

## Multiclass classification of dephasing channels

Adriano M. Palmieri<sup>1,\*</sup> Federico Bianchi<sup>2,†</sup> Matteo G. A. Paris<sup>3,4,‡</sup> and Claudia Benedetti<sup>3,§</sup>

<sup>1</sup>*Skolkovo Institute of Science and Technology, 121205 Moscow, Russia*

<sup>2</sup>*Department of Marketing, Bocconi University, 20136 Milan, Italy*

<sup>3</sup>*Quantum Technology Lab, Dipartimento di Fisica “Aldo Pontremoli, Università degli Studi di Milano, 20133 Milano, Italy*

<sup>4</sup>*Istituto Nazionale di Fisica Nucleare, Sezione di Milano, 20133 Milano, Italy*



(Received 10 August 2021; revised 19 October 2021; accepted 26 October 2021; published 10 November 2021)

We address the use of neural networks (NNs) in classifying the environmental parameters of single-qubit dephasing channels. In particular, we investigate the performance of linear perceptrons and of two nonlinear NN architectures. At variance with time-series-based approaches, our goal is to learn a discretized probability distribution over the parameters using tomographic data at just two random instants of time. We consider dephasing channels originating either from classical  $1/f^\alpha$  noise or from the interaction with a bath of quantum oscillators. The parameters to be classified are the color  $\alpha$  of the classical noise or the Ohmicity parameter  $s$  of the quantum environment. In both cases, we find that NNs are able to exactly classify parameters into 16 classes using noiseless data (a linear NN is enough for the color, whereas a single-layer NN is needed for the Ohmicity). In the presence of noisy data (e.g., coming from noisy tomographic measurements), the network is able to classify the color of the  $1/f^\alpha$  noise into 16 classes with about 70% accuracy, whereas classification of Ohmicity turns out to be challenging. We also consider a more coarse-grained task and train the network to discriminate between two macroclasses corresponding to  $\alpha \leq 1$  and  $s \leq 1$ , obtaining up to 96% and 79% accuracy using single-layer NNs.

DOI: [10.1103/PhysRevA.104.052412](https://doi.org/10.1103/PhysRevA.104.052412)

### I. INTRODUCTION

Noisy intermediate-scale quantum devices are currently available [1], but the technology does need to overcome this intermediate stage to allow their full scalability [2]. In this endeavor, the ability to fully characterize the unavoidable impact of the environment on the quantum system of interest is a cornerstone of the implementation of quantum information tasks [3,4], including redundancy calibration or high-fidelity quantum gates. In particular, high-performance quantum simulators [5], sensors [6], and computers [7] rely on the precise knowledge of the surrounding environment [8], which is usually modeled as a complex system, made of many and uncontrollable degrees of freedom.

Dephasing channels describe the open dynamics of quantum systems that do not exchange energy with their surrounding environments, but are nevertheless affected in their coherence [9]. If a qubit goes through a dephasing channel, the environment, in particular the spectral features of the noise, strongly affects the reduced dynamics of the qubit. Upon accessing the quantum state of the qubit, it is thus possible to infer the spectral parameters of the noise. The use of a single qubit as a quantum probe for environmental parameters has been extensively analyzed [10–15]. However, in all these studies the qubit probe must be measured at a very specific

time in order to achieve high precision. In this paper, we go beyond this constraint and design a classification technique which requires tomographic data at just two random instants of times, within a fixed time window.

We consider two paradigmatic models of dephasing. The first is the celebrated  $1/f^\alpha$  classical noise originating from the interaction with a set of classical fluctuators [16]. The second, referred to as quantum dephasing noise, comes from the interaction of the qubit with a bath of bosonic oscillators [3]. The  $1/f^\alpha$  noise is a ubiquitous model of noise, which affects both quantum and classical systems. The value of the exponent  $\alpha$  is usually referred to as the color of the noise and depends on the details of the system under investigation [17–23]. If the environment is a collection of bosonic modes, the spin-boson model is instead suitable to describe the dephasing dynamics of a single qubit interacting with the quantum bath.

Machine learning (ML) for physics aims at discovering and implementing data-driven-adaptive approaches to solve physical problems. Machine learning is finding a widespread, growing number of applications in quantum physics including quantum tomography [24–29], metrology, sensing and probing [30–33], entanglement classification [34,35], quantum thermodynamics [36], quantum control [37], communication [38], and renormalization group. In turn, attention has been devoted also to open quantum system dynamics and noise mitigation. Restricted Boltzmann machines and neural quantum states have found several applications in dynamical systems [39–41] undergoing dephasing, and more recently autoregressive models over experimental data have been discussed [42]. Other techniques have been considered also for simulation of open quantum systems, as convex

\*adriano.palmieri@skoltech.ru

†f.bianchi@unibocconi.it

‡matteo.paris@fisica.unimi.it

§claudia.benedetti@unimi.it

optimization for Lindbladian unsupervised tensor network learning [26], adaptive regression strategy [43], and deep evolutionary approaches. In contrast, long short-term memory architecture has been shown to offer large improvements in the arduous problems of noise mitigation and spectrum understanding [44].

In this paper, we address discrimination (i.e., classification) problems for single-qubit dephasing channels, originating from the interaction with either a classical or a quantum environment. Specifically, we want the network to distinguish among a discretized set of spectral parameters, i.e., the color of the noise for classical dephasing and the Ohmicity for the quantum baths. Our approach relies on the fact that deep neural networks (NNs) are effective at learning functions or at replicating the dynamical parameters thereof [45–49]. Our goal is to infer the environmental parameters by feeding the network with tomographic data taken at (only) two random instants of time. In this way, we lift the constraint of the qubit state being measured at a specific time. We show that NNs are able to perfectly discriminate between different values of the spectral parameters as long as we use noiseless data. Noisy data instead reduce the performance. More specifically, classification of the color of classical noise is still feasible, whereas boson baths are more difficult to classify. Higher levels of accuracy are still accessible by reframing the problem into a two-class discrimination task, which is of interest in several potential applications [10,36,50].

The paper is structured as follows. In Sec. II we introduce the physical model for a dephasing qubit stemming from either a classical or a quantum environment. In Sec. III we describe the data-set preparation and we briefly review the learning models' salient characteristics and metrics. In Sec. IV we present our results for two main families of data: noiseless and noisy data. We address the accuracy and macro  $F1$  score to evaluate the classification performance. Section V summarizes and discusses the results. All codes for data generation and numerical experiments are available [51].

## II. MODEL

A qubit dephasing channel is described by the quantum map

$$\rho(t) = \frac{1 + \Lambda(t)}{2} \rho(0) + \frac{1 - \Lambda(t)}{2} \sigma_z \rho(0) \sigma_z, \quad (1)$$

where  $\sigma_z$  is a Pauli matrix,  $\rho(0)$  is the initial density matrix of the qubit, and  $\Lambda(t)$  is referred to as the dephasing function. The explicit expression of  $\Lambda(t)$  depends on the specific microscopic model describing the system-environment interaction. The dephasing map in Eq. (1) may originate from different mechanisms. Here we consider two of them, which are relevant to different branches of quantum technology, corresponding to situations where the qubit interacts with a classical fluctuating environment or with a bosonic bath [52,53]. If the qubit interacts with a classical bistable fluctuator having random switching rates  $\gamma$ , the dephasing coefficient takes the expression [54]

$$\Lambda_c(t, \alpha) = \int_{\gamma_1}^{\gamma_2} G(t, \gamma) p_\alpha(\gamma) d\gamma, \quad (2)$$

with

$$G(t, \gamma) = e^{-\gamma t} \left[ \cosh(\delta t) + \frac{\gamma}{\delta} \sinh(\delta t) \right], \quad (3)$$

where  $\delta = \sqrt{\gamma^2 - 4}$  and  $t$  is dimensionless time. If the probability distribution  $p_\alpha(\gamma)$  is defined as

$$p_\alpha(\gamma) = \begin{cases} \frac{1}{\gamma \ln(\gamma_2/\gamma_1)}, & \alpha = 1 \\ \frac{\alpha-1}{\gamma^\alpha} \left[ \frac{(\gamma_1 \gamma_2)^{\alpha-1}}{\gamma_2^{\alpha-1} - \gamma_1^{\alpha-1}} \right], & \alpha \neq 1, \end{cases} \quad (4)$$

the overall spectrum of the noise corresponds to a  $1/f^\alpha$  distribution. As mentioned above, the parameter  $\alpha$  is referred to as the color of the noise, since it determines the weight of the different frequencies  $f$ .

If the qubit is interacting with a bath of quantum oscillators at zero temperature, the dephasing function has an exponential decaying form [3]

$$\Lambda_q(t, s) = e^{-\Gamma(t, \omega_c, s)}, \quad (5)$$

with

$$\Gamma(t, \omega_c, s) = \begin{cases} \frac{1}{2} \ln(1 + \omega_c^2 t^2), & s = 1 \\ \left(1 - \frac{\cos[(s-1) \arctan(\omega_c t)]}{[1 + \omega_c^2 t^2]^{(s-1)/2}}\right) \bar{\Gamma}[s-1], & s \neq 1, \end{cases} \quad (6)$$

where  $\bar{\Gamma}[x] = \int_0^\infty t^{x-1} e^{-t} dt$ . The quantity  $\omega_c$  is usually referred to as the cutoff frequency of the bath and  $s$  is the so-called Ohmicity parameter. The Ohmicity governs the behavior of the spectrum at low frequency and identifies three regimes: For  $s < 1$  the spectrum is referred to as sub-Ohmic,  $s = 1$  represents the Ohmic case, and  $s > 1$  corresponds to a super-Ohmic environment.

Our aim is to infer the properties of the environment by exploiting ML techniques. Since the estimation task (regression in ML jargon) is challenging because it requires inferring the value of the parameter belonging to a continuous interval, here we address the problem of classification, i.e., the ability to discriminate among  $m$  different values, called classes, for the estimated parameter. In particular, we want to discriminate different values of the parameters  $\alpha$  and  $s$  in the classical and quantum cases, respectively. In order to overcome the constraint of measuring the qubit at a very precise instant of time, which could be a burden for the experimentalists, we consider the situation where we have little or no control over the measurement time. We thus feed the network only with two qubit states at two random instants of time.

## III. LEARNING TASK AND MODELS

### A. Data description

For a generic state  $|\phi\rangle = \beta_1|0\rangle + \beta_2|1\rangle$ , randomly sampled from the Haar distribution in  $\mathcal{C}^2$ , the associated density matrix is  $\rho = |\phi\rangle\langle\phi|$  and the state evolved in a dephasing channel is given by Eq. (1), i.e.,

$$\rho(t, v_l) = \begin{pmatrix} |\beta_1|^2 & \beta_1 \beta_2^* \Lambda_l(t, v_l) \\ \beta_1^* \beta_2 \Lambda_l(t, v_l) & |\beta_2|^2 \end{pmatrix}, \quad (7)$$

where  $\Lambda_l(t, v_l)$  is the dephasing function with  $l = c, q$  and we identify the noise parameters  $v_c = \alpha$  and  $v_q = s$ . In order to feed the network, we need to know the density matrix

$\rho(t)$  at some time. To this aim we should consider a reliable tomographic measurement, e.g., a symmetric informationally complete (SIC) positive-operator-valued measure (POVM), i.e., the set  $\{\mathcal{M}_k = |\mathcal{M}_k\rangle\langle\mathcal{M}_k|, \sum_{k=1}^{d^2} \mathcal{M}_k = d\mathbb{I}, \text{rank}\mathcal{M}_k = 1\}_{k=1}^{d^2}$ , where the  $|\mathcal{M}_k\rangle$  are  $d$ -dimensional states with the Hilbert-Schmidt product  $\text{Tr}(\mathcal{M}_k\mathcal{M}_j) = \frac{1+d\delta_{kj}}{d+1}$  [55,56]. Symmetric informationally complete operators are maximally efficient at estimating the quantum state. For this very reason they are largely employed in tomography and cryptography [55,56] and this makes them good candidates for us as well.

To build our data sets, we leverage the injective mapping between the convex set of density operators and the set of probability distributions over  $d^2$  outcomes  $p_k(t) = \text{Tr}[\mathcal{M}_k\rho(t)]$  [57]. A quantum state of a  $d$ -dimensional system can be thus expressed as a  $d^2$ -dimensional real vector on a regular simplex  $S$  immersed in  $\mathcal{R}^{d^2-1}$ ,  $\|\mathbf{p}_{\Lambda_l(t, v_l)}\rangle\rangle = (p_1, p_2, \dots, p_{d^2})_{\Lambda_l}^T$ , with  $l = (c, q)$  (i.e., classical or quantum). Indeed, the subset of the points representing density matrices in the  $d^2 - 1$  simplex corresponds to the  $d^2 - 1$  generalized Bloch sphere vectors [58]. In particular, we consider the qubit systems, for which the SIC basis can be directly expressed in term of Pauli matrices [57].

For a given number  $N$  of initial density matrices and a fixed dimensionless-time window  $D$  of arbitrary width, we consider the set of pairs  $\{(t_1, t_2), t_1, t_2 \in D\}$ . The input states  $\mathbf{x}_{v_l}$  for the networks are devised as

$$\{t_1, t_2\} \rightarrow \{\mathbf{x}_{v_l}\} = \{\|\mathbf{p}_{v_l}(t_1)\rangle\rangle \oplus \|\mathbf{p}_{v_l}(t_2)\rangle\rangle\}.$$

We create a selection of  $m$  arbitrarily chosen classes of bath parameters  $\vec{v}_c = (\alpha_1, \dots, \alpha_m)$  and  $\vec{v}_q = (s_1, \dots, s_m)$ , for the classical and quantum environments, respectively. To build our data set  $\mathcal{D}_l$  we collect and shuffle the inputs  $\mathbf{x}_{v_l}$  for the selected values of the parameter  $v_l \in \vec{v}_l$ ,

$$\mathcal{D}_l = \bigcup_{k=1}^N \bigcup_{v_l \in \vec{v}_l} \{\mathbf{x}_{v_l}, v_l\}_{\rho_k}, \quad l = c, q, \quad (8)$$

where  $\rho_k$  is the  $k$ th randomly generated density matrix of Eq. (7) and we use the shorthand notation  $v_l \in \vec{v}_l$  to indicate the different elements of the set  $\vec{v}_l$ .

### B. Learning models

Our task is to learn the function

$$f : S \oplus S \rightarrow \mathcal{R}^m, \quad \mathbf{x}_{v_l} \rightarrow \mathbf{y}_{\text{out}}$$

mapping the input  $\mathbf{x}_{v_l}$  to the vector  $\mathbf{y}_{\text{out}}$ , i.e., the probability distribution of the parameter  $v_l$  over the  $m$  classes. The class with highest probability is then the estimated value of the bath parameter. The explicit equations for a feedforward neural network of  $n - 2$  hidden layer are given by

$$\begin{aligned} h_1 &= \sigma(W_1\mathbf{x}_{v_l} + b_1), \\ &\vdots \\ h_n &= \sigma(W_{n-1}h_{n-1} + b_{n-1}), \\ \mathbf{y}_{\text{out}} &= \text{softmax}(h_n), \end{aligned} \quad (9)$$

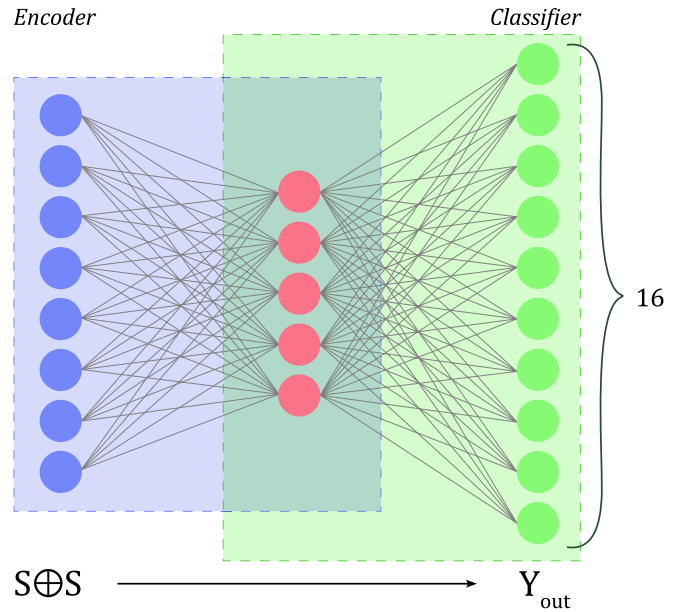


FIG. 1. Schematic diagram of our NN-based classification methods. We address the use of a linear perceptron, a single-inner-layer 8-5-16 feedforward neural network (in the picture), and a 8-30-30-30-30-16 feedforward NN.

where  $W_i$  ( $i = 1, \dots, n - 1$ ) are the weights that the network is going to learn during the training phase,  $h_i$  is the  $i$ th network layer,  $b_i$  is the bias vectors, and  $\sigma$  is the activation function, which is the rectified linear unit (ReLU) function  $\text{ReLU}(x) = \max(0, x)$ . When the network maps a function from  $h_{i-1} \rightarrow h_i$  with  $\dim(h_{i-1}) < \dim(h_i)$ , the network is compressing information by mapping the input into a lower space. Otherwise the network is mapping the input to a space of higher dimension.

We use three different models for our analysis: a simple linear perceptron (LP), which is just an input and an output layer; a single inner-layer 8-5-16 feedforward neural network, which is an encoder with a stacked classifier; and an 8-30-30-30-30-16 feedforward NN. By construction, the LP model will just be able to provide a linear discrimination between the classes [59]. As usual in machine learning classification tasks, we employ the softmax activation function on the output layer

$$\text{softmax}(y_i^{\text{out}}) = \frac{e^{y_i^{\text{out}}}}{\sum_j^{d^2} e^{y_j^{\text{out}}}}, \quad (10)$$

which takes the last layer's output and turns it into a probability distribution over the target classes (see Fig. 1). This setup allows us to solve our multinomial logistic regression problem. As a loss function to train our network, we use the categorical cross entropy, which quantifies how far the prediction of the models is from the correct class.

In order to assess the predictions made by our classification models, we use two performance metrics: accuracy and the macro  $F1$  score. Both are built from the confusion matrix  $C$ , which is an  $m \times m$  table where the entry  $C_{jk}$  contains the number of events where the  $j$ th actual class leads to a predicted  $k$ th class. Given a data set of  $N$  elements, the first measure is the accuracy, naturally defined as  $A = \frac{1}{N} \sum_j C_{jj}$ . However, we cannot rely on this measure alone, especially when it comes

to comparing different models. We need a measure that takes into account the imbalances in the class distribution [60]. In turn, the precision is defined as the number of corrected predicted instances out of all predicted instances. For each actual class  $j$ , precision is defined as  $P_j = \frac{C_{jj}}{\sum_k C_{jk}}$ . The recall is instead defined as the number of correctly predicted instances out of the number of true ones. For each predicted class  $k$ , we have  $R_k = \frac{C_{kk}}{\sum_j C_{jk}}$ . Finally, we define the  $F1$  score for each class as the harmonic mean  $F1_j = 2 \frac{P_j R_j}{P_j + R_j}$ . The overall metric we use to assess performance is then the macro averaged  $F1$  score

$$F1 = \frac{1}{m} \sum_{j=1}^m F1_j, \quad (11)$$

which we just refer to as the  $F1$  score hereafter.

## IV. RESULTS

### A. Data-set preparation

For each data set, we employ  $N = 2500$  Haar-generated density matrices undergoing the dephasing channel of Eq. (7). Then we create 1000 time-ordered couples  $\{(t_1, t_2)\}$  using 110 random times, uniformly distributed in the selected time window  $D$ . In addition, we set the following physical parameters: The frequency parameter for colored noise lies in the interval  $[\gamma_1, \gamma_2] = [10^{-4}, 10^4]$ , whereas the cutoff frequency for the spin-boson model is set to  $\omega_c = 1$ . All the functions are implemented using a PYTHON library for real and complex floating-point arithmetic with arbitrary precision [61].

We create different data sets for our experiments as follows. First, we want to test how our models work with noiseless data. This is relevant in order to gauge the absolute performance of our learning models. Then we introduce Gaussian measurement noise into our data  $\mathbf{x}_{v_l}$ . In particular, we assume that each  $p_i(t)$  is affected by a random Gaussian perturbation, taken from the normal distribution  $\mathcal{N}(0, 0.01)$ .

We then focus on two different classification tasks. First, for noiseless and noisy data, we establish  $m = 16$  classes  $\vec{v}_l$ . In particular, we divide the interval  $\alpha = [0.5, 2]$  into 16 equally spaced classes, while for the Ohmic parameter  $s$  we picked unequally spaced ones.<sup>1</sup> Then we reshape our noisy data into two coarse-grained classes  $\vec{\mu}_l = (v_l \leq 1, v_l > 1)$  with  $l = c, q$ . The two-class data sets allow us to assess whether the network is able to classify data according to a leading physical property of the environment, i.e., being more or less colored for classical noise or being sub- or super-Ohmic for the quantum bath.

All data sets consist of  $15.3 \times 10^5$  shuffled-input training data and  $3.6 \times 10^5$  shuffled-input data for (both) validation and the test. The time windows are chosen to cover the relevant dynamics of the system such that the curves  $\Lambda_l(t, v_l)$  overlap as little as possible for different values of the parameter  $v_l$ . The selected time windows are  $D = [0.2, 3.14]$  for noiseless colored data,  $D = [0.4, 2.15]$  for noisy colored

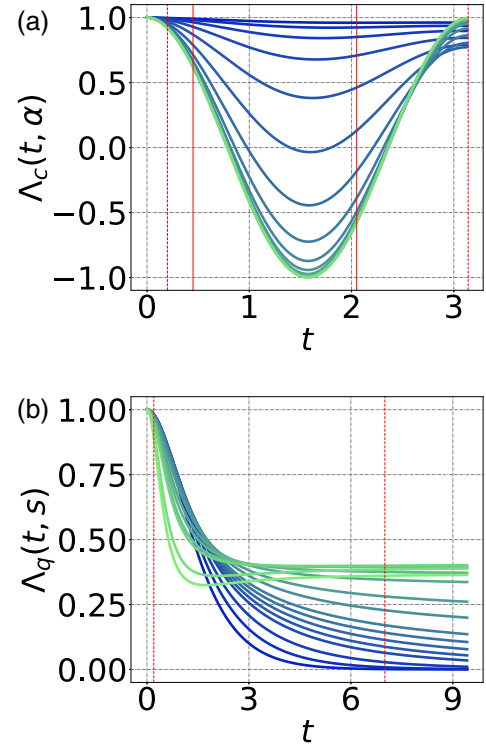


FIG. 2. Dephasing terms  $\Lambda_c(t, \alpha)$  and  $\Lambda_q(t, s)$  for (a) classical colored noise and (b) Ohmic quantum noise. For classical noise we show  $\Lambda_c(t, \alpha)$  for  $\alpha$  in the range  $[0.5, 2]$  (from dark blue to light green curves). For quantum noise we show  $\Lambda_q(t, s)$  for  $s$  in the range  $[0.1, 3]$  (from dark blue to light green curves). The vertical lines in both panels refer to the time windows  $D$  used in the data generation stage (chosen to contain the relevant dynamics of the system). For classical noise we have  $D = [0.2, 3.14]$  (dashed vertical lines) in the noiseless case and  $D = [0.4, 2.15]$  (solid vertical lines) for the noisy scenario. For quantum noise we have  $D = [0.2, 7]$  in both cases.

data, and  $D = [0.2, 7]$  for the bosonic bath (noiseless and noisy cases; see Fig. 2). We use pure initial states, again to test the model in ideal conditions, and mixed ones with purity  $\text{Tr}[\rho(0, v_l)^2] = 0.72$ , to mimic realistic noisy conditions. The mixed states are obtained by depolarizing the initial pure states.

### B. Noiseless measurement data

The noise-free scenario aims to test whether, and to what extent, a LP or a NN are able to solve the classification problem using 16 classes for the dephasing parameters. For this task the NN architecture is set to 8-5-16 with ReLU activation. For both NN and LP we use the following specifications: batch size of 300 for training and the data set split into test, train, and validation, with the validation data set monitoring the early-stopping function, set at one, to avoid overfitting [62].

For classical colored noise the problem is linear. Our experiments confirm that the problem may be easily solved with both (LP and NN) models, achieving 100% accuracy (and  $F1$  score) with either pure or mixed initial states. In the Ohmic case, we observe a difference in the performance of the two classification models. The LP model cannot achieve perfect classification using pure initial state states, while a

<sup>1</sup>The Ohmic classes are  $s = \{0.1, 0.3, 0.5, 0.7, 0.8, 0.9, 1.0, 1.1, 1.3, 1.5, 1.8, 2.0, 2.2, 2.3, 2.8, 3\}$ .

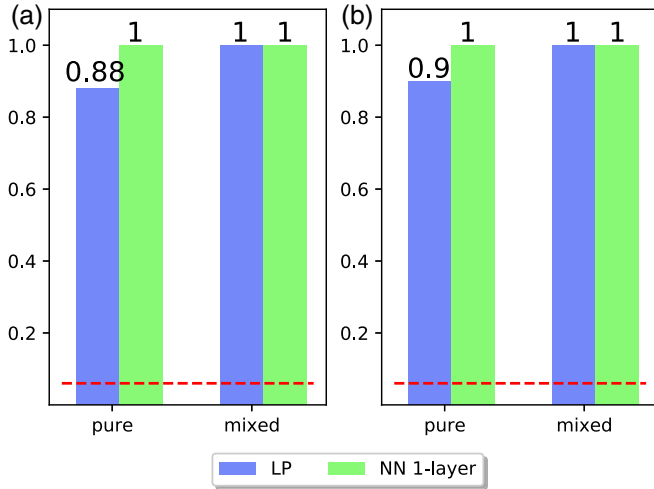


FIG. 3. (a) The  $F1$  score and (b) accuracy for Ohmic dephasing classification for the 16-class problem. For the linear projector, the  $F1$  score and accuracy values are [0.88, 0.9] and [1, 1] for pure and mixed probes, respectively. The NN model achieves perfect classification for both types of probes. The red dashed line marks the random guess for the  $F1$  score and accuracy.

single-inner-layer NN achieves perfect environment discrimination. Results are summarized in Fig. 3. We anticipate that the small differences between the performance of the different models will be amplified in the realistic scenario, i.e., when measurement noise is added.

Additionally, we want to understand how our network is generalized with respect to novel data, by training and testing the network with data coming from different regions of the Bloch sphere, e.g., states with a constraint in the component along the  $z$  axis. The train and the test data sets are given by

$$D_{v_l} = \{\mathbf{x}_{v_l} : |\text{Tr}[\sigma_z \rho(0, v_l)]| < 0.7\} \quad (\text{train and validation}),$$

$$\bar{D}_{v_l} = \{\mathbf{x}_{v_l} : 0.7 < |\text{Tr}[\sigma_z \rho(0, v_l)]| < 0.8\} \quad (\text{test}).$$

In this way, we can ensure that the test states are always far enough and separated from our train or validation set. In this case, we use the 8-5-16 architecture and the experiment aims to assess whether the network is truly inferring the bath parameters from the  $\{\mathbf{x}_{v_l}\}$  inputs or we are just overfitting. Results for the single-layer NN confirm that our model is able to classify novel data. Indeed, we obtain perfect scoring for accuracy and macro  $F1$  (both 100%). This provides further evidence in favor of our basic hypothesis: The network is learning a function able to catch the hallmark of the baths.

Finally, we also consider a situation in which we train the network over initial pure states and test it with mixed initial states. Specifically, the data sets are

$$D_{v_l} = \{\mathbf{x}_{v_l} : \text{Tr}[\rho^2(0, v_l)] = 1\} \quad (\text{train}),$$

$$\bar{D}_{v_l} = \{\mathbf{x}_{v_l} : \text{Tr}[\rho^2(0, v_l)] = 0.72\} \quad (\text{test}).$$

The test data set is obtained by depolarizing the initial pure states. We consider the 16-class problem and use the 8-5-16 model. The batch size is equal to 300 and early stopping is set to one. The above single-layer NN is indeed able to

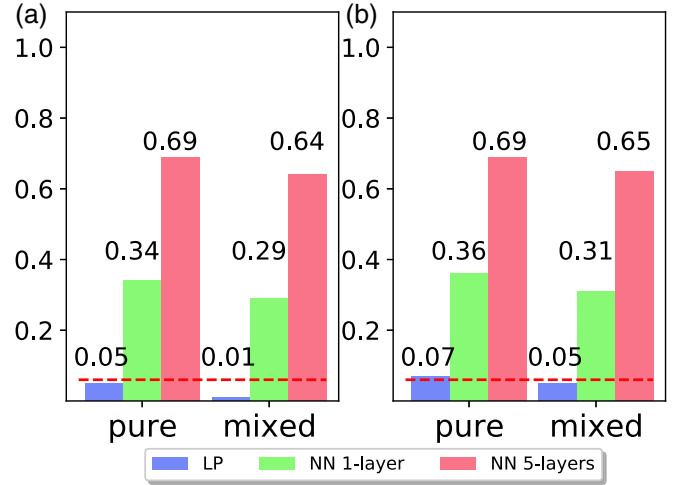


FIG. 4. (a) The  $F1$  score and (b) accuracy for noisy classification for colored dephasing using 16 classes. For this task the three different models increase in complexity stepwise, offering us a global view of the increasing level of difficulty of the task. The LP model  $F1$  score and accuracy values are [0.05, 0.07] for pure initial states and [0.01, 0.05] for mixed ones. The single-layer NN model values are [0.34, 0.36] for initial pure states and [0.29, 0.31] for mixed. With the five-layer model NN we achieve [0.69, 0.695] for pure states and [0.64, 0.65] for mixed ones. The red dashed line marks the random guess for the  $F1$  score and accuracy.

discriminate the bath parameters, obtaining an accuracy  $\sim 100\%$  for the color  $\alpha$  and  $\sim 88\%$  for the Ohmicity  $s$ .

### C. Noisy measurement data

If data are affected by Gaussian noise, the performance of LP and NN models degrades. In particular, LPs fail to classify parameters due to the strong nonlinearity of the problem. For the 16-class problem, we employ both the 8-5-16 NN model and 8-30-30-30-30-30-16 NN architecture. Batch size and early stopping parameters are left unchanged. Similar architectures have been exploited in other applications facing noisy data sets [27, 63].

Results for the colored noise classification case are shown in Fig. 4. The LP model apparently fails at classifying noise, reaching at most the order of the random guess ( $\sim 0.06$ ). The nonlinear NN architectures can instead achieve higher values of  $F1$  score and accuracy by increasing the number of layers. We find that for the 16-class problem, the noisy case requires wider inner layers. In other words, data are mapped into a space of higher dimension than their initial one. In order to illustrate the model statistics, we report the confusion matrix for the 16-class colored classical noise (mixed initial state) in Appendix B. The confusion matrix suggests that with the addition of the noise the models tend to confuse some of the classes. The classification task is instead challenging in the Ohmic noisy case, where the best performer is the five-inner-layer model and is able to achieve an accuracy of 0.25, corresponding to an  $F1$  score equal to 0.22.

Our results show that the classification task for the Ohmic noise parameters is in general a harder problem. This behavior

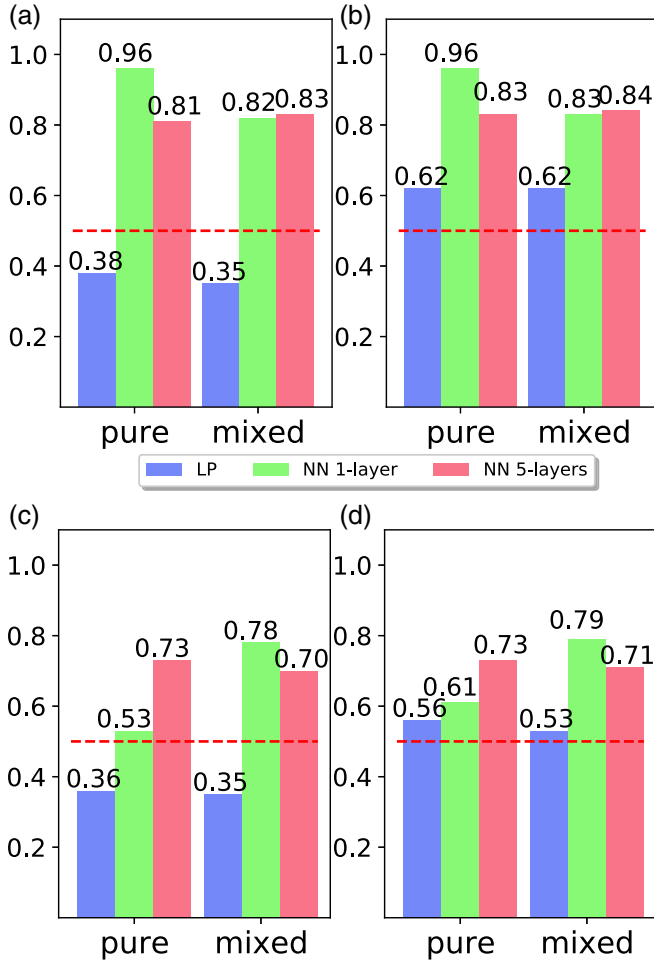


FIG. 5. (a) and (c) The  $F1$  score and (b) and (d) accuracy for the two-class classification problem involving (a) and (b) colored and (c) and (d) Ohmic noise. For colored noise, the largest values of the  $F1$  score and accuracy are obtained by the single-layer NN (0.96%) for pure states, while for mixed states the two models are almost equivalent. For Ohmic dephasing, we obtain the best results with the five-layer NN with pure states and with the single-layer NN for mixed states. The red dashed line denotes the random guess threshold for the  $F1$  score and accuracy.

is due to the fact that the 16 classes in the Ohmic case are closer to each other, in terms of SIC POVMs.

#### Two-class classification problem

In order to investigate the problem from a physically driven perspective, we also consider the case where the network is given the problem to distinguish between the  $\alpha \leq 1$  regimes for the classical noise and sub- or super-Ohmic dephasing ( $s \leq 1$ ) for the quantum bath. Results are illustrated in Fig. 5, which shows how the problem is nonlinear and highly dependent on the classical or quantum nature of the environment. For classical colored noise with pure initial states, the best performer is the single-layer NN, able to achieve 96% in both accuracy and  $F1$ . For mixed states, the corresponding values of  $F1$  and accuracy are 83% and 82%, just 1% less with respect to the five-layer NN, however obtained with a simpler architecture. Indeed, the model deals with the noise

perturbations efficiently. This is further evidence that the network captures the physically identifying characteristic of the environment parameters.

For the quantum bath with mixed initial states, a 8% advantage is gained by the one-layer model. In contrast, for pure initial states the five-layer model leads to a higher accuracy and  $F1$  score. This behavior may be understood by considering the data sparseness. Indeed, if we add an  $L1$  regularization term [64] the network may reach up to 81% accuracy and 80% for  $F1$  score. Overall, the single-layer network turns out to be the most efficient model for tackling the two-class task.

#### V. CONCLUSION

In this paper, we have analyzed the use of neural networks to classify the noise parameter of dephasing channels. We have considered channels originating either from classical  $1/f^\alpha$  noise or from the interaction with a bath of quantum oscillators. Our strategy involves the use of a single qubit to probe the channel and requires, to train the classifying network, the knowledge of the qubit density matrix at just two random instants of time within a fixed time window.

First we showed that the qubit dephasing problem may be recast into a linear discrimination task, solvable with minimal resources, e.g., basic ML models. In particular, we showed that our network is able to exactly classify spectral parameters into 16 classes using noiseless data. Then, in order to assess the performance in a more realistic scenario, we added Gaussian noise to data. For the 16-class task linearity is lost for both colored and Ohmic dephasing. The classical noise classification is still feasible with a five-layer NN model, leading to relatively high accuracy, whereas quantum dephasing classification is prone to fail. We also trained the network to discriminate between two macro classes, involving either  $\alpha \leq 1$  or  $s \leq 1$ . In these cases the single-layer model outpaces the five-layer one, reaching levels of  $F1$  and accuracy of 96% for the colored and 79% for the Ohmic environment.

Our results confirm the feasibility of the approach, i.e., the use of neural networks in classifying the environmental parameters of single-qubit dephasing channels. This is remarkable in view of the small number of experimental measurements required (the state at two random instants of time) to use the network. Our results show that, in general, bosonic baths are more demanding to discriminate compared to classical  $1/f^\alpha$  noise and that NNs may not be able to achieve multiclass classification of quantum noise. For the simpler task of two-class classification a single-layer NN is able to effectively discriminate noise, either classical or quantum.

Our results show that quantum probing may be effectively enhanced by the use of NNs, which provide a way to effectively extract information from tomographic data, and pave the way for future investigations devoted to more advanced and in-depth architectures for realistic use cases, e.g., to reveal the reason for the bosonic bath hardships. Future outlooks also include the possibility to give a solid contribution to the task of environmental characterization, by overcoming the assumption of the decoherence model and investigate whether it is possible to train a network without any previous knowledge about the noise. This is a big challenge that needs to be

addressed in steps, each requiring a tailored solver. From this perspective, this work sets the ground to attack this problem.

### APPENDIX A: SINGLE-QUBIT SIC POVM

Symmetric informationally complete positive-operator-valued measurements are described by the set of operators  $\{\mathcal{M}_k = |M_k\rangle\langle M_k|\}_{k=1}^{d^2}$  that satisfy the following properties:  $\sum_{k=1}^{d^2} \mathcal{M}_k = d \mathbb{I}$  and  $\text{Tr}(\mathcal{M}_k \mathcal{M}_j) = \frac{1+d\delta_{kj}}{d+1}$ . The  $|M_k\rangle$  are normalized  $d$ -dimensional vectors. For a single qubit, i.e.,  $d = 2$ , we choose

$$\begin{aligned} |M_1\rangle &= |0\rangle, & |M_2\rangle &= \frac{1}{\sqrt{3}}|0\rangle + \frac{\sqrt{2}}{\sqrt{3}}|1\rangle, \\ |M_3\rangle &= \frac{1}{\sqrt{3}}|0\rangle + \frac{\sqrt{2}}{\sqrt{3}}e^{i(2\pi/3)}|1\rangle, \\ |M_4\rangle &= \frac{1}{\sqrt{3}}|0\rangle + \frac{\sqrt{2}}{\sqrt{3}}e^{i(4\pi/3)}|1\rangle \end{aligned}$$

in the  $\{|0\rangle, |1\rangle\}$  basis.

### APPENDIX B: 16-CLASS CONFUSION MATRICES

For the sake of completeness, we report here (see Fig. 6) an example of a confusion matrix for a 16-class classification task involving classical colored noise and initially mixed states in the presence of noise. Similar patterns may be observed for pure initial states. The good accuracy level is witnessed by the higher values of the diagonal entries, whereas the good  $F1$  score is achieved because the off-diagonal elements are smaller. Notice that the accuracy starts to drop from the class  $\alpha = 1.2$  and a larger drop may be seen in the range  $[1.7, 2]$ , where the off-diagonal entries take non-negligible values. This introduces a geometrical consideration inspired

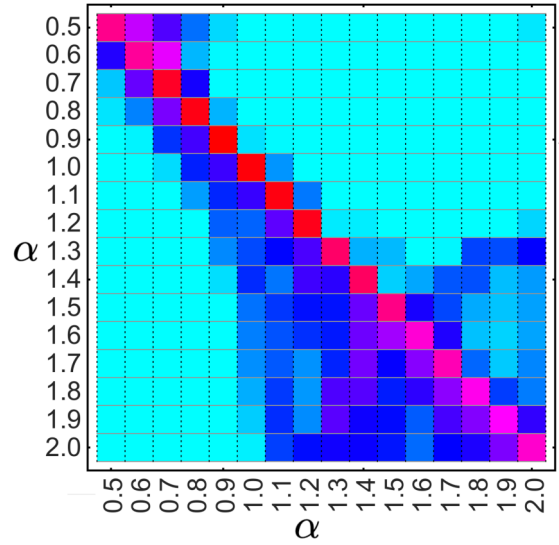


FIG. 6. Confusion matrix for a 16-class classification model involving mixed states undergoing a colored noise dephasing and subject to Gaussian noise in the tomographic stage (red denotes higher values and light blue lower ones). The higher contributions to the confusion correspond to the region  $\alpha = [1.7, 2]$ , where the dephasing factors become closer and closer, which makes it very difficult for the network to find optimal separating hyperplanes.

by the behavior of the dephasing factor  $\Lambda_c(t, \alpha)$ : When noise is added, the data are no longer well separated. The smaller the difference  $|\Lambda_c(t, \alpha_j) - \Lambda_c(t, \alpha_{j+1})|$  is, the harder it is for the network to identify an optimal separating hyperplane between classes  $j$  and  $j + 1$ . This is especially true for the classes  $\alpha \in [1.7, 2]$ .

- [1] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, G. Yang, L. You, Z. Wang *et al.*, *Science* **370**, 1460 (2020).
- [2] A. Acín, I. Bloch, H. Buhrman, T. Calarco, C. Eichler, J. Eisert, D. Esteve, N. Gisin, S. J. Glaser, F. Jelezko, S. Kuhr, M. Lewenstein, M. F. Riedel, P. O. Schmidt, R. Thew, A. Wallraff, I. Walmsley, and F. K. Wilhelm, *New J. Phys.* **20**, 080201 (2018).
- [3] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2002).
- [4] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, Cambridge, 2009).
- [5] T. H. Johnson, S. R. Clark, and D. Jaksch, *EPJ Quantum Technol.* **1**, 10 (2014).
- [6] M. A. C. Rossi, F. Albarelli, D. Tamascelli, and M. G. Genoni, *Phys. Rev. Lett.* **125**, 200505 (2020).
- [7] G. García-Pérez, M. A. C. Rossi, and S. Maniscalco, *npj Quantum Inf.* **6**, 1 (2020).
- [8] C. Foti, T. Heinosaari, S. Maniscalco, and P. Verrucchi, *Quantum* **3**, 179 (2019).
- [9] S. Cialdi, M. A. C. Rossi, C. Benedetti, B. Vacchini, D. Tamascelli, S. Olivares, and M. G. A. Paris, *Appl. Phys. Lett.* **110**, 081107 (2017).
- [10] C. Benedetti, F. Buscemi, P. Bordone, and M. G. A. Paris, *Phys. Rev. A* **89**, 032114 (2014).
- [11] C. Benedetti and M. G. Paris, *Phys. Lett. A* **378**, 2495 (2014).
- [12] C. Benedetti, F. Salari Sehdaran, M. H. Zandi, and M. G. A. Paris, *Phys. Rev. A* **97**, 012126 (2018).
- [13] Q. Bouton, J. Nettersheim, D. Adam, F. Schmidt, D. Mayer, T. Lausch, E. Tiemann, and A. Widera, *Phys. Rev. X* **10**, 011018 (2020).
- [14] D. Tamascelli, C. Benedetti, H.-P. Breuer, and M. G. A. Paris, *New J. Phys.* **22**, 083027 (2020).
- [15] I. Gianani, D. Farina, M. Barbieri, V. Cimini, V. Cavina, and V. Giovannetti, *Phys. Rev. Research* **2**, 033497 (2020).
- [16] E. Paladino, Y. M. Galperin, G. Falci, and B. L. Altshuler, *Rev. Mod. Phys.* **86**, 361 (2014).
- [17] P. Kumar, S. Sendelbach, M. A. Beck, J. W. Freeland, Z. Wang, H. Wang, C. C. Yu, R. Q. Wu, D. P. Pappas, and R. McDermott, *Phys. Rev. Appl.* **6**, 041001(R) (2016).
- [18] P. H. Handel, *2017 Joint Conference of the European Frequency and Time Forum and IEEE International Frequency Control Symposium (EFTF/IFCS)* (IEEE, Piscataway, 2017), pp. 257–260.
- [19] L. A. Pachón, A. Relaño, B. Peropadre, and A. Aspuru-Guzik, *Phys. Rev. E* **98**, 042213 (2018).

- [20] J. Bergli, Y. M. Galperin, and B. L. Altshuler, *New J. Phys.* **11**, 025002 (2009).
- [21] K. A. Kazakov, *Phys. Lett. A* **384**, 126812 (2020).
- [22] M. M. Glazov and E. Y. Sherman, *Phys. Rev. Lett.* **107**, 156602 (2011).
- [23] A. V. Shumilin, E. Y. Sherman, and M. M. Glazov, *Phys. Rev. B* **94**, 125305 (2016).
- [24] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, *Nat. Phys.* **14**, 447 (2018).
- [25] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, *Nat. Mach. Intell.* **1**, 155 (2019).
- [26] G. Torlai, C. J. Wood, A. Acharya, G. Carleo, J. Carrasquilla, and L. Aolita, [arXiv:2006.02424](https://arxiv.org/abs/2006.02424).
- [27] A. M. Palmieri, E. Kovlakov, F. Bianchi, D. Yudin, S. Straupe, J. D. Biamonte, and S. Kulik, *npj Quantum Inf.* **6**, 20 (2020).
- [28] S. Ahmed, C. S. Muñoz, F. Nori, and A. F. Kockum, *Phys. Rev. Lett.* **127**, 140502 (2021).
- [29] D. Bondarenko and P. Feldmann, *Phys. Rev. Lett.* **124**, 130502 (2020).
- [30] L. J. Fiderer, J. Schuff, and D. Braun, *PRX Quantum* **2**, 020303 (2021).
- [31] V. Cimini, I. Gianani, N. Spagnolo, F. Leccese, F. Sciarrino, and M. Barbieri, *Phys. Rev. Lett.* **123**, 230502 (2019).
- [32] V. Cimini, E. Polino, M. Valeri, I. Gianani, N. Spagnolo, G. Corrielli, A. Crespi, R. Osellame, M. Barbieri, and F. Sciarrino, *Phys. Rev. Appl.* **15**, 044003 (2021).
- [33] G. Garau Estarellas, G. L. Giorgi, M. C. Soriano, and R. Zambrini, *Adv. Quantum Technol.* **2**, 1800085 (2019).
- [34] Y. Ma and M. Yung, *npj Quantum Inf.* **4**, 34 (2018).
- [35] C. Harney, S. Pirandola, A. Ferraro, and M. Paternostro, *New J. Phys.* **22**, 045001 (2020).
- [36] P. Sgroi, G. M. Palma, and M. Paternostro, *Phys. Rev. Lett.* **126**, 020601 (2021).
- [37] M. Bukov, A. G. R. Day, D. Sels, P. Weinberg, A. Polkovnikov, and P. Mehta, *Phys. Rev. X* **8**, 031086 (2018).
- [38] J. Walln fer, A. A. Melnikov, W. D r, and H. J. Briegel, *PRX Quantum* **1**, 010301 (2020).
- [39] M. J. Hartmann and G. Carleo, *Phys. Rev. Lett.* **122**, 250502 (2019).
- [40] S. Bandyopadhyay, Z. Huang, K. Sun, and Y. Zhao, *Chem. Phys.* **515**, 272 (2018).
- [41] F. Vicentini, A. Biella, N. Regnault, and C. Ciuti, *Phys. Rev. Lett.* **122**, 250503 (2019).
- [42] E. Flurin, L. S. Martin, S. Haco en-Gourgy, and I. Siddiqi, *Phys. Rev. X* **10**, 011006 (2020).
- [43] A. Goldschmidt, E. Kaiser, J. L. DuBois, S. L. Brunton, and J. N. Kutz, *New J. Phys.* **23**, 033035 (2021).
- [44] D. F. Wise, J. J. L. Morton, and S. Dhomkar, *PRX Quantum* **2**, 010316 (2021).
- [45] J. Hermann, Z. Sch tzle, and F. No , *Nat. Chem.* **12**, 891 (2020).
- [46] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, *Phys. Rev. Lett.* **120**, 024102 (2018).
- [47] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, *Chaos* **27**, 121102 (2017).
- [48] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, *Sci. Rep.* **8**, 6085 (2018).
- [49] A. P. Trischler and G. M. D’Eleuterio, *Neural Netw.* **80**, 67 (2016).
- [50] S. Hesabi, D. Afshar, and M. G. Paris, *Opt. Commun.* **437**, 377 (2019).
- [51] <https://github.com/AdriQD/multiclass-classification-dephasing-channel>.
- [52] C. Benedetti, F. Buscemi, P. Bordone, and M. G. Paris, *Int. J. Quantum Inf.* **10**, 1241005 (2012).
- [53] S. Razavian, C. Benedetti, M. Bina, Y. Akbari-Kourbolagh, and M. G. A. Paris, *Eur. Phys. J. Plus* **134**, 284 (2019).
- [54] C. Benedetti, M. G. A. Paris, and S. Maniscalco, *Phys. Rev. A* **89**, 012114 (2014).
- [55] A. E. Rastegin, *Phys. Scr.* **89**, 085101 (2014).
- [56] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, *J. Math. Phys.* **45**, 2171 (2004).
- [57] C. A. Fuchs and R. Schack, *Rev. Mod. Phys.* **85**, 1693 (2013).
- [58] J. I. Rosado, *Found. Phys.* **41**, 1200 (2011).
- [59] M. Minsky and S. A. Papert, *Perceptron: An Introduction to Computational Geometry* (MIT Press, Cambridge, 1969).
- [60] C. D. Manning, R. Prabhakar, and H. Sch tze, *Introduction to Information Retrieval* (Cambridge University Press, Cambridge, 2009).
- [61] F. Johansson *et al.*, mpmath: A Python library for arbitrary-precision floating-point arithmetic, version 0.18, 2013, available at <http://mpmath.org/>.
- [62] *Neural Networks: Tricks of the Trade*, edited by G. Montavon, G. B. Orr, and K. R. M ller, Lecture Notes in Computer Science Vol. 7700 (Springer, Berlin, 2012).
- [63] T. Xin, S. Lu, N. Cao, G. Anikeeva, D. Lu, J. Li, G. Long, and B. Zeng, *npj Quantum Inf.* **5**, 109 (2019).
- [64] P. B hlmann and S. van de Geer, *Statistics for High-Dimensional Data* (Springer, Berlin, 2011).