

Quantum State Reconstruction in a Noisy Environment via Deep Learning

Angela Rosy Morgillo^{1,2,*} Stefano Mangini^{1,2,†} Marco Piastra³ and Chiara Macchiavello^{1,2}

¹*Dipartimento di Fisica, Università di Pavia, Via Bassi 6, I-27100, Pavia, Italy*

²*INFN Sezione di Pavia, Via Bassi 6, I-27100, Pavia, Italy*

³*Dipartimento di Ingegneria Industriale e dell'Informazione, Università di Pavia, via Ferrata 5, I-27100, Pavia, Italy*

(Dated: September 22, 2023)

Quantum noise is currently limiting efficient quantum information processing and computation. In this work, we consider the tasks of reconstructing and classifying quantum states corrupted by the action of an unknown noisy channel using classical feedforward neural networks. By framing reconstruction as a regression problem, we show how such an approach can be used to recover with fidelities exceeding 99% the noiseless density matrices of quantum states of up to three qubits undergoing noisy evolution, and we test its performance with both single-qubit (bit-flip, phase-flip, depolarising, and amplitude damping) and two-qubit quantum channels (correlated amplitude damping). Moreover, we also consider the task of distinguishing between different quantum noisy channels, and show how a neural network-based classifier is able to solve such a classification problem with perfect accuracy.

I. INTRODUCTION

One of the main problems in quantum information processing and computation is that quantum systems can be corrupted by unwanted interactions with the environment. Therefore, the incorporation of robust quantum error correction and mitigation strategies is of paramount importance to realize the full potential of quantum information processing.

Despite the effectiveness of quantum error correction protocols in preserving information, they often require significant overhead and resources. Quantum error mitigation techniques, on the other hand, focus on reducing the impact of noise without fully correcting it, making them more feasible for near-term quantum devices [1, 2]. Examples are readout mitigation techniques to correct measurement errors [3–5], noise deconvolution methods to retrieve ideal expectation values of generic observables evaluated on a system subject to a known noise before measurement [6, 7], probabilistic error cancellation [8] and data driven approaches such as zero-noise extrapolation [9] and Clifford data regression [10–12] to mitigate noise happening during a quantum computation.

Another area of great interest is deep learning, which has achieved impressive successes over the past years, with generative pre-trained large language models now leading the way [13–15]. Deep learning models have excelled in diverse areas, from image and speech recognition models [16, 17] to playing games [18], reaching and often surpassing human-level performances. These advancements highlight the vast potential of deep learning in

revolutionizing numerous fields, including quantum computation and information.

Indeed, deep learning techniques have shown great promises also for quantum information processing applications, as they were leveraged successfully in, e.g., experimental phase estimation tasks [19], automating the development of QCVV (Quantum characterization, validation and verification) protocols [20], learning quantum hardware-specific noise models [21], increasing measurement precision of quantum observables with neural networks [22], quantum error mitigation [23–25], identifying quantum protocols such as teleportation or entanglement purification [26], classification and reconstruction of optical quantum states [27], and quantum state estimation [28].

In this work, we leverage machine learning techniques based on feed-forward neural networks to deal with the task of recovering noise-free quantum states when they undergo an undesired noise evolution. In fact, while it is well known that quantum noisy channels cannot be physically inverted in general, this may be achieved by means of classical post-processing methods [6–8]. In particular, since neural networks are universal approximators [29], they can be used to learn a mapping that effectively inverts that effect of noise, and hence they can be used to reconstruct noiseless quantum states. Specifically, let $\tilde{\mathbf{r}}$ indicate the (generalised) Bloch components of a noisy quantum state, our goal is to train a neural network $h_{\mathbf{w}}(\cdot)$ to output the Bloch vector of the ideal noiseless state $\tilde{\mathbf{r}} \rightarrow h_{\mathbf{w}}(\tilde{\mathbf{r}}) = \mathbf{r}$, where \mathbf{r} is the ideal Bloch vector of the state before it undergoes noise process. We explore several combinations of single- and two-qubit noisy channels acting on systems of up to three qubits, and also study the effect of using different loss functions for training, and show that our neural network-based method can reach quantum state reconstruction fidelities higher than 99.9%. The main idea of the proposed method is summarised in Figure I.

In addition to regression tasks, we also show how feed-

* angelarosy.morgillo01@universitadipavia.it

† Current address: ¹*Algorithmiq Ltd, Kanavakatu 3C 00160 Helsinki, Finland.* ²*QTF Centre of Excellence, Department of Physics, University of Helsinki, P.O. Box 43, FI-00014 Helsinki, Finland.*

forward neural networks can be used for classifying different quantum channels based on the effect they have on quantum states. In particular, using as inputs Bloch vectors $[\tilde{\mathbf{r}}, \mathbf{r}]$ obtained with different channels, the network will output a label corresponding to the quantum channel that has been applied to \mathbf{r} in order to produce $\tilde{\mathbf{r}}$. Also in this case, we achieve almost perfect channel classification accuracy.

The rest of the manuscript is organised as follows. In Section II we formally introduce the problem and the neural network used to obtain the quantum state reconstruction. In Section III we present the results obtained for the reconstruction of pure and mixed states, and we introduce the noise classification problems that can be solved similarly with neural networks. In Section IV we summarise all our results and possible improvements of our method.

II. METHODS

In this section, we formalise the quantum communication problem we want to tackle, that is, the recovery of noiseless quantum states undergoing an undesired noisy evolution. We first start by introducing the notation for describing an n -qubit quantum state in terms of its Bloch components, and then move on discussing the neural network approach used in this work, including details on the optimisation procedure and the construction of the training and test dataset.

A. Reconstruction of noisy Bloch vectors

The state of an n -qubit quantum system is described by its density matrix $\rho \in \mathbb{C}^{2^n \times 2^n}$, which can be expressed in the Pauli basis as follows [30, 31]

$$\rho = \frac{1}{2^n} (\mathbb{I}_{2^n} + \mathbf{r} \cdot \mathbf{P}) \quad (1)$$

where $\mathbf{r} \in \mathbb{R}^{4^n - 1}$ is the *generalised Bloch vector*, and $\mathbf{P} = (P_1, \dots, P_{4^n - 1})$ is a vector containing the multi-qubit Pauli basis, obtained by considering tensor products of single-qubit Pauli matrices, that is $P_i = \sigma_1^{(i)} \otimes \dots \otimes \sigma_n^{(i)}$, with $\sigma_k \in \{\mathbb{I}, X, Y, Z\}$.

Quantum channels are completely positive trace preserving (CPTP) maps whose action on a state ρ can be expressed in Kraus form as [30]

$$\rho \longrightarrow \tilde{\rho} = \mathcal{N}(\rho) = \sum_i E_i \rho E_i^\dagger, \quad (2)$$

where $\{E_i\}$ are the Kraus operators of the channel $\mathcal{N}(\cdot)$, satisfying the trace preserving condition $\sum_i E_i^\dagger E_i = \mathbb{I}$. In our experiments, we consider various single-qubit noisy channels (bit-flip \mathcal{X}_p , phase-flip \mathcal{Z}_p , bit-phase-flip \mathcal{Y}_p , general Pauli \mathcal{P}_p , depolarizing \mathcal{D}_p and amplitude damping $\mathcal{A}_{p\gamma}$), as well as a correlated two-qubit amplitude

damping channel. We refer to Appendix A for an extended discussion on the quantum noise models used in this work.

Given a noisy channel $\mathcal{N}(\cdot)$, our goal is to obtain through a learning procedure an optimised neural network that receives noisy Bloch vectors $\{\tilde{\mathbf{r}}_k\}$, and outputs the corresponding noiseless vectors $\{\mathbf{r}_k\}$. In other words, we are looking for the function $h(\cdot)$ which inverts the action of the noise on the Bloch components of the quantum states, namely

$$\begin{aligned} h(\tilde{\mathbf{r}}) &= \mathbf{r}, \quad \text{with} \\ \rho &= \frac{1}{2^n} (\mathbb{I}_{2^n} + \mathbf{r} \cdot \mathbf{P}) \\ \tilde{\rho} &= \mathcal{N}(\rho) = \frac{1}{2^n} (\mathbb{I}_{2^n} + \tilde{\mathbf{r}} \cdot \mathbf{P}). \end{aligned} \quad (3)$$

B. Reconstruction with neural networks

We provide a concise overview of the fundamental aspects of neural networks, discussing their relevance in addressing the task of quantum state reconstruction.

1. Generation of the training set

The initial phase of addressing a regression problem involves constructing a valid dataset. In our specific case, the training (and validation) set consists of pairs of noisy and noiseless Bloch vectors

$$\mathcal{T} = \{(\tilde{\mathbf{r}}_m, \mathbf{r}_m)\}_{m=1}^M, \quad (4)$$

which are obtained by evolving some input quantum states ρ_m through the noisy channel under investigation, thus obtaining the noisy states $\tilde{\rho}_m$. The Bloch components of these density matrices are then computed as $r_i = \text{Tr}[\rho P_i]$ for $i = 1, \dots, 4^n - 1$ (1).

The choice of the (generalized) Bloch vector as element of the dataset is twofold: first, each quantum state is characterized by its own vector, which grants a unique representation of the state; and secondly a vector as input fits naturally in the processing structure of a feed-forward neural network.

The input quantum states we consider are uniformly distributed in the space of quantum states. For the case of pure states $\rho_m = |\psi_m\rangle\langle\psi_m|$, these are obtained by sampling states $|\psi_m\rangle$ from the Haar distribution [32, 33], while for uniformly distributed mixed states, these can be generated either starting from uniformly distributed pure states by means of an appropriate rescaling [34, 35] or by using the Ginibre ensemble [36–38].

The cardinality $|\mathcal{T}| = M$ of the dataset is contingent upon the specific problem under consideration and, as demonstrated in Sec. III, has a direct impact on the performance of the network. As the quantum computational resources to generate the training set \mathcal{T} may be

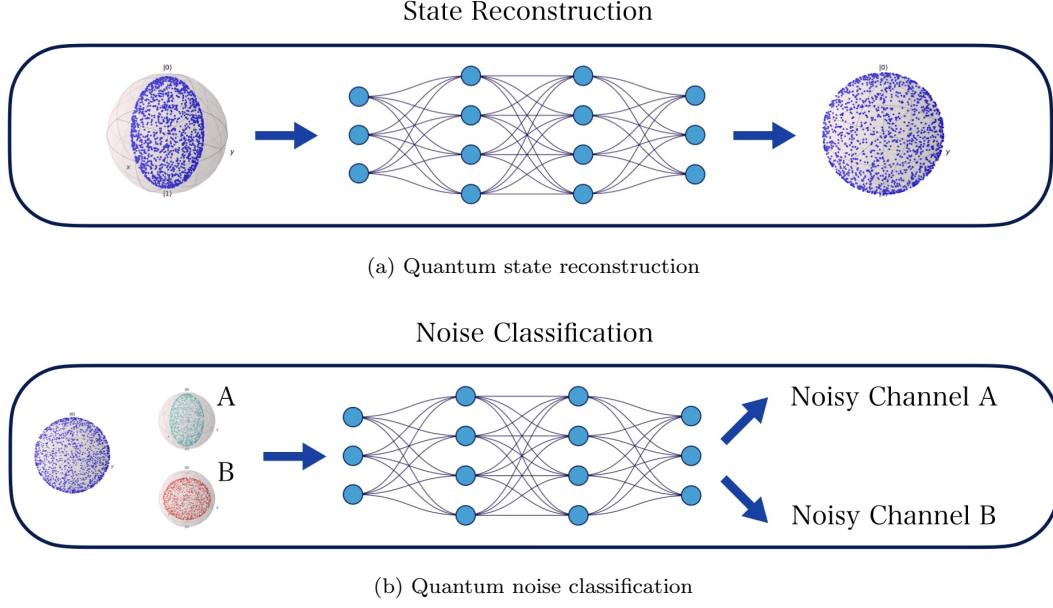


FIG. 1. Outline of the neural network-based noise reconstruction and classification protocols. **(a)** Noisy Bloch vectors, representing quantum states affected by noise, serve as input to a feed-forward neural network. The quantum state reconstruction protocol aims to recover the original noiseless quantum states from the observed noisy Bloch vectors, utilizing the neural network model. **(b)** Both noisy and noiseless Bloch vectors are fed into the neural network as input. The network is specifically designed for a classification task, where the output provides a label representing the type of noise acting on the noiseless quantum state.

demanding experimentally, one generally has to find a compromise between achieving high reconstruction accuracies, and the number of samples (i.e. quantum states) included in the dataset.

2. Feed-forward neural networks

In this work, we analyze data using deep feed-forward neural networks, which are parametric models that process information in layer-wise fashion through a repeated application of similar operations, as shown in Fig. 1(a).

These neural networks consist of an *input layer* responsible for data loading, followed by multiple *hidden layers* to process the information, and finally, an *output layer* to obtain the computation's result. Each layer consists of a set of individual nodes known as *neurons*, and while input and output layers have a number of neurons matching the dimension of the input and the output respectively, the number of neurons in the hidden layers is an architectural hyperparameter to be chosen in advance by the designer. For example, the action of a feed-forward neural network with two hidden layers and trainable parameters θ , can be expressed as

$$\begin{aligned} \hat{\mathbf{y}} &= \text{NN}_{\theta}(\mathbf{x}) \\ &= \mathbf{w} \cdot g\left(\mathbf{W}^{[2]} g\left(\mathbf{W}^{[1]} \mathbf{x} + \mathbf{b}^{[1]}\right) + \mathbf{b}^{[2]}\right) + b \end{aligned} \quad (5)$$

where $\mathbf{x} \in \mathbb{R}^d$ and $\hat{\mathbf{y}} \in \mathbb{R}^p$ are the input and output vectors, $\mathbf{W}^{[1]} \in \mathbb{R}^{h_1 \times d}$ and $\mathbf{b}^{[1]} \in \mathbb{R}^{h_1}$ are the trainable

parameters for the first hidden layer, $\mathbf{W}^{[2]} \in \mathbb{R}^{h_2 \times h_1}$ and $\mathbf{b}^{[2]} \in \mathbb{R}^{h_2}$ are the trainable parameters for the second hidden layer, $\mathbf{w} \in \mathbb{R}^p$ and $b \in \mathbb{R}$ are the trainable parameters for the output layer, and $g(\cdot)$ is a non-linear activation function which is applied element-wise to the entries of the vectors. As previously mentioned, h_1 and h_2 are hyperparameters that represent the number of hidden neurons for each respective layer.

In our simulations, we explore different architectures using networks with 2 or 3 hidden layers using $h_i \in \{64, 128\}$ hidden neurons per layer, while the input and output layers have dimension $d = p = 4^n - 1$, as they are employed to represent the components of the Bloch vectors. For the activation function, as customary in machine learning, we adopt the Rectified Linear Unit (ReLU), defined as $g(x) = \text{ReLU}(x) := \max(0, x)$.

3. Performance metrics

Given the dataset and the trainable model, we discuss the figures of merit employed for training and evaluating the neural network's performance in quantum state reconstruction and noise classification tasks. In the context of quantum state reconstruction, we have tested two possible alternatives coming from the classical and quantum information domain respectively, namely the Mean Squared Error (MSE) between the reconstructed and ideal Bloch vectors, and the quantum infidelity between the reconstructed and original quantum states.

In noise classification problems, we used both categorical cross-entropy (15) and accuracy metrics (16) in order to assess how effectively the neural network can distinguish between different types of noisy channels.

MSE.— The mean squared error is the most common measure of performances for regression problems in machine learning and consists of the Euclidean distance between vectors, which in our case becomes

$$\ell(\boldsymbol{\theta}, \mathbf{r}_i) = \|\mathbf{r}_i - \hat{\mathbf{r}}_i(\boldsymbol{\theta})\|^2 = \|\mathbf{r}_i - \text{NN}_{\boldsymbol{\theta}}(\tilde{\mathbf{r}}_i)\|^2, \quad (6)$$

where \mathbf{r}_i is the noiseless Bloch vector (see Eq. (4)), and $\hat{\mathbf{r}}_i(\boldsymbol{\theta}) = \text{NN}_{\boldsymbol{\theta}}(\tilde{\mathbf{r}}_i)$ is the one predicted by the neural network, with trainable parameters $\boldsymbol{\theta}$, when receiving as input the noisy Bloch vector $\tilde{\mathbf{r}}_i$. Then, the mean squared error function over the entire dataset \mathcal{T} of size $|\mathcal{T}| = M$, is

$$\begin{aligned} \mathcal{L}_{\text{MSE}}(\boldsymbol{\theta}; \mathcal{T}) &= \frac{1}{M} \sum_{i=1}^M \ell(\boldsymbol{\theta}, \mathbf{r}_i) \\ &= \frac{1}{M} \sum_{i=1}^M \|\mathbf{r}_i - \hat{\mathbf{r}}_i\|^2 \quad (\tilde{\mathbf{r}}_i, \mathbf{r}_i) \in \mathcal{T}. \end{aligned} \quad (7)$$

Infidelity.— The quantum fidelity is a measure of distance for quantum states, and given the quantum nature of the data under investigation, it is particularly suited to assess the reconstruction performances of the neural network. Given two density matrices ρ and σ , their fidelity is defined as [30]

$$F(\rho, \sigma) := \text{Tr} \left[\sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \right]^2, \quad (8)$$

with $0 \leq F(\rho, \sigma) \leq 1$, where the second equality holds if and only if the states are equal $\rho = \sigma$. The *infidelity* between two quantum states is then defined as $I(\rho, \sigma) := 1 - F(\rho, \sigma)$.

Despite being suited to measure the distance of quantum states, the complex functional dependence of the infidelity on the Bloch vectors of the density matrices — hence the parameters of the neural network — often leads to numerical instabilities when it is used as the loss function to drive the training of the neural network, eventually impairing the optimisation process. For this reason, when ρ and σ are pure states, we instead use directly the simplified but equivalent expression for the fidelity

$$F(\rho, \sigma) = \text{Tr}[\rho \sigma], \quad \text{for } \rho, \sigma \text{ pure}, \quad (9)$$

while if the states are mixed, we use an alternative measure of distance proposed in [39]

$$F(\rho, \sigma) = \frac{|\text{Tr}[\rho \sigma]|}{\sqrt{\text{Tr}[\rho^2] \text{Tr}[\sigma^2]}}, \quad \text{for } \rho, \sigma \text{ mixed}. \quad (10)$$

The fidelity in Eq. (10) reaches the maximum 1 if and only if $\rho = \sigma$. However, it differs from the standard fidelity reported in Eq. (8), since it is not monotonic under quantum operations, meaning it is neither concave

nor convex, and it includes a normalization factor, resulting in a scaled fidelity measure, when one of the two states is pure, i.e. if $\sigma = |\psi\rangle\langle\psi|$, then $F(\rho, |\psi\rangle\langle\psi|) = \langle\psi|\rho|\psi\rangle / \text{Tr} \rho^2$.

From the fidelity expressions in Eqs. (9) and (10), we calculate the corresponding infidelity and perform an average over the entire dataset, yielding

$$\mathcal{L}_{\text{INF}}(\boldsymbol{\theta}; \mathcal{T}) = \frac{1}{M} \sum_{i=1}^M 1 - F(\rho_i, \sigma_i), \quad (11)$$

where ρ_i and σ_i are the density matrices computed respectively from the Bloch vectors \mathbf{r}_i and $\hat{\mathbf{r}}_i = \hat{\mathbf{r}}_i(\boldsymbol{\theta}) = \text{NN}_{\boldsymbol{\theta}}(\tilde{\mathbf{r}}_i)$, with $(\tilde{\mathbf{r}}_i, \mathbf{r}_i) \in \mathcal{T}$.

Moreover, it is worth noticing that for single-qubit density matrices the fidelity (8) can be further simplified and expressed in terms of the Bloch vectors as [40]

$$F(\mathbf{r}, \mathbf{s}) = \frac{1}{2} \left(1 + \mathbf{r} \cdot \mathbf{s} + \sqrt{(1 - \|\mathbf{r}\|^2)(1 - \|\mathbf{s}\|^2)} \right), \quad (12)$$

where $\mathbf{r}, \mathbf{s} \in \mathbb{R}^3$ are the Bloch vectors for ρ and σ respectively. Specifically, for the particular case when both states are pure $\|\mathbf{r}\|^2 = \|\mathbf{s}\|^2 = 1$, then the expression for the infidelity corresponds to the mean squared error up to a prefactor

$$I(\mathbf{r}, \mathbf{s}) = 1 - \frac{1}{2}(1 + \mathbf{r} \cdot \mathbf{s}) = \frac{\|\mathbf{s} - \mathbf{r}\|^2}{4}. \quad (13)$$

We refer to Appendix B for more details on the derivation of Eqs.(12) and (13). Notably, to the best of our knowledge, apart for single-qubit states, there is no straightforward connection between the fidelity of quantum states and the Euclidean distance of their generalised Bloch vectors.

Finally, in order to assess the performance of the optimised neural network we introduce the Average Test Fidelity (ATF), defined as the mean fidelity between the predicted quantum states and their corresponding ideal counterparts averaged over a test dataset $\tilde{\mathcal{T}}$ which was not used during training. The ATF is calculated as

$$\text{ATF}(\boldsymbol{\theta}; \tilde{\mathcal{T}}) = \frac{1}{N} \sum_{i=1}^N F(\rho_i, \sigma_i), \quad (14)$$

where N is the cardinality of the test set, $F(\cdot, \cdot)$ is the fidelity in Eq. (8), and ρ_i and σ_i are the density matrices computed respectively from the Bloch vectors \mathbf{r}_i and $\hat{\mathbf{r}}_i = \text{NN}_{\boldsymbol{\theta}}(\tilde{\mathbf{r}}_i)$, with $(\tilde{\mathbf{r}}_i, \mathbf{r}_i) \in \tilde{\mathcal{T}}$. A high average test fidelity, typically exceeding 99.9%, indicates that our neural network is capable of accurately reconstructing the corrupted quantum states.

Categorical Cross-Entropy.— The categorical cross-entropy is one of the most common measures to evaluate the performance of classification models [41]. Given a classification task with C different classes, the categorical cross-entropy quantifies the disparity between the predicted probability distribution of the classes and the true

class labels, and it is mathematically defined as

$$\text{CCE} := - \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log(p_{ij}), \quad (15)$$

where M is the number of samples to be classified, $y_i = (0, 0, \dots, 1_{c(i)}, \dots, 0) \in \{0, 1\}^C$ is the true probability distribution for the samples indicating that the i -th sample belongs to class $c(i)$, and $p_{ij} \in [0, 1]$ is the model's predicted probability distribution for the i -th sample to belong to the j -th class. This metric penalizes the deviation between predicted and actual distributions, with lower values indicating better alignment between the model's predictions and the true labels.

Accuracy.— For classification tasks the evaluation of performance commonly revolves around the metric of accuracy, which quantifies the effectiveness of a model in correctly categorizing samples within a dataset. Mathematically, the accuracy is defined as the ratio of the number of correctly classified samples to the total number of samples, namely

$$\text{ACC} := \frac{\text{number of correct predictions}}{\text{total number of predictions}}, \quad (16)$$

where a value of 1 indicates a perfect classification.

4. Optimization

Training the neural network means solving the minimisation problem

$$\theta_{\text{opt}} = \arg \min_{\theta} \mathcal{L}(\theta; \mathcal{T}), \quad (17)$$

where θ are the trainable parameters of the neural network, \mathcal{T} is the training dataset as defined in Eq. (4), and \mathcal{L} is the loss function driving the learning process, which, as discussed in the previous section, in our case is either the mean squared error (7) or the average infidelity (11).

We optimize the neural networks using Adam, a variant of a stochastic gradient descent with adaptive estimation of first and second order moments [42]. For the results shown in Sec. III, we report the combination of training hyperparameters (number of training epochs, batch size, and learning rate) that attains the best performances.

III. RESULTS

We now present the results obtained for quantum state reconstruction and quantum noise classification using the proposed neural network methods. We start discussing the performance of the neural network in reconstructing noise-free quantum states corrupted by various single- and two-qubit noisy channels, and then proceed showcasing the network's capability to classify different quantum

noisy channels accurately. Our results demonstrate high-fidelity state reconstruction and robust channel classification, thus revealing the potential of machine learning techniques for quantum information processing and computation.

All simulations performed in this work are run with Qiskit [43] and TensorFlow [44].

A. Quantum State Reconstruction

In order to explore the reconstruction capabilities of the neural network approach, we first start by studying its performances in the simpler case of learning noisy single-qubit states, and then move on to the more complex case of multi-qubit systems. In both scenarios, we see a clear dependence of the performances on the amount of available data to train the model (size of the training set), and provided that data is sufficient the network is always able to restore noiseless quantum states from their noisy counterparts.

Whenever we consider the task of reconstructing initially pure states, an auxiliary *normalization layer* is added at the end of the neural network (5) so that the outputs always consists of (Bloch) vectors with unit norm (for single-qubit states), as desired for pure states. Such normalisation constraint enforces the generation of physically consistent output states, which in turns effectively constrains the value of the infidelity loss function to remain in the physical regime¹. It's worth noting that even when the MSE is employed as the loss function, the same normalization layer is used to ensure that the predicted states maintain their physical integrity and adhere to the desired constraints. The effect of the normalization layer is simply to rescale each output as follows

$$\text{NN}_{\theta}(\tilde{\mathbf{r}}) = \hat{\mathbf{r}} \longrightarrow \hat{\mathbf{r}} / \|\hat{\mathbf{r}}\|_2. \quad (18)$$

In addition, whenever we consider the case of reconstructing already mixed states ρ of given purity $\text{Tr}[\rho^2]$, the output of the neural network is additionally rescaled with the purity of the initial mixed state, $\hat{\mathbf{r}} \rightarrow \sqrt{\text{Tr}[\rho^2]} \hat{\mathbf{r}}$.

Single-qubit states.— In Table I we summarize the results of the reconstruction process for various noisy channels, using different loss functions to drive the training process, and using pure or mixed states as inputs to the neural network. In all cases we observe a very good average test fidelity (ATF) (14) at the end of training exceeding 99.6%, thus showing the effectiveness of the proposed approach for reconstructing noisy states. As we now see, equally good performances are also obtained in the more complex task of inverting noise in multi-qubit systems.

¹ The fidelity (8) can be larger than 1 if the Bloch vectors have norms exceeding one, which is not possible for density matrices of quantum systems. If no constraint on the Bloch vectors is present, then the training process will simply push the neural network to output states of larger and larger norm to maximise the overlap.

Channel	ATF (MSE)	ATF (INF)
$ \mathcal{T} = 30$ Pure States		
\mathcal{X}_p ($p = 0.2$)	0.998	0.997
\mathcal{Z}_p ($p = 0.2$)	0.997	0.998
\mathcal{Y}_p ($p = 0.2$)	0.998	0.998
\mathcal{P}_p ($p_0 = 0.7, p_{1,2,3} = 0.1$)	0.998	0.998
\mathcal{D}_p ($p = 0.3$)	0.996	0.997
$\mathcal{A}_{p\gamma}$ ($p = 0.5, \gamma = 0.3$)	0.998	0.997
$ \mathcal{T} = 30$ Mixed States		
\mathcal{Z}_p ($p = 0.2$)	0.999	0.999

TABLE I. Reconstruction of single-qubit quantum states. Training with both the MSE (7) and infidelity (11) as loss functions yields good average test fidelities (ATF) at the end of the optimisation process. Results are reported for different noisy channels (bit-flip \mathcal{X}_p , phase-flip \mathcal{Z}_p , bit-phase flip \mathcal{Y}_p , general Pauli \mathcal{P}_p , depolarizing \mathcal{D}_p , and general amplitude damping channels $\mathcal{A}_{p\gamma}$), and for both pure and mixed initial states.

Channel	ATF (MSE)	ATF (INF)
$ \mathcal{T} = 300$ Two-qubit States		
$\mathcal{Z}_p \otimes \mathcal{I}$ ($p = 0.2$)	0.994	0.993
$\mathcal{Z}_p \otimes \mathcal{Z}_p$ ($p = 0.2$)	0.994	0.995
$\mathcal{Z}_p \otimes \mathcal{X}_p$ ($p = 0.2$)	0.994	0.995
\mathcal{C}_{AD} ($\eta = 0.1, \mu = 0.2$)	0.995	0.999
$ \mathcal{T} = 900$ Three-qubit States		
$\mathcal{X}_p \otimes \mathcal{Z}_p \otimes \mathcal{Y}_p$ ($p = 0.2$)	0.999	0.999

TABLE II. Results of pure multi-qubit states reconstruction using MSE and infidelity as loss functions. For two-qubit states, we studied scenarios involving: a phase-flip channel on the first qubit $\mathcal{Z}_p \otimes \mathcal{I}$, phase-flip channel on both qubits $\mathcal{Z}_p \otimes \mathcal{Z}_p$, phase-flip and bit-flip channels respectively on the first and second qubit $\mathcal{Z}_p \otimes \mathcal{X}_p$, and correlated amplitude damping $\mathcal{C}_{AD}(\eta, \mu)$. For three-qubit states, we considered the scenario characterized by bit-, phase-, and bit-phase-flip channels applied distinctly to all three qubits $\mathcal{X}_p \otimes \mathcal{Z}_p \otimes \mathcal{Y}_p$.

Multi-qubit states.— We tested the reconstruction procedure also on systems of $n = 2$ and $n = 3$ qubit systems undergoing several noisy evolutions with both uncorrelated and correlated noisy channels, and we summarize the results in Table II. For two-qubit systems, we considered the following noise maps: (i) phase-flip channel with $p = 0.2$ applied to the first qubit and nothing on the second indicated as $\mathcal{Z}_p \otimes \mathcal{I}$; (ii) a phase-flip channel applied to both qubits with $p = 0.2$, denoted as $\mathcal{Z}_p \otimes \mathcal{Z}_p$; (iii) phase-flip channel on the first qubit and a bit-flip channel on the second qubit, both with $p = 0.2$, denoted as $\mathcal{Z}_p \otimes \mathcal{X}_p$; (iv) and finally we also have investigated a scenario where a correlated two-qubit amplitude damping channel \mathcal{C}_{AD} ($\eta = 0.1, \mu = 0.2$), was applied to the system (see Eq.(A14) in Appendix A for a definition of the channel).

For $n = 3$ qubit systems instead, we tested the reconstruction performance with states subject to the composite channel $\mathcal{X}_p \otimes \mathcal{Z}_p \otimes \mathcal{Y}_p$. In this configuration, each

qubit experiences a distinct quantum channel, including a bit-flip, phase-flip, and bit-phase-flip channel with a common noise parameter of $p = 0.2$.

The results reported in Table II again reveal a successful reconstruction of ideal density matrices through the use of a relatively simple feed-forward neural network, and thus confirm the effectiveness of the proposed method. As already mentioned previously, we stress again that to ensure the production of pure quantum states as output, a normalization layer has been incorporated. Specifically, by appropriately multiplying the normalisation layer in (18), the norm of the output Bloch vectors is constrained to $\sqrt{3}$ for two-qubit states, while it is set to $\sqrt{7}$ [45] for three-qubit states. Comparing the cardinality of the training set $|\mathcal{T}|$ used for the simulations in Tab. I and Tab II, we see that more samples in the training dataset are generally needed to ensure a good reconstruction for larger system sizes, as one would expect. A discussion on the impact of the available information on the reconstruction performances is the topic of the next section.

In Fig 2 we report the evolution of the MSE (7) and infidelity (11) loss functions during training for the case of $\mathcal{Z}_p \otimes \mathcal{X}_p$ applied to a two-qubit system. As clear from the picture, the optimisation process of both cost functions is straightforward, and interestingly they follow a similar minimisation behaviour, despite there is not a trivial relation between the two, as instead it happens for single qubit states, see Eq. 13.

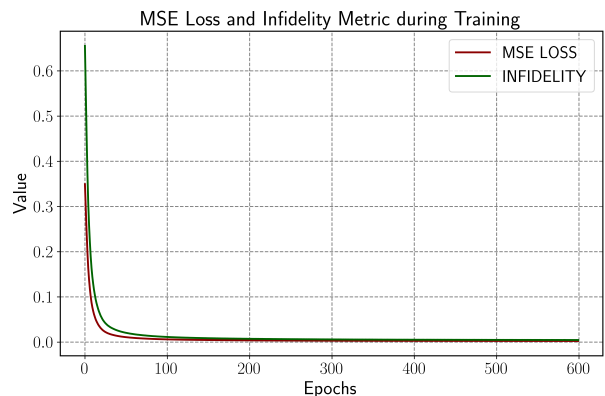


FIG. 2. Optimisation process of the neural network to reconstruct a two-qubit state under application of the noisy channel $\mathcal{Z}_p \otimes \mathcal{X}_p$, using the MSE (7) and the infidelity (11) as loss functions. Both metrics display a similar minimisation behaviour.

We conclude by noticing that, taking into account that an increase in the number of qubits leads to a proportional increase in the dataset cardinality and the neural network complexity, the reconstruction approach proposed in this work can be straightforwardly applied to any n -qubit quantum state.

1. Impact of the dataset size on the reconstruction performances

So far we have focused on assessing the reconstruction performances of the neural network assuming that enough data is available, and here instead analyse how such performances depends on the size of the training set.

In Fig. 3 we report the average test fidelity obtained at the end of training, for neural networks optimised using training sets of different cardinality. In panel 3a we show the data for the reconstruction of single-qubit states undergoing a phase-flip channel $\mathcal{Z}_{0.2}$, and in panel 3b the reconstruction of two-qubit states undergoing an uncorrelated two-qubit phase-flip channel $\mathcal{Z}_{0.2} \otimes \mathcal{Z}_{0.2}$.

In both cases —and for both the considered loss functions—, we observe that the use of a larger training set yields better reconstruction performances, up until a plateau is reached, but importantly also that satisfactory results can be achieved even with a limited number of samples.

B. Classification of noisy channels

The application of neural networks to quantum information processing can be extended beyond quantum state reconstruction to that of classification of quantum channels. In particular, in this section we show how a neural network can be trained to discriminate between noisy channels based on the effect they have on some input states, a scenario which is graphically depicted in Fig. I. Our exploration encompasses a series of classification scenarios, including binary and multi-class classification.

In such classification problems, each data item in the training set is constructed by considering as input the noiseless Bloch vector \mathbf{r}_m appended to the noisy one $\tilde{\mathbf{r}}_m$, and as output an integer label y_m encoding which type of error was applied to \mathbf{r}_m to obtain $\tilde{\mathbf{r}}_m$. Formally, given a classification task with C possible classes, the training dataset is then defined as

$$\mathcal{T}_{IN} = \{([\tilde{\mathbf{r}}_m, \mathbf{r}_m], y_m)\}_{m=1}^M, \quad (19)$$

where $[\tilde{\mathbf{r}}_m, \mathbf{r}_m] \in \mathbb{R}^{2 \times 3}$ is the input to the neural network, and $y_m \in \{1, \dots, C\}$ is the desired output. We refer to this training dataset with \mathcal{T}_{IN} , where the subscript indicates that we use as input the extended vector obtained by merging the *ideal* and *noisy* Bloch vector.

Furthermore, we consider the more complex scenario where every training data point comprises only the noisy vector $\tilde{\mathbf{r}}_m$, accompanied with its associated noise label y_m . As the neural network is now provided with less information, this scenario presents a higher level of complexity and, as will become clear from the following results, it generally necessitates a larger corpus of data points. We refer to this type of training dataset with

TABLE III. Results for quantum channel classification tasks, with corresponding accuracies (16) obtained at the end of training and evaluated on test sets $\tilde{\mathcal{T}}$. We studied the binary classification task of distinguishing channels \mathcal{Z}_p ($p = 0.2$) vs. $\mathcal{A}_{p\gamma}$ ($p = 0.5, \gamma = 0.3$) in both variations “IN” and “N” for the training dataset (19), and the three-class classification problem for channels \mathcal{Z}_p ($p = 0.2$) vs. $\mathcal{A}_{p\gamma}$ ($p = 0.5, \gamma = 0.3$) vs. \mathcal{D}_p ($p = 0.3$) in the variant “IN”.

Dataset cardinality	Accuracy (on test set $\tilde{\mathcal{T}}$)
Binary Classification	
\mathcal{Z}_p vs $\mathcal{A}_{p\gamma}$	
$ \mathcal{T}_{IN} = 300$	1 ($ \tilde{\mathcal{T}} = 100$)
$ \mathcal{T}_N = 300$	0.92 ($ \tilde{\mathcal{T}} = 100$)
Ternary Classification	
\mathcal{Z}_p vs $\mathcal{A}_{p\gamma}$ vs \mathcal{D}_p	
$ \mathcal{T}_{IN} = 960$	1 ($ \tilde{\mathcal{T}} = 120$)

\mathcal{T}_N , where the subscript indicates that we use as input exclusively the noisy Bloch vector.

As standard with classification tasks in machine learning, for both scenarios we use a *one-hot encoding* of the labels and train the neural networks using the categorical cross entropy (15) as loss function [41], and then measure the final performances with the accuracy metric (16).

Binary classification.— We consider the binary classification problem ($C = 2$) of discriminating single-qubit states subject either to phase-flip channel \mathcal{Z}_p , or amplitude damping channel $\mathcal{A}_{p\gamma}$. The training set is generated by sampling a number $|\mathcal{T}|$ of random uniform single-qubit states and evolving half of them with the phase-flip channel, and the remaining half with the amplitude damping one. Classification accuracies at the end of the training procedure are reported in Table III.

Remarkably, with a training set containing $|\mathcal{T}_{IN}| = 300$ samples, our model exhibits very good classification performances, reaching perfect accuracy $\text{ACC} = 1$. In this case, we noticed that even with a reduced dataset comprising just a few dozen samples, the model is able to achieve almost perfect accuracy, even though the learning process becomes perceptibly less stable.

On the other hand, with the noisy training set \mathcal{T}_N with $|\mathcal{T}_N| = 300$ samples, the best accuracy obtained was $\text{ACC} = 0.92$. As expected, as the noisy dataset \mathcal{T}_N contains only information about the noisy Bloch vectors (and a label for the noisy channel that created them), more data is needed to reach good classification performances. Indeed, higher accuracies can then be obtained by using larger training datasets: for example, an accuracy of $\text{ACC} = 0.98$ can be achieved with a training set containing 800 samples.

Note that when using dataset \mathcal{T}_N , learning the relation between noisy Bloch vectors belonging to the same channel implicitly translates to that of reconstructing the shape of the Bloch sphere generated by the two channels, which are reported in Fig. 4. However, as clear from the graphical representation, there is a nontrivial in-

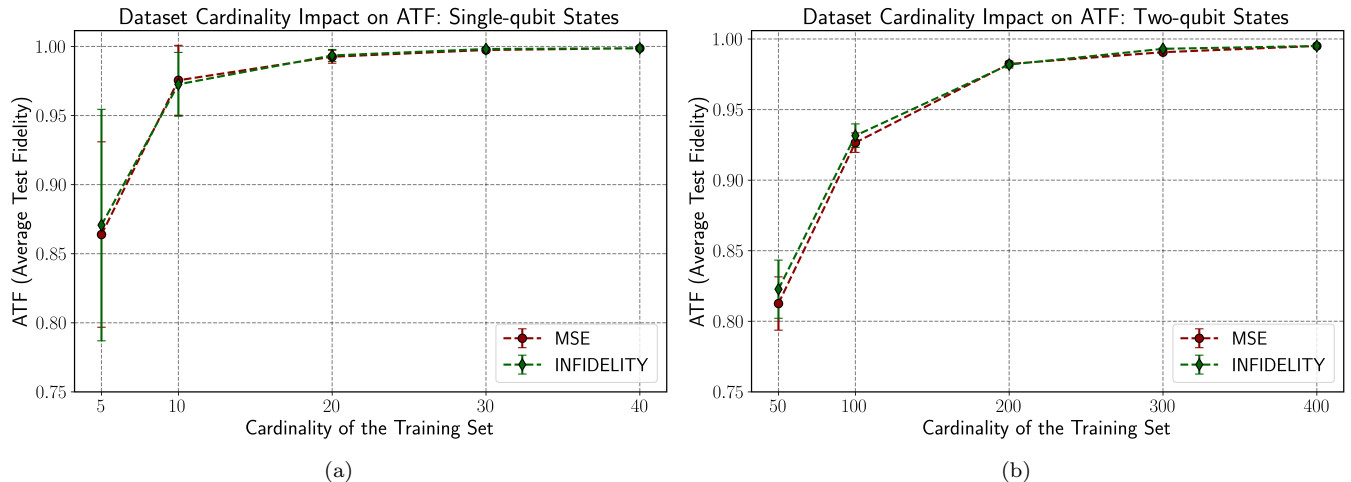


FIG. 3. Average Test Fidelity (ATF) obtained at the end of training when optimising the neural networks with training sets of different cardinality. For each cardinality, we repeat the training process 5 times using different initialisation of the parameters and different training data, and report the mean value and the standard deviation of the resulting ATFs. (a) Reconstruction of single-qubit states undergoing a phase-flip channel $\mathcal{Z}_p(p = 0.2)$. (b) Reconstruction of two-qubit states undergoing the uncorrelated phase-flip channel $\mathcal{Z}_p \otimes \mathcal{Z}_p(p = 0.2)$.

tersection between the two Bloch spheres, which implies that certain samples could reasonably be assigned to either classes. The imperfect accuracy and the general decrease in classification accuracy for the noise-only training dataset \mathcal{T}_N shown in Table III, can then be explained by such inherent ambiguity in the dataset which makes the classification task more difficult—if not impossible—to solve exactly.

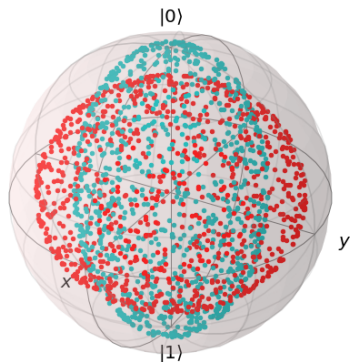


FIG. 4. Deformed Bloch spheres obtained by applying a phase-flip channel $\mathcal{Z}_p(p = 0.2)$ (light blue), and a generalised amplitude damping channel $\mathcal{A}_{p\gamma}(p = 0.5, \gamma = 0.3)$, to a set of uniformly distributed pure states. Note the non-trivial intersection between the two ellipsoids.

Multi-class classification.— As a straightforward extension to the previous analysis, we also report results for the case of a multi-class classification task, where the network is asked to classify states generated by three different channels (phase-flip, amplitude damping and de-

polarizing channel), using a dataset of type \mathcal{T}_{IN} , that is containing both ideal and noisy Bloch vectors. We find again that the network is able to perfectly classify all the states, reaching a perfect final accuracy on a test set $\text{ACC} = 1$, as reported in Table III.

IV. CONCLUSION

In conclusion, our research underscores the remarkable effectiveness of deep neural networks in quantum information processing tasks, specifically in reconstructing and classifying quantum states undergoing unknown noisy evolutions.

This study, as exemplified in our results, showcases the successful recovery of ideal (generalized) Bloch vectors with fidelities exceeding 0.99, even for quantum states involving up to three qubits, under different correlated and uncorrelated noisy channels, and using both classical (mean squared error) and quantum-inspired (infidelity) loss functions for training.

Furthermore, our investigation demonstrates the versatility of our neural network approach in classification problems, adeptly handling a wide range of noise patterns and consistently achieving remarkable classification accuracy across all test samples. Notably, in the context of discriminating between phase-flip and amplitude damping channels, our model achieves an outstanding classification accuracy of 98%, highlighting its remarkable capacity to discern the relationships between states affected by similar noise sources, even when presented with the noisy vectors alone.

As we look ahead, an intriguing avenue for further exploration lies in examining the intricate connections be-

tween various fidelity measures [46] used as training loss functions and their impact on the resulting test fidelities. This pursuit aims to identify the fidelity metric best suited to the specific characteristics of the problem at hand. Such investigations promise to advance our understanding of quantum information processing and open new horizons for practical applications in quantum technology.

V. ACKNOWLEDGEMENTS

A.R.M. acknowledges support from the PNRR MUR Project PE0000023-NQSTI. C.M. acknowledges support from the National Research Centre for HPC, Big Data and Quantum Computing, PNRR MUR Project CN0000013-ICSC, and from the EU H2020 QuantERA ERA-NET Cofund in Quantum Technologies project QuICHE.”

Appendix A: Noise Channels

In this appendix we present the noise quantum channels used in the simulations to generate the datasets containing noisy Bloch vectors.

a. Bit-flip Channel.— The bit-flip channel flips the qubit state from $|0\rangle$ to $|1\rangle$ and vice versa with probability p . Given the operator-sum representation in Eq. (2), its Kraus operators are

$$E_0 = \sqrt{1-p} \mathbb{I} \quad E_1 = \sqrt{p} X, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad (\text{A1})$$

It is possible to represent the deformation that occurs on the Bloch sphere after the bit-flip noise occurred: the states on the \hat{x} axis are left alone while the $\hat{y} - \hat{z}$ plane is uniformly contracted by a factor $2p$.

b. Phase-flip channel.— This channel changes the sign of the component associated to the element of the computational basis $|1\rangle$ of the qubit state. The channel is represented by the operation elements

$$E_0 = \sqrt{1-p} \mathbb{I} \quad E_1 = \sqrt{p} Z, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (\text{A2})$$

On the Bloch sphere, while the $\hat{x} - \hat{y}$ plane is contracted by a factor $2p$, the states on the \hat{z} axis are left untouched.

c. Bit-phase-flip channel.— This channel is a combination of a bit-flip and a phase-flip channel. Recalling that $Y = iXZ$, the Kraus operators of this channel are

$$E_0 = \sqrt{1-p} \mathbb{I} \quad E_1 = \sqrt{p} Y, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \quad (\text{A3})$$

This noise acts on the Bloch sphere by leaving alone the states on the \hat{y} axis and contracting the $\hat{x} - \hat{z}$ plane by a factor $2p$.

d. General Pauli channel.— The general Pauli channel is a combination of bit-, phase- and bit-phase-flip channels, each one with its own intensity on the correspondent axis [6]. In this case, the operators are

$$E_0 = \sqrt{p_0} \mathbb{I} \quad E_1 = \sqrt{p_1} X, \quad (\text{A4})$$

$$E_2 = \sqrt{p_2} Y \quad E_3 = \sqrt{p_3} Z, \quad (\text{A5})$$

with p_1, p_2 and p_3 probabilities of each error such that $p_0 + p_1 + p_2 + p_3 = 1$.

e. Depolarizing Channel.— The depolarizing channel acts on a quantum state by leaving it untouched with probability $1 - p$, or replacing it with the completely mixed state $\mathbb{I}/2$ with probability p . Its Kraus operators are

$$E_0 = \sqrt{1 - \frac{3p}{4}} \mathbb{I} \quad E_1 = \frac{\sqrt{p}}{2} X \quad (\text{A6})$$

$$E_2 = \frac{\sqrt{p}}{2} Y \quad E_3 = \frac{\sqrt{p}}{2} Z. \quad (\text{A7})$$

In this case the entire Bloch sphere is uniformly contracted by a factor which depends on p . The channel can be generalized for d -dimensional ($d = 2^n$ for n qubits) quantum systems as

$$\mathcal{E}(\rho) = (1 - p)\rho + p \frac{\mathbb{I}_d}{d}. \quad (\text{A8})$$

f. Generalized amplitude damping channel.— This channel can be used to describe the energy dissipation in a quantum system to an environment at finite temperature [30], and can be described with Kraus operators

$$E_0 = \sqrt{p} \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \quad (\text{A9})$$

$$E_1 = \sqrt{p} \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}, \quad (\text{A10})$$

$$E_2 = \sqrt{1-p} \begin{pmatrix} \sqrt{1-\gamma} & 0 \\ 0 & 1 \end{pmatrix}, \quad (\text{A11})$$

$$E_3 = \sqrt{1-p} \begin{pmatrix} 0 & 0 \\ \sqrt{\gamma} & 0 \end{pmatrix}, \quad (\text{A12})$$

where it is possible to define the stationary state

$$\rho_\infty = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix}. \quad (\text{A13})$$

This channel leads to a deformation of the Bloch sphere, where γ regulates the shrinking of each component of the Bloch vector and p indicates which is the fixed point.

g. Correlated amplitude damping channel.— This is a two-qubit noise channel defined as the convex combination of two channels \mathcal{N}_0 and \mathcal{N}_1 [47]

$$\mathcal{N}(\rho) = (1 - \mu)\mathcal{N}_0(\rho) + \mu\mathcal{N}_1(\rho), \quad (\text{A14})$$

where $\mu \in [0, 1]$ is a correlation parameter between the qubits, and $\mathcal{N}_0 = \sum_{j=0}^3 A_j \rho A_j^\dagger$ and $\mathcal{N}_1 = \sum_{j=0}^1 B_j \rho B_j^\dagger$ are noisy channels defined by Kraus operators $A_0 = E_0 \otimes E_0$, $A_1 = E_0 \otimes E_1$, $A_2 = E_1 \otimes E_0$ and $A_3 = E_1 \otimes E_1$ with E_0 and E_1 being the operators in Eqs. (A9) and (A10) with $p = 1$, and

$$B_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \sqrt{\gamma} \end{pmatrix}, \quad (\text{A15})$$

$$B_1 = \begin{pmatrix} 0 & 0 & 0 & \sqrt{1-\gamma} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (\text{A16})$$

Appendix B: Fidelity of single qubit states

As proved in [40], the fidelity formula (8) can be much simplified for one-qubit states and expressed in terms of Bloch vectors. For an hermitian 2×2 matrix M with positive eigenvalues, it holds

$$(\text{Tr} \sqrt{M})^2 = \text{Tr}[M] + 2(\det M)^{1/2}, \quad (\text{B1})$$

and with $M = \sqrt{\rho} \sigma \sqrt{\rho}$ as in the definition of the fidelity (8), one obtains

$$F(\rho, \sigma) = \text{Tr}[\rho\sigma] + 2(\det \rho \det \sigma)^{1/2}. \quad (\text{B2})$$

Finally, expressing the density matrices in terms of their Bloch vectors, namely $\rho = (\mathbb{I} + \mathbf{r} \cdot \mathbf{P})/2$ and $\sigma = (\mathbb{I} + \mathbf{s} \cdot \mathbf{P})/2$ with $\mathbf{P} = (X, Y, Z)$, by explicit calculation one arrives at

$$F(\mathbf{r}, \mathbf{s}) = \frac{1}{2} \left(1 + \mathbf{r} \cdot \mathbf{s} + \sqrt{(1 - \|\mathbf{r}\|^2)(1 - \|\mathbf{s}\|^2)} \right). \quad (\text{B3})$$

With such expression at hand, it is then possible to prove that for pure single-qubit states minimizing the mean squared error loss is equivalent to minimizing infidelity. In fact, consider the Euclidean distance between the Bloch vectors

$$\ell(\mathbf{r}, \mathbf{s}) = \|\mathbf{r} - \mathbf{s}\|^2 = \|\mathbf{r}\|^2 + \|\mathbf{s}\|^2 - 2\mathbf{r} \cdot \mathbf{s}, \quad (\text{B4})$$

and the infidelity

$$I(\mathbf{r}, \mathbf{s}) = 1 - \frac{1}{2} \left(1 + \mathbf{r} \cdot \mathbf{s} + \sqrt{(1 - \|\mathbf{r}\|^2)(1 - \|\mathbf{s}\|^2)} \right). \quad (\text{B5})$$

Then, since we are computing the loss functions with Bloch vectors \mathbf{r} and \mathbf{s} of pure states—the ideal noise-free Bloch vector and the predicted one by the neural network, see (4)—, it holds that $\|\mathbf{r}\| = \|\mathbf{s}\| = 1$, thus obtaining

$$I(\mathbf{r}, \mathbf{s}) = 1 - \frac{1}{2}(1 + \mathbf{r} \cdot \mathbf{s}) \longrightarrow \mathbf{r} \cdot \mathbf{s} = -2I(\mathbf{r}, \mathbf{s}) + 1, \quad (\text{B6})$$

and substituting in Eq. (B4) one finally arrives at

$$\ell(\mathbf{r}, \mathbf{s}) = 4I(\mathbf{r}, \mathbf{s}). \quad (\text{B7})$$

-
- [1] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Hugins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, arXiv preprint arXiv:2210.00921 (2022).
 - [2] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, *Nature* **567**, 491 (2019).
 - [3] E. van den Berg, Z. K. Mineev, and K. Temme, Model-free readout-error mitigation for quantum expectation values, *Phys. Rev. A* **105**, 032620 (2022).
 - [4] A. W. R. Smith, K. E. Khosla, C. N. Self, and M. S. Kim, Qubit readout error mitigation with bit-flip averaging, *Science Advances* **7**, eabi8009 (2021), <https://www.science.org/doi/pdf/10.1126/sciadv.abi8009>.
 - [5] S. Bravyi, S. Sheldon, A. Kandala, D. C. McKay, and J. M. Gambetta, Mitigating measurement errors in multiqubit experiments, *Physical Review A* **103**, 042605 (2021).
 - [6] S. Mangini, L. Maccone, and C. Macchiavello, Qubit noise deconvolution, *EPJ Quantum Technology* **9**, 29 (2022).
 - [7] S. Roncallo, L. Maccone, and C. Macchiavello, Multi-qubit noise deconvolution and characterization, *Phys. Rev. A* **107**, 022419 (2023).
 - [8] E. Van Den Berg, Z. K. Mineev, A. Kandala, and K. Temme, Probabilistic error cancellation with sparse pauli-lindblad models on noisy quantum processors, *Nature Physics*, 1 (2023).
 - [9] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, Evidence for the utility of quantum computing before fault tolerance, *Nature* **618**, 500 (2023).
 - [10] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Error mitigation with Clifford quantum-circuit data, *Quantum* **5**, 592 (2021).
 - [11] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, and

- W. J. Zeng, Digital zero noise extrapolation for quantum error mitigation, in *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, 2020) pp. 306–316.
- [12] A. Lowe, M. H. Gordon, P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Unified approach to data-driven quantum error mitigation, *Physical Review Research* **3**, 033098 (2021).
- [13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need (2023), arXiv:1706.03762 [cs.CL].
- [14] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, Language models are few-shot learners (2020), arXiv:2005.14165 [cs.CL].
- [15] J. Deng and Y. Lin, The benefits and challenges of chatgpt: An overview, *Frontiers in Computing and Intelligent Systems* **2**, 81 (2022).
- [16] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016) pp. 770–778.
- [17] U. Kamath, J. Liu, and J. Whitaker, *Deep learning for NLP and speech recognition*, Vol. 84 (Springer, 2019).
- [18] J. Schrittwieser, I. Antonoglou, T. Hubert, K. Simonyan, L. Sifre, S. Schmitt, A. Guez, E. Lockhart, D. Hassabis, T. Graepel, T. Lillicrap, and D. Silver, Mastering atari, go, chess and shogi by planning with a learned model, *Nature* **588**, 604 (2020).
- [19] A. Lumino, E. Polino, A. S. Rab, G. Milani, N. Spagnolo, N. Wiebe, and F. Sciarrino, Experimental phase estimation enhanced by machine learning, *Physical Review Applied* **10**, 044033 (2018).
- [20] T. Scholten, Y.-K. Liu, K. Young, and R. Blume-Kohout, Classifying single-qubit noise using machine learning (2019).
- [21] A. Zlokapa and A. Gheorghiu, A deep learning model for noise prediction on near-term quantum devices (2020).
- [22] G. Torlai, G. Mazzola, G. Carleo, and A. Mezzacapo, Precise measurement of quantum observables with neural-network estimators, *Physical Review Research* **2**, 022060 (2020).
- [23] J. Kim, B. Oh, Y. Chong, E. Hwang, and D. K. Park, Quantum readout error mitigation via deep learning, *New Journal of Physics* **24**, 073009 (2022).
- [24] C. Kim, K. D. Park, and J.-K. Rhee, Quantum error mitigation with artificial neural network, *IEEE Access* **8**, 188853 (2020).
- [25] S. H. Sack and D. J. Egger, Large-scale quantum approximate optimization on non-planar graphs with machine learning noise mitigation (2023).
- [26] J. Walln fer, A. A. Melnikov, W. D r, and H. J. Briegel, Machine learning for long-distance quantum communication, *PRX Quantum* **1**, 010301 (2020).
- [27] S. Ahmed, C. S. Mu oz, F. Nori, and A. F. Kockum, Classification and reconstruction of optical quantum states with deep neural networks, *Physical Review Research* **3**, 033278 (2021).
- [28] S. Lohani, B. T. Kirby, M. Brodsky, O. Danaci, and R. T. Glasser, Machine learning assisted quantum state estimation, *Machine Learning: Science and Technology* **1**, 035007 (2020).
- [29] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, *Neural networks* **2**, 359 (1989).
- [30] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge university press, 2010).
- [31] M. Ozols and L. Man cinska, Generalized bloch vector and the eigenvalues of a density matrix (2007), available online at: <http://home.lu.lv/~sd20008/papers/essays.html>.
- [32] A. Edelman and N. R. Rao, Random matrix theory, *Acta Numerica* **14**, 233–297 (2005).
- [33] E. S. Meckes, *The Random Matrix Theory of the Classical Compact Groups*, Cambridge Tracts in Mathematics (Cambridge University Press, 2019).
- [34] R. Harman and V. Lacko, On decompositional algorithms for uniform sampling from n-spheres and n-balls, *Journal of Multivariate Analysis* **101**, 2297 (2010).
- [35] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo method* (John Wiley & Sons, 2016).
- [36] K. Życzkowski, K. A. Penson, I. Nechita, and B. Collins, Generating random density matrices, *Journal of Mathematical Physics* **52**, <https://doi.org/10.1063/1.3595693> (2011).
- [37] J. Ginibre, Statistical ensembles of complex, quaternion, and real matrices, *Journal of Mathematical Physics* **6**, 440 (1965).
- [38] K. Gulshen, J. Combes, M. P. Harrigan, P. J. Karalekas, M. P. da Silva, M. S. Alam, A. Brown, S. Caldwell, L. Capelluto, G. Crooks, D. Girshovich, B. R. Johnson, E. C. Peterson, A. Polloreno, N. C. Rubin, C. A. Ryan, A. Staley, N. A. Tezak, and J. Valery, Forest Benchmarking: QCVV using PyQuil (2019).
- [39] X. Wang, C.-S. Yu, and X. Yi, An alternative quantum fidelity for mixed states of qudits, *Physics Letters A* **373**, 58 (2008).
- [40] R. Jozsa, Fidelity for mixed quantum states, *Journal of Modern Optics* **41**, 2315 (1994).
- [41] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016).
- [42] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization (2014).
- [43] Qiskit contributors, Qiskit: An open-source framework for quantum computing (2023).
- [44] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems (2015), software available from tensorflow.org.
- [45] J. Avron and O. Kenneth, An elementary introduction to the geometry of quantum states with pictures, *Reviews in Mathematical Physics* **32**, 2030001 (2020).
- [46] Y.-C. Liang, Y.-H. Yeh, P. E. Mendon a, R. Y. Teh, M. D. Reid, and P. D. Drummond, Quantum fidelity measures for mixed states, *Reports on Progress in Physics* **82**, 076001 (2019).
- [47] A. D’Arrigo, G. Benenti, G. Falci, and C. Macchiavello, Classical and quantum capacities of a fully correlated amplitude damping channel, *Phys. Rev. A* **88**, 042337 (2013).