to recognize patterns to classify the classical data. In the top-right panel, the model learns the probability distribution of the training data and can generate new synthetic data accordingly. For supervised learning of quantum data, bottom-left panel, the model assists the compilation of a high-level algorithm to low-level gates. Finally, for unsupervised learning of quantum data, bottom-right panel, the model performs lossy compression of a quantum state.

highly entangling variational circuits is inspired by tensor networks. Grant *et al.* [17] use TTN and MERA variational circuits to perform binary classification on qubit-encoded Iris and MNIST datasets. In their simulations, MERA always outperforms TTN. One of their simplest models is efficiently trained classically and then deployed on the IBM Q5 Tenerife quantum computer with significant resilience to noise.

Stoudenmire *et al.* [43] train a 2D TTN to perform pairwise classification of the MNIST image data. Although a fully classical experiment, they use quantum fidelity to measure the inherent difficulty to distinguish two classes, and entanglement entropy as quantifying the amount of information about one part of an image that can be gained by knowing the rest. Defining and testing these measures is an important step towards the use of quantum properties to model classical datasets.

Schuld *et al.* [63] propose a VQM classifier for real-valued data assuming they have amplitude encoding. Since this encoder circuit may be very expensive, the authors aim to keep the variational circuit low-depth and highly expressive at the same time. This is achieved through a systematic use of entangling gates, and by keeping the number of parameters polynomial in the number of qubits. Their simulations on several real-world datasets show performance comparable to that of off-the-shelves models while using significantly fewer parameters.

In summary, the quantum advantage in supervised learning of classical data may stem from classically intractable PQC models operating in high-dimensional feature spaces. To date all demonstrations have been on scaled-down, often trivial, problems due to the limitation of available quantum hardware. Therefore, a convincing demonstration of reasonable scale is urgently needed. Another largely undeveloped area is that of regularization techniques specifically designed for PQC models.

B. Generative modeling

We now discuss generative modeling, an unsupervised learning task where the goal is to model an unknown probability distribution and generate synthetic data accordingly. Generative models have been successfully applied in computer vision, speech synthesis, inference of missing text, de-noising of images, chemical design, and many other automated tasks. It is believed that they will play a key role in the development of general artificial intelligence; a model that can generate realistic synthetic samples is likely to 'understand' its environment.

Concretely, the task is to learn a model distribution $q_{\boldsymbol{\theta}}$ that is close to a target distribution p. The closeness is defined in terms of a divergence D on the statistical manifold, and learning consists of minimizing this divergence; that is,

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} D(p, q_{\boldsymbol{\theta}}). \tag{8}$$

Since the target probability distribution is unknown, it is approximated using a dataset $\mathcal{D} = \{\boldsymbol{v}^{(i)}\}_{i=1}^N$ which we have access to and which is distributed according to the target distribution. As an example, $\boldsymbol{v}^{(i)}$ could be natural images extracted from the Internet.

The probabilistic nature of quantum mechanics suggests that a model distribution can be encoded in the wave function of a quantum system [83, 84]. Let us see how a simple adaptation of the model shown in Fig. 2 gives a generative model for n-dimensional binary data $\boldsymbol{v}^{(i)} \in \{0,1\}^n$. First, we set the encoder circuit to the identity $U_{\phi(\boldsymbol{x})} = I$ since in this problem there is no input data. Second, we apply a variational circuit $U_{\boldsymbol{\theta}}$ to the initial state $|0\rangle^{\otimes n}$. Finally, we perform a measurement in the computational basis, i.e., we measure the set of operators $\{\langle M_{\boldsymbol{v}}\rangle_{\boldsymbol{\theta}}\}_{\boldsymbol{v}}$ where $M_{\boldsymbol{v}} = |\boldsymbol{v}\rangle\langle\boldsymbol{v}|$ are projectors for the bitstrings. The resulting generative model, known as the quantum circuit Born machine [20, 41], implements the probability distribution

$$q_{\boldsymbol{\theta}}(\boldsymbol{v}) = \mathrm{tr}\Big(M_{\boldsymbol{v}} U_{\boldsymbol{\theta}} \,|0\rangle\langle 0|\, U_{\boldsymbol{\theta}}^\dagger\Big). \tag{9}$$

Since the target data is binary, no post-processing is needed and each measurement outcome $\boldsymbol{v} \sim q_{\boldsymbol{\theta}}$ is an operational output. If the target data were instead real-valued, we could interpret bitstrings as discretized outputs and use a post-processing function to recover real-values.

As one does not have access to the wave function, characterizing the distribution $q_{\boldsymbol{\theta}}$ may be intractable for all but the smallest circuits. For this reason, QCBMs belong to the class of *implicit models*, models where it is easy to obtain a sample $\boldsymbol{v} \sim q_{\boldsymbol{\theta}}$, but may be hard to estimate the likelihood $q_{\boldsymbol{\theta}}(\boldsymbol{v})$. The ML community has become increasingly interested in implicit models because of their generality, expressive power, and success in practice [85]. Interestingly, Du *et al.* [86] show that QCBMs have strictly more expressive power than classical models such as deep Boltzmann machines, when only a polynomial number of parameters are allowed. Coyle *et al.* [87] show that some QCBMs cannot be efficiently simulated by classical means in the worst case, and that this holds for all the circuit families encountered during training.

Benedetti *et al.* [20] build low-depth QCBMs using variational circuits suitable for trapped ion computers (see Fig. 4 (b) for an example). They use PSO to minimize an approximation to the Kullback-Leibler divergence [90] $D(p, q_{\boldsymbol{\theta}}) = \sum_{\boldsymbol{v}} p(\boldsymbol{v}) \ln \frac{p(\boldsymbol{v})}{q_{\boldsymbol{\theta}}(\boldsymbol{v})}$. In their simulations they successfully train models for the canonical Bars-and-Stripes dataset and for Boltzmann distributions, and use them to design a performance indicator for hybrid quantum-classical systems. Zhu *et al.* [22] implement this schema on four qubits of an actual trapped ion computer and experimentally demonstrate convergence of the model to the target distribution.

Liu and Wang [41] propose the use of GD to minimize the maximum mean discrepancy [91] $D(p, q_{\boldsymbol{\theta}}) = \|\sum_{\boldsymbol{v}} p(\boldsymbol{v})\phi(\boldsymbol{v}) - \sum_{\boldsymbol{v}} q_{\boldsymbol{\theta}}(\boldsymbol{v})\phi(\boldsymbol{v})\|^2$, where ϕ is a classical feature map, and the expectations are estimated from samples. Their approach allows for gradient estimates with discrete target data, which is often not possible in classical implicit models. In their simulations they successfully train QCBMs for the Bars-and-Stripes dataset and for discretized Gaussian distributions. Hamilton *et al.* [21] implement this schema on the IBM Q20 Tokyo computer, and examine how statistical and hardware noise affect convergence. They find that the generative performance of state-of-the-art hardware is usually significantly worse than that of the numerical simulations. Leyton-Ortega *et al.* [23] perform a complementary experimental study on the Rigetti 16Q-Aspen computer. They argue that due to the many components involved in hybrid quantum-classical systems (e.g., choice for the entangling layers, optimizers, post-processing strategy, etc.), the performance ultimately depends on the ability to correctly set hyperparameters. Research on automated hyperparameter setting will be key to the success of QCBMs.

In summary, studies have validated the QCBM under different settings for both quantum and classical parts of the hybrid system. Arguably, the main challenge has been in designing suitable loss functions. Non-differentiable loss functions are often hard to optimize; one may resort to gradient-free methods, but these are likely to struggle as the number of parameters becomes large. Differentiable loss functions are often hard to design; since QCBM are implicit models, one does not even have access to the likelihood $q_{\boldsymbol{\theta}}(\boldsymbol{v})$. Borrowing ideas from deep learning, such as *adversarial* methods, we can potentially overcome these limitations. Figure 8 (a) shows the intuition; the adversarial method introduces a discriminative model whose task is to distinguish between true data coming from the dataset and synthetic data coming from the generative model. This creates a 'game' where the two players, i.e., the models, compete. The advantage is that both models are trained at the same time, with the discriminator providing a differentiable loss function for the generator.

Lloyd and Weedbrook [92] put forward the quantum generative adversarial network (QGAN) and theoretically examine variants where target data, generator and discriminator are either classical or quantum. We discuss the case of quantum data in the next Section while here we focus on classical data. Both Situ *et al.* [93] and Zeng *et al.* [94] couple a PQC generator to an ANN discriminator and successfully reproduce the statistics of discrete target distributions. Romero and Aspuru-Guzik [95] ex-

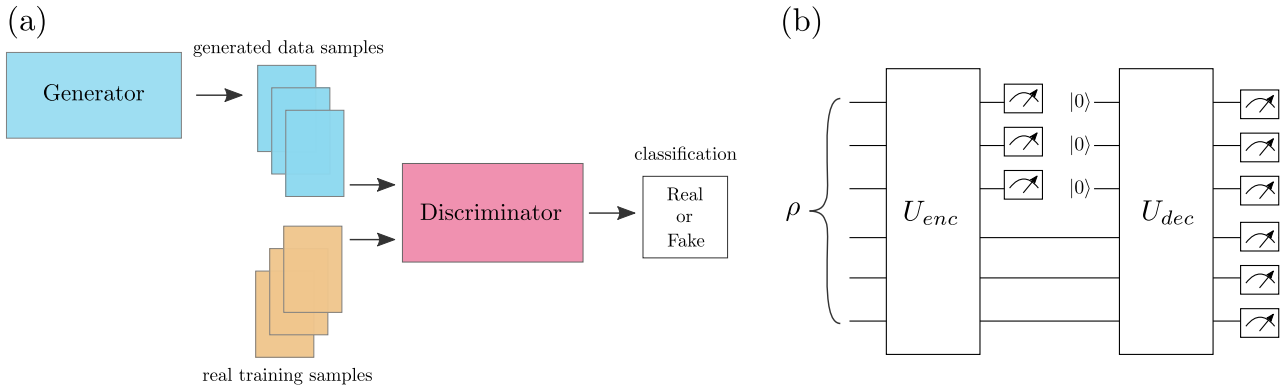

FIG. 8. Illustration of quantum generative models. (a) In the quantum generative adversarial network the generator creates synthetic samples and the discriminator tries to distinguish between the generated and the real samples. The network is trained until the generated samples are indistinguishable from the training samples. In this method the target data, the generator, and the discriminator can all be made quantum or classical. (b) The quantum autoencoder reduces the dimensionality of quantum data by applying an encoder circuit U_{enc}, tracing over a number of qubits and finally reconstructing the state with a decoder circuit U_{dec}. Panels (a) and (b) are adapted from Refs. [88] and [89], respectively.

tend this to continuous target distributions using a suitable post-processing function. Zoufal *et al.* [88] propose a QGAN to approximately perform amplitude encoding. While the best known generic method has exponential complexity, their circuit uses a polynomial number of gates. If both the cost of training and the required precision are kept low, this method has the potential to facilitate algorithms that require amplitude encoding.

One key aspect of generative models is their ability to perform *inference*. That is, when some of the observable variables are 'clamped' to known values, one can infer the expectation value of all other variables by sampling from the conditional probability. For example, inpainting, the process of reconstructing lost portions of images and videos, can be done by inferring missing values from a suitable generative model. Low *et al.* [96] use Grover's algorithm to perform inference on quantum circuits and obtain a quadratic speedup over naïve methods, although the overall complexity remains exponential. Zeng *et al.* [94] propose to equip QCBMs with this method, although this requires amplitude amplification and estimation methods that may be beyond NISQ hardware capabilities. It is an open question how to perform inference on QCBMs in the near term.

C. Quantum learning tasks

We finally consider learning tasks that are inherently quantum mechanical. As discussed in the Introduction, early hybrid approaches [13, 14] were proposed to assist the implementation of quantum algorithms (e.g. Deutsch's, Grover's, and Shor's) from datasets of input-output pairs. *Quantum algorithm learning* has been recently rediscovered by the community.

Morales *et al.* [97] train PQCs corresponding to diffusion and oracle operators for Grover's algorithm. When specifically using three and four qubits their approach finds new operators that outperform standard Grover search in terms of success probability. Wan *et al.* [98] train a PQC to solve the hidden subgroup problem studied by Simon [99]. In their simulations, they recover the original Simon's algorithm with equal performance. Anschuetz *et al.* [100] use known techniques to map integer factoring to an Ising model, then train a PQC to find the ground state of the corresponding Hamiltonian. Cincio *et al.* [101] train circuits to implement the SWAP test (see Fig. 3) and find solutions with a smaller number of gates than the known circuits.

These methods promise to assist the implementation of algorithms on near-term computers. Experimental studies will be needed to assess their scaling under realistic NISQ constraints and noise. Theoretical studies will be needed to understand their *sample complexity*, that is, the number of training samples required in order to successfully learn the target algorithm. Even in small-scale computers, we shall avoid exponential sampling complexity if we want these methods to be practical.

In the context of *quantum state classification*, Grant *et al.* [17] simulate the training of a TTN variational circuit for the classification of pure states that have different levels of entanglement. They found that, if the unitary operations in the TTN are too simple, classification accuracy on their synthetic dataset is no better than random class assignments. When using more complex operations involving ancilla qubits the TTN is able to classify quantum states with some accuracy. Chen *et al.* [102] simulate the training of variational circuits to classify quantum states as pure or mixed. Their construction relies on layers of gates that are conditioned on measurement outcomes, with the purpose of introducing non-linear behaviour similar to that of ANNs. Their binary classification model includes a third possible output associated with an inconclusive result. The authors

show that this feature can affect generalization because it introduces a non-trivial trade-off between classification accuracy and inconclusiveness.

State tomography is another ubiquitous task aiming at predicting the outcome probabilities of any measurement performed on an unknown state. To completely model the unknown state, one would require a number of measurements growing exponentially with the number of qubits. However, this can be formulated as a *quantum state learning* problem with the hope of minimizing the number of required measurements. Aaronson [103] studies the sampling complexity of this problem under Valiant's probably approximately correct (PAC) learning model [104]. They find that for practical purposes one needs a number of measurements scaling linearly with the number of qubits. Rocchetto *et al.* [25] experimentally verify the linear scaling on a custom photonic computer and extrapolate the value of the scaling constant. In terms of methodology, Lee *et al.* [105] propose to train a variational circuit $U_{\boldsymbol{\theta}}$ that transforms the unknown state $|\psi\rangle$ to a known fiducial state $|f\rangle$. The unknown state can be reproduced by evaluating the adjoint circuit on the fiducial state, that is, $|\psi\rangle \approx U_{\boldsymbol{\theta}}^{\dagger}|f\rangle$. A related learning tasks is that of *quantum state diagonalization*. LaRose *et al.* [106] propose to train a variational circuit $U_{\boldsymbol{\theta}}$ such that the density matrix $\tilde{\rho} = U_{\boldsymbol{\theta}}\rho U_{\boldsymbol{\theta}}^{\dagger}$ is diagonalized, hence representing a classical probability distribution.

In the previous Section and in Fig. 8 (a) we introduced QGANs for classical data. We now discuss the case where all components are quantum mechanical, hence enabling the generative modeling of quantum data. The discriminator, now taking an input target and fake quantum states, aims at modeling the measurement for optimal distinguishability, also known as the Helstrom measurement [107]. In turn, the generator tries to make the task of distinguishing more difficult by minimizing its distance from the target state [92, 108]. In practice, this game can be implemented by coupling two PQCs and optimizing them in tandem. For example, Dallaire-Demers and Killoran [62] propose a QGAN that generates states conditioned on labels. This may find application in chemistry where the label is clamped to a desired physical property and the model generates new molecular states accordingly. Benedetti *et al.* [108] propose a QGAN that generates approximations of pure states. They numerically show how the depths of generator and discriminator impact the quality of approximation. They also design a heuristic for stopping training, which is a non-trivial problem even in classical adversarial methods. Hu *et al.* [24] experimentally demonstrate adversarial learning on a custom superconducting qubit.

Finally, PQCs can be used to attack well-known problems in quantum information (e.g. compression, error correction and compilation) from a novel machine learning perspective. Romero *et al.* [89] propose a quantum autoencoder (QAE) to reduce the amount of resources needed to store quantum data. As shown in Fig. 8 (b) an encoder circuit U_{enc} is applied to the quantum data stored on n qubits. After tracing out $n-k$ qubits, a decoder circuit U_{dec} is used to reconstruct the initial state. The circuits are trained to maximize the expected fidelity between inputs and outputs, effectively performing a lossy compression of an n-qubit state into a k-qubit state.

Fault-tolerant quantum computers require error correction schemes that can deal with noisy and faulty operations. Leading proposals such as the color code and the surface code devote a large number of physical qubits to implement error-corrected logical qubits (see Gottesman [109] for an introduction to quantum error correction). Johnson *et al.* [110] suggest that a reduced overhead could be achieved in NISQ devices by training encoding and recovery circuits to optimize the average code fidelity.

The implementation of a quantum algorithm is also limited by the available gate set and qubit-to-qubit connectivity of the underlying hardware. This is where quantum compilers come into play, by abstracting the user from the low-level details. Khatari *et al.* [111] propose to train a hardware-efficient variational circuit $U_{\boldsymbol{\theta}}$ to approximately execute the same action as a target unitary $U = e^{-i\mathcal{H}t}$.

IV. OUTLOOK

Researchers have begun to explore connections between quantum supremacy proposals and quantum algorithms for optimization [112], getting us closer to practical utility if some key requirements can be met [113–115]. It is natural to explore similar connections between quantum supremacy and machine learning.

We have seen that parameterized quantum circuits (PQCs) can implement classically intractable feature maps and estimate intractable kernel functions. Further studies will be needed to assess whether these can improve the performance of supervised learning, in particular, kernel-based models such as support vector machines, Gaussian processes and principal component analysis.

We also know that PQCs inspired by instantaneous quantum polynomial-time circuits are classically intractable to simulate in the average case. A natural application for them is in generative modeling where the task itself requires sampling from complex probability distributions. But does classical intractability of these circuits imply an advantage in practice? To answer this question the community needs a demonstration on a real-world dataset of practical utility.

PQC models can also help in the study of quantum mechanical systems. For systems that exhibit quantum supremacy, a classical model cannot learn to reproduce the statistics unless it uses exponentially scaling resources. Provided that we can efficiently load or prepare quantum data in a qubit register, PQC models will deliver a clear advantage over classical methods for quantum learning tasks.

From the machine learning practitioner's point of view, there are several desirable properties that are naturally captured by PQC models. For example, recurrent neural networks may suffer from the exploding gradient problem. This can be prevented by constraining the operations to be unitary and much work has been done to efficiently parameterize the unitary group [116, 117]. PQC models have the advantage of naturally implementing unitary operations on an exponentially large Hilbert space. As another example, state-of-the-art classical generative models may not allow gradient-based training when the data is discrete [85]. In PQCs discrete data arises from measurements on the qubits and, as we have seen, this does not preclude the computation of gradients. We believe that this is only the tip of the iceberg and that there a number of research opportunities in this field. Largely unexplored aspects of PQC models include Vapnik-Chervonenkis dimensions, regularization techniques, Bayesian inference, and applications to reinforcement learning.

Finally, hybrid systems based on PQCs provide a framework for the incremental development of algorithms. In the near term, algorithms shall rely heavily on classical resources. As quantum hardware improves, classical resources shall gradually be replaced by quantum resources and generic methods. For example, Wang *et al.* [118] propose a method that interpolates between the near-term variational quantum eigensolver and the long-term quantum phase estimation. Similarly, destructive SWAP and Hadamard tests [64, 119] could be gradually replaced by their corresponding interference circuits. Hardware-efficient circuits shall be replaced by new parameterizations driven by the theory of tensor networks. Quantum compilers [120, 121] will enable the implementation of these higher level constructions on existing devices.

Ideas and examples presented in this Review show the remarkable flexibility of the hybrid framework and its potential to use existing hardware to its full extent. If PQC models can be shown to scale well to realistic ML tasks, they may become an integral part of automated forecasting and decision-making systems.

V. ACKNOWLEDGEMENTS

The authors would like to thank Tiya-Renee Jones for her help with Figures 1 and 7, Ilyas Khan for his support, and Mattia Fiorentini for helpful discussions. M.B. is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) and by Cambridge Quantum Computing Limited (CQC).

[1] John Preskill, "Quantum computing in the NISQ era and beyond," Quantum **2**, 79 (2018).
[2] Masoud Mohseni, Peter Read, Hartmut Neven, Sergio Boixo, Vasil Denchev, Ryan Babbush, Austin Fowler, Vadim Smelyanskiy, and John Martinis, "Commercialize quantum technologies in five years," Nature **543**, 171–174 (2017).
[3] AP Lund, Michael J Bremner, and TC Ralph, "Quantum sampling problems, bosonsampling and quantum supremacy," npj Quantum Information **3**, 15 (2017).
[4] Aram W Harrow and Ashley Montanaro, "Quantum computational supremacy," Nature **549**, 203 (2017).
[5] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L Obrien, "A variational eigenvalue solver on a photonic quantum processor," Nature communications **5**, 4213 (2014).
[6] PJJ O'Malley, Ryan Babbush, ID Kivlichan, Jonathan Romero, JR McClean, Rami Barends, Julian Kelly, Pedram Roushan, Andrew Tranter, Nan Ding, *et al.*, "Scalable quantum simulation of molecular energies," Physical Review X **6**, 031007 (2016).
[7] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M Chow, and Jay M Gambetta, "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets," Nature **549**, 242 (2017).
[8] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann, "A quantum approximate optimization algorithm," arXiv preprint arXiv:1411.4028 (2014).
[9] Nikolaj Moll, Panagiotis Barkoutsos, Lev S Bishop, Jerry M Chow, Andrew Cross, Daniel J Egger, Stefan Filipp, Andreas Fuhrer, Jay M Gambetta, Marc Ganzhorn, Abhinav Kandala, Antonio Mezzacapo, Peter Mller, Walter Riess, Gian Salis, John Smolin, Ivano Tavernelli, and Kristan Temme, "Quantum optimization using variational algorithms on near-term quantum devices," Quantum Science and Technology **3**, 030503 (2018).
[10] JS Otterbach, R Manenti, N Alidoust, A Bestwick, M Block, B Bloom, S Caldwell, N Didier, E Schuyler Fried, S Hong, *et al.*, "Unsupervised machine learning on a hybrid quantum computer," arXiv preprint arXiv:1712.05771 (2017).
[11] Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre GR Day, Clint Richardson, Charles K Fisher, and David J Schwab, "A high-bias, low-variance introduction to machine learning for physicists," Physics Reports (2019), https://doi.org/10.1016/j.physrep.2019.03.001.
[12] Michael A. Nielsen and Isaac L. Chuang, "Quantum computation and quantum information: 10th anniversary edition," (2011).
[13] Jeongho Bang, James Lim, MS Kim, and Jinhyoung Lee, "Quantum learning machine," arXiv preprint arXiv:0803.2976 (2008).
[14] Søren Gammelmark and Klaus Mølmer, "Quantum learning by measurement and feedback," New Journal of Physics **11**, 033017 (2009).
[15] Mark Fingerhuth, Tom Babej, and Peter Wittek, "Open source software in quantum computing," PLOS ONE **13**, 1–28 (2018).

[16] Maria Schuld, M Fingerhuth, and F Petruccione, "Implementing a distance-based classifier with a quantum interference circuit," Europhys. Lett. **119** (2017), 10.1209/0295-5075/119/60002.

[17] Edward Grant, Marcello Benedetti, Shuxiang Cao, Andrew Hallam, Joshua Lockhart, Vid Stojevic, Andrew G. Green, and Simone Severini, "Hierarchical quantum classifiers," npj Quantum Information **4**, 65 (2018).

[18] Vojtěch Havlíček, Antonio D Córcoles, Kristan Temme, Aram W Harrow, Abhinav Kandala, Jerry M Chow, and Jay M Gambetta, "Supervised learning with quantum-enhanced feature spaces," Nature **567**, 209 (2019).

[19] Dario Gerace Francesco Tacchino, Chiara Macchiavello and Daniele Bajoni, "An artificial neuron implemented on an actual quantum processor," npj Quantum Information **5**, 26 (2019).

[20] Marcello Benedetti, Delfina Garcia-Pintos, Oscar Perdomo, Vicente Leyton-Ortega, Yunseong Nam, and Alejandro Perdomo-Ortiz, "A generative modeling approach for benchmarking and training shallow quantum circuits," npj Quantum Information **5**, 45 (2019).

[21] Kathleen E Hamilton, Eugene F Dumitrescu, and Raphael C Pooser, "Generative model benchmarks for superconducting qubits," arXiv preprint arXiv:1811.09905 (2018).

[22] D Zhu, NM Linke, M Benedetti, KA Landsman, NH Nguyen, CH Alderete, A Perdomo-Ortiz, N Korda, A Garfoot, C Brecque, *et al.*, "Training of quantum circuits on a hybrid quantum computer," arXiv preprint arXiv:1812.08862 (2018).

[23] Vicente Leyton-Ortega, Alejandro Perdomo-Ortiz, and Oscar Perdomo, "Robust implementation of generative modeling with parametrized quantum circuits," arXiv preprint arXiv:1901.08047 (2019).

[24] Ling Hu, Shu-Hao Wu, Weizhou Cai, Yuwei Ma, Xianghao Mu, Yuan Xu, Haiyan Wang, Yipu Song, Dong-Ling Deng, Chang-Ling Zou, *et al.*, "Quantum generative adversarial learning in a superconducting quantum circuit," Science advances **5**, eaav2761 (2019).

[25] Andrea Rocchetto, Scott Aaronson, Simone Severini, Gonzalo Carvacho, Davide Poderini, Iris Agresti, Marco Bentivegna, and Fabio Sciarrino, "Experimental learning of quantum states," Science Advances **5**, eaau1946 (2019).

[26] Yongcheng Ding, Lucas Lamata, Mikel Sanz, Xi Chen, and Enrique Solano, "Experimental implementation of a quantum autoencoder via quantum adders," Advanced Quantum Technologies , 1800065 (2019).

[27] Diego Ristè, Marcus P da Silva, Colm A Ryan, Andrew W Cross, Antonio D Córcoles, John A Smolin, Jay M Gambetta, Jerry M Chow, and Blake R Johnson, "Demonstration of quantum advantage in machine learning," npj Quantum Information **3**, 16 (2017).

[28] Gadi Aleksandrowicz, Thomas Alexander, Panagiotis Barkoutsos, Luciano Bello, Yael Ben-Haim, David Bucher, Francisco Jose Cabrera-Hernádez, Jorge Carballo-Franquis, Adrian Chen, Chun-Fu Chen, Jerry M. Chow, Antonio D. Córcoles-Gonzales, Abigail J. Cross, Andrew Cross, Juan Cruz-Benito, Chris Culver, Salvador De La Puente González, Enrique De La Torre, Delton Ding, Eugene Dumitrescu, Ivan Duran, Pieter Eendebak, Mark Everitt, Ismael Faro Sertage, Albert Frisch, Andreas Fuhrer, Jay Gambetta, Borja Godoy Gago, Juan Gomez-Mosquera, Donny Greenberg, Ikko Hamamura, Vojtěch Havlíček, Joe Hellmers, Łukasz Herok, Hiroshi Horii, Shaohan Hu, Takashi Imamichi, Toshinari Itoko, Ali Javadi-Abhari, Naoki Kanazawa, Anton Karazeev, Kevin Krsulich, Peng Liu, Yang Luh, Yunho Maeng, Manoel Marques, Francisco Jose Martín-Fernández, Douglas T. McClure, David McKay, Srujan Meesala, Antonio Mezzacapo, Nikolaj Moll, Diego Moreda Rodríguez, Giacomo Nannicini, Paul Nation, Pauline Ollitrault, Lee James O'Riordan, Hanhee Paik, Jesús Pérez, Anna Phan, Marco Pistoia, Viktor Prutyanov, Max Reuter, Julia Rice, Abdón Rodríguez Davila, Raymond Harry Putra Rudy, Mingi Ryu, Ninad Sathaye, Chris Schnabel, Eddie Schoute, Kanav Setia, Yunong Shi, Adenilton Silva, Yukio Siraichi, Seyon Sivarajah, John A. Smolin, Mathias Soeken, Hitomi Takahashi, Ivano Tavernelli, Charles Taylor, Pete Taylour, Kenso Trabing, Matthew Treinish, Wes Turner, Desiree Vogt-Lee, Christophe Vuillot, Jonathan A. Wildstrom, Jessica Wilson, Erick Winston, Christopher Wood, Stephen Wood, Stefan Wörner, Ismail Yunus Akhalwaya, and Christa Zoufal, "Qiskit: An open-source framework for quantum computing," (2019).

[29] Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, and Nathan Killoran, "Pennylane: Automatic differentiation of hybrid quantum-classical computations," arXiv preprint arXiv:1811.04968 (2018).

[30] Viral B. Shah Jeffrey Sarnoff Roger Luo, Leo GiggleLiu, "Yao," `https://github.com/QuantumBFS/Yao.jl` (2018).

[31] Edwin Stoudenmire and David J Schwab, "Supervised learning with tensor networks," in *Advances in Neural Information Processing Systems 29*, edited by D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (Curran Associates, Inc., 2016) pp. 4799–4807.

[32] Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii, "Quantum circuit learning," Physical Review A **98**, 032309 (2018).

[33] Javier Gil Vidal and Dirk Oliver Theis, "Input redundancy for parameterized quantum circuits," arXiv preprint arXiv:1901.11434 (2019).

[34] CM Wilson, JS Otterbach, Nikolas Tezak, RS Smith, GE Crooks, and MP da Silva, "Quantum kitchen sinks: An algorithm for machine learning on near-term quantum computers," arXiv preprint arXiv:1806.08321 (2018).

[35] Ali Rahimi and Benjamin Recht, "Random features for large-scale kernel machines," in *Advances in neural information processing systems* (2008) pp. 1177–1184.

[36] Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd, "Quantum support vector machine for big data classification," Phys. Rev. Lett. **113**, 130503 (2014).

[37] Scott Aaronson, "Read the fine print," Nature Physics **11**, 291 EP – (2015).

[38] Martin Rötteler, "Quantum algorithms for highly nonlinear boolean functions," in *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete algorithms* (Society for Industrial and Applied Mathematics, 2010) pp. 448–457.

[39] Kurt Hornik, Maxwell Stinchcombe, and Halbert White, "Multilayer feedforward networks are universal

approximators," Neural networks **2**, 359–366 (1989).
[40] Henry W Lin, Max Tegmark, and David Rolnick, "Why does deep and cheap learning work so well?" Journal of Statistical Physics **168**, 1223–1247 (2017).
[41] Jin-Guo Liu and Lei Wang, "Differentiable learning of quantum circuit born machines," Phys. Rev. A **98**, 062324 (2018).
[42] C Chow and Cong Liu, "Approximating discrete probability distributions with dependence trees," IEEE transactions on Information Theory **14**, 462–467 (1968).
[43] Ding Liu, Shi-Ju Ran, Peter Wittek, Cheng Peng, Raul Blázquez García, Gang Su, and Maciej Lewenstein, "Machine learning by unitary tensor network of hierarchical tree structure," arXiv preprint arXiv:1710.04833 (2017).
[44] William Huggins, Piyush Patil, Bradley Mitchell, K Birgitta Whaley, and E Miles Stoudenmire, **4**, 024001 (2019).
[45] Pankaj Mehta and David J Schwab, "An exact mapping between the variational renormalization group and deep learning," arXiv preprint arXiv:1410.3831 (2014).
[46] Yoav Levine, Or Sharir, Nadav Cohen, and Amnon Shashua, "Quantum entanglement in deep learning architectures," Phys. Rev. Lett. **122**, 065301 (2019).
[47] Giuseppe Carleo and Matthias Troyer, "Solving the quantum many-body problem with artificial neural networks," Science **355**, 602–606 (2017).
[48] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven, "Barren plateaus in quantum neural network training landscapes," Nature communications **9**, 4812 (2018).
[49] Kurt Hornik, "Approximation capabilities of multilayer feedforward networks," Neural networks **4**, 251–257 (1991).
[50] Yudong Cao, Gian Giacomo Guerreschi, and Alán Aspuru-Guzik, "Quantum neuron: an elementary building block for machine learning on quantum computers," arXiv preprint arXiv:1711.11240 (2017).
[51] E. Torrontegui and J. J. García-Ripoll, "Unitary quantum perceptron as efficient universal approximator," EPL (Europhysics Letters) **125**, 30004 (2019).
[52] Geoffrey E Hinton, DE Rumelhart, and Ronald J Williams, "Learning representations by back-propagating errors," Nature **323**, 533–536 (1986).
[53] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione, "The quest for a quantum neural network," Quantum Information Processing **13**, 2567–2586 (2014).
[54] Guillaume Verdon, Jason Pye, and Michael Broughton, "A universal training algorithm for quantum deep learning," arXiv preprint arXiv:1806.09729 (2018).
[55] Kerstin Beer, Dmytro Bondarenko, Terry Farrelly, Tobias J Osborne, Robert Salzmann, and Ramona Wolf, "Efficient learning for deep quantum neural networks," arXiv preprint arXiv:1902.10445 (2019).
[56] James C Spall, "A one-measurement form of simultaneous perturbation stochastic approximation," Automatica **33**, 109–112 (1997).
[57] James C Spall, "Adaptive stochastic approximation by the simultaneous perturbation method," IEEE transactions on automatic control **45**, 1839–1853 (2000).
[58] Atilim Gunes Baydin, Barak A Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind, "Automatic differentiation in machine learning: a survey," Journal of Marchine Learning Research **18**, 1–43 (2018).
[59] Jun Li, Xiaodong Yang, Xinhua Peng, and Chang-Pu Sun, "Hybrid quantum-classical approach to quantum optimal control," Physical review letters **118**, 150503 (2017).
[60] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran, "Evaluating analytic gradients on quantum hardware," arXiv preprint arXiv:1811.11184 (2018).
[61] Edward Farhi and Hartmut Neven, "Classification with quantum neural networks on near term processors," arXiv:1802.06002 (2018).
[62] Pierre-Luc Dallaire-Demers and Nathan Killoran, "Quantum generative adversarial networks," Physical Review A **98**, 012324 (2018).
[63] Maria Schuld, Alex Bocharov, Krysta Svore, and Nathan Wiebe, "Circuit-centric quantum classifiers," arXiv preprint arXiv:1804.00633 (2018).
[64] Kosuke Mitarai and Keisuke Fujii, "Methodology for replacing indirect measurements with direct measurements," arXiv preprint arXiv:1901.00015 (2018).
[65] Aram Harrow and John Napp, "Low-depth gradient measurements can improve convergence in variational hybrid quantum-classical algorithms," arXiv preprint arXiv:1901.05374 (2019).
[66] Scott Wisdom, Thomas Powers, John R. Hershey, Jonathan Le Roux, and Les Atlas, "Full-capacity unitary recurrent neural networks," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16 (Curran Associates Inc., USA, 2016) pp. 4887–4895.
[67] M. Ledoux, *The Concentration of Measure Phenomenon*, Mathematical surveys and monographs (American Mathematical Society, 2001).
[68] Edward Grant, Leonard Wossnig, Mateusz Ostaszewski, and Marcello Benedetti, "An initialization strategy for addressing barren plateaus in parametrized quantum circuits," arXiv preprint arXiv:1903.05076 (2019).
[69] Herbert Robbins and Sutton Monro, "A stochastic approximation method," The annals of mathematical statistics , 400–407 (1951).
[70] Martin Riedmiller and Heinrich Braun, "A direct adaptive method for faster backpropagation learning: The rprop algorithm," in *IEEE International Conference on Neural Networks* (1993) pp. 586–591 vol.1.
[71] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).
[72] Russell C Eberhart and Xiaohui Hu, "Human tremor analysis using particle swarm optimization," in *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)*, Vol. 3 (IEEE, 1999) pp. 1927–1930.
[73] Peter I Frazier, "A tutorial on bayesian optimization," arXiv preprint arXiv:1807.02811 (2018).
[74] Yu-Ren Liu, Yi-Qi Hu, Hong Qian, Yang Yu, and Chao Qian, "Zoopt: Toolbox for derivative-free optimization," arXiv preprint arXiv:1801.00329 (2017).
[75] Kumara Sastry, David Goldberg, and Graham Kendall, "Genetic algorithms," in *Search methodologies* (Springer, 2005) pp. 97–125.
[76] Lucas Lamata, Unai Alvarez-Rodriguez, José David Martín-Guerrero, Mikel Sanz, and Enrique Solano, "Quantum autoencoders via quantum adders with genetic algorithms," Quantum Science and Technology **4**,

014007 (2018).
[77] Mateusz Ostaszewski, Edward Grant, and Marcello Benedetti, "Quantum circuit structure learning," arXiv preprint arXiv:1905.09692 (2019).
[78] Ken M Nakanishi, Keisuke Fujii, and Synge Todo, "Sequential minimal optimization for quantum-classical hybrid algorithms," arXiv preprint arXiv:1903.12166 (2019).
[79] Esma Aïmeur, Gilles Brassard, and Sébastien Gambs, "Machine learning in a quantum world," in *Conference of the Canadian Society for Computational Studies of Intelligence* (Springer, 2006) pp. 431–442.
[80] Maria Schuld and Nathan Killoran, "Quantum machine learning in feature hilbert spaces," Phys. Rev. Lett. **122**, 040504 (2019).
[81] Bernhard Schölkopf, Ralf Herbrich, and Alex J. Smola, "A generalized representer theorem," in *Computational Learning Theory*, edited by David Helmbold and Bob Williamson (Springer Berlin Heidelberg, Berlin, Heidelberg, 2001) pp. 416–426.
[82] Maxwell Henderson, Samriddhi Shakya, Shashindra Pradhan, and Tristan Cook, "Quanvolutional neural networks: Powering image recognition with quantum circuits," arXiv preprint arXiv:1904.04767 (2019).
[83] Song Cheng, Jing Chen, and Lei Wang, "Information perspective to probabilistic modeling: Boltzmann machines versus born machines," Entropy **20**, 583 (2018).
[84] Zhao-Yu Han, Jun Wang, Heng Fan, Lei Wang, and Pan Zhang, "Unsupervised generative modeling using matrix product states," Physical Review X **8** (2018), 10.1103/physrevx.8.031012.
[85] Ian Goodfellow, "Nips 2016 tutorial: Generative adversarial networks," arXiv preprint arXiv:1701.00160 (2016).
[86] Yuxuan Du, Min-Hsiu Hsieh, Tongliang Liu, and Dacheng Tao, "The expressive power of parameterized quantum circuits," arXiv preprint arXiv:1810.11922 (2018).
[87] Brian Coyle, Daniel Mills, Vincent Danos, and Elham Kashefi, "The born supremacy: Quantum advantage and training of an ising born machine," arXiv preprint arXiv:1904.02214 (2019).
[88] Christa Zoufal, Aurélien Lucchi, and Stefan Woerner, "Quantum generative adversarial networks for learning and loading random distributions," arXiv preprint arXiv:1904.00043 (2019).
[89] Jonathan Romero, Jonathan P Olson, and Alan Aspuru-Guzik, "Quantum autoencoders for efficient compression of quantum data," Quantum Science and Technology **2**, 045001 (2017).
[90] Solomon Kullback and Richard A Leibler, "On information and sufficiency," The annals of mathematical statistics **22**, 79–86 (1951).
[91] Arthur Gretton, Karsten M. Borgwardt, Malte Rasch, Bernhard Scholköpf, and Alexander J. Smola, "A kernel approach to comparing distributions," in *Proceedings of the 22Nd National Conference on Artificial Intelligence - Volume 2*, AAAI'07 (AAAI Press, 2007) pp. 1637–1641.
[92] Seth Lloyd and Christian Weedbrook, "Quantum generative adversarial learning," Physical review letters **121**, 040502 (2018).
[93] Haozhen Situ, Zhimin He, Lvzhou Li, and Shenggen Zheng, "Quantum generative adversarial network for generating discrete data," arXiv preprint arXiv:1807.01235 (2018).
[94] Jinfeng Zeng, Yufeng Wu, Jin-Guo Liu, Lei Wang, and Jiangping Hu, "Learning and inference on generative adversarial quantum circuits," Phys. Rev. A **99**, 052306 (2019).
[95] Jonathan Romero and Alan Aspuru-Guzik, "Variational quantum generators: Generative adversarial quantum machine learning for continuous distributions," arXiv preprint arXiv:1901.00848 (2019).
[96] Guang Hao Low, Theodore J Yoder, and Isaac L Chuang, "Quantum inference on bayesian networks," Physical Review A **89**, 062315 (2014).
[97] Mauro ES Morales, Timur Tlyachev, and Jacob Biamonte, "Variational learning of grover's quantum search algorithm," Physical Review A **98**, 062333 (2018).
[98] Kwok Ho Wan, Feiyang Liu, Oscar Dahlsten, and MS Kim, "Learning simon's quantum algorithm," arXiv preprint arXiv:1806.10448 (2018).
[99] Daniel R Simon, "On the power of quantum computation," SIAM journal on computing **26**, 1474–1483 (1997).
[100] Eric Anschuetz, Jonathan Olson, Alán Aspuru-Guzik, and Yudong Cao, "Variational quantum factoring," in *International Workshop on Quantum Technology and Optimization Problems* (Springer, 2019) pp. 74–85.
[101] 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).
[102] Hongxiang Chen, Leonard Wossnig, Simone Severini, Hartmut Neven, and Masoud Mohseni, "Universal discriminative quantum neural networks," arXiv preprint arXiv:1805.08654 (2018).
[103] Scott Aaronson, "The learnability of quantum states," Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **463**, 3089–3114 (2007).
[104] Leslie G Valiant, "A theory of the learnable," in *Proceedings of the sixteenth annual ACM symposium on Theory of computing* (ACM, 1984) pp. 436–445.
[105] Sang Min Lee, Jinhyoung Lee, and Jeongho Bang, "Learning unknown pure quantum states," Physical Review A **98**, 052302 (2018).
[106] Ryan LaRose, Arkin Tikku, Étude O'Neel-Judy, Lukasz Cincio, and Patrick J Coles, "Variational quantum state diagonalization," arXiv preprint arXiv:1810.10506 (2018).
[107] Carl W Helstrom, "Quantum detection and estimation theory," Journal of Statistical Physics **1**, 231–252 (1969).
[108] Marcello Benedetti, Edward Grant, Leonard Wossnig, and Simone Severini, "Adversarial quantum circuit learning for pure state approximation," New Journal of Physics (2019), 10.1088/1367-2630/ab14b5.
[109] Daniel Gottesman, "An introduction to quantum error correction and fault-tolerant quantum computation," in *Quantum information science and its contributions to mathematics, Proceedings of Symposia in Applied Mathematics*, Vol. 68 (2010) pp. 13–58.
[110] 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).
[111] Sumeet Khatri, Ryan LaRose, Alexander Poremba, Lukasz Cincio, Andrew T. Sornborger, and Patrick J. Coles, "Quantum-assisted quantum compiling," Quantum **3**, 140 (2019).
[112] Edward Farhi and Aram W Harrow, "Quantum supremacy through the quantum approximate optimization algorithm," arXiv preprint arXiv:1602.07674 (2016).
[113] G. G. Guerreschi and A. Y. Matsuura, "Qaoa for max-cut requires hundreds of qubits for quantum speed-up," Scientific Reports **9**, 6903 (2019).
[114] Leo Zhou, Sheng-Tao Wang, Soonwon Choi, Hannes Pichler, and Mikhail D Lukin, "Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices," arXiv preprint arXiv:1812.01041 (2018).
[115] Gavin E Crooks, "Performance of the quantum approximate optimization algorithm on the maximum cut problem," arXiv preprint arXiv:1811.08419 (2018).
[116] Li Jing, Yichen Shen, Tena Dubcek, John Peurifoy, Scott Skirlo, Yann LeCun, Max Tegmark, and Marin Soljačić, "Tunable efficient unitary neural networks (eunn) and their application to rnns," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70* (JMLR. org, 2017) pp. 1733–1741.
[117] Stephanie L Hyland and Gunnar Rätsch, "Learning unitary operators with help from u (n)," in *Thirty-First AAAI Conference on Artificial Intelligence* (2017).
[118] Daochen Wang, Oscar Higgott, and Stephen Brierley, "Accelerated variational quantum eigensolver," Phys. Rev. Lett. **122**, 140504 (2019).
[119] Juan Carlos Garcia-Escartin and Pedro Chamorro-Posada, "Swap test and hong-ou-mandel effect are equivalent," Physical Review A **87**, 052330 (2013).
[120] Alexander Cowtan, Silas Dilkes, Ross Duncan, Alexandre Krajenbrink, Will Simmons, and Seyon Sivarajah, "On the qubit routing problem," arXiv preprint arXiv:1902.08091 (2019).
[121] Raban Iten, Oliver Reardon-Smith, Luca Mondada, Ethan Redmond, Ravjot Singh Kohli, and Roger Colbeck, "Introduction to universalqcompiler," arXiv preprint arXiv:1904.01072 (2019).