This is a repository copy of *Mixed State Entanglement Classification using Artificial Neural Networks*.

PAPER • OPEN ACCESS

Mixed state entanglement classification using artificial neural networks

View the article online for updates and enhancements.

New Journal of Physics
The open access journal at the forefront of physics

PAPER

Mixed state entanglement classification using artificial neural networks

Cillian Harney[1],[*] , Mauro Paternostro[2] and Stefano Pirandola[1]

1   Department of Computer Science, University of York, York YO10 5GH, United Kingdom
2   Centre for Theoretical Atomic, Molecular and Optical Physics, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, United Kingdom
*   Author to whom any correspondence should be addressed.

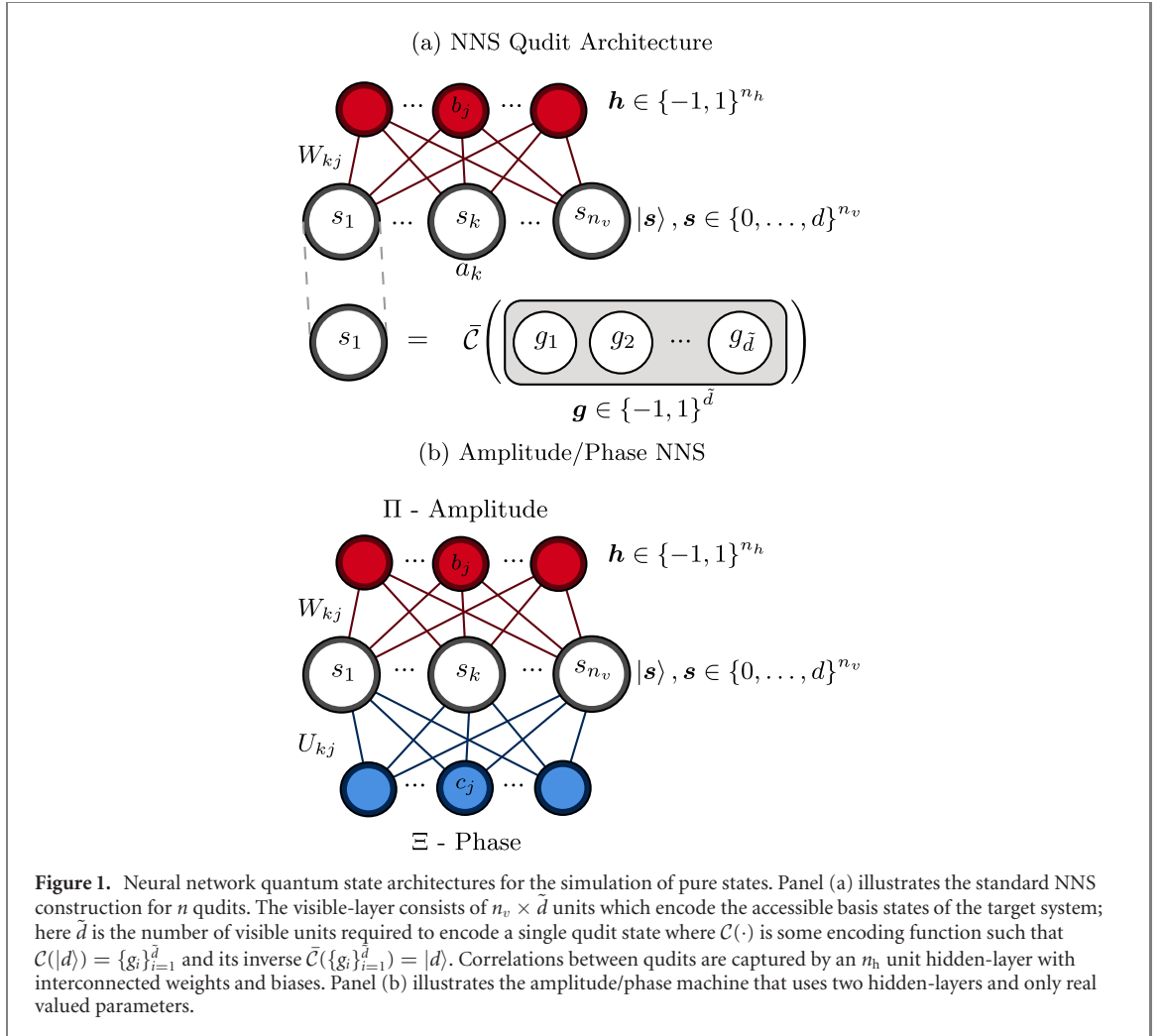**E-mail:** cth528@york.ac.uk

Abstract

Reliable methods for the classification and quantification of quantum entanglement are fundamental to understanding its exploitation in quantum technologies. One such method, known as separable neural network quantum states (SNNS), employs a neural network inspired parameterization of quantum states whose entanglement properties are explicitly programmable. Combined with generative machine learning methods, this ansatz allows for the study of very specific forms of entanglement which can be used to infer/measure entanglement properties of target quantum states. In this work, we extend the use of SNNS to mixed, multipartite states, providing a versatile and efficient tool for the investigation of intricately entangled quantum systems. We illustrate the effectiveness of our method through a number of examples, such as the computation of novel tripartite entanglement measures, and the approximation of ultimate upper bounds for qudit channel capacities.

The core tasks of entanglement *classification* [1–3] and *quantification* [4–6] are essential for future quantum technologies, and ask the seemingly straightforward questions: given a quantum state ρ, is it entangled? If so, by how much is it entangled? As the system size or dimension of a quantum system grows, these questions become highly non-trivial and in general there are no universal criteria or methods to provide answers. The most popular mathematical recipe for classification, the positive partial transposition (PPT) criterion (or Peres–Horodecki criterion) [7, 8], applies only to $(2 \otimes 2)$ or $(2 \otimes 3)$ bipartite systems. As one extends to multipartite, higher-dimensional quantum systems more sophisticated tools are required.

The application of classical machine learning tools for the study of quantum systems, such as artificial neural networks, have seen a surge of interest due to their remarkable expressive power and efficiency [9–11]. In particular, Carleo and Troyer [12] showed that restricted Boltzmann machines offer a resoundingly appropriate classical representation of quantum states, due to their ability to perform dimensionality reduction, their non-local information distribution, and optimization capacity [13]. Ansatzes based on this architecture are known as neural network quantum states (NNS), and they have been a successful classical simulation tool in a variety of contexts such as tomography [14–17], open quantum system dynamics [18–22], and the study of quantum technologies [23–26].

The versatility of NNS also provides an excellent framework for the study of entanglement [27]. As introduced for pure, qubit states in reference [28], it is possible to manipulate and constrain these neural networks in a way that guarantees a strict form of separability. These constrained variational states are known as separable neural network states (SNNS). Combined with a quantum state reconstruction algorithm, this introduces a unique entanglement witness protocol based on the reconstructive performance of an SNNS with a target state.

In this paper, we generalize these results to mixed, *d*-dimensional quantum states. We show how SNNS can be used to perform highly specific entanglement classification, and approximate entanglement measures to a very high degree of accuracy. The ability to implicitly characterize the space of separable states is

Figure 1. Neural network quantum state architectures for the simulation of pure states. Panel (a) illustrates the standard NNS construction for *n* qudits. The visible-layer consists of $n_v \times \tilde{d}$ units which encode the accessible basis states of the target system; here $\tilde{d}$ is the number of visible units required to encode a single qudit state where $\mathcal{C}(\cdot)$ is some encoding function such that $\mathcal{C}(|d\rangle) = \{g_i\}_{i=1}^{\tilde{d}}$ and its inverse $\bar{\mathcal{C}}(\{g_i\}_{i=1}^{\tilde{d}}) = |d\rangle$. Correlations between qudits are captured by an n_h unit hidden-layer with interconnected weights and biases. Panel (b) illustrates the amplitude/phase machine that uses two hidden-layers and only real valued parameters.

extremely valuable, and allows one to compute entanglement measures that are otherwise extremely difficult to measure, such as the relative entropy of entanglement (REE) [29].

This paper is structured as follows: in section 1 we revise the NNS architecture and its variants for pure and mixed states. Section 2 overviews separable architectures, and shows how specific forms of entanglement can be guaranteed. In section 3 the methods of classification and quantification using SNNS are discussed. Section 4 provides numerical evidence for their utility through a number of relevant examples, with interesting applications in the study of noisy tripartite entanglement, bound entanglement, and quantum channel capacities. Finally, conclusions and future directions are addressed in section 5.

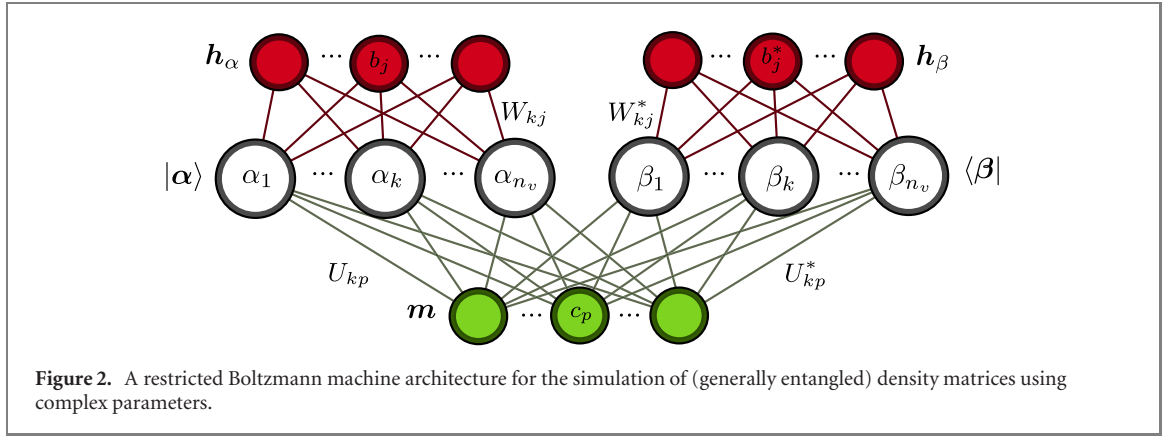## 1. Neural network quantum states

1.1. Pure states

The simplest neural network model we can introduce is the positive, real NNS. This model uses a real valued restricted Boltzmann machine (RBM) architecture, with n_v visible units $\boldsymbol{s} = \{s_1, \ldots, s_{n_v}\}$ representing the number of qudits being modelled within the target quantum system, fully interconnected with n_h hidden units $\boldsymbol{h} = \{h_1, \ldots, h_{n_h}\}$. The visible units are typically binary valued to study $d = 2$-dimensional systems, $s_i \in \{-1, 1\}$ as are the hidden units $h_j \in \{-1, 1\}$; however this depends on the system being modelled. This network architecture allows us to capture the correlations of the objective quantum system through network parameters:

$$\Pi = \{a_k, b_j, W_{kj}\} \quad \text{for } k \in [1, n_v], \quad j \in [1, n_h], \tag{1}$$

$$\boldsymbol{a} \in \mathbb{R}^{n_v}, \quad \boldsymbol{b} \in \mathbb{R}^{n_h}, \quad W \in \mathbb{R}^{n_v \times n_h}, \tag{2}$$

where $\boldsymbol{a}$ are visible biases, $\boldsymbol{b}$ are hidden biases, and W is the network weight matrix. The total number of parameters is $|\Pi| = n_h \cdot n_v + n_h + n_v$ (see figure 1).

Figure 2. A restricted Boltzmann machine architecture for the simulation of (generally entangled) density matrices using complex parameters.

The inherent advantage offered by the RBM architecture for generative modelling is that there are no *intra*-layer connections (i.e. there are no connections between adjacent visible units or hidden units). This allows for an ansatz that is independent from the activations of the hidden state space. Thus, one can define a positive NNS wavefunction as [12]

$$\Psi_\Pi(\boldsymbol{s}) = e^{\sum_{k=1}^{n_v} a_k s_k} \prod_{j=1}^{n_h} 2 \cosh\left(\sum_k W_{kj} s_k + b_j\right),\qquad(3)$$

and therefore the NNS is $|\Psi_\Pi\rangle = \sum_{\boldsymbol{s}} \Psi_\Pi(\boldsymbol{s})|\boldsymbol{s}\rangle$.

Whilst NNS have typically been applied to qubit systems using binary visible units, one can extend the modelling to *d*-dimensional qudits by using a set of visible binary neurons that collectively represent a single qudit [17]. One may choose to encode *d*-dimensional states using a collection of $\tilde{d}$ visible, binary neurons via an encoding function $\mathcal{C}$, i.e.

$$|s\rangle \mapsto \mathcal{C}(s) = \{g_1, g_2, \ldots, g_{\tilde{d}}\} = \boldsymbol{g}.\qquad(4)$$

The n_v qudit visible-layer can then be encoded into $\tilde{n}_v = \tilde{d} n_v > n_v$ visible neurons,

$$\boldsymbol{s} = \{s_1, s_2, \ldots, s_{n_v}\} \mapsto \{\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_{\tilde{n}_v}\}.\qquad(5)$$

We may identically define the qudit decoding function $\bar{\mathcal{C}}$ such that $\bar{\mathcal{C}}(\boldsymbol{g}) = |s\rangle$. One may encode qudits into binary codes on the visible-layer $|s\rangle \mapsto \mathrm{bin}(s)$, requiring $\tilde{n}_v = \lceil \log_2 d \rceil n_v$ visible binary neurons, which however requires $d = 2^r$ for some integer r in order to admit a complete basis set. For arbitrary d it may be more useful to utilize one-hot encoding such that $|s\rangle \mapsto \mathrm{one-hot}(s) = \boldsymbol{e}_s^d$ where $\boldsymbol{e}_s^d$ is a d-length vector that is zero at all indices except index s.

In order to study non-positive quantum states one can introduce complex network parameters. Letting $a_k = \alpha_k + \mathrm{i}\beta_k$, $b_j = \gamma_j + \mathrm{i}\lambda_j$, and $W_{kj} = \Gamma_{kj} + \mathrm{i}\Lambda_{kj}$, then the NNS wavefunction is

$$\Psi_\Pi(\boldsymbol{s}) = e^{\sum_{k=1}^{n_v} (\alpha_k + \mathrm{i}\beta_k) s_k} \prod_{j=1}^{n_h} 2 \cosh\left(\theta_j^\gamma + \mathrm{i}\theta_j^\lambda\right),\qquad(6)$$

where $\theta_j^\gamma = \sum_k \Gamma_{kj} s_k + \gamma_j$, and $\theta_j^\lambda = \sum_k \Lambda_{kj} s_k + \lambda_j$. Thus the NNS can exhibit phase properties of quantum states. The network parameter set extends to $\Pi = \{a_k, b_j, W_{kj}\} \in \mathbb{C}$.

Alternatively one can preserve reality of network parameters by restructuring the nature of the NNS ansatz itself. In particular we can construct an ansatz that uses two RBMs that unify to represent a complete state. Defining a variational phase state $\Phi_\Xi(\boldsymbol{s})$, and amplitude state $\Psi_\Pi(\boldsymbol{s})$, this network ansatz is given as [14]

$$|\Psi_{\Pi,\Xi}\rangle = \sum_{\boldsymbol{s}} e^{\mathrm{i} \log \Phi_\Xi(\boldsymbol{s})} \Psi_\Pi(\boldsymbol{s})|\boldsymbol{s}\rangle.\qquad(7)$$

Therefore both the variational phase and amplitude networks need only be real valued, since the complex/phase properties of the state are managed through the complex exponential. The state is now defined by two parameter sets, $\Pi = \{a_k, b_j, W_{kj}\} \in \mathbb{R}$ and $\Xi = \{c_k, d_j, U_{kj}\} \in \mathbb{R}$.

1.2. Mixed states

To extend the variational ansatz to mixed states requires the addition of a hidden mixing-layer with n_m hidden units, capable of encoding the classical probability distribution of the mixed quantum state [19–21]. The network state can be constructed from two sets of variational network parameters: $\Pi = \{c_p, U_{kp}\}$, $c_p \in \mathbb{R}^{n_m}$ and $U_{kp} \in \mathbb{C}^{n_v \times n_m}$ encoding the mixing probabilities [30] and the previously defined $\Xi = \{a_k, b_j, W_{kj}\} \in \mathbb{C}$ which encodes the pure state probability distribution. Let the density-matrix row and column degrees of freedom be described by basis vectors $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$ respectively. As these parameter sets are independent, we may describe a density-matrix element as a contribution from a *classical mixing state* $\mathcal{P}_\Pi$ and a *pure state* σ_Ξ.

The contribution from a classical mixing network is given by

$$\mathcal{P}_\Pi^{\boldsymbol{\alpha},\boldsymbol{\beta}} = \prod_{p=1}^{n_m} \cosh(\phi_p(\boldsymbol{\alpha},\boldsymbol{\beta})), \tag{8}$$

$$\phi_p(\boldsymbol{\alpha},\boldsymbol{\beta}) = c_p + \sum_k U_{kp}\alpha_k + U_{kp}^*\beta_k, \tag{9}$$

where x^* denotes complex conjugation. Meanwhile the pure state contribution is

$$\sigma_\Xi^{\boldsymbol{\alpha},\boldsymbol{\beta}} = e^{\omega(\boldsymbol{\alpha},\boldsymbol{\beta})} \prod_{j=1}^{n_h} \cosh(\theta_j(\boldsymbol{\alpha})) \cosh(\theta_j^*(\boldsymbol{\beta})), \tag{10}$$

$$\omega(\boldsymbol{\alpha},\boldsymbol{\beta}) = \sum_k a_k\alpha_k + a_k^*\beta_k, \tag{11}$$

$$\theta_j(\boldsymbol{x}) = b_j + \sum_k W_{kj}x_k. \tag{12}$$

The complete variational state can therefore be constructed as a sum over all density-matrix elements,

$$\rho_{\Pi,\Xi} = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} \mathcal{P}_\Pi^{\boldsymbol{\alpha},\boldsymbol{\beta}} \cdot \sigma_\Xi^{\boldsymbol{\alpha},\boldsymbol{\beta}} |\boldsymbol{\alpha}\rangle\langle\boldsymbol{\beta}| = \mathcal{P}_\Pi \odot \sigma_\Xi, \tag{13}$$

where $\odot$ is the Hadamard product. This architecture is presented in figure 2. It is important to emphasize that by construction, the classical mixing state $\mathcal{P}_\Pi$ cannot simulate quantum correlations, only classical correlations (see appendix A). The pure state density-matrix σ_Ξ alone is able to simulate the quantum correlations within the global network state. Just as a mixed state can be decomposed via a statistical ensemble of pure states $\{p_i; |\phi_i\rangle\}$, where $\rho = \sum_i p_i |\phi_i\rangle\langle\phi_i|$, equation (13) can be considered as a matrix element-wise decomposition of the density-matrix which is readily accessible via NNS.

The network parameters in this ansatz are necessarily complex, thus combining the control of phase and amplitude contributions much like equation (6). However, it may be desirable to formulate an ansatz that is similar to equation (7) in which phase/amplitude contributions are controlled by different networks. One *could* use the NNS in equation (7) to learn a vectorised density-matrix $\boldsymbol{\rho}_{\Pi,\Xi} = \text{vec}(\rho_{\Pi,\Xi}) = |\rho_{\Pi,\Xi}\rangle$, where the function $\text{vec}(\cdot)$ simply reshapes an n-qudit, $d^n \times d^n$ density-matrix into a d^{2n} column vector. It follows that two real parameter RBMs could then be used to learn phase and amplitude properties respectively, as with pure states. Whilst optimal convergence towards a target vectorised mixed state is possible in this way, the ansatz itself is neither Hermitian or positive semi-definite under reshaping to a matrix. That is, given an inverse vectorisation function $\text{vec}^{-1}(\cdot)$ which reshapes a d^{2n} column vector into $d^n \times d^n$ density-matrix, then $\rho_{\Pi,\Xi} = \text{vec}^{-1}(\boldsymbol{\rho}_{\Pi,\Xi})$ is not a valid density-matrix. Therefore, this form of ansatz may represent states that are non-physical, which is clearly not desirable.

Instead, we can restructure the mixed state ansatz in order to take a closer form to the complex exponential format utilized in the previous section. Let the real parameter sets Ξ, Π be used to describe the pure state phase and amplitude networks respectively, and the complex parameter set Ω used to describe the mixing network. Recall a pure state wavefunction in complex exponential form $\Psi_{\Pi,\Xi}(\boldsymbol{\alpha}) = e^{i \log \varphi_\Xi(\boldsymbol{\alpha})} \sigma_\Pi(\boldsymbol{\alpha})$. It is useful to define the following functions of our pure density-matrix phase/amplitude wavefunctions

$$\Phi_\Xi^{\boldsymbol{\alpha},\boldsymbol{\beta}} = \frac{\varphi_\Xi(\boldsymbol{\alpha})}{\varphi_\Xi(\boldsymbol{\beta})}, \qquad \Gamma_\Pi^{\boldsymbol{\alpha},\boldsymbol{\beta}} = \sigma_\Pi(\boldsymbol{\alpha})\sigma_\Pi(\boldsymbol{\beta}). \tag{14}$$

In order to incorporate classical mixing we need a mixing-layer that takes a similar vectorized form.

Omitting the visible biases which are already possessed by the pure states, the mixing-layer takes the form

$$\mathcal{P}_{\Omega}^{\alpha,\beta} = \prod_{p=1}^{n_m} \cosh(\mu_p + i\psi_p) = \prod_{p=1}^{n_m} r_p^{\alpha,\beta} \, e^{i \log \vartheta_p^{\alpha,\beta}}, \tag{15}$$

$$\mu_p(\boldsymbol{\alpha}, \boldsymbol{\beta}) = c_p + \sum_k R_{kp}(\alpha_k + \beta_k), \tag{16}$$

$$\psi_p(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_k I_{kp}(\alpha_k - \beta_k), \tag{17}$$

where $R_{kp} = \text{Re}(U_{kp})$ and $I_{kp} = \text{Im}(U_{kp})$ denote the real and imaginary components of the mixing network respectively. One can then construct the following phase and amplitude functions for the classical mixing

$$r_{\Omega}^{\alpha,\beta} = \prod_{p=1}^{n_m} \sqrt{\cosh(\mu_p + i\psi_p)\cosh(\mu_p - i\psi_p)}, \tag{18}$$

$$\vartheta_{\Omega}^{\alpha,\beta} = \prod_{p=1}^{n_m} \exp\left[\frac{1}{2i} \log\left(\frac{-\cosh(\mu_p + i\psi_p)}{\cosh(\mu_p - i\psi_p)}\right)\right], \tag{19}$$

such that the vectorized mixing state takes the form $e^{i \log |\vartheta_\Omega\rangle}|r_\Omega\rangle$. This allows for any element of the complete mixed state to be expressed according to

$$\rho_{\Omega,\Pi,\Xi}^{\alpha,\beta} = e^{i \log\left(\Phi_\Xi^{\alpha,\beta} \vartheta_\Omega^{\alpha,\beta}\right)} \Gamma_\Pi^{\alpha,\beta} r_\Omega^{\alpha,\beta}. \tag{20}$$

2. Separable neural network architectures

2.1. Separable pure network states

Through restrictions on the connectivity of the weight matrix W_{kj}, one can guarantee separability of the generative network state. Let us define $\mathcal{K}$ as a collection of K-disjoint subsets $\mathcal{K} = \{\boldsymbol{k}_l\}_{l=1}^K$, that collect qudit indices from an n-qudit system. More precisely,

$$\mathcal{K} = \bigcup_{l=1}^K \boldsymbol{k}_l, \quad \text{s.t. } \{1, \ldots, n\} \subseteq \mathcal{K}, \tag{21}$$

$$\boldsymbol{k}_m \cap \boldsymbol{k}_l = \varnothing, \quad \forall \, m \neq l \in \{1, \ldots, n\}. \tag{22}$$
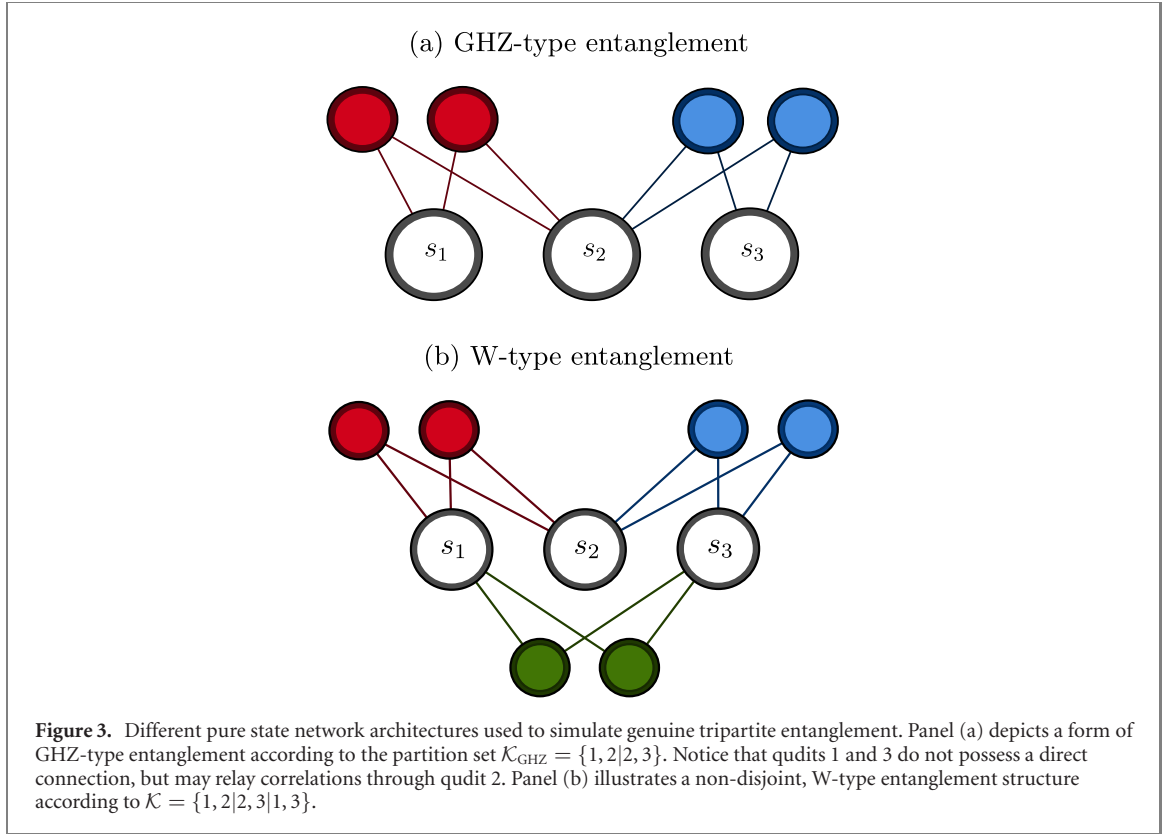
In equation (21) we have demanded that the global partition set necessarily contains all n-qudits in the system, and that subsets of qudits are disjoint in equation (22). Hence, an n-qudit, pure state $|\Psi\rangle$ is defined to be $\mathcal{K}$-separable if it can expressed as a tensor-product of sub-states $|\Psi\rangle = \bigotimes_{\boldsymbol{k}\in\mathcal{K}}|\psi_{\boldsymbol{k}}\rangle$, i.e. it is separable with respect to the partition set $\mathcal{K}$. This is a very precise format of separability, as it precisely specifies the arrangement of entangled parties. If we were to disregard specific party orderings we would refer to $(|\mathcal{K}| = K)$-separability.

Disjointedness in this definition of $\mathcal{K}$-separability ensures that each qudit is only entangled with respect to a single subset of the quantum system. This provides a specific level of detail to the entanglement structure, while also degenerating many forms of entanglement that we may not be interested in. For example, genuine tripartite entanglement under disjoint $\mathcal{K}$-separability allows for only a single set $\mathcal{K} = \{\boldsymbol{k}_1\} = \{1, 2, 3\}$ with no partitions. We may then define *non-disjoint* $\mathcal{K}$-separability as an extension of the previous definition simply by removing the conditions in equation (22). Using this non-disjoint definition, genuine tripartite entanglement allows for many more definitions, $\mathcal{K} = \{1, 2, 3\}, \{1, 2|2, 3\},$ $\{1, 2|2, 3|1, 3\}, \ldots$, which is studied in later sections (see figure 3 for an example).

To strictly impose either type of separability on an NNS, the goal is to express the wavefunction of the network state in the following form

$$\Psi_\Pi(\boldsymbol{s}) = \prod_{l=1}^K \psi_\Pi^{\boldsymbol{k}_l}(\boldsymbol{s}), \tag{23}$$

where $\psi_\Pi^{\boldsymbol{k}_l}$ are separable sub-wavefunctions that describe the behaviour of qudits in the partition $\boldsymbol{k}_l$. We may then construct an analogous hidden-layer partition set $\mathcal{H} = \{\boldsymbol{h}_l\}_{l=1}^K$, which assigns a subset of hidden units

Figure 3. Different pure state network architectures used to simulate genuine tripartite entanglement. Panel (a) depicts a form of GHZ-type entanglement according to the partition set $\mathcal{K}_{\mathrm{GHZ}} = \{1, 2|2, 3\}$. Notice that qudits 1 and 3 do not possess a direct connection, but may relay correlations through qudit 2. Panel (b) illustrates a non-disjoint, W-type entanglement structure according to $\mathcal{K} = \{1, 2|2, 3|1, 3\}$.

to each visible subset of entangled qudits $\mathcal{K} = \{\boldsymbol{k}_l\}_{l=1}^{K}$. By segmenting the layer of hidden units into these K-subsets and applying the following restriction to the weight matrix

$$W_{ij} = 0 \quad \text{for } i \in \boldsymbol{k}_l, \quad j \notin \boldsymbol{h}_l, \quad \forall \, l \in \{1, \ldots, K\}, \tag{24}$$

this condition then provides the complete, $\mathcal{K}$-separable network state

$$\Psi_{\Pi|\mathcal{K}}(\boldsymbol{s}) = \prod_{l=1}^{K} e^{\tilde{\omega}_l(\boldsymbol{s})} \prod_{j \in \boldsymbol{h}_l} 2 \, \cosh\left(\theta_l^j(\boldsymbol{s})\right),$$

$$\theta_l^j(\boldsymbol{s}) = \sum_{i \in \boldsymbol{k}_l} W_{ij} s_i + b_j, \qquad \tilde{\omega}_l(\boldsymbol{s}) = \sum_{i \in \boldsymbol{k}_l} a_i s_i. \tag{25}$$

2.2. Separable neural network density matrices

Whilst pure states are $\mathcal{K}$-separable when they can be expressed as the tensor product of $|\mathcal{K}| = k$ local sub-states, a mixed state possesses a form of separability iff it can be expressed as a convex combination of local sub-states $\rho^{\{\boldsymbol{k}_l\}_{l=1}^{K}}$. It is now useful to define two distinct forms of separability; *consistent* and *inconsistent* mixed-multipartite separability.

A state is consistently $\mathcal{K}$-separable if it can be expressed as a convex combination of states which *all* admit an identical form of separability,

$$\rho^{\mathcal{K}} = \sum_j p_j \bigotimes_{\boldsymbol{k} \in \mathcal{K}} \rho_j^{\boldsymbol{k}}. \tag{26}$$

On the contrary, a state is inconsistently $\{\mathcal{K}_j\}$-separable if it is a mixture of states with different entanglement properties,

$$\rho^{\{\mathcal{K}_j\}} = \sum_j p_j \bigotimes_{\boldsymbol{k} \in \mathcal{K}_j} \rho_j^{\boldsymbol{k}}, \tag{27}$$

so its entanglement properties are defined by a combination of constituent $\mathcal{K}_j$-separabilities. Precise classification methods are much more difficult for mixed states, however there are still some very useful approaches that can be introduced using NNS.

Consistently $\mathcal{K}$-separable states require a direct application of the separability conditions given by equation (24) onto the pure state of the NNS. Since the mixing state cannot capture quantum correlations,

it is already separable and requires no restrictions. It is thus expedient to apply the separability conditions of equation (24) onto the pure states of the mixed NNS, restricting the capacity of the neural network to simulate quantum correlations. Enforcing separability on the pure density-matrix in this way

$$\sigma_{\Xi|\mathcal{K}}^{\boldsymbol{\alpha},\boldsymbol{\beta}} = \prod_{l=1}^{K} e^{\omega_l(\boldsymbol{\alpha},\boldsymbol{\beta})} \prod_{j \in \boldsymbol{h}_l} \cosh(\theta_l^j(\boldsymbol{\alpha})) \cosh(\theta_l^{j*}(\boldsymbol{\beta})),$$

$$\omega_l(\boldsymbol{\alpha},\boldsymbol{\beta}) = \sum_{i \in \boldsymbol{k}_l} a_i \alpha_i + a_i^* \beta_i, \tag{28}$$

thus provides an NNS guaranteed to be consistently $\mathcal{K}$-separable

$$\rho_{\Pi,\Xi}^{\mathcal{K}} = \mathcal{P}_\Pi \odot \sigma_{\Xi|\mathcal{K}}. \tag{29}$$

If one wishes to enforce complete separability such that for an n-qudit state $\rho = \sum_j p_j \bigotimes_{m=1}^{n} \rho_j^m$, one can of course just apply consistent separability onto the network state via the separability set $\mathcal{K} = \{1|2|,\ldots,|n\}$ in an identical manner as before. However, as the state is completely separable, there are no quantum correlations and the pure states in the network ansatz are not necessary for simulation of the state. It can then be simplified to $\rho_\Pi = \mathcal{P}_\Pi$, and we can simulate completely separable mixed quantum systems using an RBM with a classical mixing-layer only [31]

$$\rho_\Pi^{\text{Sep}} = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} e^{\omega(\boldsymbol{\alpha},\boldsymbol{\beta})} \prod_{p=1}^{n_{\text{m}}} \cosh(\phi_p(\boldsymbol{\alpha},\boldsymbol{\beta})) |\boldsymbol{\alpha}\rangle\langle\boldsymbol{\beta}|. \tag{30}$$

Unfortunately, it is not possible to *strictly* classify an inconsistently separable mixed state according to ansatzes discussed in this section. Take the tripartite example

$$\rho = \sum_j p_j \rho_j^{\{1,2|3\}} + \sum_k p_k \rho_k^{\{1|2,3\}} + \sum_m p_m \rho_m^{\{1,3|2\}}, \tag{31}$$

which can be thought of as 'cheap' genuine tripartite entangled state. We can certainly define an NNS that can reconstruct a state of this form (trivially, one can utilize a fully connected NNS that can reconstruct ρ); however we cannot specify all three forms of separability in ρ without also allowing the NNS to potentially manifest genuine, pure tripartite entanglement. One can instead utilize independent consistently separable NNS according to the partitions $\{1,2|3\}$, $\{1,3|2\}$ and $\{2,3|1\}$ in order to quantify the amount of entanglement in the target state with respect to each partition.

3. Classifying and quantifying entanglement

3.1. Learning of quantum states
We present a learning protocol for a pure NNS $|\Psi_{\Pi,\Xi}\rangle$ to reconstruct a target state $|\varphi\rangle$ using the ansatz from equation (7), which is then extendible to mixed states. We employ a unified learning approach, where the variational state optimizes the global, vectorized fidelity with a target state, rather than separate phase and amplitude fidelities. We may define the loss function as the negative logarithmic fidelity between two pure states as a function of our set of variational parameters

$$\mathcal{L} = -\log \sqrt{\frac{|\langle \Psi_{\Pi,\Xi}|\varphi\rangle|^2}{\langle \Psi_{\Pi,\Xi}|\Psi_{\Pi,\Xi}\rangle\langle\varphi|\varphi\rangle}}. \tag{32}$$

Splitting these wavefunctions into respective phase and amplitude functions,

$$\Psi_{\Pi,\Xi}(\boldsymbol{s}) = \psi_\Pi(\boldsymbol{s}) e^{i \log(\phi_\Xi(\boldsymbol{s}))}, \qquad \varphi(\boldsymbol{s}) = \lambda(\boldsymbol{s}) e^{i \log(\xi(\boldsymbol{s}))}, \tag{33}$$

we wish to compute the derivatives of the unified cost function with respect to the parameter sets $\{\Pi,\Xi\}$. Since these wavefunctions utilize only real parameters, it is expedient to compute the derivatives using the following chain rule formulation,

$$\nabla_k^{\psi_\Pi} \mathcal{L} = \frac{\partial \mathcal{L}}{\partial |\psi_\Pi\rangle} \cdot \frac{\partial |\psi_\Pi\rangle}{\partial \Pi_k}, \qquad \nabla_k^{\phi_\Xi} \mathcal{L} = \frac{\partial \mathcal{L}}{\partial |\phi_\Xi\rangle} \cdot \frac{\partial |\phi_\Xi\rangle}{\partial \Xi_k}. \tag{34}$$

Computing these gradients will provide the necessary parameter update rules at the mth iteration to the kth network parameter by gradient descent, taking the form

$$\Pi_k^{m+1} = \Pi_k^m - \eta \nabla_k^{\psi_\Pi} \mathcal{L}, \qquad \Xi_k^{m+1} = \Xi_k^m - \eta \nabla_k^{\phi_\Xi} \mathcal{L}, \tag{35}$$

where η is some learning rate small enough such that the network state converges to the target state over sufficient iterations of the learning scheme.

Defining the quantity

$$\Delta(s) = \langle \Psi_{\Pi,\Xi} | \varphi \rangle^{-1} \, e^{i \log \frac{\phi_\Xi(s)}{\xi(s)}}, \tag{36}$$

complete gradients with respect to variational parameters can therefore be computed as

$$\nabla_k^{\psi_\Pi} \mathcal{L} = \sum_s \left[\frac{\psi_\Pi(s)}{|\Psi_{\Pi,\Xi}|^2} - \lambda(s) \mathrm{Re}\left[\Delta(s)\right] \right] \mathcal{O}_k^\Pi |\psi_\Pi\rangle, \tag{37}$$

$$\nabla_k^{\phi_\Xi} \mathcal{L} = -\sum_s \left[\frac{\lambda(s)\psi_\Pi(s)}{\phi_\Xi(s)} \, \mathrm{Im}\left[\Delta(s)\right] \right] \mathcal{O}_k^\Xi |\phi_\Xi\rangle, \tag{38}$$

where $\mathcal{O}_k^\Pi = \mathrm{diag}\left(\partial_{\Pi_k} \log |\psi_\Pi\rangle\right)$, $\mathcal{O}_k^\Xi = \mathrm{diag}\left(\partial_{\Xi_k} \log |\phi_\Xi\rangle\right)$ denote diagonal matrices containing the logarithmic derivatives of the network state with respect to the kth amplitude and phase network parameters respectively. Utilizing equation (38) in the update rule given by equation (35), the phase and amplitude properties will optimize in a unified manner, maximizing the fidelity between the network and the target state endowed with non-trivial phase structure.

Fortunately this learning procedure is readily extended to mixed states via the ansatz in equation (20). Since the variational state is in a complex exponential format, one then formulates a cost function based on the fidelity between the vectorized density-matrix and the vectorized target state. The extension is straightforward and explained in appendix B.

As shown in reference [28] separable neural network states can be used to perform entanglement classification and provide entanglement measures of pure, two-dimensional quantum states. Using qudit sub-encoding and the mixed state architectures discussed in the previous sections, these ideas can be extended to classification of more complex quantum systems.

Let us devise a precise decision rule for classification. Consider a target n-qudit state σ, a $\mathcal{K}$-separable learner $\rho_\Omega^\mathcal{K}$, and a free, entangled learner $\rho_\Omega^{\mathrm{Ent}}$ which have both been optimized with respect to reconstructing σ. Using the Bures fidelity, $F(\sigma,\rho) = \mathrm{Tr}\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}$, we denote the *reconstruction fidelity* of a learning process as the final/optimal fidelity achieved after a given number of learning iterations. A target σ is learnable via $\rho_\Omega^{\mathrm{Ent}}$ iff its reconstruction fidelity satisfies

$$F(\sigma, \rho_\Omega^{\mathrm{Ent}}) \geqslant F_{\mathrm{opt}} = 1 - \epsilon, \tag{39}$$

for a sufficiently small threshold ϵ. The choice of F_{opt} determines the reliability of classification, and in our numerical experiments we fix $\epsilon \leqslant 10^{-4}$. The accuracy of this reconstruction via free learning also benchmarks the satisfactory computational resources required in the network, informing the separable reconstruction.

One can reliably infer that a target state is $\mathcal{K}$-separable if it is learnable by both a free NNS ($\rho_\Omega^{\mathrm{Ent}}$), and a $\mathcal{K}$-separable NNS ($\rho_\Omega^\mathcal{K}$). Then the NNS reconstruction fidelities must satisfy

$$F(\sigma, \rho_\Omega^\mathcal{K}) \geqslant F(\sigma, \rho_\Omega^{\mathrm{Ent}}) \geqslant F_{\mathrm{opt}}. \tag{40}$$

Otherwise, the state is entangled to a higher degree. One may then quantify the entanglement content of the target by investigating the distance between σ and an approximation to the closest $\mathcal{K}$-separable state.

3.2. Quantifying entanglement

The most difficult aspect of quantifying entanglement stems from the complicated nature of characterising the space of separable quantum states. Thanks to the implicit guarantee of specific separability, SNNS offer an extremely useful tool to help with this, and provide the opportunity to study a variety of entanglement measures that are otherwise much too difficult to explore.

Let us consider measures E that satisfy the general properties of a valid entanglement measure [4]. Many important types of E are constructed as a geometric optimization problem with respect to the space of all fully separable states $\mathcal{D}_{\mathrm{Sep}}$. That is, given a target state σ and a distance measure (possibly quasi-distance measure) f,

$$E(\sigma) = \min_{\rho \in \mathcal{D}_{\mathrm{Sep}}} f(\sigma, \rho), \tag{41}$$

$$\text{if } \sigma \in \mathcal{D}_{\mathrm{Sep}} \implies E(\sigma) = 0, \tag{42}$$

$$\text{if } \sigma \notin \mathcal{D}_{\mathrm{Sep}} \implies E(\sigma) > 0. \tag{43}$$

These are entanglement measures which are computed by locating the closest separable state (CSS) $\sigma^\star$ to σ, with respect to the distance measure f. For such measures, the employment of SNNS to parameterize the separable states $\rho_\Omega \in \mathcal{D}_{\text{Sep}}$ is extremely useful, as it offers an efficient way to perform this optimization. Furthermore, since SNNS are inherently separable, they will always approximate an upper bound on E, since they are certifiably limited in the quantum correlations that they are able to simulate. This is,

$$E(\sigma) \leqslant E_\Omega(\sigma) = \min_{\rho_\Omega \in \mathcal{D}_{\text{Sep}}} f(\sigma, \rho_\Omega). \tag{44}$$

To generalize, we may construct a measure $E^\mathcal{K}$ which is analogous to E, but is defined with respect to the space of all states which are at most $\mathcal{K}$-separable. Defining the set of all states that are $\mathcal{K}$-separable as $\mathcal{D}_\mathcal{K}$, then the set of all states that are at most $\mathcal{K}$-separable is given by [32]

$$\tilde{\mathcal{D}}_\mathcal{K} = \mathcal{D}_\mathcal{K} \bigcup_{|\mathcal{K}'|>|\mathcal{K}|} \mathcal{D}_{\mathcal{K}'}. \tag{45}$$

Assuming a measure of the form equation (41), then we can define

$$E^\mathcal{K}(\sigma) = \min_{\rho \in \tilde{\mathcal{D}}_\mathcal{K}} f(\sigma, \rho) \leqslant E_\Omega^\mathcal{K}(\sigma), \tag{46}$$

$$\text{if } \sigma \in \tilde{\mathcal{D}}_\mathcal{K} \implies E^\mathcal{K}(\sigma) = 0, \tag{47}$$

$$\text{if } \sigma \notin \tilde{\mathcal{D}}_\mathcal{K} \implies E^\mathcal{K}(\sigma) > 0. \tag{48}$$

$E^\mathcal{K}$ satisfies all the general properties of an entanglement measure, but now with respect to $\tilde{\mathcal{D}}_\mathcal{K}$, and is therefore able to classify/quantify more complex forms of entanglement.

Let us specify some important entanglement measures which SNNS can utilize, starting from the geometric measure of entanglement (GME) [33]. For pure states, the GME is the maximum fidelity that can be obtained between a target state $|\sigma\rangle$ and the set of pure, at most $\mathcal{K}$-separable states $\tilde{\mathcal{B}}_\mathcal{K}$

$$E_{\text{G}}(|\sigma\rangle) = \max_{|\varphi\rangle \in \tilde{\mathcal{B}}_\mathcal{K}} F(|\sigma\rangle, |\varphi\rangle). \tag{49}$$

For more sophisticated mixed state approaches, it is expedient to employ any number of density-matrix distance measures. Several important examples include the trace distance

$$E_{\text{C}_1}(\sigma) = \frac{1}{2} \min_{\rho \in \mathcal{D}_{\text{Sep}}} \|\sigma - \rho\|_1, \tag{50}$$

where $\|X\|_1 = \text{Tr}\sqrt{X^\dagger X}$ or the Bures metric

$$E_{\text{B}}(\sigma) = \min_{\sigma \in \mathcal{D}_{\text{Sep}}} \left[1 - F^2(\rho, \sigma)\right], \tag{51}$$

where F is the Bures fidelity as before. These quantities are readily approximated via SNNS, and easily specified to different forms of $\mathcal{K}$-separability.

Of particular interest is the REE [29], an entanglement measure that has many applications in quantum communications and channel capacities [34]. The REE is based on the quantum relative entropy (QRE), a kind of distance measure between two quantum states where
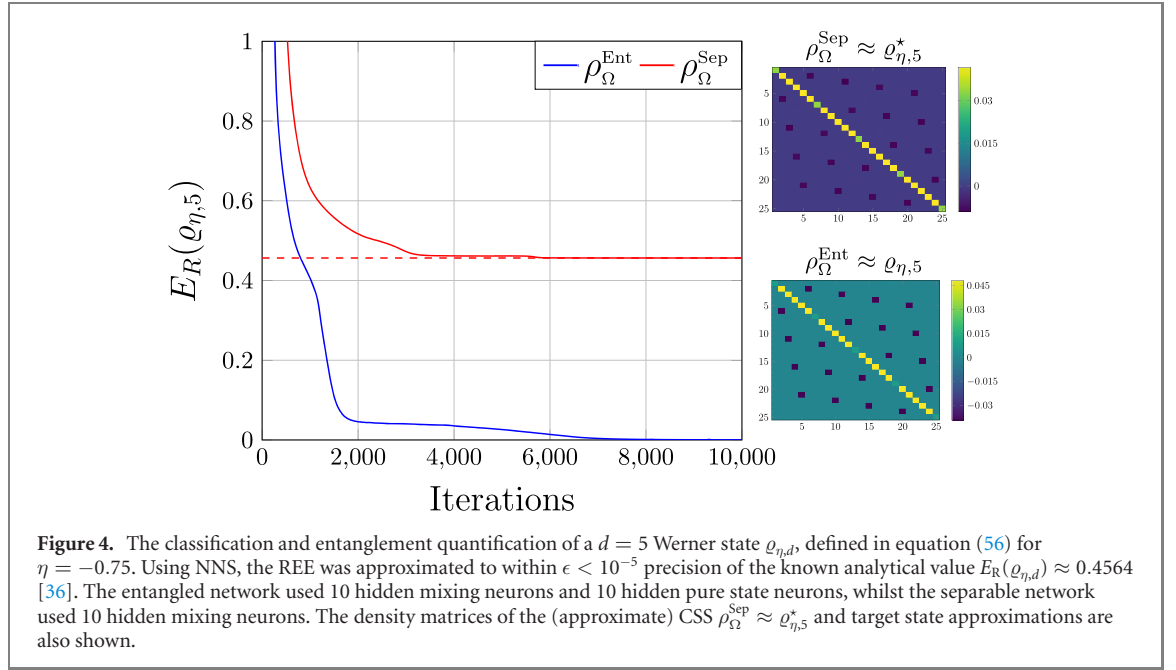
$$S(\rho\|\sigma) = \text{Tr}\left[\rho\left(\log\rho - \log\sigma\right)\right], \tag{52}$$

such that $S(\rho\|\sigma) \in [0, +\infty)$. Due to its asymmetry and the fact that it is infinite on pure states, it is not a true metric. However, the QRE is an important distinguishability measure between quantum states which provides access to important entropic quantities such as the Shannon entropy. Minimizing the relative entropy with respect to the set of all separable quantum states results in the REE

$$E_{\text{R}}(\rho) = \min_{\sigma \in \mathcal{D}_{\text{Sep}}} S(\rho\|\sigma), \tag{53}$$

which can be readily employed with respect to parameterized NNS. This can of course generalize to $E_{\text{R}}^\mathcal{K}(\sigma)$ given a form of separability. Interestingly, the REE is sub-additive and in general

$$E_{\text{R}}(\rho \otimes \sigma) \leqslant E_{\text{R}}(\rho) + E_{\text{R}}(\sigma). \tag{54}$$

Figure 4. The classification and entanglement quantification of a $d = 5$ Werner state $\varrho_{\eta,d}$, defined in equation (56) for $\eta = -0.75$. Using NNS, the REE was approximated to within $\epsilon < 10^{-5}$ precision of the known analytical value $E_R(\varrho_{\eta,d}) \approx 0.4564$ [36]. The entangled network used 10 hidden mixing neurons and 10 hidden pure state neurons, whilst the separable network used 10 hidden mixing neurons. The density matrices of the (approximate) CSS $\rho_\Omega^{\text{Sep}} \approx \varrho_{\eta,5}^\star$ and target state approximations are also shown.

This lets us define a regularized n-shot REE

$$E_R^n(\rho) = \frac{1}{n} \min_{\sigma \in \mathcal{D}_{\text{Sep}}} S(\rho^{\otimes n} \| \sigma) \leqslant E_R(\rho). \tag{55}$$

The single-shot, standard REE alone is an extremely difficult quantity to compute, largely due to the characterization of $\mathcal{D}_{\text{Sep}}$ and the unruliness of the QRE. Its computation has recently been explored using an active learning strategy [35], in which the authors use active learning to compress $\mathcal{D}_{\text{Sep}}$ into a more relevant subset of the separable state space that contributes strongly to the REE. Thanks to the implicit separability of NNS, we may choose an alternative approach where it is possible to optimise some other cost function such as fidelity/trace distance that will simultaneously minimise the QRE towards the optimal REE. In doing so, SNNS should allow for the accurate and efficient approximation of E_R, and previously unexplored REEs with respect to other forms of separability $E_R^{\mathcal{K}}$.

4. Applications and results

4.1. Mixed states in *d*-dimensions

The most substantial generalisation of the methods introduced in reference [28] is the ability to classify and quantify entanglement in mixed, d-dimensional states. To illustrate this improvement, consider the d-dimensional Werner state, parameterized by
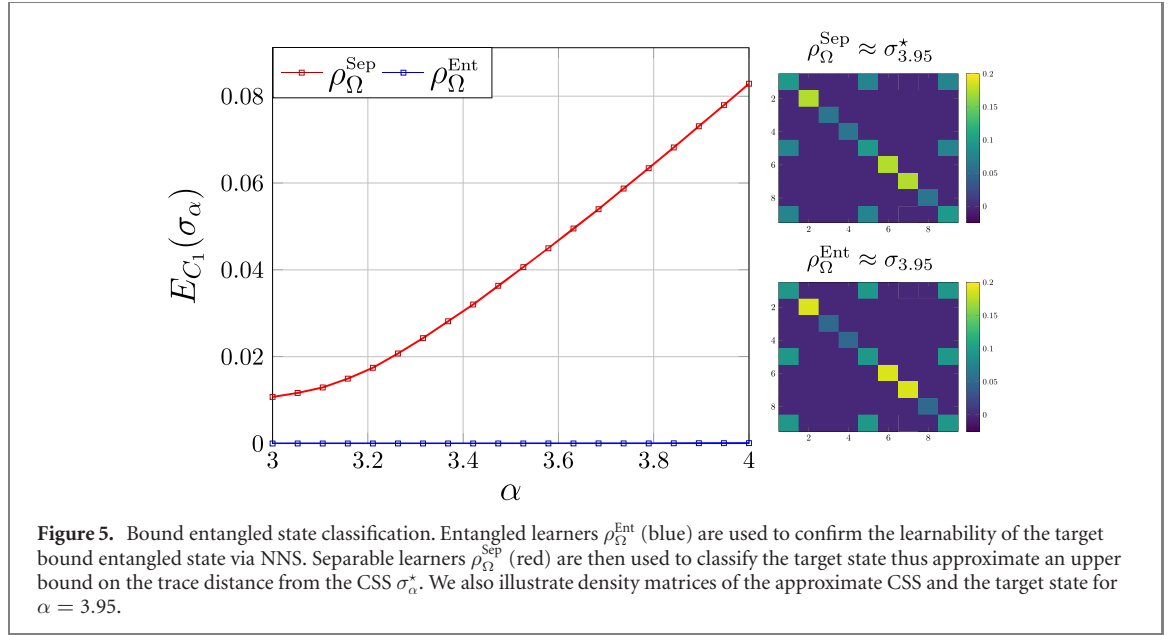
$$\varrho_{\eta,d} = \frac{(d - \eta)\mathbb{I}_d^{\otimes 2} + (d\eta - 1)\mathbb{F}_d}{d(d^2 - 1)}, \tag{56}$$

where $\mathbb{F}_d = \sum_{i,j=0}^{d-1} |ij\rangle\langle ji|$ is the two-qudit flip operator, $\mathbb{I}_d$ is the d-dimensional identity operator, and η characterizes the entanglement properties of the state. For $\eta \in [-1, 0]$ the state is entangled, and we can easily quantify this entanglement using the analytically known REE [36],

$$E_R(\varrho_{\eta,d}) = \frac{1 + \eta}{2} \log_2(1 + \eta) + \frac{1 - \eta}{2} \log_2(1 - \eta). \tag{57}$$

In figure 4 we display an optimization procedure for $d = 5, \eta = -0.75$ using an entangled learner ρ_Ω^{Ent} and a fully separable learner ρ_Ω^{Sep}. The free, entangled learner is able to reconstruct the target Werner state with ease, and an extremely high fidelity, while the fully separable learner correctly classifies the target as entangled.

Beyond the obvious entanglement classification, the SNNS is able to quantify the REE of the state, by monitoring the relative entropy $E_R^\Omega(\varrho_{\eta,d}) = S(\varrho_{\eta,d} \| \rho_\Omega^{\text{Sep}})$ throughout the learning process. As the optimization converges, $E_R^\Omega \to E_R$, we gather an approximation to the REE of the state. Indeed, under typical optimization settings, the REE is approximated to within $\epsilon < 10^{-5}$ precision of the known analytical value $E_R(\varrho_{-0.75,5}) \approx 0.4564$, reinforcing the strength of this approach.

Figure 5. Bound entangled state classification. Entangled learners ρ_Ω^{Ent} (blue) are used to confirm the learnability of the target bound entangled state via NNS. Separable learners ρ_Ω^{Sep} (red) are then used to classify the target state thus approximate an upper bound on the trace distance from the CSS $\sigma_\alpha^\star$. We also illustrate density matrices of the approximate CSS and the target state for $\alpha = 3.95$.

4.2. Classification of bound entangled states

The positivity of a partially transposed quantum system *can* be a signature of separability. However it is not universal, and there exist classes of states which are PPT but are entangled, known as bound entangled (BE) states. Here we consider the following two-qutrit state,

$$\sigma_+ = -\frac{1}{3}(|01\rangle\langle01| + |12\rangle\langle12| + |20\rangle\langle20|),$$

$$\sigma_- = \frac{1}{3}(|10\rangle\langle10| + |21\rangle\langle21| + |02\rangle\langle02|),$$

$$\sigma_\alpha = \frac{2}{7}|\Phi^+\rangle\langle\Phi^+| + \frac{\alpha}{7}\sigma_+ + \frac{5-\alpha}{7}\sigma_-, \tag{58}$$

where $|\Phi^+\rangle = \frac{1}{\sqrt{3}}(|00\rangle + |11\rangle + |22\rangle)$ is a $d = 3$-dimensional Bell state. This state is known to satisfy the following entanglement properties [37]:

$$\sigma_\alpha \text{ is } \begin{cases} \text{Separable} & \text{if } 2 \leqslant \alpha \leqslant 3, \\ \text{Bound entangled} & \text{if } 3 < \alpha \leqslant 4, \\ \text{Free entangled} & \text{if } 4 < \alpha \leqslant 5. \end{cases} \tag{59}$$
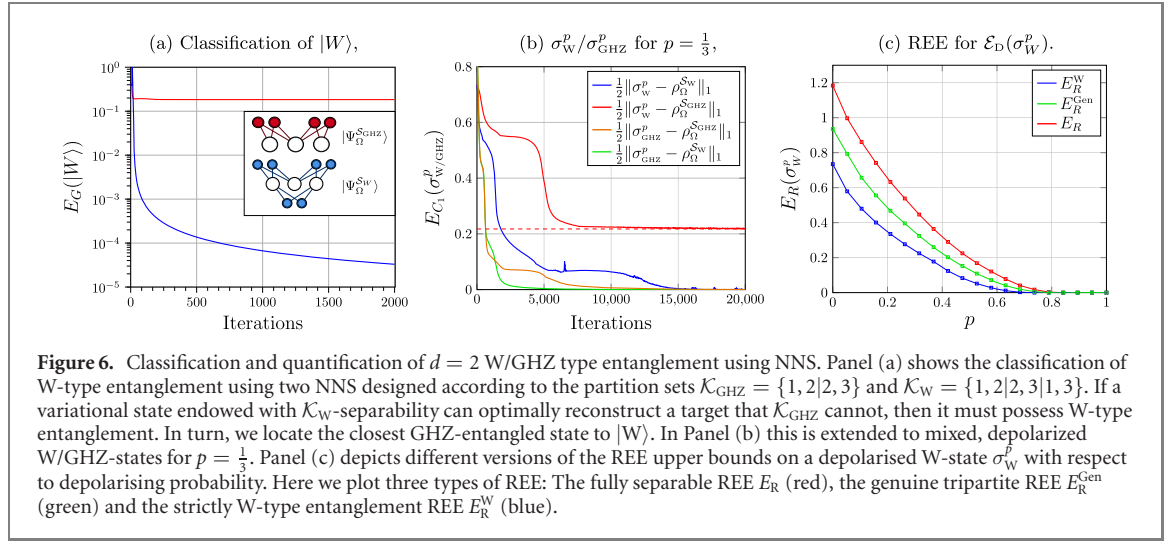
Here we investigate the target state in the BE region, and show that this bipartite state cannot be optimally reconstructed via SNNS. Figure 5 depicts the employment of entangled learners ρ_Ω^{Ent} (blue), and fully separable learners ρ_Ω^{Sep} (red) to reconstruct σ_α across the domain $3 < \alpha \leqslant 4$.

For all values of α, ρ_Ω^{Ent} is able to reconstruct the state to a high degree of precision such that the trace distance is $\|\sigma_\alpha - \rho_\Omega^{\text{Ent}}\|_1 \leqslant 10^{-4}$. However, the separable learners are unable to reach this level of reconstruction accuracy. Hence, since σ_α are learnable via free NNS, the inability of ρ_Ω^{Sep} to reconstruct σ_α implies that these states are entangled in this region. Since they are also PPT in this region, we have successfully shown the ability of SNNS to classify bound entanglement.

During each constrained optimization we gather an upper bound on the distance between the target BE state, and its CSS. As said before, this is an upper bound since ρ_Ω^{Sep} offers an approximation to the CSS, and is potentially loose. Nonetheless the inferred classification is informative. Figure 5 plots the trace distance $\|\sigma_\alpha - \rho_\Omega^{\text{Sep}}\|_1$, shown to steadily rise as α increases, which is expected as σ_α becomes freely entangled for $4 < \alpha \leqslant 5$.

4.3. Detection and measurement of multipartite entanglement

The versatility of the $\mathcal{K}$-separable state design means that we can explore entanglement classification and quantification methods that are otherwise very difficult. In particular, we may construct an NNS protocol that is able to witness W/GHZ-state entanglement, and measure W/GHZ-type correlations in both pure

Figure 6. Classification and quantification of $d = 2$ W/GHZ type entanglement using NNS. Panel (a) shows the classification of W-type entanglement using two NNS designed according to the partition sets $\mathcal{K}_{\text{GHZ}} = \{1, 2|2, 3\}$ and $\mathcal{K}_{\text{W}} = \{1, 2|2, 3|1, 3\}$. If a variational state endowed with $\mathcal{K}_{\text{W}}$-separability can optimally reconstruct a target that $\mathcal{K}_{\text{GHZ}}$ cannot, then it must possess W-type entanglement. In turn, we locate the closest GHZ-entangled state to $|W\rangle$. In Panel (b) this is extended to mixed, depolarized W/GHZ-states for $p = \frac{1}{3}$. Panel (c) depicts different versions of the REE upper bounds on a depolarised W-state σ_W^p with respect to depolarising probability. Here we plot three types of REE: The fully separable REE E_R (red), the genuine tripartite REE E_R^{Gen} (green) and the strictly W-type entanglement REE E_R^{W} (blue).

and mixed quantum states. Consider the three-qubit W and GHZ states respectively [38, 39]

$$|\text{W}\rangle = \frac{1}{\sqrt{3}} \left(|001\rangle + |010\rangle + |100\rangle \right),$$

$$|\text{GHZ}\rangle = \frac{1}{\sqrt{2}} \left(|000\rangle + |111\rangle \right).$$

These are both maximally entangled three party states. However they possess two inequivalent forms of tripartite entanglement, such that $|\text{W}\rangle$ cannot be transformed into $|\text{GHZ}\rangle$ by means of LOCC (local operations and classical communications) strategies. The key difference in these forms of entanglement is their *robustness* i.e. when a party is removed from a GHZ state the remaining states are separable, whilst a W-state remains entangled. Therefore a W-state possesses strict bipartite entanglement between all three parties, whereas GHZ entanglement can be achieved via 'relayed entanglement' [40].

To classify between these states, we must define a partition set that is capable of capturing GHZ correlations, but incompletely capture W-type correlations. The non-disjoint separability set

$$\mathcal{K}_{\text{W}} = \{1, 2|2, 3|1, 3\}, \tag{60}$$

is capable of learning both W and GHZ entangled states, as it strictly specifies bipartite entanglement between all parties. However, one can construct the partition set

$$\mathcal{K}_{\text{GHZ}} = \{i, j|i, k\}, \quad i \neq j \neq k \in \{1, 2, 3\}, \tag{61}$$

which is any possible permutation of two subsets of $\mathcal{K}_{\text{W}}$. Programming an NNS according to $\mathcal{K}_{\text{GHZ}}$ does not allow the network to capture direct correlations between qubits j and k, and will therefore provide an insufficient ansatz to reconstruct W-states. This forms a witness for W-type entanglement; if a target state is learnable via an NNS endowed with $\mathcal{K}_{\text{W}}$-separability, but is not learnable via $\mathcal{K}_{\text{GHZ}}$-separability, then the state is verified as possessing W-type entanglement. Furthermore, by constructing entanglement measures $E_\Omega^{\mathcal{K}_{\text{GHZ}}}$ we are able to measure the amount of W-type correlations within a target state.

Figure 6(a) shows the pure state classification of a three-qubit W-state, where the non-disjoint network architectures perform classification easily. Note that these three-qubit partitions can be analogously embedded into larger, n-qudit systems in order to study more complex forms of entanglement.

Realistically, multipartite entangled resources for future quantum communication/computing protocols will be noisy and imperfect. Generating and distributing multipartite entanglement over noisy quantum channels is fundamental for many future quantum technologies, particularly for secure communications and quantum networks [41–48]. Therefore it is a more interesting challenge to consider the classification and quantification of tripartite entanglement subject to decoherence. For instance, one can consider versions of $|W\rangle/|\text{GHZ}\rangle$ in which each qudit has been passed through a depolarizing channel

$$\mathcal{E}_{\text{D}}(\rho) = (1 - p)\rho + \frac{p}{d^n} \mathbb{I}_d^{\otimes n}, \tag{62}$$

where n denotes the number of qudits being acted on (in this case $n = 3$). We denote these noisy, three-qubit states as

$$\sigma_{\mathrm{W}}^{p} = (1-p)|\mathrm{W}\rangle\langle\mathrm{W}| + \frac{p}{8}\mathbb{I}_2^{\otimes 3}, \tag{63}$$

$$\sigma_{\mathrm{GHZ}}^{p} = (1-p)|\mathrm{GHZ}\rangle\langle\mathrm{GHZ}| + \frac{p}{8}\mathbb{I}_2^{\otimes 3}. \tag{64}$$

Using mixed NNS programmed with different separabilities, we may then easily distinguish between the entanglement properties of noisy W/GHZ-states subject to depolarizing channels. Indeed, figure 6(b) shows that for $p = \frac{1}{3}$ we can perform this classification. Given two learners $\rho_{\Omega}^{\mathcal{K}_{\mathrm{W}}}$ and $\rho_{\Omega}^{\mathcal{K}_{\mathrm{GHZ}}}$, it is clear that both are able to optimally reconstruct the noisy GHZ-state, whilst only $\rho_{\Omega}^{\mathcal{K}_{\mathrm{W}}}$ is able to optimally reconstruct the noisy W-state, completing the classification.

This is taken a step further in figure 6(c) where different versions of the REE of σ_{W}^{p} is monitored for various depolarizing probabilities. This plot describes three forms of REE:

- The standard E_{R} (red) defined on the space of all fully separable states (using the partition set $\mathcal{K}_{\mathrm{FS}} = \{1|2|3\}$) which measures the amount of any entanglement present.
- The genuine tripartite entangled REE, $E_{\mathrm{R}}^{\mathrm{Gen}}$ (green), using the bi-separable partition sets $\mathcal{K}_{\mathrm{BS}} = \{i,j|k\}, i \neq j \neq k \in \{1,2,3\}$, which measures the amount of genuine tripartite entanglement in the state (W or GHZ correlations).
- The W-REE, $E_{\mathrm{R}}^{\mathrm{W}}$ (blue) using the partition set $\mathcal{K}_{\mathrm{GHZ}}$ in equation (61), which measures the amount of genuine, tripartite, strictly W-type entanglement within the state.

By employing more complex separable architectures, we may study how different forms of entanglement behave with respect to environmental properties, such as depolarization. By measuring $E_{\mathrm{R}}^{\mathrm{Gen}}$ and $E_{\mathrm{R}}^{\mathrm{W}}$ for instance, we may monitor the decoherence of genuine tripartite entanglement, rather than *any* entanglement as done so by E_{R}. Such characterizations could prove very useful in communication/networking scenarios, where genuine multipartite entanglement is critical to performance.

It is important to remind the reader that these are upper bounds. The standard REE upper bound is expected to be tight, as fully separable NNS architectures precisely capture full separability. However, $\mathcal{K}_{\mathrm{BS}}$ and $\mathcal{K}_{\mathrm{GHZ}}$ are degenerate, e.g. $\mathcal{K}_{\mathrm{BS}} = \{i,j|k\}$ has 3 unique forms. Since mixed SNNS are restricted to consistent separabilities, there may be convex combinations of states of these separabilities that produce tighter bounds. It is unknown if this is the case, nonetheless $E_{\mathrm{R}}^{\mathrm{Gen}}$ and $E_{\mathrm{R}}^{\mathrm{W}}$ provide informative upper bounds on these unique entanglement measures.

4.4. Ultimate limits for channel capacities

We may provide a more practical example for the use of SNNS in the realm of quantum communications, using them to approximate upper bounds of quantum channel capacities. Introduced in reference [34], the Pirandola–Laurenza–Ottaviani–Banchi (PLOB) bound is an ultimate upper bound on the two-way assisted quantum (and secret-key) capacity $\mathcal{C}(\mathcal{E})$ for a given quantum channel $\mathcal{E}$. Its derivation is based on the techniques of channel simulation and teleportation stretching, which have proven to be extremely versatile in a number of settings [42, 49–53]. An essential class of quantum channels are those which are teleportation covariant, meaning that they satisfy the condition

$$\mathcal{E}(U\rho U^{\dagger}) = V\mathcal{E}(\rho)V^{\dagger}, \tag{65}$$

for some pair of teleportation unitaries $\{U, V\}$. Let us define the Choi matrix of a d-dimensional channel $\mathcal{E}$ as the result of passing one mode of a maximally entangled state Φ^{+} through the $\mathcal{E}$, and the other through an identity channel $\mathcal{I}$
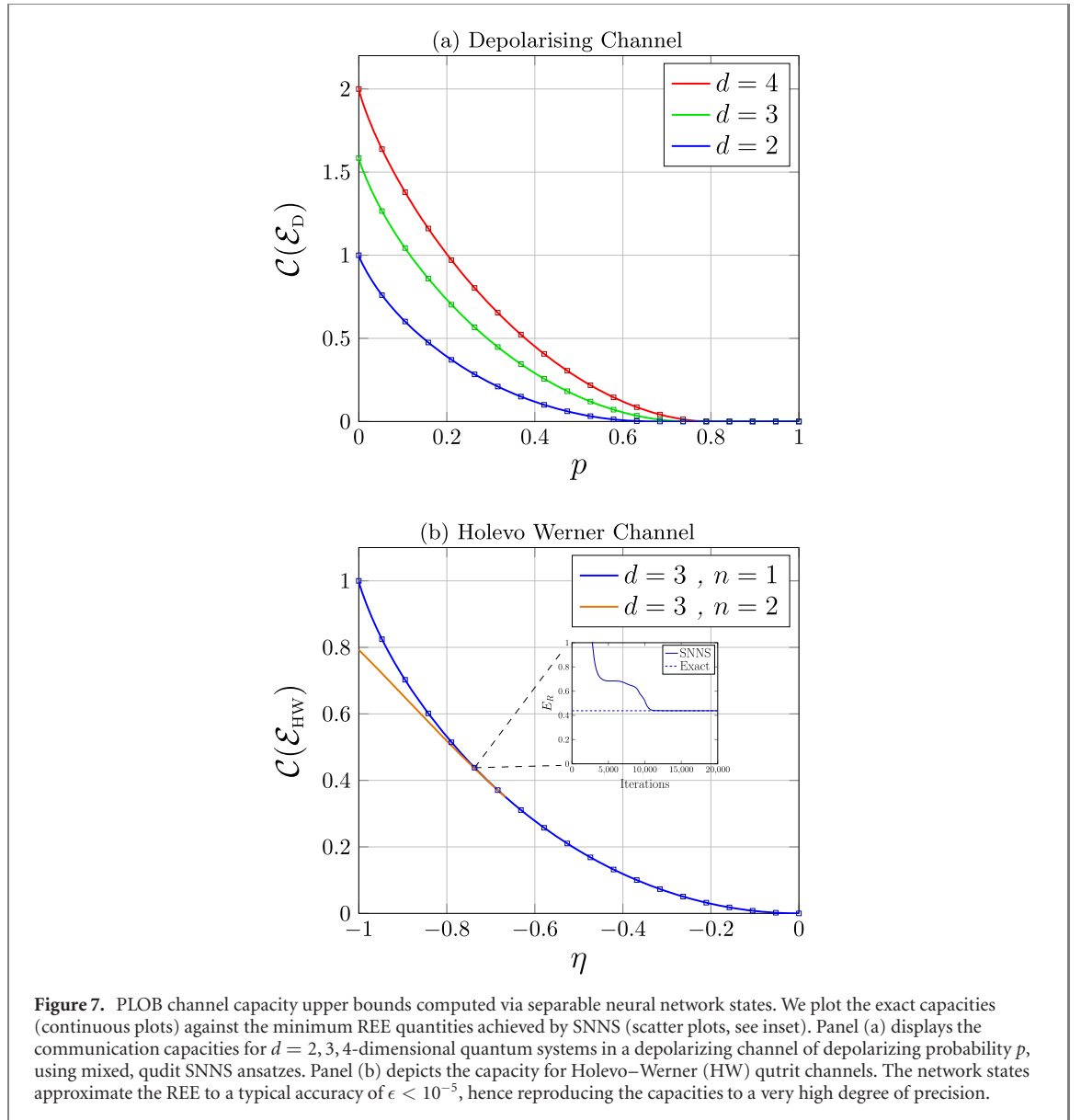
$$\rho_{\mathcal{E}} = \mathcal{I} \otimes \mathcal{E}[\Phi^{+}], \tag{66}$$

where the maximally entangled state may take the form $\Phi^{+} = \frac{1}{d}\sum_{i,j=0}^{d-1}|ii\rangle\langle jj|$. For teleportation covariant channels, the ultimate channel capacity can then be upper bounded in a remarkably simple way [34]

$$\mathcal{C}(\mathcal{E}) \leqslant E_{\mathrm{R}}^{n}(\rho_{\mathcal{E}}) \leqslant E_{\mathrm{R}}(\rho_{\mathcal{E}}), \tag{67}$$

where E_{R} is the standard REE (and E_{R}^{n} its n-shot version). SNNS can be used to approximate upper bounds on these channel capacities, via constrained reconstruction of the Choi state of the desired quantum channel.

We consider two important, teleportation covariant, d-dimensional quantum channels in an effort to illustrate the effectiveness of our approach: the depolarizing channel considered in equation (62), and the HW channel [54–56]. The Choi states of these channels are the classes of isotropic states and Werner states

Figure 7. PLOB channel capacity upper bounds computed via separable neural network states. We plot the exact capacities (continuous plots) against the minimum REE quantities achieved by SNNS (scatter plots). Panel (a) displays the communication capacities for $d = 2, 3, 4$-dimensional quantum systems in a depolarizing channel of depolarizing probability p, using mixed, qudit SNNS ansatzes. Panel (b) depicts the capacity for Holevo–Werner (HW) qutrit channels. The network states approximate the REE to a typical accuracy of $\epsilon < 10^{-5}$, hence reproducing the capacities to a very high degree of precision.

respectively, whose REE bounds are known analytically. Therefore, we can compare the numerical performance of computing the REE via SNNS with the known, exact bounds.

Figure 7(a) reports REE bounds on the capacity of depolarising channels for dimensions $d = 2, 3, 4$. Approximating these bounds via separable network states requires the targeted reconstruction of the isotropic state,

$$\rho_{\mathcal{E}_D} = (1 - p)\Phi^+ + \frac{p}{d^2}\mathbb{I}_d^{\otimes 2}. \tag{68}$$

Using a bipartite SNNS ρ_Ω^{Sep}, and attempting to learn the target Choi state leads to an approximation of the REE of said state. Performing this optimization for many depolarizing probabilities p, the results in figure 7(a) can be produced. This is be achieved to a very high degree of accuracy, reproducing the analytical bounds with an average error $\sim\epsilon < 10^{-5}$. Furthermore, these bounds can be computed very efficiently by performing each optimization sequentially, initializing the network parameters using the results of previous optimizations (see appendix C).

In figure 7(b) we give REE upper bounds for the HW channel, which takes the form

$$\mathcal{E}_{\text{HW}}^{\eta,d}(\rho) = \frac{(d - \eta)\mathbb{I}_d^{\otimes 2} + (d\eta - 1)\rho^{\text{T}}}{d^2 - 1}, \tag{69}$$

such that T superscript denotes the transposition. The Choi state of the HW channel is the d-dimensional Werner state, introduced in equation (56). The single shot REE bounds for the HW channel are analytically

known and given in equation (57), and are independent of dimension d. Again, this single shot bound is approximated to a good precision, as shown in the results.

For Werner states of dimension $d > 2$, their REE is known to be strictly sub-additive when $\eta < -\frac{d}{2}$, and previous studies have explored the two-shot REE for these Choi states [55], which can therefore be used to tighten these upper bounds. For instance, in figure 7(b) the two-shot capacity can be seen to significantly tighten the bounds for $d = 3$. In order to compute these tighter bounds, one must modify the definition of the n-shot quantities slightly. Now the minimization is performed with respect to the space of all locally bi-separable states. Consider the n-copy Werner state, and let us label each copy with indices of its modes $\{i,j\}$,

$$\varrho_{\eta,d}^{\otimes n} = \varrho_{\eta,d}^{\{1,2\}} \otimes \varrho_{\eta,d}^{\{3,4\}} \otimes \ldots \otimes \varrho_{\eta,d}^{\{2n-1,2n\}}. \tag{70}$$

The goal is now to find the CSS that possesses the following bi-separability

$$\sigma^n = \sigma_a^{\{1,3,5,\ldots,2n-1\}} \otimes \sigma_b^{\{2,4,6,\ldots,2n\}}, \tag{71}$$

where we have permuted the labels into a bi-separable decomposition such that each state belongs to exclusively even or odd mode labels. This corresponds to a situation where two users each possess n local modes, and their goal is to produce the closest state to $\varrho_{\eta,d}^{\otimes n}$ that is bi-separable between them. In general this is a very difficult task, and while beyond the scope of this paper, poses as an interesting future application for SNNS.

5. Conclusions and outlook

We have generalized the concept of NNS with programmable separability to mixed, d-dimensional quantum states. We discussed a number of neural network architectures for the description of quantum states, and detailed how their entanglement properties may be controlled via constraints placed on network connectivity. It was shown that network connectivity controls entanglement structure on a very specific level, requiring distinctions between certain forms of entanglement. Outlining one of many possible optimisation protocols, methods of classification and quantification via SNNS have been logically developed, and applied in a number of important settings. We then studied a practical application of these tools in the bounding of ultimate quantum channel capacities, showing that they can reproduce the PLOB bounds for DV channels with high precision.

There are a number of valuable future directions in which SNNS may be explored and expanded. While an optimization scheme based on the vectorized fidelity is effective for a variety of applications (as shown in this work) more sophisticated optimization protocols could enhance performance for more specific entanglement measures. In particular, a gradient descent method that directly minimizes the relative entropy (or some variant thereof) would provide a more effective computation of the REE for complex states. This would also lend well to the study of n-shot REE quantities with applications in quantum channel capacities, and the characterization of more complex BE states (such as those constructed from un-extendible product bases). Combining these tools with those from practical quantum tomography could also be extremely useful, e.g. where SNNS may be used to certify the effectiveness an entanglement distribution protocol.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Appendix A. Neural network mixed state ansatz

We briefly review the construction of the mixed NNS ansatz (see reference [19–21] for more detailed derivations) to illustrate the emergence of the classical mixing state and pure state ansatzes. A generic density-matrix element with respect to row and column basis vectors $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$ can be expressed as

$$\rho^{\boldsymbol{\alpha},\boldsymbol{\beta}} = \sum_n p_n \phi_n(\boldsymbol{\alpha}) \phi_n^*(\boldsymbol{\beta}), \tag{A1}$$

where p_n is the classical probability of a pure state ϕ_n existing within ensemble, and the sum $\sum_n$ may run over many contributing states.

We can use NNS in order to translate this expression into a variational ansatz. As stated in the main text, the inherent advantage to a pure NNS is that its output is independent from the activations of the hidden layer $\boldsymbol{h} \in \{-1, 1\}^{n_h}$, which consists of n_h neurons. Prior to tracing out this hidden layer, a pure NNS wavefunction is given by,

$$\Psi_{\Pi}(\boldsymbol{s}) = \sum_{\boldsymbol{h}} \exp\left(\sum_{k=1}^{n_v} a_k s_k + \sum_{j=1}^{n_h} b_j h_h + \sum_{k,j=1}^{n_v, n_h} W_{kj} h_j s_k \right). \tag{A2}$$

Using this NNS wavefunction, it is then easy to construct a pure density-matrix, such that $\sigma^{\boldsymbol{\alpha},\boldsymbol{\beta}} = \Psi_{\Pi}(\boldsymbol{\alpha}) \Psi_{\Pi}^*(\boldsymbol{\beta})$, using two visible layers in order to encode density-matrix entries, as shown in figure (2).

In order to construct the mixed state ansatz, we introduce an additional mixing layer $\boldsymbol{m} \in \{-1, 1\}^{n_m}$ which is used to represent the classical probabilities $p_n = \exp\left(\sum_p c_p m_p \right)$, where $c_p \in \mathbb{R}$ are the real-valued hidden mixing neural biases. This mixing layer is interconnected with the visible layers in order to capture classical correlations, mediated via the weight matrix $U_{kp} \in \mathbb{C}^{n_v \times n_m}$. Combining all the RBM contributions the ansatz reads,

$$
\rho^{\boldsymbol{\alpha},\boldsymbol{\beta}} = \sum_{\boldsymbol{m}} \sum_{\boldsymbol{h}_{\boldsymbol{\alpha}}, \boldsymbol{h}_{\boldsymbol{\beta}}} \exp\left(\sum_{p=1}^{n_m} c_p m_p \right) \times \exp\left(\sum_k a_k \alpha_k + \sum_j b_j h_{\alpha j} + \sum_{k,j} W_{kj} h_{\alpha j} \alpha_k + \sum_{k,p} U_{kp} m_p \alpha_k \right)
$$
$$
\times \exp\left(\sum_k a_k^* \beta_k + \sum_l b_l^* h_{\beta l} + \sum_{k,l} W_{kl}^* h_{\beta l} \beta_k + \sum_{k,p} U_{kp}^* m_p \beta_k \right). \tag{A3}
$$

Since there are no intra-layer connections, the hidden layers can be effectively traced out, leaving the mixed state ansatz used in equation (13).

In this work, we make use of a matrix element-wise version of the mixed state decomposition in equation (A1), such that

$$\rho = \sum_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \sum_n p_n \phi_n(\boldsymbol{\alpha}) \phi_n^*(\boldsymbol{\beta}) |\boldsymbol{\alpha}\rangle\langle\boldsymbol{\beta}| = \mathcal{P} \odot \sigma, \tag{A4}$$

where $\mathcal{P}$ describes classical contributions to the density-matrix, while σ describes quantum contributions from pure states. This representation is readily accessible via the NNS ansatz, and extremely useful for programming forms of entanglement. Importantly, it is by construction that the contributions from the mixing layer are purely classical. On its own, the mixing layer is capable of simulating classical correlations only, and is therefore implicitly separable.

Appendix B. Learning with complex-exponential ansatz for mixed states

As discussed in section 1.1, one can make use of a restructuring of the mixed state ansatz into complex exponential form in order to take better control of the learning procedure. Indeed, the total mixed state can be expressed as

$$\rho_{\Omega,\Pi,\Xi}^{\boldsymbol{\alpha},\boldsymbol{\beta}} = e^{i \log\left(\Phi_{\Xi}(\boldsymbol{\alpha},\boldsymbol{\beta})\vartheta_{\Omega}(\boldsymbol{\alpha},\boldsymbol{\beta})\right)} \Gamma_{\Pi}(\boldsymbol{\alpha},\boldsymbol{\beta}) r_{\Omega}(\boldsymbol{\alpha},\boldsymbol{\beta}), \tag{B1}$$

such that the state is constructed from three variational parameter sets, where r_{Ω} and Γ_{Π} assume responsibility for the magnitude of any element of the density-matrix, while functions Φ_{Ξ} and ϑ_{Ω} are responsible for the complex phase of such elements. Consider a target state χ which also admits the following decomposition

$$\chi^{\boldsymbol{\alpha},\boldsymbol{\beta}} = \lambda(\boldsymbol{\alpha},\boldsymbol{\beta}) e^{i \log \xi(\boldsymbol{\alpha},\boldsymbol{\beta})}. \tag{B2}$$

The pure density-matrix phase/amplitude functions Φ_Ξ and Γ_Π respectively, are parameterized by real valued parameter sets. Furthermore, they are decomposed with respect to their pure state wavefunctions, as shown in equation (14). The logarithmic derivatives of the pair of pure state phase functions take the form

$$\frac{\partial \log |\Phi_\Xi\rangle}{\partial \Xi_k} = \sum_{\alpha,\beta} \left(\frac{\partial \log \varphi(\alpha)}{\partial \Xi_k} - \frac{\partial \log \varphi(\beta)}{\partial \Xi_k} \right), \tag{B3}$$

while the amplitude function derivatives are

$$\frac{\partial \log |\Gamma_\Pi\rangle}{\partial \Pi_k} = \sum_{\alpha,\beta} \left(\frac{\partial \log \sigma(\alpha)}{\partial \Pi_k} + \frac{\partial \log \sigma(\beta)}{\partial \Pi_k} \right). \tag{B4}$$

Meanwhile, the mixing state phase/amplitude wavefunctions ϑ_Ω and r_Ω respectively are based on complex parameters. In this case, it is expedient to take derivatives with respect to real and imaginary components, i.e. $\frac{\partial \log |r_\Omega\rangle}{\partial \operatorname{Re}(\Omega_k)}, \frac{\partial \log |r_\Omega\rangle}{\partial \operatorname{Im}(\Omega_k)}, \frac{\partial \log |\vartheta_\Omega\rangle}{\partial \operatorname{Re}(\Omega_k)}$ and $\frac{\partial \log |\vartheta_\Omega\rangle}{\partial \operatorname{Re}(\Omega_k)}$ which can be treated separately. All these derivatives take real, compact and easily derived forms with respect to the neural network parameters, making gradient computations straightforward.

The learning procedure of minimising the negative logarithmic fidelity between a target vectorized density-matrix $|\chi\rangle$ and the mixed NNS is given by the usual update rule in section 3. Defining the quantity

$$\Delta(\alpha, \beta) = \langle \rho_{\Omega,\Pi,\Xi} | \chi \rangle^{-1} \, \mathrm{e}^{\mathrm{i} \log \frac{\Phi_\Xi(\alpha,\beta)\vartheta_\Omega(\alpha,\beta)}{\xi(\alpha,\beta)}}, \tag{B5}$$

where $\langle \rho_{\Omega,\Pi,\Xi} | \chi \rangle$ is the vectorized overlap between the variational and target state, we can then make use of the following gradients,

$$\nabla_k^{\Gamma_\Pi} \mathcal{L} = \sum_{\alpha,\beta} \left[\frac{r_\Omega^2(\alpha,\beta)\Gamma_\Pi(\alpha,\beta)}{|\rho_{\Omega,\Pi,\Xi}|^2} - \lambda(\alpha,\beta)r_\Omega(\alpha,\beta)\operatorname{Re}\left[\Delta(\alpha,\beta)\right] \right] \cdot \mathcal{O}_k^\Pi |\Gamma_\Pi\rangle, \tag{B6}$$

$$\nabla_k^{r_\Omega} \mathcal{L} = \sum_{\alpha,\beta} \left[\frac{\Gamma_\Pi^2(\alpha,\beta)r_\Omega(\alpha,\beta)}{|\rho_{\Omega,\Pi,\Xi}|^2} - \lambda(\alpha,\beta)\Gamma_\Pi(\alpha,\beta)\operatorname{Re}\left[\Delta(\alpha,\beta)\right] \right] \cdot \mathcal{O}_k^{\Omega_r} |r_\Omega\rangle, \tag{B7}$$

$$\nabla_k^{\Phi_\Xi} \mathcal{L} = -\sum_{\alpha,\beta} \left[\frac{r_\Omega(\alpha,\beta)\lambda(\alpha,\beta)\Gamma_\Pi(\alpha,\beta)}{\Phi_\Xi(\alpha,\beta)} \operatorname{Im}\left[\Delta(\alpha,\beta)\right] \right] \cdot \mathcal{O}_k^\Xi |\Phi_\Xi\rangle, \tag{B8}$$

$$\nabla_k^{\vartheta_\Omega} \mathcal{L} = -\sum_{\alpha,\beta} \left[\frac{r_\Omega(\alpha,\beta)\lambda(\alpha,\beta)\Gamma_\Pi(\alpha,\beta)}{\vartheta_\Omega(\alpha,\beta)} \operatorname{Im}\left[\Delta(\alpha,\beta)\right] \right] \cdot \mathcal{O}_k^{\Omega_\vartheta} |\vartheta_\Omega\rangle. \tag{B9}$$

Here, $|\rho_{\Omega,\Pi,\Xi}|^2$ is the magnitude of the vectorized density-matrix. Furthermore $\mathcal{O}_k^{\Omega_r} = \operatorname{diag}\left(\partial_{\Omega_k} \log |r_\Omega\rangle\right)$ and $\mathcal{O}_k^{\Omega_\vartheta} = \operatorname{diag}\left(\partial_{\Omega_k} \log |\vartheta_\Omega\rangle\right)$ are the diagonal matrices with mixing layer gradients. Again, these are treated separately with respect to real and imaginary valued parameters in Ω.

Appendix C. Details on numerical simulation

The gradient descent optimization procedures utilized throughout this work were facilitated by an adaptive learning rate scheme using the AdaMax optimizer [57] with a typical initial learning rate of the order $\eta_{\mathrm{init}} \in [10^{-4}, 10^{-3}]$. The number of learning iterations varied dependent on the complexity of the target state, i.e. complexity of entanglement needed to be simulated/classified, the dimension of the qudit system being considered (and therefore size of the target density-matrix). Since the time-to-convergence is shorter for states with smaller degrees of entanglement, it is intuitively more efficient to perform classification with a separable NNS than to explicitly reconstruct an entangled state.

A scenario in which the efficiency of learning can be greatly enhanced is the study of evolving, or 'nearby' states. Consider the results from figures 5–7. In a number of instances, we are classifying/quantifying the entanglement of a target state which is changing incrementally (and by a small amount) throughout an interval. Consider an NNS ρ_Ω that learns a state σ. It is logical to assume that if the target state is perturbed/evolved by some small amount, $\sigma' = \sigma + \delta\sigma$, the network Ω will only need to be optimized by a small amount $\Omega' = \Omega + \delta\Omega$. Therefore, when studying evolving target states, it is extremely useful to initialize each state using the parameter distribution of the previous learner. This not only simplifies learning and performance, but increases efficiency dramatically; the initial target can be reconstructed over a number of optimization steps S, but subsequent alterations to the network only require a fraction of S steps.

Importantly, when performing this method with SNNS one should ensure that the chosen initialization complies with the entanglement properties of the separable variational state, i.e. a separable network should be initialized with a nearby separable network state. If an SNNS is initialized in with the network parameters of a nearby entangled NNS, when separability conditions are imposed the network state will change rapidly and potentially end up in a state that is very different to the target, contrary to the desired effect.

ORCID iDs

Cillian Harney ⓘ https://orcid.org/0000-0002-9333-3407
Mauro Paternostro ⓘ https://orcid.org/0000-0001-8870-9134
Stefano Pirandola ⓘ https://orcid.org/0000-0001-6165-5615

References

[1] Gühne O and Tóth G 2009 *Phys. Rep.* **474** 1
[2] Horodecki R, Horodecki P, Horodecki M and Horodecki K 2009 *Rev. Mod. Phys.* **81** 865
[3] Nielsen M A and Chuang I L 2011 *Quantum Computation and Quantum Information: 10th Anniversary Edition* 10th edn (Cambridge: Cambridge University Press)
[4] Vedral V, Plenio M B, Rippin M A and Knight P L 1997 *Phys. Rev. Lett.* **78** 2275
[5] Vedral V and Plenio M B 1998 *Phys. Rev.* A **57** 1619
[6] Plenio M B and Virmani S 2007 *Quantum Inf. Comput.* **7** 1
[7] Peres A 1996 *Phys. Rev. Lett.* **77** 1413
[8] Horodecki P 1997 *Phys. Lett.* A **232** 333
[9] Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L and Zdeborová L 2019 *Rev. Mod. Phys.* **91** 045002
[10] Bharti K, Haug T, Vedral V and Kwek L-C 2020 *AVS Quantum Sci.* **2** 034101
[11] Carrasquilla J 2020 *Adv. Phys.: X* **5** 1797528
[12] Carleo G and Troyer M 2017 *Science* **355** 602
[13] Melko R G, Carleo G, Carrasquilla J and Cirac J I 2019 *Nat. Phys.* **15** 887
[14] Torlai G, Mazzola G, Carrasquilla J, Troyer M, Melko R and Carleo G 2018 *Nat. Phys.* **14** 447
[15] Torlai G and Melko R G 2018 *Phys. Rev. Lett.* **120** 240503
[16] Tiunov E S, Tiunova V V, Ulanov A E, Lvovsky A I and Fedorov A K 2020 *Optica* **7** 448
[17] De Vlugt I J S, Iouchtchenko D, Merali E, Roy P-N and Melko R G 2020 *Phys. Rev.* B **102** 035108
[18] Yuan D, Wang H, Wang Z and Deng D-L 2020 arXiv:2009.00019
[19] Hartmann M J and Carleo G 2019 *Phys. Rev. Lett.* **122** 250502
[20] Nagy A and Savona V 2019 *Phys. Rev. Lett.* **122** 250501
[21] Vicentini F, Biella A, Regnault N and Ciuti C 2019 *Phys. Rev. Lett.* **122** 250503
[22] Yoshioka N and Hamazaki R 2019 *Phys. Rev.* B **99** 214306
[23] Jónsson B, Bauer B and Carleo G 2018 arXiv:1808.05232
[24] Medvidovic M and Carleo G 2020 arXiv:2009.01760
[25] Torlai G, Mazzola G, Carleo G and Mezzacapo A 2020 *Phys. Rev. Res.* **2** 022060
[26] Bausch J and Leditzky F 2020 *New J. Phys.* **22** 023005
[27] Deng D-L, Li X and Das Sarma S 2017 *Phys. Rev.* X **7** 021021
[28] Harney C, Pirandola S, Ferraro A and Paternostro M 2020 *New J. Phys.* **22** 045001
[29] Vedral V 2002 *Rev. Mod. Phys.* **74** 197
[30] The mixing hidden biases c_p are set as real, while the network weights U_{kp} are complex. This is just a simplification, since imaginary components of c_p are negated in the network output functions. If c_p were complex, it would be clear in equation (9) that only the real components would be utilized, since $\phi_p(\alpha, \beta) = c_p + c_p^* + \sum_k U_{kp}\alpha_k + U_{kp}^*\beta_k = 2\mathrm{Re}(c_p) + \sum_k U_{kp}\alpha_k + U_{kp}^*\beta_k$.
[31] We are also free to combine visible layer biases $\boldsymbol{a}$ into the parameter set Π, since they are inherently local and thus invoke separable correlations
[32] Note that in equation (45) the union runs over all $|\mathcal{K}'| > |\mathcal{K}|$. This is necessarily a strict inequality. Suppose $|\mathcal{K}| = \mathrm{k}$ such that $\mathcal{K}$ describes a form of k-separability. The set of all $\mathcal{K}$-separable states thus inherits all states which are $(l < k)$-separable, but there are other forms of $|\mathcal{K}'| = \mathrm{k}$ separability which it will not inherit. This is true for disjoint and non-disjoint constructions, however in some non-disjoint cases these sets coincide.
[33] Wei T-C and Goldbart P M 2003 *Phys. Rev.* A **68** 042307
[34] Pirandola S, Laurenza R, Ottaviani C and Banchi L 2017 *Nat. Commun.* **8** 15043
[35] Hou S-Y, Cao C, Zhou D L and Zeng B 2020 *Quantum Sci. Technol.* **5** 045019
[36] Vollbrecht K G H and Werner R F 2001 *Phys. Rev.* A **64** 062307
[37] Horodecki P, Horodecki M and Horodecki R 1999 *Phys. Rev. Lett.* **82** 1056
[38] Dür W, Vidal G and Cirac J I 2000 *Phys. Rev.* A **62** 062314
[39] Greenberger D M, Horne M A, Shimony A and Zeilinger A 1990 *Am. J. Phys.* **58** 1131
[40] We refer to relayed entanglement as that which is caused indirectly through mutually entangled parties. For example, xmltex $\mathcal{K} = \{1,2|2,3\}$, entanglement is indirect between parties 1 and 3-it is relayed.
[41] Pirandola S *et al* 2020 *Adv. Opt. Photon.* **12** 1012
[42] Pirandola S 2019 *Commun. Phys.* **2** 51
[43] Pirandola S 2019 *Quantum Sci. Technol.* **4** 045006
[44] Pirandola S 2020 *IET Quantum Commun.* **1** 22
[45] Grasselli F, Kampermann H and Bruß D 2019 *New J. Phys.* **21** 123002

[46] Murta G, Grasselli F, Kampermann H and Bruß D 2020 *Adv. Quantum Technol.* **3** 2000025
[47] Lipinska V, Murta G and Wehner S 2018 *Phys. Rev.* A **98** 052320
[48] Unnikrishnan A, MacFarlane I J, Yi R, Diamanti E, Markham D and Kerenidis I 2019 *Phys. Rev. Lett.* **122** 240501
[49] Pirandola S, Braunstein S L, Laurenza R, Ottaviani C, Cope T P W, Spedalieri G and Banchi L 2018 *Quantum Sci. Technol.* **3** 035009
[50] Pirandola S and Lupo C 2017 *Phys. Rev. Lett.* **118** 100502
[51] Laurenza R, Lupo C, Lloyd S and Pirandola S 2020 *Phys. Rev. Res.* **2** 023023
[52] Pirandola S, Laurenza R, Lupo C and Pereira J L 2019 *npj Quantum Inf.* **5** 50
[53] Banchi L, Pereira J, Lloyd S and Pirandola S 2020 *npj Quantum Inf.* **6** 42
[54] Werner R F and Holevo A S 2002 *J. Math. Phys.* **43** 4353–7
[55] Cope T P W, Goodenough K and Pirandola S 2018 *J. Phys. A: Math. Theor.* **51** 494001
[56] Cope T P W and Pirandola S 2017 *Quantum Meas. Quantum Metrol.* **4** 44
[57] Kingma D P and Ba J L 2015 *Proceedings of the 3rd ICLR: Int. Conf. on Learning Representations* (San Diego) pp 1–15