

## Neural Networks for Detecting Multimode Wigner Negativity

Valeria Cimini<sup>1,\*</sup>, Marco Barbieri<sup>1</sup>, Nicolas Treps<sup>2</sup>, Mattia Walschaers<sup>2</sup>, and Valentina Parigi<sup>2</sup>

<sup>1</sup>*Dipartimento di Scienze, Università degli Studi Roma Tre, Via della Vasca Navale 84, 00146 Rome, Italy*

<sup>2</sup>*Laboratoire Kastler Brossel, Sorbonne Université,*

*CNRS, ENS-PSL Research University, Collège de France, 4 place Jussieu, F-75252 Paris, France*

 (Received 9 March 2020; revised 20 July 2020; accepted 16 September 2020; published 16 October 2020)

The characterization of quantum features in large Hilbert spaces is a crucial requirement for testing quantum protocols. In the continuous variable encoding, quantum homodyne tomography requires an amount of measurement that increases exponentially with the number of involved modes, which practically makes the protocol intractable even with few modes. Here, we introduce a new technique, based on a machine learning protocol with artificial neural networks, that allows us to directly detect negativity of the Wigner function for multimode quantum states. We test the procedure on a whole class of numerically simulated multimode quantum states for which the Wigner function is known analytically. We demonstrate that the method is fast, accurate, and more robust than conventional methods when limited amounts of data are available. Moreover, the method is applied to an experimental multimode quantum state, for which an additional test of resilience to losses is carried out.

DOI: [10.1103/PhysRevLett.125.160504](https://doi.org/10.1103/PhysRevLett.125.160504)

The ability to engineer large and scalable multiparty quantum states is at the core of future quantum technologies. In particular, large entangled structures are essential for measurement-based quantum computing protocols [1,2]. Multimode quantum optics represents a powerful platform for generating large entangled networks in the continuous variables (CV) regime. Over the last decade the generation of up to thousands of multimode entangled states has been experimentally demonstrated [3–7] in optical parametric processes. These quantum states, which are characterized by Gaussian statistics, are necessary but not sufficient alone to perform quantum computing protocols, as they can be efficiently simulated via classical resources. The implementation of unconditional non-Gaussian operations is a much more demanding task, as it requires strong nonlinear interactions. In the perspective of investigating intermediate-scale systems in the near term, de-Gaussification via heralded photon subtraction or addition operations has been demonstrated [8–22]. If the generation of non-Gaussian multimode entangled states is within reach of state-of-the-art experimental platforms, the full characterization of their quantum state remains a hard task.

In the CV picture Gaussian quantum states are completely characterized by the mean values of two conjugated quadratures per mode, plus their covariance matrix. Beyond Gaussian statistics, the complete quantum description of the optical system in terms of the density matrix or its equivalent Wigner representation, may be recovered via quantum homodyne tomography. In usual maximum likelihood (MaxLik) procedures [23,24] good accuracy in the state reconstruction requires a large number of

measurements, which scales exponentially with the number of involved optical modes. This implies that the setup should be stable until the whole set of measurements is taken, and that the algorithm for reconstructing the state becomes computationally too heavy.

A particular feature that we are interested to recover for non-Gaussian quantum states is the negativity of their Wigner representation, as it is accounted to be a pivotal quantum resource [25–27]. This can be tested after its tomographic reconstruction via the usual MaxLik procedure, but this is not a viable option in the multimode scenario. Moreover, when interested in a particular quantum feature, like the Wigner negativity, and not in the complete knowledge of the Wigner function, it is worth finding a more direct approach to link the measurements and the specific property we are interested in.

In this Letter, we discuss an alternative approach, easily scalable with the number of optical modes of the system, that aims to specifically identify the presence of Wigner negativity, given a set of quadrature data. Our method relies on machine learning algorithms [28], which have been demonstrated to be particularly powerful for characterization and optimization of quantum systems in different contexts [29–39]. In particular, artificial neural networks (NNs) already offer an alternative and efficient strategy to represent quantum many-body states, enabling us to perform quantum state tomography for high dimensional states from a limited number of experimental data [40,41]. Most of the protocols have been implemented in a discrete variables framework, one approach for quantum homodyne tomography has been proposed [42], and experimentally tested in the single-mode configuration. Our algorithm

allows, in a supervised learning approach, the discrimination between multimode optical states presenting negative or positive Wigner function and it is the first application of a machine learning algorithm to CV multimode optical states.

Compared to standard quantum homodyne tomography protocols, our method is more robust when limited data are available. Furthermore, it allows us to identify the Wigner negativity for states up to ten modes, a task too hard to accomplish with current  $\text{MaxLik}$  procedures. In this Letter, we first test the method with simulated data and then we apply it to classify the Wigner negativity of experimental quantum states.

Testing the negativity of the Wigner function, namely study if the function is either always positive or shows some negative regions, can be seen as a binary classification problem [43]. This is a very common use case of machine learning algorithms, which are daily used to solve tasks like email spam filtering, document categorization, as well as speech, image, and handwriting recognition. The use of NNs is in fact suitable in problems where the outcomes of the observed variables span in a large space. This is the case for the measurement of optical quadratures whose continuous values span, in principle, over an infinite phase space. Indeed, thanks to the network's ability to dynamically create complex prediction functions, there is no need for modeling [44–47]. Also, in standard tomographic methods a considerable number of data have to be collected for every new instance to be classified and the algorithm is run with no memory of previous samples. On the contrary in the NN approach, after a first stage of training (learning) on simulated quadrature measurements of hundreds of different states, the network is able to give a fast classification of new samples. The performance of this new method is then compared to the Wigner negativities that are found when applying the  $\text{MaxLik}$  protocol to the same quadrature data for few-mode states.

The success of the NN approach in classifying states with Wigner negativity relies on our ability to generate training data, i.e., to simulate quadrature outcomes of various states. Therefore, we limit ourselves to the actual operations that are at hand in the experiments of interest: the ability to prepare an arbitrary  $m$ -mode Gaussian state, the ability to act with a non-Gaussian operation on these states through either photon addition or photon subtraction, and the occurrence of losses. To generate the quadrature data, we use the Wigner function that can be obtained analytically [48–50]. We simulate squeezed states over  $m = 3, 5$ , or 10 modes (with randomly chosen squeezing between 0 and 8 dB for each mode), we randomly choose whether or not to add or subtract a photon, and we add a randomly chosen percentage of losses and thermal noise (see Supplemental Material for details [51]). We generate 4000 of these states, with approximately the same number of positive and negative Wigner functions. For each we

choose a random mode basis in which we perform  $k = 1000$  repetitions of joint quadrature measurements. Each of them contains three detectors outcome per mode associated with three different phases, chosen randomly within three fixed phase intervals. This leads to a total of  $3000 \times m$  quadrature measurements per state.

Because the training data are simulated starting from an analytical Wigner function, we know its minimal value  $W_{\min}$ . This value can be converted to a binary classifier: the target output  $W_0$  is set to  $W_0 = 0$  if  $W_{\min} \geq C$ , while  $W_0 = 1$  if  $W_{\min} < C$ . The constant  $C < 0$  represents a cutoff allowing us to exclude limiting cases where the Wigner negativity is too close to zero to be considered significant or to look specifically for highly negative Wigner functions. Thus, for every simulated measurement we add the label  $W_0$  to highlight whether or not the data correspond to a nonpositive Wigner function.

Feeding in the complete set of joint homodyne detection events would require a NN of thousands of nodes. Instead, in order to pass the information contained in the quadrature distribution to the input layer, we binned the data for each individual mode and choice of phase, and evaluate the occurrence frequencies in each bin, normalized to the total number of measurements. In practice, we process the data with five bins relative to the three different phase values for each mode, which corresponds to having an input layer with only  $15 \times m$  nodes. We feed the network with a matrix of all combined training data with  $15 \times m$  columns and as many rows as the different states that we use for the training. With this method of data processing, we do not require joint quadrature measurements. The correlations between the different modes are effectively integrated out, thus the method scales linearly with the number of modes rather than exponentially. This also makes the generation of training data less demanding.

To identify the Wigner negativity of an  $m$ -mode optical state we use a feed-forward NN with three hidden layers, with, respectively, 30, 20, and 10 nodes, all activated by a rectified linear unit function. This architecture of the NN has been selected among the different tested configurations as the one yielding the least means squared error on the validation set. The output layer consists of one node activated by sigmoid function. The tuning of all the hyperparameters of the network, namely, all those parameters that are set before the learning process begins, is done using a grid search in order to determine the optimal values for a given model. The training is performed minimizing the loss function, that in our case is a simple mean squared error between the target value  $W_0$  and the network output. The latter, that corresponds to the probability of having a negative minimum value of the Wigner function, is then used to classify the state as negative if it exceeds the threshold  $P_{\text{th}} = 50\%$ .

As is customary in evaluating the NN performance, we use a cross-validation procedure, in which we split the data

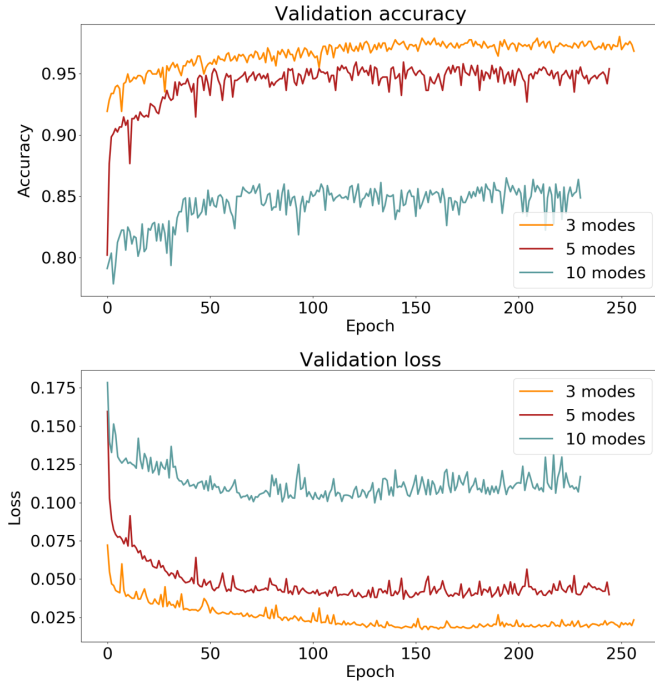


FIG. 1. NN performances on the validation set. Top: accuracy of the model as a function of the epoch, i.e., each iteration of the training. Bottom: loss function as a function of the epoch. In both panels, the orange, red and cyan curves refer, respectively, to the performances on  $m = 3, 5$ , and  $10$  mode states.

into two parts—the training set (80% of total data), and the validation set (20% of total data). The training set is used to train the model, while the validation set is used to evaluate the model’s performance on a different sample of data and it is used to stop the network training as soon as the loss function evaluated for this set stops decreasing in a sufficiently long interval, in order to avoid overfitting. This is referred to as early stopping.

In Fig. 1 we show how the model performs, after it has been trained, on the independent validation set. To this purpose, we evaluate the NN accuracy defined as the fraction of instances in the validation set which are correctly identified. The correct identification of states with a negative Wigner function happens with good accuracy for all the tested cases. The accuracy for the  $m = 3$  modes case is particularly high, exceeding 95% of correct identifications, however, even for the  $m = 10$  modes states, our method correctly identifies more than 85% of the states in the validation set. We found that the discrimination is optimized using the cutoff  $C = -0.1/(2\pi)^m$ , where  $-1/(2\pi)^m$  is the maximal negativity attained by an  $m$ -mode state. We refer the reader to the Supplemental Material [51] for further details on the algorithm and its appraisal.

We remark that even starting with the same batch of training data, the optimization of the NN will end up in slightly different configurations, since the weights of the

connections are randomly initialized. In addition, there is an element of potential variability in using different training sets taken from the same class of states. However, we have checked that this level of accuracy is reproducible and independent on the choice of the training set.

We investigated the comparison between the NN performance in identifying the Wigner negativity with the one of the standard state tomography, based on the MaxLik algorithm. This is carried out only for  $m = 3$ , for which this procedure is computationally feasible. Even with a limited amount of data, this method will provide a density matrix  $\rho_*$  for the state, represented in the Fock basis. It is reasonable to assume that  $\rho_*$  already manifests Wigner negativity long before the full tomography has converged. When we limit ourselves to photon-added or -subtracted Gaussian states that do not have any initial mean field, it can be proven that the Wigner function achieves its most negative value in the origin of phase space [48]. This value can be calculated [54] via the parity operator:

$$W_{\min} = \sum_{n_1, n_2, n_3=0}^{N_{\text{ph}}} \frac{(-1)^{n_1+n_2+n_3}}{8\pi^3} \langle n_1, n_2, n_3 | \rho_* | n_1, n_2, n_3 \rangle, \quad (1)$$

where  $|n_1, n_2, n_3\rangle$  denotes the state of the three-mode Fock basis, and  $N_{\text{ph}}$  is the maximal photon number that is chosen as a trade-off between accuracy and computing time for the MaxLik procedure. Here, we set  $N_{\text{ph}} = 5$  and run 15 iterations of the algorithm. In Fig. 2, the performance of the NN approach is compared to that of MaxLik for the same number of measurements  $k = 1000, 100, 30, 10$ , on 100 different states. The uncertainties correspond to the variability observed over 6 different batches. As expected, the MaxLik estimation improves its performance with the number of measurements and its variability improves as well. The adoption of NN results in a twofold advantage when a reduced number of measurements is available:

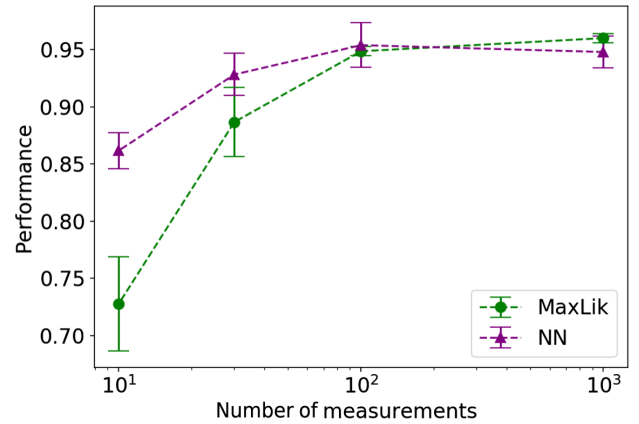


FIG. 2. Comparison between MaxLik and NN performance as a function of the input data size. In both approaches, this is quantified as the fraction of states correctly classified. Error bars reflect the variability over six repetitions.

the value of the performance is more robust and also its variability remains more confined, with an improvement of a factor of about 40% in standard deviation. We propose that the same method can also be applied with similar success on actual experimental data. The key ability is then to know a sensible approximation of the main characteristics of the experimental setup such as losses, the maximum level of squeezing, and the noise present in the state. By these, it becomes possible to train the NN using simulated data, and still obtain good accuracy in identifying the interesting feature of the experimental state.

We test these ideas with an experimental single-photon subtracted two-mode entangled state whose generation has been reported in [55]. The initial state before subtraction shows the entanglement correlations of an EPR state [56] between quadratures of two given frequency modes. The single-photon subtraction takes place on one of the two-given frequency modes, so that, in case of high purity and low losses, the first mode is left in a vacuum state while the second mode is left in a state with a negative Wigner function (see Supplemental Material [51] for more details). We train the network using 1000 simulated quadrature measurements, calibrated with the known imperfections of our experimental setup, namely thermal noise equal to 1.11 (directly linked with the purity of the initial Gaussian state), a factor of 12% of losses, and a maximum level of squeezing of 3 dB. The data are binned as before in order to build the input to the NN. Using the actual experimental data we can compute the complete Wigner function and the density matrix of the state using a maximal photon number of only  $N_{\text{ph}} = 3$ , giving  $W_{\text{min}} = -0.03$  (see Supplemental Material [51]). We now take 15 000 experimental quadratures, arranged in  $15 \times 2$  histograms, and we feed them to the NN. Our algorithm is able to detect the negativity present in the experimental state. Since imperfections limit the minimum of the Wigner function, no cutoff, i.e.,  $C = 0$ , is used in this case.

The characteristics of the experiment can be obtained within a certain accuracy and precision. Misrepresentation of the experimental parameters may lead to a failure of the NN classification. On the other hand, these networks are known to be able to work reliably even in the presence of noise. This suggests that discrepancies between actual and simulation parameters can be tolerated.

We gathered evidence of such resilience by testing the consistency of the network as we introduce extra losses in the experimental data. We monitored the transition to a positive minimum value of the Wigner function as the losses increase. This can be simulated replacing a fraction of the quadrature data with data sampled from the vacuum state. We used the MaxLik method as a benchmark: it gives the results in Fig. 3, that shows  $W_{\text{min}}$  as a function of the introduced losses. Each point represents the average over 100 sets of 1000 quadratures, extracted at random from the same state. The error bar is the standard deviation on these

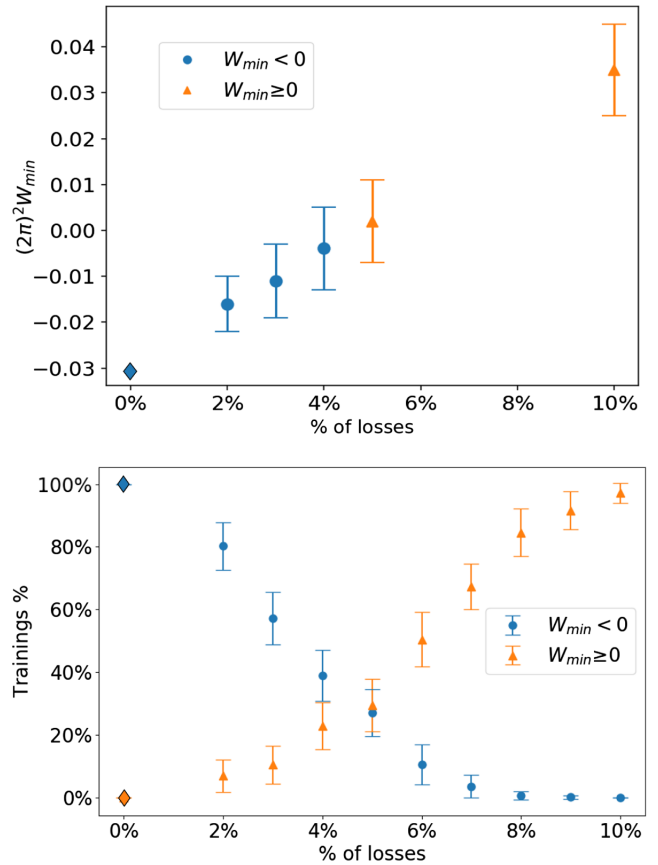


FIG. 3. Effect of extra losses. Top:  $(2\pi)^2 W_{\text{min}}$  for different percentages of introduced losses as estimated by the MaxLik algorithm. Error bars are computed over 100 repetitions. Bottom: Percentage of training sets identifying a negative state (blue circles) or a positive state (orange triangles) for at least 95 out of 100 inputs. Error bars are computed over 50 runs of the whole process. In both panels diamonds correspond to the experimental state, the circles (triangles) refer to states with  $W_{\text{min}} < 0$  ( $W_{\text{min}} \geq 0$ ).

replicas. As expected the negativity decreases from the initial value, reaching positive values above  $\sim 5\%$ .

We studied the same problem with the use of the NN algorithm. The training is operated generating the training data as before. Since we now underestimate the level of losses, the NN will not be optimized for this task, but we can rely on some robustness on its part. However, this property will strongly depend on the actual configuration reached with the training, hence on the random initial conditions. This implies that the test should be run over many instances of the training, so that one can assess the typical behavior of the NN.

For each value of the loss, we carry out our analysis by feeding quadrature binned histograms derived from 100 different repetitions of the lossy state to the NN (the same one as for MaxLik benchmark). We then record the fraction of states  $f_p$  with  $W_{\text{min}} \geq 0$ , and  $f_n$  with  $W_{\text{min}} < 0$  out of the 100 attempts. To account for the variability of the training,

we feed the same input data to 30 differently trained NNs, and register the corresponding values of  $f_p$  and  $f_n$ . In Fig. 3, we report the percentage of training instances for which  $f_n \geq 0.95$ , and the same for  $f_p \geq 0.95$ . These percentages do not sum to one, due to the presence of inconclusive results (see Supplemental Material [51]). The NN method is thus able to identify reliably state with a negative Wigner function even when the actual amount of loss does not correspond exactly to that in the simulated data used for its training. Negativity is witnessed up to approximately 5%: remarkably, this is the threshold value at which the MaxLik shows the transition to a positive Wigner function.

In conclusion, we have found that machine learning techniques can provide meaningful information on the Wigner negativity even when limited data are available.

The routine adoption of this method on large quantum cluster states is conditioned on the reliability of the data used for the training. We have shown that there exist specific instances in which the NN enjoys a given degree of flexibility. The described method relies in fact on numerical simulations, since no encompassing analytical description is possible. Investigation of more generic states is the scope of future works, as the use of NN appears as a promising avenue for studying the behavior of large quantum states for which state tomography becomes impractical. In particular, NN seem to be particularly useful to directly test specific quantum features of large multipartite systems without requiring the full reconstruction of the quantum states. This unleashes the potential of NN in quantum enabled technologies.

We thank I. Gianani for stimulating discussions. V.P. acknowledges financial support from the European Research Council under the Consolidator Grant COQCOoN (Grant No. 820079). N.T. acknowledges financial support of the Institut Universitaire de France.

---

\*valeria.cimini@uniroma3.it

- [1] H. J. Briegel, D. E. Browne, W. Dür, R. Raussendorf, and M. Van den Nest, *Nat. Phys.* **5**, 19 (2009).
- [2] M. Gu, C. Weedbrook, N. C. Menicucci, T. C. Ralph, and P. van Loock, *Phys. Rev. A* **79**, 062318 (2009).
- [3] S. Yokoyama, R. Ukai, S. C. Armstrong, C. Sornphiphatphong, T. Kaji, S. Suzuki, J.-i. Yoshikawa, H. Yonezawa, N. C. Menicucci, and A. Furusawa, *Nat. Photonics* **7**, 982 (2013).
- [4] J. Roslund, R. Medeiros de Araújo, S. Jiang, C. Fabre, and N. Treps, *Nat. Photonics* **8**, 109 (2014).
- [5] M. Chen, N. C. Menicucci, and O. Pfister, *Phys. Rev. Lett.* **112**, 120505 (2014).
- [6] W. Asavanant, Y. Shiozawa, S. Yokoyama, B. Charoensombutamon, H. Emura, R. N. Alexander, S. Takeda, J.-i. Yoshikawa, N. C. Menicucci, H. Yonezawa, and A. Furusawa, *Science* **366**, 373 (2019).
- [7] M. V. Larsen, X. Guo, C. R. Breum, J. S. Neergaard-Nielsen, and U. L. Andersen, *Science* **366**, 369 (2019).
- [8] J. Wenger, R. Tualle-Brouiri, and P. Grangier, *Phys. Rev. Lett.* **92**, 153601 (2004).
- [9] A. Zavatta, S. Viciani, and M. Bellini, *Science* **306**, 660 (2004).
- [10] J. S. Neergaard-Nielsen, B. M. Nielsen, C. Hettich, K. Mølmer, and E. S. Polzik, *Phys. Rev. Lett.* **97**, 083604 (2006).
- [11] A. Ourjoumtsev, R. Tualle-Brouiri, J. Laurat, and P. Grangier, *Science* **312**, 83 (2006).
- [12] V. Parigi, A. Zavatta, M. Kim, and M. Bellini, *Science* **317**, 1890 (2007).
- [13] M. Sasaki, M. Takeoka, and H. Takahashi, *Phys. Rev. A* **77**, 063840 (2008).
- [14] A. Ourjoumtsev, F. Ferreyrol, R. Tualle-Brouiri, and P. Grangier, *Nat. Phys.* **5**, 189 (2009).
- [15] E. Bimbard, N. Jain, A. MacRae, and A. I. Lvovsky, *Nat. Photonics* **4**, 243 (2010).
- [16] N. Namekata, Y. Takahashi, G. Fujii, D. Fukuda, S. Kurimura, and S. Inoue, *Nat. Photonics* **4**, 655 (2010).
- [17] J. S. Neergaard-Nielsen, M. Takeuchi, K. Wakui, H. Takahashi, K. Hayasaka, M. Takeoka, and M. Sasaki, *Phys. Rev. Lett.* **105**, 053602 (2010).
- [18] T. Gerrits, S. Glancy, T. S. Clement, B. Calkins, A. E. Lita, A. J. Miller, A. L. Migdall, S. W. Nam, R. P. Mirin, and E. Knill, *Phys. Rev. A* **82**, 031802(R) (2010).
- [19] N. Lee, H. Benichi, Y. Takeno, S. Takeda, J. Webb, E. Huntington, and A. Furusawa, *Science* **332**, 330 (2011).
- [20] W. Asavanant, K. Nakashima, Y. Shiozawa, J.-i. Yoshikawa, and A. Furusawa, *Opt. Express* **25**, 32227 (2017).
- [21] N. Biagi, L. S. Costanzo, M. Bellini, and A. Zavatta, *Phys. Rev. Lett.* **124**, 033604 (2020).
- [22] A. I. Lvovsky, P. Grangier, A. Ourjoumtsev, V. Parigi, M. Sasaki, and R. Tualle-Brouiri, *arXiv:2006.16985*.
- [23] A. I. Lvovsky, *J. Opt. B* **6**, S556 (2004).
- [24] D. Mogilevtsev, J. Řeháček, and Z. Hradil, *Phys. Rev. A* **75**, 012112 (2007).
- [25] A. Mari and J. Eisert, *Phys. Rev. Lett.* **109**, 230503 (2012).
- [26] R. Takagi and Q. Zhuang, *Phys. Rev. A* **97**, 062337 (2018).
- [27] F. Albarelli, M. G. Genoni, M. G. A. Paris, and A. Ferraro, *Phys. Rev. A* **98**, 052350 (2018).
- [28] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [29] G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
- [30] E. Magesan, J. M. Gambetta, A. D. Córcoles, and J. M. Chow, *Phys. Rev. Lett.* **114**, 200501 (2015).
- [31] V. Dunjko and H.-J. Briegel, *Rep. Prog. Phys.* **81**, 074001 (2018).
- [32] A. Lumino, E. Polino, A. S. Rab, G. Milani, N. Spagnolo, N. Wiebe, and F. Sciarrino, *Phys. Rev. Applied* **10**, 044033 (2018).
- [33] A. Hentschel and B. C. Sanders, *Phys. Rev. Lett.* **104**, 063603 (2010).
- [34] A. Rocchetto, S. Aaronson, S. Severini, G. Carvacho, D. Poderini, I. Agresti, M. Bentivegna, and F. Sciarrino, *Sci. Adv.* **5**, eaau1946 (2019).
- [35] S. Yu, F. Albarran-Arriagada, J. C. Retamal, Y.-T. Wang, W. Liu, Z.-J. Ke, Y. Meng, Z.-P. Li, J.-S. Tang, E. Solano, L.

- Lamata, C.-F. Li, and G.-C. Guo, *Adv. Quantum Technol.* **2**, 1800074 (2019).
- [36] A. A. Melnikov, H. Poulsen Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger, and H. J. Briegel, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1221 (2018).
- [37] M. Bukov, A. G. R. Day, D. Sels, P. Weinberg, A. Polkovnikov, and P. Mehta, *Phys. Rev. X* **8**, 031086 (2018).
- [38] I. Agresti, N. Viggianiello, F. Flamini, N. Spagnolo, A. Crespi, R. Osellame, N. Wiebe, and F. Sciarrino, *Phys. Rev. X* **9**, 011013 (2019).
- [39] V. Gebhart and M. Bohmann, *Phys. Rev. Research* **2**, 023150 (2020).
- [40] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, *Nat. Phys.* **14**, 447 (2018).
- [41] G. Torlai, B. Timar, E. P. L. van Nieuwenburg, H. Levine, A. Omran, A. Keesling, H. Bernien, M. Greiner, V. Vuletić, M. D. Lukin, R. G. Melko, and M. Endres, *Phys. Rev. Lett.* **123**, 230504 (2019).
- [42] E. Tiunov, V. Tiunova (Vyborova), A. Ulanov, A. Lvovsky, and A. Fedorov, *Optica* **7**, 448 (2020).
- [43] K. P. Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press, Cambridge, 2012).
- [44] V. Cimini, I. Gianani, N. Spagnolo, F. Leccese, F. Sciarrino, and M. Barbieri, *Phys. Rev. Lett.* **123**, 230502 (2019).
- [45] A. Macarone-Palmier, E. Kovlakov, F. Bianchi, D. Yudin, S. Straupe, J. Biamonte, and S. Kulik, *npj Quantum Inf.* **6**, 20 (2020).
- [46] T. Fösel, P. Tighineanu, T. Weiss, and F. Marquardt, *Phys. Rev. X* **8**, 031084 (2018).
- [47] M. August and X. Ni, *Phys. Rev. A* **95**, 012335 (2017).
- [48] M. Walschaers, C. Fabre, V. Parigi, and N. Treps, *Phys. Rev. Lett.* **119**, 183601 (2017).
- [49] M. Walschaers, C. Fabre, V. Parigi, and N. Treps, *Phys. Rev. A* **96**, 053835 (2017).
- [50] M. Walschaers, Y.-S. Ra, and N. Treps, *Phys. Rev. A* **100**, 023828 (2019).
- [51] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.125.160504> for more details on the generation of simulated data and NN performances, which includes Refs. [52,53].
- [52] D. P. Kingma and J. Ba, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [53] Y.-S. Ra, C. Jacquard, A. Dufour, C. Fabre, and N. Treps, *Phys. Rev. X* **7**, 031012 (2017).
- [54] A. Royer, *Phys. Rev. A* **15**, 449 (1977).
- [55] Y.-S. Ra, A. Dufour, M. Walschaers, C. Jacquard, T. Michel, C. Fabre, and N. Treps, *Nat. Phys.* **16**, 144 (2020).
- [56] Z. Y. Ou, S. F. Pereira, H. J. Kimble, and K. C. Peng, *Phys. Rev. Lett.* **68**, 3663 (1992).