

PAPER • OPEN ACCESS

Benchmarking quantum tomography completeness and fidelity with machine learning

To cite this article: Yong Siah Teo *et al* 2021 *New J. Phys.* **23** 103021

View the [article online](#) for updates and enhancements.

You may also like

- [Practical Bayesian tomography](#)
Christopher Granade, Joshua Combes and D G Cory
- [Focus on quantum tomography](#)
K Banaszek, M Cramer and D Gross
- [Optimal evolution models for quantum tomography](#)
Artur Czerwinski



PAPER

Benchmarking quantum tomography completeness and fidelity with machine learning

OPEN ACCESS

RECEIVED
27 May 2021REVISED
9 August 2021ACCEPTED FOR PUBLICATION
20 August 2021PUBLISHED
14 October 2021

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the
title of the work, journal
citation and DOI.



Yong Siah Teo¹, Seongwook Shin¹, Hyunseok Jeong¹, Yosep Kim², Yoon-Ho Kim²,
Gleb I Struchalin³, Egor V Kovlakov³, Stanislav S Straupe³, Sergei P Kulik³,
Gerd Leuchs^{4,5} and Luis L Sánchez-Soto^{4,6,*}

¹ Department of Physics and Astronomy, Seoul National University, 08826 Seoul, Republic of Korea

² Department of Physics, Pohang University of Science and Technology (POSTECH), 37673 Pohang, Republic of Korea

³ Quantum Technology Centre, and Faculty of Physics, Moscow State University, 119991 Moscow, Russia

⁴ Max-Planck-Institut für die Physik des Lichts, Staudtstraße 2, 91058 Erlangen, Germany

⁵ Institute of Applied Physics, Russian Academy of Sciences, 603950 Nizhny Novgorod, Russia

⁶ Departamento de Óptica, Facultad de Física, Universidad Complutense, 28040 Madrid, Spain

* Author to whom any correspondence should be addressed.

E-mail: yong.siah.teo@gmail.com, h.jeong37@gmail.com, yoohnho72@gmail.com and lsanchez@fis.ucm.es

Keywords: quantum tomography, convolutional networks, compressive

Supplementary material for this article is available [online](#)

Abstract

We train convolutional neural networks to predict whether or not a set of measurements is informationally complete to uniquely reconstruct any given quantum state with no prior information. In addition, we perform fidelity benchmarking based on this measurement set without explicitly carrying out state tomography. The networks are trained to recognize the fidelity and a reliable measure for informational completeness. By gradually accumulating measurements and data, these trained convolutional networks can efficiently establish a compressive quantum-state characterization scheme by accelerating runtime computation and greatly reducing systematic drifts in experiments. We confirm the potential of this machine-learning approach by presenting experimental results for both spatial-mode and multiphoton systems of large dimensions. These predictions are further shown to improve when the networks are trained with additional bootstrapped training sets from real experimental data. Using a realistic beam-profile displacement error model for Hermite–Gaussian sources, we further demonstrate numerically that the orders-of-magnitude reduction in certification time with trained networks greatly increases the computation yield of a large-scale quantum processor using these sources, before state fidelity deteriorates significantly.

1. Introduction

Recent advances in quantum algorithms and error correction [1–6] have fueled the development of noisy intermediate-scale quantum computing devices. This progress requires an efficient assessment of the relevant quantum systems [6–8], gates [9–13] and measurements [14–21]. Toolkits developed in quantum tomography [22–33] have concomitantly evolved into modern schemes appropriate for characterizing those components efficiently. A notable branch of schemes attempt to cope with a large number of qubits by directly estimating quantum properties [34–42].

As typical quantum tasks involve pure states, unitary gates, and projective measurements, there also exists a series of compressed-sensing-related proposals [43–51] that fully reconstruct low-rank quantum components with few measurement resources. However, they rely on prior knowledge about the rank, which often turns out to be unreliable in practice because of noise. Very recently, compressive-tomography methods without assuming any prior information has been developed and applied to the individual low-resource characterization of quantum states, processes and measurements [52–56]. A crucial ingredient

in these methods is informational completeness certification (ICC) that determines whether or not a given measurement set and its corresponding data is informationally complete (IC). This is done by computing a uniqueness measure based on the given measurements. Such a computation can be performed with classical semidefinite programs (SDPs) [57] of (worst-case) polynomial-time complexities.

Like any tomography scheme that invokes rounds of optimization routines, an accumulation of errors can occur in real experiments while running SDPs on-the-fly. As a practically feasible solution, we propose to train an artificial neural network to verify the IC property for a set of quantum measurements and corresponding raw data. We further introduce an auxiliary network to be used concurrently for us to validate the fidelity of the unknown state for the given measurement set without carrying out explicit reconstruction. Once a set of IC measurement data is collected, it takes only one final round of state reconstruction to obtain the unique physical estimator, if so desired.

Network training can be done offline using simulated noisy datasets, and the stored network model can later be retrieved and used in real experiments with statistical noise. More specifically, a convolutional neural-network (CNN) architecture shall be used for training and prediction. Among other kinds of networks that have been widely adopted by the quantum-information community [58–63], this is a popular network architecture that is used in image-pattern recognition [64–67], with boosted support by a recent universality proof [68] that such networks can indeed forecast any continuous function mapping. Both its classical application and quantum analog have also gained traction in quantum-information science [69–72].

In this work, we train an *informational completeness certification net* (ICCNNet) and a *fidelity prediction net* (FidNet), each made up of a sequence of convolution and pooling neural layers that is reasonably deep. Partnered with FidNet for direct fidelity benchmarking without the need for state tomography, ICCNet constitutes the foundational core for deciding if the given measurement resources are sufficient to uniquely characterize any unknown state in real experimental situations. Neural-network (NN) training is versatile in the sense that noise models may be incorporated into the training procedure to improve the predictive power of the networks. After offline training, the network models can heavily reduce the computation time of the uniqueness certification by orders of magnitude for large dimensions while running the experiments. This essentially realizes a compressive tomography scheme that is drift-proof, comprising a highly efficient uniqueness certification and fidelity-benchmarking protocols.

Apart from Monte Carlo simulations, we also use real data obtained from two separate groups of experiments to demonstrate that the resulting trained network models can predict the average behaviors of both the IC property and fidelity very well, despite the presence of errors and experimental imperfections. We also show that performances in predicting both properties can be further boosted when the neural networks are trained with additional bootstrapped experimental datasets. Finally, simulations on a time-dependent error model relevant to Hermite–Gaussian sources are performed as an example to illustrate the effectiveness of our NN certification tools in suppressing systematic drifts during quantum computation.

2. Background

2.1. Ascertaining informational completeness

The main procedure for certifying whether a generic measurement dataset is sufficient to unambiguously determine an unknown quantum state can be represented as a simple flowchart in figure 1(a). Given a positive operator-valued measure (POVM) that models the measurements performed, the corresponding data counts are noisy due to statistical fluctuation arising from finite data samples. Proper data analysis first entails the extraction of physical probabilities from the accumulated data, which can be done with well-established statistical methods, such as those of maximum likelihood (ML) [23, 24, 73–75] and least squares (LS) [76, 77], subject to the physical constraint of density matrices (refer also to section 2.1).

After obtaining the physical probabilities, one may proceed to evaluate the measurements and find out whether they are IC. More specifically, a uniqueness indicator $0 \leq s_{CVX} \leq 1$ can be directly computed from the POVM and data with the help of SDPs—the ICC. When $s_{CVX} > 0$, there is equivalently a convex set of state estimators that are consistent with the physical probabilities. It can be shown [52] that a unique estimator is obtained from the measured POVM and corresponding data if and only if $s_{CVX} = 0$.

Bottom-up resource-efficient quantum-state tomography is thus an iterative program involving rounds of extracting physical probabilities from the measurement data and certifying uniqueness based on these probabilities. At each round, the computed s_{CVX} is used to decide whether new measurements are needed in the next one. In this manner, the POVM outcomes may be accumulated bottom-up until $s_{CVX} = 0$, after which a physical state reconstruction using either the ML or LS scheme is carried out to obtain the unique

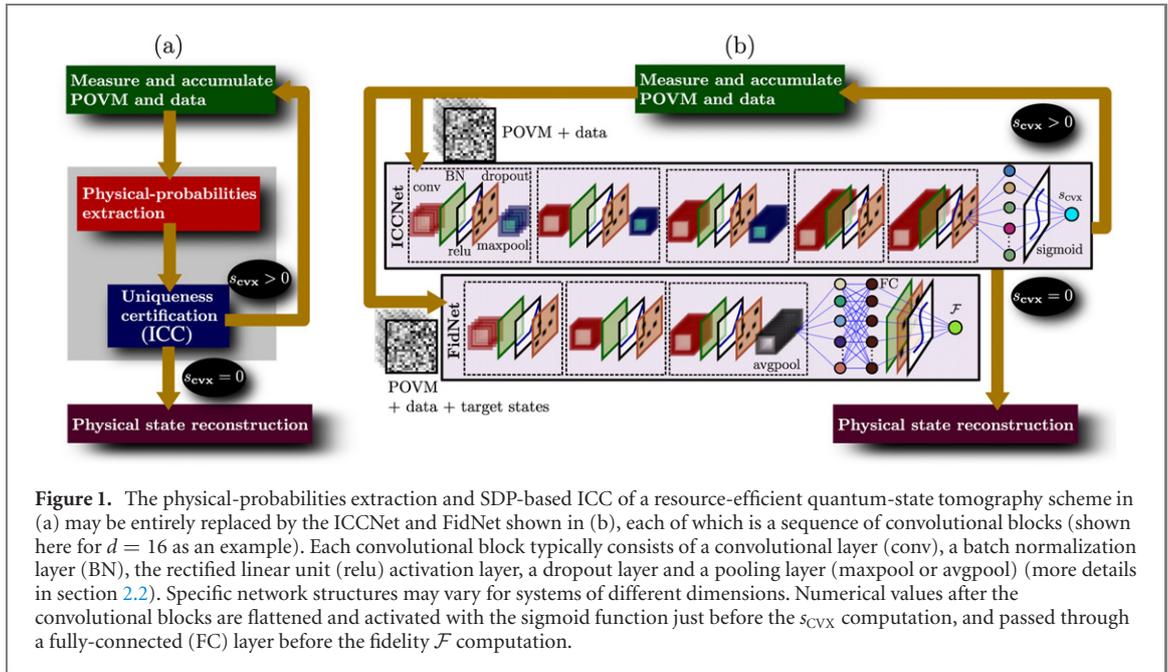


Figure 1. The physical-probabilities extraction and SDP-based ICC of a resource-efficient quantum-state tomography scheme in (a) may be entirely replaced by the ICCNet and FidNet shown in (b), each of which is a sequence of convolutional blocks (shown here for $d = 16$ as an example). Each convolutional block typically consists of a convolutional layer (conv), a batch normalization layer (BN), the rectified linear unit (relu) activation layer, a dropout layer and a pooling layer (maxpool or avgpool) (more details in section 2.2). Specific network structures may vary for systems of different dimensions. Numerical values after the convolutional blocks are flattened and activated with the sigmoid function just before the s_{cvx} computation, and passed through a fully-connected (FC) layer before the fidelity \mathcal{F} computation.

estimator. The size of the resulting IC POVM is minimized accordingly. We remark that ICC turns out to be the limiting procedure in practical implementation relative to a typical quantum-state reconstruction. This is because an estimation over the space of quantum states can be very efficiently implemented with an iterative scheme, where each step involves a regular gradient update and just one round of convex projection [75]. (The case for quantum processes has also been discussed [78].) On the other hand, satisfying both Born's rule and quantum positivity constraint in ICC requires a separate iterative procedure just to carry out the correct convex projection onto their intersection [79]. To date, there exists no efficient way to perform projections of these constraints to the authors' knowledge.

By recalling the results in references [52, 53], we briefly describe the simple procedures that deterministically verify whether a set of POVM outcomes $\{\Pi_j \geq 0\}$ is IC given their corresponding set of relative frequency data $\{\nu_j\}$. The first necessary step is to acquire the physical probabilities from ν_j . To this end, we consider two popular choices often considered in quantum tomography, namely the ML and LS methods. In ML, we maximize the log-likelihood function $\log L$ that best describes the physical scenario over the quantum state space. Since we predominantly discuss von Neumann measurement bases, each basis induces a multinomial distribution such that we have the form $\log L \propto \sum_j \nu_j \log p'_j$, where $p'_j = \text{tr}\{\rho' \Pi_j\}$ are our sought-after physical probabilities to be optimized over the operator space in which $\rho' \geq 0$ and $\text{tr}\{\rho'\} = 1$. In LS, which we have adopted to deal with arbitrary projective measurements that do not sum to the identity operator in general, the distance $\mathcal{D} = \|\nu_j - p'_j\|^2$ is minimized with respect to p'_j over the space of $\rho' \geq 0$, this time with the unit-trace constraint relaxed and later reinstated after the minimization is completed.

Using the obtained physical probability estimators \hat{p}_j through the aforementioned optimization strategies, we can now define and fix a randomly generated full-rank state Z and conduct the following two SDPs:

$$\begin{aligned} & \text{minimize/maximize } f = \text{tr}\{\rho' Z\} \text{ over } \rho' \\ & \text{subject to:} \\ & \rho' \geq 0, \quad \text{tr}\{\rho'\} = 1, \quad \text{tr}\{\rho' \Pi_j\} = \hat{p}_j. \end{aligned} \quad (2.1)$$

It is now clear why the SDPs are to be carried out with the physical probabilities \hat{p}_j instead of raw data ν_j : the relative frequencies ν_j are statistically noisy and in general do not correspond to a feasible solution set in (2.1). It has been shown that when $s_{\text{cvx}} \equiv f_{\text{max}} - f_{\text{min}}$ is zero, this implies that any quantum-state estimator reconstructed from $\{\Pi_j\}$ and $\{\nu_j\}$ is unique and equal to the solution for (2.1).

2.2. Training the ICCNet and FidNet

We propose to tackle the combined problem of physical probabilities extraction and uniqueness certification by predicting with trained neural networks. We also demonstrate the possibility of performing fidelity evaluation on the reconstruction with such neural networks without explicitly carrying out physical

state tomography. To do this, we introduce the ICCNet and FidNet, illustrated in figure 1(b), where each possesses a convolutional network architecture that analyzes the given POVM and data by regarding them as images. Such a treatment allows one to train the networks with far less trainable parameters to recognize s_{CVX} and fidelity $\mathcal{F}(\hat{\rho}, \rho_{\text{targ}}) = \text{tr} \left\{ \sqrt{\sqrt{\hat{\rho}} \rho_{\text{targ}} \sqrt{\hat{\rho}}} \right\}^2$ between the state estimator $\hat{\rho}$ and some target state ρ_{targ} as compared to using, for instance, the multilayer perceptron (feed-forward) architecture [80, 81] that consists only of FC or dense layers.

The purpose of FidNet is to assess the quality of the reconstruction after each measurement set is made. Before the point of informational completeness, the reconstructed state $\hat{\rho}$ is not unique by definition. Throughout this article, for consistency, $\hat{\rho}$ shall always be taken as the ML estimator that minimizes the linear function in (2.1). This is only a particular choice used to define FidNet that is chosen as a standard. One may end up with a slightly more conservative FidNet by setting $\hat{\rho}$ to be the minimum-fidelity estimator with respect to the SDP constraints stated in the last line of (2.1). This would require another round of fidelity minimization in every step of the training-data-generation phase.

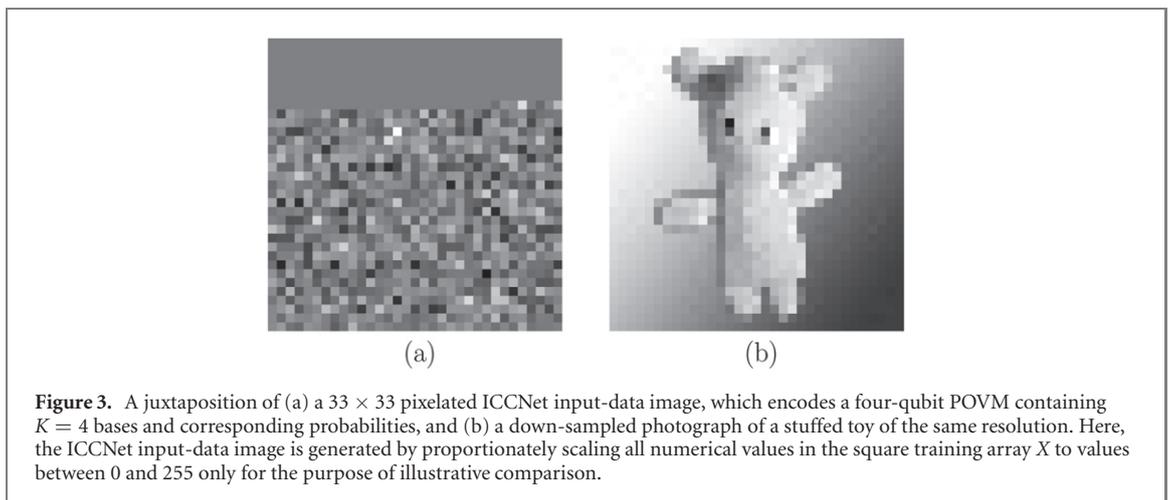
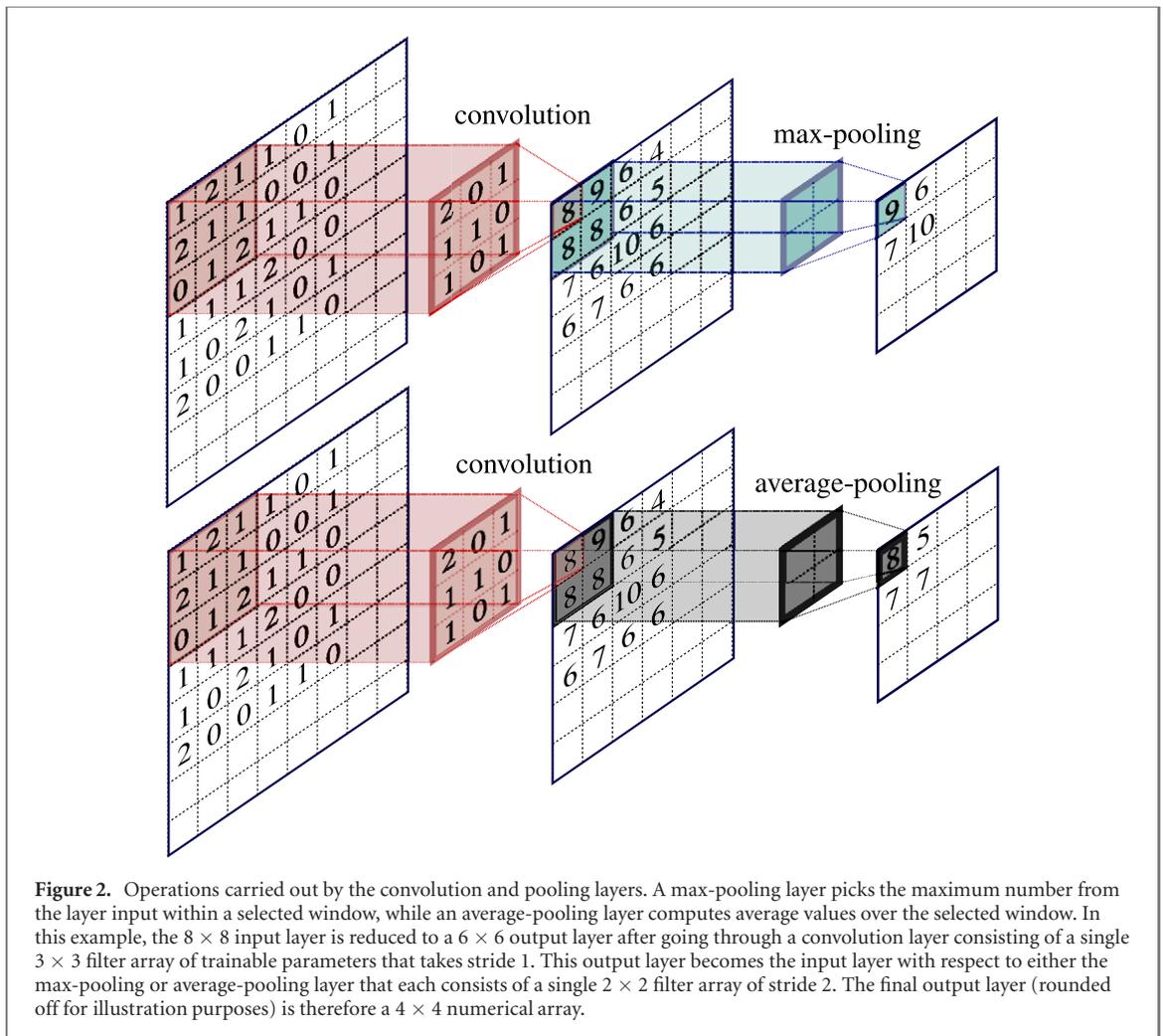
For predicting s_{CVX} and \mathcal{F} , both ICCNet and FidNet employ a sequence of two-dimensional array manipulating layers. Two important types of layers responsible for these operations are the convolution layer, which are two-dimensional filters that carry out multiplicative convolutions with the layer input numerical array, and the pooling layer that generally down-samples a layer input array into a smaller output array with a simple numerical-summarizing computation. To each convolution layer, an activation function is applied to further introduce nonlinear characteristics for predicting general network output functions.

The convolutional ICCNet and FidNet take on a similar architecture, which consists of convolution, max-pooling and average-pooling layers. Each convolution layer consists of n_f filters, where each filter is a 3×3 array window that slides vertically and across layer input arrays with stride 1 in both directions. We design the sequence of convolution layers to have an exponentially increasing n_f with the network depth. These pooling layers are generally responsible for shrinking the layer input array to a smaller layer output array. The actions of all types of layers are summarized in figure 2. We insert the default ‘relu’ activation function after every convolution layer, which is defined as $f_{\text{relu}}(x) = \max(0, x)$. At the end of ICCNet and FidNet, the respective output values are computed with the sigmoid activation function given by $f_{\text{sigmoid}}(x) = 1/(1 + e^{-x})$.

Overfitting can be an issue in machine learning, in which case the neural networks are prone to fitting training datasets much better than unseen ones. It is therefore essential to regulate network training by keeping the problem of overfitting in check so that the resulting trained models have high predictive power. This problem often arises when the neural network is deep. The addition of dropout layers has been proven to be an effective method for combating overfitting [82–84], which randomly exclude trainable parameters. More recently, it has been demonstrated that NN training can be further enhanced by adding BN layers. This was supported not only by the initial observation that the distribution of layer input values are stabilized with BN [85], but also by the even more relevant finding that the gradient landscape of the network loss function (the figure of merit quantifying the difference between the actual output value and that computed by the network) seen by the optimization routine that trains the network becomes smoother [86], making training much more stable.

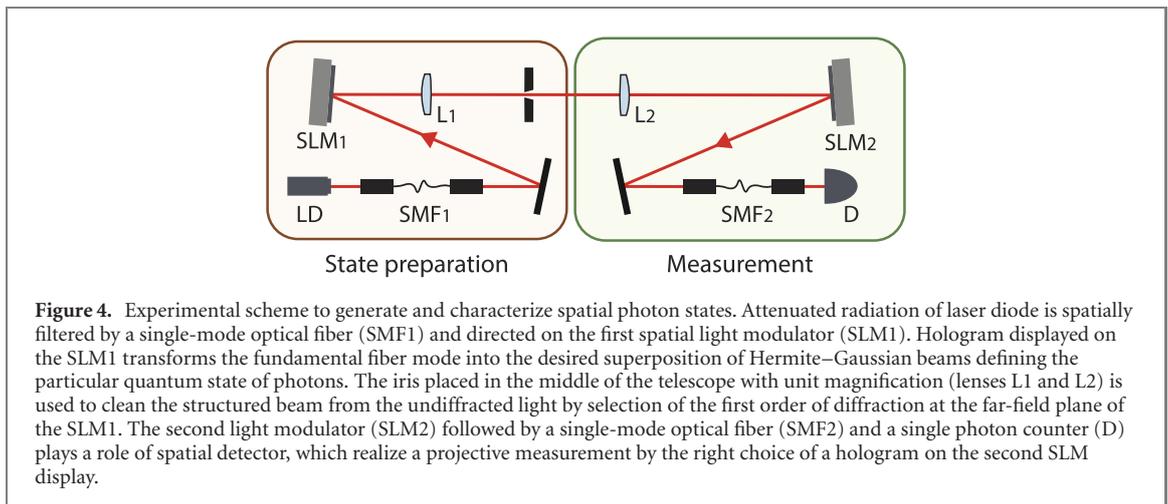
All trainable parameters in the relevant neural layers of ICCNet and FidNet are optimized using a variant of stochastic gradient descent known as NAdam [87], where the network gradients are computed in batches of the training data. To prepare ICCNet training input datasets, for von Neumann measurements of a fixed number (K) of bases considered in sections 4.1 and 4.2, the initial network input X is an $m \times K(d^2 + d)$ matrix that contains m training datasets, each recording the K measured bases and corresponding relative frequencies $\{\nu_{jk}\}_{j=0}^{d-1} \text{ }_{k=1}^K$ ($\sum_j \nu_{jk} = 1$). To encode the measurement bases, we regard all bases as some unitary rotation $U_k |j\rangle \langle j| U_k^\dagger$ of the standard computational basis $\{|j\rangle\}_{j=0}^{d-1}$, where $U_1 = 1$. These unitary operators are then logarithmized in order to obtain their Hermitian exponents $H_k = -i \log U_k$ ($H_1 = 0$), from which the diagonals and upper triangular real and imaginary matrix elements are extracted. Each row of X is thus a flattened $K(d^2 + d)$ -dimensional row of real numerical values formatted properly to encode $U_1, U_2, \dots, U_K, \nu_{01}, \dots, \nu_{d-1,1}, \nu_{02}, \dots, \nu_{d-1,2}, \dots, \nu_{0K}, \dots, \nu_{d-1K}$ in this order. This input matrix is then processed into a $\lceil \sqrt{K(d^2 + d)} \rceil \times \lceil \sqrt{K(d^2 + d)} \rceil$ square training array of elements, which is then fed to ICCNet (see figure 3). Zeros are padded to this array in order to complete the square. Similarly, for a fixed set of L projective measurements discussed in section 4.3, analogous arguments lead to the necessary $\lceil \sqrt{L(d^2 + 1)} \rceil \times \lceil \sqrt{L(d^2 + 1)} \rceil$ input square training array. For each dimension, the randomly generated full-rank state Z needed to solve (2.1) is fixed during the training and testing stages.

On the other hand, training the FidNet requires input information about not only the measured bases (or projectors) and their corresponding data, but also the additional m target states to be included



as inputs, one for each dataset. The correct dimensions of the training arrays are $\lceil \sqrt{(K+1)d^2 + Kd} \rceil \times \lceil \sqrt{(K+1)d^2 + Kd} \rceil$ or $\lceil \sqrt{(L+1)d^2 + L} \rceil \times \lceil \sqrt{(L+1)d^2 + L} \rceil$ respectively for basis and projective measurements. We note that to predict fidelities for simulated test datasets of $d = 16, 32$ and 64 as shown in figures 7 and 8,

For all purely-simulation figures, FidNet training is done with target states defined by the true states that generated the simulated training datasets. On the other hand, for all experimental results in figure 99, FidNet training is carried out simultaneously with the target states derived from the corresponding true states and those that deviate from them in order to account for systematic errors more effectively and improve average prediction accuracy. The list of hyperparameters that define the architectures of ICCNet



and FidNet, as well as the technical analyses of network input-data generation and network training are given in sections II and III of supplementary material (<https://stacks.iop.org/NJP/23/103021/mmedia>).

Once an IC set of measurements are performed and assessed with ICCNet and FidNet, the density matrix representing the final state estimator may be obtained using the accelerated projected-gradient algorithm developed in [75]. Alternatively, it is possible to append our networks with additional conditional generative networks to yield the density matrix [62, 63].

3. Experiments

3.1. Spatial-mode photonic systems

Apart from evaluating simulation test datasets, we also run the trained ICCNet and FidNet models to benchmark real experimental datasets. In the first group of experiments, we showcase the accuracy of ICCNet and FidNet predictions on experimental data acquired from an attenuated laser source prepared in quantum states projected onto Hermite–Gaussian spatial-mode bases of various dimensions d . With this group of experiments, for the sake of variety, we shall consider measurement bases that are obtained from adaptive compressive tomography (ACT). These are eigenbases of the state that minimizes the von Neumann entropy subject to the same SDP constraints in (2.1). It has been demonstrated that successive measurements of such eigenbases result in a fast convergence of s_{CVX} [52, 53]. An explicit protocol to construct these bases is given in section I of supplementary material.

The Hilbert space of photonic spatial degrees of freedom is typically discretized using an appropriate basis of transverse modes. To produce high-dimensional quantum states we attenuate an 808-nm diode laser, filter the resulting radiation with a single-mode optical fiber and then adjust the spatial structure of the light field with a spatial light modulator (SLM, see figure 4). The holographic approach [88] allows us to transform the incident light into arbitrary transverse modes by controlling the phase pattern on the SLM's display.

We work with Hermite–Gaussian (HG) modes $\text{HG}_{nm}(x, y)$, which are the solutions to the Helmholtz equation in Cartesian coordinates (x, y) and form a complete orthonormal basis. By bounding the sum of beam orders $n + m$, we restrict the dimension of the generated quantum systems. Since holograms displayed on the SLM make use of a blazed grating, in order to select the first diffraction order, we place an iris in the middle of the telescope, where different diffraction orders are well separated. Using a second SLM, a single-mode optical fiber, followed by a single photon counting module, we realize a well-known technique of projective measurements in the spatial-mode space [89]. These allows us to also conveniently implement general ACT basis measurements in arbitrary dimensions.

3.2. Multiphoton systems

In the second group of experiments, we switch to a different flavor of informational completeness by discussing two-mode photon-number states. In particular, we look at quantum states of up to three photons occupying two optical modes. Such three-photon states were of interest in the study of high-order quantum polarization properties beyond the Stokes vectors [90]. The resulting Hilbert space is effectively four-dimensional and spanned by the set $\{|n_{\text{H}}, n_{\text{V}}\}_{n_{\text{H}}+n_{\text{V}}=3} = \{|0, 3\rangle, |1, 2\rangle, |2, 1\rangle, |3, 0\rangle\}$. Here n_{H} and n_{V} denote the number of photons in the horizontal and vertical polarization modes, respectively.

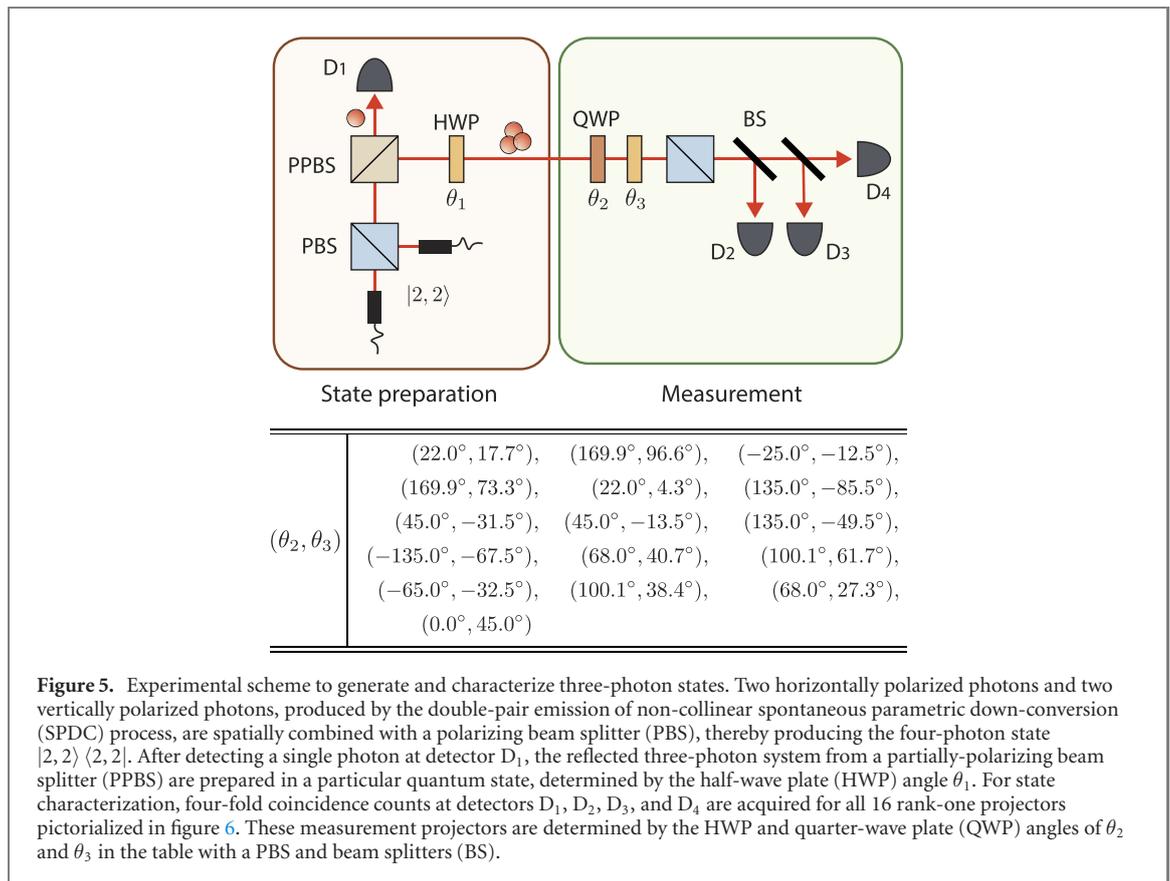


Figure 5. Experimental scheme to generate and characterize three-photon states. Two horizontally polarized photons and two vertically polarized photons, produced by the double-pair emission of non-collinear spontaneous parametric down-conversion (SPDC) process, are spatially combined with a polarizing beam splitter (PBS), thereby producing the four-photon state $|2, 2\rangle$ $\langle 2, 2|$. After detecting a single photon at detector D_1 , the reflected three-photon system from a partially-polarizing beam splitter (PPBS) are prepared in a particular quantum state, determined by the half-wave plate (HWP) angle θ_1 . For state characterization, four-fold coincidence counts at detectors D_1 , D_2 , D_3 , and D_4 are acquired for all 16 rank-one projectors pictorialized in figure 6. These measurement projectors are determined by the HWP and quarter-wave plate (QWP) angles of θ_2 and θ_3 in the table with a PBS and beam splitters (BS).

To perform tomography on the multiphoton quantum states, expectation values of a set of 16 rank-one projectors are measured. In principle, any set of 16 linearly independent projectors are suitable for a complete characterization of arbitrary four-dimensional states *without* ICC. For these experiments, we define each projector Π_j by a ket $b_j^{\dagger 3} |0, 0\rangle / \sqrt{6}$, where b_j^\dagger and the other unobserved counterpart c_j^\dagger are photonic creation operators derived from an SU(2) unitary operator \tilde{U}_j according to the transformation

$$\begin{pmatrix} b_j^\dagger \\ c_j^\dagger \end{pmatrix} = \tilde{U}_j \begin{pmatrix} a_H^\dagger \\ a_V^\dagger \end{pmatrix}, \quad (3.1)$$

and a_H^\dagger and a_V^\dagger are the creation operators of the horizontal and vertical polarization modes [91]. Clearly, $\sum_j \Pi_j \neq 1$ this time, as the projectors are independently measured.

Figure 5 depicts the experimental setup to generate and characterize three-photon states. Four photons are produced through double pair emission of non-collinear SPDC process. The initial state is prepared in $|2, 2\rangle$ by combining two horizontally polarized photons and two vertically polarized photons with a PBSTo ensure that the photons are indistinguishable in the frequency domain, interference filters of 3 nm bandwidth centered at 780 nm are placed before sending the photons into the PBS. The four photons are then reduced into three photons by detecting a photon at D_1 and the reflected three photons from a PPBS are in the state of $|1, 2\rangle$ $\langle 1, 2|$. The PPBS perfectly reflects vertically polarized photons and reflects 1/3 of horizontally polarized photons. The HWP setting of $\theta_1 = 0^\circ$ leaves the state unchanged, whereas the setting of $\theta_1 = 45^\circ$ transforms the state into $|2, 1\rangle$ $\langle 2, 1|$. In addition, the mixed state $(|1, 2\rangle \langle 1, 2| + |2, 1\rangle \langle 2, 1|)/2$ is obtained by incoherently adding the relevant pure states through post-processing. These three-photon states are used to demonstrate the performances of ICCNet and FidNet in figure 9(b).

Experimentally [90], the three-photon states were characterized by acquiring the four-fold coincidence counts at D_1 , D_2 , D_3 , and D_4 for 16 rank-one projectors after passing through a PBS and BS. The SU(2) unitary operators \tilde{U}_j that define the projectors $\Pi_j = b_j^{\dagger 3} |0, 0\rangle \frac{1}{6} \langle 0, 0| b_j^3$ according to rule (3.1) are determined by the QWP and HWP angles of θ_2 and θ_3 inasmuch as $\tilde{U}_j = H(\theta_3)Q(\theta_2)$, where the matrix

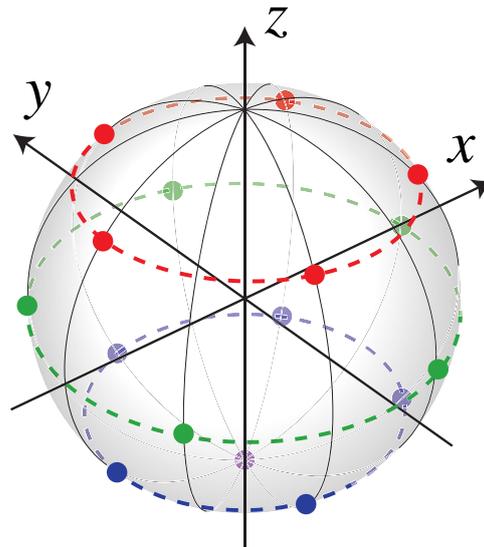


Figure 6. Reduced visualization of the 16 two-photon measurement projectors on the single-qubit Bloch sphere. The projectors of three-photon states are defined as $b_j^{i3} |0, 0\rangle / \sqrt{6}$ in accordance with equation (3.1). The projection states are chosen to equally distribute the corresponding single-photon component pure states $b_j^i |0, 0\rangle$ on the equatorial great circle and two small circles on the Bloch sphere, together with the south pole.

representations for the wave plates are given by

$$Q(\theta) \cong \frac{1}{\sqrt{2}} \begin{pmatrix} 1 - i \cos(2\theta) & -i \sin(2\theta) \\ -i \sin(2\theta) & 1 + i \cos(2\theta) \end{pmatrix},$$

$$H(\theta) \cong \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix}. \quad (3.2)$$

In our experiments, we consider $SU(2)$ rotations that fairly distribute the single-photon component $b_j^i |0, 0\rangle$ on three Bloch-spherical circles parallel to the equatorial plane [90, 92] as shown in figure 6. The measurement angles that realize these projectors are given in figure 5.

4. Results

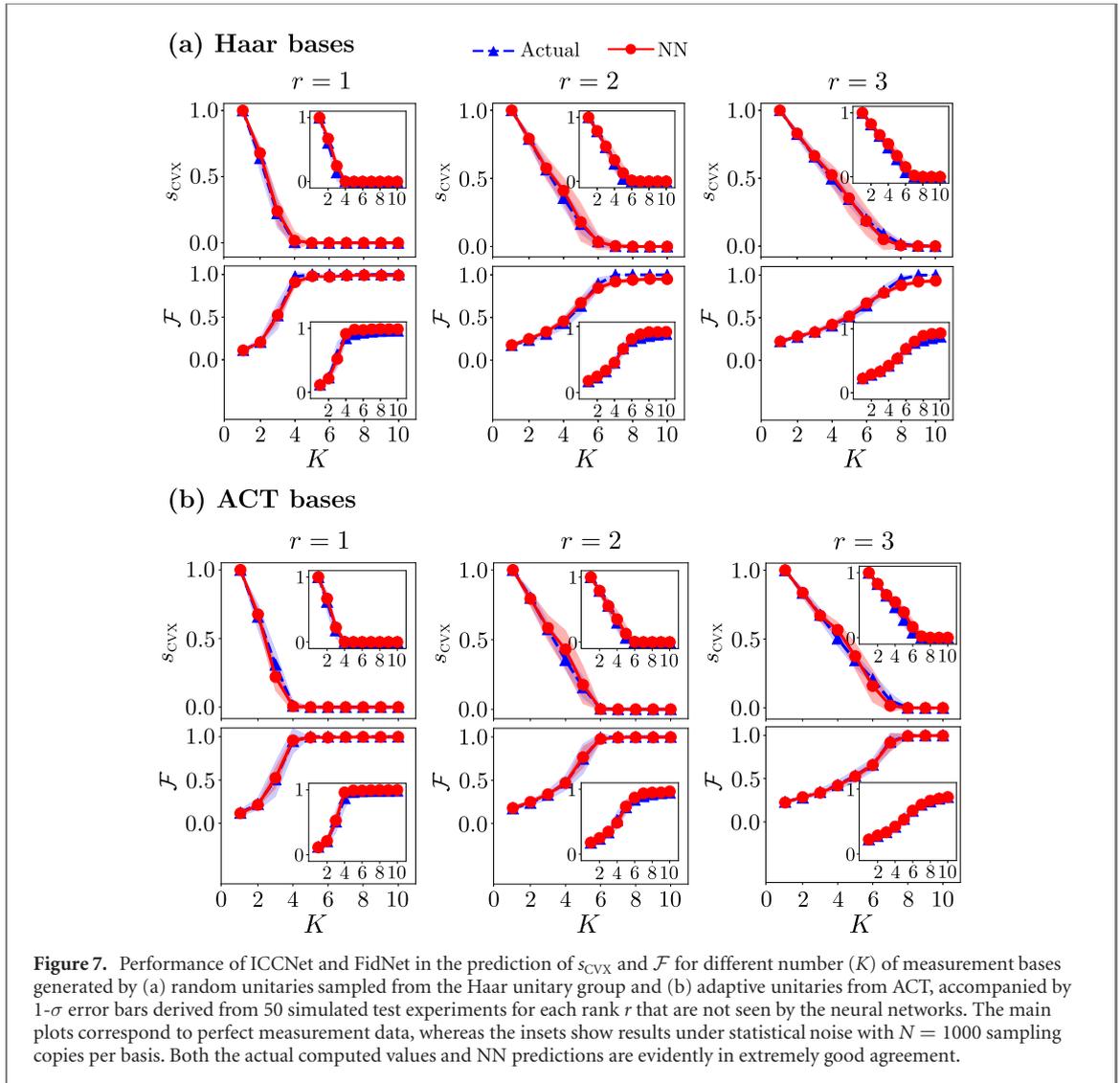
4.1. Simulations—neural-network performances

We first present performance graphs of ICCNet and FidNet in figure 7 based on two sets of simulations on four-qubit states ($d = 16$) using random measurement bases generated with the Haar measure for the unitary group (see section I of supplementary material), and bases found using ACT. In each set of simulations, for both cases where statistical noise is either absent or present, we collect simulation data of various number (K) of bases (s_{CVX} is normalized to 1 at $K = 1$ by default), each case recording measurements of 5000 randomly-generated quantum states of uniformly distributed rank $1 \leq r \leq 3$. The explicit CNN architecture employed is specified in section 2.2. The accurate fit between the actual computed values and those predicted by ICCNet and FidNet suggests that faithful neural network predictions of both the degree of informational completeness and fidelity are a definite possibility in both noiseless and statistically noisy environments. Sample codes for network training and evaluation with four-qubit simulation datasets are available online [93].

In separate simulations on four- ($d = 16$), five- ($d = 32$) and six-qubit ($d = 64$) systems with random Haar measurement bases, numerical evidence presented in figure 8 shows that the computation times in s_{CVX} NN predictions can be significantly reduced by about four orders of magnitude relative to ordinary SDP calculations, and this difference grows wider with larger dimensions. The corresponding ICCNet and FidNet performance graphs similar to figure 7 are given in section V of supplementary material.

4.2. Experimental performance with spatial-mode photonic states

For each value of d , we experimentally generated random pure states and construct their respective ACT measurement bases in order to evaluate the performance of ICCNet and FidNet, which were previously trained with 10 000 simulation datasets of random quantum states of uniformly distributed rank $1 \leq r \leq 3$

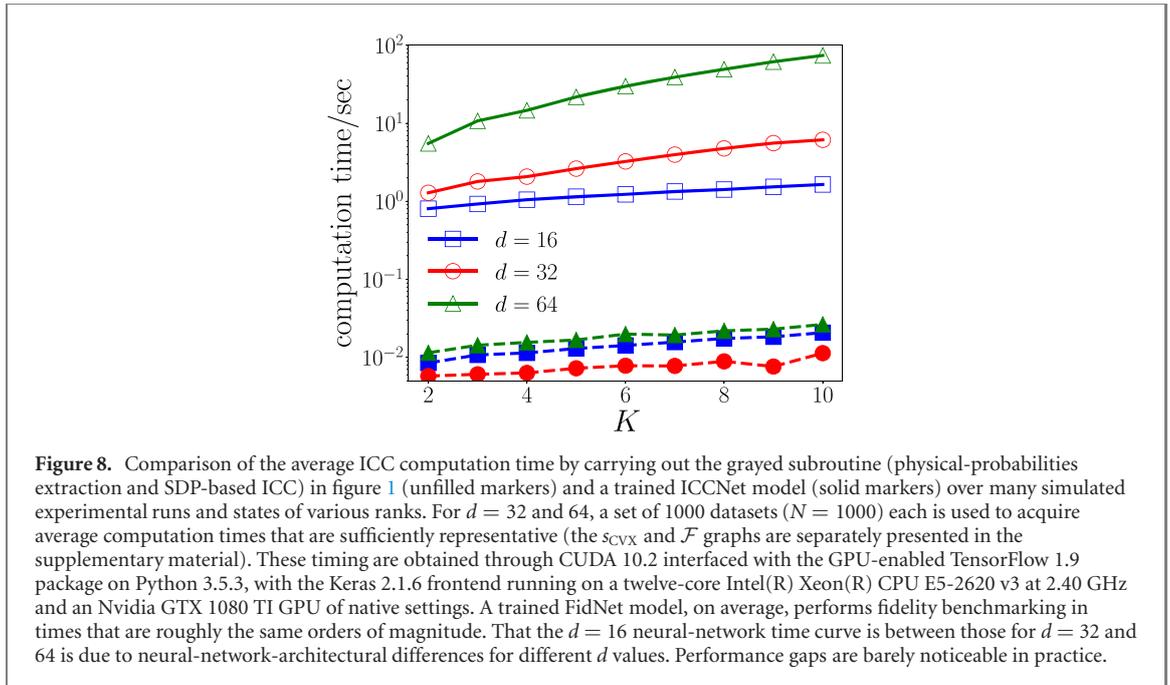


and different K values. These simulated training datasets are modeled with statistical noise arising from a multinomial distribution defined by $N = 5000$ sampling copies per basis, which is close to the experimental average.

Owing to experimental noise, the resulting spatial-mode quantum states are, as a matter of fact, nearly pure but sufficiently low-rank. Figure 9(a) confirms that ICC and fidelity benchmarking with simulation-trained neural network models are accurate even with real experimental test data. One can observe the relative network-prediction stability of s_{CVX} in contrast with that of \mathcal{F} . This coincides with the expectation that while the fidelity is strongly affected by statistical noise and other imperfections such as systematic errors, the degree of informational completeness is more intimately related to the quantum measurements and rank of the quantum state, such that noise only introduces perturbations on the functional behavior of s_{CVX} . Regardless, figure 9(a) shows that all predictions made by the simulation-trained ICCNet and FidNet models remain roughly within the error margins of actual computed values.

4.3. Experimental performance with multiphoton states

For every fixed number (L) of projectors chosen from the complete set of 16 defined in section 3.2, simulation datasets of 10 000 random $d = 4$ quantum states of uniformly distributed r are fed into both ICCNet and FidNet for training. These datasets are obtained from randomized sequences of the 16 projectors described above. Statistical noise is introduced into the simulation with multinomial distributions defined by $N = 500$ per projective measurement. To test the trained models and acquire prediction results depicted in figure 9(b), we make use of three different sets of 20 experimental runs outside the training datasets, each set corresponding to a different quantum state.



4.4. Noise training and reduction

Experimental noise due to imperfections and systematic errors are always present in any real dataset. Fluctuating deviations of NN predicted values from actual ones as observed in figure 9 arise from the lack of such experimental noise in all simulated training datasets, apart from purely statistical fluctuations, used to train ICCNet and FidNet.

When more knowledge about the noisy environment is acquired, data simulation from such knowledge may be carried out to improve the network predictions under such an environment. Here, we show that when some samples of experimental data that are sufficiently representative of the overall noise behavior can be spared for training, it is possible to train ICCNet and FidNet with both statistically-noisy simulated datasets and bootstrapped experimental datasets in order to learn the experimental noise effects approximately well and improve network predictions.

Bootstrapping entails using a given experimental dataset to generate numerous mock datasets using Monte Carlo procedures. More specifically, in the multinomial setting, the column \mathbf{v}_k of relative frequencies for the k th basis possess a Gaussian distribution of mean \mathbf{p}_k and covariance matrix $\Sigma_p^{(k)} = [\text{diag}(\mathbf{p}_k) - \mathbf{p}_k \mathbf{p}_k^T]/N$ for sufficiently large N owing to the central limit theorem, where $\text{diag}(\cdot)$ forms a diagonal matrix whose diagonals are defined by the argument. A direct substitution of \mathbf{v}_k for \mathbf{p}_k leads to the following simple rule for bootstrapping experimental ACT datasets from Hermite–Gaussian mode photonic system: $\mathbf{v}_k' = \mathcal{N}_{\geq 0}\{\mathbf{v}_k + \mathbf{w}_k\}$, where \mathbf{w}_k is a column of random variables collectively distributed according to the Gaussian distribution of zero mean and covariance matrix $\Sigma_v^{(k)}$ where $\Sigma_v^{(k)}$ is to be evaluated with the measurement relative frequencies of the particular k th basis and N is set to 5000, which is the estimated number of copies per ACT basis considered in section 4.2. The operation $\mathcal{N}_{\geq 0}$ is a composition of absolute value of the argument followed by its sum normalization over $0 \leq j \leq d - 1$ for the k th ACT basis. Finally, the states that produce the bases relative frequencies used in the bootstrapping procedure are different from the test states used to evaluate the network predictions.

Owing to a limited set of three-photon states, we adopt a different method to bootstrap experimental datasets acquired from these states. Since these datasets are obtained from measuring independent projectors, we randomly permute these projectors and their corresponding relative (unnormalized) frequencies in order generate new measurement sequences as mock datasets. The 16 projectors offer us a total of $16!$ permutations for each state, allowing us to conveniently generate an abundance of bootstrapped training datasets that are clearly different from those used for testing. By a similar token to the spatial-mode photonic systems, each relative frequency ν_l here is a binomial random variable normalized by the number of copies N used to measure the l th projector. Therefore, bootstrapping these relative frequencies may be carried out by additive Gaussian random variables inasmuch as $\nu_l' = \nu_l + w_l \sqrt{\nu_l(1 - \nu_l)/N}$, where w_l is a standard Gaussian random variable of zero mean and unit variance, and $N = 500$ is fixed as the estimated number of copies used to obtain the measured relative frequency for each projector, consistent with figure 9(b).

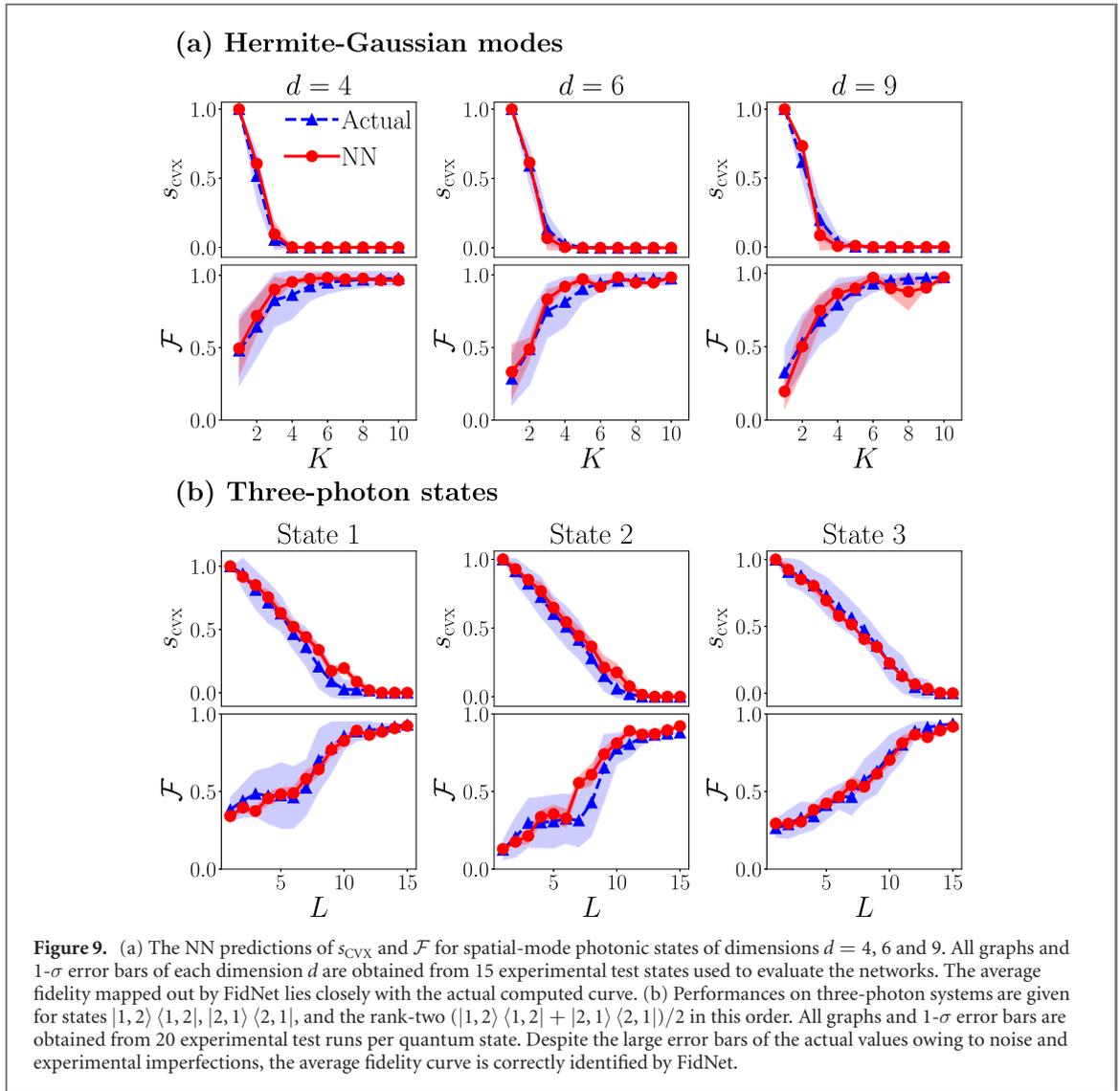


Figure 10 shows the enhanced prediction performances of ICCNet and FidNet. To generate this figure, a total of 5000 simulated and 5000 bootstrapped datasets are employed ($m = 10\,000$) for each group of experiments to train the networks for every value of K and L . These new plots indicate that slightly fluctuating NN prediction curves on noisy experimental data can be smoothed when bootstrapped information about the noisy environment is incorporated into the training.

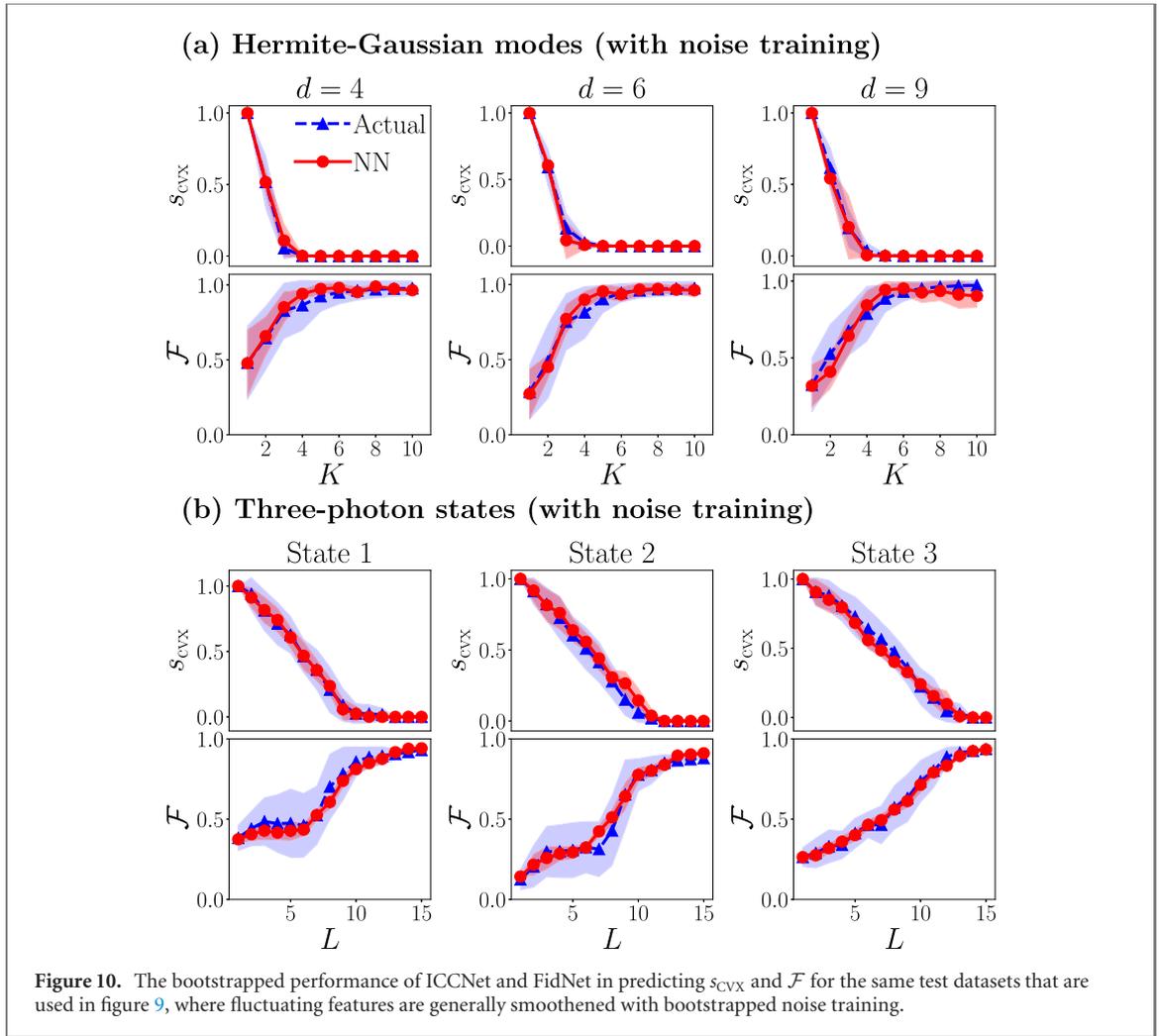
4.5. Suppression of systematic errors

To end this section, we shall now discuss the implications of all presented results, especially the computation performance graphs shown in figure 8, as far as real-time experimental systematic errors are concerned [94, 95]. As analytical results are unavailable, we resort to numerical analyses on the effects of ICC-computation-time reduction on such errors. For this purpose, we provide an important example of a kind of systematic drift phenomenon that is highly typical in optical fibers that carry spatial-mode photons, such as those of Hermite–Gaussian modes discussed in this article.

Focusing only on the transverse plane relative to the propagation direction of a laser beam, a given Hermite–Gaussian mode function $u_m(x)$ of order m in the spatial x -coordinate is given by [96]

$$u_m(x) = \left(\frac{2}{\pi}\right)^{1/4} \frac{1}{\sqrt{2^m m! w_0}} H_m\left(\frac{\sqrt{2}x}{w_0}\right) e^{-\frac{x^2}{w_0^2}}, \quad (4.1)$$

with w_0 being the beam waist, and $H_m(x)$ a degree- m Hermite polynomial in x . Upon using the ket notation, one can express the mode function in the familiar manner inasmuch as $\langle x|m; 0\rangle_{\text{HG}} = u_m(x)$. An ideal fiber would carry spatial-mode photons of a mode function that has a stable center point, which is usually set at the Cartesian origin as in (4.1). In real experiments, however, a main source of



time-dependent systematic errors is transversal displacements of $u_m(x)$ away from the origin [97, 98], which we may approximately model as random Wiener processes. Such displacements would distort the originally intended true state. Suppose that a basis ket $|m\rangle_{HG} \equiv |m; 0\rangle_{HG}$ is displaced away from the origin by a , it is shown in section IV of supplementary material that the resulting displaced $|m; a\rangle_{HG} = e^{-iaP}|m; 0\rangle_{HG}$, where $\langle x|e^{-iaP} = \langle x - a|$, possesses the following transformation function

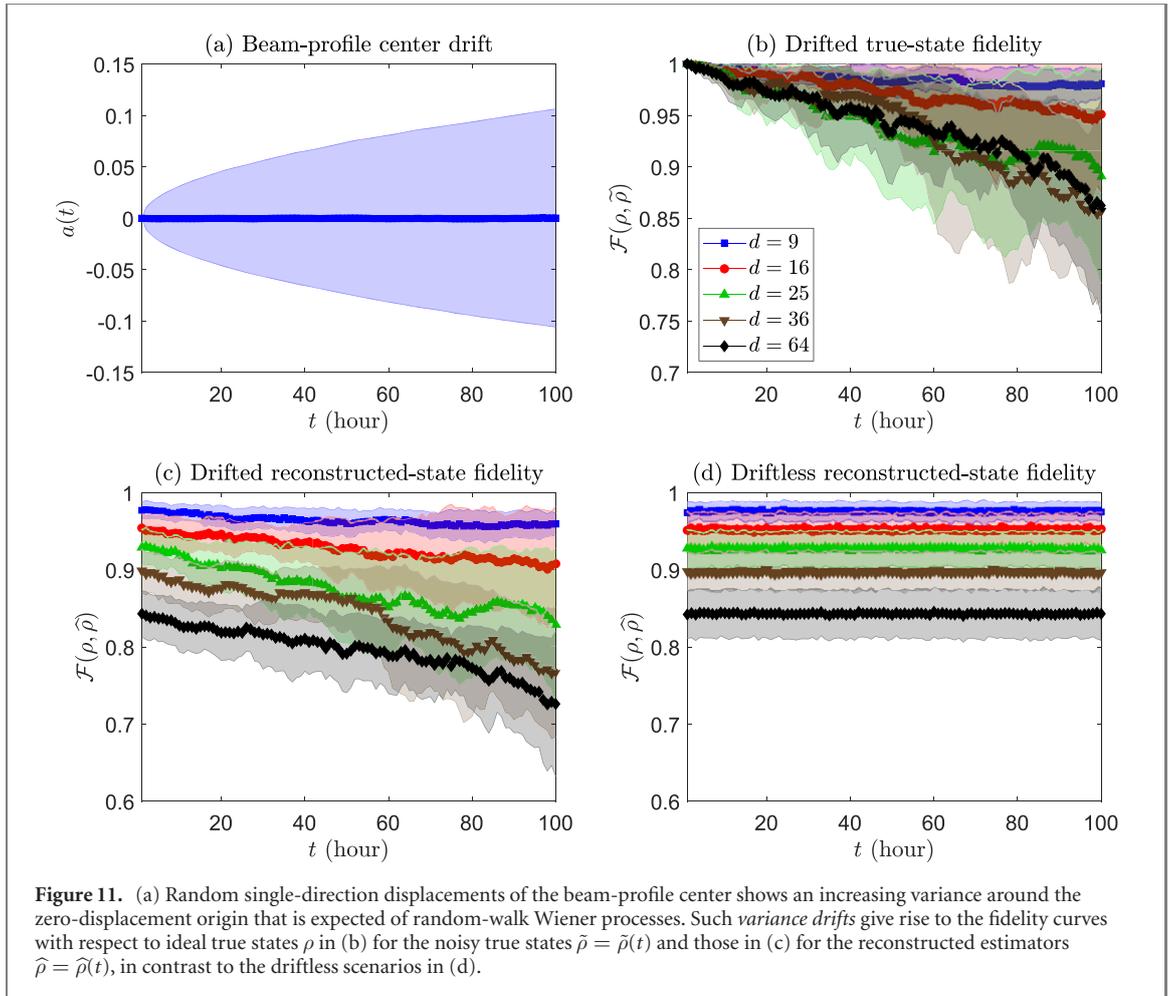
$$\begin{aligned}
 |m; a\rangle_{HG} &= \sum_l |l; 0\rangle_{HG} \langle l; 0|m; a\rangle_{HG}, \\
 \langle l; 0|m; a\rangle_{HG} &= \int dx' u_m(x' - a) u_l(x') \\
 &= \frac{e^{-\frac{a^2}{2w_0^2}}}{\sqrt{m!l!}} (-1)^m \left(\frac{a}{w_0}\right)^{n_>-n_<} U(-n_<, 1+n_>-n_<, a^2/w_0^2), \tag{4.2}
 \end{aligned}$$

where $n_> = \max\{m, l\}$, $n_< = \min\{m, l\}$, and

$$U(-n_<, 1+n_>-n_<, x^{-2}) = x^{-2n_<} \sum_{l=0}^{n_<} (-x^2)^l \binom{n_<}{l} \frac{n_>!}{(n_>-l)!} \tag{4.3}$$

is Kummer's confluent hypergeometric function. This result reduces to that in [97] for the $m = 0$ special case. The complete two-dimensional transverse profile of a Hermite–Gaussian mode in space is therefore described by the mode function $u_{m,n}(x, y) \equiv u_m(x)u_n(y)$, where the corresponding displaced kets $|m, n; \mathbf{a}\rangle_{HG} \equiv |m; a_1\rangle_{HG}|n; a_2\rangle_{HG}$ exhibit the transformation

$$|m, n; \mathbf{a}\rangle_{HG} = \sum_{l, l'} |l, l'; \mathbf{0}\rangle_{HG} \langle l; 0|m; a_1\rangle_{HG} \langle l'; 0|n; a_2\rangle_{HG}, \tag{4.4}$$

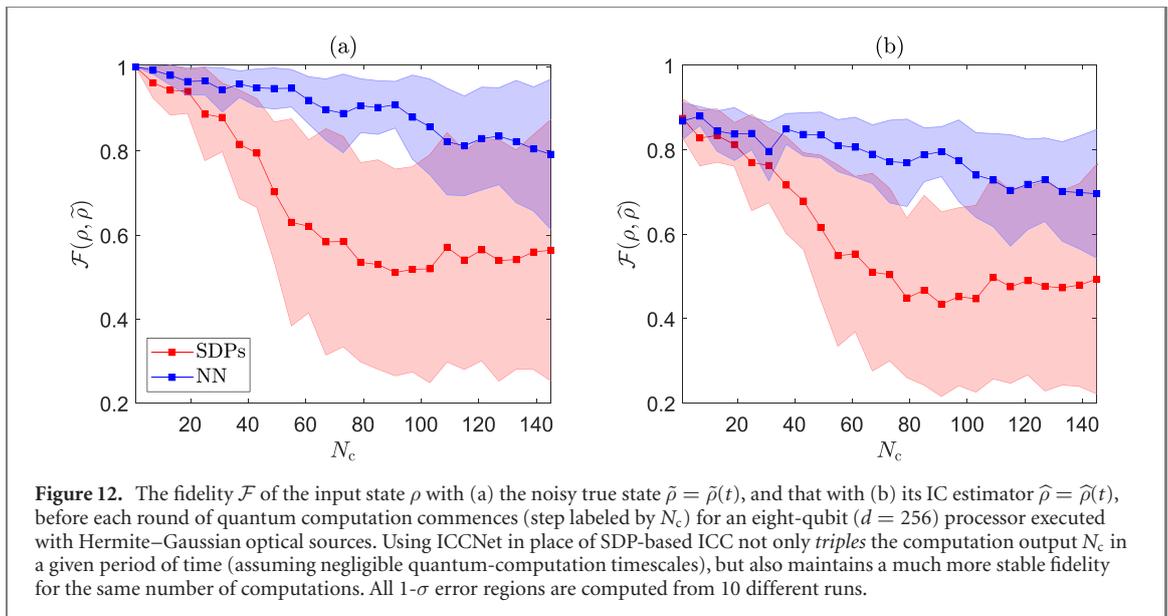


where the coefficients are pair-products of one-dimensional transformation functions as in (4.2). Components of the two-dimensional displacement $\mathbf{a} = (a_1 \ a_2)^\top$ are assumed to be independent. After a period of time t , the noisy true state $\tilde{\rho} = \sum_{m,n,m',n'} |m, n; \mathbf{a}(t)\rangle_{\text{HG}} \rho_{mn,m'n'} |m', n'; \mathbf{a}(t)\rangle$ is now defined in the displaced Hermite–Gaussian basis of some displacement $\mathbf{a}(t)$. The same type of disturbances apply to the POVM outcomes since they are implemented digitally using SLMs with the same beams.

In the experiments that gathered data used in plotting figure 9(a), the root-mean-square displacement of the Hermite–Gaussian beam center was measured to be about 5% of the beam waist w_0 after a period of 24 hours. This approximately coincides with a model of a Wiener process specified by the random displacement variable $\mathbf{a}(t) = \mathbf{a}(t-1) + \mathbf{b}(t-1)$, which is a cumulative temporal sum of random variables $\mathbf{b}(t)$ that are each distributed according to the standard Gaussian distribution defined by the standard deviation $\sigma \approx w_0/95$ when the time coordinate t is in units of an hour. Figure 11 shows the Hermite–Gaussian beam-profile center displacement and fidelity curves in time t for various dimensions d , where $d = d_0^2$ is the product of the individual dimensions d_0 of the truncated Hilbert space spanned by the finite set of Hermite–Gaussian basis kets of orders $0 \leq m, n \leq d_0 - 1$. In actual experiments, there most likely exist other time-dependent sources of errors not modeled here that could worsen the fidelities.

To appreciate the significance of systematic-error suppression, we consider a quantum processor that utilizes Hermite–Gaussian beams as a source for producing high-dimensional initial states for quantum computation. Suppose that the processor is running continuously under server conditions, and maintenance is carried out before significant systematic drifts are anticipated. Whenever the processor refreshes after each set of computations is completed, the newly prepared initial state ρ undergoes compressive state tomography to ensure that it is within the expected error margins.

As an instructive example, we consider at an eight-qubit processor controlled by a $d = 256$ Hermite–Gaussian source [99]. Each projector exposure time is about 1.5 s so that the measurement time of K von Neumann bases is $t_{\text{meas}} = 1.5Kd = 1920$ s. Preliminary calibration indicates that such an exposure period yields $N = 1000d = 2.56 \times 10^5$ state copies per basis. For a realistic error modeling, the expressions of t_{meas} and N in d were calibrated from the actual setup used to collect our experimental data for figures 9 and 10. We focus on the preparation of pure states, each of which requires only $K = 5$ random bases for an



IC reconstruction [53], the average time, estimated over five random sets of $K = 5$ bases, for an ICC verification (two SDPs) is $t_{\text{SDP}} = 4000$ s using the personal computer with hardware specification given in the caption of figure 8. After carrying out the basis measurements and ICC, the final state estimator $\hat{\rho}$ is given by the ML estimator that takes an average of $t_{\text{ML}} \approx 240$ s to generate using the accelerated projected-gradient algorithm in [75], which is insignificant in comparison to t_{SDP} . The time for each round of quantum computation depends on the actual application. For simplicity, we assume that each round of quantum computation is executed almost instantly since no classical post-processing is needed. From these specifications, we note that $t_{\text{SDP}} \approx 2t_{\text{meas}}$, and the prefactor grows with $d > 256$. Therefore, in the case of $d = 256$, replacing the two SDP algorithms in ICC with a trained ICCNet would shave about 66% of the total computation time off. For larger dimensions, if so desired, the ML estimation procedure may be completely replaced with trained conditional generative networks [62, 63] to eliminate t_{ML} .

Figure 12 shows the fidelity of the estimator $\hat{\rho}$ with the generated initial state ρ before every quantum-computation step (N_c of them in total). The graphs highlight the adverse effects of systematic errors when initial-state verification takes too long. Hence, compressive tomography performed with trained neural networks provides a better solution to real-time device certification with a higher fidelity stability, so that quantum computation can run much more smoothly with a greater N_c output before drift maintenance is applied.

5. Concluding remarks

We took advantage of the universality of convolutional networks to train two neural networks that can very efficiently certify a low-measurement-cost quantum-state characterization scheme. These networks can respectively benchmark the quantum completeness of a given set of measurement outcomes and corresponding data for reconstructing an unknown quantum state, as well as the resulting fidelity without explicitly carrying out the state reconstruction. Our machine-learning-assisted scheme therefore allows experimentalists to rapidly assess the sufficiency of measurement resources for an unambiguous characterization of arbitrary quantum states and achieve accelerated real-time verification without having to perform any optimization routine during the experiment. This becomes essential for many practical quantum tasks that do require fast execution times to avoid noise accumulation and drifts.

An arguably interesting problem would be to minimize the time required to acquire the trained neural networks. This includes both the training time and the generation of adequate training datasets. While the former can now be easily parallelized with graphics processing units, the latter involves two rounds of semidefinite programming per dataset as discussed in section 2.1, the acceleration of which is still a subject of ongoing research [100].

On the other hand, while classical algorithms for these procedures have worst-case polynomial time complexities in the dimension of the Hilbert space, it is known [101] that quantum algorithms can execute SDPs with polylogarithmic time complexities in the dimension. This immediately reveals the possibility of completely transforming the neural networks employed here, or part thereof, into their quantum

counterparts (fused with the training-data processing procedures that use quantum semidefinite programming) that could assimilate into a much larger set of networks for a grander purpose. Practical feasibility in implementing such extended quantum neural networks still remains to be seen.

Note.—Nearing the submission of our work, we discovered another very recent preprint reference [102] that purely discusses the estimation of the fidelity with FC networks. Apart from the clear distinction in architectures, we remark that while the networks in this reference were specifically trained for Pauli measurements, our CNN-based FidNet is compatible with generalized measurement inputs that can be readily used in parallel with ICCNet or any other quantum task that relies on arbitrary measurements. The objectives of both works are hence very different. The next key distinction is network training, which in this reference is based on categorical training that splits the continuous fidelity range into small intervals. As mentioned in the preprint itself, network training can be slow when the intervals are too small. In our current work, FidNet directly computes the fidelity values without such output splitting, and hence training efficiency is not sacrificed for prediction accuracy.

Acknowledgments

YST, SS and HJ acknowledge support by the National Research Foundation of Korea (Grant Nos. 2019R1A6A1A10073437, 2019M3E4A1080074, 2020R1A2C1008609, and 2020K2A9A1A06102946) via the Institute of Applied Physics at Seoul National University, and by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (Grant Nos. 2020-0-01606 and 2021-0-01059). GL acknowledges support by the Center of Excellence «Center of Photonics» funded by the Ministry of Science and Higher Education of the Russian Federation, Contract No. 075-15-2020-906. YK and Y-HK acknowledge support by the National Research Foundation of Korea (Grant No. 2019R1A2C3004812) and the ITRC support program (IITP-2020-0-01606). LLSS acknowledges support from European Union's Horizon 2020 research and innovation program (ApresSF and STORMYTUNE) and the Ministerio de Ciencia e Innovación (PGC2018-099183-B-I00). The MSU team acknowledges support from the Russian Foundation for Basic Research (RFBR Project No. 19-32-80043 and RFBR Project No. 19-52-80034) and support under the Russian National Technological Initiative via MSU Quantum Technology Centre. SSS and SPK acknowledge support by the Development Program of the Interdisciplinary Scientific and Educational School of Lomonosov Moscow State University 'Photonic and quantum technologies: Digital medicine'.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Yong Siah Teo  <https://orcid.org/0000-0002-1766-6402>
Gleb I Struchalin  <https://orcid.org/0000-0002-8887-3361>
Stanislav S Straupe  <https://orcid.org/0000-0001-9810-1958>
Gerd Leuchs  <https://orcid.org/0000-0003-1967-2766>
Luis L Sánchez-Soto  <https://orcid.org/0000-0002-7441-8632>

References

- [1] Grimsley H R, Economou S E, Barnes E and Mayhall N J 2019 *Nat. Commun.* **10** 3007
- [2] Arute F *et al* 2019 *Nature* **574** 505
- [3] Hu L *et al* 2019 *Nat. Phys.* **15** 503
- [4] Havlíček V, Córcoles A D, Temme K, Harrow A W, Kandala A, Chow J M and Gambetta J M 2019 *Nature* **567** 209
- [5] Beer K, Bondarenko D, Farrelly T, Osborne T J, Salzmann R, Scheiermann D and Wolf R 2020 *Nat. Commun.* **11** 808
- [6] Gard B T, Zhu L, Barron G S, Mayhall N J, Economou S E and Barnes E 2020 *npj Quantum Information* **6** 10
- [7] Plesch M and Brukner Č 2011 *Phys. Rev. A* **83** 032302
- [8] Holmes A and Matsuura A Y 2020 *2020 IEEE Int. Conf. on Quantum Computing and Engineering (QCE)* pp 169–79
- [9] Schäfer V M, Ballance C J, Thirumalai K, Stephenson L J, Ballance T G, Steane A M and Lucas D M 2018 *Nature* **555** 75
- [10] Shi X-F 2018 *Phys. Rev. Appl.* **10** 034006
- [11] Ono T, Okamoto R, Tanida M, Hofmann H F and Takeuchi S 2017 *Sci. Rep.* **7** 45353
- [12] Patel R B, Ho J, Ferreyrol F, Ralph T C and Pryde G J 2016 *Sci. Adv.* **2** e1501531

- [13] Fiurášek J 2008 *Phys. Rev. A* **78** 032317
- [14] Wootters W K and Fields B D 1989 *Ann. Phys.*, NY **191** 363
- [15] Nielsen M A 2003 *Phys. Lett. A* **308** 96
- [16] Raussendorf R and Briegel H J 2001 *Phys. Rev. Lett.* **86** 5188
- [17] Briegel H J, Browne D E, Dür W, Raussendorf R and Van den Nest M 2009 *Nat. Phys.* **5** 19–26
- [18] Durt T, Englert B-G, Bengtsson I and Życzkowski K 2010 *Int. J. Quant. Inf.* **08** 535
- [19] Scott A J 2006 *J. Phys. A: Math. Gen.* **39** 13507
- [20] Zhu H and Englert B-G 2011 *Phys. Rev. A* **84** 022327
- [21] Zhu H 2014 *Phys. Rev. A* **90** 012115
- [22] Chuang I and Nielsen M 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
- [23] Paris M G A and Řeháček J (ed) 2004 *Quantum State Estimation (Lecture Notes in Physics)* vol 649 (Berlin: Springer)
- [24] Teo Y S 2015 *Introduction to Quantum-State Estimation* (Singapore: World Scientific)
- [25] O’Brien J L, Pryde G J, Gilchrist A, James D F V, Langford N K, Ralph T C and White A G 2004 *Phys. Rev. Lett.* **93** 080502
- [26] Poyatos J F, Cirac J I and Zoller P 1997 *Phys. Rev. Lett.* **78** 390
- [27] Teo Y S, Englert B-G, Řeháček J and Hradil Z 2011 *Phys. Rev. A* **84** 062125
- [28] Luis A and Sánchez-Soto L L 1999 *Phys. Rev. Lett.* **83** 3573
- [29] Fiurášek J 2001 *Phys. Rev. A* **64** 024102
- [30] D’Ariano G M, Maccone L and Presti P L 2004 *Phys. Rev. Lett.* **93** 250407
- [31] Chen Y, Farahzad M, Yoo S and Wei T-C 2019 *Phys. Rev. A* **100** 052315
- [32] Zhang L, Datta A, Coldenstrodt-Ronge H B, Jin X-M, Eisert J, Plenio M B and Walmsley I A 2012 *New J. Phys.* **14** 115005
- [33] Altorio M, Genoni M G, Somma F and Barbieri M 2016 *Phys. Rev. Lett.* **116** 100802
- [34] Kim Y, Kim Y-S, Lee S-Y, Han S-W, Moon S, Kim Y-H and Cho Y-W 2018 *Nat. Commun.* **9** 192
- [35] Gaikwad A, Rehal D, Singh A, Arvind and Dorai K 2018 *Phys. Rev. A* **97** 022311
- [36] Bendersky A and Paz J P 2013 *Phys. Rev. A* **87** 012122
- [37] Schmiegelow C T, Bendersky A, Larotonda M A and Paz J P 2011 *Phys. Rev. Lett.* **107** 100502
- [38] Bendersky A, Pastawski F and Paz J P 2009 *Phys. Rev. A* **80** 032116
- [39] Bendersky A, Pastawski F and Paz J P 2008 *Phys. Rev. Lett.* **100** 190403
- [40] Proctor T, Rudinger K, Young K, Sarovar M and Blume-Kohout R 2017 *Phys. Rev. Lett.* **119** 130502
- [41] Helsen J, Xue X, Vandersypen L M K and Wehner S 2019 *npj Quantum Information* **5** 71
- [42] Lu Y, Sim J Y, Suzuki J, Englert B-G and Ng H K 2020 *Phys. Rev. A* **102** 022410
- [43] Gross D, Liu Y-K, Flammia S T, Becker S and Eisert J 2010 *Phys. Rev. Lett.* **105** 150401
- [44] Kalev A, Kosut R L and Deutsch I H 2015 *npj Quantum Information* **1** 15018
- [45] Baldwin C H, Deutsch I H and Kalev A 2016 *Phys. Rev. A* **93** 052105
- [46] Steffens A, Riofrío C A, McCutcheon W, Roth I, Bell B A, McMillan A, Tame M S, Rarity J G and Eisert J 2017 *Quantum Sci. Technol.* **2** 025005
- [47] Schwemmer C, Tóth G, Niggebaum A, Moroder T, Gross D, Gühne O and Weinfurter H 2014 *Phys. Rev. Lett.* **113** 040503
- [48] Riofrío C A, Gross D, Flammia S T, Monz T, Nigg D, Blatt R and Eisert J 2017 *Nat. Commun.* **8** 15305
- [49] Baldwin C H, Kalev A and Deutsch I H 2014 *Phys. Rev. A* **90** 012110
- [50] Rodionov A V, Veitia A, Barends R, Kelly J, Sank D, Wenner J, Martinis J M, Kosut R L and Korotkov A N 2014 *Phys. Rev. B* **90** 144504
- [51] Shabani A, Kosut R L, Mohseni M, Rabitz H, Broome M A, Almeida M P, Fedrizzi A and White A G 2011 *Phys. Rev. Lett.* **106** 100401
- [52] Ahn D et al 2019 *Phys. Rev. Lett.* **122** 100404
- [53] Ahn D, Teo Y S, Jeong H, Koutný D, Řeháček J, Hradil Z, Leuchs G and Sánchez-Soto L L 2019 *Phys. Rev. A* **100** 012346
- [54] Teo Y S, Struchalin G I, Kovlakov E V, Ahn D, Jeong H, Straupe S S, Kulik S P, Leuchs G and Sánchez-Soto L L 2020 *Phys. Rev. A* **101** 022334
- [55] Kim Y, Teo Y S, Ahn D, Im D-G, Cho Y-W, Leuchs G, Sánchez-Soto L L, Jeong H and Kim Y-H 2020 *Phys. Rev. Lett.* **124** 210401
- [56] Gianani I, Teo Y, Cimini V, Jeong H, Leuchs G, Barbieri M and Sánchez-Soto L 2020 *PRX Quantum* **1** 020307
- [57] Vandenbergh L and Boyd S 1996 *SIAM Rev.* **38** 49
- [58] Torlai G, Mazzola G, Carrasquilla J, Troyer M, Melko R and Carleo G 2018 *Nat. Phys.* **14** 447
- [59] Palmieri A M, Kovlakov E, Bianchi F, Yudin D, Straupe S, Biamonte J D and Kulik S 2020 *npj Quantum Information* **6** 20
- [60] Neugebauer M, Fischer L, Jäger A, Czischek S, Jochim S, Weidemüller M and Gärtner M 2020 *Phys. Rev. A* **102** 042604
- [61] Lohani S, Kirby B T, Brodsky M, Danaci O and Glasser R T 2020 *Mach. Learn.: Sci. Technol.* **1** 035007
- [62] Ahmed S, Muñoz C S, Nori F and Kockum A F 2020 Quantum state tomography with conditional generative adversarial networks (arXiv:2008.03240 [quant-ph])
- [63] Ahmed S, Muñoz C S, Nori F and Kockum A F 2020 Classification and reconstruction of optical quantum states with deep neural networks (arXiv:2012.02185 [quant-ph])
- [64] Lecun Y, Bottou L, Bengio Y and Haffner P 1998 *Proc. IEEE* **86** 2278
- [65] Krizhevsky A, Sutskever I and Hinton G E 2012 *Advances in Neural Information Processing Systems* vol 25 ed F Pereira, C J C Burges, L Bottou and K Q Weinberger (Curran Associates, Inc.) pp 1097–105
- [66] Russakovsky O et al 2015 *Int. J. Comput. Vis.* **115** 211
- [67] He K, Zhang X, Ren S and Sun J 2016 *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 770–8
- [68] Zhou D-X 2020 *Appl. Comput. Harmon. Anal.* **48** 787
- [69] Ming Y, Lin C-T, Bartlett S D and Zhang W-W 2019 *npj Comput. Mater.* **5** 88
- [70] Cong I, Choi S and Lukin M D 2019 *Nat. Phys.* **15** 1273
- [71] Melnikov A A, Fedichkin L E and Alodjants A 2019 *New J. Phys.* **21** 125002
- [72] Tsai Y-H, Yu M-Z, Hsu Y-H and Chung M-C 2020 *Phys. Rev. B* **102** 054512
- [73] Řeháček J, Hradil Z, Knill E and Lvovsky A I 2007 *Phys. Rev. A* **75** 042108
- [74] Teo Y S, Zhu H, Englert B-G, Řeháček J and Hradil Z 2011 *Phys. Rev. Lett.* **107** 020404
- [75] Shang J, Zhang Z and Ng H K 2017 *Phys. Rev. A* **95** 062336
- [76] Kariya T and Kurata H 2004 *Generalized Least Squares* (New York: Wiley)
- [77] Rencher A C and Christensen W F 2012 *Methods of Multivariate Analysis* (New York: Wiley)
- [78] Knee G C, Bolduc E, Leach J and Gauger E M 2018 *Phys. Rev. A* **98** 062336

- [79] Boyle J P and Dykstra R L 1986 *Advances in Order Restricted Statistical Inference* ed R Dykstra, T Robertson and F T Wright (New York: Springer) pp 28–47
- [80] Rumelhart D E and McClelland J L 1987 Learning internal representations by error propagation *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations* (Cambridge, MA: MIT Press) pp 318–62
- [81] Hastie T, Tibshirani R and Friedman J 2009 *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (New York: Springer)
- [82] Hinton G E, Srivastava N, Krizhevsky A, Sutskever I and Salakhutdinov R R 2012 Improving neural networks by preventing co-adaptation of feature detectors (arXiv:1207.0580[cs.NE])
- [83] Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R 2014 *J. Mach. Learn. Res.* **15** 19291929
- [84] Warde-Farley D, Goodfellow I J, Courville A C and Bengio Y 2014 *2nd Int. Conf. on Learning Representations, ICLR 2014, Conf. Track Proc.* (Banff, AB, Canada 14–16 April 2014) ed Y Bengio and Y LeCun
- [85] Ioffe S and Szegedy C 2015 *Proc. of the 32nd Int. Conf. Machine Learning—Volume 37, ICML'15* (JMLR.org) pp 448–56
- [86] Santurkar S, Tsipras D, Ilyas A and Mądry A 2018 *Proc. of the 32nd Int. Conf. on Neural Information Processing Systems, NIPS'18* (Red Hook, NY, USA: Curran Associates Inc.) pp 2488–98
- [87] Dozat T 2016 *ICLR Workshop* pp 2013–6
- [88] Bolduc E, Bent N, Santamato E, Karimi E and Boyd R W 2013 *Opt. Lett.* **38** 3546
- [89] Mair A, Vaziri A, Weihs G and Zeilinger A 2001 *Nature* **412** 313
- [90] Kim Y, Björk G and Kim Y-H 2017 *Phys. Rev. A* **96** 033840
- [91] Schilling U, von Zanthier J and Agarwal G S 2010 *Phys. Rev. A* **81** 013826
- [92] Israel Y, Afek I, Rosen S, Ambar O and Silberberg Y 2012 *Phys. Rev. A* **85** 022115
- [93] Codes written in MATLAB and Python are uploaded onto Github at <https://github.com/ACAD-repo/ICCNNet-FidNet>
- [94] Proctor T, Revelle M, Nielsen E, Rudinger K, Lobser D, Maunz P, Blume-Kohout R and Young K 2020 *Nat. Commun.* **11** 5396
- [95] Kelly J, O'Malley P, Neeley M, Neven H and Martinis J M 2018 Physical qubit calibration on a directed acyclic graph (arXiv:1803.03226 [quant-ph])
- [96] Siegman A E 1986 *Lasers* (Mill Valley, CA: University Science Books)
- [97] Hsu M T L, Delaubert V, Lam P K and Bowen W P 2004 *J. Opt. B: Quantum Semiclass. Opt.* **6** 495
- [98] Ndagano B, Mphuthi N, Milione G and Forbes A 2017 *Opt. Lett.* **42** 4175
- [99] Grandstrand O 2012 *Advanced Photonic Sciences* ed E Nagali and F Sciarrino (Oxford: InTechOpen)
- [100] Majumdar A, Hall G and Ahmadi A A 2020 *Annual Review of Control, Robotics, and Autonomous Systems* **3** 331
- [101] Brandão F G S L, Kalev A, Li T, Lin C Y-Y, Svore K M and Wu X 2019 *46th Int. Coll. on Automata, Languages, and Programming (ICALP 2019), Leibniz Int. Proc. in Informatics (LIPIcs)* vol 132 ed C Baier, I Chatzigiannakis, P Flocchini and S Leonardi (Dagstuhl, Germany: Schloss Dagstuhl—Leibniz-Zentrum fuer Informatik)
- [102] Zhang X, Luo M, Wen Z, Feng Q, Pang S, Luo W and Zhou X 2021 Direct fidelity estimation of quantum states using machine learning (arXiv:2102.02369 [quant-ph])