Entanglement entropy production in Quantum Neural Networks

Marco Ballarin[1,2,3], Stefano Mangini[1,4,5], Simone Montangero[2,3,6], Chiara Macchiavello[4,5,7], and Riccardo Mengoni[8]

[1]These authors contributed equally to this work

[2]Dipartimento di Fisica e Astronomia "G. Galilei", via Marzolo 8, I-35131, Padova, Italy

[3]INFN, Sezione di Padova, via Marzolo 8, I-35131, Padova, Italy

[4]Dipartimento di Fisica, Università di Pavia, Via Bassi 6, I-27100, Pavia, Italy

[5]INFN Sezione di Pavia, Via Bassi 6, I-27100, Pavia, Italy

[6]Padua Quantum Technologies Research Center, Università degli Studi di Padova

[7]CNR-INO - Largo E. Fermi 6, I-50125, Firenze, Italy

[8]CINECA Quantum Computing Lab,Via Magnanelli, 6/3, 40033 Casalecchio di Reno, Bologna, Italy

Quantum Neural Networks (QNN) are considered a candidate for achieving quantum advantage in the Noisy Intermediate Scale Quantum computer (NISQ) era. Several QNN architectures have been proposed and successfully tested on benchmark datasets for machine learning. However, quantitative studies of the QNN-generated entanglement have been investigated only for up to few qubits. Tensor network methods allow to emulate quantum circuits with a large number of qubits in a wide variety of scenarios. Here, we employ matrix product states to characterize recently studied QNN architectures with random parameters up to fifty qubits showing that their entanglement, measured in terms of entanglement entropy between qubits, tends to that of Haar distributed random states as the depth of the QNN is increased. We certify the randomness of the quantum states also by measuring the expressibility of the circuits, as well as using tools from random matrix theory. We show a universal behavior for the rate at which entanglement is created in any given QNN architecture, and consequently introduce a new measure to characterize the entanglement production in QNNs: the entangling speed. Our results characterise the entanglement properties of quantum neural networks, and provides new evidence of the rate at which these approximate random unitaries.

Marco Ballarin: marco.ballarin.6@phd.unipd.it

1 Introduction

Nowadays quantum computing is a well-established research field where quantum phenomena like superposition and entanglement are exploited in order to process information, possibly more efficiently than standard classical data processing [1]. The aim of quantum computing is to devise quantum algorithms capable of generating a target quantum state representing the solution of a given problem. In the last decade, the community has put a large effort into the realization of hardware able to perform quantum computation.

Accompanying the rise of quantum computing, another research area, namely Machine Learning (ML), has gained a lot of popularity. We live undoubtedly in the era of big data, where information is collected by the most disparate devices. In this context, ML constitutes a set of techniques for efficiently identifying patterns in huge datasets and for inferring input-output relations in data, even in the case of previously unseen inputs [2, 3]. ML proves to be a powerful tool with a wide range of applications: from image classifications [4], over devising playing strategies for complex games [5], to controlling nuclear fusion reactors [6].

Inspired by some of these outstanding results, a new interdisciplinary research topic that goes by

arXiv:2206.02474v3 [quant-ph] 24 May 2023

the name of Quantum Machine Learning (QML) has recently begun to combine quantum computing and machine learning techniques in various ways, with the hope of achieving improvements in both fields [7, 8, 9, 10, 11]. As small-scale quantum devices start to be available [12], a new class of quantum procedures called *variational quantum algorithms* have been developed to take advantage of current and near-term quantum hardware, by trading theoretical guarantees of success with feasibility of execution [9, 11, 13]. Generally speaking, a variational quantum circuit is a hybrid quantum-classical algorithm employing a quantum circuit $U(\boldsymbol{\theta})$ that depends on a set of parameters $\boldsymbol{\theta}$, which are adjusted in order to minimize a given objective function. While the objective function is evaluated by measuring outcomes of the variational circuit, optimization is performed by a classical iterative optimization algorithm that proposes better candidates for the parameters $\boldsymbol{\theta}$, starting from random (or pre-trained) initial values.

Within the domain of variational quantum circuits, quantum versions of neural networks, often referred to as quantum neural networks (QNNs), represent a promising quantum alternative for classical supervised learning [10, 14]. An efficient encoding of input data is key to perform computations in a high dimensional (possibly even infinite) Hilbert space. In fact, it is possible to encode classical inputs $\boldsymbol{x}$ into a quantum state $|\mathcal{F}(\boldsymbol{x})\rangle$ using a parameterized quantum circuit (PQC), a procedure which goes by the name of *feature encoding*. Thus, the goal of the feature map $\mathcal{F}$ is to map classical vectors to the qubits' Hilbert space. This feature map is accompanied by a layered structure of additional variational PQCs, which are trained in order to solve the desired learning task. Recently, QNNs gained a lot of attention after it was shown that they could be more expressive and efficiently trained than their classical counterparts [14]. Still, the dispute on how to achieve quantum advantages over machine learning is still far from being settled [15, 16, 17]. As for classical neural networks, the type of implementation of parameterized quantum circuits has a profound impact on the QNN performances, both in terms of trainability and classification accuracy [18, 19, 20, 21, 22]. Thus, characterizing parameterized quantum circuits in terms of their expressibility and entangling capability is key to

selecting a good *ansatz*, i.e. structure, for a QNN.

Following and expanding the investigation pioneered in [23], in this work we study the entanglement properties of quantum neural networks initialized with random parameters. We employ methods from the tensor network literature, namely Matrix Product States (MPS), to study the entanglement generated in various QNNs architectures composed of up to 50 qubits. Since MPS are a very powerful tool for simulating quantum systems with bounded entanglement, if a quantum neural network can only access low entangled states, it can be easily simulated, which spoils any hope of achieving a concrete quantum advantage. Thus, using entanglement entropy among qubits as a figure of merit, we evaluate the entanglement capabilities of some of the most common and promising QNN architectures [14, 23]. We consider several QNNs with different combinations of feature maps $\mathcal{F}$ and variational forms V and perform an extended numerical analysis varying: (i) the number of qubits n, (ii) the number of layers L in the network, (iii) the entangling topology of the circuit, (iv) the data re-uploading [24, 25] structure being either alternated or sequential. In this respect, we focus our analysis on data re-uploading quantum circuits because, as extensively discussed later, recent results in the quantum machine learning literature highlight the need for such a circuit structure to increase the expressibility of the parametric models implemented by quantum neural networks. Thus, we consider this class of parametrized quantum circuits due to their practical relevance in quantum machine learning tasks. Nonetheless, in Sec. 3.4, we also analyze instances of random quantum circuits where parameters are not shared between layers (hence no data reuploading is used) and show that, as long as entanglement is involved, the results presented in this paper depend primarily on the architecture of the parametric quantum circuit, and not on the presence of shared parameters. A summary of the circuit templates analyzed in this work is shown in Fig. 1.

For all the considered QNNs with nearest neighbour connectivity, as the number of layers L is increased, the entanglement generated inside the circuit grows, eventually reaching a plateau when $L \approx n$, where n is the number of qubits. This behavior is associated with the typical en-
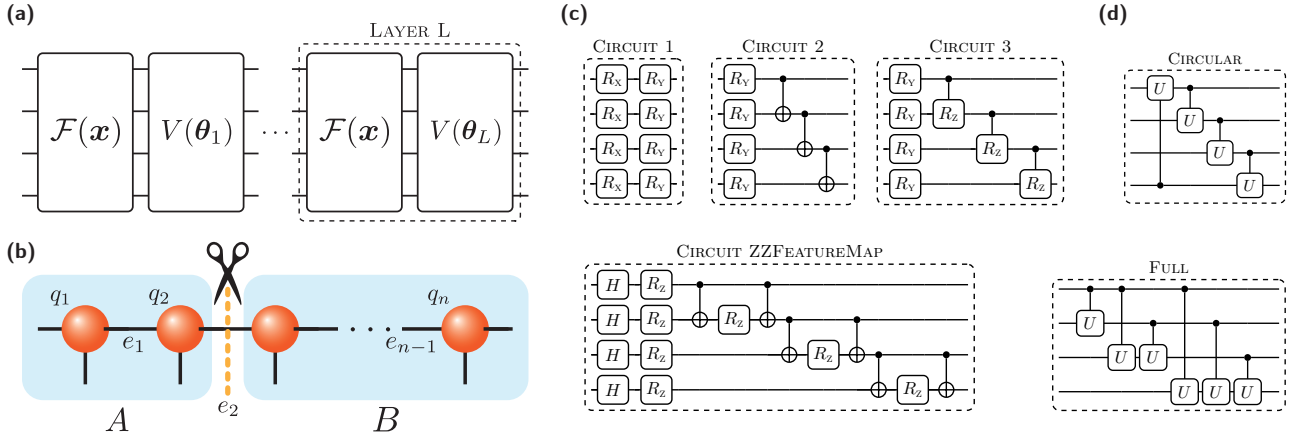
Figure 1 – Graphical representation of QNN and MPS. **(a)** QNN structure with alternating feature map $\mathcal{F}$ and variational ansatz V. Note that the ansatz parameters are different in each layer, while the feature map parameters are the same throughout the whole circuit. **(b)** MPS diagram. Each sphere is a tensor, representing a qubit q_j. The entanglement entropy between bi-partitions A and B is computed by "cutting" the connecting edge e_j. **(c)** Circuits analyzed in the manuscript, depicted with a linear entanglement topology, i.e. entangling gates are only applied between nearest neighbors on a line. **(d)** Different entanglement topologies: circular, with the first and last qubit of the line connected, and full, where the entangling gates are applied between each pair of qubits (see Appendix C for a clear definition and discussion). When using parameterized two qubits gates, like the controlled rotations in circuit 2, the entanglement maps are generalized to their parameterized version by substituting X gates on the controlled qubits with the corresponding parameterized operation. Note that the circuit templates 2 and ZZFEATUREMAP are those used in the QNN of [14], and also that circuits 1, 2 and 3 share similarities with circuits 1, 15, and 13 of [23], respectively.

tanglement of a random Haar-distributed quantum state. The choice of the entangling topology (nearest neighbors, circular, or all to all) clearly affects the rate of creation of entanglement in the circuit. We also point out that a careless definition of a full, i.e. all to all, connectivity map can effectively result in a linear nearest-neighbors interaction if unparameterized two qubits gates (CNOTs) are used, something apparently overlooked in the recent literature using this type of ansatz [14, 26, 27]. By bounding the entanglement generated by the circuit, we are able to simulate QNNs with MPS up to $n = 50$ qubits. It should be stressed that such simulations are exact up to a given number of layers, after which a truncation of the entanglement via MPS is applied. By appropriately normalizing the entanglement produced we show that all the points for a given QNN architecture follow the same curve, independently from the number of qubits. Thus, we exploit this behavior to define a universal figure of merit given the QNN architecture, the entangling speed. This figure of merit characterizes how fast the entanglement is produced by the QNN, with respect to the number of layers L.

In addition, we evaluate the expressibility mea-

sure of the considered QNNs as defined in [23] and argue that the optimality of the QNN introduced in [14] may be related to its good trade-off between mild entanglement production and high expressibility. Finally, we employ tools from random matrix theory, specifically convergence to the Marčenko-Pastur distribution, to further characterise the resemblance of the deep enough quantum neural networks to random unitary matrices. At last, we note that differently from [23] which bases their analysis on the Meyer-Wallach entanglement measure [28], in this work we make use of the entanglement entropy among subsystems, which allows for a more careful analysis of the entanglement distribution in the system, and it is also readily accessed in an MPS simulation with no computational overhead.

The manuscript is organized as follows. In Sec. 2 we review the basis of tensor networks and MPS, and introduce the Von Neumann entropy as an entanglement measure. We then discuss the entanglement entropy properties of random quantum states. We proceed by discussing the most recent results on parameterized quantum circuits and QNNs, especially, on the relation between randomness, trainability, and entangle-

ment found in these circuits. In Sec. 3 we show the results of our analysis for various QNN architectures, and discuss the results in Sec. 4. Finally, we discuss the implications of our work and possible routes for future investigations in Sec. 5.

2 Methods

2.1 Tensor Networks and Matrix Product States

An n-qubit quantum state is defined in a Hilbert space $\mathcal{H}$ of dimension $\dim(\mathcal{H}) = 2^n$. The exponential scaling of $\mathcal{H}$ with n makes the classical description of quantum states an exponentially expensive task. This problem is widely known in many-body quantum physics, and many different techniques have been developed to alleviate the issue, like the Density Matrix Renormalization Group (DMRG) or Tensor Network (TN) techniques [29, 30].

In this work, we use Tensor Network methods to efficiently describe the n-qubit state. In particular, we employ Matrix Product States (MPS), which are a specific tensor network ansatz particularly suited to represent 1-dimensional (i.e. like atoms on a chain, as in Fig. 1) quantum states [31]. The power of tensor networks lies in the assumption that we are only interested in a *tiny subspace* of the entire Hilbert space, namely the states that display a limited amount of entanglement.

An n-qubit pure state $|\psi\rangle \in \mathcal{H}$ can be written as a MPS as follows [31]

$$|\psi\rangle = \sum_{s_1,\ldots,s_n=0}^{1} \sum_{\alpha_1,\ldots,\alpha_n=1}^{\chi} \mathbf{M}_{1\alpha_1}^{[1],s_1} \mathbf{M}_{\alpha_1\alpha_2}^{[2],s_2} \cdots$$
$$\mathbf{M}_{\alpha_{n-2}\alpha_{n-1}}^{[n-1],s_{n-1}} \mathbf{M}_{\alpha_{n-1}1}^{[n],s_n} |s_1 s_2 \ldots s_n\rangle . \quad (1)$$

Each tensor $\mathbf{M}_{\alpha_i\alpha_{i+1}}^{[i],s_i}$ is a local description for the $[i]$-th site, which allows one to apply a *local* operator to a certain site without the need to change all the other coefficients. For a fixed s_i, $\mathbf{M}_{\alpha_i\alpha_{i+1}}^{[i],s_i}$ is a $\chi \times \chi$ complex matrix, meaning that Eq. (1) is the sum of basis elements weighted by matrix products. The integer χ is called the MPS *bond dimension*, and a sufficiently high χ is needed to express a general $|\psi\rangle$ in such form. However, MPS with a *lower* χ can still encode all the meaningful states, albeit clearly not *all* possible states. In particular, to correctly describe any quantum

state the bond dimension needed is $\chi = d^{\lfloor \frac{n}{2} \rfloor}$, where d is the local dimension of the degrees of freedom ($d = 2$ for qubits). We can also efficiently evolve the state under the application of 2-qubit gates, using an approach known in the literature as time-evolving block decimation [32], and perform measurements. Simulations using MPS are not bounded by the number of qubits in the system, but by the amount of entanglement generated inside it, as we explain in detail in Section 2.2.

Nonetheless, while the use of an MPS simulation imposes some constraints on the maximum entanglement that it is possible to represent, this issue is relevant only for very deep circuits involving many qubits. Indeed, we reliably simulate circuit instances involving up to $n = 50$ qubits and moderate depth, which is already sufficient to provide clear insights on the entanglement entropy generated in such circuits. Moreover, as explained below in Sec. 2.2, during an MPS simulation one has constant access to the singular values of the quantum state, so the entanglement of the state can be calculated on the fly without any computational overhead. Thus, MPS are an effective tool to study the entanglement properties of quantum circuits, especially in regimes that cannot be easily accessed with a full-scale simulation of the statevector of the system.

2.2 Entanglement measure in Matrix Product States

Entanglement in quantum states can be evaluated using the so-called Von Neumann *entanglement entropy*. Let $\rho = |\psi\rangle\langle\psi|$ be the quantum state of a system of n qubits, and consider a bipartition A, B of such system of qubits n_A and $n_B = n - n_A$ respectively, like the one shown in Fig. 1(b). The entanglement entropy of the subsystem A having reduced density matrix $\rho_A = \text{Tr}_B[\rho]$, is defined as

$$S(\rho_A) = -\text{Tr}[\rho_A \log \rho_A], \quad (2)$$

and quantifies the amount of entanglement shared between the parties A and its complement B. Note that throughout the whole manuscript we consider logarithms in natural base e. If A and B are in a product state then $S(\rho_A) = 0$, while if the two subsystems share maximal entanglement one has $S(\rho_A) = n_A \log(2)$ [31]. An important property of Eq. (2) is that the entanglement

entropy of the two subsystems is equal, namely $S(\rho_A) = S(\rho_B)$, as it can be easily checked using the Schmidt decomposition of the pure global state $\rho = |\psi\rangle\langle\psi|$ (see below).

It turns out that matrix product states are a natural tool to characterize the entanglement entropy of a quantum system. This can be illustrated by considering the simple case of a state of $n = 2$ qubits. The statevector

$$|\psi\rangle = \sum_{i,j=0}^{1} c_{ij} |ij\rangle, \text{ with } \sum_{i,j=0}^{1} |c_{ij}|^2 = 1, \quad (3)$$

can be expressed in the Schmidt decomposition as

$$|\psi\rangle = \sum_{\alpha=1}^{\chi_s} \lambda_\alpha |\xi_\alpha\rangle_1 |\eta_\alpha\rangle_2, \quad (4)$$

where χ_s is the *Schmidt rank*, λ_α are the Schmidt coefficients, and $\{|\xi_\alpha\rangle_1\}, \{|\eta_\alpha\rangle_2\}$ are orthonormal bases in the space of the first and second qubit respectively. Using the decomposition (4) in Eq. (2), the entanglement entropy between the two qubits then amounts to

$$S(\rho_A) = -\sum_{\alpha=1}^{\chi_s} \lambda_a^2 \log \lambda_\alpha^2, \quad (5)$$

In an MPS simulation we always have access to a subset of the Schmidt coefficients, since such representation is built by iteratively applying the Singular Value Decomposition (SVD), a procedure equivalent to Schmidt-decomposing a quantum state. The reason why we have access only to subsets of them is that we impose the following conditions on the Schmidt coefficients. Listing the coefficients in ascending order, i.e. $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{\chi_s}$, then:

- Schmidt coefficients whose ratio with λ_0 is smaller than ϵ are discarded. The value of ϵ in this work is fixed at $\epsilon = 10^{-9}$;

- only the first largest χ_{max} coefficient are retained. The value χ_{max} is called maximum *bond dimension*.

The approximation we are performing is the optimal one in terms of the represented entanglement. Then, the measure of entanglement for the MPS now becomes

$$S(\rho_A) = -\sum_{\alpha=1}^{\chi_{max}} \lambda_a^2 \log \lambda_\alpha^2. \quad (6)$$

As explained in detail in Appendix F, despite the approximations, the faithfulness of the simulation can be easily monitored. Finally, we remark that since we have constant access to the considered subset of Schmidt coefficients during the state evolution, we are able to compute the entanglement entropy of a quantum state without any computational overhead.

2.3 Entanglement entropy in random quantum states

In this section we briefly describe the entanglement features of uniformly distributed random pure quantum states, that is quantum states sampled according to the unique unitarily invariant probability distribution induced by the Haar measure. Denoting by $\mathbb{U}(n)$ the group of $2^n \times 2^n$ unitary matrices, there is a unique unitarily invariant probability measure $\mu(U)$ defined on the group, and such measure is called *Haar measure* [33, 34, 35]. Unitary invariance corresponds to the requirement that the measure is invariant under translations in the space of unitary matrices, that is

$$\mu(MU) = \mu(UM) = \mu(U) \quad U, M \in \mathbb{U}(n).$$

The Haar measure induces a uniform probability distribution in the space of unitary matrices so that sampling a quantum state according to the Haar measure means randomly picking a state uniformly from the space of quantum states. We denote with $\mathcal{P}(n)$ such probability distribution.

We are interested in the entanglement features of random quantum states, particularly in the entanglement entropy. Let $|\psi\rangle \in (\mathbb{C}^2)^{\otimes n}$ be a quantum state of n qubits sampled from the uniform distribution $|\psi\rangle \sim \mathcal{P}(n)$, and a bipartition of the n qubits system in two subsystems A and B, of size n_A and $n_B = n - n_A$ respectively. Then, for $n_A \leq n_B$, the expected value of the entanglement entropy (2) corresponding to this cut amounts to the Page value [33, 36]

$$\mathbb{E}[S(\rho_A)] = \sum_{j=d_B+1}^{d_A d_B} \frac{1}{j} - \frac{d_A - 1}{2d_B}, \quad (7)$$

where $d_B = 2^{n_B}$, $d_A = 2^{n_A}$ are the local dimensions of the two subsystems, and the expectation value is over the uniform probability distribution $\mathbb{E}(\cdot) = \mathbb{E}_{|\psi\rangle \sim \mathcal{P}(n)}(\cdot)$. One can check that the entanglement is highest whenever the two partitions

have equal size $n_A = n_B = n/2$ (for n even, and similarly for n odd, $n_A = \lfloor \frac{n}{2} \rfloor$ and $n_B = \lceil \frac{n}{2} \rceil$).

From Eq. (7) it follows that $\mathbb{E}[S(\rho_A)] \geq \log d_A - d_A/2d_B$ [33], and since the maximum value of the entanglement entropy for such bipartition is $\log d_A$, obtained if the subsystems A and B share maximal entanglement, one concludes that random states are generally highly entangled. Indeed, in ref. [33] it was shown that the probability that a random pure state has entanglement entropy lower than $\log d_A - d_A/2d_B$ is exponentially small. Thus, with very high probability, random quantum pure states are almost maximally entangled.

2.4 Quantum Neural Networks as Parameterized Quantum Circuits

Currently available quantum devices are still too small and noisy to perform relevant fault-tolerant computations of notorious and efficient quantum algorithms, like Shor's factoring [1, 12]. For this reason, recent research has focused on a new paradigm of quantum computation based on so-called variational quantum algorithms (VQAs), which trade theoretical success guarantees with feasibility of execution, and are thought to be the most effective way to reach a quantum advantage in the near term, already with small quantum devices [9, 11, 37].

Variational quantum algorithms are based on PQCs, which are quantum circuits in which some of the unitary operations are characterized by *variational* parameters to be adjusted in order to solve an optimization problem. The optimal parameters are found by minimizing a properly chosen cost (or loss) function encoding the task to be solved. Let $U_{\boldsymbol{\theta}}$ be the unitary evolution implemented by a quantum circuit with tunable parameters $\boldsymbol{\theta}$, and O a Hermitian operator (an observable). The goal of variational quantum algorithms is to optimize the quantum circuit parameters $\boldsymbol{\theta}$ in order to minimize the expectation value (or variations thereof)

$$f(\boldsymbol{\theta}) = \langle \mathcal{O} \rangle_{\boldsymbol{\theta}} = \mathrm{Tr}\left[O\, U_{\boldsymbol{\theta}} \rho U_{\boldsymbol{\theta}}^{\dagger} \right] \qquad (8)$$

where ρ is an initial quantum state, generally set to the ground state $\rho = |\mathbf{0}\rangle\langle\mathbf{0}|$. This is achieved by means of an iterative hybrid quantum-classical approach where the quantum computer is used to estimate the cost function (8), and given such

value, the classical computer proposes new variational parameters according to an optimization method, the most common one being gradient descent.

There is freedom in the choice of the gate sequence defining the parameterized unitary $U_{\boldsymbol{\theta}}$, and a choice of its structure is referred to as variational *ansatz*. For example, the unitary could be composed of a layer of Pauli rotations around the X-axis on each qubit $R(\theta) = \exp(-i\theta X/2)$, followed by a layer of CNOTs acting on pairs of neighboring qubits. This is in fact the general blueprint of variational quantum circuits, as they are generally created by repeating single-qubits parameterized rotations followed by multi-qubits operations which introduce entanglement into the computation. Examples of parameterized quantum circuits are shown in Fig. 1.

Quantum Neural Networks. As it is often the case with learning tasks, either classical or quantum, the goal is to solve a problem given access to a dataset of inputs $\mathcal{X} = \{\boldsymbol{x}_i\}_i$, representative of the task to be solved. Whenever data is involved, variational quantum circuits are often referred to as Quantum Neural Networks, as they share the very same idea as their classical counterpart: learn patterns in input data by adjusting tunable parameters through the iterative minimization process known as *training*. In this case, the quantum circuit of the neural network depends on two sets of parameters $\boldsymbol{x}$ and $\boldsymbol{\theta}$, the former being the input data to be analyzed, and the latter the variational parameters to be adjusted (i.e. the *weights* of the neural network). In the quantum machine learning jargon, the encoding scheme used to load the input data onto the quantum computer is known as *feature map*, and consists of a unitary operation parameterized by $\boldsymbol{x}$. We will denote such feature encoding gate with $\mathcal{F}(\boldsymbol{x})$, where $\boldsymbol{x} \in \mathcal{X}$. As with the variational unitary, there is no standard choice for a feature map, and one has to pick a specific ansatz, ideally biasing the choice towards architectures built using knowledge of the problem to be solved [22, 21]. Summing up, a general QNN can be then expressed as

$$\begin{aligned} U_{\mathrm{QNN}}(\boldsymbol{x}; \boldsymbol{\theta}) &= \prod_{i=L}^{1} V(\boldsymbol{\theta}_i)\mathcal{F}(\boldsymbol{x}) \\ &= V(\boldsymbol{\theta}_L)\mathcal{F}(\boldsymbol{x}) \cdots V(\boldsymbol{\theta}_1)\mathcal{F}(\boldsymbol{x}), \end{aligned} \qquad (9)$$

where $\mathcal{F}(\boldsymbol{x})$ is the feature map ansatz depending on the input data $\boldsymbol{x}$; $V(\boldsymbol{\theta}_i)$ is a variational ansatz depending on trainable parameters $\boldsymbol{\theta}_i \in \boldsymbol{\theta} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_L)$ with $\boldsymbol{\theta}_i \in \mathbb{R}^p$; and L is the number of repetitions (also referred to as *layers*) of the such layered structure.

It was recently shown that uploading the input data multiple times throughout the circuit is essential for quantum neural networks to model higher-order functions of the inputs [25, 38]. Such procedure is now standard practice in the quantum neural network-based quantum machine learning, and it is called *data reuploading* [39, 24]. Notice that the input data in the feature map in Eq. (9) is the same in every layer, while the variational blocks V use a different parameter vector in every layer. In Fig. 1(a) we give a graphical representation of the general structure of QNNs. As for the explicit implementation of $\mathcal{F}$ and V, there is no fixed choice and these are usually composed of single qubit rotations followed by entangling operations, either fixed (e.g. CNOTs) or themselves parameterized (e.g. controlled rotations). See Fig. 1(c) for some prototypical examples of parameterized blocks proposed in the literature [14, 23], which we will consider throughout the manuscript.

2.5 Randomness, Entanglement and Trainability.

One of the hardest theoretical challenges affecting quantum machine learning models is the emergence of so-called *barren plateaus* (BPs) in their optimisation landscape [40]. BPs are regions in parameter space where the loss function is essentially flat, with no interesting minimising direction, so that it is not possible to train the model efficiently and independently of the optimization methods used, be it gradient-based [41] or gradient-free [42]. Different sources can lead to the unfolding of barren plateaus, and these can be broadly grouped into three main categories: randomness-induced BP [40, 43, 44], BP induced by *global* cost functions defined with observables having support on a large number of qubits [19], and eventually noise-induced BP [45].

In this work we are concerned with the former type of barren plateaus, that roughly occur when parameterized quantum circuits, when initialised with random parameters, resemble general random unitaries. Indeed, despite being quite lim-

ited in terms of qubits connectivity and gate operations, common instances of parameterized quantum circuits are often found to behave as unitary 2-designs, that is they efficiently approximate the statistics of Haar-random unitaries up to the second moment [46]. In this case, then one can prove that the variance of the gradients of any cost function $f(\boldsymbol{\theta})$ defined on the circuit will vanish exponentially with the number of qubits n, namely [43]

$$\mathrm{Var}_{\boldsymbol{\theta}}[\partial_k f(\boldsymbol{\theta})] \in \mathcal{O}(b^{-n}) \quad b > 1, \qquad (10)$$

where $f(\boldsymbol{\theta})$ is as in Eq. (8). Specifically, the cost function concentrates around its mean value and stays constant almost everywhere in parameter space [47], which makes training unfeasible.

Vanishing gradients are used as a *witness* to assess whether a parameterized quantum circuit resembles a unitary 2-designs. Of course, this is only *necessary* but not *sufficient* condition, as one can easily devise a circuit that is not a 2-design but has vanishing gradients, for example using a global cost with a shallow circuit [19].

In addition to vanishing gradients, another witness of randomness is the entanglement generated inside the circuit [44]. Indeed, as discussed previously in Sec. 2.3, random quantum states are almost maximally entangled, so one can use the maximality of entanglement generated by a parameterized circuit as an indicator of the resemblance to a random unitary evolution. As for vanishing gradients, the presence of large entanglement is however only a necessary but not sufficient condition for randomness, as a simple shallow circuit composed of Hadamards and CNOTs can create maximally entangled states (GHZ states), which are clearly not random. As discussed in [48], the so-called entanglement-induced BPs [44, 49] provide an alternative yet equivalent description of local cost barren plateaus (circuits with global costs always suffer of vanishing gradients [19], regardless of randomness), as they both stem from the proximity of parameterized quantum circuits to unitary 2-designs.

Indeed, if a circuit is a unitary 2-design, then the average entanglement entropy of any subsystem A of dimension d_A ($d_A \leq d_B$) will be already very close to its maximal value [48, 50]

$$\log d_A - 1 \leq \mathbb{E}_{\boldsymbol{\theta}}[S(\rho_A)] \leq \log d_A, \qquad (11)$$

and approaches the Page value (7) for truly Haar-random states. We provide a proof of Eq. (11) based on the Rényi 2-entropy in Appendix A.

Recent investigations quantify the tight connection between trainability and randomness in terms of the *expressibility*, roughly defined as the ability of a parameterized quantum circuit to address the full unitary space [23], and show that highly expressible ansätze have flatter loss landscape, hence they are harder to train [43]. We further discuss the expressibility measure in Sec. 4.

To summarise, while the presence of entanglement is a necessary ingredient to avoid classical simulability, its uncontrolled growth is likely to signal the emergence of barren plateaus. The evaluation of the entangling capabilities of parameterized quantum circuits is then a valuable diagnostic tool to provide information both on the classical simulability and trainability issues of quantum machine learning models. At last, we note that although various methods have been put forward to mitigate the occurrence of BPs [51, 52, 53], including proposals based on entanglement control [48, 49, 54], these remain a bottleneck for scaling up quantum machine learning computations based on variational circuits.

3   Results

We now proceed to analyze the entanglement production in various quantum neural network architectures with different feature maps and variational ansatz, obtained composing the circuit blocks shown in Fig. 1. In particular, we take as a prototypical example the QNN introduced in [14], argued as a good candidate for quantum machine learning applications in terms of capacity and expressibility, possibly achieving an advantage over classical counterparts. Such QNN model uses as feature-map $\mathcal{F}(\boldsymbol{x})$ the so-called ZZFEATUREMAP firstly introduced in [55] as a classically-hard map to load classical data on a quantum state in a non-linear fashion. The variational block $V(\boldsymbol{\theta})$ is instead composed of single qubit rotations followed by entangling operations. In order to better understand the effect of every single operation in the quantum circuit, we also consider variations of the QNN introduced above, varying both the feature map, the variational form, and the entangling topology. All considered circuit blocks are graphically represented in Fig. 1.

Be $U_L(\boldsymbol{x}, \boldsymbol{\theta})$ the unitary representing a specific quantum neural network with L layers with input data $\boldsymbol{x} = (x_1, \ldots, x_m) \in \mathbb{R}^m$, and variational parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p$, see Eq. (9). We consider random instances of such QNN by sampling both the inputs and the variational parameters according to the uniform distribution $x_i, \theta_i \sim \text{Unif}(0, \pi)$, hence obtaining a collection of QNNs $U_L = \{U_L(\boldsymbol{x}_i, \boldsymbol{\theta}_i), i = 1, \ldots, M\}$. Then, we study the entanglement entropy properties of each of these instances and average the result over the M trials (unless stated otherwise, we take $M = 100$). Thus, when in the following we refer to the entanglement entropy of a quantum circuit, we are always denoting the average over M realizations of that circuit. In order to evaluate the influence of the depth on the entanglement, we repeat this analysis by increasing the number of layers in the quantum neural network $L = 1, \ldots, L_{max}$.

Note that although the total number of parameters (inputs and parameters) depends on the specific feature map and variational form used, for the considered circuits such difference generally amounts to a constant and does not have a relevant impact on the results. In Tab. 1 we report the number of parameters in each circuit template analyzed in this work. We anticipate that while the number of parameters in the considered quantum circuits only scales polynomially with the system size n, these are found to be sufficient to reproduce some entanglement features of random unitaries, which are instead characterized by an exponential number parameters. This is in agreement with results on random quantum circuits that states that polynomial resources are sufficient to approximate unitary designs [56, 57]. We refer to Sec. 4 for an extended discussion.

3.1   Alternating vs. Sequential data reuploading

As a first analysis, we study the difference in entanglement growth between a standard QNN using an *alternated* repetitions of feature maps and variational forms (as in Fig. 1), and one in which we have first L repetitions of the feature map followed by L repetitions of the variational form. We call this structure *sequential*. The former leverages an alternated evolution of the quantum state which is typical of quantum neural networks using a data reupload-

	Abbr.	Number of parameters		
		LINEAR	CIRCULAR	FULL
CIRCUIT 1	C_1	$2n$	$2n$	$2n$
CIRCUIT 2	C_2	n	n	n
CIRCUIT 3	C_3	$2n-1$	$2n$	$\frac{n^2+n}{2}$
CIRCUIT ZZFEATUREMAP	C_{zz}	n	n	n

Table 1 – Number of parameters for each considered circuit and their relative entanglement topology. Notice that, while the number of parameters remains constant for C_{zz} as shown in the table, the number of parametric gates varies analogously to C_3.
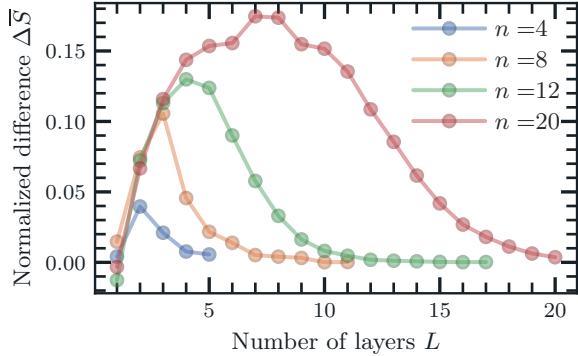
Figure 2 – Normalized entanglement difference $\Delta \overline{S}$, as defined in Eq. (12) for different numbers of qubits. The used QNN is defined by $\mathcal{F} = C_{zz}$ and $V = C_2$. All the data points are obtained by averaging over 10^3 realizations.

ing scheme [24, 25, 38]. The latter instead uses an initial data-dependent evolution followed by a trainable unitary, thereby creating an architecture similar to quantum kernel machines [58]. While the two structures (alternated and sequential) may be mapped to each other using ancillary qubits [59], they can have rather different performances, and we hereby show how they also create entanglement in a different way. Specifically, given the two unitary evolutions, namely the fixed input-dependent feature map $\mathcal{F}(\boldsymbol{x})$ and the varying parameterized variational form $V(\boldsymbol{\theta}_i)$, one expects the alternated dynamics

$$U_{alt} = V(\boldsymbol{\theta}_L)\mathcal{F}(\boldsymbol{x}) \cdots V(\boldsymbol{\theta}_1)\mathcal{F}(\boldsymbol{x}),$$

to introduce randomness at a faster rate than the sequential process

$$U_{seq} = \prod_{i=1}^{L} V(\boldsymbol{\theta}_i) \prod_{i=1}^{L} \mathcal{F}(\boldsymbol{x}),$$

and hence introduce more entanglement in the system. Such intuition is confirmed by the nu-

merical results, and may be understood as a consequence of the universality of the alternating dynamics proved for example for QAOA circuits [60, 61].

Here we use $\mathcal{F} = C_{zz}$ and $V = C_2$, as defined in Fig. 1, both with linear topology. Be S^{alt} and S^{seq} the entanglement entropy of the bipartition with an equal number of qubits, which is generally the highest, for the alternating and sequential structure, respectively. We define the normalized difference as

$$\Delta \overline{S} = \frac{S^{alt} - S^{seq}}{(S^{alt} + S^{seq})/2}, \qquad (12)$$

and study its behavior as the depth of the quantum circuit is increased, as shown in Fig. 2.

The metric is always positive and features a maximum, implying that the alternated structure is creating entanglement *faster* (i.e. with fewer layers) than its non-alternated counterpart. Note that for $L = 1$ layers the two structures are identical, so the generated entanglement is the same up to the statistical error, which explains why all the curves start around zero. At a high number of repetitions, the two structures tend to the same value, showing a $\Delta \overline{S} \simeq 0$, which can be understood in light of the results presented in the following sections: as the number of layers of a QNN is increased, the entanglement rapidly converges to that of a Haar-distributed random state, thus the alternated and non-alternated structure eventually converge to the same value. Given the higher entanglement production rate of the alternated structure, in the following analysis, we shall focus on this structure only.

3.2 Entanglement distribution across bonds

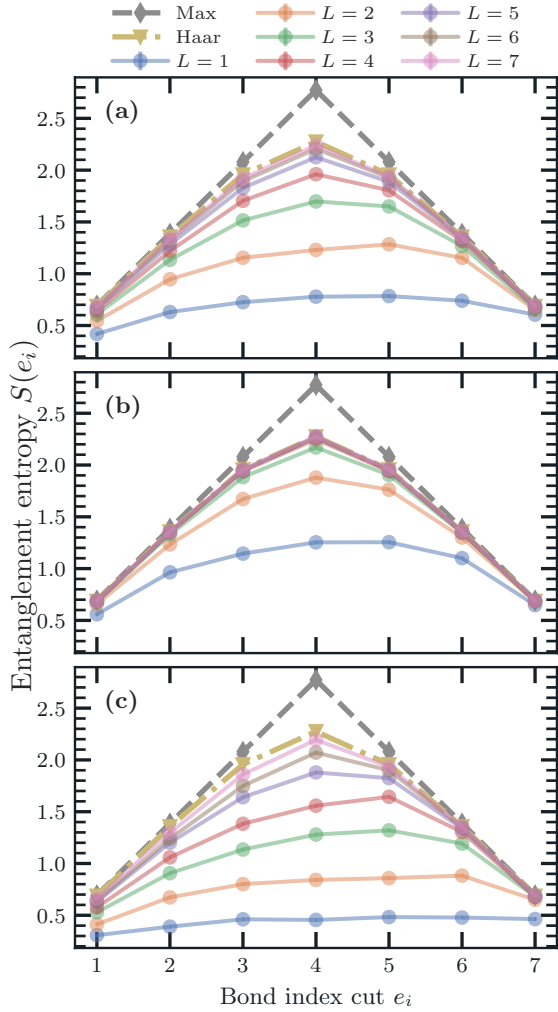It is natural to ask how the choice of the feature map, the variational form, and the entangling

Figure 3 – Average entanglement entropy across bonds for a system composed of $n = 8$ qubits, where e_i is the bond connecting qubit i and $i+1$, as in Fig. 1. The curves represent different numbers of layers L in the quantum neural network. **(a)** QNN with structure: $\mathcal{F} = \mathrm{C_{zz}}$, $V = \mathrm{C_2}$, both with linear entanglement. **(b)** QNN with structure: same as in (a) but with circular entanglement. **(c)** QNN with structure: $\mathcal{F} = \mathrm{C_1}$ which has no entangling gates, and $V = \mathrm{C_2}$ with linear entanglement.

topology impact the growth of entanglement of the quantum state. In this section, we start to explore this question by studying how entanglement is distributed across all possible ordered bipartitions of the n qubits in the network. That is, given an MPS representation as in Fig. 1(b), we study the entanglement entropy corresponding to each bond in the linear chain. Denoting with e_i the bond connecting qubit q_i and q_{i+1}, the entan-

glement entropy of that bond is (see Eq. (2))

$$S(e_i) = - \mathrm{Tr}\Big[\rho_{[1:i]} \log \rho_{[1:i]}\Big] \,, \tag{13}$$
$$\rho_{[1:i]} = \mathrm{Tr}_{i+1,\dots,n}[\rho]$$

where $\rho_{[1:i]}$ is the reduced density matrix of all the qubits up to the i-th one, and ρ is the state obtained from the quantum neural network $\rho = U_L(\boldsymbol{x}, \boldsymbol{\theta}) |\mathbf{0}\rangle\langle\mathbf{0}| U_L(\boldsymbol{x}, \boldsymbol{\theta})^\dagger$.

In Fig. 3 we show the entanglement entropy distribution for the case of $n = 8$ qubits using three different quantum neural networks architectures: in panel (a) the one proposed in [14] with feature map $\mathcal{F} = \mathrm{C_{zz}}$, variational ansatz $V = \mathrm{C_2}$, both with linear entanglement; in (b) same as before but using a circular entanglement topology; and eventually in panel (c) a simpler circuit using a tensor product feature map $\mathcal{F} = \mathrm{C_1}$ which encodes data independently on each qubit, followed by the same variational ansatz $V = \mathrm{C_2}$, again with linear entanglement both. For reference, it is also shown the expectation value of the entanglement entropy for Haar-random quantum states evaluated with Eq. (7), as well as an upper bound given by the highest possible entanglement $\log(\min(d_A, d_B))$, obtained if the two partitions A and B were maximally entangled. Note that while we report only the simulation data for $n = 8$, the discussion has general validity as identical results hold for all tested numbers of qubits, $n = 2, \dots, 20$.

First of all, the findings agree with the intuition that deeper circuits are able to create higher entangled states with respect to shallower ones, in accordance with results from [23]. In particular, the entanglement entropy is higher at the center of the chain. Clearly, depending on the specifics of the QNN, the entanglement grows faster in certain architectures with respect to others. Regarding the effect of the entangling topology, comparing panels (a) and (b) we see that circular connections produce greater entanglement compared to the nearest-neighbors interaction and that such entanglement grows at a faster rate as the number of layers is increased. As for the choice of the feature map, since the QNN in panel (c) produces entanglement only through the entangling gate in the variational blocks, its entanglement is lower and also grows slower with respect to the QNN in panel (a), even though it has twice the number of parameters in the feature map.

Interestingly, however, as the number of layers approaches the number of qubits $L \approx n$, all investigated QNNs converge to the same values, that is those obtained for random states sampled from the uniform Haar distribution. Deep enough QNNs are then flexible enough to reproduce the same entanglement spectrum of a random state, which, as discussed in section 2.3, are very highly entangled. Again, even though the measure of entanglement is different, this is in agreement with the results presented in [23], where the convergence to the Haar distribution is encountered for various parameterized quantum circuits, and also with other results in the literature regarding the properties of random quantum circuits to approximate the Haar distribution [62, 63]. We will discuss this more in detail in Sec. 4. A more in-depth analysis of the convergence is the subject of the next section.

3.3 Entanglement scaling with increasing depth

In order to better understand the entanglement scaling properties of QNNs, we introduce a new quantity, defined as the total entanglement entropy S_{tot} created in the MPS chain

$$S_{tot} = \sum_{i=1}^{n-1} S(e_i), \qquad (14)$$

which is the sum of the entanglement entropy of all the ordered bipartitions of the quantum state. We use this global measure to quantify how fast QNNs approach the Haar distribution in terms of overall entanglement production. In particular, we define a new figure of merit, the entangling layers $\widetilde{L}$, defined as the number of layers needed by an architecture to reach 90% of the total entanglement of a Haar distributed state S_{tot}^{Haar}, namely

$$\widetilde{L} = \min \# \text{ of layers } \text{ s.t. } S_{tot} \geq 0.9\, S_{tot}^{\mathrm{Haar}}. \qquad (15)$$

The choice of the 90% threshold allows to select states that are already very close to the Haar-random value, and avoids undesired oscillating behaviours obtained when higher thresholds are used, e.g. 99%, which are caused by statistical fluctuations (recall that every QNN is sampled multiple times with different parameters to calculate averages).

In Fig. 4 we show the behavior of $\widetilde{L}$ for four different QNNs as the number of qubits is increased. Note that each QNN is considered with all the three possible entangling topologies (*linear, circular* and *full* as defined in Fig. 1). At last, note that all QNNs leverage the same variational form $V = \mathrm{C}_2$, while the feature map is changed, as reported in the legend.

First, we observe that the entangling layers display a linear behavior when a linear entanglement topology is used. This means that the number of layers needed to entangle the system scales linearly with the size of the system. The behavior changes abruptly when we move to a circular or full entangling topology. All architectures display a faster entanglement production when passing from a linear to a circular topology, as can be seen from the lower slope of the curves. The all-to-all connectivity speeds up entanglement production only for $\mathcal{F} = \mathrm{C}_{zz}, \mathrm{C}_3$, while the circuits $\mathcal{F} = \mathrm{C}_2, \mathrm{C}_1$ show essentially the same behavior of the linear case. We now proceed to discuss more in detail such results.

We start comparing the entangling capabilities of C_{zz} vs. C_2. Both with linear and circular entangling topology, C_2 is able to produce entanglement essentially at the same rate as C_{zz}, despite C_2 being of a much simpler structure, with half the number of two-qubit gates. However, things change dramatically using a full entangling map, as the QNN reaches the 90% threshold already at $\widetilde{L} = 1$, while C_2 needs more layers, showing the same dependence of a linear connectivity. While counter-intuitive at first, is it easy to see that the entanglement generated by C_2 with a *full* architecture is indeed equivalent to the *linear* one. This is due to a simple circuit identity regarding networks of CNOTs reported in Fig. 5. Such circuital identity holds for any number of qubits, which makes the full entangling map as shown in Fig. 1 just as a linear entangling map in disguise (in particular, it is the inverse of the linear entangling map). See Appendix C for a more precise statement, discussion and proof. Such circuital identity thus explains the equivalence of the yellow ($\mathcal{F} = \mathrm{C}_2$) and red ($\mathcal{F} = \mathrm{C}_1$) curves between the first and last plot of Fig. 4.

Such equivalence clearly does not hold if controlled rotations are used instead of CNOTs. Indeed, the feature map $\mathcal{F} = \mathrm{C}_3$ uses controlled rotations with independent random parameters,
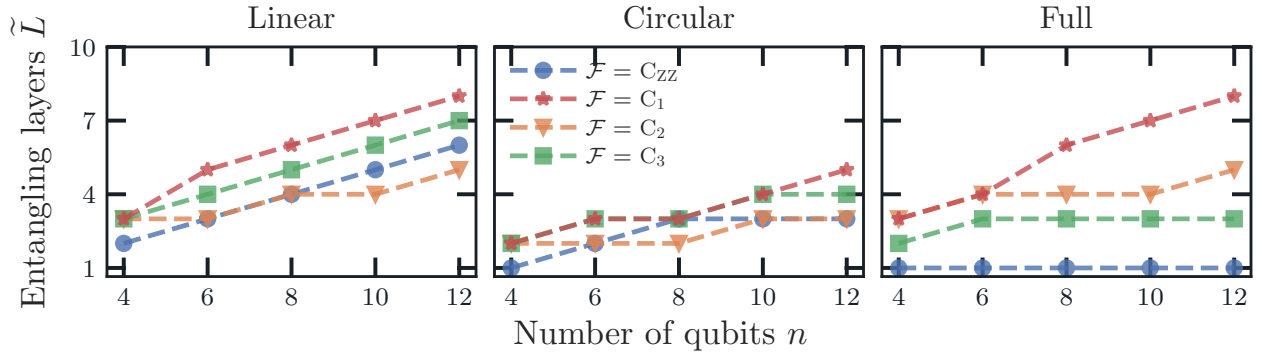
Figure 4 – Entangling layers $\widetilde{L}$, i.e number of layers to reach 90% of the Haar entanglement, versus the number of qubits. The analysis is carried out over four different QNN architectures, each evaluated with different entangling topologies (*linear, circular,* and *full*). The architectures leverage the same variational form V, while the feature map $\mathcal{F}$ is changed, as reported in the legend.
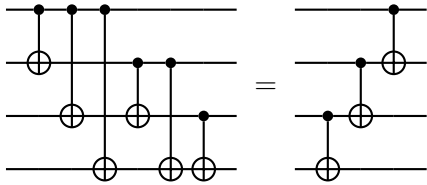

Figure 5 – Circuital identity between a full entangling map made of only CNOTs and the adjoint of a linear entangling map.

and given that these gates do not cancel out, the entanglement is always increasing going from low to high connectivity. Note that such increase is mainly due to the feature map, as the variational ansatz $V = C_2$ is the same as other structures, suffering from the CNOTs cancellation issue described above.

For comparison, we also show the performances of a QNN with the tensor product feature encoding $\mathcal{F} = C_1$, using no entangling operations. Interestingly, even if this QNN uses two-qubit interactions only inside the variational blocks, these are sufficient to create entanglement similar to other considered QNNs, even at a slower yet comparable rate.

We report in Appendix E the complete simulation results detailing the evolution of the entanglement with the depth of the circuit, for different numbers of qubits.

3.4 Entanglement Speed

So far we have presented numerical evidence for the entanglement production in QNNs up to a maximum of 20 qubits. In the following we extend the analysis leveraging MPS to simulate quantum systems of bigger size up to 50 qubits, with a maximum bond dimension of $\chi_{\max} = 4096$. More importantly, we show how the entanglement growth follows a behavior that is specific to each particular QNN architecture and the number of layers considered, but independent of the number of qubits in the circuit. We can thus uniquely assign an *entanglement speed* value to each QNN, which, we stress again, only depends on the choice of the ansatz, and holds identically for any instantiation of that QNN with arbitrary number of qubits.

Taking into account the entanglement growth discussed in Sec. 3.3, we restrict the analysis to a linear architecture, to increase as much as possible the number of layers we can correctly simulate with tensor networks techniques. Indeed, the entanglement production with a circular or full topology is too fast to allow for a convergent simulation with MPS for deep circuits.

Furthermore, we introduce the maximum Haar entanglement entropy, defined as the maximum across all bond entropies for a given number of qubits, as

$$S_{n,\max}^{\mathrm{Haar}} = \max_A \left(\mathbb{E}[S(\rho_A)] \right)$$
$$\approx \frac{n}{2} \log 2 - \frac{1}{2} \quad \text{for } n_A = \frac{n}{2} \gg 1, \tag{16}$$
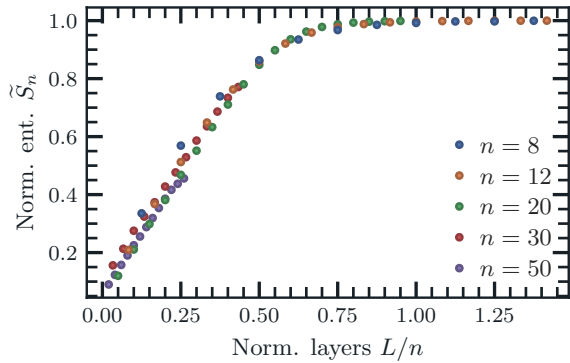
where the approximation in the second line has an

Figure 6 – Normalized entanglement $\widetilde{S}_n$ (17) versus the normalized number of layers L/n, for different number of qubits n, and for the QNN defined by $\mathcal{F} = \mathrm{C_{zz}}$, $V = \mathrm{C_2}$ with linear connectivity. All the normalized entanglement points follow the same curve, independently of the system size n. The points for $n = 8$, 12, 16, 20 are obtained by averaging over 10^3 samples, while for $n = 30$, 50 the averages are over 10 samples.

errors that scales as $\mathcal{O}\!\left(2^{-n/2}\right)$, see Appendix B for its derivation. Thus, for $n \geq 30$ qubits, when the exact computation of the Haar entanglement entropy is unfeasible, we employ the approximated Eq. (16). Finally, we define the normalized entanglement entropy $\widetilde{S}_n$ as:

$$\widetilde{S}_n = \frac{\max_{e_i}\left[S(e_i)\right]}{S_{n,\mathrm{max}}^{\mathrm{Haar}}} \, . \qquad (17)$$

We stress that $\widetilde{S}_n$ is normalized to the maximum Haar entanglement for a fixed n, not to the real maximum of the entanglement, which would be $S = \frac{n}{2}\log 2$ for the equal size bipartition.

In Fig. 6 we show the evolution of $\widetilde{S}_n$ versus the normalized number of layers L/n for $n \in \{8, 12, 16, 20, 30, 50\}$ qubits, for the QNN defined with $\mathcal{F} = \mathrm{C_{zz}}$, $V = \mathrm{C_2}$ with linear connectivity. We note that all the points, independently of the system size n, follow the same curve: an initial linear growth of the entanglement is followed by a saturation to the Haar-random value for the entanglement entropy (7). In particular, we check this behaviour also at large system sizes with $n = 30$, 50 qubits and circuits with up to $L = 11$ layers, and confirm that such scaling is indeed size independent. See Appendix F for a discussion on the errors introduced by truncation in the MPS representation for simulations with $n = 30$, 50 qubits.

Inspired by the behavior of $\widetilde{S}_n$, we introduce a measure for the entanglement production which is specific to a given QNN architecture (feature map plus variational ansatz) and independent of the number of qubits. Borrowing from the literature on random quantum circuits, it is known that the entanglement of a system undergoing random evolution initially grows linearly in time (depth of the circuit) before reaching the plateau of Haar random states [64, 65, 66, 54]. Indeed, as clear from Fig. 6, we observe the same initial linear growth, and thus we define the *entangling speed* v_s as

$$\widetilde{S}_n \propto v_s \cdot \left(\frac{L}{n}\right) \ \text{ for } \widetilde{S}_n \leq 0.5, \qquad (18)$$

where 0.5 is a threshold such that the linear behavior holds. The entangling speed can thus be obtained by fitting the curve in Fig. 6 with the linear function (18) in the appropriate range. We report in Tab. 2 the entangling speed for a subset of the inspected architectures, and notice that entanglement is produced at sensibly different rates. In agreement with the findings of Sec. 3.3, we see that for a linear topology the circuit $\mathrm{C_2}$ builds the entanglement at the fastest rate. Indeed, fixing the feature map to $\mathcal{F} = \mathrm{C_{zz}}$, $\mathrm{C_2}$ produces entanglement 3 times faster than $\mathrm{C_3}$.

To further characterise the applicability of the entangling speed, we show that the behavior of Fig. 6 evaluated for random circuits also holds when the input data $\boldsymbol{x} \in \mathbb{R}^n$ in the feature map $\mathcal{F}(\boldsymbol{x})$ are not drawn from the uniform distribution, but rather from real-world datasets. In particular, we select two common datasets in the machine learning literature, the wine [67] and breast cancer [68] datasets, and calculate the entanglement generated in the circuit when these data are fed into the feature maps (variational blocks are still populated with random parameters as before). The results presented in Fig. 7 are obtained by rescaling all the features of the datasets in the interval $[0, \pi]$. For each sample in the dataset, we average over $M = 10$ runs with randomly drawn parameters for the variational ansatz. The results shown in the figure are then obtained as the average over the whole dataset. The wine dataset ($n = 13$ features, hence $n = 13$ qubits, and 178 samples) follows perfectly the theoretical curve, and the breast cancer ($n = 9$ features, hence $n = 9$ qubits, and 286 samples) only slightly deviates from it, producing entanglement

Feature map $\mathcal{F}$	Variational ansatz V	Entangling speed v_s
C_{zz}	C_2	(1.8 ± 0.1)
C_{zz}	C_3	(0.59 ± 0.02)
C_1	C_3	(0.316 ± 0.006)

Table 2 – Entangling speed, i.e. a measure of how fast the entanglement is created, for different QNN architectures. These results are obtained using up to 16 qubits.

at a smaller rate. We then conclude that the entangling speed depends primarily on the architecture of the circuit rather than the actual values of the parameters. Clearly, this holds for reasonably distributed data features, that is excluding pathological cases of values being either zero or concentrating around it. Finally, to verify that the QNN architecture is ultimately responsible for the entanglement speed, we analyze random circuits where the encoding blocks do not share the parameters, but these are sampled independently for each layer, thus effectively removing the data-reuploading feature. This case is portrayed in Figure 7 with yellow square markers, each obtained by averaging over $M = 2000$ realizations of the random circuit, from which it is clear the normalized entanglement $\widetilde{S}_n$ again follows the same behavior of the previous scenarios.

Thus, the entangling speed can be used as a good estimate of the entanglement generated in a QNN also in real use cases, especially at the start of optimisation, when trainable parameters are usually initialised at random. For example, one could measure the entangling speed of the architecture of interest on a random quantum circuit of just a few qubits, and then estimate the entanglement generated with the same architecture on an arbitrary number of qubits and circuit layers, especially in regimes where simulations are no longer computationally feasible.

3.5 Expressibility

In addition to entanglement, another useful quantity to characterize parametrized quantum circuits is the expressibility, as defined by authors in [23]. Such measure quantifies how well the QNN is able to explore the Hilbert space by comparing the distribution of fidelities of states generated by the QNNs with that of randomly Haar-distributed ones (see Appendix D for a formal definition and explanation).

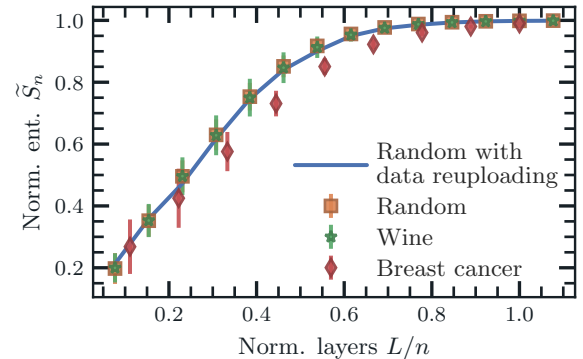Thus, in order to have a comprehensive under-

Figure 7 – Normalized entanglement using three different ways to sample the inputs $\boldsymbol{x}$ in the data encoding blocks. (i) "Random with data reuploading" indicates the case of random synthetic inputs with data reuploading, as shown also in Fig. 6. (ii) "Random" indicates the case of random synthetic inputs without data reuploading, that is encoding blocks in different layers have different random parameters. (iii) "Wine" and "Breast cancer" indicates inputs drawn from the corresponding real-world dataset, used with data reuploading. In all cases, the parameters in the variational blocks are sampled from the uniform distribution Unif$(0, \pi)$. Such distribution is used also to sample the synthetic random inputs. Data points for real-world datasets are obtained by first averaging over 10 realization for each sample in the dataset, and then averaging again over the whole dataset. Results for random inputs without data reuploading (yellow square markers) are obtained by averaging over 2000 realizations of the circuits. The error bars show the standard deviation of the mean. Error bars associated with random inputs with data reuploading (blue curve) are not shown to avoid cluttering but are of comparable size with the other points.
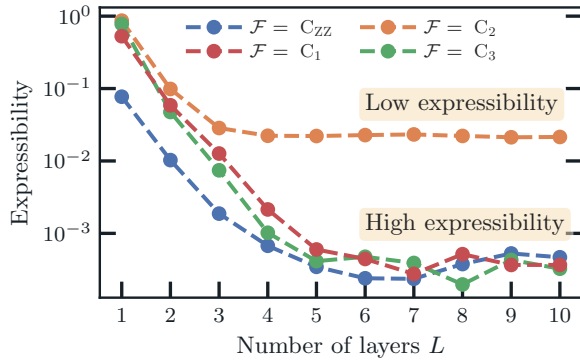
Figure 8 – Expressibility of the QNNs analyzed in Fig. 4, for $n = 8$ qubits with linear entanglement. The expressibility measures how well a variational circuit is able to address the unitary space (the lower, the better). All QNNs use the same variational form $V = C_2$, but with different feature maps. As the number of layers is increased, QNNs become more expressible, eventually reaching a plateau.
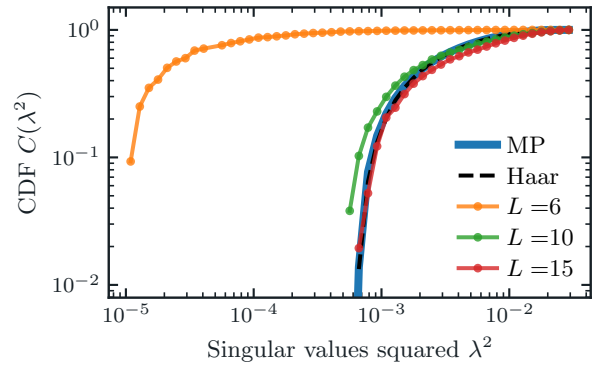

Figure 9 – Convergence to the Marčenko-Pastur (MP) distribution of the eigenvalues. The cumulative distribution function (CDF) $C(\lambda)$ of the eigenvalues λ^2 corresponding to the central bond of a quantum state of $n = 15$ qubits, generated with $\mathcal{F} = C_{zz}$, $V = C_2$, and linear connectivity. We show the behavior for different numbers of layers L, and for truly Haar-random states which, as expected, exactly follows the MP curve.

standing of the factors at play in the behavior of QNNs, in Fig. 8 we show the expressibility measure for the QNNs analyzed in Fig. 4 with a linear connectivity. As one would expect, the expressibility increases as the number of layers is increased, up until a plateau is reached.

Interestingly, the structure with $\mathcal{F} = V = C_2$ turns out to be the least expressible of all the structures considered, even if it is the one producing entanglement at the fastest rate, in agreement with the results reported in [23], as such QNN is indeed very similar to the parameterized circuit labeled '15' in [23]. On the contrary, the QNN with $\mathcal{F} = C_{zz}$, and $V = C_2$ proposed in [14] is able to reach high expressibility while producing entanglement at a controlled pace. As the presence of high entanglement is correlated with trainability issues [44], this QNN attains an optimal balance of mild entanglement with high expressibility even at low depth, which could be related to its good performances in quantum machine learning task [14, 55]. However, a similar, yet less favorable balance, is achieved by the other two architectures, so further investigation is needed to discriminate where the optimality comes from.

In this respect, the authors in [18] found the expressibility to be correlated with the classification accuracy of QNNs in supervised learning tasks, while weak correlation was found with the

entanglement generated inside the circuit, in line with the observations regarding entanglement-induced barren plateaus [44]. As discussed earlier in Sec. 2.5, both expressibility and high entanglement are related to the resemblance of the circuit to a random unitary, but while the former provides a more direct evidence, the latter gives an indirect indication. Indeed, there are cases of circuits having low expressibility but high entanglement, indicating that such circuits selectively explore only some highly-entangled regions of the Hilbert space [23].

3.6 Distribution of the singular values

The randomness of a quantum state can also be probed using tools from random matrix theory. Specifically, this can be done by studying the distribution of the eigenvalues of the reduced density matrices, which are known to follow the Marčenko-Pastur (MP) law when pure random quantum states are considered [69, 70]. More in detail, let $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be an Haar-random bipartite quantum state with Schmidt decomposition $|\psi\rangle = \sum_{i=1}^{d} \lambda_i \, |\xi_i\rangle_A \otimes |\eta_i\rangle_B$, where $d = \min(d_A, d_B)$ and $d_{A,B}$ is the dimension of the Hilbert space $\mathcal{H}_{A,B}$. The reduced density matrix $\rho_A = \text{Tr}_B [|\psi\rangle\langle\psi|]$ has eigenvalues λ_i^2 given by the square of the singular values, and for large system size their distribution is described by the MP distribution [71, 72].

In Figure 9 we show the cumulative distribution function of the eigenvalues $C(\lambda^2)$ of the reduced state of the first half of the qubits, obtained with a QNN with $n = 15$ qubits, feature map $\mathcal{F} = C_{zz}$, variational ansatz $V = C_2$, and linear connectivity. The distribution of the singular values for the QNN is obtained by running the circuit 10^2 times with different sets of parameters, and storing the singular values corresponding to the central cut. Then, we construct the cumulative distribution from the histogram of all the singular values obtained from the simulations. As the number of layers L is increased, the distribution of the eigenvalues approaches the theoretical MP distribution, eventually matching it when the number of layers is equal to the number of qubits. This behavior is displayed also by other QNN architectures. For completeness, we also show the distribution of the eigenvalues of a truly Haar-random quantum state, generated by sampling its entries independently from a normal complex distribution and then normalising it [69], which, as expected, follows perfectly the MP curve.

4 Discussion

Moments of the Haar distribution can be approximated efficiently using local random quantum circuits of sufficient depth. Depending on the connectivity dimension D of the qubits, defined as the number of other qubits that are connected to each qubit, order $\mathcal{O}(\text{poly}(t) \cdot n^{1/D})$-depth random circuits are sufficient to create approximate unitary t-designs [62, 63, 57, 56], that is circuits that generate a distribution of unitaries which approximately matches moments of the Haar distribution up to order t [46]. Numerical studies suggest that these results also hold for random parameterized quantum circuits of various forms [40, 19, 43, 23].

We extend these results by showing similar results also for quantum neural networks featuring data re-uploading, both for random instances using random inputs and parameters, and also for real-world dataset when these are used as inputs in the feature map. In particular, for a linear connectivity, as the number of layers approaches to the number of qubits $L \approx n$, QNNs display the same entanglement entropy properties of Haar-distributed random states, a fact which can be taken as a proxy for QNNs approximating unitary designs. Such behaviour was also confirmed by studying the randomness of the circuits with other metrics, namely the expressibility and the convergence to the Marčenko-Pastur distribution of the eigenvalues of the reduced states. In both cases, we find strong evidence of the QNN reproducing the same features of random quantum states as the number of layers approaches the system size using a linear connectivity.

Our analysis also underlines the importance of the entangling operations, as careless use of an all-to-all connection can result in unwanted simplifications, making the effective connectivity identical to a nearest-neighbors one. Parameterized two-qubit interactions can solve the problem, even though they may be challenging to implement on real hardware. A good trade-off is achieved with a circular entangling topology, which is immune to simplifications and shows remarkable entangling capabilities. Indeed, from the results of Fig. 4, we see that such connectivity is able to create high multiparite entanglement between qubits already at shallow depth, and with only minor additional hardware resources compared to the linear connectivity. An all-to-all topology instead reaches typical values for entanglement of random states essentially at constant number of layers $L \in \mathcal{O}(1)$ — implying in general $\mathcal{O}(n^2)$ entangling operations —, independently of the system size, and the architecture used (when non-trivial feature maps and variational ansätze are used).

While limiting the entanglement inside a quantum neural network may be necessary to ensure its trainability [44], low entanglement makes the circuit prone to be simulated exactly with an MPS, as discussed in Sec. 3.4. Thus, we envision that a sweet spot should be found in order for QNNs to show signs of quantum advantage: not too high to preclude trainability, not to low to escape triviality.

At last, the introduction of the entangling speed v_s (18) can be used as a figure of merit for the entanglement production of a given QNN, independent of the size of the system. Indeed, the entangling speed can be studied and assigned to an architecture in the simulable regime (low number of qubits n), and then used to estimate the number of layers to achieve a well-determined quantity of the entanglement, for any system size. We also stress that v_S characterizes the most in-

teresting interval of layers in a circuit. As discussed earlier, a value of the entanglement too high might be connected to barren plateaus, underlying the importance of exploring the regime where the entanglement has not saturated yet, and the linear regime still holds.

We now briefly comment on future interesting investigation directions regarding entanglement and QNNs. The focus of this work was to carefully study the entanglement features of common quantum ansätze, specifically when they are initialized with random parameters and no optimization has yet started. A natural followup is to ask whether entanglement plays any role also during the optimization process, which is at core of variational quantum algorithms. While for some specific variational procedures like QAOA [73] or VQE-based ground state solvers [74] one has some knowledge of the structure of the target solution, and hence can infer the behaviour of the entanglement created in the circuit, this is not the case for quantum machine learning tasks, as they are usually very task-dependent. Indeed, current proposals for QML advocate for the use of constrained quantum ansätze specifically tailored to the problem under investigation [22, 21, 75], and then one expects the depth of the circuit and the entanglement generated inside it to highly depend on the specific task to be solved, and dataset to fit, either classical or quantum [76]. Moreover, while the use of deep QNN ansätze (with arguably more entanglement) could offer some optimization advantages due to overparametrization [77, 78, 79], the emergence of barren plateaus suggests using shallow circuits instead [19, 45]. The characterisation of the role played by entanglement in QNNs, and how it may be leveraged to achieve a quantum advantage over classical methods will be objects of future studies.

5 Conclusion

In this paper we discussed in detail the entanglement generated by different promising Quantum Neural Networks (QNNs) when these are initialised with random parameters, and showed that they reproduce the same properties of random quantum states under various measures.

We employed a Matrix Product States (MPS) simulation of the quantum circuits, which guarantees an easy computation of the entanglement

in the circuits, and let us study systems of large system size composed of up to $n = 50$ qubits.

We showed that while all the architectures tend to a Haar entanglement distribution for a sufficiently high number of layers, the speed of convergence strongly depends on the specific circuit ansatz. This result highlights the universal behavior of the normalized entanglement production (17) for a given architecture, so we introduced a new measure to characterize a QNN in terms of its entanglement production: the entangling speed (18).

Finally, we argued that a trade-off between expressibility and entanglement is the key to a better understanding of QNN performances and an auspicious target for the search of quantum advantage. While high entanglement is a necessary condition to avoid classical simulability, a too-large entanglement is detrimental to the training procedure due to its tight connection with barren plateaus, as discussed in Sec. 2.2. A promising future direction is to extend the entanglement analysis of QNNs not only at initialization but also during the training procedure [48, 54, 74]. These tests would help to understand if QNNs really are a suitable platform for proving quantum advantage.

Code availability

All simulations with a number of qubits $n \leq 12$ were performed using Qiskit [80], while larger systems were simulated with Quantum Matcha Tea package [81] available at `https://baltig.infn.it/quantum_tea/quantum_tea`. The simulations for the training sections were performed using Pennylane [82]. All the code for reproducing the results presented here is available at the repository: `https://github.com/mballarin97/mps_qnn`.

Acknowledgment

References

[1] Michael A. Nielsen and Isaac L. Chuang. "Quantum computation and quantum information". Cambridge University Press. Cambridge, UK (2010). 10th anniversary ed. (2010) edition.

[2] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. "Deep learning". MIT Press. (2016). url: http://www.deeplearningbook.org.

[3] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". Nature 521, 436–444 (2015).

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks". In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1. Page 1097–1105. NIPS'12Red Hook, NY, USA (2012). Curran Associates Inc.

[5] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. "Mastering the game of Go with deep neural networks and tree search". Nature 529, 484–489 (2016).

[6] Jonas Degrave, Federico Felici, Jonas Buchli, Michael Neunert, Brendan Tracey, Francesco Carpanese, Timo Ewalds, Roland Hafner, Abbas Abdolmaleki, Diego de las Casas, Craig Donner, Leslie Fritz, Cristian Galperti, Andrea Huber, James Keeling, Maria Tsimpoukelli, Jackie Kay, Antoine Merle, Jean-Marc Moret, Seb Noury, Federico Pesamosca, David Pfau, Olivier Sauter, Cristian Sommariva, Stefano Coda, Basil Duval, Ambrogio Fasoli, Pushmeet Kohli, Koray Kavukcuoglu, Demis Hassabis, and Martin Riedmiller. "Magnetic control of tokamak plasmas through deep reinforcement learning". Nature 602, 414–419 (2022).

[7] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. "Quantum machine learning". Nature 549, 195–202 (2017).

[8] Vedran Dunjko and Peter Wittek. "A non-review of quantum machine learning: trends and explorations". Quantum 4, 32 (2020).

[9] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and et al. "Variational quantum algorithms". Nature Reviews Physics 3, 625–644 (2021).

[10] S. Mangini, F. Tacchino, D. Gerace, D. Bajoni, and C. Macchiavello. "Quantum computing models for artificial neural networks". Europhysics Letters 134, 10002 (2021).

[11] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S. Kottmann, Tim Menke, Wai-Keong Mok, Sukin Sim, Leong-Chuan Kwek, and Alán Aspuru-Guzik. "Noisy intermediate-scale quantum algorithms". Rev. Mod. Phys. 94, 015004 (2022).

[12] John Preskill. "Quantum Computing in the NISQ era and beyond". Quantum 2, 79 (2018).

[13] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. "A variational eigenvalue solver on a photonic quantum processor". Nat. Commun.5 (2014).

[14] Amira Abbas, David Sutter, Christa Zoufal, Aurelien Lucchi, Alessio Figalli, and Stefan Woerner. "The power of quantum neural networks". Nature Computational Science 1, 403–409 (2021).

[15] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Information-theoretic bounds on quantum advantage in machine learning". Phys. Rev. Lett. 126, 190505 (2021).

[16] Hsin-Yuan Huang, Michael Broughton, Masoud Mohseni, Ryan Babbush, Sergio Boixo, Hartmut Neven, and Jarrod R. McClean. "Power of data in quantum machine learning". Nature Communications 12, 2631 (2021).

[17] Franz J. Schreiber, Jens Eisert, and Johannes Jakob Meyer. "Classical surrogates for quantum learning models" (2022) arXiv:2206.11740.

[18] Thomas Hubregtsen, Josef Pichlmeier, Patrick Stecher, and Koen Bertels. "Evaluation of parameterized quantum circuits: On the relation between classification accuracy, expressibility, and entangling capability". Quantum Machine Intelligence **3**, 9 (2021).

[19] M. Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J. Coles. "Cost function dependent barren plateaus in shallow parametrized quantum circuits". Nat. Commun.**12** (2021).

[20] Iris Cong, Soonwon Choi, and Mikhail D. Lukin. "Quantum convolutional neural networks". Nature Physics **15**, 1273–1278 (2019).

[21] Johannes Jakob Meyer, Marian Mularski, Elies Gil-Fuster, Antonio Anna Mele, Francesco Arzani, Alissa Wilms, and Jens Eisert. "Exploiting symmetry in variational quantum machine learning". PRX Quantum **4**, 010328 (2023).

[22] Andrea Skolik, Michele Cattelan, Sheir Yarkoni, Thomas Bäck, and Vedran Dunjko. "Equivariant quantum circuits for learning on weighted graphs". npj Quantum Information **9**, 47 (2023).

[23] Sukin Sim, Peter D. Johnson, and Alán Aspuru-Guzik. "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms". Adv. Quantum Technol. **2**, 1900070 (2019).

[24] Adrián Pérez-Salinas, Alba Cervera-Lierta, Elies Gil-Fuster, and José I. Latorre. "Data re-uploading for a universal quantum classifier". Quantum **4**, 226 (2020).

[25] Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. "Effect of data encoding on the expressive power of variational quantum-machine-learning models". Phys. Rev. A **103**, 032430 (2021).

[26] Francesco Tacchino, Stefano Mangini, Panagiotis Kl. Barkoutsos, Chiara Macchiavello, Dario Gerace, Ivano Tavernelli, and Daniele Bajoni. "Variational learning for quantum artificial neural networks". IEEE Transactions on Quantum Engineering **2**, 1–10 (2021).

[27] B Jaderberg, L W Anderson, W Xie, S Albanie, M Kiffner, and D Jaksch. "Quantum self-supervised learning". Quantum Science and Technology **7**, 035005 (2022).

[28] David A. Meyer and Nolan R. Wallach. "Global entanglement in multiparticle systems". Journal of Mathematical Physics **43**, 4273–4278 (2002).

[29] Pietro Silvi, Ferdinand Tschirsich, Matthias Gerster, Johannes Jünemann, Daniel Jaschke, Matteo Rizzi, and Simone Montangero. "The tensor networks anthology: Simulation techniques for many-body quantum lattice systems". SciPost Physics Lecture Notes (2019).

[30] S. Montangero. "Introduction to tensor network methods". Springer International Publishing. Cham, CH (2018).

[31] J. Eisert. "Entanglement and tensor network states" (2013). arXiv:1308.3318.

[32] Sebastian Paeckel, Thomas Köhler, Andreas Swoboda, Salvatore R. Manmana, Ulrich Schollwöck, and Claudius Hubig. "Time-evolution methods for matrix-product states". Annals of Physics **411**, 167998 (2019).

[33] Patrick Hayden, Debbie W. Leung, and Andreas Winter. "Aspects of Generic Entanglement". Communications in Mathematical Physics **265**, 95–117 (2006).

[34] Elizabeth S. Meckes. "The Random Matrix Theory of the Classical Compact Groups". Cambridge Tracts in Mathematics. Cambridge University Press. Cambridge (2019).

[35] Alan Edelman and N. Raj Rao. "Random matrix theory". Acta Numerica **14**, 233–297 (2005).

[36] Don N. Page. "Average entropy of a subsystem". Phys. Rev. Lett. **71**, 1291–1294 (1993).

[37] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. "The theory of variational hybrid quantum-classical algorithms". New J. Phys. **18**, 023023 (2016).

[38] Francisco Javier Gil Vidal and Dirk Oliver Theis. "Input redundancy for parameterized quantum circuits". Front. Phys. **8**, 297 (2020).

[39] E. Torrontegui and J. J. Garcia-Ripoll. "Uni-

tary quantum perceptron as efficient universal approximator". EPL **125**, 30004 (2019).

[40] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. "Barren plateaus in quantum neural network training landscapes". Nat. Commun. **9**, 4812 (2018).

[41] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. "Evaluating analytic gradients on quantum hardware". Phys. Rev. A **99**, 032331 (2019).

[42] Andrew Arrasmith, M. Cerezo, Piotr Czarnik, Lukasz Cincio, and Patrick J. Coles. "Effect of barren plateaus on gradient-free optimization". Quantum **5**, 558 (2021).

[43] Zoë Holmes, Kunal Sharma, M. Cerezo, and Patrick J. Coles. "Connecting ansatz expressibility to gradient magnitudes and barren plateaus". PRX Quantum **3**, 010313 (2022).

[44] Carlos Ortiz Marrero, Mária Kieferová, and Nathan Wiebe. "Entanglement-induced barren plateaus". PRX Quantum **2**, 040316 (2021).

[45] Samson Wang, Enrico Fontana, M. Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J. Coles. "Noise-induced barren plateaus in variational quantum algorithms". Nature Communications **12**, 6961 (2021).

[46] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. "Exact and approximate unitary 2-designs and their application to fidelity estimation". Physical Review A**80** (2009).

[47] Andrew Arrasmith, Zoë Holmes, Marco Cerezo, and Patrick J Coles. "Equivalence of quantum barren plateaus to cost concentration and narrow gorges". Quantum Science and Technology **7**, 045015 (2022).

[48] Stefan H. Sack, Raimel A. Medina, Alexios A. Michailidis, Richard Kueng, and Maksym Serbyn. "Avoiding barren plateaus using classical shadows". PRX Quantum **3**, 020365 (2022).

[49] Taylor L. Patti, Khadijeh Najafi, Xun Gao, and Susanne F. Yelin. "Entanglement devised barren plateau mitigation". Phys. Rev. Research **3**, 033090 (2021).

[50] Zi-Wen Liu, Seth Lloyd, Elton Zhu, and Huangjun Zhu. "Entanglement, quantum randomness, and complexity beyond scram-

bling". Journal of High Energy Physics **2018**, 41 (2018).

[51] Edward Grant, Leonard Wossnig, Mateusz Ostaszewski, and Marcello Benedetti. "An initialization strategy for addressing barren plateaus in parametrized quantum circuits". Quantum **3**, 214 (2019).

[52] Tyler Volkoff and Patrick J Coles. "Large gradients via correlation in random parameterized quantum circuits". Quantum Science and Technology **6**, 025008 (2021).

[53] Andrea Skolik, Jarrod R McClean, Masoud Mohseni, Patrick van der Smagt, and Martin Leib. "Layerwise learning for quantum neural networks". Quantum Machine Intelligence **3**, 1–11 (2021).

[54] Joonho Kim and Yaron Oz. "Entanglement diagnostics for efficient vqa optimization". Journal of Statistical Mechanics: Theory and Experiment **2022**, 073101 (2022).

[55] Vojtěch Havlíček, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta. "Supervised learning with quantum-enhanced feature spaces". Nature **567**, 209–212 (2019).

[56] Aram W. Harrow and Richard A. Low. "Random Quantum Circuits are Approximate 2-designs". Communications in Mathematical Physics **291**, 257–302 (2009).

[57] Jonas Haferkamp and Nicholas Hunter-Jones. "Improved spectral gaps for random quantum circuits: Large local dimensions and all-to-all interactions". Phys. Rev. A **104**, 022417 (2021).

[58] Maria Schuld. "Supervised quantum machine learning models are kernel methods" (2021) arXiv:2101.11020.

[59] Sofiene Jerbi, Lukas J Fiderer, Hendrik Poulsen Nautrup, Jonas M Kübler, Hans J Briegel, and Vedran Dunjko. "Quantum machine learning beyond kernel methods". Nature Communications **14**, 517 (2023).

[60] Seth Lloyd. "Quantum approximate optimization is computationally universal" (2018) arXiv:1812.11075.

[61] M. E. S. Morales, J. D. Biamonte, and Z. Zimborás. "On the universality of the quantum approximate optimization algorithm". Quantum Information Processing **19**, 291 (2020).

[62] Fernando G. S. L. Brandão, Aram W. Harrow, and Michał Horodecki. "Local Random Quantum Circuits are Approximate Polynomial-Designs". Communications in Mathematical Physics **346**, 397–434 (2016).

[63] Aram W Harrow and Saeed Mehraban. "Approximate unitary t-designs by short random quantum circuits using nearest-neighbor and long-range gates". Communications in Mathematical PhysicsPages 1–96 (2023).

[64] Pasquale Calabrese and John Cardy. "Evolution of entanglement entropy in one-dimensional systems". Journal of Statistical Mechanics: Theory and Experiment **2005**, P04010 (2005).

[65] Tianci Zhou and Adam Nahum. "Emergent statistical mechanics of entanglement in random unitary circuits". Phys. Rev. B **99**, 174205 (2019).

[66] Adam Nahum, Jonathan Ruhman, Sagar Vijay, and Jeongwan Haah. "Quantum entanglement growth under random unitary dynamics". Phys. Rev. X **7**, 031016 (2017).

[67] M. Aeberhard, Stefan & Forina. "Wine". UCI Machine Learning Repository (1991). DOI: https://doi.org/10.24432/C5PC7J.

[68] Milan Zwitter, Matjaz & Soklic. "Breast Cancer". UCI Machine Learning Repository (1988). DOI: https://doi.org/10.24432/C51P4M.

[69] Marko Žnidarič. "Entanglement of random vectors". Journal of Physics A: Mathematical and Theoretical **40**, F105 (2006).

[70] Daniel Jaschke and Simone Montangero. "Is quantum computing green? an estimate for an energy-efficiency quantum advantage". Quantum Science and Technology (2022).

[71] V A Marčenko and L A Pastur. "Distribution of eigenvalues for some sets of random matrices". Mathematics of the USSR-Sbornik **1**, 457 (1967).

[72] Zbigniew Puchała, Łukasz Pawela, and Karol Życzkowski. "Distinguishability of generic quantum states". Physical Review A **93**, 062112 (2016).

[73] Maxime Dupont, Nicolas Didier, Mark J. Hodson, Joel E. Moore, and Matthew J. Reagor. "Entanglement perspective on the quantum approximate optimization algorithm". Phys. Rev. A **106**, 022423 (2022).

[74] Andreas J. C. Woitzik, Panagiotis Kl. Barkoutsos, Filip Wudarski, Andreas Buchleitner, and Ivano Tavernelli. "Entanglement production and convergence properties of the variational quantum eigensolver". Phys. Rev. A **102**, 042402 (2020).

[75] Michael Ragone, Paolo Braccia, Quynh T. Nguyen, Louis Schatzki, Patrick J. Coles, Frederic Sauvage, Martin Larocca, and M. Cerezo. "Representation theory for geometric quantum machine learning" (2022) arXiv:2210.07980.

[76] Kunal Sharma, M. Cerezo, Zoë Holmes, Lukasz Cincio, Andrew Sornborger, and Patrick J. Coles. "Reformulation of the no-free-lunch theorem for entangled datasets". Phys. Rev. Lett. **128**, 070501 (2022).

[77] Martin Larocca, Nathan Ju, Diego García-Martín, Patrick J. Coles, and M. Cerezo. "Theory of overparametrization in quantum neural networks" (2021) arXiv:2109.11676.

[78] Bobak Toussi Kiani, Seth Lloyd, and Reevu Maity. "Learning unitaries by gradient descent" (2020) arXiv:2001.11897.

[79] Eric R. Anschuetz and Bobak T. Kiani. "Quantum variational algorithms are swamped with traps". Nature Communications**13** (2022).

[80] Md Sajid Anis et al. "Qiskit: An open-source framework for quantum computing". Zenodo (2021).

[81] Marco Ballarin. "Quantum computer simulation via tensor networks". Università degli Studi di Padova, Master Thesis (2021). url: https://hdl.handle.net/20.500.12608/21799.

[82] Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, M Sohaib Alam, Shahnawaz Ahmed, Juan Miguel Arrazola, Carsten Blank, Alain Delgado, Soran Jahangiri, et al. "Pennylane: Automatic differentiation of hybrid quantum-classical computations" (2018). arXiv:1811.04968.

[83] Julian Havil. "Gamma: exploring euler's constant". The Australian Mathematical SocietyPage 250 (2003). url: https://ieeexplore.ieee.org/document/9452347.

[84] Juan Carlos Garcia-Escartin and Pedro Chamorro-Posada. "Equivalent quantum circuits" (2011). arXiv:1110.2998.

[85] Karol Życzkowski and Hans-Jürgen Som-

mers. "Average fidelity between random quantum states". Phys. Rev. A **71**, 032313 (2005).

A Lower bound on entanglement entropy for unitary 2-designs

The presented derivation is a straightforward application of known results on the entanglement of random states and properties of Rényi-entropies [50, 48]. Rényi α-entropies of a density operator ρ are defined as

$$S_\alpha(\rho) = \frac{1}{1-\alpha} \log \mathrm{Tr}[\rho^\alpha], \tag{19}$$

where $\lim_{\alpha \to 1} S_\alpha(\rho) = S(\rho)$ is the Von Neumann entropy of Eq. (2), and it holds that $S_\beta(\rho) \leq S_\alpha(\rho)$ for $\beta \geq \alpha$. Of particular interest is the Rényi 2-entropy $S_2(\rho) = -\log \mathrm{Tr}[\rho^2]$ depending on the purity $\mathrm{Tr}[\rho^2]$ of the system, which is much easier to computer and it can be used to lower bound the Von Neumann entropy via $S(\rho) > S_2(\rho)$.

Let $|\psi\rangle \in (\mathbb{C}^2)^{\otimes n}$ be the state of a composite system made of subsystems A and B with dimensions $d_A = 2^{n_A}$ and $d_B = 2^{n-n_A}$, respectively. Suppose $|\psi\rangle$ is a random state $|\psi\rangle = U|\psi_0\rangle$, where U is sampled from an ensemble of unitaries that constitutes at least a unitary 2-design. Then, the average value of the purity of the reduced density matrix $\rho_A = \mathrm{Tr}_B[|\psi\rangle\langle\psi|]$ amounts to [50, 48]

$$\mathbb{E}_{\text{2-design}} \mathrm{Tr}\left[\rho_A^2\right] = \frac{d_A + d_B}{d_A d_B + 1}. \tag{20}$$

By the convexity of Rényi-entropies with respect to $\mathrm{Tr}[\rho^\alpha]$, and using Jensen's inequality ($\mathbb{E} f \geq f \mathbb{E}$), one can lower bound the average Rényi 2-entropy as $\mathbb{E}_{\text{2-design}}[S_2(\rho_A)] \geq -\log \mathbb{E}_{\text{2-design}}[\rho_A^2]$, hence

$$\mathbb{E}_{\text{2-design}}[S_2(\rho_A)] \geq -\log \frac{d_A + d_B}{d_A d_B + 1} > \log d_A - \log \frac{d_A + d_B}{d_B} > \log d_A - 1. \tag{21}$$

Then, since $S(\rho) \geq S_2(\rho) \; \forall \; \rho$, taking the expectation value on both sides yields a lower bound on the average Von Neumann entropy of ρ_A, namely

$$\log d_A - 1 < \mathbb{E}_{\text{2-design}}[S_2(\rho_A)] \leq \mathbb{E}_{\text{2-design}}[S(\rho_A)] \leq \log d_A. \tag{22}$$

which is the bound shown in Eq. (11) in the main text.

If the state $|\psi\rangle$ is instead a truly Haar-random state, that is U is sampled from the uniform Haar distribution and not just from a 2-design, the entanglement entropy is given by the Page value of Eq. (7) in the main text, which is itself lower bounded by [33]

$$\mathbb{E}_{\text{Haar}}[S(\rho_A)] > \log d_A - \frac{1}{2}\frac{d_A}{d_B} > \log d_A - \frac{1}{2} \qquad (d_A < d_B). \tag{23}$$

Summarising, for $d_A < d_B$, putting together the bounds (22) and (23) one has

$$\log d_A - 1 < \mathbb{E}_{\text{2-design}}[S(\rho_A)] < \log d_A \tag{24}$$

$$\log d_A - \frac{1}{2} < \mathbb{E}_{\text{Haar}}[S(\rho_A)] < \log d_A, \tag{25}$$

Alternatively, in the limit when the subsystem B is much larger than A, $d_B \gg d_A$, then by approximating the logarithm $\log(1+x) \approx x$ in (21) one also has

$$\log d_A - \frac{d_A}{d_B} < \mathbb{E}_{\text{2-design}}[S(\rho_A)] < \log d_A \tag{26}$$

$$\log d_A - \frac{1}{2}\frac{d_A}{d_B} < \mathbb{E}_{\text{Haar}}[S(\rho_A)] < \log d_A, \tag{27}$$

Thus, the entanglement entropy of a state sampled from a 2-design is close to that of a truly Haar-random state, with both achieving near-maximal entanglement. Of course, one also expects the Von Neumann entropy of a general t-design to be upper bounded by the Page value, $\mathbb{E}_{t\text{-design}}[S(\rho_A)] < \mathbb{E}_{\text{Haar}}[S(\rho_A)]$, with equality obtained in the limit $t \gg 1$.

B Computation of the Haar entanglement distribution

While Eq. (7) is the theoretical definition of the Haar entanglement entropy, it is not possible to exactly compute it, due to the exponential number of terms in the sum. However, it is possible to exploit the similarity of the sum with the harmonic series to obtain a good approximation. First, we denote with H_n the truncated harmonic series:

$$H_n = \sum_{k=1}^{n} \frac{1}{k}. \tag{28}$$

Then, we rewrite in a more convenient way the sum in Eq. (7):

$$\sum_{j=d_B+1}^{d_A d_B} \frac{1}{j} = \sum_{j=1}^{d_A d_B} \frac{1}{j} - \sum_{j=1}^{d_B} \frac{1}{j} = H_{d_A d_B} - H_{d_B}. \tag{29}$$

Using well-known results for the truncated Harmonic series [83]:

$$H_n = \log n + \gamma + \frac{1}{2n} - \epsilon_n, \tag{30}$$

where $\gamma \simeq 0.5772$ is the Euler-Mascheroni constant, and $0 \le \epsilon_n \le 1/8n^2$. Thus, the correction ϵ_n goes to zero as the number of terms in the sum n increases, allowing for a meaningful approximation of the value. Using this technique, we are able to estimate the Haar entanglement entropy of a 50-qubits state with an error of the order 10^{-16}.

We now proceed to compute the maximum and average of the distribution with a fixed number of qubits n. Using Eq. 7 and recalling $d_{A(B)} = 2^{n_{A(B)}}$, $n_B = n - n_A$, $n_A \in [1, n/2]$ we can write:

$$\mathbb{E}[S(\psi_A)] = H_{d_A d_B} - H_{d_B} - \frac{d_A - 1}{2 d_B} \tag{31}$$

$$= H_{2^n} - H_{2^{n-n_A}} - \frac{2^{n_A} - 1}{2^{n-n_A+1}} \tag{32}$$

$$= \log 2^{n_A} - \frac{2^{n_A} - 1}{2^{n-n_A+1}} + O\left(\frac{1}{2^{n-n_A}}\right). \tag{33}$$

We are now interested in the maximum and average of the distribution. It is easy to see that the maximum is achieved for $n_A = n/2$. In this scenario $2^{n_A} \gg 1$:

$$\max_A \left(\mathbb{E}[S(\psi_A)]\right) = \frac{n}{2} \log 2 - \frac{1}{2} + O\left(\frac{1}{2^{n/2}}\right). \tag{34}$$

Taking into account that for an n qubit system the maximum of the entanglement entropy is $S = \frac{n}{2} \log 2$ we can state that, in the large n limit, a Haar state presents a maximally entangled bond.

C Triviality of the full entangling map

The full entangling map defined as


Algorithm 1 Full entangling map

**Input:** $q_1, \ldots q_n$, qubits
Output: Quantum circuit
  1: **for** $i = 1$ to n **do**
  2:     **for** $j = i$ to n **do**
  3:         CNOT(q_i, q_j)
  4:     **end for**
 5: **end for**

---

can be shown to be equivalent to a nearest neighbors entangling map with the gates in reversed order, see Fig. 10. The proof is straightforward and obtained by direct evaluation, making use of some circuit identities for networks of CNOTs [84]. In particular, (*i*) a CNOT can be distributed into four CNOTs acting on an additional intermediate qubit
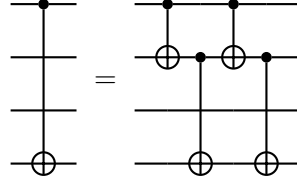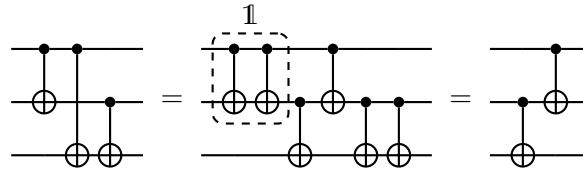
(*ii*) CNOTs having different controls and targets commute with each other
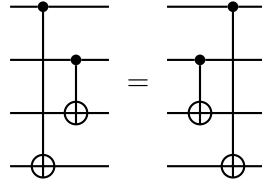
(*iii*) a cascade of CNOTs can be decomposed as
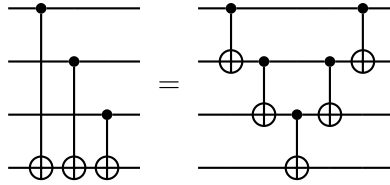
The full entangling map can be highly simplified using these three rules, reducing it to a simple sequence of nearest-neighbors interactions. For example, for $n = 3$ qubits, using (*i*) to distribute the long-range CNOT, one obtains

The simplification process can be iterated for a higher number of qubits by first commuting long range CNOTs at the end of the circuit to create a final cascade, and then making use of the result from the lower dimension case. In Fig. 10 the simplification process for $n = 4, 5$ qubits is explicitly shown, and it is directly generalized for all numbers of qubits.

Clearly, these results only hold for networks composed of plain CNOTs, and do not apply for general two-qubit interactions made of controlled unitaries.

D Expressibility of Parameterized Quantum Circuits

The expressibility introduced in [23] quantifies how well the QNN is able to explore the unitary space by comparing the distribution of fidelities of states generated by the QNN with that of randomly Haar-distributed ones.

Let $U(\boldsymbol{\phi})$ be the unitary operation implemented by a parameterized quantum circuit (PQC) with parameters $\boldsymbol{\phi}$ (in our case, we would have $\boldsymbol{\phi} = (\boldsymbol{x}, \boldsymbol{\theta})$), and be $|\psi_{\boldsymbol{\phi}}\rangle = U(\boldsymbol{\phi}) |\mathbf{0}\rangle$. Given two realizations of the PQC with parameters $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$, consider the fidelity $F = |\langle \psi_{\boldsymbol{\phi}_1} | \psi_{\boldsymbol{\phi}_2} \rangle|^2$. By repeatedly sampling
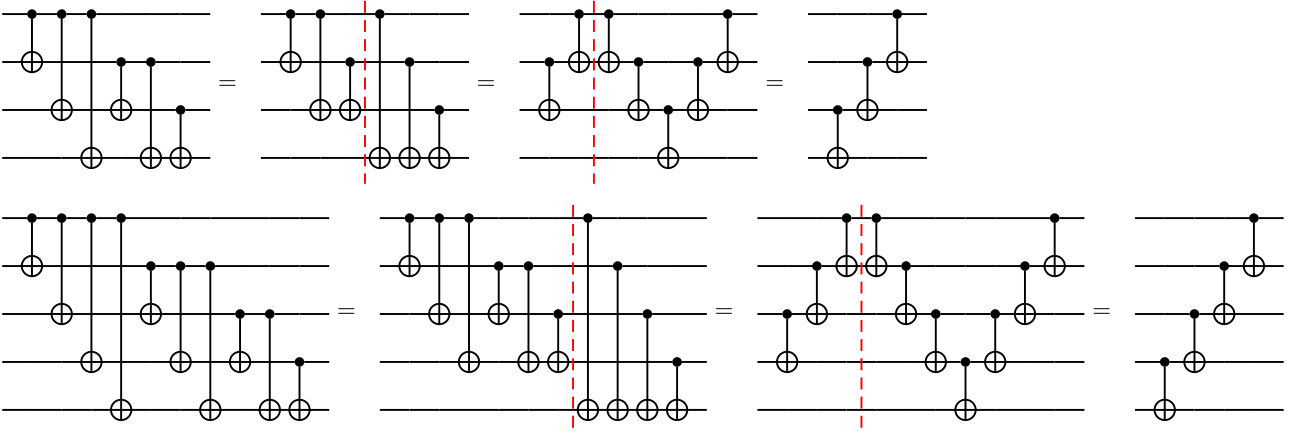
Figure 10 – Equivalence of the full entangling map with a nearest-neighbors scheme. Using the circuit identities discussed in the main text, it is straightforward to check that the all-to-all entangling scheme as defined in Alg. 1 is equivalent to a nearest-neighbors interaction.

two sets of parameters and evaluating the corresponding fidelity F, one can construct a histogram approximating the probability distribution $\hat{P}_{\mathrm{PQC}}(F)$ of the fidelity for states generated by the PQC. For Haar random quantum states, the probability density function of fidelities is known and amounts to $P_{\mathrm{Haar}}(F) = (N-1)(1-F)^{N-1}$, where $N = 2^n$ is the dimension of the Hilbert space [85].

The expressibility is then defined as the Kullback–Leibler divergence D_{KL} between the estimated fidelity distribution and that of a Haar-distributed ensemble, namely

$$\text{Expressibility} := D_{KL}\Big(\hat{P}_{\mathrm{PQC}}(F)||P_{\mathrm{Haar}}(F)\Big). \tag{35}$$

E Extensive analysis of the entanglement scaling with the increasing depth

In Fig. 11 we show the behavior of the total entanglement S_{tot} defined in Eq. (14) for four different QNNs as the depth of the quantum circuit is increased. Note that each QNN is considered with all the three possible entangling topologies (*linear*, *circular* and *full* as defined in Fig. 1), and the results are shown for several numbers of qubits $n = 4, 6, 8, 10, 12$. At last, note that all QNNs leverage the same variational form $V = C_2$, while the feature map is changed, $\mathcal{F} = C_{zz}, C_2, C_3, C_1$ for panels (a), (b), (c) and (d), respectively. See main text for comments on results.

F Convergence of MPS simulations

Using tensor network, specifically MPS, methods we perform an approximation to simulate large systems, in this work up to $n = 50$ qubits. However, the error introduced by the approximations can be monitored, so one always has an estimate of the faithfulness of the tensor network simulation [70]. Let $|\psi_{\mathrm{exact}}\rangle$ be the true state of the quantum system after the i-th two qubit gates in the circuit is applied (one qubit gates do not imply additional approximation errors), and let $|\psi_{\mathrm{trunc}}\rangle$ denote the truncated quantum state represented by the MPS. The fidelity between these two states evaluated on the i-th step of the computation is

$$F_i = |\langle\psi_{\mathrm{exact}}|\psi_{\mathrm{trunc}}\rangle|^2 = \left|\sum_{\alpha=1}^{\chi_{\mathrm{exact}}} \lambda_\alpha \langle\xi_\alpha|_1 \otimes \langle\eta_\alpha|_2 \sum_{\beta=1}^{\chi_s} \lambda_\beta |\xi_\beta\rangle_1 \otimes |\eta_\beta\rangle_2\right|^2 \tag{36}$$

$$= \left|\sum_{\alpha=1}^{\chi_s} \lambda_\alpha^2\right|^2 = \left|1 - \sum_{\alpha=\chi_s+1}^{\chi_{\mathrm{exact}}} \lambda_\alpha^2\right|^2, \tag{37}$$
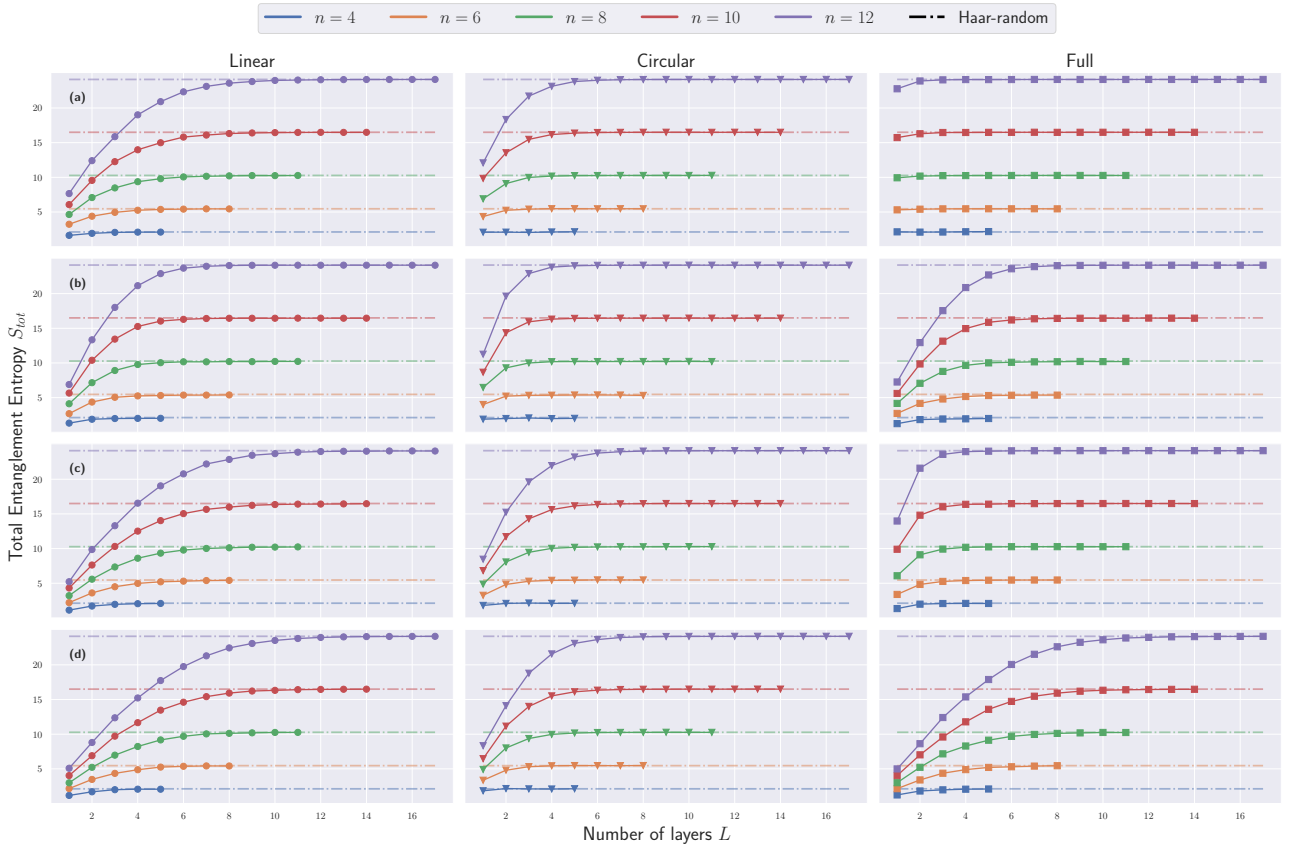
Figure 11 – Total entanglement S_{tot} (14) for four different QNN architectures, each evaluated with different entangling topologies (*linear*, *circular* and *full*), shown for increasing number of layers L and for several numbers of qubits n. QNNs architecture given by: **(a)** $\mathcal{F} = C_{zz}$, $V = C_2$; **(b)** $\mathcal{F} = C_2$, $V = C_2$; **(c)** $\mathcal{F} = C_3$, $V = C_2$; **(d)** $\mathcal{F} = C_1$, $V = C_2$. Note that QNNs leverage the same variational form V, while the feature map $\mathcal{F}$ is changed. See the main text for a discussion of the results.

where we represented the states in the Schmidt decomposition with respect to the bond where the i-th two-qubit gate was applied, and χ_s is the bond dimension of the MPS state. The fidelity F_t of the simulation after application of the t-th two-qubit gate is lower bounded by the product of the previous fidelities F_i, as [70]

$$F_t \geq \prod_{i=1}^{t-1} F_i. \tag{38}$$

where we note that the single step fidelities F_i are readily accessed during the MPS simulation, since one calculates the fidelity before the truncation of the singular values takes place. Equation (38) gives a lower bound to the error introduced by truncation in terms of the fidelity between the true state and the one evolved using an MPS simulation, and one can then control the faithfulness of the simulation at any given time step of the circuit.

In Figure 12 we show the infidelity $1 - F$ of the final state from the circuit for $n = 30, 50$ with a maximum bond dimension $\chi_s = 4096$. The plotted result is the average over $M = 10$ realization of the quantum circuit with different sets of parameters. Defining reliable results with the infidelity of at most $1 - F = 10^{-4}$ we observe that, for $n = 50$, we reliably describe circuits up to 11 layers, while for $n = 30$ we can reach $L = 12$ layers.
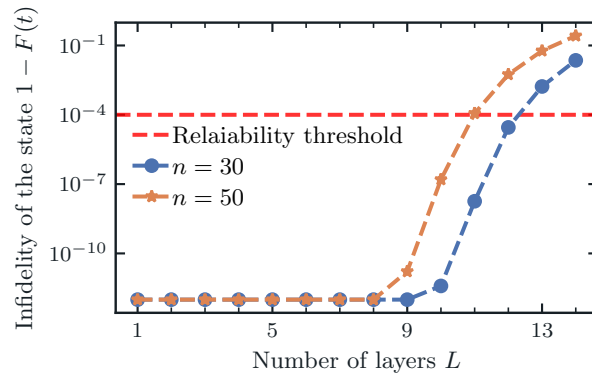
Figure 12 – We show the infidelity of the state as a function of the number of layers for $n = 30, 50$ qubits. The results are reliable up to $L = 11$ layers for $n = 50$ and up to $L = 12$ for $n = 30$.