Article

# Quantum process tomography with unsupervised learning and tensor networks

Giacomo Torlai [1,2] ✉, Christopher J. Wood [3], Atithi Acharya[2,4], Giuseppe Carleo [2,5], Juan Carrasquilla[6] & Leandro Aolita[7,8]

The impressive pace of advance of quantum technology calls for robust and scalable techniques for the characterization and validation of quantum hardware. Quantum process tomography, the reconstruction of an unknown quantum channel from measurement data, remains the quintessential primitive to completely characterize quantum devices. However, due to the exponential scaling of the required data and classical post-processing, its range of applicability is typically restricted to one- and two-qubit gates. Here, we present a technique for performing quantum process tomography that addresses these issues by combining a tensor network representation of the channel with a data-driven optimization inspired by unsupervised machine learning. We demonstrate our technique through synthetically generated data for ideal one- and two-dimensional random quantum circuits of up to 10 qubits, and a noisy 5-qubit circuit, reaching process fidelities above 0.99 using several orders of magnitude fewer (single-qubit) measurement shots than traditional tomographic techniques. Our results go far beyond state-of-the-art, providing a practical and timely tool for benchmarking quantum circuits in current and near-term quantum computers.

Digital quantum computers and analog quantum simulators are entering regimes outside the reach of classical computing hardware[1]. Coherent manipulation of complex quantum states with dozens of qubits have been realized across several platforms, including trapped ions[2,3], Rydberg atom arrays[4], cold atoms in optical lattices[5], and superconducting qubit circuits[6,7]. In particular, programable 2D quantum circuits with between 53[6] and 65[7] qubits, and tens of gate layers, have been run with high fidelity in the latter platforms. Over the next few years, it is expected that quantum devices will attain hundreds of qubits, unlocking a variety of quantum computing applications with far-reaching scientific and technological ramifications.

As the size and complexity of quantum hardware continues to grow, techniques capable of characterizing complex multi-qubit error processes are essential for developing error mitigation for near-term

applications[8–11]. Recent efforts have focused on generalizations of randomized benchmarking[12] to recover partial information about the strength and locality of correlated errors in multi-qubit devices[13–15]. However, these approaches capture only averaged features (their so-called Pauli projections[14]) of the noise channel, leaving aside, e.g., all non-unital channels, including the important amplitude damping noise. Other approaches exist for the validation of average fidelities of a prepared quantum state through a reduced set of measurements[16–20], but only provide limited information about the nature of the noise in the preparation circuit. None of the above methods tackles the characterization of arbitrary quantum processes.

Quantum process tomography (QPT)[21,22], a procedure that reconstructs an unknown quantum process from measurement data, is a fundamental tool for diagnostic and full characterization of quantum

[1]AWS Center for Quantum Computing, Pasadena, CA, USA. [2]Center for Computational Quantum Physics, Flatiron Institute, New York, NY 10010, USA. [3]IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA. [4]Physics and Astronomy Department, Rutgers University, Piscataway, NJ 08854, USA. [5]Institute of Physics, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland. [6]Vector Institute, MaRS Centre, Toronto, Ontario M5G 1M1, Canada. [7]Quantum Research Centre, Technology Innovation Institute, Abu Dhabi, UAE. [8]Instituto de Física, Federal University of Rio de Janeiro, 21941-972, P. O. Box 68528, Rio de Janeiro, Brazil. ✉e-mail: gttorlai@amazon.com

gates and circuits. A direct approach to QPT relies on a informationally-complete (IC) set of measurement settings, which inevitably leads to an algorithmic complexity—in terms of number of measurements and classical post-processing—that scales exponentially with the number of qubits. Due to these limitations, QPT has only been experimentally implemented on up to 3 qubits[23-30].

In most practical scenarios, however, a process to be characterized in a quantum computer typically contains structures that may facilitate its reconstruction. The origin of these structures can be traced back to, e.g., the limited availability and degree of locality of the Hamiltonians used to implement the unitary set of operations in a quantum computer, as well as the nature and strength of the inherent noise of the device, which is often local and exhibits weak correlations among the different qubits. While manipulating and reconstructing a fully generic process requires exponential classical resources[28], these observations suggest that it may be possible to accurately describe relevant quantum channels implemented in real devices by means of classical resources with only polynomial overhead. In fact, similar insights have been leveraged successfully in quantum state tomography, the data-driven reconstruction of a quantum state. Notable examples include matrix product state (MPS) tomography[31-33], exploiting low-entanglement representations of quantum states, and compressed sensing[28,34], relying on the assumption of sparsity of the measurement data.

More recently, an alternative theoretical framework for quantum state tomography based on machine learning has been put forward[35-37], and implemented in a cold-atom experiment[38]. This approach leverages the effectiveness of unsupervised machine learning in extracting high-dimensional probability distributions from raw data[39], combined with the high expressivity of neural networks for capturing highly-entangled quantum many-body states[40-43]. In contrast, approximate algorithms for QPT applicable to near-term quantum devices are currently lacking. While progress has been made in the context of learning non-Markovian dynamics[44,45], the question of a scalable method capable of reconstructing noisy quantum circuits remains wide open.

In this work, we present a technique to perform QPT of quantum circuits of sizes well beyond state-of-the-art. By exploiting the structure of the problem, our approach alleviates important scaling issues of standard QPT. We combine elements of two state-of-the-art classes of algorithms, namely a tensor-network representation of a quantum channel and a data-driven global optimization inspired by

unsupervised learning algorithms. The latter is in stark contrast with previous approaches where the optimisation is driven from local reconstructions on system sub-blocks[30-33]; and is key for the scalability of our method. We show numerical experiments on synthetic data for the computationally challenging case of random unitary circuits, reaching reconstruction fidelities above 0.99 for 2D 10-qubit depth-4 instances using less than $10^5$ single-shot measurements out of a tomographycally complete set of ~$10^{12}$ settings. We also demonstrate the reconstruction of a single 5-qubit parity-check measurement in the surface code undergoing amplitude damping noise. Our proposed method paves the way to the robust and scalable verification of quantum circuits implemented in current experimental hardware.

## Results
### Quantum process tomography
The unavoidable interaction of a quantum device with its environment typically introduces non-unitary dynamics in the underlying quantum state. The time evolution of the corresponding density operator $\rho$ is generated by a *quantum channel*, a linear map $\mathcal{E}: \rho \longrightarrow \mathcal{E}(\rho)$ that is *completely-positive* (CP) ($\mathcal{E} \otimes \mathbb{1} \geq 0$) and *trace-preserving* (TP) ($\mathrm{tr}(\mathcal{E}(\rho)) = \mathrm{tr}(\rho)$)[46]. There are several equivalent mathematical representations of a CPTP map (see ref. [47] for summary). One example is the *Kraus* representation, where the channel is expressed as a set of Kraus operators $\{K_i\}$, leading to the dynamics $\mathcal{E}(\rho) = \sum_i K_i \rho K_i^\dagger$ (Fig. 1a). From the CPTP nature of the channel, it follows $\sum_i K_i^\dagger K_i = \mathbb{1}$.

In the context of quantum tomography, it is natural to instead consider the *Choi matrix* of the channel[48,49], a positive semidefinite operator

$$\Lambda_\mathcal{E} = (\mathbb{1} \otimes \mathcal{E}) \left[ \bigotimes_{j=1}^{N} \left| \Phi_j^+ \right\rangle \left\langle \Phi_j^+ \right| \right], \tag{1}$$

where $\mathcal{E}$ is applied to one half of the tensor product of $N$ unnormalized Bell pairs $|\Phi_j^+\rangle = \sum_{\sigma_j = \tau_j} |\sigma_j \tau_j\rangle$, and $|\sigma_j\rangle$ and $|\tau_j\rangle$ are the input and the output degrees of freedom to the channel (Fig. 1b). The channel $\mathcal{E}$ is CP if and only if the Choi matrix is positive-semidefinite ($\Lambda_\mathcal{E} \geq 0$). It follows that $\Lambda_\mathcal{E}$ is isomorphic to an unnormalized density operator over an extended (bipartite) 2$N$-qubit Hilbert space ($\mathrm{Tr}\,\Lambda_\mathcal{E} = d^N$, with $d$ the dimension of the local Hilbert space, i.e., $d = 2$ for qubits). The TP condition of the channel $\mathcal{E}$ requires that the partial trace of the Choi matrix over the output indices should yield the identity over the input indices: $\mathrm{Tr}_\tau \Lambda_\mathcal{E} = \mathbb{1}_\sigma$[47]. The evolution of a generic quantum state $\rho$ under the channel $\mathcal{E}$ is obtained through the Choi matrix as[47] (Fig. 1c)

$$\mathcal{E}(\rho) = \mathrm{Tr}_\sigma \left[ (\rho^T \otimes \mathbb{1}_\tau) \Lambda_\mathcal{E} \right], \tag{2}$$

where $\rho^T$ denotes matrix transposition.

Unlike the Kraus representation, the Choi matrix of a quantum channel is unique. This implies that QPT simply accounts of fitting the matrix elements of $\Lambda_\mathcal{E}$ to the data, which consists of a special set of prepared input states to the channel and a set of measurements on the output states. In particular, a set of input states and measurements is called *informationally-complete* (IC) if the inputs $\{\rho_\alpha\}$ and the measurement operators $\{M_\beta\}$ span in full the input and the output Hilbert spaces of the quantum channel, respectively. In this case, the probability distribution

$$\begin{aligned} P_\mathcal{E}(\beta|\alpha) &= \mathrm{Tr}_\tau \left[ M_\beta \mathcal{E}(\rho_\alpha) \right] \\ &= \mathrm{Tr}_{\tau,\sigma} \left[ (\rho_\alpha^T \otimes M_\beta) \Lambda_\mathcal{E} \right] \end{aligned} \tag{3}$$

that a measurement on the output state $\mathcal{E}(\rho_\alpha)$ of the channel applied to the input state $\rho_\alpha$ yields outcome $M_\beta$ contains complete information on the channel. That is, $P_\mathcal{E}(\beta|\alpha)$ uniquely characterizes the channel,
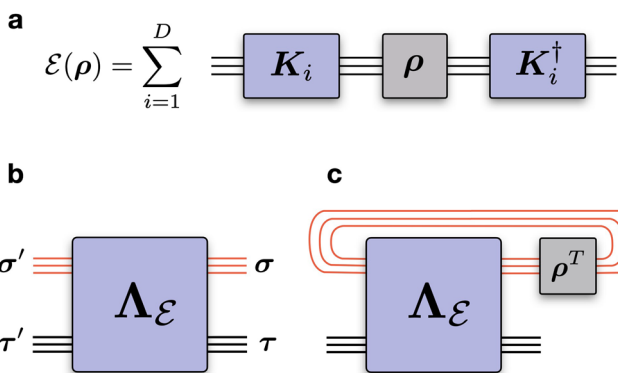


**Fig. 1 | Representations of a quantum channel (with $N = 3$ qubits). a** Evolution of a density operator $\rho$ under a quantum channel $\mathcal{E}$ in the Kraus representation, where the channel has decomposition over $D$ Kraus operators. **b** Representation of the channel with the Choi matrix $\Lambda_\mathcal{E}$, a rank-4$N$ tensor with the upper and lower 2$N$ indices corresponding, respectively, to the input $\{|\sigma\rangle\}$ and output $\{|\tau\rangle\}$ Hilbert spaces. **c** Evolution of $\rho$ using Choi representation. The output state of the channel $\mathcal{E}(\rho)$ is obtained by first contracting the input space $\{|\sigma\rangle\}$ with the transpose state $\rho^T$, followed by a trace, resulting into $\mathcal{E}(\rho)$ over the output space $\{|\tau\rangle\}$.
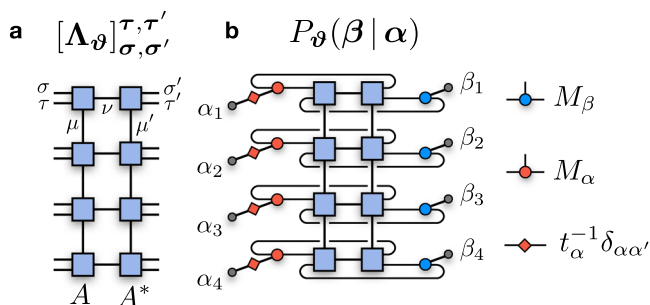
**Fig. 2 | Quantum process tomography with tensor networks. a** The quantum process ($N = 4$) is represented by a Choi matrix $\mathbf{\Lambda}_\vartheta$, parametrized by a locally-purified density operator (LPDO). The input and output indices of the process are $\{\sigma_j\}$ and $\{\tau_j\}$, respectively. **b** Tensor contraction evaluating the conditional probability distribution $P_\vartheta(\boldsymbol{\beta}|\boldsymbol{\alpha})$, i.e., the probability that the LPDO Choi matrix associates with the measurement $\boldsymbol{M}_\beta$ given the state $\boldsymbol{\rho}_\alpha = t_\alpha^{-1} \boldsymbol{M}_\alpha$ at the input of the channel.

and can be used to reconstruct the corresponding (unknown) Choi matrix $\boldsymbol{\Lambda}_\mathcal{E}$.

The standard approach to perform QPT consists of parametrizing the Choi matrix (i.e., using a $4^N \times 4^N$ matrix) and extracting its elements by solving the maximum likelihood estimation problem[22]. There are two fundamental limitations of this approach. First, it requires the parametrization of the full Choi matrix, which scales exponentially with the number of qubits. Second, in order to achieve a high-fidelity fit, the full IC set of input states and measurements is required, which also scales exponentially with $N$. For these reasons, full QPT has remained restricted to very small system sizes.

## Tensor-network Choi matrix

In order to mitigate the exponential complexity of full QPT, we first introduce an efficient representation of Choi matrices in terms of tensor networks, whose total number of parameters is small compared to the dimension of the process Hilbert space. Specifically, we consider a parametrization of the Choi matrix $\boldsymbol{\Lambda}_\vartheta$ (with $\vartheta$ the set of variational parameters) in terms of locally-purified density operator (LPDO), a class of matrix product operators that are non-negative by construction[50] (Fig. 2a). Given a basis for the input $\{|\boldsymbol{\sigma}\rangle\}$ and the output $\{|\boldsymbol{\tau}\rangle\}$ Hilbert spaces, the matrix elements $\langle \boldsymbol{\sigma}, \boldsymbol{\tau} | \boldsymbol{\Lambda}_\vartheta | \boldsymbol{\sigma}', \boldsymbol{\tau}' \rangle$ of the LPDO Choi matrix are given by

$$[\boldsymbol{\Lambda}_\vartheta]_{\boldsymbol{\sigma},\boldsymbol{\sigma}'}^{\boldsymbol{\tau},\boldsymbol{\tau}'} = \sum_{\{\boldsymbol{\mu},\boldsymbol{\mu}'\}} \sum_{\{\boldsymbol{\nu}\}} \prod_{j=1}^N [A_j]_{\mu_{j-1},\nu_j,\mu_j}^{\tau_j,\sigma_j} [A_j^*]_{\mu_{j-1}',\nu_j,\mu_j'}^{\tau_j',\sigma_j'}, \quad (4)$$

where $\vartheta = \{A_j\}$. Here, we assume that $\{A_j\}$ already incorporate the proper normalization $\mathrm{Tr}_{\boldsymbol{\sigma},\boldsymbol{\tau}} \boldsymbol{\Lambda}_\vartheta = d^N$. Each tensor $A_j$ has input index $\sigma_j$, output index $\tau_j$, bond indices ($\mu_{j-1}, \mu_j$) and Kraus index $\nu_j$. The bond and Kraus dimensions of the LPDO are defined as $\chi_\mu = \max_j\{\chi_{\mu_j} = \dim[\mu_j]\}$ and $\chi_\nu = \max_j\{\chi_{\nu_j} = \dim[\nu_j]\}$. By setting $\chi_\nu = 1$, the resulting Choi matrix is rank-1, $\boldsymbol{\Lambda}_\vartheta = |\boldsymbol{\Psi}_\vartheta\rangle\langle\boldsymbol{\Psi}_\vartheta|$, where $|\boldsymbol{\Psi}_\vartheta\rangle$ is an MPS with physical dimension $d^2$ and bond dimension $\chi_\mu$. This corresponds to $\mathcal{E}$ being a unitary channel.

## QPT via unsupervised learning

To perform process tomography with LPDOs, we consider the standard QPT setup of positive operator valued measures (POVM) $\boldsymbol{M}_\beta = \bigotimes_{j=1}^N M_{\beta_j}$, where $\{M_{\beta_j}\}_{\beta_j=1}^{K_m}$ are single-qubit POVMs with $K_m$ measurement outcomes ($M_{\beta_j} \geq 0$ and $\sum_{\beta_j} M_{\beta_j} = \mathbb{1}_j$). As input states to the channel, we take product states $\boldsymbol{\rho}_\alpha = \bigotimes_{j=1}^N \rho_{\alpha_j}$. The preparation states and output measurements are identified by the classical strings $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_N)$, respectively. In the following, we use

for convenience the same POVM set for the input states $\boldsymbol{\rho}_\alpha = t_\alpha^{-1} \boldsymbol{M}_\alpha$ (i.e., $K_m = K_p \equiv K$), where $t_\alpha = \mathrm{Tr}\, \boldsymbol{M}_\alpha = \prod_j \mathrm{Tr}\, M_{\alpha_j}$ is a normalization factor.

We generate a training data set by first preparing a finite set of $M$ input states $\{\boldsymbol{\rho}_\alpha^{(k)}\}_{k=1}^M$, randomly sampled according to a fixed prior distribution $Q(\boldsymbol{\alpha})$. We then apply the channel to each state, and perform a measurement at its output, recording the outcomes $\{\boldsymbol{M}_\beta^{(k)}\}_{k=1}^M$. The resulting data set is specified by $M$ strings of $2NK$-valued integers, $\mathcal{D} = \{(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)})\}_{k=1}^M$, with joint probability distribution $P_\mathcal{D}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = Q(\boldsymbol{\alpha}) P_\mathcal{E}(\boldsymbol{\beta}|\boldsymbol{\alpha})$. Similarly, we can estimate the corresponding probability distribution $P_\vartheta(\boldsymbol{\beta}|\boldsymbol{\alpha})$ for the Choi matrix $\boldsymbol{\Lambda}_\vartheta$. Since both input states and output POVMs factorize over the extended Hilbert space, estimating the probability translates into local contractions of the tensors $A_j$ with the tensor product $\rho_{\alpha_j}^T \otimes M_{\beta_j}$ at all sites $j$ (Fig. 2b). The cost of this operation is $\mathcal{O}(d^2 N \chi_\nu \chi_\mu^3)$, remaining efficient as long as the bond dimensions ($\chi_\mu, \chi_\nu$) are sufficiently small.

The learning procedure, inspired by generative modeling of neural networks in machine learning applications[39], consists of varying the parameters $\vartheta$ to minimize the distance between the LPDO distribution $P_\vartheta(\boldsymbol{\beta}|\boldsymbol{\alpha})$ and the target distribution $P_\mathcal{E}(\boldsymbol{\beta}|\boldsymbol{\alpha})$, averaged over the input prior $Q(\boldsymbol{\alpha})$. As distance, we adopt the Kullbach-Leibler divergence[51]:

$$D_{KL} = \sum_{\{\boldsymbol{\alpha}\}} Q(\boldsymbol{\alpha}) \sum_{\{\boldsymbol{\beta}\}} P_\mathcal{E}(\boldsymbol{\beta}|\boldsymbol{\alpha}) \log \frac{P_\mathcal{E}(\boldsymbol{\beta}|\boldsymbol{\alpha})}{P_\vartheta(\boldsymbol{\beta}|\boldsymbol{\alpha})}, \quad (5)$$

Minimizing this quantity is equivalent to minimizing the negative log-likelihood

$$\mathcal{C}(\vartheta) = -\frac{1}{M} \sum_{k=1}^M \log P_\vartheta(\boldsymbol{\beta}_k|\boldsymbol{\alpha}_k), \quad (6)$$

where the average is taken over the data set $\mathcal{D}$. This is the cost function of our optimization problem. This type of tensor network optimization, also explored for quantum state tomography[52], is in contrast with the local optimization used in the original formulation of MPS tomography, which relies on measurements of local subsystems and entails exponential scaling with the size of the subsystems[30,33].

The LPDO parameters are iteratively updated using gradient descent $\vartheta \to \vartheta - \eta \nabla_\vartheta \mathcal{C}(\vartheta)$ (or a variation thereof), where $\eta$ is the size of the gradient update (i.e., the *learning rate*). In our simulations, we optimize the LPDO using automatic differentiation software[53], a framework that is being increasingly explored in tensor networks applications[54,55]. However, the gradients of the cost function can also be derived analytically[56,57], and are shown in the Supplementary Material. We also point out that, due to the tensor-network parametrization of the Choi matrix, the optimization landscape is non-convex, which means that there is no a priori guarantee that the training will yield the exact target Choi matrix in the limit of infinite data.

In defining our parametrized model $\boldsymbol{\Lambda}_\vartheta$, we exploited the fact that Choi matrices are isomorphic to density operators, which justifies the use of LPDOs. However, while $\boldsymbol{\Lambda}_\vartheta = \boldsymbol{\Lambda}_\vartheta^\dagger$ and $\boldsymbol{\Lambda}_\vartheta \geq 0$ by construction, the LPDO is not inherently TP. That is, the condition $\mathrm{Tr}_\tau \boldsymbol{\Lambda}_\vartheta = \mathbb{1}_\sigma$ is not enforced at the level of the elementary tensors $\{A_j\}$. We expect that, if $M$ is large enough and the model faithfully learns the quantum channel underlying the training data set, this property should also be approximately satisfied. Nonetheless, we can approximately impose the TP constraint by adding a *regularization* term to $\mathcal{C}(\vartheta)$, which induces a bias towards trace-preserving matrices. We define this regularization term as

$$\Gamma_\vartheta = \sqrt{d^{-N}}\, \|\boldsymbol{\Delta}_\vartheta\|_F = \sqrt{d^{-N}} \sqrt{\mathrm{Tr}_\sigma\left(\boldsymbol{\Delta}_\vartheta \boldsymbol{\Delta}_\vartheta^\dagger\right)}, \quad (7)$$

where $\mathbf{\Delta}_\vartheta = \mathrm{Tr}_\tau \mathbf{\Lambda}_\vartheta - \mathbb{1}_\sigma$. The final cost function becomes $\mathcal{C}(\vartheta) = -\langle \log P_\vartheta(\boldsymbol{\beta}|\boldsymbol{\alpha})\rangle_\mathcal{D} + \kappa\,\Gamma_\vartheta$, where $\kappa$ is a hyper-parameter of the optimization.

## Numerical experiments

We study the performance of LPDO-based QPT for unitary and noisy quantum channels. We adopt, for both the input states and measurements, the POVM set built out of the rank-1 projectors of the $K=6$ eigenstates of the Pauli matrices. This POVM is informationally over-complete and experimentally-friendly, as it can be implemented with routinely available single-qubit measurements. For all the instances described, we generate the training data set $\mathcal{D}$ using a uniform prior distribution $Q(\boldsymbol{\alpha}) = K^{-N}$. We split the data set into a training set and a validation set, containing, respectively, 80% and 20% of the total data. The training data set contains the measurements used to compute the gradients and train the LPDO. The remaining held-out data is used for cross-validation for selecting the optimal model. That is, the cost function computed on the validation data set is used to verify that the model is not overfitting the training data set, and to choose the optimal training epoch (see "Methods"). Details on the data generation and the LPDO trainings are provided in the Supplementary Material.

We start by studying the case of a unitary channel characterized by a rank-1 Choi matrix $\mathbf{\Lambda}_\mathcal{E} = |\boldsymbol{\Psi}_\mathcal{E}\rangle\langle\boldsymbol{\Psi}_\mathcal{E}|$. We perform QPT by setting the Kraus dimension to $\chi_\nu = 1$, leading to the parametrized Choi matrix $\mathbf{\Lambda}_\vartheta = |\boldsymbol{\Psi}_\vartheta\rangle\langle\boldsymbol{\Psi}_\vartheta|$ expressed in terms of an MPS $\boldsymbol{\Psi}_\vartheta$. We also set the bond dimension of the LPDO $\chi_\mu$ equal to the bond dimension $\chi_\mathcal{E}$ of $\boldsymbol{\Psi}_\mathcal{E}$. Thus, there is no approximation in the representation of the channel, and any reconstruction error generates solely from the finite size of the data set and any potential inefficiency of the optimization procedure. We point out that, when the ideal target quantum circuit is known, it is possible to estimate what is the minimum value of $\chi_\mathcal{E}$ leading to a faithful tensor-network representation of the quantum circuit. Both conditions on $\chi_\mu$ and $\chi_\nu$ will be lifted for the reconstruction of a noisy channel, later in this section.

During the training, we measure the cost function computed on both the training and validation data sets. The former monitors the learning progress, while the latter monitors the overfitting and is used to select the optimal parameters (as those in the training epoch that minimize the validation cost function). In addition, we also measure the *quantum process fidelity* $\mathcal{F}(\mathbf{\Lambda}_\mathcal{E} \mathbf{\Lambda}_\vartheta)$ of the reconstruction to the true channel used to generate the data, which measures the average-case performance of the reconstruction. The process fidelity is equivalent to the quantum state fidelity between the two (properly normalized) Choi matrices

$$\mathcal{F}(\mathbf{\Lambda}_\mathcal{E}, \mathbf{\Lambda}_\vartheta) = d^{-2N}\left(\mathrm{Tr}\sqrt{\sqrt{\mathbf{\Lambda}_\mathcal{E}}\mathbf{\Lambda}_\vartheta\sqrt{\mathbf{\Lambda}_\mathcal{E}}}\right)^2, \qquad (8)$$

which reduces to $\mathcal{F}(\mathbf{\Lambda}_\vartheta, \mathbf{\Lambda}_\mathcal{E}) = d^{-2N}\langle\boldsymbol{\Psi}_\mathcal{E}|\mathbf{\Lambda}_\vartheta|\boldsymbol{\Psi}_\mathcal{E}\rangle$ when the target Choi matrix $\mathbf{\Lambda}_\mathcal{E}$ is rank-1. In addition, in the latter rank-1 case, the process fidelity also directly gives other process-closeness quantifiers, such as the Frobenius-norm distance between the Choi states in question[58]. We note that—being an average-case metric—the process fidelity can lead to reconstruction error estimates that may differ from worst-case estimates (as quantified, for instance, by the diamond-norm distance) by orders of magnitude (see e.g., ref. [59] and refs therein). Nevertheless, the fidelity is significantly easier to estimate than other more stringent metrics, which makes it one of the most practical and commonly used metric for experimental state or process reconstruction.

In the numerical experiments, we show the fit fidelity between the target and the learned Choi states, since it provides a direct measure of the quality of the tomographic reconstruction. However, QPT is typically used to extract valuable information about a device, rather than returning a single figure of merit (which can be obtained with more efficient methods[16]). One important example is whether the Choi state
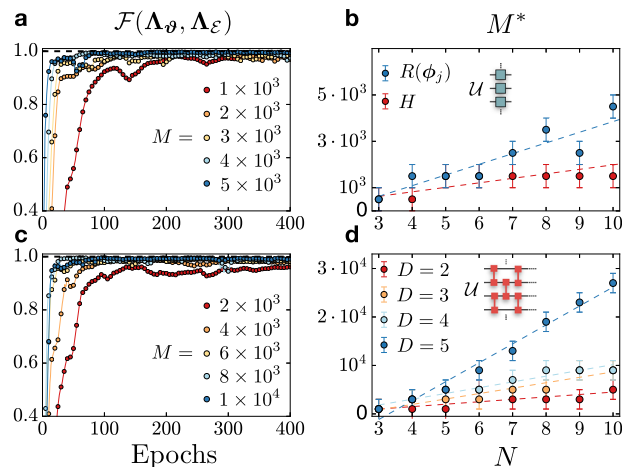


**Fig. 3 | Process reconstruction for unitary quantum circuits containing single-qubit and two-qubit quantum gates. a** Reconstruction fidelity during the LPDO training for a circuits with $N=4$ qubits containing a single layer of Hadamard gates. Different curves corresponds to an increasing size $M$ of the data set. **b** Scaling of the minimum number of samples $M^*$ as a function of $N$ to reach a reconstruction infidelity of $\varepsilon = 0.025$ (i.e., the sample complexity) for a circuit with Hadamard gates (red) and a circuit with random single-qubit rotations $R(\boldsymbol{\phi}_j)$ (blue). **c** Reconstruction fidelity for a circuit with $N=4$ qubits containing 4 layers of controlled-not (CX) gates, for various data set sizes $M$. **d** Sample complexity for quantum circuits with different depths $D$ containing layers of CX gates. For the sample complexity plots, the value $M^*$ is obtained by sequentially increasing $M$ until the threshold in accuracy is met. Error bars are given by the step-size in $M$, and dashed lines are linear fits.

factorizes over a specific partition of the device. Within our framework, this can be easily checked in a scalable manner by simply tracing out local tensors in the Choi LPDO (if averaging over both input and output states).

The first test-case is a unitary quantum circuit containing a single layer of Hadamard gates acting on all qubits. We train LPDOs for different sizes $M$ of the training data set, and we show in Fig. 3a the corresponding reconstruction fidelities measured at each training iteration (*epoch*), for $N=4$ qubits. From this data, we can compute the minimum number of training samples $M^*$ required to reach a fixed accuracy $\varepsilon$ in the reconstruction infidelity $1 - \mathcal{F}(\mathbf{\Lambda}_\vartheta, \mathbf{\Lambda}_\mathcal{E})$. By repeating the same experiment for several systems sizes up to $N=10$ (with $\varepsilon = 0.025$), we show the sample complexity in Fig. 3b—the value $M^*$ as a function of $N$—observing a favorable scaling consistent with a linear behavior. We repeat the same experiment for a single layer of random single-qubit rotations $R(\boldsymbol{\phi}_j)$, observing a similar scaling with a steeper slope.

We also consider a quantum circuit containing $D$ layers of controlled-NOT (CX) gates applied between neighboring qubits in a one-dimensional geometry. Each layer is applied in a staggered manner (inset of Fig. 3d). We perform the same analysis as for the one-qubit circuits, and plot the fidelity curves for various $M$ for a circuit with $N=4$ qubits and depth $D=4$ (Fig. 3c). The sample complexity, computed in an analogous manner, is shown in Fig. 3d for different depths $D$. As expected, the threshold $M^*$ increases with the depth of the circuit.

We now move to the more challenging case of 10-qubit random quantum circuits with depth $D$, for both one- and two-dimensional qubit arrays. Each layer in the circuit consists of $N$ random single-qubit rotations followed by a layer of CX gates. For the one-dimensional circuit, the CX gates alternate between even and odd layers (Fig. 4a). For the two-dimensional circuit, the CX gates are applied in a sequence according to the colors shows in Fig. 4b. In the plots of Fig. 4c, d we show the process infidelity during the training for depth-4 circuits and different values of the data set size $M$. We observe that, with enough number of single-shot samples $M$, the reconstructions surpass a fidelity of $\mathcal{F} = 0.99$.
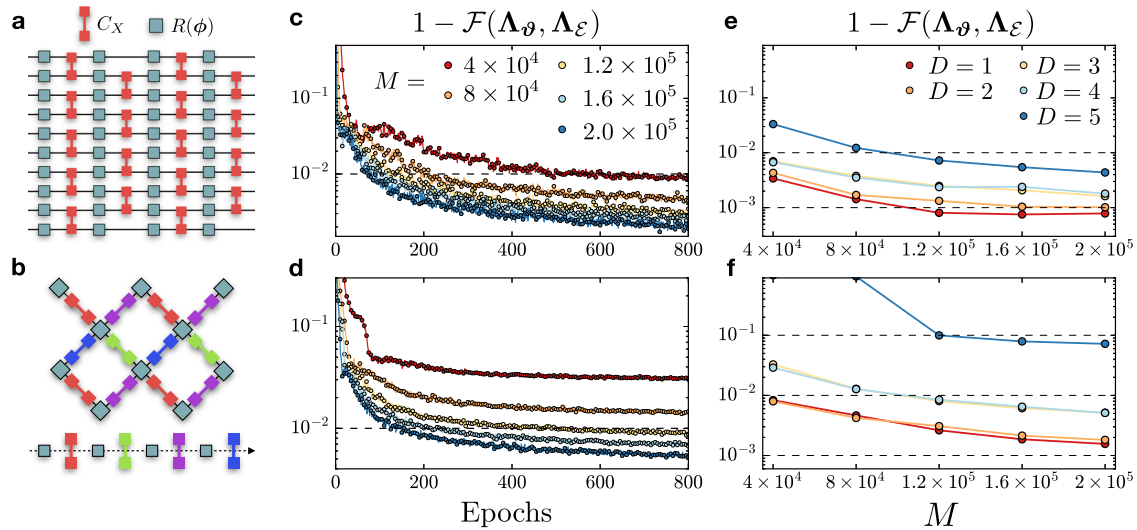
**Fig. 4 | Random quantum circuits. a** One-dimensional quantum circuit with $N = 10$ qubits and $D = 4$ layers, each one consisting of random single-qubit rotations and CX gates, the latter applied in a staggered pattern between even and odd layers. **b** Two-dimensional random quantum circuit, where each layer applies random single-qubit rotations and CX gates according to the colored sequence shown at the bottom of the image. In the panels (**c**) and (**d**), we show the reconstruction infidelity at each epoch, respectively, for a one- and two-dimensional quantum circuit with depth $D = 4$, for various data set sizes $M$. Subplots (**e**) and (**f**) show the lowest infidelities, obtained via cross-validation on held-out data, as a function of the data set size $M$ for different depths.
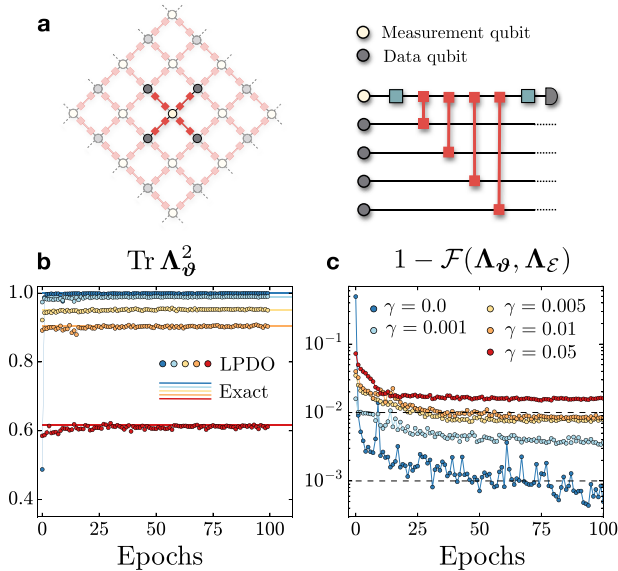


**Fig. 5 | Noisy stabilizer in the surface code. a** A $X$-stabilizer plaquette embedded into the surface code (left) and the quantum circuit performing the parity-check measurement (right), containing Hadamard and CX gates. **b** Purity of the LPDO Choi matrix during training (markers), compared to the exact Choi matrix (solid lines). **c** Infidelity measurement during training for a data set size of $M = 5 \times 10^5$ single-shot measurement outcomes.

We evaluate the optimal LPDO parameters using cross-validation on the held-out data, a metric that does not rely on any prior information about the process and available in an experimental setting. We show in Fig. 4e, f the corresponding lowest infidelities obtained during the training as a function of $M$. As in the previous case, the number of samples to reach a given accuracy increases with the depth of the circuit. For the one-dimensional circuit, the fidelity reaches $\mathcal{F} > 0.99$ with $4 \times 10^4$ measurements up to $D = 4$, and converges to $\mathcal{F} \approx 0.999$ and $\mathcal{F} \approx 0.998$ for $D = 2$ and $D = 4$, respectively. For the two-dimensional circuit, the fidelity converges to $\mathcal{F} > 0.99$ up to $D = 4$ at

$M = 2 \times 10^5$, while $\mathcal{F} \approx 0.93$ for $D = 5$. In this case, the bond dimension of the target circuit is $\chi_\mu = 32$, a four-fold increase from $\chi_\mu = 8$ of the $D = 4$ circuit. We emphasize that the size $M$ of the data set used is very small in comparison with any IC set of input states and measurement settings. For instance, for our choice of POVMs, the total number of experimental configurations for a 10-qubit circuit is $6^N 3^N \sim 10^{12}$.

Finally, we turn to the case of a quantum circuit undergoing a noise channel. As a test case, we study a single $X$-stabilizer measurement of the surface code, a paradigmatic model of topological quantum computation[60,61]. The circuit contains a total of $N = 5$ qubits, where a single measurement qubit is used to stabilize the $X$ parity-check between four data qubits. The quantum circuit for the stabilizer measurement consists of a Hadamard gate on the measurement qubit, four CX gates between the measurement qubit and each data qubit, followed by an additional Hadamard gate on the measurement qubit (Fig. 5a). We apply a single-qubit amplitude damping channel to each qubit involved in a quantum gate after its application, with a fixed decay probability $\gamma \in [0, ..., 0.05]$.

We perform the reconstruction by varying both the bond dimension and the Kraus dimension, until convergence is found, and we show the results for $\chi_\mu = \chi_\nu = 6$. During the training, we measure the reconstruction fidelity, as well as the purity $\mathrm{Tr} \Lambda_\vartheta^2$ of the LPDO. For all values of the decay probability $\gamma$, we observe that the purity converges to the correct value (solid lines) computed from the exact Choi matrix (Fig. 5b), suggesting that the Kraus dimension of the LPDOs is sufficient to capture the target noisy channel. We also show the process infidelity curves obtained using a total of $M = 5 \times 10^5$ measurement samples, for different values of $\gamma$ (Fig. 5c). While for the noiseless channel the fidelity reaches $\mathcal{F} > 0.999$, the learning appears to become increasingly harder for larger values of $\gamma$. The lowest fidelity $\mathcal{F} \approx 0.985$ is found at $\gamma = 0.05$, which is a fairly large decay probability for current experiments. For lower levels of noise, the reconstruction reaches remarkably high fidelities $\mathcal{F} > 0.99$.

## Discussion

We introduced a procedure for quantum process tomography that integrates a tensor network representation of the Choi matrix in terms of a locally-purified matrix product operator[50], and an optimization strategy based on by machine learning algorithms for generative
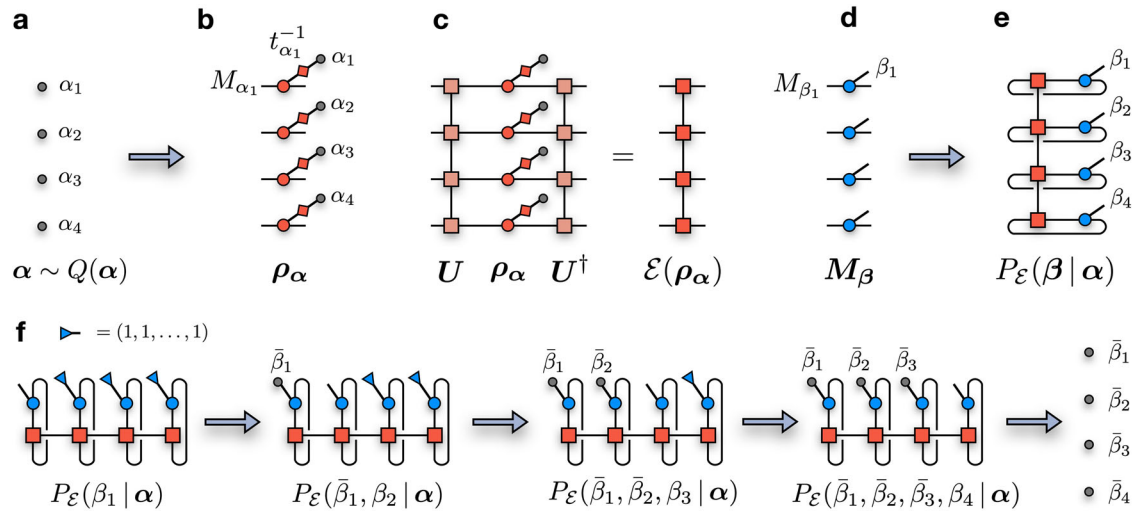
**Fig. 6 | Generation of one training data sample. a** First, we sample a random input POVM state $\alpha = (\alpha_1, \alpha_2, ..., \alpha_N)$ from a reference prior distribution $Q(\alpha)$. **b** The string $\alpha$ specifies an input product state $\rho_\alpha = t_\alpha^{-1} M_\alpha$ to the channel. **c** The output state of the channel is obtained by contracting the input state with the circuit MPO $U$, resulting into a new MPO $\mathcal{E}(\rho_\alpha)$. **d** The measurement POVM $M_\beta$. **e** The process probability distribution $P_\mathcal{E}(\beta|\alpha) = \mathrm{Tr}_\tau[M_\beta \mathcal{E}(\rho_\alpha)]$. **f** Sampling scheme to obtain a single measurement outcome $\beta$ from $P_\mathcal{E}(\beta|\alpha)$. By tracing the indices $\beta_2, ..., \beta_N$ (i.e.,

contracting with a vector [1, ..., 1]), the resulting tensor network with one open index is the probability $P_\mathcal{E}(\beta_1 | \alpha)$, which can be sampled to generate a measurement outcome $\bar{\beta}_1$. By sweeping left to right, this procedure is repeated for each qubits, generating an outcome $\bar{\beta}$ from the correct probability distribution $P_\mathcal{E}(\beta|\alpha)$. The final result of this procedure is one single training sample $(\alpha, \beta)$. The data set are generated by repeating these steps consecutively.

modeling of high-dimensional probability distributions[39]. We demonstrated the power and scalability of the technique using simulated data for unitary random quantum circuits, reaching system sizes of up to 10 qubits and depth 5, and a stabilizer measurement of the surface code undergoing amplitude damping noise. In both cases, the resulting process fidelities reach values close to $\mathcal{F} = 0.99$, using single-shot samples corresponding to a small fraction of the total number of preparation and measurements in the corresponding informationally-complete set, amenable to current experiments.

Due to the entanglement structure induced by a tensor network with small bond dimension, our technique lends itself extremely well to the characterization of quantum hardware operating circuits of sufficiently low depth. The stringent limitation of standard process tomography in the accessible number of qubits is lifted, allowing the reconstruction of large quantum circuits for the case of one-dimensional geometries, as well as two-dimensional thin strips.

Our work demonstrates how infusing state-of-the-art tensor network algorithms with machine learning ideas has the potential to unlock progress in the validation and characterization of currently available quantum devices, and in the design of better error mitigation protocols. This combination makes our techniques relevant for tackling several key obstacles to realizing large-scale quantum computation, including the need for quantum error correction and fault tolerance, which naturally calls for the systematic characterization of effective error terms in large quantum circuits such as the ones studied here.

We anticipate that our strategy will enable progress in the ongoing push for the construction of quantum hardware with lower gate error rates, which will decrease the overhead cost of quantum error correction. This, in turn, will facilitate the faithful execution of more sophisticated quantum algorithms beyond the capabilities of modern classical computers, and help materialize the scientific and technological promises of the nascent second quantum revolution.

## Methods
### Data sets generation
In our numerical experiments, we adopted, both for input states and measurement operators, the set of the rank-1 projector into the

eigenstates of the Pauli matrices:

$$M_0 = p_z |0\rangle \langle 0|, \quad M_1 = p_z |1\rangle \langle 1|, \tag{9}$$

$$M_2 = p_x |+\rangle \langle +|, \quad M_3 = p_x |-\rangle \langle -|, \tag{10}$$

$$M_4 = p_y |+i\rangle \langle +i|, \quad M_5 = p_y |-i\rangle \langle -i| \tag{11}$$

We assume throughout equal probabilities $p_x = p_y = p_z = 1/3$. The full set for the $N$-qubit system is obtained from the tensor product of the operators single-qubit operators

$$M_\alpha = M_{\alpha_1} \otimes M_{\alpha_2} \otimes \cdots \otimes M_{\alpha_N}, \tag{12}$$

and it is specified by a string $\alpha = (\alpha_1, ..., \alpha_N)$, with $\alpha_j = 0, ..., 5$. The input states are simple product states $\rho_\alpha = t_\alpha^{-1} M_\alpha$ with proper normalization $t_\alpha = \mathrm{Tr} M_\alpha = \prod_j \mathrm{Tr} M_{\alpha_j}$. The measurement operators $M_\beta$ are defined analogously, and identified by a string $\beta = (\beta_1, ..., \beta_N)$.

We now provide the step-by-step procedure used to generate the training data for the case of the unitary quantum circuits. Even though the operators we implement are rank-1, we give a description for a more general case of an IC positive operator valued measures (POVM) $M$ beyond the standard projective measurements. For a given circuit architecture, containing a set of single-qubit and two-qubit gates, we first contract each gate together to obtain the MPO corresponding to the full circuit unitary $U$. After each application of a two-qubit gate, we restore the tensor network into an MPO structure by means of singular value decomposition. During this step, we only discard zero singular values, which implies that there is no approximation in the unitary MPO, and that the bond dimension $\chi_U$ generally grows exponentially with the depth of the circuit.

Next, we fix a uniform prior distribution $Q(\alpha) = K^{-N}$ for the input states, where $K$ is the cardinality of the single-qubit POVM (e.g., $K = 6$ for the Pauli projectors). The POVM string $\alpha$ is randomly sampled from

$Q(\boldsymbol{\alpha})$ (Fig. 6a), which defines a specific input state (Fig. 6b)

$$\boldsymbol{\rho_\alpha} = \frac{\boldsymbol{M_\alpha}}{t_{\boldsymbol{\alpha}}} = \frac{M_{\alpha_1}}{t_{\alpha_1}} \otimes \frac{M_{\alpha_2}}{t_{\alpha_2}} \otimes \cdots \otimes \frac{M_{\alpha_N}}{t_{\alpha_N}} \quad (13)$$

For the set of Pauli eigenstates projectors, this translates into applying one layer of single-qubit gates, according to the string $\boldsymbol{\alpha}$. The output state of the channel is then estimated by contracting $\boldsymbol{\rho_\alpha}$ with the circuit MPO $\boldsymbol{U}$, $\mathcal{E}(\boldsymbol{\rho_\alpha}) = \boldsymbol{U}\boldsymbol{\rho_\alpha}\boldsymbol{U}^\dagger$ (Fig. 6c). The output state $\mathcal{E}(\boldsymbol{\rho_\alpha})$ is itself an MPO describing a properly normalized density operator.

Given the output state and the measurement operator $\boldsymbol{M_\beta}$ (Fig. 6d), the process probability $P_\mathcal{E}(\boldsymbol{\beta}|\boldsymbol{\alpha})$ is obtained by contracting (and tracing) these two objects together (Fig. 6e). This probability can then be exactly sampled using the chain rule of probabilities[37,62]. The measurement probability for qubit 1 is computed as

$$p(\beta_1) = \sum_{\beta_2, \beta_3, \ldots, \beta_N} p(\beta_1, \beta_2, \beta_3, \ldots, \beta_N) \quad (14)$$

where we introduced the short-hand notation $p(\boldsymbol{\beta}) = P_\mathcal{E}(\boldsymbol{\beta}|\boldsymbol{\alpha})$. The probability $p(\beta_1)$ is calculated by tracing out each local POVM subspace via a contraction of the tensor network for $P_\mathcal{E}(\boldsymbol{\beta}|\boldsymbol{\alpha})$ with constant vectors $(1_1, 1_2, \ldots, 1_K)$ (blue triangles) at each site $j = 2, \ldots, N$ (Fig. 6f). Once known, the distribution can be sampled to generate measurement outcome $\bar{\beta}_1 \sim P(\beta_1)$. Next, the probability distribution $p(\beta_2|\bar{\beta}_1)$ for the second qubit, conditional on the measurement of the first qubit, is calculated as the ratio between $p(\bar{\beta}_1, \beta_2)$ (shown in the second network of Fig. 6f) and $p(\bar{\beta}_1)$. By repeating this procedure, one obtains a final configuration $\bar{\boldsymbol{\beta}}$ sampled from the correct probability distribution $p(\boldsymbol{\beta}) = P_\mathcal{E}(\boldsymbol{\beta}|\boldsymbol{\alpha})$. Importantly, each $N$-qubit measurement outcome is completely uncorrelated from any other.

For the noisy quantum channels studied in the paper, since there are only $N = 5$ qubits, we perform a direct simulation of the channel to obtain the full Choi matrix. The training data is obtained directly from the Choi matrix, using input states and measurement operators identical to the ones described above.

## Trace-preserving regularization

In general, the LPDO representation does not enforce the TP condition on the corresponding quantum channel, i.e., $\mathrm{Tr}_{\boldsymbol{\tau}} \boldsymbol{\Lambda}_\vartheta \neq \mathbb{1}_{\boldsymbol{\sigma}}$. However, this condition can be easily added to the cost function as a *regularization term*, which biases the optimization to yield a set of optimal parameters $\vartheta$ that minimizes the negative log-likelihood, while also minimizing the *distance* between $\mathrm{Tr}_{\boldsymbol{\tau}} \boldsymbol{\Lambda}_\vartheta$ and $\mathbb{1}_{\boldsymbol{\sigma}}$. As a distance measure, we choose the Frobenius norm of the difference $\boldsymbol{\Delta}_\vartheta = \mathrm{Tr}_{\boldsymbol{\tau}} \boldsymbol{\Lambda}_\vartheta - \mathbb{1}_{\boldsymbol{\sigma}}$:

$$\|\boldsymbol{\Delta}_\vartheta\|_F = \sqrt{\mathrm{Tr}_{\boldsymbol{\sigma}}\left(\boldsymbol{\Delta}_\vartheta \boldsymbol{\Delta}_\vartheta^\dagger\right)}. \quad (15)$$

The tensor network for $\boldsymbol{\Delta}_\vartheta$ can be easily computed by performing an MPO subtraction[63], which in this case it increases the bond dimension of $\boldsymbol{\Lambda}_\vartheta$ by 1 (Fig. 7a). The regularization term is then

$$\Gamma_\vartheta = \sqrt{d^{-N}} \sqrt{\mathrm{Tr}_{\boldsymbol{\sigma}}\left(\boldsymbol{\Delta}_\vartheta \boldsymbol{\Delta}_\vartheta^\dagger\right)}, \quad (16)$$

where we introduced a normalization pre-factor $\sqrt{d^{-N}}$. This leads to the final cost function

$$\mathcal{C}(\vartheta) = \log Z_\vartheta - \langle \log \widetilde{P}_\vartheta(\boldsymbol{\beta}|\boldsymbol{\alpha}) \rangle_\mathcal{D} + \kappa \Gamma_\vartheta, \quad (17)$$

where $\kappa$ is an additional hyper-parameter.

We show the measurement of the regularization term $\Gamma_\vartheta$ (Fig. 7b) at each training iteration for the reconstruction of one-dimensional random quantum circuits of different depths. By comparing these
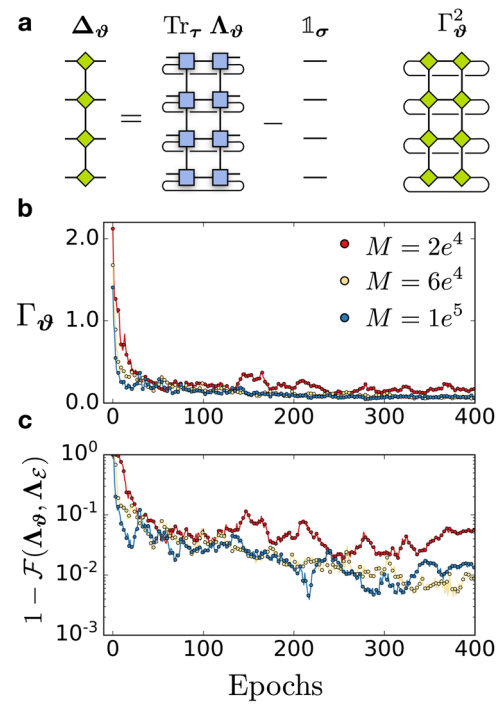


**Fig. 7 | Trace-preserving regularization. a** Tensor network for $\boldsymbol{\Delta}_\vartheta$, obtained by subtracting the identity MPO (with bond dimension 1) to the (properly normalized) LPDO $\boldsymbol{\Lambda}_\vartheta$, and tensor contraction required to compute $\Gamma_\vartheta$. We show the measurement of the regularization $\Gamma_\vartheta$ (**b**) and the reconstruction infidelity (**c**) during the training for a one-dimension random quantum circuit with $N = 10$ qubits and depth $D = 2$, for different number of the total data set size $M$.

curves with the reconstruction infidelities (Fig. 7c), one can clearly see the correlation between the accuracy of the reconstruction and the amount of violation of the TP condition.

## Overfitting and model selection

The goal of training the LPDO using unsupervised learning is to efficiently extract the relevant structure and features characterizing the unknown channel from a limited set of measurements. In other words, the model needs to be able to *generalize* beyond the measurements provided for its training. If the number of samples in the data set $\mathcal{D}$ is too low, it is likely that the LPDO training leads to *overfitting*, i.e., the LPDO learns features present in the data that are not representative of the unknown channel, but only stems from the limited number of training samples.

A strategy to monitor the overfitting, routinely used in the training of deep neural networks, is to divide the data set into two sub-sets: a training data set $\mathcal{D}_T$ and a validation data set $\mathcal{D}_V$. Here, we do so using a 80%/20% split ratio. The training data set $\mathcal{D}_T$ is used for the learning procedure, i.e., the calculation of the gradients used to update the model. During training, we compute the training loss (i.e., the average of the cost function on the training data set)

$$\mathcal{L}_T(\vartheta) = -\frac{1}{|\mathcal{D}_T|} \sum_{(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathcal{D}_T} \log P_\vartheta(\boldsymbol{\beta}|\boldsymbol{\alpha}), \quad (18)$$

which signals whether the model is actively learning (i.e., a decreasing $\mathcal{L}_T(\vartheta)$). At the same time, we also compute the validation loss on the held-out data

$$\mathcal{L}_V(\vartheta) = -\frac{1}{|\mathcal{D}_V|} \sum_{(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathcal{D}_V} \log P_\vartheta(\boldsymbol{\beta}|\boldsymbol{\alpha}). \quad (19)$$

Here, $|\mathcal{D}_T|$ and $|\mathcal{D}_V|$ are the size of the training and validation data sets, respectively.

Generally, in the early stage of the training, the validation loss decreases hand-in-hand with the training loss. However, if the model starts to overfit spurious features in the training data, the validation loss will invert its trend and start increasing, an indication that more training data is needed. We stress that both of these measurements are available in a practical experimental setting, since no information about the channel is being used.

The validation loss $\mathcal{L}_V(\vartheta)$ is also a useful metric to perform the model selection, i.e., to pick a specific set of parameters $\vartheta^{(t)}$ at epoch $t$ to be considered the optimal solution of the optimization problem. In our numerical simulations, we select the optimal parameters as the ones at the training epochs $t$ where the measurement of the validation loss returned its lowest value. This is also a model selection procedure that can be used in an experimental setting.

### Specifics of the numerical experiments

In this final section, we provide details on the numerical experiments presented in the main text. In all cases, the LPDO tensors $\{\widetilde{A}_j\}$ are initialized randomly, with each tensor component set to

$$[\widetilde{A}_j]^{\tau_j,\sigma_j}_{\mu_{j-1},\nu_j,\mu_j} = a_r + ia_i \tag{20}$$

where $a_r$ and $a_i$ are drawn from a uniform distribution centered around zero with width 0.2. We compute the gradients on batches of data containing $M_B = 800$ samples. Once the gradients are collected, we update the LPDO tensors using the Adam optimization with parameters $\eta = 0.005$, $\xi_1 = 0.9$, $\xi_2 = 0.999$, and $\epsilon = 10^{-7}$.

*Figure 2.* The first set of quantum channels investigated are unitary quantum circuits containing one layer of single-qubit gates. We study two types of circuits, containing either Hadamard gates

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \tag{21}$$

or random single-qubit rotations

$$R(\theta,\phi,\lambda) = \begin{pmatrix} \cos\frac{\theta}{2} & -e^{i\lambda}\sin\frac{\theta}{2} \\ e^{i\phi}\sin\frac{\theta}{2} & e^{i(\phi+\lambda)}\cos\frac{\theta}{2} \end{pmatrix}. \tag{22}$$

To obtain the sample complexity curves shows in Fig. 2b, we perform the reconstruction for an increasing number $N$ of qubits. For each $N$, we start using a small data set size $M$, and increase it with a fixed size-step until the threshold $\varepsilon = 0.025$ in infidelity is met. The result is a value $M^*$ with an error bar given by the size-step.

We repeat the same scaling study for quantum circuits containing $D$ layers of controlled-NOT (CX) gates

$$CX = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{23}$$

For a quantum circuit with depth $D$, the odd and even layers apply two-qubit gates with the control qubit having odd and even qubit-index, respectively. Here, the bond dimension of the LPDO Choi matrix is set to the bond dimension of the circuit MPO.

*Figure 3.* We reconstruct random quantum circuits in both one and two dimensions. In both cases, each layer of the quantum circuit consists of one layer with $N$ single-qubit random rotations $R(\theta,\phi,\lambda)$ (defined above) and one layer of CX gates. In the one-dimensional geometry, the CX gates alternates as in the previous case. For the two-dimensional quantum circuit, they are applied according to the color scheme shown in Fig. 3b. For the simulation of the quantum circuit and

the data generation, the circuit MPO has a "snake-shape" as per usual in MPS simulations of two-dimensional geometries. After applying the CX gates, the circuit tensor network is restored into a local form by means of singular value decomposition, where only zero singular values are discarded. This means that the representation of the target quantum circuit is exact.

We first set of the bond dimension of the LPDO Choi matrix equal to the bond dimension of the circuit MPO, and set the Kraus dimension to $\chi_\nu = 1$. All the data shown in Fig. 3 has been collected under this condition. However, additional simulations have also been performed using larger values of the LPDO bond dimension, obtaining comparable results. During the training, we monitor the training loss, the validation loss, the TP regularizer, and the reconstruction fidelity. We use cross-validation on the held-out data set $\mathcal{D}_V$ to select the best models for each circuit configuration and for each data set size $M$. The curves in Fig. 3e, f show the reconstruction infidelities of these selected models.

*Figure 4.* Finally, we reconstruct a noisy quantum channel. We consider the X-stabilizer measurement in the surface code, where the parity-check between four data qubits is measured using an additional (measurement) qubit with the quantum circuit shown in Fig. 4a. The circuit contains two Hadamard gates and four CX gates. We apply an amplitude-damping channel, characterized by the Kraus operators

$$K_0 = |0\rangle\langle0| + \sqrt{1-\gamma}|1\rangle\langle1| \tag{24}$$

$$K_1 = \sqrt{\gamma}|0\rangle\langle1| \tag{25}$$

where $\gamma$ is the decay probability. The channel is applied to each quantum gate in the circuit, where for the two-qubit gates the channel is just the tensor product of the single-qubit channel shown above.

We now relax any prior information on both the quantum circuit and the noise channel. We perform the reconstruction by varying the bond dimension $\chi_\mu$ and the Kraus dimension $\chi_\nu$ of the LPDO. The only setting where convergence in the training metrics is found already for $\chi_\nu = 0$ is the noiseless channel $\gamma = 0$, as expected. Nonetheless, even by increasing $\chi_\nu$, the noiseless channel is still properly reconstructed. This can be seen in Fig. 4b, where the purity of the reconstruction LPDO Choi matrix for $\gamma = 0$ and $\chi_\nu = 6$ reaches the correct value of $\mathrm{Tr}\Lambda_\vartheta^2 \approx 1$. The infidelity curves are shown for a fixed data set size of $M = 5 \times 10^5$.

## Data availability

The data sets generated during and/or analyzed during the current study are available from the corresponding author on request.

## Code availability

The software used in this work has been re-based into the Julia package PastaQ[65].

## References

1. Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018).
2. Smith, J. et al. Many-body localization in a quantum simulator with programmable random disorder. *Nat. Phys.* **12**, 907–911 (2016).
3. Friis, N. et al. Observation of entangled states of a fully controlled 20-qubit system. *Phys. Rev. X* **8**, 021012 (2018).
4. Bernien, H. et al. Probing many-body dynamics on a 51-atom quantum simulator. *Nature* **551**, 579–584 (2017).
5. Trotzky, S. et al. Probing the relaxation towards equilibrium in an isolated strongly correlated one-dimensional Bose gas. *Nat. Phys.* **8**, 325–330 (2012).
6. Arute, F. et al. Quantum supremacy using a programmable super-conducting processor. *Nature* **574**, 505–510 (2019).

7. Mooney, G. J., White, G. A. L., Hill, C. D. & Hollenberg, L. C. L. Whole-Device Entanglement in a 65-Qubit Superconducting Quantum Computer. *Adv. Quantum Technol.*, **4**, 2100061 (2021).

8. Kandala, A. et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242–246 (2017).

9. Kokail, C. et al. Self-verifying variational quantum simulation of lattice models. *Nature* **569**, 355–360 (2019).

10. Havlíček, V. et al. Supervised learning with quantum-enhanced feature spaces. *Nature* **567**, 209–212 (2019).

11. Harrigan, M. P. et al. Quantum approximate optimization of non-planar graph problems on a planar superconducting processor. *Nat. Phys.* **17**, 332–336 (2021).

12. Magesan, E., Gambetta, J. M. & Emerson, J. Characterizing quantum gates via randomized benchmarking. *Phys. Rev. A* **85**, 042311 (2012).

13. Erhard, A. et al. Characterizing large-scale quantum computers via cycle benchmarking. *Nat. Commun.* **10**, 5347 (2019).

14. Harper, R., Flammia, S. T. & Wallman, J. J. Efficient learning of quantum noise. *Nat. Phys.* **16**, 1184–1188 (2020).

15. McKay, D. C., Cross, A. W., Wood, C. J. & Gambetta, J. M. Correlated randomized benchmarking. Preprint at http://arxiv.org/abs/2003.02354 (2020).

16. Flammia, S. T. & Liu, Y.-K. Direct fidelity estimation from few Pauli measurements. *Phys. Rev. Lett.* **106**, 230501 (2011).

17. da Silva, M. P., Landon-Cardinal, O. & Poulin, D. Practical characterization of quantum devices without tomography. *Phys. Rev. Lett.* **107**, 210404 (2011).

18. Aolita, L., Gogolin, C., Kliesch, M. & Eisert, J. Reliable quantum certification of photonic state preparations. *Nat. Commun.* **6**, 8498 (2015).

19. Gluza, M., Kliesch, M., Eisert, J. & Aolita, L. Fidelity witnesses for fermionic quantum simulations. *Phys. Rev. Lett.* **120**, 190501 (2018).

20. Roth, I. et al. Recovering quantum gates from few average gate fidelities. *Phys. Rev. Lett.* **121**, 170502 (2018).

21. Chuang, I. L. & Nielsen, M. A. Prescription for experimental determination of the dynamics of a quantum black box. *J. Mod. Opt.* **44**, 2455–2467 (1997).

22. D'Ariano, G. M. & Lo Presti, P. Quantum tomography for measuring experimentally the matrix elements of an arbitrary quantum operation. *Phys. Rev. Lett.* **86**, 4195–4198 (2001).

23. O'Brien, J. L. et al. Quantum process tomography of a controlled-not gate. *Phys. Rev. Lett.* **93**, 080502 (2004).

24. Riebe, M. et al. Process tomography of ion trap quantum gates. *Phys. Rev. Lett.* **97**, 220407 (2006).

25. Weinstein, Y. S. et al. Quantum process tomography of the quantum Fourier transform. *J. Chem. Phys.* **121**, 6117–6133 (2004).

26. Bialczak, R. C. et al. Quantum process tomography of a universal entangling gate implemented with Josephson phase qubits. *Nat. Phys.* **6**, 409–413 (2010).

27. Chow, J. M. et al. Universal quantum gate set approaching fault-tolerant thresholds with superconducting qubits. *Phys. Rev. Lett.* **109**, 060501 (2012).

28. Shabani, A. et al. Efficient measurement of quantum dynamics via compressive sensing. *Phys. Rev. Lett.* **106**, 100401 (2011).

29. Krinner, S. et al. Benchmarking coherent errors in controlled-phase gates due to spectator qubits. *Phys. Rev. Applied* **14**, 024042 (2020).

30. Govia, L. C. G., Ribeill, G. J., Ristè, D., Ware, M. & Krovi, H. Bootstrapping quantum process tomography via a perturbative ansatz. *Nat. Commun.* **11**, 1084 (2020).

31. Cramer, M. et al. Efficient quantum state tomography. *Nat. Commun.* **1**, 149 (2009).

32. Baumgratz, T., Gross, D., Cramer, M. & Plenio, M. B. Scalable reconstruction of density matrices. *Phys. Rev. Lett.* **111**, 020401 (2013).

33. Lanyon, B. P. et al. Efficient tomography of a quantum many-body system. *Nat. Phys.* **13**, 1158–1162 (2017).

34. Gross, D., Liu, Y.-K., Flammia, S. T., Becker, S. & Eisert, J. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* **105**, 150401 (2010).

35. Torlai, G. & Melko, R. G. Machine-learning quantum states in the NISQ era. *Annu. Rev. Condens. Matter Phys.* **11**, 325–344 (2020).

36. Torlai, G. et al. Neural-network quantum state tomography. *Nat. Phys.* **14**, 447–450 (2018).

37. Carrasquilla, J., Torlai, G., Melko, R. G. & Aolita, L. Reconstructing quantum states with generative models. *Nat. Mach. Intell.* **1**, 155–161 (2019).

38. Torlai, G. et al. Integrating neural networks with a quantum simulator for state reconstruction. *Phys. Rev. Lett.* **123**, 230504 (2019).

39. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016).

40. Carleo, G. & Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **355**, 602–606 (2017).

41. Gao, X. & Duan, L.-M. Efficient representation of quantum many-body states with deep neural networks. *Nat. Commun.* **8**, 662 (2017).

42. Glasser, I., Pancotti, N., August, M., Rodriguez, I. D. & Cirac, J. I. Neural-network quantum states, string-bond states, and chiral topological states. *Phys. Rev. X* **8**, 011006 (2018).

43. Carleo, G., Nomura, Y. & Imada, M. Constructing exact representations of quantum many-body systems with deep neural networks. *Nat. Commun.* **9**, 5322 (2018).

44. Guo, C., Modi, K. & Poletti, D. Tensor network based machine learning of non-markovian quantum processes. *Phys. Rev. A* **102**, 062414 (2020).

45. White, G .A. L., Hill, C. D., Pollock, F. A., Hollenberg, L. C. L. & Modiet K. Demonstration of non-Markovian process characterisation and control on a quantum processor. *Nat. Commun.* **11**, 6301 (2020).

46. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).

47. Wood, C. J., Biamonte, J. D. & Cory, D. G. Tensor networks and graphical calculus for open quantum systems. *Quantum Inf. Computat.* **15**, 0579–0811 (2015).

48. Jamiołkowski, A. Linear transformations which preserve trace and positive semidefiniteness of operators. *Rep. Math. Phys.* **3**, 275–278 (1972).

49. Choi, M. D. Completely positive linear maps on complex matrices. *Linear Algebra Its Appl.* **10**, 285–290 (1975).

50. Werner, A. H. et al. Positive tensor network approach for simulating open quantum many-body systems. *Phys. Rev. Lett.* **116**, 237201 (2016).

51. Kullback, S. & Leibler, R. A. On information and sufficiency. *Ann. Math. Statist.* **22**, 79–86 (1951).

52. Wang, J. et al. Scalable quantum tomography with fidelity estimation. *Phys. Rev. A* **101**, 032321 (2020).

53. Abadi, M. et al. TensorFlow: large-scale machine learning on heterogeneous systems. https://www.tensorflow.org/ (2015).

54. Liao, H.-J., Liu, J.-G., Wang, L. & Xiang, T. Differentiable programming tensor networks. *Phys. Rev. X* **9**, 031041– (2019).

55. Torlai, G., Carrasquilla, J., Fishman, M. T., Melko, R. G. & Fisher, M. P. A. Wavefunction positivization via automatic differentiation. *Phys. Rev. Research* **2**, 032060(R) (2020).

56. Han, Z.-Y., Wang, J., Fan, H., Wang, L. & Zhang, P. Unsupervised generative modeling using matrix product states. *Phys. Rev. X* **8**, 031012– (2018).

57. Glasser, I., Sweke, R., Pancotti, N., Eisert, J. & Cirac, J. I. Expressive power of tensor-network factorizations for probabilistic modeling, with applications from hidden Markov models to quantum machine learning. Preprint at http://arxiv.org/abs/1907.03741 (2019).

58. Coles, P. J., Cerezo, M. & Cincio, L. Strong bound between trace distance and Hilbert-Schmidt distance for low-rank states. *Phys. Rev. A* **100**, 022103 (2019).

59. Wallman, J. J. Bounding experimental quantum error rates relative to fault-tolerant thresholds. Preprint at https://doi.org/10.48550/ARXIV.1511.00727 (2015).

60. Dennis, E., Kitaev, A., Landahl, A. & Preskill, J. Topological quantum memory. *J. Math. Phys.* **43**, 4452–4505 (2002).

61. Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: towards practical large-scale quantum computation. *Phys. Rev. A* **86**, 032324 (2012).

62. Ferris, A. J. & Vidal, G. Perfect sampling with unitary tensor networks. *Phys. Rev. B* **85**, 165146– (2012).

63. Hubig, C., McCulloch, I. P. & Schollwöck, U. Generic construction of efficient matrix product operators. *Phys. Rev. B* **95**, 035129 (2017).

64. Héctor, A. et al. Qiskit: an open-source framework for quantum computing. https://doi.org/10.5281/zenodo.2562110 (2019).

65. Torlai, G. & Fishman, M. PastaQ: a package for simulation, tomography and analysis of quantum computers. https://github.com/GTorlai/PastaQ.jl/ (2020).

## Author contributions

G.T., C.J.W., A.A., G.C., J.C., and L.A. contributed equally to this work.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41467-023-38332-9.

**Correspondence** and requests for materials should be addressed to Giacomo Torlai.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.