







Quantum state discrimination using noisy quantum neural networks

Andrew Patterson ^{1,*} Hongxiang Chen ^{2,3} Leonard Wossnig ^{2,3} Simone Severini ^{3,4} Dan Browne ¹ and Ivan Rungger ^{5,1,†}

¹Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, England, United Kingdom

²Rahko, Ltd., Finsbury Park, London N4 3JP, England, United Kingdom

³Department of Computer Science, University College London, Gower Street, London WC1E 6BT, England, United Kingdom

⁴AWS Center for Quantum Computing, Pasadena, California 91125, USA

⁵National Physical Laboratory, Teddington, London TW11 0LW, England, United Kingdom



(Received 16 June 2020; accepted 22 December 2020; published 20 January 2021)

Near-term quantum computers are noisy, and therefore must run algorithms with a low circuit depth and qubit count. Here we investigate how noise affects a quantum neural network (QNN) for state discrimination, which is applicable on near-term quantum devices as it fulfils the above criteria. We find that for the required gradient calculation on a noisy device a quantum circuit with a large number of parameters is disadvantageous. By introducing a smaller circuit ansatz we overcome this limitation, and find that the QNN performs well at noise levels of current quantum hardware. We present a model showing that the main effect of the noise is to increase the overlap between the states as circuit gates are applied, hence making discrimination more difficult. Our findings demonstrate that noisy quantum computers can be used for state discrimination and other applications, such as classifiers of the output of quantum generative adversarial networks.

DOI: [10.1103/PhysRevResearch.3.013063](https://doi.org/10.1103/PhysRevResearch.3.013063)

I. INTRODUCTION

Quantum state discrimination is important in many emerging quantum technologies: quantum cryptography [1], entanglement concentration [2], quantum cloning [3], and quantum metrology and sensing [4,5]. Quantum circuits trained for classification could also be used in quantum machine learning problems as a classifier of quantum data. They could classify the output of other quantum circuits, e.g., the output of a quantum generative adversarial network [6]. Current quantum computing devices are subject to non-negligible amounts of noise [7–9], and therefore algorithm design for devices in the near future must take this into account. Here we present an extension to noisy devices of the approach for quantum state discrimination outlined in [10], a quantum analog of a neural network used for state discrimination. In [10] simulations of shallow quantum circuits were trained to find the optimal positive operator-valued measure (POVM), or measurement, to distinguish between two families of nonorthogonal quantum states. Given an input state chosen randomly from one of the families, the output of the network should indicate which family the input was chosen from. To do this the network is trained on a set of labeled data, performing supervised learning [11]. The ideal POVM was learned via a classical optimizer using a

gradient descent algorithm on the quantum parameters, which correspond to the rotation gates in the quantum circuit. This POVM is distinct from the error minimizing POVMs, as it also attempts to minimize the occurrence of inconclusive results.

There are similarities between a classical unitary neural network [12] and this algorithm. It contains a layer of symmetric fully connected neurons, followed by arbitrary numbers of nonlinear layers, or dropout layers. The nonlinearity in the quantum network is introduced by measurement of some of the qubits.

In this paper we extend the simulations done previously from pure vector states to simulations of states represented as density matrices, so that we can model noise in the quantum device. We also simulate calculation of the parameter gradients on the quantum device, which would also be subject to noise in a real machine. We find that with these extensions including the effect of noise the previous algorithm proposed for noiseless systems no longer performs optimally. To recover performance we reduce the number of trainable parameters through consideration of the circuit structure.

This paper is structured as follows: we begin by outlining the theory of state discrimination and the QNN. We then discuss the simulation methods, gradient calculation, and measurement. In Sec. III we present the results, including the effect of reducing the number of parameters and the effects of noise on training the circuits.

II. METHODS

A. Quantum state discrimination

We wish to discriminate a two-qubit input state, $|\psi_{in}\rangle$, which in general can be represented as a normalized vector

*andrew.patterson.10@ucl.ac.uk

†ivan.rungger@npl.co.uk

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

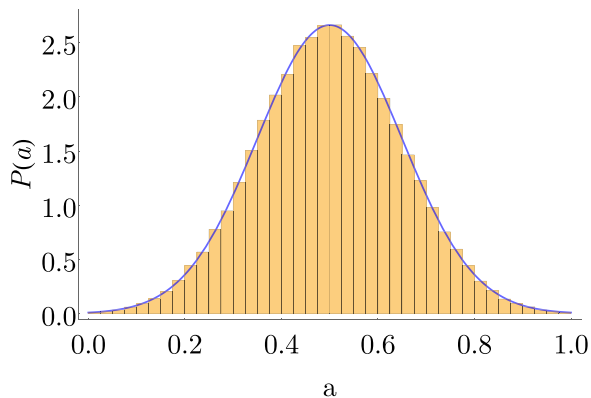


FIG. 1. Distribution $P(a)$ of the parameter a for the input states in the first set [Eq. (1)], for an average value $\mu_a = 0.5$ and a standard deviation $\sigma_a = 0.15$. These values of μ_a and σ_a are used for the data in Figs. 4–8 and 11. Due to the rather large value of σ_a the most probable values of a are not confined only in a thin range around μ_a , but instead cover almost all the range between 0 and 1.

with four complex components. The state is chosen randomly from two sets of states, labeled a and b , and elements defined by

$$|\psi_{in,a}\rangle = (\sqrt{1 - a^2}, 0, a, 0), \tag{1}$$

$$|\psi_{in,b}\rangle = \left(0, \pm \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right), \tag{2}$$

where Eq. (1) defines the states in the first set, $|\psi_{in,a}\rangle$, where by setting the continuous real parameter $a \in (0, 1]$ the specific state is chosen. The number of states within this first set is therefore infinite due to the fact that a can be varied continuously. Equation (2) defines the second set, $|\psi_{in,b}\rangle$, which has two elements, given by $|\psi_{in,b,+}\rangle = (0, 1/\sqrt{2}, 1/\sqrt{2}, 0)$ and $|\psi_{in,b,-}\rangle = (0, -1/\sqrt{2}, 1/\sqrt{2}, 0)$. The discrimination in our paper is performed between these two sets of states: one with an infinitely large continuous set of states, and one set with two states in it. State discrimination between general sets of states is discussed in [13]. We set the probability that a state from the a set appears as input state to $p_a = 1/3$, the probability that a b state with positive sign appears to $p_{b+} = 1/3$, and the probability that a b state with a negative sign appears to $p_{b-} = 1/3$. Note however that our results are expected to be applicable also to other sets of states, since the methods presented here are based on variational algorithms, which can be formulated for any target state.

To fully define the first set one further needs to define the probability distribution, $P(a)$, that describes how the parameter a of the first set of input states is statistically distributed in the range from 0 to 1. We use a Gaussian distribution with an average value, μ_a , and a standard deviation around this average value, σ_a . As illustrative example in Fig. 1 we show $P(a)$ for $\mu_a = 0.5$ and $\sigma_a = 0.15$. The set of states defined by this distribution has about 68% of the states in the range $a = 0.35$ to 0.65 , 25% of the states within the range $a = 0.1$ to 0.35 and $a = 0.65$ to 0.9 , and still a non-negligible probability of about 5% of finding a state in the range $a = 0$ to 0.1 and $a = 0.9$ to 1.0 .

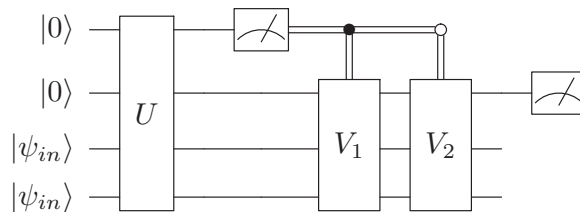


FIG. 2. The general form of the quantum circuits used in this paper. The input state is on the bottom two qubits, and measuring the first qubit introduces a nonlinear dropout layer. The subcircuits U and $V_{1,2}$ are shown in Fig. 3.

Primarily our specific sets of states a and b are chosen to reproduce the work in [10] and in [14], where the state discrimination was performed in a laboratory. Secondly, these states are nonorthogonal and therefore cannot be distinguished perfectly without some probability of erroneous or inconclusive outcomes, making the problem harder for the algorithm. It is therefore an ideal case to verify the method, since the level of nonorthogonality can be tuned by choosing μ_a and σ_a in the distribution function of a in Eq. (1), with a value of a closer to 1 being more difficult to discriminate.

B. The quantum neural network

Classical artificial neural networks (ANNs) consist of a set of artificial neurons, also called nodes, connected via links [11,15]. Typically the nodes are arranged in layers, where the output of all nodes in one layer is passed via the links to the inputs of all the nodes in the next layer. The links connecting the nodes have an adjustable weight associated to them, which determines the strength of the influence of the output of one node on the node in the next layer. Within each node all the inputs with their given weights are added up, and a so-called activation function is performed on this sum, which determines the output strength of the node. Importantly, the activation function is usually nonlinear. This nonlinearity gives the ANN the capability to perform complex tasks when the weights of the links are adjusted according to the task at hand. Such ANNs have been shown to perform extremely well on a variety of tasks such as recognizing objects in images [16] to generating novel writing when provided with a prompt [17].

In a classical neural network some nodes are discarded during training, which is called dropout and stops the network from overfitting [18]. We can also think of dropout as the introduction of nonlinearity into the network. A network with dropout cannot be represented by any smaller, linear network, whereas a many-layered linear network can always be reduced to a single linear layer. Quantum evolution is unitary and linear, so if we wish to introduce nonlinearity into a quantum neural network we need to include a measurement. Figure 2 shows the structure of the QNN, where the choice of the second step of the circuit, $V_{1,2}$, is conditioned on the outcome of a measurement on the first qubit. The measurement results are then used as the output of the neural network [10].

There are two nonorthogonal states to discriminate, so if we wish to have a network that can be trained to not commit

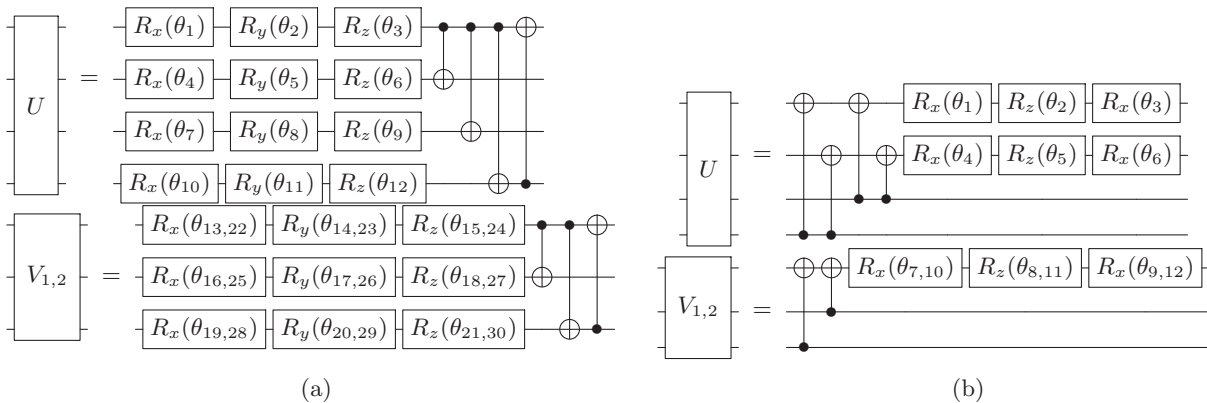


FIG. 3. The circuits showing the trainable parameters, which are used in this paper. Comparison of results obtained for the circuits in Figs. 3(b) and 3(a) is made in Sec. III A. (a) The U and V circuit blocks originally used in [10]. (b) The form of the U and V blocks with a reduced number of parameters.

any errors, we must allow for it to produce an inconclusive result [14]. This allows the network to give a “don’t know” result as opposed to an erroneous one. Therefore we have a minimum of three outputs, necessitating two measurement qubits.

The output of the network is determined by the measurement outcome. As we begin in a random configuration and are training the system, we can arbitrarily select which label a measurement outcome corresponds to:

$$\{|00\rangle : a, |01\rangle : b, |10\rangle : a, |11\rangle : \text{inconclusive}\}. \quad (3)$$

The choice of unbalanced labels may have an effect upon the outcome. For a random measurement outcome the probability to guess the right state is $1/2$ for a and $1/4$ for b . We partly mitigate this bias setting the probability of a and b states to appear as input to the values specified in the previous subsection, namely, $p_a = 1/3$ and $p_b = p_{b+} + p_{b-} = 2/3$.

This results in the probability of correctly guessing the input state for a fully random measurement outcome to be $1/2p_a + 1/4p_b = 1/3$, and correspondingly the probability for an incorrect guess is equal to $2/3$. In general one might adapt the assignment of measurement outcomes to labels according to the considered specific task.

The structure of the U and $V_{1,2}$ circuit blocks is given in Fig. 3, where Fig. 3(a) shows the same circuits used in [10] and Fig. 3(b) shows the reduced circuits introduced here, which we will discuss in more detail below. These circuits are small and have low depth, so that they can be ran on a quantum computer which supports measurement as the circuit is running and classical feedback. This requires fast measurement and fast classical processing which is not possible in many current systems, but has been achieved in an ion-trap device [19], meaning this algorithm could run on a current device.

The state discrimination task is then as follows: input states are drawn from two sets of states, and the classical optimizer must optimize the rotation angles $\theta_{1..n}$ of the quantum circuit to maximize the likelihood of a correct determination of the state. In our specific case it has to determine whether an input state is from the a set of states or the b set of states. Note that only these states are allowed as input states during both training and testing of the circuit. A correct determination is found when the measurement output of the quantum circuit

is equal to the corresponding input state label as defined in Eq. (3).

This task is the minimum error discrimination problem, which has been solved classically using semidefinite programming methods [20–22]. Here we solve the same problem using a noisy quantum circuit, and demonstrate the usefulness of using quantum measurement as a form of dropout. Note that each step of the semidefinite programming method requires $O(m^4)$ classical steps, where m is the length of the input vector, which scales exponentially with the number of qubits required to represent the states. The scaling of variational quantum circuits such as the ones used in our approach is not known currently, especially as the number of trainable parameters increases, and is a topic of ongoing research [23–26].

C. Optimization

Since the input states are initially labeled, the task for the classical optimizer is a supervised learning task [11]. The optimizer used in this experiment is ADAM [27], which has been found to work well in a number of quantum variational algorithms [10,28–31]. It has also been shown classically that ADAM deals well with noisy gradients [32], which will be the output of our noisy quantum computer. This is possible since ADAM uses the concept of momentum, where the gradients of past steps contribute to the current step. Other optimizers such as ROTOSOLVE [33] have been proposed, and a comparison of performance can be made in future work.

Noisy gradients are a feature of the work here: as gradient calculation must be performed on the noisy quantum device, we expect that the output gradients will be noisy. We also expect that there will be nonoptimal local minima in our loss landscape, as this is also a feature of the loss function in the noiseless case [10]. Finally we also expect that the loss landscape may feature “barren plateaus,” as these have been shown to be a feature of quantum optimization problems [34]. This further motivates the choice of a gradient-based optimizer such as ADAM.

We define the function to minimize, the cost function, as

$$C = \alpha_{\text{err}}P_{\text{err}} + \alpha_{\text{inc}}P_{\text{inc}}, \quad (4)$$

where the positive real numbers α_{err} and α_{inc} are the cost parameters used to bias the network towards minimizing errors or inconclusive results (P_{err} and P_{inc} are defined below). If for example we require the network to produce fewer errors, we can do this at the cost of recording more inconclusive results by increasing the value of α_{err} relative to the value of α_{inc} . Such intermediate optimization strategies based on changing the cost parameters, thereby allowing for a tradeoff between minimum-error and unambiguous state discrimination, were studied in [35–38]. Here we use a quantum algorithm to solve this problem, and perform the analysis of how noise in a quantum device affects the results. We discuss the effect of changing the cost parameters in Sec. III A.

While the input states are always pure states, as they are processed in our quantum circuit the noise causes them to become mixed states. To simulate this effect of the noise in the quantum circuit we use density operators, ρ , to represent quantum states inside the quantum circuit. The measurement probabilities of a state, ρ , for a generalized measurement, $M = |\phi\rangle\langle\phi|$, are given by

$$\langle\rho\rangle = \text{Tr}(|\phi\rangle\langle\phi|\rho), \quad (5)$$

and the quantum state after measurement is given by

$$\rho_{\text{measured}} = \frac{|\phi\rangle\langle\phi|\rho|\phi\rangle\langle\phi|}{\text{Tr}(|\phi\rangle\langle\phi|\rho)}. \quad (6)$$

Using this we can find the probability of an erroneous or inconclusive measurement:

$$P_{\text{err}} = \sum_{\rho_i \in b} (\langle\rho_i\rangle_{00} + \langle\rho_i\rangle_{10}) + \sum_{\rho_i \in a} \langle\rho_i\rangle_{01}, \quad (7)$$

$$P_{\text{inc}} = \sum_{\rho_i \in a, b} \langle\rho_i\rangle_{11}, \quad (8)$$

where ρ_i is the input state, and $\langle\rho_i\rangle_{jk}$ refers to the probability of obtaining a measurement of $|jk\rangle$ from the circuit.

Discrimination of these states, without the use of a variational algorithm, has been shown in the laboratory to reach the theoretical best success probability, P_{suc} of 0.833 for $\mu_a = 0.25$, $\sigma_a = 0.01$ [14]. This is a minimum loss, $L = 1 - P_{\text{suc}} = P_{\text{err}} + P_{\text{inc}}$, of 0.166. For the equal probability case, $P(|00\rangle) = P(|01\rangle) = P(|10\rangle) = P(|11\rangle) = 0.25$, the success rate is 0.385, which translates into a loss of 0.635. This gives us lower and upper expected bounds to compare our results for the loss to.

D. Gradient calculation

Unlike gradient-free optimizers (such as Nelder-Mead [39]) the ADAM optimizer requires the calculation of parameter derivatives ($\frac{\partial\langle C\rangle}{\partial\theta_i}$). In the previous work this was done using the forward differences formula [10], which requires direct access to the components of the wave function. In a real quantum computer this is difficult to achieve, and hence here we use a more practical approach. Calculation of the gradients of quantum parameters has received attention recently [40–42] due to the introduction of variational methods such as the variational quantum eigensolver [43]. The gradient of the loss function with respect to a parameter θ_i is calculated by the method outlined in [40], which requires two extra repetitions

of the circuit for each θ_i :

$$\frac{\partial\langle C\rangle}{\partial\theta_i} = \frac{1}{2}(\langle C\rangle^+ - \langle C\rangle^-), \quad (9)$$

where $\langle C\rangle^\pm$ is calculated by changing θ_i by $\pm\frac{\pi}{2}$, and leaving all other parameters in the circuit constant.

E. Reduced circuit

For the probability distribution of a determined by $\mu_a = 0.25$ and $\sigma_a = 0.01$ the maximum theoretical success rate (P_{suc}) is 0.8333 [14], which was obtained with the long circuit in [10]. However, after optimization of circuit parameters for our larger circuit in Fig. 3(a) we reach only 0.72, which is significantly smaller than the theoretical limit. We attribute this discrepancy to the different implementations of the optimization procedure, and to the different calculation of the gradients. To overcome this suboptimal result we designed the shorter circuits in Fig. 3(b). The choice of the reduced circuit is motivated by the consideration that for this task the rotations on the state qubits have a smaller effect on the measurement outcomes than rotations on the measurement qubits. This choice of structure is so that the input states are entangled with both output qubits, and then the measurement qubits are rotated. The choice of rotations about the x axis, followed by the z axis, and then again the x axis allows for the initial state to be transformed to any other state on the surface of the Bloch sphere [44]. With this short circuit [Fig. 3(b)] we obtain a success rate of 0.826, close to optimal performance. This is the circuit used for the results presented, except where we explicitly note that the longer circuit is used.

We note that as the shorter circuits do not explore the full Hilbert space of all the qubits, they may not be necessarily optimal for all discrimination tasks. Investigations into the capability of different variational quantum circuits have been made in [45]. Here we present evidence that when used on a noisy device, the smaller variational circuit converges to better results than the larger circuit. In general a tradeoff needs to be made between this better resilience to noise and the ability of the circuit to distinguish very complex states.

F. Noise

Noise in quantum computers can be modeled by a superoperator, $\mathcal{E}(\rho)$, which is a completely positive, trace-preserving map on the state ρ [44]. We can give the operator-sum representation of \mathcal{E} by introducing the Kraus operators, E_k :

$$\mathcal{E}(\rho) = \sum_k E_k \rho E_k^\dagger, \quad (10)$$

and to preserve the trace of ρ , they must obey the relation

$$\sum_k E_k^\dagger E_k = \mathbb{1}. \quad (11)$$

For the single-qubit noise channel our operators are the single-qubit Pauli operators, modified by the noise probability, p , to

give the depolarizing channel:

$$E_0 = \sqrt{1 - \frac{3p}{4}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, E_1 = \sqrt{\frac{p}{4}} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

$$E_2 = \sqrt{\frac{p}{4}} \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, E_3 = \sqrt{\frac{p}{4}} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (12)$$

For the two-qubit noise channel, which is applied after a two-qubit gate, the Kraus operators are tensor products of the combinations of these operators, i.e., $E_0 \otimes E_0, \dots, E_1 \otimes E_2, \dots, E_3 \otimes E_3$.

Since in our circuit we have four qubits, these one- and two-qubit Kraus operators have to be expanded using tensor products with identity operators. For example, if the Kraus operator i is acting on the first qubit, and the Kraus operator j is acting on the third qubit (e.g., after an entangling gate between those two qubits), then the total Kraus operator used in our simulations is $E_i \otimes \mathbb{I} \otimes E_j \otimes \mathbb{I}$. In matrix form this is of dimension 16×16 .

The probability of the single-qubit noise channel is $p_{1q} = \frac{4}{5}p_{2q}$. This is the one-qubit marginal probability of error for the two-qubit gates [46], i.e., the probability of a single-qubit error without condition of an error on the other qubit. This is a commonly used assumption in the quantum error correction literature [47], which assumes that the error process in single- and two-qubit gates is the same. In real devices the process can be quite different, but we nevertheless choose this method as it is an upper limit on the error probability of the single-qubit gate. When quoting the noise level in this paper, we will always refer to p_{2q} . We set the highest noise level in our simulations to $p_{2q} = 0.1$, as this is an upper limit on two-qubit gate fidelities reported on current quantum hardware [7–9].

Note that here we have not considered asymmetric noise or different quality qubits. However, we believe that correcting for a systematic bias such as this is possible for a variational algorithm, as seen in [48]. Furthermore, in actual devices the single-qubit noise probability reported is much lower than $4/5$ of the two-qubit gate noise level. For example, the single-qubit gate error rate reported in [9] is 1.4×10^{-3} , whereas the two-qubit gate fidelity is 9.3×10^{-3} , and the ratio between these is approximately $3/20$, at least a factor of 5 lower. In our simulations the single-qubit noise is set to the higher limit of $4/5$, so that we are more demanding of the algorithm.

G. Simulation

Simulations of the quantum device were performed on a simulator built using the Tensorflow machine learning package [49], and verified with the Cirq [50] quantum simulation package. In our simulations we set the initial angles, which are our parameters to be optimized, at random values. The labeled quantum state is an input to the circuit in Fig. 2 ; that circuit is ran and the measurement probabilities calculated and with them the cost. The gradient of the cost with respect to each parameter is then calculated by the method described in Sec. IID, and the parameters are updated according to the ADAM optimizer to minimize the cost. This routine is repeated until the cost no longer significantly decreases.

During the training process of the optimal quantum circuit rotation angles, $\{\theta_i\}$, at each iteration we evaluate the cost

function in Eq. (4) with a number of randomly chosen values of a within a given distribution $P(a)$. This number is 20 for each training step, of which there are typically 1000–3000 until convergence. Due to this rather large number of samples used in the training optimization it is ensured that the whole distribution of a according to $P(a)$ is covered. Hence the circuit is trained to discriminate between the whole continuous set $\{|\psi_{in,a}\rangle\}$, with a distributed according to $P(a)$, and the two states in set $\{|\psi_{in,b}\rangle\}$. Once the circuit parameters are trained, for the testing step we use 250 samples of a values, distributed again following $P(a)$. Therefore the QNN is trained not only for one specific given value of a , but rather for the chosen distribution of a values.

Measurements here are calculated in the “infinite-shot” regime, where the representation of the quantum state at the end of the circuit is used to extract exact measurement probabilities. The inclusion of statistical measurement noise can be expected to result in a slower rate of convergence than obtained here. We note that in [29] it was demonstrated that convergence of variational algorithms is guaranteed even for single-shot measurements of the gradient. We indicate that convergence can also be achieved using this method in presence of measurement noise, although with a higher number of iterations.

III. RESULTS

A. Effect of cost function choice and circuit depth

In Fig. 4 we compare the obtained optimized P_{err} and P_{inc} for an error minimizing cost function ($\alpha_{err} = 60, \alpha_{inc} = 10$) and a balanced cost function ($\alpha_{err} = 40, \alpha_{inc} = 40$). The error minimizing cost function often results in a practically unusable network, because while it gives a low probability of error, the probability of inconclusive results is too high, as seen for an extreme case in the inset of Fig. 4. Note that in this particular case all b states are detected as inconclusive, and one could in principle switch the *inconclusive* and b labels to obtain a good discrimination. However, for the more general case this will not be possible.

In comparison to the error minimizing setting, the results for the balanced cost function are stable and generally give both small P_{err} and P_{inc} , with some P_{err} comparable to the error minimizing setting. For the remaining analysis we therefore use the balanced cost function ($\alpha_{err} = \alpha_{incon} = 40$). We note that as the noise level is increased, P_{inc} and P_{err} progressively tend to larger values. The effect of noise will be analyzed in detail in the next section.

We next investigate the influence of the number of parameters in the quantum circuit on the loss. In Fig. 5 we compare the distributions of loss between the circuit with more trainable parameters in Fig. 3(a) to the circuit with fewer parameters in Fig. 3(b). It can be seen that the reduced circuits consistently perform better than the long circuits. It is more difficult to train circuits with a large number of parameters both without and with noise, as seen in Fig. 6. We see that the higher noise cases always converge to a higher loss, and that the reduced circuits perform better in both cases. These results show that in practice increasing the number of parameters used in a quantum circuit does not always have a

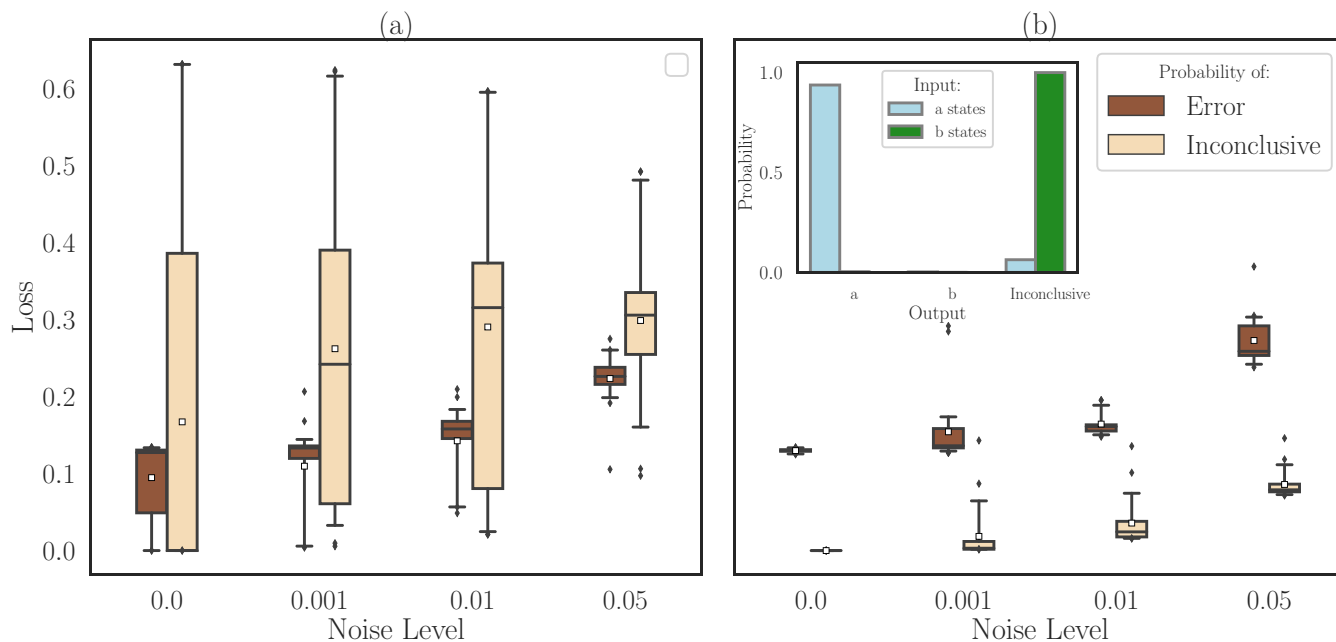


FIG. 4. The distribution of P_{inc} and P_{err} from 25 repeats of (a) a network biased towards reducing errors $\alpha_{err} = 60$ and $\alpha_{inc} = 10$ and (b) a network with a balanced cost function, $\alpha_{err} = \alpha_{inc} = 40$, both with values $\mu_a = 0.5$ and $\sigma_a = 0.15$. An example undesirable output for a single minimizing error run is in the inset, where no b states are measured correctly, but the network still converges (the x axis shows the output label and the color is the input state). The interquartile range is contained within the box, and the 5th and 95th percentiles are marked by the whiskers. Outliers of this range are marked by a diamond. The mean is marked with a white square, and the median is the line across the box.

beneficial effect. Importantly, even in the noiseless case the circuit with less parameters leads to better results. Further-

more the reduced number of parameters also significantly lowers the required run time.

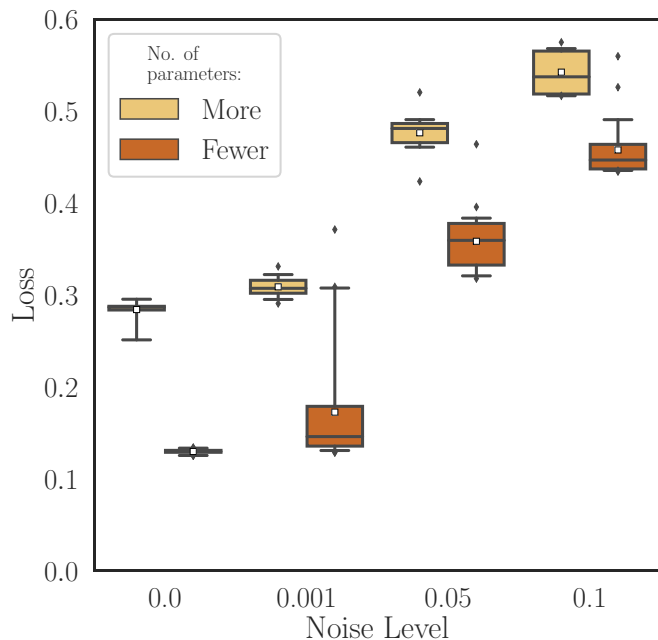


FIG. 5. The distributions of loss ($P_{err} + P_{inc}$) at different noise levels for the two circuits shown in Fig. 3. Both have other parameters fixed, $\mu_a = 0.5$, $\sigma_a = 0.15$, $\alpha_{err} = \alpha_{inc} = 40$. We observe that reducing the number of parameters is advantageous at all noise levels.

Even with very low noise, the output is worse for larger circuits. This suggests that with more parameters the algorithm struggles to optimize, when the gradient calculations are performed on the quantum device. Good performance of the short circuit in the presence of noise can be due to the noisy gradient regularizing the training, thereby optimizing performance [11]. Moreover, the ADAM optimizer has been designed to work well with noisy gradients [27]. The results seen here are indicative that a noise-resilient optimizer using gradients provided by a noisy quantum circuit can perform well.

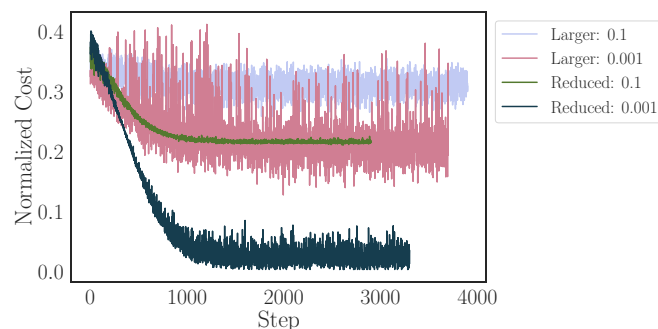


FIG. 6. Evolution of the normalized cost functions for larger and reduced circuits for $\mu_a = 0.5$ and $\sigma_a = 0.15$, with noise levels of 0.001 and 0.1. Shown here is the number of steps taken to converge. Note that the time taken to complete a single step of the longer circuit is much greater than for the reduced circuit.

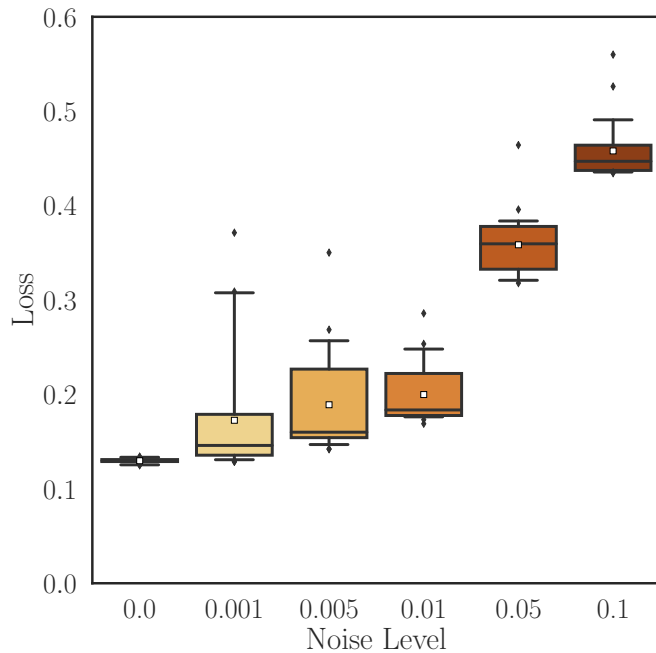


FIG. 7. The distribution of loss for 25 repeats of training the network. The cost function is balanced, $\alpha_{\text{err}} = \alpha_{\text{inc}} = 40$, $\mu_a = 0.5$, and $\sigma_a = 0.15$. At levels of noise present in current devices, 0.01, the loss value is favorable, an average of 0.2.

B. Effect of noise: Numerical analysis and model

In Fig. 7 the noiseless case is compared to resulting optimized loss for increasing noise levels (note that in this section we always use the reduced circuit). It can be seen that using this algorithm with zero noise produces the lowest loss, as one expects intuitively. With increasing noise the average loss increases continuously. In presence of noise there are a few high-loss outliers, which we attribute to the optimizer becoming stuck in local minima of the cost function. As the noise is increased, performance deteriorates, but is no worse than the random output limit of $2/3 \approx 0.67$ (see Sec. II B).

Importantly, at noise levels comparable to current devices, $p_{2q} = 0.01$, the algorithm is still performing well, at an average loss of 0.2.

In general a high level of noise always leads to a higher loss. However, we find that when noise is applied only during the training of the parameters, the optimized parameters are rather resilient to this training noise. To show this in Fig. 8 we present the results when training the device at one noise level, and validating at another. We see that even with high levels of training noise the optimizer converges onto good parameters, as we find comparably low loss levels when validating those parameters trained at a high noise level with low noise in the validation step. Also here we find that when validating at noise levels seen in current devices, $p_{2q} = 0.01$, the average loss does not increase above 0.25, which would be acceptable to use for state discrimination.

In order to provide an understanding of the numerically found changes of the loss with noise, in what follows we present a simple model that can describe the results. It is based on the notion that a larger overlap between the states to be dis-

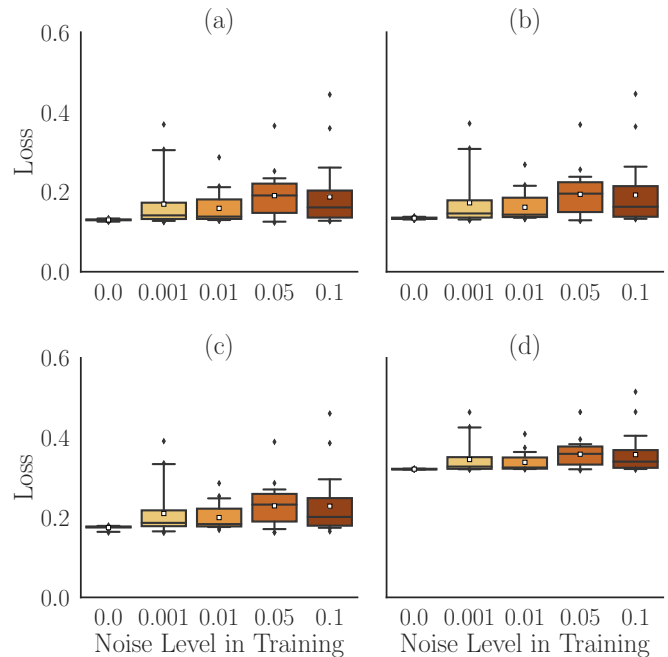


FIG. 8. Distribution of loss ($P_{\text{err}} + P_{\text{inc}}$) against training noise for different noise levels in the validation circuit: (a) 0.0, (b) 0.001, (c) 0.01, and (d) 0.05.

criminated generally makes discrimination more difficult. As outlined in Sec. II A, with our chosen set of states this overlap can be tuned by setting the value of a , and is equal to $a/\sqrt{2}$. We can therefore systematically evaluate the effect of noise on the discrimination for increasing overlap by increasing μ_a , and the results are shown in Fig. 9. The loss increases for larger μ_a for all levels of noise. At high noise levels and high μ_a , some runs are performing even worse than the random output limit (0.67), but on average the loss remains well below that value. In general we conclude that the tolerable levels of noise depend on the overlap between the states, where small overlap allows the states to be discriminated even for higher noise in the quantum computer.

For large noise in the system the difference in loss between higher and lower values of μ_a is significantly reduced when compared to the low noise case. This seems to indicate that the noise on average reduces the difference between states as these pass the circuit, and hence effectively increases the averaged overlap. This effect can be illustrated for the ideal case of $a = 0$ and no noise, where discrimination can in principle be perfect since the states are orthogonal. However, in presence of noise there is a probability that a state is perturbed as the circuit is applied to it, and hence orthogonality between states is lost. This results in a certain probability of erroneous detection.

In order to estimate this effect on a semiquantitative level for our used circuit, shown in Figs. 2 and 3(b), we note that in absence of noise the role of the data qubits is only to store the state $|\psi_{\text{in}}\rangle$, which then controls the state of the measurement qubits. In presence of noise the three noisy two-qubit gates applied to each of the two data qubits will perturb $|\psi_{\text{in}}\rangle$ during the processing of the circuit, which in turn will affect the measurement qubits via the control operation and hence the

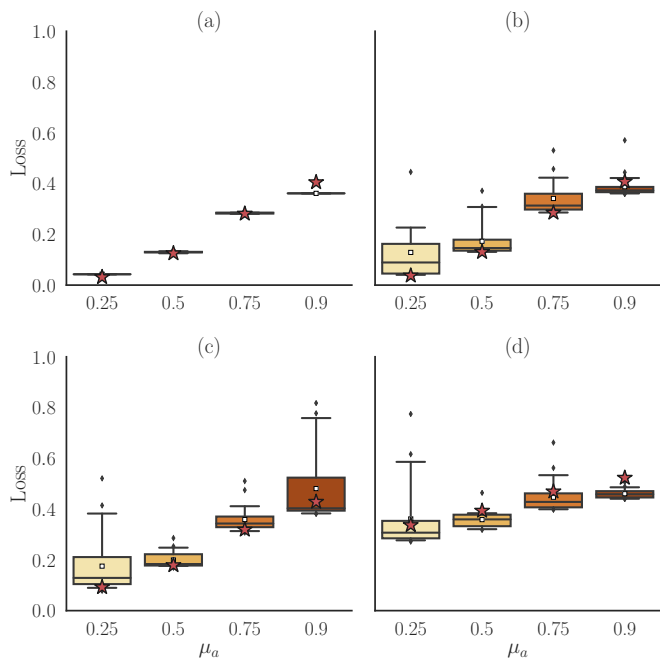


FIG. 9. The distribution of loss ($P_{\text{err}} + P_{\text{inc}}$) and the effect of different values of μ_a . The cost function is balanced, $\alpha_{\text{err}} = \alpha_{\text{inc}} = 40$, and $\sigma_a = 0.15$. The noise level is (a) 0.0, (b) 0.001, (c) 0.01, and (d) 0.05. We see that for lower values of μ_a , corresponding to smaller overlap between the states to be discriminated, the discrimination task is performed better. The red stars indicate the fidelity $F_{\tilde{a}\tilde{b}}$ between the two states after three applications of Kraus operators to each of the data qubits, as given by Eq. (14d).

outcome of the state discrimination. We therefore approximately model the effect of noise on the state discrimination by determining how much the repeated application of noisy channels at each gate affects a given $|\psi_{\text{in}}\rangle$, without considering the presence of the measurement qubits.

In what follows we quantify how the application of noise channels affects each state and its overlap with the state to be discriminated from. The quantum states with noise applied are represented by density matrices, so that the overlap between two states represented by the density matrices ρ and σ , respectively, is described by the fidelity, F , given by [51]:

$$F = \text{Tr}[\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}]^2. \tag{13}$$

Of particular interest here is the fidelity between a pure state $|\psi_a\rangle$ entering the circuit and its modified form due to the application of noise after three two-qubit gates. We denote this as $F_{\tilde{a}\tilde{a}}$, where the tilde on the second subscript indicates that the second state is the one where noise was applied. We use an analogous notation for the other relevant quantities, which are $F_{\tilde{b}\tilde{b}}$, $F_{\tilde{b}\tilde{a}}$, and $F_{\tilde{a}\tilde{b}}$. We can compute these quantities numerically, but given the rather cumbersome form of Eq. (13) it is difficult to relate the results to the fundamental parameters of the noisy discrimination process. We therefore provide a lowest-order expansion of these terms in p , which we expect to be close to the exact results since we are only dealing with

small p . We apply the noise model in Eqs. (10)–(12), to the states a and b in Eqs. (1) and (2), and obtain to first order in p

$$F_{\tilde{a}\tilde{a}} = 1 - n p, \tag{14a}$$

$$F_{\tilde{b}\tilde{b}} = 1 - \frac{3n p}{2}, \tag{14b}$$

$$F_{\tilde{a}\tilde{b}} = F_{\tilde{b}\tilde{a}} = \frac{\mu_a^2}{2} + \frac{n p}{2} (1 - 2\mu_a^2), \tag{14c}$$

$$F_{\tilde{a}\tilde{b}} = \frac{\mu_a^2}{2} + n p \left(1 + \frac{\mu_a}{\sqrt{2}} - 2\mu_a^2 + \sqrt{1 - \frac{\mu_a^2}{2}} \right), \tag{14d}$$

where $n = 0 \dots 3$ is the number of noisy channel applications, μ_a is the mean value of a , and p is the noise probability. Within this expansion order the fidelities are linear in both p and n , and the expansion coefficients are simple functions of μ_a . For the noiseless case ($p = 0$), $F_{\tilde{a}\tilde{b}} = F_{\tilde{a}\tilde{b}} = \mu_a^2/2$, which corresponds to the absolute value squared of the overlap between the a and b states.

The results for the fidelities for increasing n are shown in Fig. 10. It can be seen that the numerical results obtained directly with Eq. (13) are captured rather well with the analytical low-order expansion in Eqs. (14a)–(14d). The value of $F_{\tilde{a}\tilde{a}}$ decreases with each application of a noisy channel, showing that the purity of the state degrades as the noise channels act on the state. The value of $F_{\tilde{a}\tilde{b}}$ increases with n for lower values of μ_a , where the noise acts to increase the fidelity between the states, while it decreases with n for higher values of μ_a , where the noise reduces the orthogonal component in $|\psi_a\rangle$. In Fig. 10(c) the fidelity is plotted when the noise channel is applied to both states ($F_{\tilde{a}\tilde{b}}$), and it always increases with n . $F_{\tilde{a}\tilde{b}}$ is the relevant quantity of the influence of noise on the loss: the application of the noisy circuit on the originally pure input states causes them to degrade into mixed states. The states to be discriminated are therefore not the input pure states anymore, but these noisy states, and for each application of the noise channel they become harder to discriminate.

The minimal loss achievable in quantum state discrimination is generally a function of the fidelity between the two input states [52,53]. The exact relation depends on the cost function that is minimized, and is only known analytically for a few special cases, such as minimal error discrimination or unambiguous discrimination [52,53]. In this section we consider the case where the rate of inconclusive results and erroneous results is minimized simultaneously ($\alpha_{\text{err}} = \alpha_{\text{inc}} = 40$).

Before estimating the effect of noise we therefore need to determine the functional relation between the fidelity and the loss in the noiseless case for our circuit. In the top left panel of Fig. 9 we show the results for $F_{\tilde{a}\tilde{b}}(p = 0) = F_{ab}$ as red stars for each μ_a . One can see that the F_{ab} is approximately equal to the loss for all μ_a , so that to a good approximation for our circuit we can fit the relation as $\text{Loss} = F_{ab}$. In general the lower bound of the optimal theoretical loss is found in the minimum error discrimination setting, where there is no inconclusive measurement. In our optimization we include also the inconclusive measurement, the probability of which is minimized in the unambiguous setting, which gives an upper bound on the loss. Our minimized cost function is a

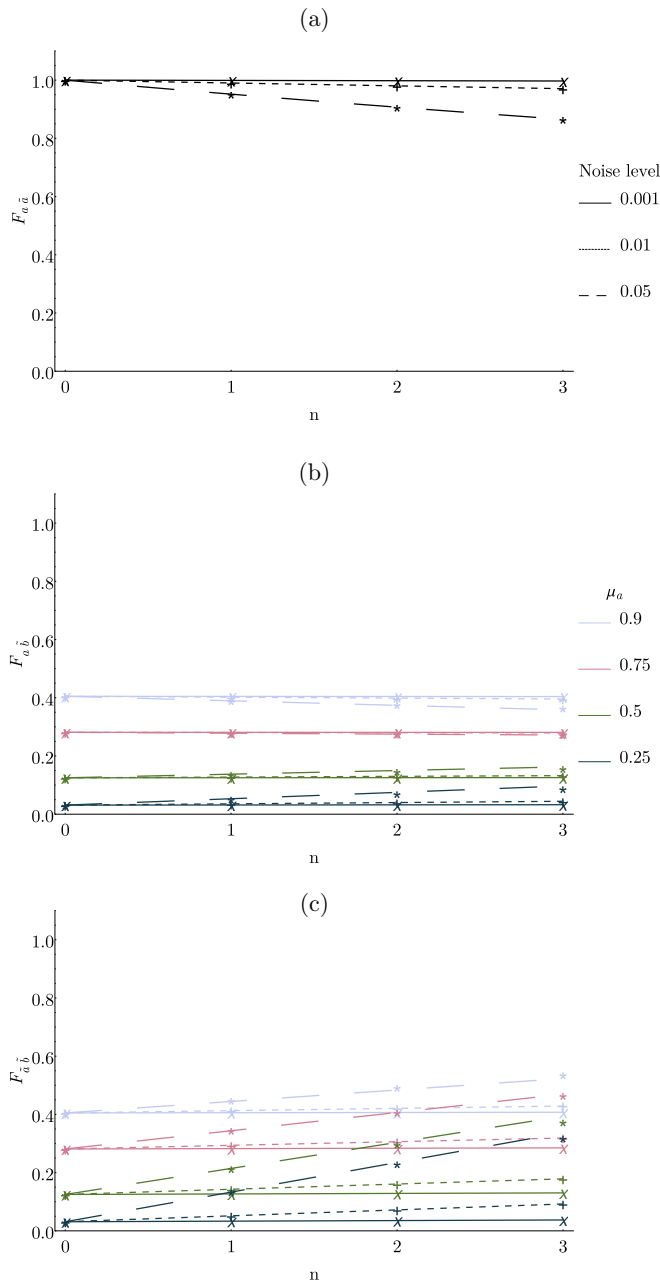


FIG. 10. Fidelities as function of the number of applied noise channels, n , (a) between the same states with noise applied to one state ($F_{a\bar{a}}$), (b) between the two different states with noise applied to one state ($F_{a\bar{b}} = F_{b\bar{a}}$), and (c) with noise applied to both states ($F_{\bar{a}\bar{b}}$). Markers show the calculated numeric fidelities using Eq. (13), and lines show the low-order expansions given by Eq. (14). The low-order expansion agrees well with the numerical results for all cases.

combination of these settings, which simultaneously minimizes errors and inconclusive results, and our relation for the loss in fact lies in between the optimum values in each of these boundary settings.

We can now verify the validity of our model for the effect of noise on the loss. To this aim we calculate $F_{a\bar{b}}$ from Eq. (14d) for $n = 3$, which corresponds to the number of entangling gates applied in our circuit to each data qubit.

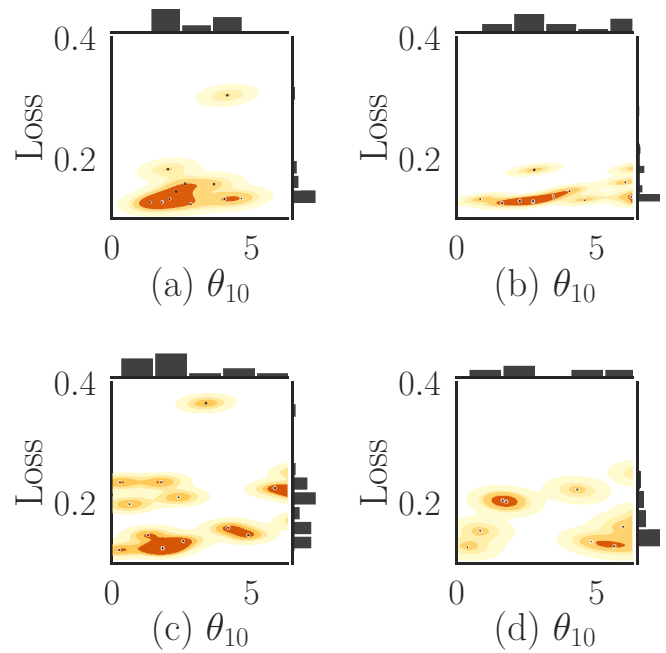


FIG. 11. The distribution of loss and θ_{10} obtained at different noise levels. Here we can see the effect of noise on the values of θ_{10} that the optimizer converges to. We only show a single representative parameter θ_{10} , since we have an approximately similar behavior for all other parameters. This is shown at different noise levels: (a) 0.0, (b) 0.001, (c) 0.01, and (d) 0.05.

These results are presented in Fig. 9 for the panels with $p \neq 0$. One can see that $F_{a\bar{b}}$ agrees rather well with the loss also in the presence of noise. This validates our model in which the effect of noise on the state discrimination is mainly determined by the noise-induced increased overlap between the states as they are processed in the circuit. In particular for the highest noise the model captures well the fact that the effect of noise is large for small μ_a , while it is reduced for larger μ_a . Equation (14d) therefore allows us to estimate the minimal loss achievable with our circuit for a given p and μ_a .

Finally, we investigate the effect of the noise during training on the actual values of the optimized parameters in the circuit. In Fig. 11 we present the distribution of θ_{10} for different values of noise. The values of θ are all taken modulo 2π , and at zero validation noise to remove the effect of validation errors on the loss. We see that the range of angles converged upon increases as the noise in the circuit increases. Some values become stuck at high loss, and there can be different values for the minimal loss parameters. The increase in noise seems to change not just the final loss, but the parameters found that minimize loss. We cannot rule out the correlation between different parameters as the noise level changes. Combined with what we see in Fig. 8, that good parameters are still found at higher noise levels, we may conclude that noise in the circuit can push the optimizer out of local minima, so that it can find some other local minima at lower loss.

From the results presented here we see that this algorithm performs well in the presence of noise in the training and validation steps (Fig. 7), and that parameters found on a noisy device work well when validated on a device with low noise

(Fig. 8). When calculating parameter gradients on a noisy quantum device, reducing the number of parameters has a positive effect, as shown in Figs. 5 and 6.

IV. CONCLUSION

We have shown that a QNN can be trained for the task of state discrimination on a noisy device, with noise levels found in current NISQ devices. We have also shown that gradient descent algorithms are viable on noisy quantum devices, given a good choice of classical algorithm. As discussed in [45], choice of training circuits in variational quantum algorithms has a large effect upon success. Here we reduced the number of parameters by removing rotation gates from the input states, and indeed show that the low circuit depth and qubit count of our algorithm is beneficial in the presence of noise. We also developed a simple model equation relating the loss to the noise level and input state, which is based on the fact that in our circuit the application of noisy gates effectively leads to an increase of the overlap between states to be discriminated during the processing of the circuit. While we specifically

considered the task of quantum state discrimination, the algorithm presented here can be equally applied to such problems as verification of general quantum machine learning outputs, and applications in sensing, imaging, and metrology.

ACKNOWLEDGMENTS

A.P. is supported by the InQUBATE Training and Skills Hub [Engineering and Physical Sciences Research Council (EPSRC) Grant No. EP/P510270/1]. I.R. acknowledges financial support from the United Kingdom Department of Business, Energy, and Industrial Strategy. H.C. acknowledges support through a Teaching Fellowship from University College London (UCL). L.W. acknowledges support through the Google PhD Fellowship in Quantum Computing. D.B. acknowledges support from the EPSRC Prosperity Partnership in Quantum Software for Modelling and Simulation (Grant No. EP/S005021/1). A.P. acknowledges the use of the UCL Myriad High Throughput Computing Facility (Myriad@UCL), and associated support services, in the completion of this work.

-
- [1] C. H. Bennett, Quantum Cryptography Using Any Two Nonorthogonal States, *Phys. Rev. Lett.* **68**, 3121 (1992).
- [2] A. Chefles, Quantum state discrimination, *Contemp. Phys.* **41**, 401 (2000).
- [3] L.-M. Duan and G.-C. Guo, Probabilistic Cloning And Identification Of Linearly Independent Quantum States, *Phys. Rev. Lett.* **80**, 4999 (1998).
- [4] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, *Nat. Photonics* **5**, 222 (2011).
- [5] S. Lloyd, Enhanced sensitivity of photodetection via quantum illumination, *Science* **321**, 1463 (2008).
- [6] M. Benedetti, E. Grant, L. Wossnig, and S. Severini, Adversarial quantum circuit learning for pure state approximation, *New J. Phys.* **21**, 043023 (2019).
- [7] IBM, Quantum systems, IBM Q (2019), <https://www.research.ibm.com/ibm-q/technology/devices/>.
- [8] S. S. Hong, A. T. Papageorge, P. Sivarajah, G. Crossman, N. Didier, A. M. Polloreno, E. A. Sete, S. W. Turkowski, M. P. da Silva, and B. R. Johnson, Demonstration of a parametrically-activated entangling gate protected from flux noise, *Phys. Rev. A* **101**, 012302 (2020).
- [9] Y. Chen, C. Neill, P. Roushan, N. Leung, M. Fang, R. Barends, J. Kelly, B. Campbell, Z. Chen, B. Chiaro, A. Dunsworth, E. Jeffrey, A. Megrant, J. Y. Mutus, P. J. J. O'Malley, C. M. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, M. R. Geller, A. N. Cleland, and J. M. Martinis, Qubit Architecture With High Coherence And Fast Tunable Coupling, *Phys. Rev. Lett.* **113**, 220502 (2014).
- [10] H. Chen, L. Wossnig, S. Severini, H. Neven, and M. Mohseni, Universal discriminative quantum neural networks, *Quantum Machine Intelligence*, **3**, 1 (2021).
- [11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT, Cambridge, MA, 2016).
- [12] M. Arjovsky, A. Shah, and Y. Bengio, Unitary evolution recurrent neural networks, in *Proceedings of the 33rd International Conference on Machine Learning*, Vol. 48, pp. 1120–1128 (2016).
- [13] J. A. Bergou, U. Herzog, and M. Hillery, Quantum Filtering And Discrimination Between Sets Of Boolean Functions, *Phys. Rev. Lett.* **90**, 257901 (2003).
- [14] M. Mohseni, A. M. Steinberg, and J. A. Bergou, Optical Realization Of Optimal Unambiguous Discrimination For Pure And Mixed Quantum States, *Phys. Rev. Lett.* **93**, 200403 (2004).
- [15] M. Nielsen, *Neural Networks and Deep Learning* (Determination Press, 2014).
- [16] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, ImageNet large scale visual recognition challenge, *International Journal of Computer Vision* **115**, 211 (2015).
- [17] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, Language models are few-shot learners, *arXiv:2005.14165* (2020).
- [18] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Repeated multi-qubit readout and feedback with a mixed-species trapped-ion register, *Journal of Machine Learning Research* **15**, 1929 (2014).
- [19] V. Negnevitsky, M. Marinelli, K. K. Mehta, H.-Y. Lo, C. Flühmann, and J. P. Home, Repeated multi-qubit readout and feedback with a mixed-species trapped-ion register, *Nature (London)* **563**, 527 (2018).
- [20] M. Ježek, J. Řeháček, and J. Fiurášek, Finding optimal strategies for minimum-error quantum-state discrimination, *Phys. Rev. A* **65**, 060301(R) (2002).
- [21] W. Wang, G. Dong, and M. Li, Minimum-error quantum state discrimination based on semidefinite programming, in *Proceed-*

- ings of the 27th Chinese Control Conference, 16-18 July 2008* (IEEE, Piscataway, NJ, 2008), pp. 521–524.
- [22] Y. C. Eldar, A semidefinite programming approach to optimal unambiguous discrimination of quantum states, *IEEE Trans. Inf. Theory* **49**, 446 (2003).
- [23] C. Cîrstoiu, Z. Holmes, J. Iosue, L. Cincio, P. J. Coles, and A. Sornborger, Variational fast forwarding for quantum simulation beyond the coherence time, *npj Quantum Inf.* **6**, 82 (2020).
- [24] C. Bravo-Prieto, J. Lumbreras-Zarapico, L. Tagliacozzo, and J. I. Latorre, Scaling of variational quantum circuit depth for condensed matter systems, *Quantum* **4**, 272 (2020).
- [25] P. K. Barkoutsos, G. Nannicini, A. Robert, I. Tavernelli, and S. Woerner, Improving variational quantum optimization using CVaR, *Quantum* **4**, 256 (2020).
- [26] N. Moll, P. Barkoutsos, L. S. Bishop, J. M. Chow, A. Cross, D. J. Egger, S. Filipp, A. Fuhrer, J. M. Gambetta, M. Ganzhorn, A. Kandala, A. Mezzacapo, P. Müller, W. Riess, G. Salis, J. Smolin, I. Tavernelli, and K. Temme, Quantum optimization using variational algorithms on near-term quantum devices, *Quant. Sci. Technol.* **3**, 030503 (2018).
- [27] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (ICLR, San Diego, 2015).
- [28] P. Gokhale, Y. Ding, T. Propson, C. Winkler, N. Leung, Y. Shi, D. I. Schuster, H. Hoffmann, and F. T. Chong, Partial compilation of variational algorithms for noisy intermediate-scale quantum machines, in *Proceedings of the 52nd Annual IEEE/ACM International Symposium on Microarchitecture: MICRO '52* (ACM, New York, 2019), pp. 266–278.
- [29] R. Sweke, F. Wilde, J. Meyer, M. Schuld, P. K. Fährmann, B. Meynard-Piganeau, and J. Eisert, Stochastic gradient descent for hybrid quantum-classical optimization, *Quantum* **4**, 314 (2020).
- [30] K. E. Hamilton, E. F. Dumitrescu, and R. C. Pooser, Generative model benchmarks for superconducting qubits, *Phys. Rev. A* **99**, 062323 (2019).
- [31] J.-G. Liu, Y.-H. Zhang, Y. Wan, and L. Wang, Variational quantum eigensolver with fewer qubits, *Phys. Rev. Research* **1**, 023025 (2019).
- [32] A. Neelakantan, L. Vilnis, Q. V. Le, I. Sutskever, L. Kaiser, K. Kurach, and J. Martens, Adding gradient noise improves learning for very deep networks, [arXiv:1511.06807](https://arxiv.org/abs/1511.06807) (2015).
- [33] M. Ostaszewski, E. Grant, and M. Benedetti, Quantum circuit structure learning, [arXiv:1905.09692](https://arxiv.org/abs/1905.09692) (2019).
- [34] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, *Nat. Commun.* **9**, 4812 (2018).
- [35] A. Hayashi, T. Hashimoto, and M. Horibe, State discrimination with error margin and its locality, *Phys. Rev. A* **78**, 012333 (2008).
- [36] T. Hashimoto, A. Hayashi, M. Hayashi, and M. Horibe, Unitary-process discrimination with error margin, *Phys. Rev. A* **81**, 062327 (2010).
- [37] E. Bagan, R. Muñoz-Tapia, G. A. Olivares-Rentería, and J. A. Bergou, Optimal discrimination of quantum states with a fixed rate of inconclusive outcomes, *Phys. Rev. A* **86**, 040303(R) (2012).
- [38] U. Herzog, Optimal measurements for the discrimination of quantum states with a fixed rate of inconclusive results, *Phys. Rev. A* **91**, 042338 (2015).
- [39] J. A. Nelder and R. Mead, A simplex method for function minimization, *Comput. J.* **7**, 308 (1965).
- [40] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, *Phys. Rev. A* **98**, 032309 (2018).
- [41] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, *Phys. Rev. A* **99**, 032331 (2019).
- [42] R. M. Parrish, E. G. Hohenstein, P. L. McMahon, and T. J. Martinez, Hybrid quantum/classical derivative theory: Analytical gradients and excited-state dynamics for the multistate contracted variational quantum eigensolver, [arXiv:1906.08728](https://arxiv.org/abs/1906.08728) (2019).
- [43] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, A variational eigenvalue solver on a photonic quantum processor, *Nat. Commun.* **5**, 4213 (2014).
- [44] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed. (Cambridge University, New York, 2011).
- [45] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms, *Adv. Quantum Technol.* **2**, 1900070 (2019).
- [46] E. Knill, Quantum computing with realistically noisy devices, *Nature (London)* **434**, 39 (2005).
- [47] D. Gottesman, An introduction to quantum error correction and fault-tolerant quantum computation, [arXiv:0904.2557](https://arxiv.org/abs/0904.2557) (2010), pp. 13–58.
- [48] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, *New J. Phys.* **18**, 023023 (2016).
- [49] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous distributed systems, [arXiv:1603.04467](https://arxiv.org/abs/1603.04467) (2016).
- [50] C. Gidney, D. Bacon, and The Cirq Developers, quantumlib/Cirq: A python framework for creating, editing, and invoking Noisy Intermediate Scale Quantum (NISQ) circuits, 2018.
- [51] R. Jozsa, Fidelity for mixed quantum states, *J. Mod. Opt.* **41**, 2315 (1994).
- [52] S. M. Barnett and S. Croke, Quantum state discrimination, *Adv. Opt. Photonics* **1**, 238 (2009).
- [53] J. Bae and L.-C. Kwek, Quantum state discrimination and its applications, *J. Phys. A: Math. Theor.* **48**, 083001 (2015).