Simplified Quantum Process Characterization by Specialised Neural Networks

M. Guarneri, 1 I. Gianani, 2 M. Barbieri, 2, 3 and A. Chiuri 1

1 ENEA - Centro Ricerche Frascati, via E. Fermi 45, 00044 Frascati, Italy
2 Dipartimento di Scienze, Università degli Studi Roma Tre, Via della Vasca Navale 84, 00146, Rome, Italy
3 Istituto Nazionale di Ottica (INO-CNR), Via E. Fermi 6, I-50125 Firenze, Italy

(Dated: November 22, 2022)

Characterization of quantum objects based on previous knowledge is a valuable approach, especially as it leads to routine procedures for real-life components. To this end, Machine Learning algorithms have demonstrated to successfully operate in presence of noise. However, there might be instances in which unknown parasitic effects occur in tandem with the sought one we aim at characterizing. Here we show that the accurate design of a two-stage neural network can account for these class of disturbances as well, applying our technique to the characterization of several quantum channels. We demonstrate that a stable and reliable characterization is achievable by training the network only with simulated data. The obtained results show the viability of this approach as an effective tool based on a completely new paradigm for the employment of NNs in the quantum domain.

Introduction. Routine characterisation of quantum objects in a laboratory takes advantage from the fact that samples can often be collected in arbitrary numbers. This prevents fluctuations in the recorded counts to hamper the reconstruction: in a frequentist view, the experimental measurement frequencies can be made to match the probabilities within arbitrary accuracy. Good level of control also ensures the object itself to belong to a known class, allowing to apply constraints and restricted methods for its characterisation [1–6].

For real-life quantum devices, however, these assets are hindered by multiple effects. A limited number of copies can be collected in reasonable times. Direct reconstruction methods are known to suffer from these fluctuations, and may even result in nonphysical outcomes: maximum likelihood algorithms help excluding these pathological instances [7–9], but do not warrant an unbiased output with small samples. Furthermore, the main result expected from the devices can be accompanied by processes of unknown nature. Nevertheless, the characterisation problem can still be cast in terms of effective parameters pertaining to the sought effect; for instance, we can be interested in assessing the depolarising action of a quantum channel, regardless of its exact form.

Between us and our aim two factors stand: the first is the uncertainty on the experimental frequency, the second is the presence of parasitic effects. A reliable procedure able to extract effective parameters, being also robust against these effects is highly desirable. Machine learning (ML) techniques have been successfully adopted to tackle the first of these two limitations for black-box characterisation of quantum objects, with realisations including state and unitary reconstruction [10–17], validation of quantum technology [18–19], identification of quantum features [20–21], and the calibration of quantum sensors [22–24]. The versatility of ML has allowed to extend beyond these scopes to the adaptive control of quantum devices [25–36], as well as to the automated design of quantum experiments [37–44]. Realising this vision in full requires further exploration to understand the capabilities and opportunities offered by progress in ML [45–46]. Specifically, the simple networks generally employed disregard the correlations existing among different elements in the state or process matrix. This could provide extra means for regularisation, a possibility currently under intense scrutiny [47].

In this article, we show how to extend the use of ML techniques coming to an intriguing solution to this problem. A specialised neural network (NN) structure is built drawing inspiration from denoising procedures for imaging: indeed, each element of a process matrix can be thought as a pixel, thus incorporating convolutional layers in the NN can efficiently address their correlations. We illustrate these ideas with numerical examples of single- and two-qubit channels to unearth the impact of the statistics. The same techniques are also applied to real data to investigate the role of systematic experimental deviations. We demonstrate that it is possible to get useful results by training the NNs exclusively with a simulated dataset, thus without implementing numerous repetitions of the experiment.

Methods. We consider the action of a quantum channel, from which we isolate relevant features. This means we are interested in providing a description as a map \( \Gamma_{(q)} \), where the set of parameters \( q_i \) captures the essentials of our description. In the practice, however, the experimental reconstruction may lead to a different map \( \Gamma' \), due to the presence of unexpected effects or simply as a consequence of the finite statistics. Our aim is then to extract viable information on effective parameters \( q_i \), i.e. on the physical effect they describe, from \( \Gamma' \), while curtailing the consequent bias. In fact, a best fit procedure, minimising the distance between \( \Gamma_{(q)} \) and \( \Gamma' \), would be affected by the aforementioned discrepancies. ML techniques may improve the results as NNs can be trained to account for these fluctuations. Since the uncertainty any specific measurement has a nontrivial influence over all the elements of the process matrix, then it would be beneficial to include the presence of correlations explicitly in the network architecture.

In the present work we consider different examples of quantum channels: depolarising channel (DC, single parameter \( p \)); generalised amplitude damping (GAD, two parameters \( \eta, \gamma \)); a control-phase two-qubit gate (CP, single parameter \( \phi \)) – their explicit form is presented in [48]. Our parameter extraction
technique is based mainly on an architecture composed by two cascaded convolutional NNs (CNNs): the first reduces noise as statistical fluctuations in the simulated data for the quantum channels, the second infers the parameters \( q_i \), characterizing the channel itself. A schematic depiction is presented in Fig. 1.

Denoising is carried out by means of an autoencoder Neural Network (aNN) architecture, composed by two main instances, called respectively encoder and decoder. We input the aNN with the noisy complex process matrix, which the network interprets as a two-channel gray-level image. The purpose of the aNN is to populate the latent space only with the relevant components extracted with 2D convolutional layers of the encoder, and then reconstruct the complex matrix processing the components in the latent space with the transposed 2D convolutional layers of the decoder \([49, 50]\). The aNN outputs a denoised matrix, which is in turn used as input for the second architecture, constituted by a Feed-Forward CNN (FFNN). The FFNN is composed by three convolutional layers followed by three dense layers and optimized for carrying out a nonlinear regression. The network structure needs to be adjusted to the specific channel: indeed, the kernel size of the convolutional layers depends on the size of the input matrix, the three dense layers are repeated a number of times equal to the number of parameters to be extracted, which also dictates the number of neurons in the final layers. The efficiency of the FFNN is shown to be strongly dependent on that of goodness of the denoising processing of the aNN.

The training of the whole NN includes 100 instances of simulated random matrices \( \hat{\rho}^{(i)} \) for each value of \( p \in [0.05, 1] \) for the DC, or 500 matrices for the GAD \( (\gamma, \eta) \in [0,1] \) in steps of 0.1, and CP \( (\phi = m_1 \cdot \pi/6 \text{ with } m_1 = [0,11] \text{ and } \phi = m_2 \cdot \pi/4 \text{ with } m_2 = 1,3,5,7) \). The matrices themselves are obtained from simulated process tomography data from a population of \( n_i \) events on average; we have then considered three signal levels \( n_i = k_i \cdot n \) with \( n = 2000 \), \( k_i = 1,0.5,1 \). Both blocks of the NN are trained using back-propagation, adopting the mean squared error (MSE) as loss function. The dataset is split between training and validation, with a 0.8 – 0.2 ratio. The two-block NN reaches a MSE as low as \((\approx 10^{-3})\) after \( \approx 50 \) epochs during validation, which can be further improved by increasing the number of epochs. More details on the generation of the dataset are contained in the Supplementary Material.

Results on DC. As a first test, we consider the DC and the instance in which the data are only affected by Poissonian noise associated to the simulated measured counts. As the relevant figure of merit, we considered the residue \(|p_{\text{meas}} - p_{\text{set}}|\) between the inferred \( p_{\text{meas}} \) and the set \( p_{\text{set}} \) values of \( p \) for three different approaches: our optimised network; a simpler network only adopting the FF block; the standard evaluation taking the highest fidelity with a state in the form \( \Gamma_{\rho|\rho} = (1 - p) \rho + p/3 \sum_{i=1}^{3} \sigma_i \rho \sigma_i \), what we term Maximum Fidelity (MF). The summary in Fig. 2 shows that our two-stage network always outperforms the simple MF, while it still offer some advantage over the use of the FF only, especially far from low values of \( p \).

The main advantage of our approach unfolds when considering instances with parasitic processes. For this second test...
case, we then employed our NN to data generated in a real experiment [31], where we expect deviations from the ideal depolarising channel [48]. As before, the training was still operated based on simulated data from the ideal process. This is aimed at accounting for the expected fluctuations in the data, with an ideal output, without incorporating knowledge on the actual process. This implies that the matrix serving as the input of the FF network should not be interpreted as the experimental \( \chi \) matrix. This is rather the representation of the closest DC to the data.

The data were collected for the expected values \( p_{\exp} = 0.09, 0.3, 0.64, 1 \). In order to investigate the impact of the Poissonian noise on the reconstruction we then generated 300 matrices from the data corresponding to each \( p_{\exp} \) for each rescaling factor \( k_i^{\exp} = 0.025, 0.05, 0.1, 0.25, 0.5, 0.75, 1 \) on the overall average number of measured counts \( n_{\exp} \), i.e. \( n_i^{\exp} = k_i^{\exp} n_{\exp} \) with \( n_i^{\exp} \approx 2000 \) comparable with \( n_s \). The results obtained with the experimental dataset are reported in Fig.3. There, we plot the estimated values of \( p \) as a function of the rescaling factor for our four cases using both our NN, and a simple procedure based on maximising the fidelity. Overall, we assess a satisfactory behaviour in the evaluation of \( p \). The best performances is achieved for small values of \( p \) (i.e. \( p = 0.09, 0.3 \)), representing the most interesting cases, in that the noise can easily overcome the channel: the adoption of the NN analysis seems to be able to mitigate biases, and this holds true down to \( k_i \approx 0.5 \) for \( p = 0.09, 0.64, 1 \) and \( k_i \approx 0.25 \) for \( p = 0.3 \), showing that the presence of parasitic processes does not introduce costs in terms of resources needed. For higher values of \( p \) (i.e. \( p = 0.64, 1 \)), instead, the performance of the two methods is aligned, with marginal amelioration still appreciable for \( k_i \geq 0.5 \). This difference can be traced to the form of the process: for low \( p \), uncertainties due to the statistics are more relevant, as diagonal terms in the matrix are smaller.

**Results for GAD and CP.** We now consider different, more involved examples of quantum channels. The first example concerns the GAD, characterised by a damping parameter \( \gamma \), and by the probability \( \eta \) of dissipating towards either level of the qubit. As a more compact figure to assess the effectiveness of our cascaded network, we adopt the success probability of obtaining residues for the parameters below 0.1 for at least 99% of the sample. This is quantified as numerical frequencies over samples of 500 instances - and repeated for different rescaling factors \( k_i \). The results are summarised in Table I presenting typical occurrences. Even for the lowest signal level, \( k_1 = 0.1 \), our combined CNNs achieve superior performance with respect to the simple application of a FFNN, thus demonstrating the usefulness of the autoencoder.

We have adopted a similar approach to test the performance on a CP gate. The figure of merit is, however, slightly modified so to make it more appropriate to the case in point: we consider the probability of achieving an error permitting a full discrimination of all our angles. Even at the lowest signal level, almost all events show such discrimination ability, with a significant improvement over the use of FFNN only

| Parameter | \( k_1 = 0.1 \) | \( k_2 = 0.5 \) | \( k_3 = 1 \) |
|-----------|-----------------|-----------------|-----------------|
| \( \eta \) | 0 % | 37.6 % | 13.6 % | 75.3 % | 34.6 % | 84 % |
| \( \gamma \) | 0 % | 16 % | 42 % | 100 % | 88.8 % | 100 % |

[See Table I. This is also indirectly inferred by comparing the matrices obtained after the aNN with those obtained with the usual maximum likelihood routine. While both algorithms perform denoising, aNNs achieve better performance in that the fidelity with the expected matrix is always \( \geq 0.99 \), with a significant fraction of values being 1, even at the lowest signal, \( k_1 = 0.1 \). With the maximum likelihood, instead, the fidelity generally remains below 0.99.]

**Table II. Summary of the results for CP.** For each value of \( \phi \), we tested our approach over a sample of 500 noisy matrices. We report the success rate for both the absolute and the relative residue (\( \leq 3\% \)).

| Residues | \( k_1 = 0.1 \) | \( k_2 = 0.5 \) | \( k_3 = 1 \) |
|----------|-----------------|-----------------|-----------------|
| \( \phi \) (Abs) | 4.7 % | 99.8 % | 10.5 % | 100 % | 11.1 % | 100 % |
| \( \phi \) (%) | 3.4 % | 92.6 % | 6.6 % | 96.3 % | 6.7 % | 98.5 % |

More detail on the analysis of GAD and CP are found in [48].

**Discussion.** In this work we have presented a model unveiling the potential of convolutional NNs to achieve robust characterisation of quantum processes from noisy measurements: the relation between the observed frequencies and underlying parameters of the measured process were approximated by a two-stage NNs. We applied the presented approach to both numerical and experimental dataset, demonstrating that the employment of these tools tolerates at least a 50% reduction in the number of collected counts even under unfavourable conditions. A key result is the adoption solely of simulated data for the training. Despite this, our NNs architecture has correctly identified the new scenarios set by \( p_{\exp} \) and \( k_i^{\exp} \). It is then possible to guarantee stability and reliability, while minimizing the resources and time necessary to train the NNs.

The NN approach can then provide accurate characterisations of the parameters, with a more satisfactory performance than the direct use of estimates based on the MF or the denoising via maximum likelihood methods. The computational resources, however, do not result heavily aggravated as a consequence. This lays a promising route for routine employment in more complex instances. Indeed, our tests have addressed different quantum channels, and the two-stage NN have been optimized for the particular process. However, they covered a broad class, making us confident that the result obtained with this demonstration may have general validity. Our approach
then allows to characterize the process, reduce the impact of the noise, and extract the parameters $q_i$, all of it optimizing
the amount of necessary resources for real-life scenarios.

Acknowledgments.- The authors thank P. Mataloni for granting access to the data and C. Macchiavello for fruitful discussion. This work was supported by the European Commission (FET-OPEN-RIA STORMYTUNE, Grant Agreement No. 899587), and by the NATO SPS Project HADES - MYP G5839.

[1] O. Gühne, P. Hyllus, D. Bruß, A. Ekert, M. Lewenstein, C. Macchiavello, and A. Sanpera, Phys. Rev. A. 66, 062305 (2002).
[2] M. Barbieri, F. De Martini, G. Di Nepi, P. Mataloni, G. M. D’Ariano, and C. Macchiavello, Phys. Rev. Lett. 91, 227901 (2003).
[3] M. Bourennane, M. Eibl, C. Kurtsiefer, S. Gaertner, H. Weinfurter, O. Gühne, P. Hyllus, D. Bruß, M. Lewenstein, and A. Sanpera, Phys. Rev. Lett. 92, 087902 (2004).
[4] M. P. A. Branderhorst, J. Nunn, I. A. Walmsley, and R. L. Kosut, New Journal of Physics 11, 115010 (2009).
[5] R. Blandino, F. Ferreyrol, M. Barbieri, P. Brangier, and R. Tualle-Brouri, New Journal of Physics 14, 013017 (2012).
[6] A. Tipsmark, R. Dong, A. Laghaout, P. Marek, M. Ježek, and U. L. Andersen, Phys. Rev. A 84, 050301 (2011).
[7] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Phys. Rev. A 64, 052312 (2001).
[8] A. I. Lvovsky, Journal of Optics B: Quantum and Semiclassical Optics 6, S556 (2004).
[9] A. Anis and A. I. Lvovsky, New Journal of Physics 14, 105021 (2012).
[10] N. Spagnolo, E. Maiorino, C. Vitelli, M. Bentivegna, A. Crespi, R. Ramponi, P. Mataloni, R. Osellame, and F. Sciarrino, Scientific Reports 7, 14316 (2017).
[11] J. Gao, L.-F. Qiao, Z.-Q. Jiao, Y.-C. Ma, C.-Q. Hu, R.-J. Ren, A.-L. Yang, H. Tang, M.-H. Yung, and X.-M. Jin, Phys. Rev. Lett. 120, 240501 (2018).
[12] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Nature Physics 14, 447 (2018).
[13] A. Rocchetto, S. Aaronson, S. Severini, G. Carvacho, D. Polder, I. Agresti, M. Bentivegna, and F. Sciarrino, Science Advances 5, eaau1946 (2019).
[14] G. Torlai, B. Timar, E. P. L. van Nieuwenburg, H. Levine, A. Omran, A. Keesling, H. Bernien, M. Greiner, V. Vuletić, M. D. Lukin, R. G. Melko, and M. Endres, Phys. Rev. Lett. 123, 230504 (2019).
[15] T. Giordani, A. Suprano, E. Polino, F. Acanfora, L. Innocenti, A. Ferraro, M. Paternostro, N. Spagnolo, and F. Sciarrino, Phys. Rev. Lett. 124, 160401 (2020).
[16] A. M. Palmieri, E. Kovlakov, F. Bianchi, D. Yudin, S. Straupe, J. D. Biamonte, and S. Kulik, npj Quantum Information 6, 20 (2020).
[17] E. S. Tiunov, V. T. (Vyborova), A. E. Ulanov, A. I. Lvovsky, and A. K. Fedorov, Optica 7, 448 (2020).
[18] I. Agresti, N. Viggianiello, F. Flamini, N. Spagnolo, A. Crespi, R. Osellame, N. Wiebe, and F. Sciarrino, Phys. Rev. X 9, 011013 (2019).
[19] F. Flamini, N. Spagnolo, and F. Sciarrino, Quantum Science and Technology 4, 024008 (2019).
[20] V. Gebhart and M. Bohmann, Phys. Rev. Research 2, 023150 (2020).
[21] V. Cimini, M. Barbieri, N. Treps, M. Walschaers, and V. Parigi, Phys. Rev. Lett. 125, 160504 (2020).
[22] A. Hentschel and B. C. Sanders, Phys. Rev. Lett. 104, 063603 (2010).
[23] V. Cimini, I. Gianani, N. Spagnolo, F. Leccese, F. Sciarrino, and M. Barbieri, Phys. Rev. Lett. 123, 230502 (2019).
[24] V. Cimini, E. Polino, M. Valeri, I. Gianani, N. Spagnolo, G. Corrèll, A. Crespi, R. Osellame, M. Barbieri, and F. Sciarrino, Phys. Rev. Applied 15, 044003 (2021).
[25] A. Hentschel and B. C. Sanders, Phys. Rev. Lett. 107, 233601 (2011).
[26] N. B. Lovett, C. Cronsier, M. Perarnau-Llobet, and B. C. Sanders, Phys. Rev. Lett. 110, 220501 (2013).
[27] C. Bonato, M. S. Blok, H. T. Dinani, D. W. Berry, M. L. Markham, D. J. Twitchen, and R. Hanson, Nature Nanotechnology 11, 247 (2016).
[28] P. Patilapongpim, P. Wittek, E. Zahedinejad, S. Vedaie, and B. C. Sanders, Neurocomputing 268, 116 (2017).
[29] J. Liu and H. Yuan, Phys. Rev. A 96, 042114 (2017).
[30] S. Paesani, A. A. Gentile, R. Santagati, J. Wang, N. Wiebe, D. P. Tew, J. L. O’Brien, and M. G. Thompson, Phys. Rev. Lett. 118, 100503 (2017).
[31] A. Lumino, E. Polino, A. S. Rab, G. Milani, N. Spagnolo, N. Wiebe, and F. Sciarrino, Phys. Rev. Applied 10, 044033 (2018).
[32] P. Patilapongpim and B. C. Sanders, Phys. Rev. A 100, 012106 (2019).
[33] H. T. Dinani, D. W. Berry, R. Gonzalez, J. R. Maze, and C. Bonato, Phys. Rev. B 99, 125413 (2019).
[34] G. Liu, M. Chen, Y.-X. Liu, D. Layden, and P. Cappellaro, Machine Learning: Science and Technology 1, 015003 (2020).
See Supplementary Information for the explicit form of the quantum channels, a description of the experiment, and additional plots and tables for the analysis.
SUPPLEMENTARY MATERIAL

Quantum channels.

A DC transforms an input qubit state $\rho$ as $\Gamma_{[\rho]}[\rho] = \sum_{i=0}^{3} p_i \sigma_i \rho \sigma_i$, with $p_0 = 1 - p$, $p_1 = p_2 = p_3 = p/3$, and $\sigma_i$ the Pauli matrices $\sigma_0 = I$, $\sigma_1 = \sigma_x$, $\sigma_2 = \sigma_y$, $\sigma_3 = \sigma_z$. The associated process matrix $\chi$ acting on one qubit of the Bell state $|\phi^+\rangle$, reads as follows:

$$\chi_{DC} = \Gamma_{[\rho]} \otimes I \left[ |\phi^+\rangle \langle \phi^+| \right] = \begin{pmatrix} 1 - p & 0 & 0 & 0 \\ 0 & p/3 & 0 & 0 \\ 0 & 0 & p/3 & 0 \\ 0 & 0 & 0 & p/3 \end{pmatrix},$$

according to the expression in the Pauli basis.

The generalized qubit amplitude damping channel (GAD) is a two parameter family of channels where the parameters $\eta$ and $\gamma$ are between zero and one. The channel has a qubit input and qubit output and its superoperator has the form:

$$\chi_{GAD} = \Gamma_{[\rho,\eta]} \otimes I \left[ |\phi^+\rangle \langle \phi^+| \right] = \sum_{i=0}^{3} K_i \rho K_i^\dagger,$$

where $K_0 = \sqrt{1 - \gamma} |0\rangle \langle 0| + \sqrt{1 - \eta} |1\rangle \langle 1|$, $K_1 = \sqrt{\eta} |1\rangle \langle 0| + |0\rangle \langle 1|$, and $K_2 = \sqrt{\gamma} |0\rangle \langle 1| + |1\rangle \langle 0|$. The C-Phase process is expressed by a superoperator $\epsilon$, which represents a quantum gate acting on an input state $\rho_{in}$:

$$\epsilon_{[\phi]}(\rho_{in}) = \sum_{m,n} \chi_{CP} \hat{A}_m \rho_{in} \hat{A}_n^\dagger$$

where $\hat{A}_m$ are a basis for operators acting on $\rho_{in}$. The matrix $\chi_{CP}$ completely describes the process of the gate operation. Ideal C-Phase gate can be decomposed as $CP = (I \otimes I + I \otimes \sigma_z + \sigma_z \otimes I - \sigma_z \otimes \sigma_z)/2$.

**Process Tomography.** The reconstruction of the quantum process matrices $\chi_{ij}$ is carried out by means of AAQPT [52] and QPT [53].

The action of the AAQPT can be described with the following procedure: i) preparing a two-qubit maximally entangled state ii) sending one of the two entangled qubits through the channel $E$; iii) reconstructing the output two-qubit state by state tomography and obtain. The density matrix $\chi$ from the two-qubit output density matrix is the process matrix.

**Process Matrices.** In Fig. 6-14 we show the matrices obtained by employing the aNN for the 4 experimental DCs, i.e. 4 different $p_{exp}$ implemented for this paper, and for several simulated noisy GAD and CP channels serving as inputs to the FFNN. For each process we also report the expected theoretical matrix. We stress that these do not aim at describing the actual process: these are only instrumental to extracting the parameters characterizing each channel in our computational routine.

![Fig. 4. Experimental Setup. The qubit encoded in the green photon is affected by a DC, implemented by suitably activating two LCs. We generated four different channels by varying the parameter $p_{exp} = 0.09, 0.3, 0.64, 1$ by controlling their relative delays and activation times. State tomography was realized by the usual arrangement of polarization analysers placed in front of each detector.](image-url)

**Experimental scheme.**

We briefly discuss the setup employed in Ref. [51]. This is based on a two-photon source producing entangled states in polarisation.

The general Pauli channel (PC) is implemented by means of a sequence of liquid crystal retarders (LC1 and LC2) in the path of one of the photons — see Fig. 4. The LCs act as phase retarders, introducing a relative phase between the ordinary and extraordinary polarisation components depending on the applied voltage $V$. Precisely, $V_x$ and $V_y$ correspond to the
case of LCs operating as a half-waveplate (HWP) and as the identity operator, respectively. The LC1 and LC2 optical axes are set at $0^\circ$ and $45^\circ$ with respect to the $V$-polarization. When the voltage $V_z$ is applied, the LC1 (LC2) acts as a Pauli $\sigma_z$ ($\sigma_x$) on the single qubit. We were able to switch between $V_H$ and $V_z$ in a controlled way and independently for both LC1 and LC2. The simultaneous application of $V_z$ on both LC1 and LC2 corresponds to the $\sigma_y$ operation. We could also adjust the temporal delay between the intervals in which the $V_\pi$ voltage is applied to the two retarders. We define $t_1$, $t_2$, $t_3$ respectively as the activation time of the operators $\sigma_x$, $\sigma_y$ or $\sigma_z$ and $T$ is the period of the LCs activation cycle.

A general PC, i.e. with anisotropic noise, was generated by varying the four time intervals $t_1$, $t_2$, $t_3$ and $T$. These intervals $t_i$ are related to the probabilities $p_i$ ($i = 1, 2, 3$) by the following expression: $p_i = \frac{t_i}{T}$. The probability $p_0$ of the identity operator is given by $p_0 = 1 - \sum p_i$ (with $\delta = t_1 + t_2 + t_3$). The condition $t_1 = t_2 = t_3$ corresponds to the depolarizing channel, i.e. i.e. isotropic noise, with the three Pauli operators acting on the single qubit with the same probability $p = \frac{\delta}{T} = \frac{t_1 + t_2 + t_3}{T}$. In the experiment, the parameter $p_{exp}$ was varied by changing the interval $T$ for a fixed period $\delta$. After the action of the DC the output state was reconstructed with usual state tomography [7].

Discussion about the NNs.

aNN training data set. During the training and validation phase of a NN it is always essential to avoid the overfitting of the network. In this work different strategies were adopted for avoiding the overfitting problem during the training phase of the aNN. For the DC, this was done taking advantage of the symmetry of the matrix Eq. (1) which was reshaped taking the elements in lexicographical order as a 16-element vector. The latter was then divided in four sub-arrays with dimensions $1\oplus5\oplus5\oplus5$ and the last three sub-arrays were exchanged in five
different combinations to give valid matrices, See Fig. 5. The GAD and the CP are not characterized by useful symmetries and a suitable number of random matrices was generated to compose the training data set. The absence of the overfitting was ensured by the loss function trend on each epoch.

Parameter estimation. The properties characterizing the DC were exploited also to further simplify the parameters estimation. Indeed in this case only the four diagonal terms were considered as input of the FFNN. Nevertheless, the results obtained with GAD and CP demonstrate that the proposed approach works also in non-trivial scenarios were the process matrix are not characterized by any symmetry and its elements are not directly related to the parameters.

Results: GAD and CP. We report here further results concerning the employed figures to assess the effectiveness of our cascaded network. For the GAD we adopt the success probability of obtaining residues for the parameters below 0.1 for at least 90% (Table III) and 95% (Table IV) of the sample. The results are summarised also in Fig. 16 where it is confirmed that, even for the lowest signal level at $k_1 = 0.1$, our combined CNNs achieve superior performance with respect to the simple application of a FFNN, thus demonstrating the usefulness of the autoencoder. The Fig. 17 shows that similar results were achieved also for the CP. Fig. 18 demonstrates the fundamental contribution of the aNN allowing to feed the FFNN with denoised matrices.

| Parameter | $k_1 = 0.1$ | $k_2 = 0.5$ | $k_3 = 1$ |
|-----------|------------|------------|------------|
| $\eta$    | 2.5 %      | 44.4%      | 61.7%      |
| $\gamma$  | 22.2%      | 96.3%      | 98.8%      |
| $\eta$    | 0 %        | 27.2 %     | 54.3 %     |
| $\gamma$  | 7.4 %      | 75.3 %     | 98.8 %     |

TABLE III. Summary of the results for GAD. For each pair $[\eta, \gamma]$ we tested our approach over a sample of $\approx 100$ noisy matrices. We report the success rate to obtain a residue $\leq 0.1$ for $> 90\%$ of the testing sample.

| Parameter | $k_1 = 0.1$ | $k_2 = 0.5$ | $k_3 = 1$ |
|-----------|------------|------------|------------|
| $\eta$    | 60.5 %     | 87.7%      | 90.1%      |
| $\gamma$  | 93.8 %     | 100%       | 100%       |
| $\eta$    | 46.9 %     | 81.5%      | 88.9 %     |
| $\gamma$  | 40.7 %     | 100%       | 100%       |

TABLE IV. Summary of the results for GAD. For each pair $[\eta, \gamma]$ we tested our approach over a sample of $\approx 100$ noisy matrices. We report the success rate to obtain a residue $\leq 0.1$ for $> 95\%$ of the testing sample.
FIG. 16. GAD: even for the lowest signal level, \( k_1 = 0.1 \), our combined CNNs achieve a success rate > 90%, i.e. probability of obtaining residues for the parameters below 0.1. We report here an example of what was obtained for all the considered values of parameters. The color map represents the number of test matrices for the particular value of residue.

FIG. 17. CP (\( \phi = 210^\circ \)): even for the lowest signal level, \( k_1 = 0.1 \), our combined CNNs achieve a success rate \( \approx 100\% \), i.e. the probability of achieving an error permitting a full discrimination of all our angles. We report here an example of the results obtained for all the considered values of \( \phi \). For each signal level \( k_i \) we show two histograms: left (right) absolute (percent) value of the residues.

FIG. 18. GAD, \( k_1 = 0.1 \). The employment of the aNN for the denoising allow to improve the success rate as reported also in Tables I and II of the main text. Similar results were obtained for the GAD (i.e. \( \gamma, \eta \)) and the CP (i.e. \( \phi \)). The color map represents the number of test matrices for the particular value of residue.