Deterministic optimal quantum cloning via a quantum-optical neural network

Denis Stanev [1,2], Nicolò Spagnolo [2], and Fabio Sciarrino [2,*]
[1]*Gran Sasso Science Institute, Viale Francesco Crispi 7, I-67100 L'Aquila, Italy*
[2]*Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 5, I-00185 Roma, Italy*

(Received 9 November 2022; revised 6 January 2023; accepted 10 January 2023; published 23 February 2023)

Quantum machine learning represents a promising approach to disclose novel possibilities for quantum information protocols. Thus, recent efforts have been dedicated to determine how to merge, or extend, machine learning approaches in the quantum domain. A relevant example is provided by a quantum-optical neural network (QONN), a class of physical systems that aim to represent the quantum-optical version of classical neural networks. This architecture was first introduced to identify the structure of optical circuits in analogy to their classical counterpart, showing the possibility of tackling different quantum tasks. However, the full potential of such an approach still needs to be explored. In this work, we provide a step forward in this direction by employing it to tackle the fundamental task of optical quantum cloning. The employed architecture has been successfully trained to perform deterministic universal quantum cloning on a photonic platform, achieving results that are very close to that of an optimal universal quantum cloner. Furthermore, we have verified numerically the robustness of this approach with respect to experimental imperfections relevant for future implementations. This work thus shows the capabilities and flexibility of QONNs, which might prove important for designing even more complex optical circuits.

DOI: 10.1103/PhysRevResearch.5.013139

I. INTRODUCTION

In recent years, several technological advances have significantly improved our ability to handle quantum systems with progressively larger complexity for quantum information applications [1]. Such applications include quantum computing, simulation, communication, and sensing. In particular, these advances in the available platforms have enabled the first experiments [2–6] tackling the quantum advantage regime, namely the scenario in which quantum devices are capable of outperforming classical computers in specific tasks. These results have thus paved the way for the development and application of noisy intermediate-scale quantum (NISQ) processors [7] for quantum-enhanced information processing.

The aforementioned technological improvements are currently accompanied by new prospects for the development of quantum algorithms. A promising direction has been identified in seeking to extend machine learning approaches in the quantum domain, thus merging that discipline with the quantum information field. Machine learning, and more specifically deep learning [8], is currently being applied to different fields, ranging from physics [9] to medicine [10]. Most machine learning algorithms start with (usually) randomized parameters and then seek to change these parameters over multiple iterations to optimize the value of a function. These methods can be applied to a large variety of tasks, and their effectiveness is based on the manifold hypothesis [11], which states that real-world high-dimensional data lie on low-dimensional manifolds embedded within high-dimensional space.

Among all the different machine learning techniques, neural networks [12], i.e., machine learning models that are inspired by the biological neural networks that constitute the brains of living beings, have attracted by far the most attention in recent years. They are currently considered to be one of the most promising approaches thanks to their good performance and high flexibility, making them well suited for a large number of applications, such as image recognition [13] and automated driving [14]. This has led to the creation of several different neural network models, each one specialized to tackle specific tasks, such as convolutional neural networks for image detection and long short-term memory for semantic parsing.

The merging of machine learning and quantum information has been pursued in different directions. On one side, conventional machine learning methods have been shown to be an effective solution to control [15–17], characterize [18–21], or analyze the output [22–25] of quantum devices. Conversely, a different approach called quantum machine learning aims at devising a novel class of algorithms directly extending the concept of machine learning in the quantum domain [26–30]. For instance, quantum machine learning algorithms for universal quantum computers have been proposed, leading to the first small-scale implementations [31]. Additionally, the first analysis has been reported for applications in other research areas [32]. Within this context, an architecture to design a

*fabio.sciarrino@uniroma1.it

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

quantum-optical neural network was proposed in Ref. [29]. This system represent a type of optical network designed specifically to represent the quantum-optical version of classical neural networks. In Ref. [29], a first investigation on the classes of tasks that can be solved via this approach was reported, with further analysis reported in Ref. [33]. However, the potentiality of such an approach still needs to be fully explored.

In this paper, we perform a step forward by applying the quantum-optical neural network (QONN) architecture to identify an appropriate circuit realizing the optimal quantum cloning process [34,35], a fundamental process in different quantum information protocols [1,36–41]. Besides a circuital approach, several optical implementations of universal quantum cloning on an optical platform have reported [41–48]. However, current optical implementations are inherently probabilistic. Recently, variational techniques have been proposed as a tool to optimize quantum circuits for quantum cloning [49]. Here, we numerically show that the QONN is capable of learning the correct structure to perform nearly optimal deterministic universal quantum cloning.

In Sec. II we provide a brief overview of the optimal quantum cloning problem. Then, in Sec. III we discuss the architecture and the employed training approach for the QONN circuit tackling the chosen task. In Sec. IV we then report the obtained numerical results, showing that the system can achieve fidelity of the operation close to the optimal value. Finally, in Sec. V we perform detailed numerical simulations to show the impact of the relevant noise factors in future implementations.

II. OPTIMAL QUANTUM CLONING IN OPTICAL PLATFORMS

Cloning is a process that takes an arbitrary, unknown state and seeks to make an exact copy without altering the original state. In classical computing, it is a fundamental operation that is at the basis of all computing architectures. In the quantum domain, however, exact cloning of an unknown quantum state is forbidden by the laws of quantum mechanics as shown by the no-cloning theorem, which states that there is no admitted operation for cloning any arbitrary state perfectly [50]. However, approximate quantum cloning is possible, and it has been shown [34,51] that the maximum $1 \rightarrow 2$ single copy fidelity that is theoretically possible for a universal quantum cloner is $F_1^{\text{opt}} = \langle\psi|\rho_1|\psi\rangle = 5/6$, this latter quantity measuring the distance between two quantum states. As such, a single copy fidelity of F_1^{opt} indicates that the clone has a $5/6$ chance of collapsing into the state of the original qubit when measured in the original qubit basis. It is also important to note that, if partial information of the input state is known, one can design a cloning machine for this kind of input state, which achieves higher fidelities than that of a universal quantum cloner [52,53]. Quantum cloning thus refers to a process that approximately copies qubit states, changing the original qubit in the process.

Subsequent studies have thus focused on the implementation of quantum devices, called quantum cloners, tailored to perform the cloning task reaching optimal fidelity. These machines obtaining the maximum theoretically possible fidelity are called optimal quantum cloners [51]. When no *a priori* information on the state is available, these devices are called universal optimal quantum cloners. A circuit model for such a machine has been defined in [34,35]. However, when talking about optical quantum cloning [54], all of the implementations that have been proposed and realized so far are probabilistic in nature [41–48], and thus typically rely on postselection costs [55], which limits the amount of useful signal. Conversely, the target operation of cloning machines to avoid such overhead is to obtain optimal performances in a deterministic fashion.

III. QUANTUM-OPTICAL NEURAL NETWORK FOR QUANTUM CLONING

An architecture to define the quantum-optical extension of neural networks has been proposed in Ref. [29]. Such an architecture is made up of multiple layers, each composed of a linear unitary transformation and a set of nonlinear single-site transformations. Input states in the network are photon states employed to encode qubits. More specifically, the dual-rail encoding is used to associate each pair of modes to a qubit, where states $\{|0\rangle, |1\rangle\}$ correspond to a single photon in the first and second mode, respectively. The QONN can be considered the quantum-optical version of the multilayer perceptron (MLP), where the linear transformations generated by the weight matrices are replaced by the unitary transformations generated by a set of matrices $U_k^{(L)}$ and the nonlinear activation functions are replaced by single-site optical nonlinearities with a Kerr-type interaction $\Sigma(\chi)$ [a schematic of the QONN structure is shown in Fig. 1(a)].

In this work, we will use the QONN architecture to implement the optical quantum cloning task. We will consider the adoption of the backpropagation method in the training process, which allows for a significant reduction in training time. It must be noted that there is currently no known way to use backpropagation when training directly on quantum hardware, and thus this technique can only be used when training the QONN using simulations. Of course, the trained network can then be physically realized using the trained parameters, so this limitation only applies to circuits that are too complex to be simulated classically. The topic of how to efficiently train larger quantum machine learning (QML) models is a very important and challenging one, requiring several investigations in the future. To fully exploit the backpropagation approach, we will also employ a detection method for the output state by measuring the output modes in pairs, as explained in Sec. III C. This allows for a faster implementation of the backpropagation approach, compared to measuring the whole state on all modes, also enabling fewer resources for the physical implementation.

A. Layer structure

The quantum-optical neural network considered here is made up of multiple layers, as shown in Fig. 1(a). Each layer of our network can be divided into two parts. The first block is a linear transformation $U_k^{(L)}$, consisting of an $M \times M$ unitary interferometer according to the universal architecture of Ref. [56], decomposed in a network of beam-splitters

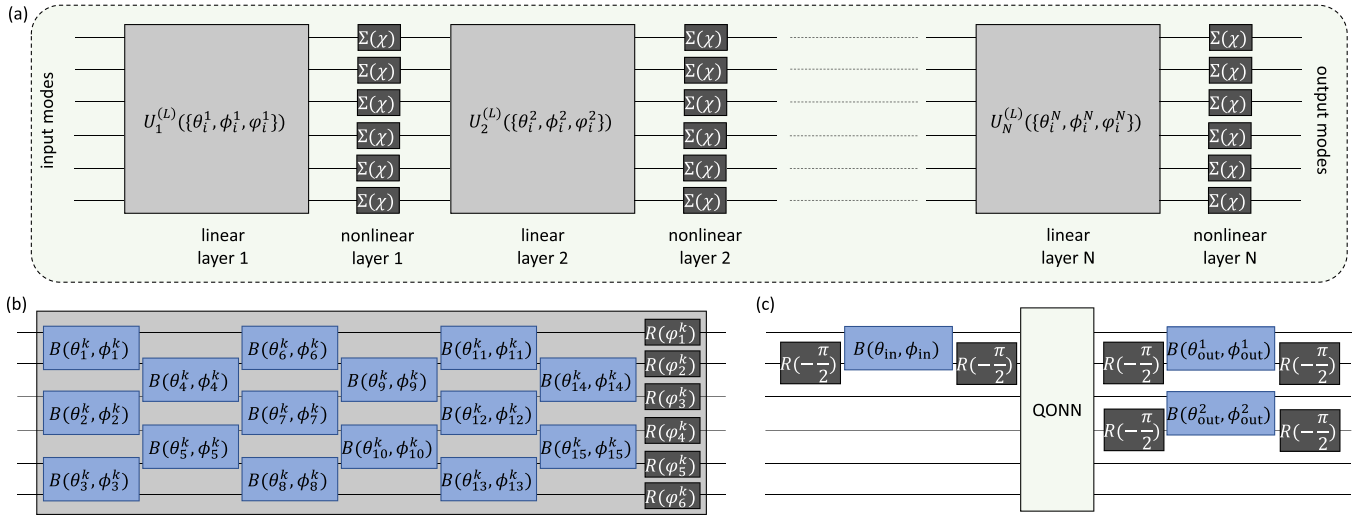


FIG. 1. (a) Structure of quantum-optical neural network architecture. This is composed of a sequence of layers performing a linear evolution, $U_k^{(L)}$, interspersed by layers composed of single-site nonlinearities with a Kerr-type interaction $\Sigma(\chi)$. (b) Implementation of a general linear transformation $U_k^{(L)}$ according to the structure defined in [56], composed of a sequence of beam-splitter transformations $B(\theta_i^k, \phi_i^k)$ and a layer of output phase $R(\varphi_i^k)$. (c) Complete circuit including the input state preparation stage [obtained via the sequence $R(-\pi/2)$, $B(\theta_{\text{in}}, \phi_{\text{in}})$, $R(-\pi/2)$] and the output state measurement one [implemented as the sequence $R(-\pi/2)$, $B(\theta_{\text{out}}^k, \phi_{\text{out}}^k)$, $R(-\pi/2)$ on each pair of modes corresponding to a logical qubit].

and phase shifters with appropriate layout. More specifically, the architecture for M modes requires $M(M-1)/2$ beam-splitters and M phase shifters as shown in Fig. 1(b). The second block is a set of nonlinear single-site transformation, in this case corresponding to Kerr interactions $\Sigma(\chi)$ on each mode:

$$\Sigma(\chi) = \sum_{n=0}^{\infty} e^{\imath n(n-1)\chi} |n\rangle\langle n|. \tag{1}$$

To reach the desired operation, the training process on the network is performed only on the linear part. More specifically, the parameters that are learned in the process are the beam-splitter transmittivities and the phase shifters, while the nonlinear phases χ for the Kerr layer is fixed to $\chi = \frac{\pi}{2}$. Indeed, each beam-splitter B is described by a 2×2 matrix depending on two parameters, θ and ϕ:

$$B(\theta, \phi) = \begin{bmatrix} \cos\theta & -e^{-\imath\phi}\sin\theta \\ e^{\imath\phi}\sin\theta & \cos\theta \end{bmatrix}. \tag{2}$$

Additionally, a single layer of phase shifts at the output $\{\varphi_i\}$ requires one additional parameter per mode. This means that in total each layer has $2M(M-1)/2 + M = M^2$ trainable parameters. In the specific application for cloning, we require the network to have six modes (see Fig. 1).

Regarding the nonlinear layers, they serve the same role as activation functions in classical MLP models, and thus it is not necessary to train them. Their role is to increase the expressivity of the network, allowing for an accurate approximation of any kind of function, not limited to only linear ones [57,58]. The usefulness of nonlinear layers in the quantum case is shown by the observation that increased performances are obtained by increasing the number of layers. It is important to note that, in principle, this is not limited to Kerr nonlinearities, but in fact any kind of nonlinearity could be used, although there would likely be differences in performance depending on the adopted evolution.

B. Network structure

As anticipated in the previous section, the QONN discussed in this manuscript is composed of a sequence of layers having the structure highlighted above. The number of layers to be included in the full network has to be chosen depending on the following considerations. On the one hand, increasing the number of layers is typically associated with improved expressivity of the network, making it possible to solve problems with higher complexity. In parallel, this corresponds to a larger number of parameters that need to be trained, increasing the training time and thus the convergence of the training process to an effective set of system parameters to solve the chosen task. Hence, it is necessary to verify the optimal compromise between the complexity of the network and the training time. Another point that is important to consider is that, when realizing the physical circuit, a larger number of layers will be associated with increasing errors caused by imperfections in the actual implementation. Thus, it is crucial to choose the minimal amount of layers that still provide enough expressivity to solve relevant tasks. For these reasons, in the optimal cloning scenario we considered as a good compromise the adoption of a seven-layer network. As shown in Fig. 1(c), the system requires the addition of a module before the QONN, used to prepare the input states, and a second module after the QONN, required to measure the output states.

Training of the input according to the employed methodology requires preparation of six input states corresponding to the three mutually unbiased orthonormal bases, namely the computational basis ($|0\rangle, |1\rangle$), the diagonal basis ($\frac{|0\rangle+|1\rangle}{\sqrt{2}}, \frac{|0\rangle-|1\rangle}{\sqrt{2}}$), and the circular basis ($\frac{|0\rangle+i|1\rangle}{\sqrt{2}}, \frac{|0\rangle-i|1\rangle}{\sqrt{2}}$). As previously introduced, encoding of the logical states is

performed by considering pairs of modes, and assigning the logical values $|0\rangle$ ($|1\rangle$) depending on the presence of a single photon in modes 1 (2) in the mode pair. The chosen bases set has the relevant property of being complete and balanced, as well as being the minimal set satisfying this feature. Being complete, this ensures that it is not necessary to expand it to obtain improved training capabilities. Furthermore, being balanced, the training process does not favor operation of the circuit on one specific basis with respect to the others. Finally, avoiding overfitting via this approach also permits us to train the network without requiring the additional use of techniques (such as early stopping [59]) that would usually be employed to prevent such an undesired effect, at the cost of reduced accuracy.

To prepare these input states, we start with either the $|0\rangle$ or the $|1\rangle$ state on the first pairs of modes, which in our encoding are represented by a photon in the first mode or a photon in the second mode, respectively. States $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$ and $\frac{|0\rangle-|1\rangle}{\sqrt{2}}$ are obtained via application of a symmetric beam-splitter and phase shifters on the second mode [60]. Preparation of states $\frac{|0\rangle+i|1\rangle}{\sqrt{2}}$ and $\frac{|0\rangle-i|1\rangle}{\sqrt{2}}$ requires an additional $\varphi_{\text{in}} = \frac{\pi}{2}$ phase shift on the second mode. This corresponds to the set of transformations shown in Fig. 1(c) before the QONN. The complete input state is thus prepared as follows. The two upper modes are employed to encode the qubit to be cloned, while the other two qubits are initialized to the state $|0\rangle$, thus requiring a single-photon in the first mode of each subsequent pairs (input modes 3 and 5 of the network).

After evolution in the QONN, the aim is to obtain an optimal copy of the input qubit in the second qubit. Thus, it is necessary to measure the output of the first four modes, corresponding to the first two qubits, while the remaining part is treated as an ancilla for the computation. Measurement of the output state can be performed with different approaches. On the one hand, complete information can be obtained via state tomography, which requires a progressively large number of measurements and thus may not represent the optimal choice for our purpose. For the chosen task (quantum cloning), knowledge of the target output state can be exploited during the training process. Indeed, a more efficient approach can be devised by projecting the output qubit onto the desired state, allowing us to measure the fidelity of the cloning with a single configuration for each state. This projection is done via a sequence of a phase shift and a Hadamard gate, thus reversing the sequence of operation for the input preparation [the complete scheme is shown in Fig. 1(c)].

C. Training the network

Once the structure of the quantum-optical neural network is defined, we can describe the training process. A custom code for the quantum simulations and the training of the QONNs was written in PYTHON 3 [61], and it was based on the libraries Pennylane [62], Strawberry Fields [63], and Tensorflow [64]. As discussed in Sec. III B, input states are prepared according to the scheme of Fig. 1(c). This permits us to prepare the set of six states belonging to three mutually unbiased orthonormal bases. A single epoch of the training process corresponds to calculating (or measuring, if performed on an actual device) the probability of each combination of photons in the first two pairs of modes. More specifically, this is performed for a fixed value of the internal parameters (beam-splitter transmittivities and phases) of the neural network. The set of probabilities y_j corresponding to a given network configuration is then compared with the expected predictions $\hat{y}_j$ via a suitably chosen cost function $C(\{\theta_i^k, \phi_i^k, \varphi_l^k\})$. In our case, the chosen cost function is the mean-squared error (MSE), defined as

$$C\left(\left\{\theta_i^k, \phi_i^k, \varphi_l^k\right\}\right) = \frac{\sum_{j=1}^{n}(y_j - \widehat{y_j})^2}{n}, \tag{3}$$

where n is the number of relevant probabilities (see below). This cost function was chosen because it leads to balanced performances with respect to the full set of states. Once we have obtained the loss, we use backpropagation to decide how to adjust the values of the network parameters in order to reduce the loss. More specifically, the new set of parameters is defined for the subsequent epoch through minimization of the cost function via the ADAM optimizer [65]. In the following, we then discuss specific details regarding such a training process.

The first relevant aspect addresses the chosen dual-rail encoding for the qubit implementation with respect to the architecture of the QONN. Indeed, such a choice of encoding requires input and output states where only one photon is present per each pair of modes, corresponding to the logical $|0\rangle$ and logical $|1\rangle$ states. Conversely, the optical architecture corresponding to the QONN provides a photon evolution that is not limited to input states that have a logical meaning in dual-rail encoding. Thus, the training process must favor a circuit operation leading to input-output states in such an encoding. To this end, it is not sufficient to measure the mean photon number in each output mode after the projection, since the process will not distinguish between states associated with logical qubits rather than other states, such as those with two photons on the same output mode or those with photons in both wires associated with the same qubit. To take such an aspect related to the logical encoding into account, the training process exploits output probabilities related to pairs of adjacent output modes. More specifically, given a pair of modes $(j,\ j+1)$ corresponding to the two rails of a logical qubit [e.g., pairs (1,2), (3,4), and (5,6)], the process includes within the set y_i all the probabilities $\Pr(n_j, n_{j+1})$ related to measuring n_j photons in mode j, and n_{j+1} photons in mode $j+1$ at the output of the network, with the constraint of having $n_j + n_{j+1} < N$, with N the total number of input photons. Then, the target probabilities $\hat{y}_i$ for all configurations (n_j, n_{j+1}) that do not correspond to a logical qubit are set to 0. Namely, only the two configurations $(n_j = 1, n_{j+1} = 0)$ and $(n_j = 0, n_{j+1} = 1)$ corresponding to logical qubits $|0\rangle$ and $|1\rangle$ have a nonzero associated target probability. In such a way, during the training process the network is driven to generate only states associated with logical qubits in dual rail encoding.

One of the main issues we encountered is the possibility that the training process will fail to reach the global minimum while getting stuck in a local one. While this possibility can be reduced by an appropriate choice of the optimizer, such as the ADAM one, the probability of this undesired outcome of the training process is still non-negligible. We then chose to divide the training process in two different steps. (i)

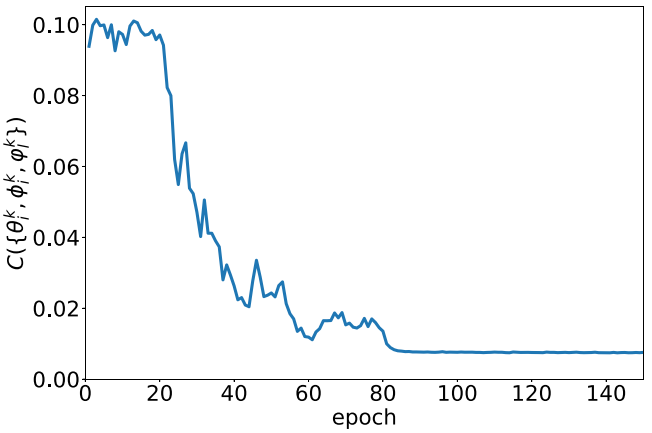


FIG. 2. Evolution of the cost function $C(\{\theta_i^k, \phi_i^k, \varphi_l^k\})$ during the epochs in a successful first step of training.

The first stage is performed by using a relatively high learning rate, with the aim of reaching the parameter space region close to the global minimum. (ii) In the second stage, we performed a longer and more refined training, with a smaller learning rate than step (i), to perform fine-tuning optimization around the global minimum. During both training steps, we use a decaying learning rate. This two-step procedure has been shown to reduce significantly the probability of reaching a local minimum, thus reducing the effective time required for the overall process.

In the first stage of the training process, we employed the MSE cost function to avoid obtaining asymmetric results with respect to the target values $\hat{y}_i$ for the probabilities $\Pr(n_j, n_{j+1})$. In particular, we set to 1 only the probabilities for the configurations corresponding to the output associated with perfect clones. This choice is performed to provide an additional boost to the learning process to approach the region corresponding to a global minimum. Note that these values are effectively forbidden by the laws of quantum mechanics, since they correspond to a perfect (unitary fidelity for the clones) universal cloning machine. A first interesting thing to note is that, even though we initially set the target values to be those corresponding to a physically impossible scenario, we find that the training process reaches the correct configuration. Additionally, when the probability of cloning the correct state approaches the (physically accessible) optimal value of $5/6$, the probability of obtaining the orthogonal state converges to the value of $1/6$. Hence, the system naturally converges to a configuration close to the expected result for an optimal cloning machine. This also motives the choice of the target probabilities to be inserted in the second stage. In Fig. 2 we show graph of the loss, i.e., the cost function $C(\{\theta_i^k, \phi_i^k, \varphi_l^k\})$, during the epochs in a successful first step of training.

For the second stage of training, we first keep the MSE cost function and the target to be a clone with unitary fidelity. Then, after some training with those conditions we switch the cost function to MAE (mean absolute error), and we switch the targets to the theoretical physical accessibility maximum fidelity, meaning that we set the probability $\Pr(n_j, n_{j+1})$ of a perfect clone to $5/6$ and the probability of the orthogonal state to $1/6$. This choice, in parallel to the adoption of a low learning rate, reduces the probability that the training process exits from the region corresponding to a global minimum in this second stage.

Finally, we have considered also the possibility of extending the depth of the network to verify whether it can lead to significant improvements in the system performances. In particular, after training the seven-layer network and setting the parameters of the first section accordingly, we included three additional layers by initializing their parameter such that the added part performs the identity operation. This new 10-layer network is then trained with a low learning rate. We discuss the output of this configuration in the following section by analyzing the interplay between the slightly improved fidelities and the unavoidable errors introduced by an actual implementation of the network.

IV. RESULTS

Once we complete the training, we stop updating the parameters and we test the trained network. More specifically, the trained circuit can be now used to act as a universal optimal cloning machine on unknown input states. To measure the performance of the trained network, we tested the fidelity of the output clones with 1000 different input states, randomly generated over the Bloch sphere. These states are generated and measured via the two blocks shown in Fig. 1(c). More specifically, the input states are prepared via a beam-splitter $B(\theta_{\rm in}, \phi_{\rm in})$ with $\phi_{\rm in} = \frac{\pi}{2}$ and $\theta_{\rm in} \in [0, 2\pi)$ taken randomly from 0 to 2π, with a phase shift $R(-\frac{\pi}{2})$ on the second mode both before and after the beam-splitter. This gives us the matrix

$$U(\theta_{\rm in}) = \begin{bmatrix} \cos\theta_{\rm in} & \sin\theta_{\rm in} \\ \sin\theta_{\rm in} & -\cos\theta_{\rm in} \end{bmatrix}. \tag{4}$$

Finally, a phase shifter $R(\varphi_{\rm in})$ with $\varphi_{\rm in} \in [0, 2\pi)$ is inserted in the lower mode. The circuit is then injected with the logical state $|000\rangle$, corresponding to a three-photon input state in modes (1,3,5). The above circuit permits us to generate a logical input state of the form $|\psi 00\rangle$, where $|\psi\rangle = \cos\theta_{\rm in}|0\rangle + e^{i\varphi_{\rm in}} \sin\theta_{\rm in}|1\rangle$ is the qubit to be cloned.

The output fidelity of the clones is then measured by projecting the output qubits on the specific basis corresponding to the desired output [see Fig. 1(c)]. For each clone, i.e., for each pair of modes, this directly provides an estimate of the fidelity as the probability of getting one photon on the first mode and zero photons on the second mode, which is equivalent to the qubit state $|0\rangle$. As a relevant note, such a figure of merit provides the probability that we obtain the desired output given an arbitrary input without applying postselection. Thus, reaching performances close to the optimal value provides an indication that the device is acting as a deterministic quantum cloner.

We then tested the performances of a seven-layer QONN by considering both the results obtained after the first state of training and after the second part acting as a refinement. The obtained results (see Table I) show the capability of reaching performances that are close to the optimal fidelity value $F_1^{\rm opt} = 5/6$. As a second study, we verified whether additional layers can lead to increased performances. The

TABLE I. Fidelities obtained for the different trained QONNs. In the table we report the average single-copy fidelity $\langle F_1 \rangle$, estimated as described in the main text, over all tested random states, and the corresponding maximum $F_1^{\max}$ and minimum $F_1^{\min}$ values. Furthermore, we also report the value for the two-copy fidelity $\langle F_2 \rangle$. The obtained values should be compared with the optimal single-copy $F_1^{\text{opt}} = 5/6 \sim 0.8333$ and two-copy $F_2^{\text{opt}} = 2/3 \sim 0.6667$ fidelities.

Network	$\langle F_1 \rangle$	$F_1^{\max}$	$F_1^{\min}$	$\langle F_2 \rangle$
seven layers, first stage	0.8243	0.8379	0.8092	0.6576
seven layers, second stage	0.8269	0.8371	0.8143	0.6605
ten layers	0.8297	0.8364	0.8230	0.6633

ten-layer QONN, after the training, was able to achieve slightly improved performances compared with its seven-layer counterpart (see Table I). Given the larger amount of parameters, imperfections in a physical circuit implementing these QONNs could make the ten-layer network perform worse than the seven-layer one. For this reason, it is important to choose the appropriate network depth depending on the quality of the available optical elements. It is important to note that the input states we used in this test were mostly states on which the network was not trained. These results show that the system is not memorizing the input-output combinations that we fed it during training, but is actually learning how to perform genuine optimal quantum cloning.

As an additional analysis, we report in Fig. 3 a fidelity heatmap over the input states for the fully trained seven-layer and the ten-layer QONN. Such heatmaps are obtained by choosing 25 equally spaced values of both $\theta_{\text{in}} \in [0, 2\pi)$ and $\varphi_{\text{in}} \in [0, 2\pi)$, and then evaluating the fidelity for each of those different input states distributed over the Bloch sphere. We observe that the ten-layer QONN improved average performance comes mainly from an increase for those states having lower values of the fidelity, although the pattern of low-fidelity and high-fidelity zones remains mostly unchanged. This behavior is likely due to the low learning rate used when training the ten-layer QONN, which leads only to small corrections on the performances of the seven-layer network.

All these results thus show that QONNs can be successfully trained to be a deterministic optical universal quantum cloner.

V. IMPACT OF IMPERFECTIONS

We investigate in this section the impact on the performances derived from imperfections when seeking an experimental realization of the QONN scheme. In particular, in our case we consider how the fidelity of the trained optimal quantum cloner is affected by noisy circuit implementations. In this analysis, we will not consider the Kerr parameters χ, which are not part of the training process while being fixed for the network. Starting from an *a priori* characterization, small deviations from the ideal $\chi = \pi/2$ may be expected to be (at least partially) compensated during the training process. Nevertheless, we leave this aspect open for future analyses.

As a main source of error, we focused on the implementation of the tunable beam-splitters, in particular when using integrated photonic technologies. The equivalent

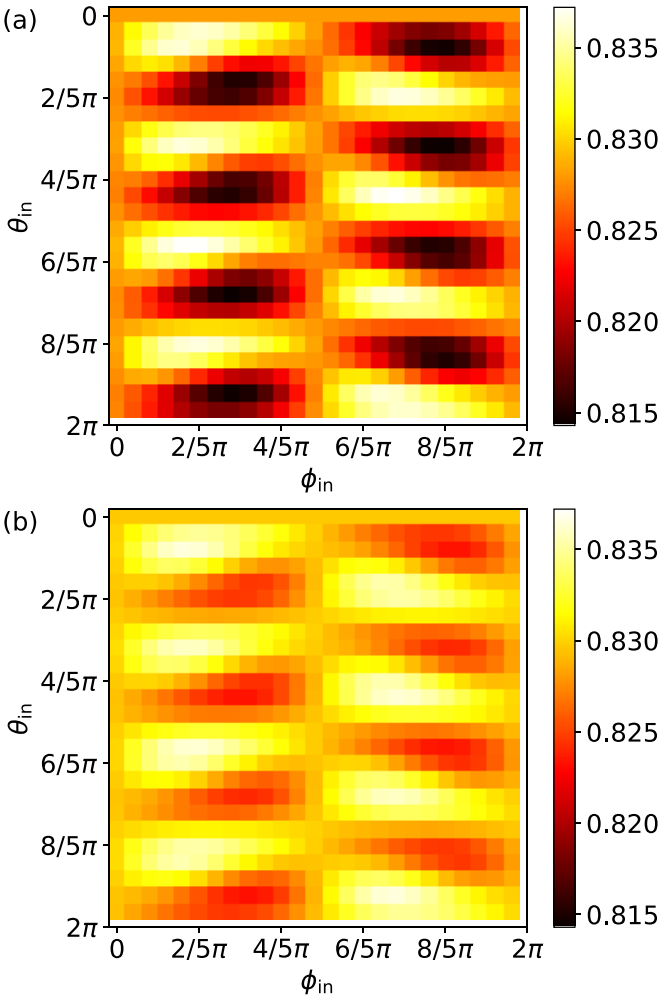


FIG. 3. Fidelity heatmap as a function of the input state parameters $(\theta_{\text{in}}, \phi_{\text{in}})$ for (a) the seven-layer after both stages of training, and (b) the ten-layer QONN.

operation of a tunable beam-splitter is commonly obtained by the circuit shown in Fig. 4, composed of two symmetric directional couplers, and two phase-shifters that are typically implemented exploiting the thermo-optic effect. In an actual implementation, one needs to consider the common experimental fabrication errors on such fundamental building

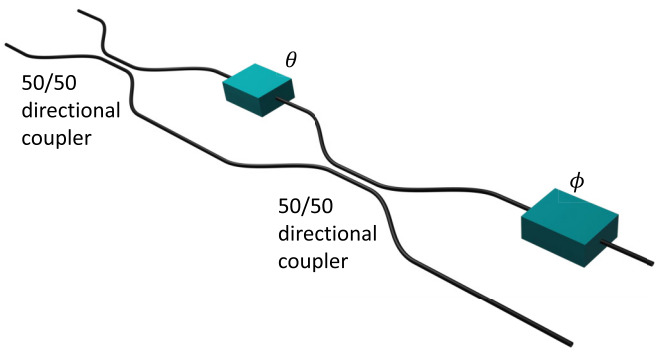


FIG. 4. Integrated decomposition of a tunable beam-splitter operation into two 50-50 directional couplers and two tunable phase-shifters.

blocks. On the one hand, thermal phase-shifters have been demonstrated in several platforms (see Ref. [66] for a recent review) to be capable of reaching the full $[0, 2\pi)$ interval. Their correct operation then shifts the problem to the characterization of the response function between voltage applied to the thermal elements, and the actual phase added in the circuit [19]. For a sufficiently large characterization data set, therefore, it is in principle possible to obtain a reconstruction of such a response function with enough accuracy. However, it is often the case that the directional couplers that are fabricated do not achieve perfect symmetric operations. Overall, this leads to a restricted range of transmittivity for the tunable beam-splitter operation. Indeed, by replacing the transmission coefficients of the symmetric couplers from the ideal value $t_i = r_i = 1/\sqrt{2}$, we find that the overall transmissivity of the block of Fig. 4 reads $T = |t_1 t_2 e^{i\theta} - r_1 r_2|^2$, where (t_1, r_1) and (t_2, r_2) are the parameters of the first and second coupler. However, as soon as the values of t_1 and t_2 stray from their perfect value, the maximum (minimum) of the transmittivity of the tunable beam-splitter operation decreases (increases), limiting the capabilities of its operation. Overall, this effect reduces the interval achievable from [0,1] to $[T_{\min}, T_{\max}]$, with $T_{\min} > 0$ and $T_{\max} < 1$.

We have performed some numerical simulations to investigate the impact of such imperfections in the seven-layer architecture. This can be simulated by evaluating the accessible values of T for the overall tunable beam-splitter operation for a given fabrication error ΔT_j, where $T_j = |t_j|^2$ is the transmittivity of the directional couplers. This is performed by sampling over values of the transmittivities T_j according to a Gaussian distribution centered around $T_j = 0.50$ (the optimal case), and with increasing uncertainty ΔT_i. In our case, we evaluated the set $\Delta T_j = \{0.00, 0.01, 0.02, 0.03, 0.04, 0.05\}$. To evaluate the transmittivity T of the overall tunable beam-splitter operation, we sampled 20 000 values from the Gaussian distribution (10 000 for each of the two different 50-50 directional couplers in the block of Fig. 4), and from those we obtained the value of T from the expression $T = |t_1 t_2 e^{i\theta} - r_1 r_2|^2$.

To implement this limitation into the training of our QONN, we restricted the possible values that the parameters θ_i^k can have according to such previous characterization. In particular, we obtained the restricted range of values by calculating $\cos^{-1}(\sqrt{T})$ for the minimum and maximum transmittivity T of the tunable beam-splitter operation estimated from the procedure above. Since we are restricting the values of θ_i^k, we decided to initialize the values of θ_i^k in a uniform distribution around $\pi/4$ so that the values of θ always start around the middle point, and they have more freedom to both increase and decrease in value. We then first trained the QONN with the restrictions caused by the uncertainty of the transmittivity of the directional couplers according to the maximum noise considered ($\Delta T_j = 0.05$). Then, we relax those restrictions to the ones caused by a lower uncertainty, and we let the network train with a low learning rate to perform an optimization for the new limits on θ_i^k. This choice of the procedure is performed to provide a meaningful comparison of the impact of the imperfections on the resulting cloner, in particular to minimize the potential oscillations on the results due to multiple trainings.

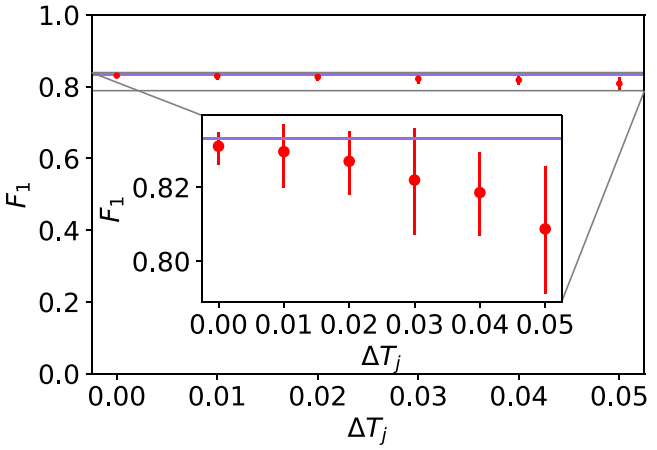


FIG. 5. Fidelity of the clones as a function of the fabrication error ΔT_j in the directional couplers. The red bars represent the range between the minimum and maximum fidelity for each QONN, while the purple line represents the theoretical maximum average fidelity.

Finally, after training each network with the corresponding restrictions due to imperfections, we evaluated the fidelity of each configuration as for the noiseless case. The results are shown in Fig. 5. As expected, decreasing the fabrication error ΔT_j leads to a better fidelity. However, even for a value of $\Delta T_j = 0.05$, well within the capabilities of current technology, we obtain only a modest reduction in the cloning fidelity ($F_1 \sim 0.8087$). A more detailed analysis is reported for $\Delta T_j = 0.05$, to be compared with the noiseless case, and it is shown in the fidelity heatmaps of Fig. 6 constructed as in Fig. 3.

A closer look at this behavior can be obtained by observing the distribution of the values of the parameters $T_i^k = \cos^2(\theta_i^k)$ for the different trained networks, corresponding to the different noise level ΔT_j (see Fig. 7). By comparing the obtained results, we find as expected that the distribution of the values T_i^k becomes progressively less concentrated over the extremal values around $T_i^k \sim 0$ and $T_i^k \sim 1$. However, the results on

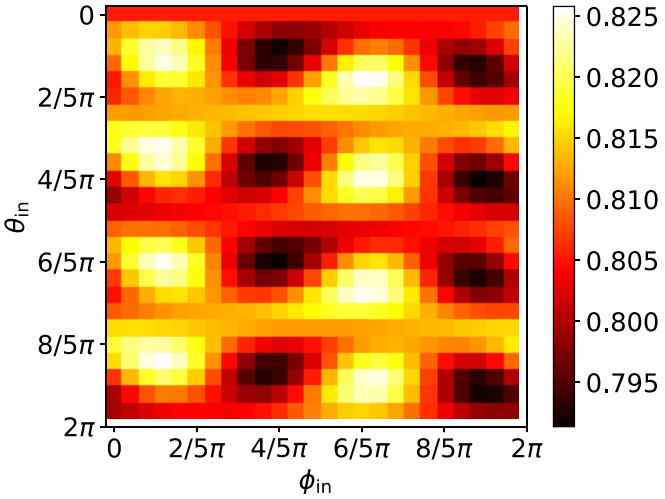


FIG. 6. Fidelity heatmap for the noisy QONN as a function of the input state parameters $(\theta_{\text{in}}, \phi_{\text{in}})$ for the case with $\Delta T_j = 0.05$.

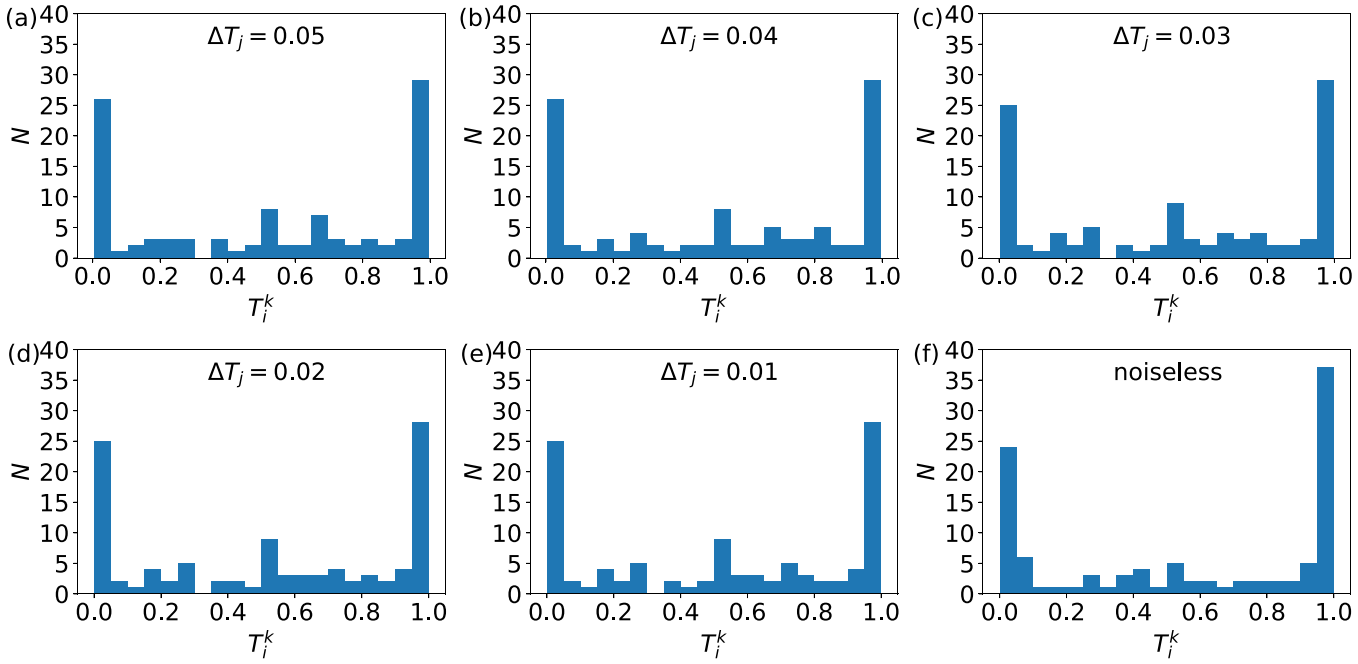


FIG. 7. Distribution of the transmittivity T_i^k of the beam-splitter operations, related to the network parameters θ_i^k, in the trained cloners for different values of the fabrication error ΔT_j.

the fidelity show that the system is capable of rearranging the configuration of the internal parameter during the training so as to minimize the impact on the output performances. Thus, all this analysis suggests that this approach presents a certain degree of robustness with respect to practical sources of noise, a crucial requirement towards future experimental implementations.

VI. CONCLUSIONS

In this work, we have shown the capabilities of quantum optical neural networks by training via numerical simulations a seven-layer network to obtain a near-optimal deterministic universal quantum cloner. The achieved average single-copy fidelity $\langle F_1 \rangle \sim 0.8269$ shows the possibility of reaching performances very close to the optimal value of F_1^{opt}. We have also shown that such an approach presents a certain degree of resilience to relevant noise sources towards future implementation. Indeed, we have quantified via numerical simulations how different levels of fabrication errors affect the performances of the trained networks. Even with significant levels of noise, well within the reach of current photonic technology, the QONN was capable of reaching high values of the operation fidelity, proving a significant resistance to noise. This result represents an essential feature for implementations of such an architecture based on NISQ technology.

Future investigations in this direction are expected, with a specific focus on a further in-depth analysis of the effects of various types of imperfections beyond those addressed in this work. These noise processes include photon losses [33], as well as imperfections in the nonlinear elements of the structure acting the counterpart of the activation nodes in classical neural networks. This latter analysis is also motivated by recent experimental advances in the implementation of nonlinear photon-photon gates [67–73]. Analysis on the effect of different types and different sizes of nonlinearities on the performance of the QONN would also be needed. There has already been a first insight on the use of Kerr nonlinearities with smaller strengths [29], which showed that a smaller strength leads to reduced performances for a fixed number of layers, but that can be at least partially compensated by using more layers in the circuit. This aspect is a relevant open challenge requiring additional investigations in the future.

Given their potential and flexibility, QONNs can find applications in different contexts, such as to perform classification tasks, to find more compressed and noise resistant versions of already existing quantum circuits, or to create small circuits that perform nontrivial operations. On a general note, their application can be foreseen as building blocks that can be combined in more complex architectures. An open challenge regards the scalability of this architecture. On the one hand, the scaling of the hardware itself will be related to the current development of photonic technologies, with the linear part of the network likely associated with current development of photonic integrated circuits. On the other hand, the scalability to larger sizes of the optimization algorithm will likely be related to the same challenges that are currently associated with all QML models. As such, scaling up to larger sizes will define whether this architecture can be used for medium-scale circuits with useful applications, as here described, or if it will be able to tackle even more complex tasks.

ACKNOWLEDGMENTS

This work is supported by the ERC Advanced Grant QUBOSS (QUantum advantage via non-linear BOSon Sampling, Grant Agreement No. 884676). D.S. acknowledges Thales Alenia Space Italia for supporting the Ph.D. fellowship.

[1] F. Flamini, N. Spagnolo, and F. Sciarrino, Photonic quantum information processing: A review, Rep. Prog. Phys. **82**, 016001 (2019).

[2] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[3] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, G. Yang, L. You, Z. Wang *et al.*, Quantum computational advantage using photons, Science **370**, 1460 (2020).

[4] Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong, C. Guo, C. Guo, S. Guo, L. Han, L. Hong, H.-L. Huang, Y.-H. Huo, L. Li, N. Li *et al.*, Strong Quantum Computational Advantage Using a Superconducting Quantum Processor, Phys. Rev. Lett. **127**, 180501 (2021).

[5] H.-S. Zhong, Y.-H. Deng, J. Qin, H. Wang, M.-C. Chen, L.-C. Peng, Y.-H. Luo, D. Wu, S.-Q. Gong, H. Su, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, G. Yang, L. You *et al.*, Phase-Programmable Gaussian Boson Sampling Using Stimulated Squeezed Light, Phys. Rev. Lett. **127**, 180502 (2021).

[6] L. S. Madsen, F. Laudenbach, M. Falamarzi. Askarani, F. Rortais, T. Vincent, J. F. F. Bulmer, F. M. Miatto, L. Neuhaus, L. G. Helt, M. J. Collins, A. E. Lita, T. Gerrits, S. W. Nam, V. D. Vaidya, M. Menotti, I. Dhand, Z. Vernon, N. Quesada, and J. Lavoie, Quantum computational advantage with a programmable photonic processor, Nature (London) **606**, 75 (2022).

[7] J. Preskill, Quantum computing in the nisq era and beyond, Quantum **2**, 79 (2018).

[8] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature (London) **521**, 436 (2015).

[9] K. Albertsson, P. Altoe, D. Anderson, J. Anderson, M. Andrews, J. P. A. Espinosa, A. Aurisano, L. Basara, A. Bevan, W. Bhimji, D. Bonacorsi, B. Burkle, P. Calafiura, M. Campanelli, L. Capps, F. Carminati, S. Carrazza, Y. fan Chen, T. Childers, Y. Coadou *et al.*, Machine learning in high energy physics community white paper, J. Phys.: Conf. Ser. **1085**, 022008 (2018).

[10] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. **42**, 60 (2017).

[11] C. Fefferman, S. Mitter, and H. Narayanan, Testing the manifold hypothesis, J. Am. Math. Soc. **29**, 983 (2016).

[12] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Netw. **61**, 85 (2015).

[13] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ, 2016), pp. 770–778.

[14] B. R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A. A. A. Sallab, S. Yogamani, and P. Pérez, Deep reinforcement learning for autonomous driving: A survey, IEEE Trans. Intell. Transp. Syst. **23**, 4909 (2022).

[15] S. Borah, B. Sarma, M. Kewming, G. J. Milburn, and J. Twamley, Measurement-Based Feedback Quantum Control with Deep Reinforcement Learning for a Double-Well Nonlinear Potential, Phys. Rev. Lett. **127**, 190403 (2021).

[16] M. Y. Niu, S. Boixo, V. N. Smelyanskiy, and H. Neven, Universal quantum control through deep reinforcement learning, npj Quantum Inf. **5**, 33 (2019).

[17] V. V. Sivak, A. Eickbusch, H. Liu, B. Royer, I. Tsioutsios, and M. H. Devoret, Model-Free Quantum Control with Reinforcement Learning, Phys. Rev. X **12**, 011059 (2022).

[18] V. Cimini, I. Gianani, N. Spagnolo, F. Leccese, F. Sciarrino, and M. Barbieri, Calibration of Quantum Sensors by Neural Networks, Phys. Rev. Lett. **123**, 230502 (2019).

[19] V. Cimini, E. Polino, M. Valeri, I. Gianani, N. Spagnolo, G. Corrielli, A. Crespi, R. Osellame, M. Barbieri, and F. Sciarrino, Calibration of Multiparameter Sensors Via Machine Learning at the Single-Photon Level, Phys. Rev. Appl. **15**, 044003 (2021).

[20] S. Xue, Y. Liu, Y. Wang, P. Zhu, C. Guo, and J. Wu, Variational quantum process tomography of unitaries, Phys. Rev. A **105**, 032427 (2022).

[21] G. Torlai, C. J. Wood, A. Acharya, G. Carleo, J. Carrasquilla, and L. Aolita, Quantum process tomography with unsupervised learning and tensor networks, arXiv:2006.02424 [quant-ph]

[22] M. Neugebauer, L. Fischer, A. Jäger, S. Czischek, S. Jochim, M. Weidemüller, and M. Gärttner, Neural-network quantum state tomography in a two-qubit experiment, Phys. Rev. A **102**, 042604 (2020).

[23] A. M. Palmieri, E. Kovlakov, F. Bianchi, D. Yudin, S. Straupe, J. D. Biamonte, and S. Kulik, Experimental neural network enhanced quantum tomography, npj Quantum Inf. **6**, 20 (2020).

[24] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Neural-network quantum state tomography, Nat. Phys. **14**, 447 (2018).

[25] Y. S. Teo, S. Shin, H. Jeong, Y. Kim, Y.-H. Kim, G. I. Struchalin, E. V. Kovlakov, S. S. Straupe, S. P. Kulik, G. Leuchs, and L. L. Sánchez-Soto, Benchmarking quantum tomography completeness and fidelity with machine learning, New J. Phys. **23**, 103021 (2021).

[26] M. Schuld, I. Sinayskiy, and F. Petruccione, An introduction to quantum machine learning, Contemp. Phys. **56**, 172 (2015).

[27] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature (London) **549**, 195 (2017).

[28] J. Adcock, E. Allen, M. Day, S. Frick, J. Hinchliff, M. Johnson, S. Morley-Short, S. Pallister, A. Price, and S. Stanisic, Advances in quantum machine learning, arXiv:1512.02900 [quant-ph].

[29] G. R. Steinbrecher, J. P. Olson, D. Englund, and J. Carolan, Quantum optical neural networks, npj Quantum Inf. **5**, 60 (2019).

[30] B. Bartlett and S. Fan, Universal programmable photonic architecture for quantum information processing, Phys. Rev. A **101**, 042319 (2020).

[31] X.-D. Cai, D. Wu, Z.-E. Su, M.-C. Chen, X.-L. Wang, L. Li, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, Entanglement-Based Machine Learning on a Quantum Computer, Phys. Rev. Lett. **114**, 110504 (2015).

[32] A. Gianelle, P. Koppenburg, D. Lucchesi, D. Nicotra, E. Rodrigues, L. Sestini, J. de Vries, and D. Zuliani, Quantum machine learning for b-jet charge identification, J. High Energy Phys. 08 (2022) 014.

[33] J. Ewaniuk, J. Carolan, B. J. Shastri, and N. Rotenberg, Imperfect quantum photonic neural networks, Adv. Quantum Technol. **2023**, 2200125 (2023).

[34] V. Bužek and M. Hillery, Quantum copying: Beyond the no-cloning theorem, Phys. Rev. A **54**, 1844 (1996).

[35] D. Bruß, D. P. DiVincenzo, A. Ekert, C. A. Fuchs, C. Macchiavello, and J. A. Smolin, Optimal universal and state-dependent quantum cloning, Phys. Rev. A **57**, 2368 (1998).

[36] D. Bruss, A. Ekert, and C. Macchiavello, Optimal Universal Quantum Cloning and State Estimation, Phys. Rev. Lett. **81**, 2598 (1998).

[37] V. Scarani, S. Iblisdir, N. Gisin, and A. Acín, Quantum cloning, Rev. Mod. Phys. **77**, 1225 (2005).

[38] N. J. Cert, S. Iblisdir, and G. V. Assche, Cloning and cryptography with quantum continuous variables, Eur. Phys. J. D **18**, 211 (2002).

[39] K. Bartkiewicz, K. Lemr, A. Cernoch, J. Soubusta, and A. Miranowicz, Experimental Eavesdropping Based on Optimal Quantum Cloning, Phys. Rev. Lett. **110**, 173601 (2013).

[40] H. Fan, Y.-N. Wang, L. Jing, J.-D. Yue, H.-D. Shi, Y.-L. Zhang, and L.-Z. Mu, Quantum cloning machines and the applications, Phys. Rep. **544**, 241 (2014).

[41] F. Bouchard, R. Fickler, R. W. Boyd, and E. Karimi, High-dimensional quantum cloning and applications to quantum hacking, Sci. Adv. **3**, e1601915 (2017).

[42] L. Ye, X.-k. Song, J. Yang, Q. Yang, and Y.-C. Ma, Optical implementation of the optimal universal and phase-covariant quantum cloning machines, Mod. Phys. Lett. B **26**, 1250102 (2012).

[43] J. Y. Haw, J. Zhao, J. Dias, S. M. Assad, M. Bradshaw, R. Blandino, T. Symul, T. C. Ralph, and P. K. Lam, Surpassing the no-cloning limit with a heralded hybrid linear amplifier for coherent states, Nat. Commun. **7**, 13222 (2016).

[44] F. De Martini, D. Pelliccia, and F. Sciarrino, Contextual, Optimal, and Universal Realization of the Quantum Cloning Machine and of the NOT Gate, Phys. Rev. Lett. **92**, 067901 (2004).

[45] M. Ricci, F. Sciarrino, C. Sias, and F. De Martini, Teleportation Scheme Implementing the Universal Optimal Quantum Cloning Machine and the Universal NOT Gate, Phys. Rev. Lett. **92**, 047901 (2004).

[46] A. Lamas-Linares, C. Simon, J. C. Howell, and D. Bouwmeester, Experimental quantum cloning of single photons, Science **296**, 712 (2002).

[47] E. Nagali, L. Sansoni, F. Sciarrino, F. De Martini, L. Marrucci, B. Piccirillo, E. Karimi, and E. Santamato, Optimal quantum cloning of orbital angular momentum photon qubits through hong-ou-mandel coalescence, Nat. Photon. **3**, 720 (2009).

[48] E. Nagali, D. Giovannini, L. Marrucci, S. Slussarenko, E. Santamato, and F. Sciarrino, Experimental Optimal Cloning of Four-Dimensional Quantum States of Photons, Phys. Rev. Lett. **105**, 073602 (2010).

[49] B. Coyle, M. Doosti, E. Kashefi, and N. Kumar, Variational quantum cloning: Improving practicality for quantum cryptanalysis, Phys. Rev. A **105**, 042604 (2022).

[50] W. K. Wooters and W. H. Zurek, A single quantum cannot be cloned, Nature (London) **299**, 802 (1982).

[51] N. Gisin and S. Massar, Optimal Quantum Cloning Machines, Phys. Rev. Lett. **79**, 2153 (1997).

[52] T. Durt and J. Du, Characterization of low-cost one-to-two qubit cloning, Phys. Rev. A **69**, 062316 (2004).

[53] L. Ye, L. Jin, and W.-H. Zhang, Optimal $1 \rightarrow 3$ economical phase-covariant quantum cloning with superconducting quantum interference devices in a cavity, J. Mod. Opt. **56**, 661 (2009).

[54] N. J. Cerf and J. Fiurasek, *Progress in Optics* (Elsevier, Amsterdam, 2006), p. 49.

[55] C. Simon, G. Weihs, and A. Zeilinger, Optimal Quantum Cloning via Stimulated Emission, Phys. Rev. Lett. **84**, 2993 (2000).

[56] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, An optimal design for universal multiport interferometers, Optica **3**, 1460 (2016).

[57] G. Cybenko, Approximation by superpositions of a sigmoidal function, Math. Control Signals Syst. **2**, 303 (1989).

[58] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, Neural Netw. **6**, 861 (1993).

[59] Y. Yao, L. Rosasco, and A. Caponnetto, On early stopping in gradient descent learning, Constr. Approx. **26**, 289 (2007).

[60] C. Adami and N. J. Cerf, Quantum computation with linear optics, in *Quantum Computing and Quantum Communications*, edited by C. P. Williams (Springer, Berlin, 1999), pp. 391–401.

[61] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).

[62] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, M. S. Alam, S. Ahmed, J. M. Arrazola, C. Blank, A. Delgado, S. Jahangiri, K. McKiernan, J. J. Meyer, Z. Niu, A. Száva, and N. Killoran, Pennylane: Automatic differentiation of hybrid quantum-classical computations, arXiv:1811.04968 [quant-ph].

[63] N. Killoran, J. Izaac, N. Quesada, V. Bergholm, M. Amy, and C. Weedbrook, Strawberry Fields: A software platform for photonic quantum computing, Quantum **3**, 129 (2019).

[64] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, *et al.*, Tensorflow: A system for large-scale machine learning, in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)* (The USENIX Association, 2016), pp. 265–283.

[65] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, in *3rd International Conference on Learning Representations, ICLR 2015 San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings*, edited by Y. Bengio and Y. LeCun (ICLR, 2015).

[66] J. Wang, F. Sciarrino, A. Laing, and M. G. Thompson, Integrated photonic quantum technologies, Nat. Photon. **14**, 273 (2020).

[67] J. Volz, M. Scheucher, C. Junge, and A. Rauschenbeutel, Nonlinear π phase shift for single fibre-guided photons interacting with a single resonator-enhanced atom, Nat. Photon. **8**, 965 (2014).

[68] A. Feizpour, M. Hallaji, G. Dmochowski, and A. M. Steinberg, Observation of the nonlinear phase shift due to single post-selected photons, Nat. Phys. **11**, 905 (2015).

[69] D. Tiarks, S. Schmidt, G. Rempe, and S. Dürr, Optical π phase shift created with a single-photon pulse, Sci. Adv. **2**, e1600036 (2016).

[70] B. Hacker, S. Welte, G. Rempe, and S. Ritter, A photon-photon quantum gate based on a single atom in an optical resonator, Nature (London) **536**, 193 (2016).

[71] T. Stolz, H. Hegels, M. Winter, B. Röhr, Y.-F. Hsiao, L. Husel, G. Rempe, and S. Dürr, Quantum-Logic Gate Between Two Optical Photons with an Average Efficiency Above 40%, Phys. Rev. X **12**, 021035 (2022).

[72] T. Kuriakose, P. M. Walker, T. Dowling, O. Kyriienko, I. A. Shelykh, P. St-Jean, N. C. Zambon, A. Lemaitre, I. Sagnes, L. Legratiet, A. Harouri, S. Ravets, M. S. Skolnick, A. Amo, J. Bloch, and D. Krizhanovskii, Few-photon all-optical phase rotation in a quantum-well micropillar cavity, Nat. Photon. **16**, 566 (2022).

[73] L. De Santis, C. Anton, B. Reznychenko, N. Somaschi, G. Coppola, J. Senellart, C. Gomez, A. Lemaitre, I. Sagnes, A. G. White, L. Lanco, A. Auffeves, and P. Senellart, A solid-state single-photon filter, Nat. Nanotechnol. **12**, 663 (2017).