Editors' Suggestion

Quantum Autoencoders to Denoise Quantum Data

Dmytro Bondarenko* and Polina Feldmann†

Institut für Theoretische Physik, Leibniz Universität Hannover, Appelstr. 2, DE-30167 Hannover, Germany

(Received 12 November 2019; accepted 24 February 2020; published 31 March 2020)

Entangled states are an important resource for quantum computation, communication, metrology, and the simulation of many-body systems. However, noise limits the experimental preparation of such states. Classical data can be efficiently denoised by autoencoders—neural networks trained in unsupervised manner. We develop a novel quantum autoencoder that successfully denoises Greenberger-Horne-Zeilinger, W, Dicke, and cluster states subject to spin-flip errors and random unitary noise. Various emergent quantum technologies could benefit from the proposed unsupervised quantum neural networks.

DOI: 10.1103/PhysRevLett.124.130502

Introduction.—With the ever-increasing complexity of systems that our society deals with, the *ab initio* understanding of important features remains a distant dream. However, one can distill many useful relations by simply collecting data about such complex systems and studying interdependencies. As our ability to gather, store and process data has rapidly progressed, the computational techniques to extract useful knowledge from data—machine learning (ML)—have become much more powerful. One of the most popular ML techniques are neural networks (NNs), which have found numerous applications, from self-driving cars to drug discovery (see, e.g., [1–3]).

Depending on the data, different learning scenarios for ML algorithms can be distinguished. If nothing is precisely known about the algorithm's desired outputs for the training inputs, the learning is called *unsupervised* or *self-supervised*. Autoencoders (AE) are a prominent example of NNs that learn without supervision, see, e.g., [3]. They have been used, e.g., to denoise bird songs in the wilderness [4].

ML could benefit from the rapid progress of quantum computing hard- and software (see, e.g., [5,6]). Moreover, there are important ML tasks where the data comes as a set of quantum, and possibly—highly entangled, states. Examples include quantum cryptography (see, e.g., [7]), metrology (see, e.g., [8,9]), and chemistry (see, e.g., [10,11]). ML is called quantum if it uses quantum algorithms or quantum data (see, e.g., [12,13]).

Virtually every experimental preparation of a quantum state introduces noise. Usually, it is hard to design a denoising protocol. First, one has to identify and characterize all noise sources. Second, one has to invent a protocol which corrects the noise without affecting any relevant features of the quantum state. ML can automate this task. As there is often no denoised reference state to compare with, unsupervised learning is required.

Various quantum neurons have been proposed in [14–28]. We follow [14], since these NNs are capable of universal quantum computation, the computational complexity per training round scales at most quadratically with the depth of the NN, the cost function has a clear operational meaning, and the authors provide an open-source implementation. The parameters of such a quantum NN (QNN) are classical variables. In general, quantum parameters may be useful [29], but for ML tasks without memory they can give only a marginal improvement [30].

Classically simulable quantum AEs have been studied in [31]. In [32] shallow quantum AEs have been introduced for data compression. The QNNs in [32] are closely related to the neurons from [14]. Contrary to a claim in [32] they are universal. However, the authors of [32] restrict the class of operations to get polynomial complexity scaling with the width of the network. Data compression via AEs has been demonstrated with photons [33]. In [34], it has been proposed to train AEs for quantum data compression using genetic algorithms on a classical computer. The trained AEs have been implemented on superconducting qubits [35]. Classical ML techniques have been used to design experiments that produce entangled states [36,37] or useful entangled states robust against noise [37,38]. The general setting of quantum unsupervised ML has been studied in [39].

In this Letter we construct quantum AEs capable of quantum advantage for the purpose of denoising quantum data. We apply them to single and continuously parametrized sets of small highly entangled states subject to different kinds of noise. We observe excellent denoising without fine tuning of the hyperparameters.

Quantum autoencoders.—ML algorithms numerically solve variational problems. An NN is a variational class of maps $f_v: X \to Y$ parametrized by a vector v. It is constructed from simpler parametrized maps called neurons. The outputs of a set of neurons—a *layer*—are fed into the next layer. If layer n gets all its inputs from layers $k < n$, the network is called *feed forward* (FF). The input $x \in X$ of the NN is the first layer, and the output $f_v(x)$—the last. The number of layers is the *depth* of an NN, and the maximal number of neurons per layer—its *width*. The geometry of the neuronal interconnections—the *topology* of an NN—and the choice of neurons determine the variational class given by the NN. With suitable neurons any map can be represented as a FFNN (FFNNs are *universal*).

Let us assume that a number of correct input-output pairs of the desired map—the *training data* $\{x_i, y_i\}_{i=1}^L \in X^L \times Y^L$—is provided. The variational parameters in v are optimized such that a *cost function* $C(\{x_i, y_i\}_{i=1}^L) = (1/L)\sum_{i=1}^L d[f_v(x_i), y_i]$ reaches a minimum. Here, d is an appropriate distance measure. Typically the optimization employs some variant of the gradient descent algorithm (see, e.g., [40]).

An AE is an FFNN for extracting the most relevant features from the input data. The network has a bottleneck—a layer with smaller width than the (equal) input and output layers. The training data is a set $\{x_i, x_i\}_{i=1}^L$ of equal training inputs and reference outputs. In general, the desired output for x is not x itself: the bottleneck (see Fig. 1) should force the AE to discard irrelevant information. Since no correct reference outputs are provided, the training of AEs is unsupervised.

We specify the quantum neuron from [14] by attributing a single qubit to every neuron. Let $\{|\uparrow\rangle, |\downarrow\rangle\}$ denote an orthonormal basis of a qubit. In each layer following the input, the jth neuron acts by a unitary U_j on its own qubit and the preceding layer. The noninput qubits are initialized in $|\downarrow\rangle$. The kth layer, $k > 1$, of m neurons maps the state ρ_{k-1} of layer $k-1$ onto

$$\mathcal{N}^k(\rho_{k-1}) \equiv \mathrm{tr}_{k-1}[U(\rho_{k-1} \otimes (|\downarrow\rangle_{\mathrm{out}}\langle\downarrow|)^{\otimes m})U^\dagger], \quad (1)$$

where the unitary $U \equiv U_m \cdots U_1$ is subject to optimization (see Fig. 1). Note that this definition is related to the general form of a quantum channel (see, e.g., [6,14,41,42]). The quantum channel describing the full network with M layers is $\mathcal{N}(\rho^{\mathrm{in}}) = \mathcal{N}^M(\cdots \mathcal{N}^2(\rho^{\mathrm{in}})\cdots)$. Our distance measure is one minus the fidelity F. For training data $\{\rho_i^{\mathrm{in}}, |\psi_i^{\mathrm{ref}}\rangle\}_{i=1}^L$ with pure desired outputs, $F(\rho, |\psi\rangle) = \langle\psi|\rho|\psi\rangle$ and the cost function reads

$$C(\{\rho_i^{\mathrm{in}}, |\psi_i^{\mathrm{ref}}\rangle\}_{i=1}^L) = 1 - \bar{F}(\{\mathcal{N}(\rho_i^{\mathrm{in}}), |\psi_i^{\mathrm{ref}}\rangle\}_{i=1}^L), \quad (2)$$

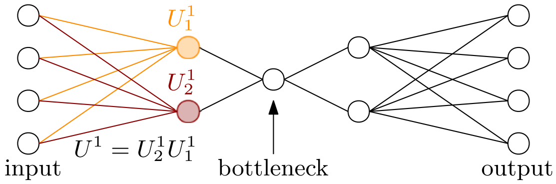

FIG. 1. Network architecture of an AE. The bottleneck prevents the AE from just copying the input data to the output so that it has to extract relevant features. Each neuron unitary acts on its qubit and the connected qubits in the previous layer (e.g., gold or red).

where $\bar{F}(\{\rho_i, |\psi_i\rangle\}_{i=1}^L) = (1/L)\sum_{i=1}^L F(\rho_i, |\psi_i\rangle) \le 1$. In the following, we abbreviate pure $\rho_i^{\mathrm{in}} = |\psi_i^{\mathrm{in}}\rangle\langle\psi_i^{\mathrm{in}}|$ by $|\psi_i^{\mathrm{in}}\rangle$.

Due to the no-cloning theorem, it is impossible to use copies of the training inputs $|\psi_i^{\mathrm{in}}\rangle$ as reference outputs $|\psi_i^{\mathrm{ref}}\rangle$. Instead, these states have to be prepared independently. If the data source is noisy, the paired states will be different due to different noise realizations. However, if these states share essential features, the AE can still be trained. Below, we use half of the noisy training data as input and half as reference output in unsupervised learning.

While, in practice, one has no access to the desired outputs of the NN $\{|\psi_i^{\mathrm{id}}\rangle\}_{i=1}^L$, the performance of AEs is best studied in a setting where these target states are known. We call the learning process successful if the mean *validation function*

$$\bar{F}_{\mathrm{val}}(\{\rho_i^{\mathrm{in}}, |\psi_i^{\mathrm{id}}\rangle\}_{i=1}^L) = \bar{F}(\{\mathcal{N}(\rho_i^{\mathrm{in}}), |\psi_i^{\mathrm{id}}\rangle\}_{i=1}^L) \quad (3)$$

is large, particularly, as compared to $\bar{F}(\{\rho_i^{\mathrm{in}}, |\psi_i^{\mathrm{id}}\rangle\}_{i=1}^L)$ before the NN is applied. We define

$$F_{\mathrm{val}}^{(i)}(\{\rho_i^{\mathrm{in}}, |\psi_i^{\mathrm{id}}\rangle\}_{i=1}^L) = F(\mathcal{N}(\rho_i^{\mathrm{in}}), |\psi_i^{\mathrm{id}}\rangle),$$
$$F^{(i)}(\{\rho_i^{\mathrm{in}}, |\psi_i^{\mathrm{id}}\rangle\}_{i=1}^L) = F(\rho_i^{\mathrm{in}}, |\psi_i^{\mathrm{id}}\rangle). \quad (4)$$

Note that the validation function, which compares $\{\mathcal{N}(\rho_i^{\mathrm{in}})\}_{i=1}^L$ with the target states, differs from the fidelity entering the cost function for training, which compares $\{\mathcal{N}(\rho_i^{\mathrm{in}})\}_{i=1}^L$ with the noisy data.

For the classical simulation of the quantum AE we have upgraded the MATLAB code from [14]. Most importantly, we now use the Nadam [40,43] gradient descent algorithm. The updated code is available at [44].

Noisy test states.—We consider highly entangled Greenberger-Horne-Zeilinger (GHZ), W, Dicke, and cluster states. Such states are important for quantum information and quantum enhanced metrology [8,45,46]. We call

$$|\mathrm{GHZ}_\phi\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle^{\otimes m} + e^{i\phi}|\downarrow\rangle^{\otimes m}) \quad (5)$$

an m-qubit GHZ state with phase ϕ or a GHZ-ϕ state. GHZ states are macroscopic superposition states with maximal entanglement depth. W, Dicke, and cluster states are discussed in the Supplemental Material [47].

For practical applications the states have to be protected from experimental noise. We investigate two complementary noise processes—spin-flip errors and small random unitary transformations (see, e.g., [5,41,42])—and show how quantum AEs can be used to denoise small GHZ, W, Dicke, and cluster states.

For spin-flip errors we assume that for a time T all qubits are flipped back and forth at some rate Γ. Thus each qubit has a probability of $p = (1 - e^{-2\Gamma T})/2 \le 0.5$ to end up in a flipped state. The flips of the jth qubit affect the density matrix ρ of the initial, noiseless, m-qubit state according to

$$\mathcal{E}_j(p,\rho) = p\sigma_j^x \rho \sigma_j^x + (1-p)\rho,$$
$$\sigma_j^x = \bigotimes_1^{j-1} Id \otimes \sigma^x \bigotimes_{j+1}^{m} Id, \tag{6}$$

where $Id = |\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|$ is the identity and $\sigma^x = |\uparrow\rangle\langle\downarrow| + |\downarrow\rangle\langle\uparrow|$ the spin-flip operator for a single qubit. The total noise channel is obtained by concatenating $\mathcal{E}_j$ for all qubits $j \in \{1, \ldots, m\}$:

$$\mathcal{E}(p,\rho) = \mathcal{E}_m(p, \mathcal{E}_{m-1}(p, \cdots \mathcal{E}_1(p,\rho)\cdots)) \tag{7}$$

We assume that in each experimental shot a subset $J \subseteq \{1, 2, \ldots, m\}$ of a total of m qubits is flipped. The probability of $\rho_J = \prod_{j\in J} \sigma_j^x \rho \prod_{j\in J} \sigma_j^x$ is $P_p(J) = p^{|J|}(1-p)^{m-|J|}$. Note that states ρ_J with different J may coincide or have nonorthogonal supports.

Our unitary noise is given by the evolution with a random time-dependent Hamiltonian. The noise strength is captured by a dimensionless parameter q. See the Supplemental Material [47] for details. A random Hamiltonian, in general, includes dephasing. However, since GHZ states are particularly sensitive to dephasing, we discuss it separately in the Supplemental Material [47].

Denoising a single state.—First, we show how well an AE can denoise four-qubit GHZ states with zero phase. We employ two AE topologies. One is the deep QNN denoted by $[4, 2, 1, 2, 4]$ and the other one is a stacked QNN: we train the AE $\sim[4, 1, 4]$ but denoise with $\sim[4, 1, 4, 1, 4]$ by applying $[4, 1, 4]$ twice. Each training employs 200 training pairs and takes 200 steps of the gradient descent algorithm (200 *training rounds*). We test the trained AEs on 200 GHZ-0 states exposed to the respective noise. The validation function, which, ideally, should reach one, is the fidelity between the denoised output of the AE and the GHZ-0 state.

Figure 2 summarizes our results in the case of spin-flip errors. For each spin-flip probability p we, first, draw the training data and one set of $L = 200$ noisy test states $\{|\psi_i\rangle\}_{i=1}^{L}$ according to the probability distribution $P_p(J)$. We independently train both AE topologies. For each topology, we apply the respective AE to every $|\psi_i\rangle$ and get outputs ρ_i. To assess the performance of the AE, we evaluate the mean validation function after denoising—$\bar{F}_{\text{val}}(\{|\psi_i\rangle, |\text{GHZ}_0\rangle\}_{i=1}^{L})$ (yellow circles and violet crosses)—and compare it to its value before denoising—$\bar{F}(\{|\psi_i\rangle, |\text{GHZ}_0\rangle\}_{i=1}^{L})$ (red dots). We find that up to $p = 0.3$ both AE topologies remove the spin-flip errors almost ideally, see the Supplemental Material [47] for a discussion.

The error bars of $\bar{F}_{\text{val}}$ indicate the standard deviation $\Delta F_{\text{val}} = \sqrt{\sum_i (F_{\text{val}}^{(i)} - \bar{F}_{\text{val}})^2}$. Note that, contrary to $F_{\text{val}}^{(i)}$, $\bar{F}_{\text{val}} + \Delta F_{\text{val}}$ can exceed one. For the input, $\Delta F = \sqrt{\bar{F}(1-\bar{F})}$ is large since $F^{(i)} \in \{0, 1\}$. Instead of adding error bars to $\bar{F}$, we show how $\{|\psi_i\rangle\}_{i=1}^{L}$ compares to the ideal probability distribution $P_p(J)$ of spin-flipped GHZ-0 states. The blue plus signs mark the expectation value of F, $\bar{F}^{\infty} = (1-p)^4 + p^4$. Their vertical bars indicate the standard deviation $\Delta F^{\infty}/\sqrt{L} = \sqrt{\bar{F}^{\infty}(1-\bar{F}^{\infty})/L}$, which characterizes the spread of the average $\bar{F}(\{|\psi_i\rangle, |\text{GHZ}_0\rangle\}_{i=1}^{L})$ for independent draws of L noisy states.

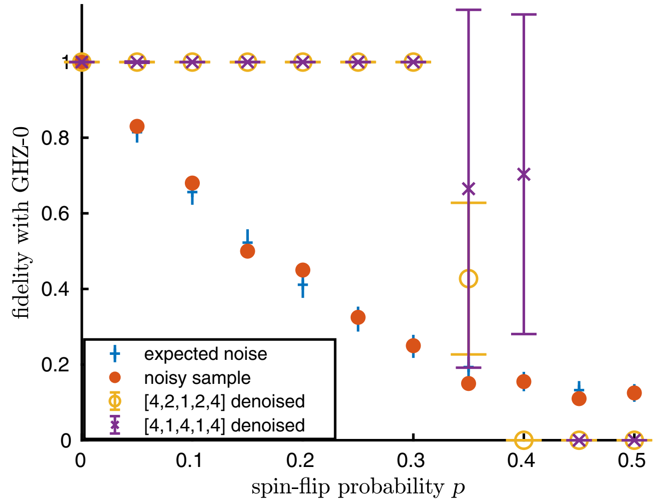

FIG. 2. Quantum AEs removing spin-flip errors from the GHZ-0 state. We show the average fidelity of noisy test states with the GHZ-0 state before denoising (red dots, $\bar{F}$) and after denoising (yellow circles and violet crosses, $\bar{F}_{\text{val}}$). Error bars display standard deviations. Blue plus signs show $\bar{F}^{\infty} \pm \Delta F^{\infty}$. The arrays $[4, 2, 1, 2, 4]$ and $[4, 1, 4, 1, 4]$ indicate different AE topologies. 200 noisy training pairs, training rounds, and noisy test states per p.

Similar results for unitary noise, its combination with spin-flip errors, and for dephasing are presented in the Supplemental Material [47]. The successful denoising of W, Dicke, and cluster states subject to spin-flip errors and unitary noise is also demonstrated in the Supplemental Material [47]. We observe that genuinely deep AEs, e.g., with a $[4, 2, 1, 2, 4]$ topology, perform better than shallow or even stacked ones.

Denoising multiple states.—So far we have demonstrated that an AE can denoise the state on which it has been trained. But it can do better. An AE can learn to denoise multiple target states, including ones not contained in the training data. It is crucial, though, that the noise process is sufficiently different from the transformations connecting the target states. Otherwise, the attribution of noisy states to target states becomes ambiguous. Assume that an experiment encodes some information into the phase of a GHZ state, and that this GHZ state is affected by spin-flip errors. We show that an AE can denoise the output of such an experiment. We consider three-qubit states and employ the simplest possible AE topology: $\sim[3, 1, 3]$.

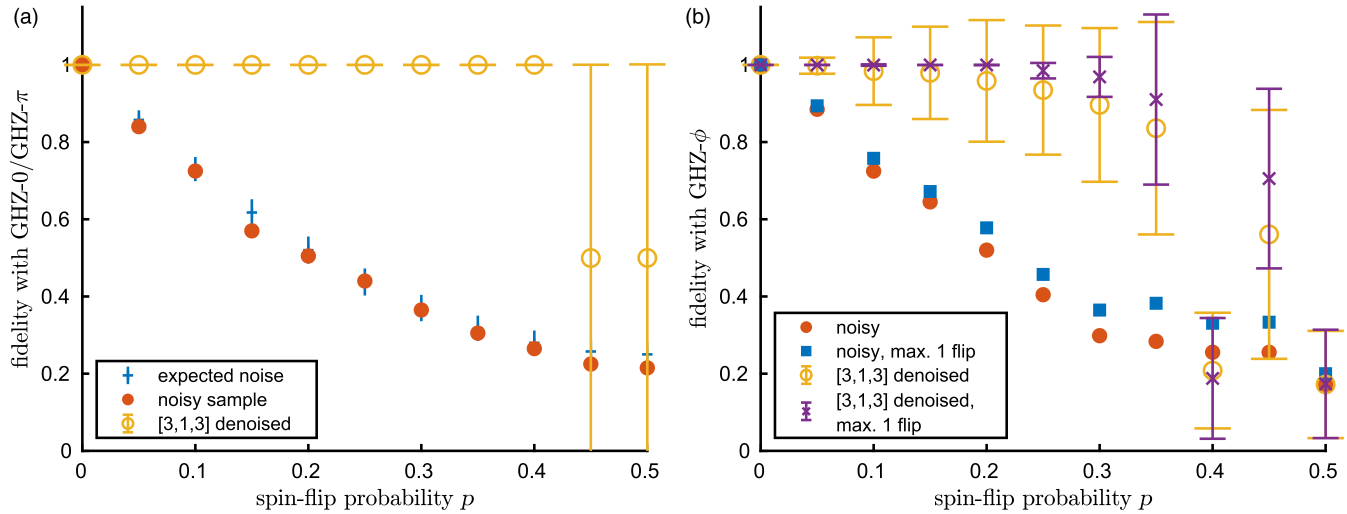

FIG. 3. $[3, 1, 3]$ quantum AEs correcting spin-flip errors in mixtures of GHZ-ϕ states with different phases ϕ. We show the average fidelity of noisy test states with the respective noiseless GHZ-ϕ states before denoising (red dots) and after denoising (yellow circles). Error bars display standard deviations. For each p: 100 training pairs per training phase, 200 training rounds, and 200 test pairs. (a) Fifty-fifty mixture of GHZ-0 and GHZ-π states, both for training and testing. Blue plus signs show $\bar{F}^{\infty} \pm \Delta F^{\infty}$. (b) Training phases $\{0, \pi/3, 2\pi/3, \pi\}$, and testing on random phases $\phi \in (0, \pi)$. Blue squares before denoising and violet crosses after denoising are obtained for the test states with $|J| \leq 1$.

As a first example, we imagine an experiment which outputs either a GHZ state with zero phase, GHZ-0, or with phase π, GHZ-π. The AE is trained for 200 rounds on 100 pairs of noisy GHZ-0 states and 100 pairs of noisy GHZ-π states. To test the performance of the trained AE, we apply it to 100 noisy GHZ-0 states and 100 noisy GHZ-π states and compare each output to the respective noiseless state. Figure 3(a) shows that the AE excellently denoises the two GHZ states up to a spin-flip probability of $p = 0.4$. Note that the AE deduces whether the experiment has given a phase of zero or π from the particular noisy input state alone.

Our second example is even more demanding. We assume that the experiment can output a GHZ state with any phase $\phi \in [0, \pi]$. We restrict the phase to $[0, \pi]$ because it is impossible to distinguish a GHZ-ϕ state with $|J|$ flipped qubits from a GHZ-($-\phi$) state with $3 - |J|$ flipped qubits. The training involves only four equidistant training phases ϕ_i between $\phi_0 = 0$ and $\phi_4 = \pi$. It, again, employs 100 training pairs per ϕ_i and takes 200 training rounds. We test the AE on 200 noisy GHZ-ϕ states with randomly chosen phases $\phi \in (0, \pi)$.

Considering a GHZ-ϕ state with $\phi \notin \pi\mathbb{Z}$ roughly doubles the number of different spin-flipped states as compared to $\phi \in \pi\mathbb{Z}$. Only for $\phi \in \pi\mathbb{Z}$, the flipped m-qubit states $\prod_{j\in J} \sigma_j^x |\mathrm{GHZ}_\phi\rangle$ and $\prod_{j\in M\setminus J} \sigma_j^x |\mathrm{GHZ}_\phi\rangle$ with $J \subseteq M = \{0, 1, \ldots, m\}$ are, up to a global phase, identical. As a consequence, for three-qubit GHZ-ϕ states with $\phi \in \pi\mathbb{Z}$, correcting spin-flip errors with $|J| = 1$ suffices for perfect denoising. For $\phi \notin \pi\mathbb{Z}$, errors with $|J| = 2$ and $|J| = 3$ need to be regarded separately.

Figure 3(b) displays the capability of the AE to denoise GHZ states with a random phase. Note that for $p = 0$ the fidelity of the outputs with the test states reaches one. Because of the bottleneck, the AE cannot learn the identity operation; nevertheless it correctly reproduces GHZ states with phases not contained in the training data. The AE improves the average value of the validation function for $p \leq 0.35$, but it leaves a considerable variance (yellow circles). However, if we keep only the test states with $|J| \leq 1$, we observe excellent denoising up to $p = 0.2$ (violet crosses).

Discussion.—We have constructed quantum AEs and have shown that these AEs can remove spin-flip errors and random unitary noise from small GHZ, W, Dicke, and cluster states. Particularly, correcting spin-flip errors has succeeded for a set of GHZ states parametrized by a continuous phase parameter. Thus, AEs for denoising can be used not only for state preparation but also for metrology. In principle, our method can be applied to any set of quantum states subject to any kind of noise. Further possible applications of quantum AEs include data compression, quantum error correction, and parametrized state preparation.

We expect that larger input states will require deeper networks. The number of quantum gates needed for one application of the fully connected AE scales exponentially with the width but only linearly with the depth of the network. The exponential scaling can be avoided by constraining the QNN, e.g., using sparse networks as discussed in the Supplemental Material [47].

Small universal quantum computers have been realized on several physical platforms, e.g., superconducting qubits and trapped ions [62,63]. If the state to be denoised is prepared on the same platform as the AE, both may be affected by equal noise, and the AE may become too noisy for denoising. However, there is a great interest in hybrid systems, which have been demonstrated, e.g., for superconducting qubits coupled to atomic and spin ensembles and for trapped ions with cold atoms [64–66]. Our proposal can help to denoise states from a noisy platform using a well-controlled one, or to remove deteriorating effects introduced at the interface between the coupled platforms. The impact of noise affecting the AE itself is discussed in [14,67].

Training an AE requires much more computational resources than testing it. To approach the experimental implementation, a small AE trained on a classical computer can be tested on a quantum computer, as has been done in [35] for data compression. Moreover, the photonic realization [33] of a compressing quantum AE suggests that also the training of our AE is within the reach of current quantum technology.

We thank Abhinav Anand, Kerstin Beer, Terry Farrelly, Carsten Klempt, Taras Kucherenko, Tobias J. Osborne, Robert Salzmann, Luis Santos, Daniel Scheiermann, Augusto Smerzi, Dorothee Tell, and Ramona Wolf for useful discussions. We acknowledge support by the SFB 1227 "DQ-mat," Projects No. A02 and No. A06, of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation).

*dimbond@live.com
†polina.feldmann@itp.uni-hannover.de

[1] M. M. Wolf, Mathematical foundations of supervised learning, https://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/\MA4801_2018S/ML_notes_main.pdf (2018).
[2] M. A. Nielsen, *Neural Networks and Deep Learning* (Determination Press, San Francisco, 2015).
[3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, 2016).
[4] I. Potamitis, Deep learning for detection of bird vocalisations, arXiv:1609.08408.
[5] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2010).
[6] R. F. Werner, Quantum information theory—An invitation, in *Quantum Information: An Introduction to Basic Theoretical Concepts and Experiments* (Springer, Berlin Heidelberg, 2001), pp. 14–57.
[7] S. Pirandola, U. L. Andersen, L. Banchi, M. Berta, D. Bunandar, R. Colbeck, D. Englund, T. Gehring, C. Lupo, C. Ottaviani, J. Pereira, M. Razavi, J. S. Shaari, M. Tomamichel, V. C. Usenko, G. Vallone, P. Villoresi, and P. Wallden, Advances in quantum cryptography, arXiv:1906.01645.
[8] L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, Quantum metrology with nonclassical states of atomic ensembles, Rev. Mod. Phys. **90**, 035005 (2018).
[9] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, Nat. Photonics **5**, 222 (2011).
[10] S. Szalay, M. Pfeffer, V. Murg, G. Barcza, F. Verstraete, R. Schneider, and Ö. Legeza, Tensor product methods and entanglement optimization for *ab initio* quantum chemistry, Int. J. Quantum Chem. **115**, 1342 (2015).
[11] R. Orús, Tensor networks for complex quantum systems, Nat. Rev. Phys. **1**, 538 (2019).
[12] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, Parameterized quantum circuits as machine learning models, Quantum Sci. Technol. **4**, 043001 (2019).
[13] Wikipedia contributors, Quantum machine learning—Wikipedia, the free encyclopedia, https://en.wikipedia.org/w/index.php?title=Quantum_machine_learning&oldid=883275117 (2019).
[14] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, Training deep quantum neural networks, Nat. Commun. **11**, 808 (2020).
[15] M. Schuld, I. Sinayskiy, and F. Petruccione, The quest for a quantum neural network, Quantum Inf. Process. **13**, 2567 (2014).
[16] M. Lewenstein, Quantum perceptrons, J. Mod. Opt. **41**, 2491 (1994).
[17] K. H. Wan, O. Dahlsten, H. Kristjánsson, R. Gardner, and M. S. Kim, Quantum generalisation of feedforward neural networks, npj Quantum Inf. **3**, 36 (2017).
[18] A. J. da Silva, T. B. Ludermir, and W. R. de Oliveira, Quantum perceptron over a field and neural network architecture selection in a quantum computer, Neural Netw. **76**, 55 (2016).
[19] M. V. Altaisky, Quantum neural network, arXiv:quant-ph/0107012.
[20] C. P. Gonçalves, Quantum neural machine learning: Backpropagation and dynamics, NeuroQuantology **15**, 22 (2017).
[21] N. Kouda, N. Matsui, H. Nishimura, and F. Peper, Qubit neural network and its learning efficiency, Neural Comput. Appl. **14**, 114 (2005).
[22] E. Torrontegui and J. J. García-Ripoll, Unitary quantum perceptron as efficient universal approximator, Europhys. Lett. **125**, 30004 (2019).
[23] E. Farhi and H. Neven, Classification with quantum neural networks on near term processors, arXiv:1802.06002.
[24] M. Schuld, A. Bocharov, K. Svore, and N. Wiebe, Circuit-centric quantum classifiers, arXiv:1804.00633.
[25] Y. Cao, G. G. Guerreschi, and A. Aspuru-Guzik, Quantum neuron: An elementary building block for machine learning on quantum computers, arXiv:1711.11240.
[26] N. Killoran, T. R. Bromley, J. M. Arrazola, M. Schuld, N. Quesada, and S. Lloyd, Continuous-variable quantum neural networks, Phys. Rev. Research **1**, 033063 (2019).
[27] J. M. Arrazola, T. R. Bromley, J. Izaac, C. R. Myers, K. Brádler, and N. Killoran, Machine learning method for state preparation and gate synthesis on photonic quantum computers, Quantum Sci. Technol. **4**, 024004 (2019).
[28] G. R. Steinbrecher, J. P. Olson, D. Englund, and J. Carolan, Quantum optical neural networks, arXiv:1808.10047.
[29] G. Verdon, J. Pye, and M. Broughton, A universal training algorithm for quantum deep learning, arXiv:1806.09729.
[30] A. Monràs, G. Sentís, and P. Wittek, Inductive Supervised Quantum Learning, Phys. Rev. Lett. **118**, 190503 (2017).

[31] E. Grant, M. Benedetti, S. Cao, A. Hallam, J. Lockhart, V. Stojevic, A. G. Green, and S. Severini, Hierarchical quantum classifiers, npj Quantum Inf. **4**, 65 (2018).

[32] J. Romero, J. P. Olson, and A. Aspuru-Guzik, Quantum autoencoders for efficient compression of quantum data, Quantum Sci. Technol. **2**, 045001 (2017).

[33] A. Pepper, N. Tischler, and G. J. Pryde, Experimental Realization of a Quantum Autoencoder: The Compression of Qutrits via Machine Learning, Phys. Rev. Lett. **122**, 060501 (2019).

[34] L. Lamata, U. Alvarez-Rodriguez, J. D. Martín-Guerrero, M. Sanz, and E. Solano, Quantum autoencoders via quantum adders with genetic algorithms, Quantum Sci. Technol. **4**, 014007 (2018).

[35] Y. Ding, L. Lamata, M. Sanz, X. Chen, and E. Solano, Experimental implementation of a quantum autoencoder via quantum adders, Adv. Quantum Technol. **2**, 1800065 (2019).

[36] A. A. Melnikov, H. P. Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger, and H. J. Briegel, Active learning machine learns to create new quantum experiments, Proc. Natl. Acad. Sci. U.S.A. **115**, 1221 (2018).

[37] R. Nichols, L. Mineh, J. Rubio, J. C. F. Matthews, and P. A. Knott, Designing quantum experiments with a genetic algorithm, Quantum Sci. Technol. **4**, 045012 (2019).

[38] P. A. Knott, A search algorithm for quantum state engineering and metrology, New J. Phys. **18**, 073033 (2016).

[39] G. Sentís, A. Monràs, R. Muñoz-Tapia, J. Calsamiglia, and E. Bagan, Unsupervised classification of quantum data, Phys. Rev. X **9**, 041029 (2019).

[40] S. Ruder, An overview of gradient descent optimization algorithms, arXiv:1609.04747.

[41] M. M. Wolf, Quantum channels & operations guided tour, https://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf (2012).

[42] C. Bény and F. Richter, Algebraic approach to quantum theory: A finite-dimensional guide, arXiv:1505.03106.

[43] T. Dozat, Incorporating nesterov momentum into adam, http://cs229.stanford.edu/proj2015/054_report.pdf (2015).

[44] https://dx.doi.org/10.5281/zenodo.3713675.

[45] W. Dür, G. Vidal, and J. I. Cirac, Three qubits can be entangled in two inequivalent ways, Phys. Rev. A **62**, 062314 (2000).

[46] R. Raussendorf, D. E. Browne, and H. J. Briegel, Measurement-based quantum computation on cluster states, Phys. Rev. A **68**, 022312 (2003).

[47] See the Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.124.130502 for unitary noise, dephasing, W, Dicke, and cluster states, denoising limitations for spin-flip errors, and sparse quantum AEs, which includes Refs. [48–61].

[48] T. Zhou and X. Chen, Operator dynamics in a brownian quantum circuit, Phys. Rev. E **99**, 052212 (2019).

[49] N. Lashkari, D. Stanford, M. Hastings, T. Osborne, and P. Hayden, Towards the fast scrambling conjecture, J. High Energy Phys. 04 (2013) 22.

[50] J. Preskill, Lecture notes for ph219/cs219: Quantum information, http://www.theory.caltech.edu/people/preskill/ph219/chap3_15.pdf (2018).

[51] R. Raussendorf and H. J. Briegel, A One-Way Quantum Computer, Phys. Rev. Lett. **86**, 5188 (2001).

[52] M. Hein, J. Eisert, and H. J. Briegel, Multiparty entanglement in graph states, Phys. Rev. A **69**, 062311 (2004).

[53] R. H. Dicke, Coherence in spontaneous radiation processes, Phys. Rev. **93**, 99 (1954).

[54] G. Tóth, Multipartite entanglement and high-precision metrology, Phys. Rev. A **85**, 022322 (2012).

[55] G. Tóth, Detection of multipartite entanglement in the vicinity of symmetric dicke states, J. Opt. Soc. Am. B **24**, 275 (2007).

[56] B. Lücke, J. Peise, G. Vitagliano, J. Arlt, L. Santos, G. Tóth, and C. Klempt, Detecting Multiparticle Entanglement of Dicke States, Phys. Rev. Lett. **112**, 155304 (2014).

[57] B. Lücke, M. Scherer, J. Kruse, L. Pezzé, F. Deuretzbacher, P. Hyllus, O. Topic, J. Peise, W. Ertmer, J. Arlt, L. Santos, A. Smerzi, and C. Klempt, Twin matter waves for interferometry beyond the classical limit, Science **334**, 773 (2011).

[58] H. J. Briegel and R. Raussendorf, Persistent Entanglement in Arrays of Interacting Particles, Phys. Rev. Lett. **86**, 910 (2001).

[59] R. B. Griffiths, Graph states and graph codes, http://quantum.phys.cmu.edu/QCQI/qitd452.pdf (2014).

[60] R. Raussendorf, J. Harrington, and K. Goyal, Topological fault-tolerance in cluster state quantum computation, New J. Phys. **9**, 199 (2007).

[61] D. Schlingemann and R. F. Werner, Quantum error-correcting codes associated with graphs, Phys. Rev. A **65**, 012308 (2001).

[62] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, Superconducting qubits: Current state of play, arXiv:1905.13641.

[63] C. D. Bruzewicz, J. Chiaverini, R. McConnell, and J. M. Sage, Trapped-ion quantum computing: Progress and challenges, Appl. Phys. Rev. **6**, 021314 (2019).

[64] G. Kurizki, P. Bertet, Y. Kubo, K. Mølmer, D. Petrosyan, P. Rabl, and J. Schmiedmayer, Quantum technologies with hybrid systems, Proc. Natl. Acad. Sci. U.S.A. **112**, 3866 (2015).

[65] Z. L. Xiang, S. Ashhab, J. Q. You, and F. Nori, Hybrid quantum circuits: Superconducting circuits interacting with other quantum systems, Rev. Mod. Phys. **85**, 623 (2013).

[66] M. Tomza, K. Jachymski, R. Gerritsma, A. Negretti, T. Calarco, Z. Idziaszek, and P. S. Julienne, Cold hybrid ion-atom systems, Rev. Mod. Phys. **91**, 035001 (2019).

[67] D. Scheiermann, Noise robustness of quantum neural networks, Bachelor's thesis, Leibniz Universität Hannover, 2019.