



Universal discriminative quantum neural networks

H. Chen^{1,2,3} · L. Wossnig^{2,3} · S. Severini^{2,4} · H. Neven⁵ · M. Mohseni⁵

Received: 5 August 2019 / Accepted: 3 September 2020 / Published online: 15 December 2020
© The Author(s) 2020

Abstract

Recent results have demonstrated the successful applications of quantum-classical hybrid methods to train quantum circuits for a variety of machine learning tasks. A natural question to ask is consequentially whether we can also train such quantum circuits to discriminate quantum data, i.e., perform classification on data stored in form of quantum states. Although quantum mechanics fundamentally forbids deterministic discrimination of non-orthogonal states, we show in this work that it is possible to train a quantum circuit to discriminate such data with a trade-off between minimizing error rates and inconclusiveness rates of the classification tasks. Our approach achieves at the same time a performance which is close to the theoretically optimal values and a generalization ability to previously unseen quantum data. This generalization power hence distinguishes our work from previous circuit optimization results and furthermore provides an example of a quantum machine learning task that has inherently no classical analogue.

Keywords Quantum computing · Quantum machine learning · Quantum data classification · Quantum sensing

1 Introduction

Quantum computation has been shown to provide speedups in several applications over classical computation in the query model. Besides the famous Shor's algorithm for prime number factorization, quantum computers can also produce statistical patterns that are hard to produce for classical devices. This raises the possibility that quantum computers can also recognize patterns that are hard to recognize for classical computers, or, in general, that quantum computers can help solve classical machine learning problems more efficiently. Recently, the intersection of

quantum computation and machine learning has received a considerable amount of attention. Using the circuit model of computation, several quantum algorithms have been designed that in principle provide quadratic to exponential speedups on classical data (Biamonte et al. 2017; Ciliberto et al. 2018).

A related area is concerned with developing novel machine learning methods that operate on quantum data. In general, any set of quantum states which encode meaningful information can be considered quantum data. To motivate this direction, we want to emphasize that using quantum states as a storage medium for information has been demonstrated to provide advantages in several ways. For example, by coupling a quantum state with another target system, we can obtain information about the target system with increased sensitivity. Quantum metrology allows for example for a quadratic improvement over classical methods in terms of the statistical sampling error, i.e., the scaling of the standard deviations in estimates obtained through repeated measurements. Another example is quantum sensing, which provides much higher sensitivity for tasks like target detections in microwaves, i.e., quantum radar (Barzanjeh et al. 2015), and in general, sensing electric or magnetic fields (Degen et al. 2017). A practical application for these methods is the reduction of damage to pictures which are sensitive to the exposure of light (Schaller and Schützhold 2006).

✉ L. Wossnig
l.wossnig@cs.ucl.ac.uk

M. Mohseni
mohseni@google.com

¹ Department of Physics & Astronomy, University College London, London, UK

² Department of Computer Science, University College London, London, UK

³ Rahko Limited, The Coalface, Clifton House, 46 Clifton Terrace, Finsbury Park, N4 3JP, London, UK

⁴ Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai, China

⁵ Google Quantum AI Laboratory, Venice, CA, USA

Certain types of datasets are inherently quantum mechanical. Such data could, for example, be the output of quantum information processing procedures such as simulation of quantum materials, or quantum chemistry more generally. For such datasets, we conjecture the inherent advantage of quantum computers to perform recognition and classification tasks. For example, topological materials made in the exotic topological phase have non-classical electronic properties and are promising materials to build fault-tolerant quantum computers (Qi and Zhang 2011; Karzig et al. 2017). Predicting the phase of topological materials has been a very challenging problem for classical approaches. However, it has recently been shown that quantum neural networks could be used to recognize the phase of a quantum state (Cong et al. 2019) and hence for predicting this phase. In addition, the promised security of quantum communication protocols and a surge of ideas in quantum communication networks (Kimble 2008; Ren et al. 2017) further stimulates the research into areas dealing with inherently quantum data.

In this work, we explore the general problem of classifying quantum data. This problem can be seen as an extension of the established field of *quantum state discrimination*, which identifies a quantum state among a set of a priori completely known candidate states. A key challenge for the discrimination of quantum states is that a deterministic discrimination is impossible when the complex vectors representing the input states are non-orthogonal, i.e., when their overlaps are non-zero. Quantum state discrimination then allows finding the measurement that optimally discriminates these states. Note that we will use in the following input data and (quantum) states interchangeably.

However, it is not possible to directly apply quantum state discrimination to classify states, i.e., quantum data. First, it is inappropriate to assume that one possesses the complete knowledge of the input data a priori, which are often only samples generated from a data collecting process. Also, even with all the input data available as quantum states, performing quantum tomography on them is prohibitively expensive. In addition, *quantum state discrimination* often fails to give the optimal discriminative measurement in an analytically closed form, unless the quantum states are already orthogonal or possess certain symmetry properties (Barnett and Croke 2009). In case it fails, one may use numerical optimization to find the optimal measurement. However, the exponential growth of the dimensionality of the density matrices renders the numerical optimization also inefficient if performed on a classical device.

Due to the limitations of quantum state discrimination, it is natural to ask whether we can use a quantum computer

to help with the optimization procedure. Since fully error corrected quantum computers are not available yet, a recent stream of works proposed various applications for circuit learning (Banchi et al. 2016; Wan et al. 2017; Innocenti et al. 2018; Romero et al. 2017; Mitarai et al. 2018; Farhi and Neven 2018; Verdon et al. 2017; Li and Benjamin 2017; Grant et al. 2018; Schuld et al. 2018; Xu et al. 2019; Khatri et al. 2019), which constitutes a form of quantum-classical hybrid neural network that have been shown to be less prone to the inherent errors of early-stage quantum hardware. In this work, we similarly utilize a hybrid approach to learn the design of a shallow quantum circuit for the classification of quantum states. Concretely, this hybrid scheme consists of a classical computer which interactively changes the parameters of a quantum circuit in order to optimize the output of the quantum computation. In other words, we train a quantum circuit to classify the states correctly.

The approach we take is novel in two ways. First, we use a quantum circuit ansatz that is designed for an implementation on near-term devices (details available in Appendix 1). This ansatz allows for a shallow circuit but is still universal; i.e., it can perform any unitary transformation allowed by quantum mechanics. It comprises gates from a universal gate set consisting of C-NOT and single-qubit gates, which is motivated by the fact that their implementations are known for the current mainstream experimental architectures. It is furthermore nearly optimal in terms of the number of C-NOT gates, which is an important feature for an implementation on near-term devices. Second, unlike previous works on quantum state discrimination, we focus on the generalization ability of our circuit; i.e., we train the circuit on a specific range of the parameters with the goal of maximizing its generalization performance, and hence in a learning setting. This distinguishes our work from the pure optimization problem for the state discrimination task, i.e., optimizing the circuit to distinguish only a concrete set of states. We show here that this universal quantum circuit can be trained as a discriminator for classification of non-orthogonal quantum data which is sampled from various different probability distributions. Our discriminator can achieve a near-zero error rate by producing inconclusive signals.

2 Dataset and framework

In this work, we propose a novel approach for training a universal quantum circuit to classify quantum data, which is stored in qubits. In this section, we first introduce the mathematical notation and description of quantum data. We then specify the quantum data we use in this work for classification. Next, we outline the approach we

take to optimize a universal quantum circuit which is used to classify the quantum data. We defer the detailed decomposition of this quantum circuit to Appendix 1.

Mathematical descriptions Quantum data are collections of quantum states which store useful information. For these data, we may assume that their density matrices ρ are parameterized by parameters a which follow a probability distribution α specific to the carried information. Then, for classification, we are normally presented with an unknown quantum state ρ_x , which belongs to a family of quantum states, each described mathematically by:

$$\rho_i(a_i), a_i \sim \alpha_i, \tag{1}$$

where i is the label for the corresponding family, and $\rho_i(a_i)$ is the density matrix describing the quantum state in the family i , parametrized by a_i , and the parameters a_i are assumed to follow the probability distribution α_i . The purpose of a classifier is to identify the family x . Note that to train the classifier, we draw samples of $\rho(a_i)$ according to the distribution α_i .

Transformations of quantum states are described by a unitary matrix U , which transforms a quantum state ρ according to the rule $\rho \rightarrow U\rho U^\dagger$.

A measurement on the quantum state ρ is described by a set of matrices $\{M_j\}$, which are Hermitian, positive semi-definite, and sum to the identity. Here, j labels the possible measurement outcomes, and the probability p_j for the measurement outcome j is given by $p_j = \text{Tr}(M_j\rho)$. Such a collection of matrices M_j is commonly called a *positive-operator valued measure* (POVM). A common example of POVM is a *projection-valued measure* (PVM). In the case of a PVM, each M_j is a projector into some linear subspace and different M_j are orthogonal to each other, i.e., $M_j M_i = \delta_{ij} M_j$. With the help of ancilla qubits, any POVM could be realized by a quantum circuit consisting of a series of unitary matrices (transformations) and measurements in the computational basis. Conversely, a quantum circuit which consists of a parameterized set of gates and measurements could also represent a range of different POVMs. There exists a quantum circuit which could represent any POVM with a fixed number of possible measurement outcomes. Such a circuit is called a universal discriminator in this paper, and the specific one we chose to use here is discussed in Appendix 1.

Dataset For this work, we restrict our attention to the classification of two families of quantum states stored in a 2-qubit system. Our first family consists of pure states, parametrized by a real number $a \in [0, 1]$:

$$\psi_1(a) = (\sqrt{1 - a^2}, 0, a, 0), \rho_1 = |\psi_1(a)\rangle\langle\psi_1(a)|. \tag{2}$$

The second family consists of mixed states $\rho_2(b)$ where $b \in [0, 1]$. Specifically,

$$\begin{aligned} \psi_{2/3} &= (0, \pm\sqrt{1 - b^2}, b, 0), \\ \rho_2(b) &= \frac{1}{2}|\psi_2\rangle\langle\psi_2| + \frac{1}{2}|\psi_3\rangle\langle\psi_3|. \end{aligned} \tag{3}$$

The overlap between ψ_1 and $\psi_{2/3}$ is ab , indicating that the two families of states are non-orthogonal. For the case of a fixed a and $b = \frac{1}{\sqrt{2}}$, the maximal success rate for unambiguously discriminating between ρ_1 and ρ_2 has been studied theoretically, and experimentally demonstrated (Mohseni et al. 2004). The specific distributions we have tested in our experiments are summarized in Table 1. To generate the data for the training, validation, and testing of our circuits, we randomly and independently sampled points from the corresponding distributions.

Approach Overall, there are two major strategies to cope with our inherent inability to perform deterministic discrimination of quantum states: (a) *Minimum-error discrimination*: In this strategy, the task is to minimize the probability that the inevitable errors occur in the classification. (b) *Unambiguous discrimination*: In this strategy, the discriminator has one more output prediction than the number of classes: the inconclusive outcome. The task is to eliminate the error rate of the discriminator while minimizing the probability of this inevitable inconclusive outcome. A pure unambiguous discrimination with strictly zero error rate is not guaranteed to be possible for arbitrary quantum data. From the perspective of numerical optimization, one hence needs to allow for some small but non-zero errors.

In this work, we use a machine learning approach for training a universal quantum circuit capable of giving any quantum measurements with four possible measurement outcomes $m_{i_2 i_1}$, where $i_1, i_2 \in \{0, 1\}$ are the measurement outcomes of the first and the second qubits respectively. The parameterization of this circuit is discussed in Appendix 1. By assuming that input $\rho_1(a)$ produces the output m_{00} or m_{10} , input ρ_2 produces the output m_{01} , and assuming that m_{11} is the inconclusive output, this circuit acts as a discriminator for our experiment datasets. Therefore, we could trivially define various probabilities (success probability P_{suc} , error probability P_{err} , and inconclusive probability P_{inc}) with respect to the input (training) data with known class label. For example, when ρ_1 is the input, the probability of detecting m_{01} is the P_{err} , and the probability of detecting m_{11} is the P_{inc} . In this work, we perform experiments on simulated quantum computers, where these probabilities are available since the whole state is stored and processed on a classical computer. We note that on real quantum computers, these probabilities need to

be estimated through repeated measurements up to some precision, and require repeated data input.

To train the circuit, we use a heuristically motivated loss function defined in Eq. 4, which is the averaged absolute difference between the desired probabilities and the measured probabilities. It contains hyperparameters α_{err} and α_{inc} to balance between the erroneous outcomes and the inconclusive outcomes:

$$\begin{aligned}
 J = & \sum_i \frac{1}{|S_i|} \sum_{a_i \in S_i} |P_{suc}(\rho_i(a_i)) - 1| \\
 & + \alpha_{err} \sum_i \frac{1}{|S_i|} \sum_{a_i \in S_i} |P_{err}(\rho_i(a_i)) - 0| \\
 & + \alpha_{inc} \sum_i \frac{1}{|S_i|} \sum_{a_i \in S_i} |P_{inc}(\rho_i(a_i)) - 0|. \tag{4}
 \end{aligned}$$

Here, we assume that for each family of quantum states, we are given a set S_i of training samples, where each class is labelled by i . We denote with $|S_i|$ the cardinality of this set, i.e., the number of samples in the training set S_i , α_{err} is the penalty for making errors, and α_{inc} is the penalty for giving inconclusive outcomes. $P_{suc}(\rho_i)/P_{err}(\rho_i)/P_{inc}(\rho_i)$ are the probabilities of giving a correct/erroneous/inconclusive measurement outcome for the specific input quantum data ρ_i . This loss function measures the performance of our quantum circuit as a minimal-error discriminator (when $\alpha_{err} < \alpha_{inc}$) or as an unambiguous discriminator (when $\alpha_{err} > \alpha_{inc}$).

To train this circuit, we use the Adam optimization algorithm (Kingma and Ba 2014), and we calculate the gradients using the forward difference formula.

For our specific problem of classifying ρ_1 and ρ_2 as defined in Eqs. 2 and 3, we define an extra set of success/erroneous/inconclusive rates in Eq. 5 to summarize and compare the performance of different instances of the training process:

$$\begin{aligned}
 P_s &= \frac{1}{3}P_s(\rho_1)_{avg} + \frac{2}{3}P_s(\rho_2)_{avg} \\
 &= \frac{1}{3}P_s(\psi_1)_{avg} + \frac{1}{3}P_s(\psi_2)_{avg} + \frac{1}{3}P_s(\psi_3)_{avg}, \tag{5}
 \end{aligned}$$

where s stands for *suc* (successful), *err* (erroneous), or *inc* (inconclusive). The subscript *avg* means that the probabilities are calculated as the average value for all samples of either the training set, or the test set (but not both). The choice of weights ($\frac{1}{3}$ and $\frac{2}{3}$) in the Eq. 5 was made to be consistent with the results in Mohseni et al. (2004).

3 Theoretical analysis

Here, we describe a theoretical result to which we will compare our numerical results. In the general case, assume

we have a family (or class) of quantum data $\rho(a)$, each one parameterized by a and occurring with a probability $P(a)$. Assume in addition that we have a quantum measurement described by a POVM with elements $\{\Pi_i\}_{i \in \mathbb{N}}$, where i labels different measurement outcomes. Then, the probability of detecting measurement outcome i , averaged over any of the input data $\rho(a)$, is:

$$\begin{aligned}
 \int \text{Tr}(\Pi_i \rho(a)) P(a) da &= \text{Tr} \left[\int \Pi_i \rho(a) P(a) daa \right] \\
 &= \text{Tr} \left[\Pi_i \int \rho(a) P(a) da \right] \\
 &= \text{Tr} [\Pi_i \rho], \tag{6}
 \end{aligned}$$

where $\rho = \int \rho(a) P(a) da$, and the integration of the matrix is done in an element-wise fashion. Therefore, if $\text{Tr}(\Pi_i \rho) = 0$ for some i , then $\int_D \text{Tr}(\Pi_i \rho(a)) P(a) = 0$ for any subset D with non-zero measure in the whole parameter space of a . This is due to the fact that $\text{Tr}[\Pi_i \rho(a)] P(a) \geq 0$ for any parameter a .

The analysis above shows that the problem of unambiguously discriminating $\rho_1 = \int_a \rho_1(a) P_1(a) da$ and $\rho_2 = \int_a \rho_2(a) P_2(a) db$, is equivalent to the problem of unambiguously discriminating the family $\rho_1(a)$, $\forall a$, from the family $\rho_2(b)$, $\forall b$, where $P_1(a)/P_2(b)$ is the probability of occurrence of $\rho_1(a)/\rho_2(b)$. That is, if $\{\Pi_1, \Pi_2, \Pi_{inc}\}$ is a POVM that unambiguously classifies all members of the two families $\rho_1(a)$ and $\rho_2(b)$, for all possible parameters, i.e., Π_{inc} corresponds to the inconclusive outcome with:

$$\begin{aligned}
 \text{Tr}(\Pi_2 \rho_1(a)) &= 0, \forall a, \\
 \text{Tr}(\Pi_1 \rho_2(b)) &= 0, \forall b,
 \end{aligned}$$

then $\text{Tr}(\Pi_1 \rho_2) = \text{Tr}(\Pi_2 \rho_1) = 0$, and vice versa. Using this formalism, we can theoretically analyze the different cases we described in Table 1 based on the works of Raynal et al. (2003) and Barnett and Croke (2009), and the results are displayed in Table 2. Note that these are average case success probabilities.

Table 1 A summary of different test cases we classify in this work

	Family 1 ($\rho_1(a)$)	Family 2 ($\rho_2(b)$)
Case 1	$a \in [0, 1]$	$b \in [0, 1]$
Case 2	$a \approx 0.25$	$b \approx \frac{1}{\sqrt{2}}$
Case 3	$a \approx 0.25$	$b \in [0, 1]$
Case 4	$a \in [0, 1]$	$b \approx \frac{1}{\sqrt{2}}$

Here, $a(b) \in [0, 1]$ signifies that $a(b)$ follows a uniform distribution in $[0, 1]$. $a(b) \approx 0.25(\frac{1}{\sqrt{2}})$ signifies that $a(b)$ follows a normal distribution with mean $0.25(\frac{1}{\sqrt{2}})$, standard deviation 0.05, and which is truncated in $[0, 1]$

Table 2 A summary of maximal success rate when the error rate is exactly 0 for the different test cases classified in this work

	Family 1 (ρ_1)	Family 2 (ρ_2)	P_{suc}
Case 1	$a \in [0, 1]$	$b \in [0, 1]$	0.67
Case 2	$a \approx 0.25$	$b \approx \frac{1}{\sqrt{2}}$	0.64
Case 3	$a \approx 0.25$	$b \in [0, 1]$	0.76
Case 4	$a \in [0, 1]$	$b \approx \frac{1}{\sqrt{2}}$	0.55

They were calculated using Eq. 5 and obtained theoretically using methods described in Section 3

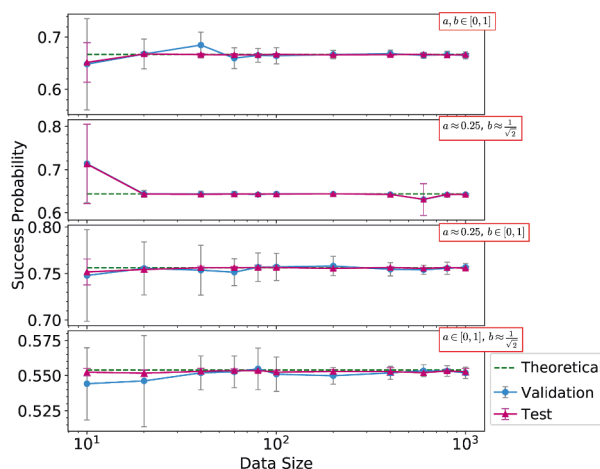
4 Numerical results

In this work, we aim to train a universal discriminator to discriminate different families of quantum data. Here, we present the results of training the universal discriminator to discriminate different distributions summarized in Table 1 on a quantum computer. The training is done by simulating the evolution of the quantum system under the parametrized circuits in a classical computer. To balance between eliminating the error rate (P_{err}) while minimizing the inconclusive rate (P_{inc}), we use a specific training strategy described in the following. We first prioritize a smaller inconclusive rate by starting with a zero penalty for erroneous outcomes ($\alpha_{inc} > \alpha_{err} = 0$), and then increase the α_{err} in a step-wise manner until a certain objective error rate is achieved. Similar optimization procedures have been used in the context of variational auto-encoders both in classical machine learning (Sønderby et al. 2016), and in quantum machine learning applications (Rocchetto et al. 2018). Using this scheme, we train our circuit to unambiguously discriminate the two families of quantum states and observe the convergence toward the theoretical success rates for the discriminator obtained in Section 3 with an increasing amount of training data. Notably, we do not observe any signs of overfitting despite the varying size of the training dataset (Fig. 1a).

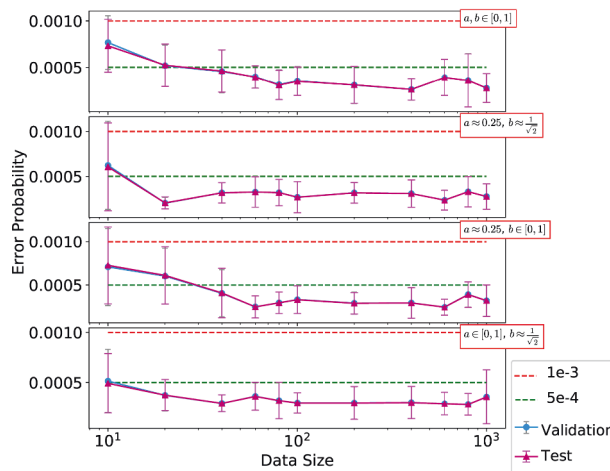
Trade-off between the error rate and the inconclusive rate

Here, we show that our model is able to obtain a much higher success rate (P_{suc}) if we allow a slightly higher error rate compared with the previous results. This hints at a trade-off between the error rate (P_{err}) and the inconclusive rate (P_{inc}) which can be utilized in real-world applications.

Specifically, for the dataset “Case 4” in Table 1, we fix the two penalties, α_{err} and α_{inc} , during the training and observed a gradual transition from unambiguous-like classification (characterized by a near-zero error probability) to minimal-error-like classification (characterized by the near-zero inconclusiveness) when we use varying penalties (Fig. 2a–c) throughout the different trainings with random initializations. Allowing a small error rate results then in



(a) The success rates converge to the theoretical result (the green dashed line in the figure) as the training dataset size is increased. The performance on the testing (purple line) and the validation dataset also converge with the growing dataset size. The abnormal deviation at the point with data size 600 for the case $a \approx 0.25$ and $b \approx 1/\sqrt{2}$ is due to unsuccessful optimization instances.



(b) Most of the error rate for the corresponding trainings are well below 0.0005, indicating an unambiguous discrimination has been achieved.

Fig. 1 Unambiguous classification of non-orthogonal quantum data sampled from different probability distributions. The data is averaged over 10 repeated trails starting with random initializations and the bars indicate the standard deviations. The training, validation, and test datasets are sampled from the corresponding distributions. The dataset size indicates the size of the training and the validation dataset. The test dataset is fixed and had a size of 10^4 samples for each family and each distribution

a much higher success rate, which has not been predicted theoretically. We note that introducing the penalty terms α_{err} and α_{inc} also makes the training process more stable (Fig. 2a). Therefore, the hyperparameters α_{err} and α_{inc} act

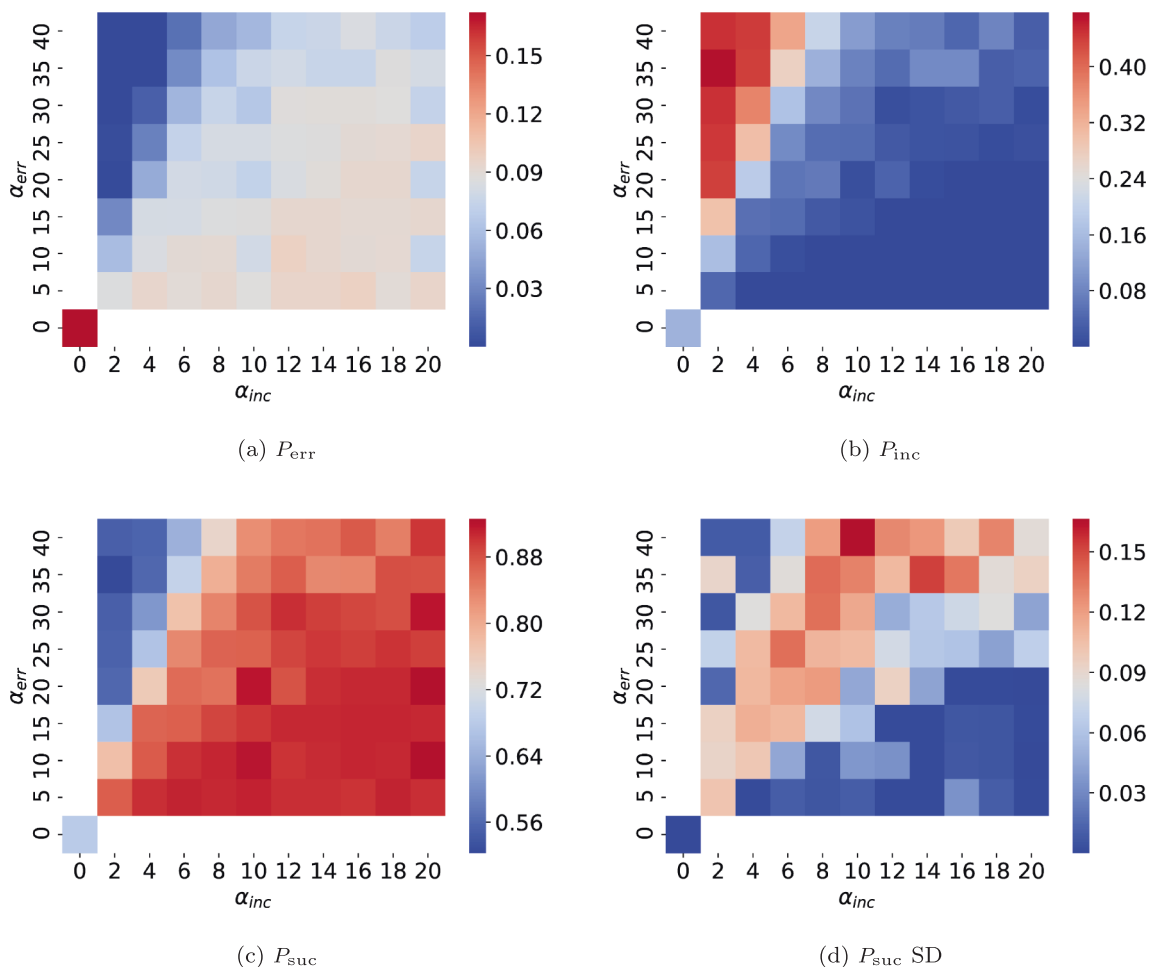
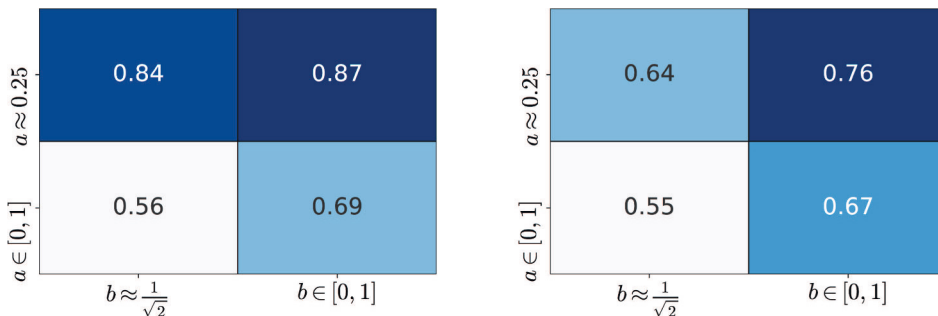


Fig. 2 With different penalties, we observe a trade-off between the error rate and the inconclusive rate. Compared with the point $\alpha_{err} = \alpha_{inc} = 0$ (bottom left corner), the added penalties improve the success probability or the inconclusiveness respectively. **a–c** The gradual transition from the unambiguous classification (near-zero error rate, top left corner) to a minimal error classification (near-zero inconclusiveness, bottom right corner) with changes in the error penalty α_{err} and the inconclusiveness penalty α_{inc} . We observe that the gain in the

success rate is around 0.32 when we make a sacrifice of only 0.1 in the error rate. We let $a \in [0, 1]$, and average over 50 repeated trails with random initializations. **d** Standard deviation for P_{suc} . With an increasing standard deviation (closer to the diagonal line), the result becomes increasingly unstable when the two penalties (α_{err} and α_{inc}) are closer in value. The standard deviations for P_{err} and P_{inc} show the same pattern as for P_{suc} (not shown)

Fig. 3 Unambiguous discrimination of data sampled from different probability distributions with higher success rate. **a** Trained quantum circuits are capable of classifying quantum data which is sampled from a variety of different mixed probability distributions for $\rho_1(a)$ and $\rho_2(b)$. The classification is done in an unambiguous manner (with error rate < 0.01). **b** For comparison, we include here the theoretical result mentioned in Table 2



(a) Success rate for trained circuit for classification of data

(b) The theoretical success rate when exact 0 error rate is achieved.

as a form of regularization and could be adjusted to give a higher success probability or a lower inconclusiveness rate for the final model (Fig. 2).

Furthermore, similar trade-off effects exhibited in all datasets are listed in Table 1. If we stop the training once the error rate drops below 0.01, we can achieve a much higher success rate than the theoretical case of exactly zero error rate (Fig. 3).

5 Learning convergence from ensemble measurements

We additionally perform experiments in which we estimate the probabilities from repeated measurements on the (simulated) quantum device. We find that the noise in gradient calculation which is caused by these estimated probabilities could be effectively countered by increasing the number of repeated measurements, using a lower error rate, and adjusting the step size in the forward difference formula. The detailed discussion is available in Appendix 2. Therefore, our study here appears to be feasible to be run on error-corrected quantum devices. We leave open the effects of machine noise (the noise caused by imperfect quantum devices), and an actual implementation as future projects.

6 Conclusions

We have developed a quantum circuit learning approach for the classification of quantum data. Specifically, we have designed a heuristically motivated loss function and used the stochastic optimization algorithm Adam in a quantum-classical hybrid scheme to train a circuit to perform quantum state discrimination. This training process generalizes well for the discrimination tasks on new data, i.e., states from the parameter range which have not been seen during the training process. This distinguishes our work from previous results on quantum circuit learning, in particular the very recent study in Fanizza et al. (2018), which only optimizes circuits for specific inputs. Note that this prior work hence does not consider the generalization ability and hence does not treat the actual learning problem, which aims at optimization as well as generalization.

In our work, we observe a trade-off between the error rates and the inconclusive rates when we penalize them differently in the loss function. Although this experiment is done on simulated quantum computers where exact measurement probabilities are available, we show that this optimization could be experimentally performed with repeated measurements of the quantum states. We note that the recent quantum methods for estimating the analytical gradient via variations in the unitaries (Mitarai et al. 2018)

can be directly applied to training our circuits; therefore, one can perform the optimization efficiently on near-term quantum devices. Also, although the Adam optimization algorithm is shown to be sufficient for the experiments conducted in this paper, several optimization algorithms specific to variational hybrid quantum-classical algorithms have been proposed and may provide improvements in more complicated cases (see for example Kübler et al. (2019)).

In this work, we have not addressed the issue of scalability of classifying quantum states. However, we expect most kinds of quantum data of interests will only require polynomial-depth circuits for classifying them. For example, it is likely that an ansatz based on the idea of tensor networks (e.g., Grant et al. (2018) and Cong et al. (2019)) can classify the different phases of ground states of quantum many-body systems in polynomial depth. Also, a scheme where one systematically increases the depth of the ansatz circuit will help explore the required circuit depth for classifying quantum data. A similar idea has been explored in the context of variational quantum eigensolver (Ostaszewski et al. 2019).

We believe that with the progress on technologies for preservation and transportation of quantum states, we will see many applications of a trained discriminative quantum circuits introduced here. Quantum state discrimination by itself plays a key role in quantum information processing protocols and is used in quantum cryptography (Bennett 1992a), quantum cloning (Duan and Guo 1998), quantum state separation, and entanglement concentration (Chefles 2000). Our work can provide improvements on these traditional areas by producing a classifier that is resilient to the statistical noise found in the actual communication. For example, we can consider an improved version of the B92 quantum key distribution protocol (Bennett 1992b) by including the noise-induced randomness in its two quantum keys and classify them with our discriminative circuit. Furthermore, we can consider training a discriminative quantum circuit used to construct quantum repeaters and state purification units within quantum communication networks. The training can take quantum data that have noise specific to the communication networks and therefore produces a discriminator that can recognize and filter those noise to provide better performance. Our discriminator can also be used to verify the output of other generative models, such as the quantum version of Boltzmann machines (Amin et al. 2018), or generative artificial neural networks (Goodfellow et al. 2014; Lloyd and Weedbrook 2018).

Acknowledgments We want to thank Raban Iten, Oliver Reardon-Smith, and Roger Colbeck for valuable insights in parametrizing the general measurement circuits and Jarrod McClean for feedback on the manuscript. This project acknowledges the use of the EPSRC funded Tier 2 facility JADE, the use of the UCL Legion High Performance

Computing Facility (Legion@UCL), and associated support services, in the completion. This work has been carried out while L.W. and S.S. participated in the workshop of *Measurement and control of quantum systems* at the Institut Henri Poincare. The financial support is kindly acknowledged.

Funding L.W. is supported by the Royal Society. S.S. is supported by the Royal Society, EPSRC, the National Natural Science Foundation of China, and the grant ARO-MURI W911NF-17-1-0304 (US DOD, UK MOD and UK EPSRC under the Multidisciplinary University Research Initiative).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix 1: Quantum circuits for POVM

This section describes the parametrization of the circuit capable of performing any quantum measurement on 2-qubit inputs with 4 possible measurement outcomes. This circuit could be represented by the following circuit diagram:

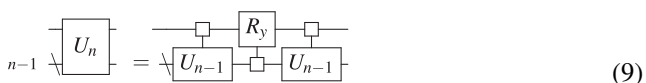


1.1 Cosine-sine decomposition

Here, we mention the cosine-sine decomposition of unitary matrices, which will be frequently used in the following sections. For every unitary matrix $U \in \mathbb{C}^{2^n \times 2^n}$, it can be decomposed as:

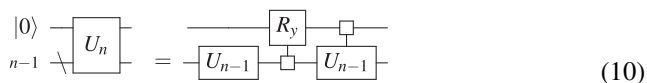
$$U_n = \begin{pmatrix} A_0 & 0 \\ 0 & A_1 \end{pmatrix} \begin{pmatrix} C & -S \\ S & C \end{pmatrix} \begin{pmatrix} B_0 & 0 \\ 0 & B_1 \end{pmatrix} \tag{8}$$

where A_0, A_1, B_0, B_1 are unitary matrices of size $2^{n-1} \times 2^{n-1}$, C and S are real diagonal matrices of size $2^{n-1} \times 2^{n-1}$ satisfying $C^2 + S^2 = \mathbb{1}$. It can be written in the following circuit equivalence diagram:



Here, a box \square represents the control part of a uniformly controlled gate; see section IV of (Iten et al. 2015) for

details. In the circuit in Eq. 7, the first qubit is initiated to be $|0\rangle$, so we have:

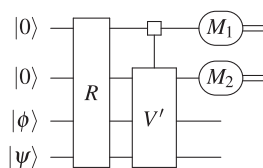


1.2 Decomposition of the circuit in Eq. 7

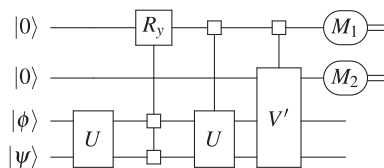
For a general measurement giving at most 4 measurement outcomes, we have the following circuit representation:



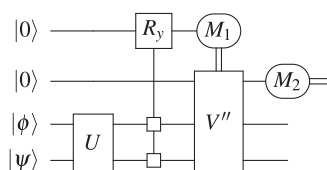
The first V could be decomposed using the circuit equivalence on page 5 of Iten et al. (2016) into:



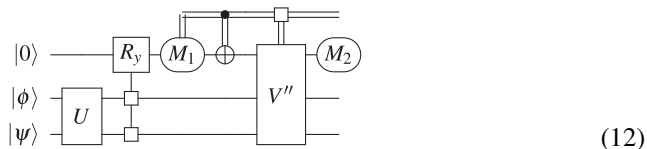
where the R gate does not act on the second qubit. Applying the cosine-sine decomposition gives:



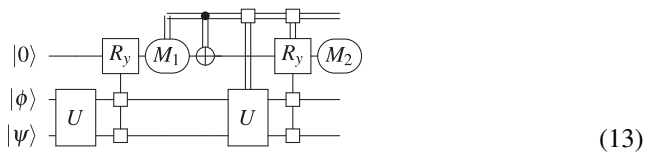
The uniformly controlled V' and U can be merged and put after the measurement of M_1 as:



The first line of the circuit could be merged with the second line as follows:



And then we can apply the cosine-sine decomposition to V'' . Throwing away the last gate on the third and the fourth qubits, we obtain:



The uniformly controlled rotations and the remaining two-qubit unitary gates could be easily parametrized by CNOTs and single-qubit rotations. For example, see Shende et al. (2006) and Shende et al. (2004).

Appendix 2: Learning convergence from ensemble measurements

Here, we simulate the process that a classical-quantum hybrid scheme would implement utilizing a quantum device and analyze its performance. These numerical simulations can in principle be validated in a physical experiment, where the measurement outcomes are used to infer the different probabilities for the cost function. To have a good estimation of the probabilities, and hence the cost function, one has to make repeated measurements to train the model, and we note that in particular better methods to evaluate the analytical gradient are available on a shallow quantum device (Mitarai et al. 2018). We first give a brief discussion

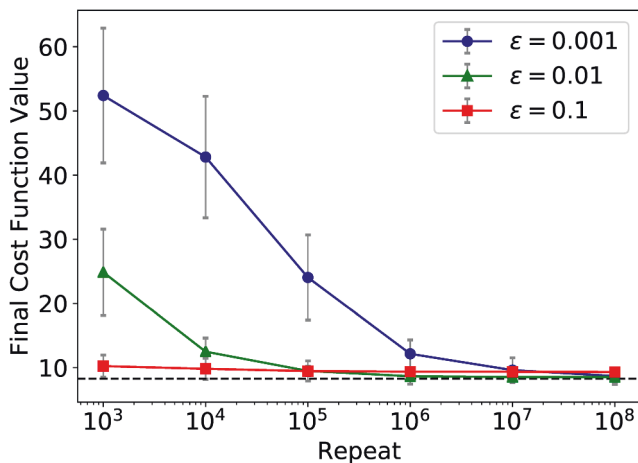
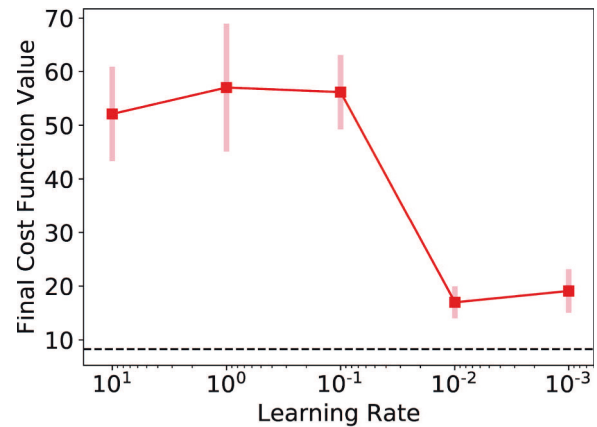
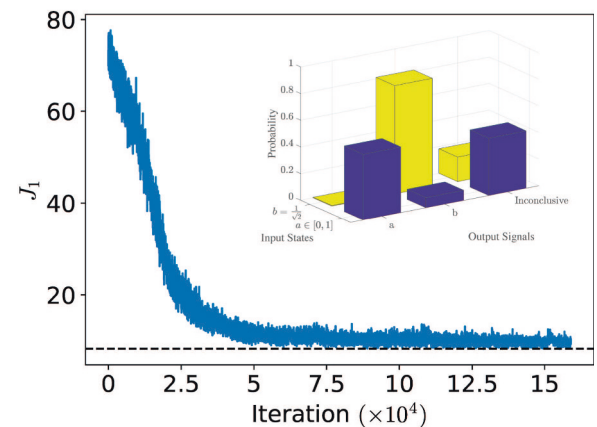


Fig. 4 The cost function after 5000 iterations. The result obtained using exact probabilities is shown by the horizontal dashed line. For a smaller step size (ϵ) for gradient calculations, we find that more repetitions are required to give a consistent result. However, a combination of $\epsilon = 10^{-2}$ and 10^5 repetitions gives a result which well approximates the result obtained using exact probabilities. Here *repeat* is the number of repeated measurements that are made each time to calculate the cost function. The cost function values are averaged over 50 repeated runs of the training process, and the bars indicate the standard deviations



(a) Lowering learning rate to counter the effect of insufficient sampling. Both the value and the standard deviation of cost function at the 20000th iteration are brought down by lowered learning rate. Here the number of repeated measurements is only 100, much smaller than $\frac{1}{\epsilon^4} = 10^8$.



(b) The noisy cost function J_1 estimated by 100 repeated measurements slowly moves to its optimal value when the learning rate is set as 0.001. The horizontal dashed line shows the minimal value 8.3 for the cost function. The inset illustrates that the trained circuit could discriminate the two quantum states. Although the error rate in the inset is not 0, we believe it could be achieved by further tuning the penalties, which is set as ($\alpha_{err} = 40, \alpha_{inc} = 5$) for this training.

Fig. 5 Small learning rates with a high number of iterations

of the estimated number of repeated measurements which are required to approximate the gradient. This follows the treatment of Farhi and Neven (2018) (Section 3). Since the gradients are calculated using the forward difference formula:

$$\frac{df}{dx}(x) = \frac{f(x + \epsilon) - f(x)}{\epsilon} + O(\epsilon) \tag{14}$$

The error in the calculation of f must be at most of the order of $O(\epsilon^2)$, in order to prevent dominating the total error. To achieve this ideally with a 99% probability, one requires

the number of repeated measurements to be of the order $\frac{1}{(\varepsilon^2)^2} = \frac{1}{\varepsilon^4}$.¹ For example, when $\varepsilon = 10^{-3}$, the ideal number of repetitions is given by 10^{12} .

In practice, we do not use $\frac{1}{\varepsilon^4}$ measurements, since the Adam optimization algorithm is designed with the noise of the cost function taken into account. To give an estimate of the number of repeated measurements which are required for the convergence of the optimization process, we perform two numerical experiments. We first look at the case when the number of repeated measurements was large ($\geq 10^3$) and $\varepsilon = 10^{-2}$. We find that 10^5 repeated measurements for each iteration are a robust configuration for a successful convergence. Second, we use a small number of repeated measurements but varied the learning rate and increased the maximal number of iterations for Adam. Setting $\varepsilon = 10^{-2}$ and taking only 100 repeated measurements, we observe that the optimizations were successful with a large number of iterations. In both experiments, the penalties are set to $\alpha_{inc} = 5$ and $\alpha_{err} = 40$.

Large number of repetitions Our results show that for a fixed maximum number of iterations (5000) for Adam, a combination of $\varepsilon = 10^{-2}$ and 10^5 repeated measurements gives robust results, i.e., the final cost function is close to the value obtained with the exact probabilities (with an error within 3%) and is stable (with a relative standard deviation of 13%). A more detailed description of the trade-off between repeated measurements and the stability of the cost function is shown in Fig. 4.

Small learning rates and high number of iterations Our numerical experiments further show that in the case of using a small number of repeated measurements, lowered learning rates could effectively counter the noisy brought by the insufficient sampling. Although in this case, the optimization requires a large number of iterations to finish. For example, with only 100 repeated measurements, the variance of cost function J_1 after 20,000 iterations decreases as we lower the learning rate (Fig. 5a). We could visually observe the optimization process where the cost function J_1 slowly approach the optimal value in Fig 5b. Here, the gradient step is taken as $\varepsilon = 10^{-2}$.

References

- Amin MH, Andriyash E, Rolfe J, Kulchytskyy B, Melko R (2018) Quantum Boltzmann machine. *Physical Review X* 8(2):021050. <https://doi.org/10.1103/physrevx.8.021050>
- ¹This assumes that the cost function follows a normal distribution with variance of the order $\frac{1}{\sqrt{N}}$, where N is the number of measurements made in each run in order to calculate the cost function.
- Banchi L, Pancotti N, Bose S (2016) Quantum gate learning in qubit networks: Toffoli gate without time-dependent control. *Npj Quantum Inf* 2:16019
- Barnett SM, Croke S (2009) Quantum state discrimination. *Adv Opt Photonics* 1(2):238. <https://www.osapublishing.org/aop/abstract.cfm?uri=aop-1-2-238>
- Barzanjeh S, Guha S, Weedbrook C, Vitali D, Shapiro JH, Pirandola S (2015) Microwave quantum illumination. *Phys Rev Lett* 114(8):080503. <https://doi.org/10.1103/physrevlett.114.080503>
- Bennett CH (1992a) Quantum cryptography using any two nonorthogonal states. *Phys Rev Lett* 68(21):3121
- Bennett CH (1992b) Quantum cryptography using any two nonorthogonal states. *Phys Rev Lett* 68(21):3121–3124. <https://doi.org/10.1103/physrevlett.68.3121>
- Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S (2017) Quantum machine learning. *Nature* 549(7671):195–202. <https://doi.org/10.1038/nature23474>
- Chefles A (2000) Quantum state discrimination. *Contemp Phys* 41(6):401–424
- Ciliberto C, Herbster M, Ialongo AD, Pontil M, Rocchetto A, Severini S, Wossnig L (2018) Quantum machine learning: a classical perspective. *Proc R Soc A* 474(2209):20170551
- Cong I, Choi S, Lukin MD (2019) Quantum convolutional neural networks. *Nat Phys* 15(12):1273–1278. <https://doi.org/10.1038/s41567-019-0648-8>
- Degen C, Reinhard F, Cappellaro P (2017) Quantum sensing. *Rev Mod Phys* 89(3):035002. <https://doi.org/10.1103/revmodphys.89.035002>
- Duan LM, Guo GC (1998) Probabilistic cloning and identification of linearly independent quantum states. *Phys Rev Lett* 80(22):4999
- Fanizza M, Mari A, Giovannetti V (2018) Optimal universal learning machines for quantum state discrimination. arXiv:180503477
- Farhi E, Neven H (2018) Classification with quantum neural networks on near term processors. arXiv:180206002
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: *Advances in neural information processing systems*, pp 2672–2680
- Grant E, Benedetti M, Cao S, Hallam A, Lockhart J, Stojevic V, Green AG, Severini S (2018) Hierarchical quantum classifiers. arXiv:180403680
- Innocenti L, Banchi L, Ferraro A, Bose S, Paternostro M (2018) Supervised learning of time-independent Hamiltonians for gate design. arXiv:180307119
- Iten R, Colbeck R, Christandl M (2016) Quantum circuits for quantum channels. *Phys Rev A Atom Mol Opt Phys* 93(3):052316. <https://doi.org/10.1103/PhysRevA.93.052316>, arXiv:1609.08103
- Iten R, Colbeck R, Kukuljan I, Home J, Christandl M (2015) Quantum circuits for isometries. *Physical Review A - Atomic, Molecular, and Optical Physics*. <https://doi.org/10.1103/PhysRevA.93.032318>. arXiv:1501.06911
- Karzig T, Knapp C, Lutchyn RM, Bonderson P, Hastings MB, Nayak C, Alicea J, Flensberg K, Plugge S, Oreg Y, Marcus CM, Freedman MH (2017) Scalable designs for quasiparticle-poisoning-protected topological quantum computation with Majorana zero modes. *Phys Rev B* 95(23):235305. <https://doi.org/10.1103/physrevb.95.235305>
- Khatri S, LaRose R, Poremba A, Cincio L, Sornborger AT, Coles PJ (2019) Quantum-assisted quantum compiling. *Quantum* 3:140. <https://doi.org/10.22331/q-2019-05-13-140>
- Kimble HJ (2008) The quantum Internet. *Nature* 453(7198):1023–1030. <https://doi.org/10.1038/nature07127>
- Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv:1412.6980

- Kübler JM, Arrasmith A, Cincio L, Coles PJ (2019) An adaptive optimizer for measurement-frugal variational algorithms. arXiv:1909.09083
- Li Y, Benjamin SC (2017) Efficient variational quantum simulator incorporating active error minimization. *Phys Rev X* 7(2):021050
- Lloyd S, Weedbrook C (2018) Quantum generative adversarial learning. *Phys Rev Lett* 121(4):040502. <https://doi.org/10.1103/physrevlett.121.040502>
- Mitarai K, Negoro M, Kitagawa M, Fujii K (2018) Quantum circuit learning. arXiv:180300745
- Mohseni M, Steinberg AM, Bergou JA (2004) Optical realization of optimal unambiguous discrimination for pure and mixed quantum states. *Phys Rev Lett* 93(20):200403. <https://doi.org/10.1103/PhysRevLett.93.200403>, 0401002
- Ostaszewski M, Grant E, Benedetti M (2019) Quantum circuit structure learning. arXiv:1905.09692
- Qi XL, Zhang SC (2011) Topological insulators and superconductors. *Rev Mod Phys* 83(4):1057–1110. <https://doi.org/10.1103/revmodphys.83.1057>
- Raynal P, Lütkenhaus N, van Enk SJ (2003) Reduction theorems for optimal unambiguous state discrimination of density matrices. *Phys Rev A* 68:022308. <https://doi.org/10.1103/PhysRevA.68.022308>. arXiv:0304179
- Ren JG, Xu P, Yong HL, Zhang L, Liao SK, Yin J, Liu WY, Cai WQ, Yang M, Li L, Yang KX, Han X, Yao YQ, Li J, Wu HY, Wan S, Liu L, Liu DQ, Kuang YW, He ZP, Shang P, Guo C, Zheng RH, Tian K, Zhu ZC, Liu NL, Lu CY, Shu R, Chen YA, Peng CZ, Wang JY, Pan JW (2017) Ground-to-satellite quantum teleportation. *Nature* 549(7670):70–73. <https://doi.org/10.1038/nature23675>
- Rocchetto A, Grant E, Strelchuk S, Carleo G, Severini S (2018) Learning hard quantum distributions with variational autoencoders. *npj Quantum Information* 4(1), <https://doi.org/10.1038/s41534-018-0077-z>, arXiv:1710.00725
- Romero J, Olson JP, Aspuru-Guzik A (2017) Quantum autoencoders for efficient compression of quantum data. *Quantum Sci Technol* 2(4):045001
- Schaller G, Schützhold R (2006) Quantum algorithm for optical-template recognition with noise filtering. *Phys Rev A* 74(1):012303. <https://doi.org/10.1103/physreva.74.012303>
- Schuld M, Bocharov A, Svore K, Wiebe N (2018) Circuit-centric quantum classifiers. arXiv:180400633
- Shende VV, Bullock SS, Markov IL (2006) Synthesis of quantum logic circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* p 18, <https://doi.org/10.1109/TCAD.2005.855930>, arXiv:0406176
- Shende VV, Markov IL, Bullock SS (2004) Smaller two-qubit circuits for quantum communication and computation. In: *Proceedings - Design automation and test in Europe conference and exhibition*, vol 2, pp 980–985. <https://doi.org/10.1109/DATE.2004.1269020>
- Sønderby CK, Raiko T, Maaløe L, Sønderby SK, Winther O (2016) Ladder variational autoencoders. arXiv:1602.02282
- Verdon G, Broughton M, Biamonte J (2017) A quantum algorithm to train neural networks using low-depth circuits. arXiv:171205304
- Wan KH, Dahlsten O, Kristjánsson H, Gardner R, Kim M (2017) Quantum generalisation of feedforward neural networks. *npj Quantum Inf* 3(1):36
- Xu X, Sun J, Endo S, Li Y, Benjamin SC, Yuan X (2019) Variational algorithms for linear algebra. arXiv:1909.03898

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