Exploring quantumness in quantum reservoir computing

Niclas Götting, Frederik Lohof, and Christopher Gies
Institute for Theoretical Physics and Bremen Center for Computational Material Science, University of Bremen, 28359 Bremen, Germany

(Received 10 February 2023; accepted 23 October 2023; published 27 November 2023)

Quantum reservoir computing is an emerging field in machine learning with quantum systems. While classical reservoir computing has proven to be a capable concept for enabling machine learning on real, complex dynamical systems with many degrees of freedom, the advantage of its quantum analog has yet to be fully explored. Here, we establish a link between quantum properties of a quantum reservoir, namely, entanglement and its occupied phase-space dimension, and its linear short-term memory performance. We find that a high degree of entanglement in the reservoir is a prerequisite for a more complex reservoir dynamics, which is key to unlocking the exponential phase space and higher short-term memory capacity. We quantify these relations and discuss the effect of dephasing in the performance of physical quantum reservoirs.

DOI: 10.1103/PhysRevA.108.052427

I. INTRODUCTION

Machine-learning models based on artificial neural networks (ANNs) have already demonstrated their transformative potential on a global scale. These models typically rely on the optimization of thousands or even billions of parameters [1], the training of which uses excessive amounts of energy. Alternative approaches lie in the implementation of ANNs as physical systems [2]. Reservoir computing (RC) is a field that has emerged from neuromorphic computing and aims at using the natural dynamics of complex systems for information processing tasks [3]. While the capability of physical reservoir computing has been proven in several key experiments [4,5], RC with quantum-mechanical systems has only recently become a research objective [6–8]. Quantum reservoir computing (QRC) is particularly suited to be implemented on noisy intermediate-scale quantum (NISQ) hardware and is considered as an alternative to quantum machine learning using variational quantum gate logic [9]. QRC has two advantages over classical RC: (i) native processing of quantum input [10,11], and (ii) exponential scaling of the phase space with system size. In principle, the exponential scaling is able to outgrow the parameter space of any classical system, wherein lies the promise of many quantum technologies. Only now, different aspects of QRC are beginning to be investigated [12,13], such as the role of the Hilbert-space dimension, [14,15], its robustness to noise [16], the origin of nonlinearity in QRC from the underlying linear quantum dynamics [14,17,18], and the role of dissipation [6,19].

In this article we address the question in how far quantum properties can improve computing performance over classical implementations by quantifying the relationship between entanglement, the utilization of the available quantum phase space, and the linear short-term memory capacity as a measure of QRC performance. Although, in principle, the Hilbert space grows exponentially, it is *a priori* not clear how much of this available space is used for computation. To answer this question, we introduce the covariance dimension to the field of QRC as a measure of the effective phase-space dimension of the quantum reservoir dynamics. We find that the degree of entanglement in the system is directly linked to the dimension of the used phase space. Furthermore, we discuss the role of dephasing mechanisms for the complexity of the reservoir dynamics and its effect on the memory capacity of the QRC.

II. QUANTUM SYSTEMS AS RESERVOIRS

We consider a transverse-field Ising model [6] as a prototypical system that is well suited to address general questions on quantum advantage without relying on specifications of certain hardware platforms. In that sense, the N-qubit quantum reservoir is represented by the Hamiltonian ($\hbar = 1$) [6,20,21]

$$H = h \sum_{i=1}^{N} \sigma_z^{(i)} + \sum_{i \neq j} J_{ij}\, \sigma_x^{(i)} \sigma_x^{(j)}, \tag{1}$$

with $2h$ being the single-qubit energy, which we choose to be equal for all qubits, and (J_{ij}) being the symmetric qubit coupling matrix. The values J_{ij} are sampled randomly from the real interval $[-1, 1]$ and are then normalized in such a way that the maximal absolute-value eigenvalue of the matrix (J_{ij}) is given by the parameter J_0, which we refer to as the *spectral radius of the coupling* or the *coupling strength*. By changing J_0, we can consistently tune the timescale on which the system evolves, even with randomly selected couplings.

While the physical implementation of the reservoir is a system of N qubits, the amount of independent internal properties exploitable as reservoir nodes is much larger. Each spin degree of freedom and correlations thereof are affected nontrivially by the system dynamics and act as reservoir nodes. These correlations are key to unlocking the exponential scaling of the phase-space dimension of the quantum reservoir [22]. The combined properties of exponential scaling and nonclassicality are key prerequisites for a possible quantum advantage.

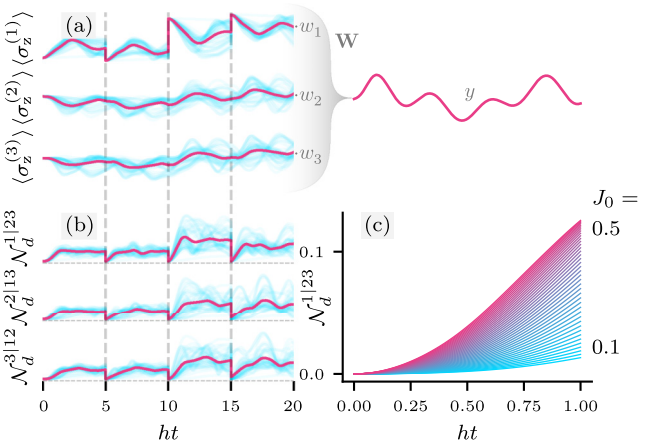

FIG. 1. Quantum dynamics of a three-qubit reservoir system. (a) Readout nodes $\langle\sigma_z^{(i)}\rangle$. The input map S_k sets the input qubit to a well-defined pure state $|\psi_{s_k}\rangle$ at time intervals $h\Delta_t = 5$. Magenta (dark) lines indicate the average over $N = 50$ randomly sampled Hamiltonians (cyan, light). The weight vector $\mathbf{W}$ combines the node trajectories into the output signal y. (b) Normalized negativity $\mathcal{N}_d$ for all possible bipartitions. After each input, the input qubit is completely disentangled from the other qubits. The unitary dynamics (re-)entangles the whole system on a timescale proportional to J_0. (c) $\mathcal{N}_d$ of partition 1|23 after input shows a larger entangling rate for increasing J_0.

To operate the QRC, a method to input information into the physical system is needed. Here, the discrete-time input signal $s_k \in [0, 1]$ is injected into the reservoir system via state initialization. In our model, this is realized mathematically by the completely positive trace preserving (CPTP) map [6]

$$\rho \mapsto |\psi_{s_k}\rangle\langle\psi_{s_k}| \otimes \mathrm{Tr}_1[\rho], \tag{2}$$

where Tr_1 denotes the partial trace over the input qubit, taken to be qubit 1, and the input-encoding pure state is given by $|\psi_{s_k}\rangle = \sqrt{1-s_k}\,|0\rangle + \sqrt{s_k}\,|1\rangle$. This operation corresponds to a projective measurement of qubit 1 and discarding the measurement outcome, and subsequently preparing the input qubit in the state $|\psi_{s_k}\rangle$. The resulting time evolution in the time interval Δ_t between two successive inputs is $\rho(t+\Delta_t) = U_{\Delta_t} S_k(\rho(t)) U_{\Delta_t}^\dagger$, where S_k is the superoperator encoding the input operation of Eq. (2), and $U_{\Delta_t} = \exp(-iH\Delta_t)$ is the unitary time evolution determined by the system Hamiltonian.

As the reservoir's readout signal we consider the expectation values of the spin components $\langle\sigma_z^{(i)}\rangle$. Their exemplary temporal behavior is shown in Fig. 1(a). Marked by the gray dashed lines are the times at which the input is injected into the first qubit. It is evident how this directly affects its state: As the input qubit is set to $|\psi_{s_k}\rangle$, the value of $\langle\sigma_z^{(1)}\rangle$ changes abruptly to the encoded input s_k. The measurement process of the $\langle\sigma_z^{(i)}\rangle$ is interpreted in an ensemble picture neglecting backaction. Protocols taking the backaction into account, either by rewinding or spatial multiplexing, as well as the influence of finite ensembles, or schemes involving weak measurements, have been put forward in the literature [23–25]. We employ a V-fold temporal multiplexing of the N readout signals by dividing the time interval between successive inputs and sampling the readout nodes at time intervals Δ_t/V. This method allows us to train on NV *virtual* readout nodes and has been shown to improve reservoir performance significantly [6]. In our experiments we choose $V = 10$. More detailed information on the technical implementation is provided in the Appendix. Furthermore, we only use the spin-z components $\langle\sigma_z^{(i)}\rangle$ of the N qubits in the network of the readout nodes for simplicity. We refrain from additionally recording two- and multiqubit correlations of the form $\langle\sigma_z^{(i)}\cdots\sigma_z^{(j)}\rangle$ [26]. In general, a variety of different state properties are thinkable as output nodes, the feasibility of which will depend on the concrete physical implementation of the reservoir [17].

The training process of the QRC in this setup is equivalent to that of a classical reservoir computer [27] in that the multiplexed readout signals are multiplied by the weight vector $\mathbf{W} = (w_0, w_1, \ldots)^\intercal$ to receive the output signal as illustrated in Fig. 1(a) (see Appendix for more details). The components of $\mathbf{W}$ are the only parameters in our QRC approach that are being trained.

III. ENTANGLEMENT IN QRC

In this article we investigate how the presence of entanglement correlates with the quantum reservoir's memory capacity as a measure of its performance. As a measure to quantify entanglement, we introduce the normalized negativity [22,28,29] with respect to the bipartition with subsystems A and B,

$$\mathcal{N}_d(\rho) = \frac{\|\rho^{T_A}\|_1 - 1}{d-1}, \tag{3}$$

where $\|\cdot\|_1$ denotes the trace norm, ρ^{T_A} is the partial transpose of ρ with respect to subsystem A, and $d = \min\{\dim(\mathcal{H}_A), \dim(\mathcal{H}_B)\}$. In comparison with the conventional definition of negativity, $\mathcal{N}_d$ has a maximal value of one irrespective of the chosen bipartition, which here facilitates better comparability (see the Appendix). $\mathcal{N}_d$ is easy to compute and provides a sufficient condition to rule out separability between two subsystems, because $\mathcal{N}_d(\rho) = 0$ for all separable states [30,31]. This implies that, while an entangled state can exhibit a negativity of zero, a finite negativity is an unambiguous sign of entanglement.

The negativity's time evolution for a QRC with input at intervals of $h\Delta_t = 5$ is shown in Fig. 1(b), averaged over 50 Hamiltonians representing randomized realizations of the quantum reservoir. At every input injection, the drop of $\mathcal{N}_d^{1|23}$ to zero is clearly visible, whereas the two other bipartitions show a finite negativity as qubits 2 and 3 remain entangled. To deal with the statistical fluctuations that come with the randomly sampled system Hamiltonians, in the following we use the average negativity $\bar{\mathcal{N}}_d$ over all bipartitions of the QRC qubits as a measure of the system's entanglement at any point in time. To obtain a single value for the negativity during the whole process of performing a memory task, the negativity is also averaged over all times, including the input and build-up stages. We use this procedure to define a measure of entanglement, which is a good indicator for the mean entanglement in the system during task execution. By tuning the coupling

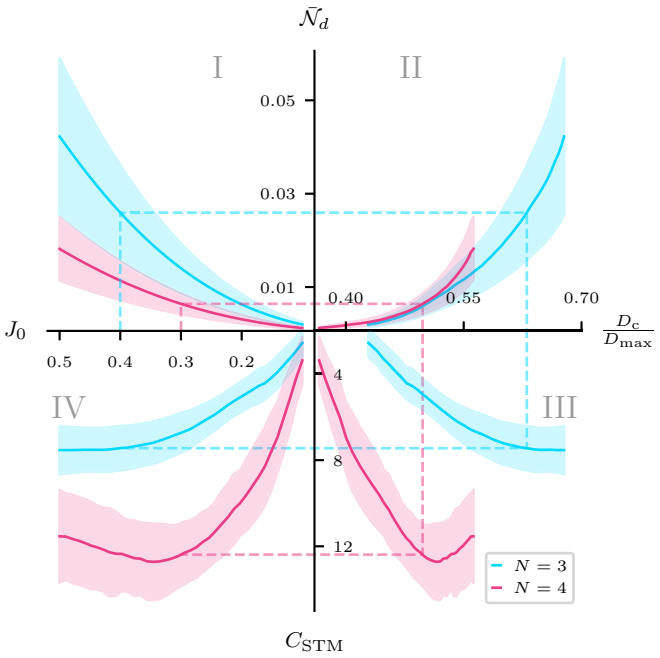

FIG. 2. Characteristic QRC properties with varying coupling strength J_0 for $N = 3$ (cyan, light) and $N = 4$ (magenta, dark) qubits averaged over 20 random realizations of the quantum reservoir. The four quadrants (I–IV) show the different functional dependence between coupling strength J_0, mean negativity $\bar{\mathcal{N}}_d$, covariance dimension D_c (given as a fraction of the theoretically available dimension D_{max}) and linear short-term memory capacity C_{STM}. Dashed lines are a guide to the eye highlighting the connection between the shown quantities. The Appendix provides an alternative representation of the same data.

strength J_0, we can control the build-up rate of entanglement, as shown in Fig. 1(c).

For 20 randomized system Hamiltonians and comparing QRC systems with $N = 3$ (cyan) and $N = 4$ (magenta) qubits corresponding to 64 and 256 internal nodes, respectively, we investigate the connection between the reservoir connectivity in terms of the spectral radius and the mean entanglement in quadrant I of Fig. 2. With J_0 increasing from 0.1 to 0.5, a monotone increase of the time-averaged mean entanglement $\bar{\mathcal{N}}_d$ is observed, underlining the connection mentioned above.

IV. PHASE-SPACE DIMENSION

QRC aims at exploiting the exponential scaling of the phase space with the system size to leverage the opportunities of NISQ devices for real-world tasks. Here, we investigate if the available phase space is used efficiently, or if the quantum dynamics is confined to a lower-dimensional manifold [32,33]. To access this information, we introduce a measure called the *covariance dimension* D_c to the field of QRC, which analyzes the trajectory of the systems' quantum dynamics in geometrical terms, a concept well known from dynamical systems theory [34]. The Appendix provides details on the calculation of D_c. Figure 3 illustrates the concept of the covariance dimension: Although the system dynamics (cyan dots) takes place in a three-dimensional (3D) phase

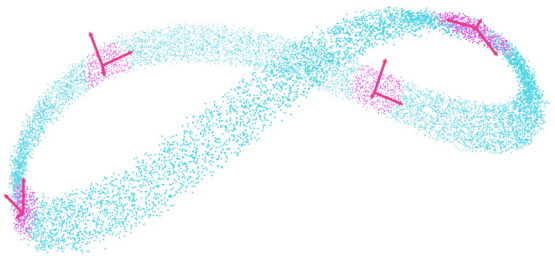

FIG. 3. Illustration of the covariance dimension using the example of a Möbius strip (cyan). While embedded in three-dimensional space, the strip itself is two dimensional, which is revealed by analysis of local properties of the system's trajectory (magenta).

space, the trajectory might be confined to a two-dimensional (2D) manifold—a Möbius strip in this example. This can be detected by analyzing the local structure (magenta clusters) of the systems' phase-space trajectories. In quadrant II of Fig. 2 we show the relation between the mean negativity $\bar{\mathcal{N}}_d$ and the fraction of the available phase-space dimension D_c/D_{max} utilized by the reservoir dynamics, where $D_{max} = 4^N - 1$ is the maximum dimension. We observe a clear positive correlation with the degree of entanglement in the network and a nearly linear increase with the coupling strength. The monotone increase is explained by the fact that stronger coupling enhances the rate of change of the system's state vector, thus allowing it to explore a higher-dimensional submanifold of the state space before collapsing again due to the input injection. Nevertheless, our analysis shows that only 40%–60% of the available phase space dimension is effectively used. Furthermore, for $N = 4$ qubits, the occupied fraction of the maximal dimension is lower than for $N = 3$ qubits, hinting at a subexponential scaling, the origin of which can most likely be attributed to an increased influence of dephasing induced by the input operation.

V. QRC PERFORMANCE

To test our initial hypothesis of a positive correlation between reservoir entanglement and QRC performance, we investigate the linear short-term memory C_{STM} as a simple but fundamental benchmark in reservoir computing [3,35]. Given an input sequence $(s_k, s_{k-1}, s_{k-2}, \dots)$, the reservoir is tasked to produce the target sequence $\hat{y}^\tau = (s_{k-\tau}, s_{k-1-\tau}, s_{k-2-\tau}, \dots)$ with $k, \tau \in \mathbb{N}$. The linear short-term memory for the time delay τ is then given by the squared Pearson correlation coefficient,

$$C^\tau_{STM} = \frac{\mathrm{cov}^2\,(y, \hat{y}^\tau)}{\sigma_y^2 \sigma^2_{\hat{y}^\tau}}, \tag{4}$$

where y is the reservoir output signal obtained after the QRC was trained on this particular task. Furthermore, σ_y is the standard deviation of y, and $\mathrm{cov}(y, \hat{y}^\tau)$ is the covariance between y and $\hat{y}^\tau$. Per definition, C^τ_{STM} lies in the interval $[0, 1]$, with 1 indicating a perfect reconstruction of the delayed input signal. As any reservoir computer has to fulfill the fading memory property [36], we can expect C^τ_{STM} to vanish for larger τ,

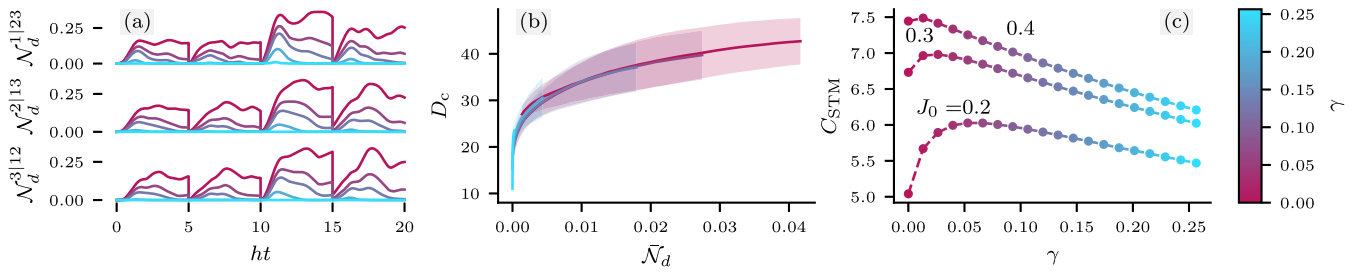

FIG. 4. (a) Normalized negativity for all three possible bipartitions in the three-qubit system averaged over 20 Hamiltonians with different coupling matrices and fixed coupling strength $J_0 = 0.4$ for different pure-dephasing rates. Input injection occurs at time intervals $h\Delta_t = 5$. (b) Relation between D_c and $\bar{\mathcal{N}}_d$ for the same dephasing rates and coupling strength as in panel (a). (c) Dependence of C_{STM} on the dephasing rate for different coupling strengths. The color bar to the right applies to all panels simultaneously.

enabling us to define the total memory capacity

$$C_{\mathrm{STM}} = \sum_{\tau=0}^{\infty} C_{\mathrm{STM}}^{\tau}. \tag{5}$$

For the perfect, noise-free system we are investigating so far, C_{STM} has to be at least 1, as the capacity C_{STM}^{0} for $\tau = 0$ is always 1. In the lower two panels of Fig. 2, we show the memory capacity of the reservoir as a function of the coupling strength J_0 (quandrant IV) and its relation to the covariance dimension D_c (quadrant III). Details on parameters such as training and test set sizes used for these results are given in the Appendix. One can see that already the weakly coupled QRC has a C_{STM} larger than one, implying an intrinsic memory capacity of the quantum network. Upon increasing the coupling strength, the mean memory capacity grows with the negativity and covariance dimension until it peaks around $C_{\mathrm{STM}} = 8$ for $N = 3$ and $C_{\mathrm{STM}} = 13$ for $N = 4$. After the initial increase, we observe a reduction of the memory capacity at larger values of the negativity, covariance dimension, and J_0. We explain this behavior by two competing effects. For small values of J_0, as explained above, the dynamics of the reservoir state between two successive inputs becomes faster, leading to a larger dimensionality (i.e., higher rank) of the multiplexed signal that is used in the STM tasks. Above $J_0 \gtrsim 0.33$, the fading memory effect becomes more pronounced, which we can infer from the Lyapunov spectrum, meaning that information loss in the QRC begins to outweigh the benefits of the quantum effects.

VI. EFFECT OF DEPHASING ON QRC PERFORMANCE

In the context of Fig. 1, we have explained that writing the input into the QRC by means of the projective measurement, expressed via the CPTP map in Eq. (2) acts as an effective dephasing mechanism that destroys both information and entanglement in the reservoir: At each input, taking the partial trace quarters the dimension of the generalized Bloch vector. Any real physical system that may serve as a hardware implementation for QRC is subject to additional degrees of dephasing due to interaction with the environment. We investigate its effect on QRC entanglement and the corresponding change in performance by introducing an additional pure-dephasing mechanism, expressed by subjecting all qubits in the QRC sequentially to the single-qubit dephasing map

$$\rho \mapsto \left(\frac{1+\mathrm{e}^{-2\gamma\Delta_t/V}}{2}\right)\rho + \left(\frac{1-\mathrm{e}^{-2\gamma\Delta_t/V}}{2}\right)\sigma_z^{(i)}\rho\sigma_z^{(i)}, \tag{6}$$

with the dephasing rate γ. In contrast to the effective dephasing induced by the input injection, this pure qubit dephasing is applied at each step of the time evolution, emulating a continuous interaction with the environment. The impact of additional pure dephasing is shown in Fig. 4(a). Its strength γ is tuned from no additional dephasing ($\gamma = 0$) to strong dephasing ($\gamma = 0.25$). As expected, stronger dephasing hinders entanglement buildup, leading to smaller values of the mean entanglement $\bar{\mathcal{N}}_d$ with increasing γ. For a quantitative analysis, Fig. 4(b) shows the relation of covariance dimension D_c and mean entanglement $\bar{\mathcal{N}}_d$ for different dephasing rates γ. We find that the relation between D_c and $\bar{\mathcal{N}}_d$ persists also in the presence of pure dephasing, i.e., dephasing decreases the mean entanglement *and* the covariance dimension to the same degree. As a result, the general shape of D_c as a function of $\bar{\mathcal{N}}_d$ changes only marginally, from which we infer a general functional dependence, the origin of which poses an open question for future work.

When investigating the linear short-term memory capacity for varying dephasing strengths, we observe an interesting effect: for most coupling strengths up to about $J_0 = 0.4$, a weak, but nonzero dephasing rate is found to increase C_{STM}, as can be seen in Fig. 4(c). While at the largest coupling strength, the effect is only marginal, it gets more pronounced for weaker coupling strengths, leading to a more than 20 % increase of the memory capacity at $J_0 = 0.2$. For such weak couplings, the STM performance of the QRC benefits even from higher dephasing rates. We conclude that stronger coupling and, with it, stronger mean entanglement and more occupied phase-space dimensions lead to better memory performance in the analyzed coupling strength interval for a fixed value of the dephasing rate.

VII. CONCLUSION

We provide results that relate the "quantumness" of a physical system to its performance as a QRC. We show that, indeed, stronger mean entanglement and more occupied phase-space dimension are beneficial to its STM performance. We find that tuning the local coupling strength within the quantum network enables one to control performance, but a

limitation is set by the balance between the speed of information spread within the QRC and the effective dephasing caused by input injection. In real physical implementations, small but nonzero additional pure dephasing can even yield a performance increase, contrasting the common perception in gate-based quantum computing and quantum machine learning that dephasing is always detrimental.

The connection between strictly quantum properties of the reservoir and their role in QRC performance stir hope for using quantum-mechanical systems in analog machine learning. For QRC to become a relevant near-term technology, we must develop a clear understanding of its potential and limitations. Future investigations will have to go beyond idealized systems and focus on actual NISQ implementations, such as ANNs based on photonic lattices or coupled-cavity arrays.

ACKNOWLEDGMENTS

This project has been supported by the Deutsche Forschungsgemeinschaft (DFG) and the Agence nationale de la recherche (ANR) via the project *PhotonicQRC* (Gi1121/6-1). F. Lohof acknowledges funding by the central research development fund (CRDF) of the University of Bremen.

APPENDIX A: ISOMETRY OF DENSITY MATRIX VECTOR SPACE TO $\mathbb{R}^d$

For efficient evaluation and analysis of the reservoir dynamics we do not solve the equation of motion for the density matrix directly. Instead, we apply a transformation to the system's density matrix ρ of N qubits that translates it into a real vector $\mathbf{x} \in \mathbb{R}^d$ with $d = 4^N$. This is well defined because the real vector space of Hermitian $2^N \times 2^N$ matrices with the metric induced by the Frobenius norm $\|\rho\|_\mathrm{F} = (\mathrm{Tr}[\rho^\dagger \rho])^{1/2}$ is isomorphic to the Euclidian metric space $\mathbb{R}^d$ with $d = 4^N$. We can see this by projecting the matrix ρ onto a generalized Pauli basis $\{P_k\}_{k=1}^{4^N} = \{\mathbb{1}, \sigma_\mathrm{x}, \sigma_\mathrm{y}, \sigma_\mathrm{z}\}^{\otimes N}$ [6]:

$$x_k = \mathrm{Tr}[\rho P_k], \quad \rho = \frac{1}{2^N} \sum_k x_k P_k, \tag{A1}$$

where x_k are the real components of the aforementioned vector $\mathbf{x}$. Indeed, we can count the degrees of freedom in ρ and find 2^N real numbers on the diagonal and $2^N(2^N - 1)/2$ nondiagonal complex entries resulting in $2^N + 2^N(2^N - 1) = 4^N$ real degrees of freedom, which is exactly the number of elements in $\mathbf{x}$. The maximum $4^N - 1$ *effective* dimensions are a result of the condition $\mathrm{Tr}\,\rho = 1$. An isometry of the different spaces is established by

$$\|\rho\|_\mathrm{F}^2 = \mathrm{Tr}[\rho^2] = \frac{1}{2^{2N}} \sum_{ij} x_i x_j \,\mathrm{Tr}[P_i P_j] = \frac{1}{2^N} \sum_i x_i^2, \tag{A2}$$

i.e., the Frobenius norm on Hermitian matrices is (up to a normalization factor) equivalent to the euclidean norm in $\mathbb{R}^d$. In Eq. (A2) we used the fact that the generalized Pauli matrices, which form the Pauli basis, fulfill the orthogonality relation $\mathrm{Tr}[P_i P_j]/2^N = \delta_{ij}$.

For density matrices representing physical states, normalization and positivity must be assured. After projection onto the Pauli basis, the vector component $x_0 = 1$ ensures normalization of the trace of the qubit state, while the components $x_{k\geqslant 1}$ can be interpreted as a N-qubit analog to the Bloch vector, which generalizes some, but not all properties of the Bloch sphere that is known from the single-qubit case. In particular, the fact that all physical pure states have a generalized Bloch vector of maximal length $\|\mathbf{x}\| = 2^{N/2}$ provides only a necessary condition because there are maximal-length vectors that do not correspond to a physical state [37] and, thus, the intuitive picture of a Bloch ball in a higher dimension with a surface representing all pure N-qubit states breaks down. The vector $\mathbf{x}$ not only contains the Bloch vectors of the reduced single-qubit states, but also correlations of the form $\langle \sigma_\mathrm{x}^{(i)} \sigma_\mathrm{x}^{(j)} \cdots \rangle$, $\langle \sigma_\mathrm{x}^{(i)} \sigma_\mathrm{y}^{(j)} \cdots \rangle$, ..., $\langle \sigma_\mathrm{z}^{(i)} \sigma_\mathrm{z}^{(j)} \cdots \rangle$.

The description of the system's state in terms of the generalized Pauli basis can also be extended to any linear superoperator $\mathcal{D} : \rho \mapsto \mathcal{D}(\rho)$ acting on a state ρ. The superoperator is projected onto the Pauli basis via

$$D_{ij} = \frac{1}{2^N} \mathrm{Tr}[P_i \mathcal{D}(P_j)], \tag{A3}$$

such that $\mathbf{x}$ is transformed under the action of the superoperator via $\mathbf{x}' = \mathbf{D}\mathbf{x}$, with $\mathbf{D} = (D_{ij})$. In particular, this includes the dynamics of the system under unitary evolution $\rho \mapsto \mathrm{e}^{-\mathrm{i}Ht} \rho \mathrm{e}^{\mathrm{i}Ht}$ which translates to the linear vector transformation $\mathbf{x}' \mapsto \mathbf{D}^t \mathbf{x}$ with $D_{ij}^t = \mathrm{Tr}[P_i \mathrm{e}^{-\mathrm{i}Ht} P_j \mathrm{e}^{\mathrm{i}Ht}]/2^N$.

APPENDIX B: TRAINING OF THE RESERVOIR SYSTEM

Here, we give further technical details on the QRC training and testing procedures that we use to obtain the short-term memory capacity C_STM. The physical output nodes of the reservoir are given by the expectation values $\langle \sigma_\mathrm{z}^{(i)} \rangle$. Multiplexing is achieved by an additional sampling of the physical nodes at intermediate time intervals Δ_t/V between two successive inputs s_k and s_{k+1} at times $k\Delta_t$ and $(k+1)\Delta_t$, respectively. Figure 5 visualizes the multiplexing scheme for $N = 3$ physical nodes and $V = 4$. Multiplexed samples are defined as $x_k^{i,j} = \langle \sigma_\mathrm{z}^{(i)} \rangle (k\Delta_t + j\Delta_t/V)$, with $k \in \mathbb{N}$, $i \in [1, N]$ and $j \in [1, V]$. Input-injection times are indicated by vertical lines (gray). The resulting $N \cdot V = 12$ virtual nodes are obtained by regrouping the subsampled physical nodes as indicated in the figure. The output of the reservoir is formed by the linear combination of all virtual nodes and a constant bias term w_0 as

$$y_k = w_0 + \sum_{l=1}^{N \cdot V} w_l x_k^l, \tag{B1}$$

where we define the index $l \in [1, N \cdot V]$ as a combination of i and j.

The reservoir is trained on an input sequence of length N_tr for a particular task by optimizing the read-out weights $\mathbf{W} = (w_0, w_1, \ldots)^\intercal$ such that the mean square error with the training sequence $\hat{y}$ is minimized, i.e.,

$$\mathbf{W} = \arg\min_{(w_0, w_1, \ldots)} \frac{1}{N_\mathrm{tr}} \sum_k [y_k(\mathbf{W}) - \hat{y}_k]^2. \tag{B2}$$

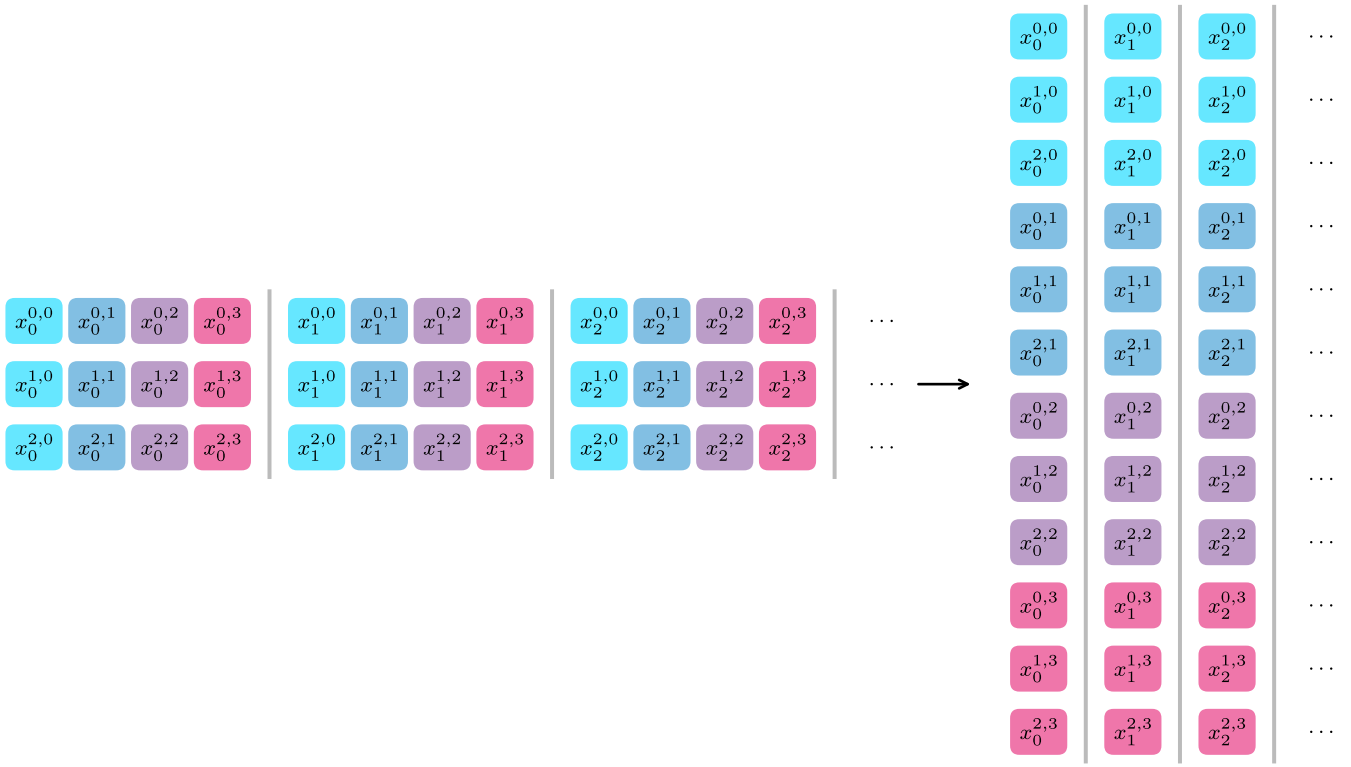

FIG. 5. Visualization of the reservoir node multiplexing. (left) Subsampling of $N = 3$ true nodes at intervals Δ_t/V between successive input steps with $V = 4$. (right) The $N \cdot V = 12$ virtual nodes are formed by regrouping the subsamples. The multiplexed samples of the physical nodes are given by $x_k^{i,j} = \langle \sigma_z^{(i)} \rangle (k\Delta_t + j\Delta_t/V)$. Gray vertical lines indicate successive inputs s_k.

For evaluation of the STM task in the paper we initially synchronize the QRC by a random input sequence of length 200 such that it becomes independent on any initial condition. Training and evaluation of the model are each performed on random input sequences of length 1000.

APPENDIX C: DERIVATION OF THE NORMALIZED NEGATIVITY

We show that the normalized negativity [28] introduced in the main text,

$$\mathcal{N}_d(\rho) = \frac{\|\rho^{T_A}\|_1 - 1}{d - 1}, \tag{C1}$$

is constructed in such a way that the maximal value of $\mathcal{N}_d(\rho)$ is bounded by 1 independent of the system size and the bipartition used. It is well known that any bipartite pure quantum state $|\psi\rangle$ has a Schmidt decomposition

$$|\psi\rangle = \sum_{i=1}^{d} \lambda_i |i_A\rangle |i_B\rangle, \tag{C2}$$

where $d = \min\{\dim(\mathcal{H}_A), \dim(\mathcal{H}_B)\}$ and the states $\{|i_A\rangle\}$ and $\{|i_B\rangle\}$ are some orthonormal basis on the Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ of the bipartition [38]. The Schmidt coefficients satisfy $\lambda_i > 0$ and $\sum_{i=1}^{d} \lambda_i^2 = 1$. Conventionally, the negativity is defined as the absolute value of the sum of the negative eigenvalues of the partial transpose ρ^{T_A}. These can be calculated as a function of the Schmidt coefficients. From

$$|\psi\rangle\langle\psi| = \sum_{i=1}^{d}\sum_{j=1}^{d} \lambda_i\lambda_j |i_A\rangle |i_B\rangle \langle j_A| \langle j_B| \tag{C3}$$

we calculate the partial transpose

$$\begin{aligned}(|\psi\rangle\langle\psi|)^{T_A} &= \sum_{i=1}^{d}\sum_{j=1}^{d} \lambda_i\lambda_j |j_A\rangle |i_B\rangle \langle i_A| \langle j_B| \qquad \text{(C4)}\\ &= \sum_{i=1}^{d} \lambda_i^2 |i_A\rangle |i_B\rangle \langle i_A| \langle i_B| \\ &\quad + \sum_{i<j}^{d} \lambda_i\lambda_j (|i_A\rangle |j_B\rangle \langle j_A| \langle i_B| \\ &\quad + |j_A\rangle |i_B\rangle \langle i_A| \langle j_B|). \qquad \text{(C5)}\end{aligned}$$

In the last step, we split diagonal and off-diagonal contributions to reveal a block-diagonal form when projected onto the basis given by $\{|i_A\rangle |j_B\rangle\}$. The first sum constitutes a diagonal block with eigenvalues λ_i^2, while each term in the second sum represents an off-diagonal 2×2-block of the form

$$\lambda_i\lambda_j \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

with eigenvalues $\pm\lambda_i\lambda_j$. Gathering all resulting negative eigenvalues shows that the *un*normalized negativity is

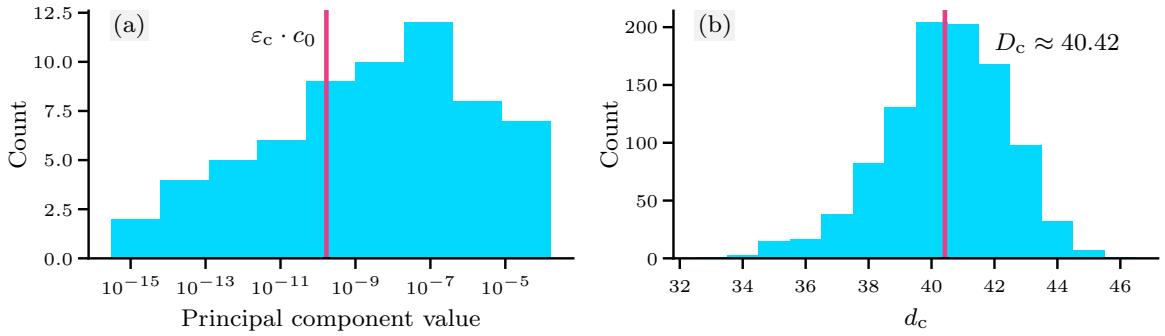

FIG. 6. (a) Principal component distribution for a single, exemplary cluster in the trajectory of a three-qubit QRC during a STM task. The magenta line indicates the cutoff threshold ($\varepsilon_c = 10^{-6}$) relative to the largest eigenvalue c_0 in the shown distribution. (b) Distribution of single-cluster covariance dimensions d_c for 1000 different clusters in the same data as in panel (a).

given by

$$\mathcal{N}(\rho) = \sum_{i<j}^{d} \lambda_i \lambda_j, \tag{C6}$$

where the sum runs over $d(d-1)/2$ individual terms. It is easy to see that under the assumption $\lambda_i > 0$, the maximum value is found when all Schmidt coefficients are equal, i.e., $\lambda_i = 1/\sqrt{d}\ \forall\ i$, which corresponds to a maximally entangled state with respect to the considered bipartition. Therefore, the maximum negativity is given by $(d-1)/2$. From this we can define the normalized negativity $\mathcal{N}_d(\rho) = \frac{2}{d-1}\mathcal{N}(\rho)$, which is equivalent to Eq. (C1). The upper bound of 1 is easily extended to mixed states by the convexity of $\mathcal{N}_d$, i.e., for any number of pure states ρ_i,

$$\mathcal{N}_d\left(\sum_i p_i \rho_i\right) \leqslant \sum_i p_i \mathcal{N}_d(\rho_i), \quad \text{with} \quad \sum_i p_i = 1, \tag{C7}$$

which quickly follows from the triangle inequality of the trace norm $\|\cdot\|_1$. We note that the result for the upper bound and the normalized negativity can be found elsewhere in the literature but we hope that our formulation provides a clear picture of the derivation.

APPENDIX D: ESTIMATION OF THE COVARIANCE DIMENSION

In the following we provide detailed information on the covariance dimension D_c that we introduce to the field of QRC in order to quantify the percentage of the theoretically available Hilbert-space dimension that is effectively utilized by the reservoir dynamics. We note that similar approaches are well known, e.g., from dynamical systems analysis [34] and also from classical reservoir computing [32]. For this work we have adapted the technique for our purpose.

In Appendix A we explained how the description of the quantum system in terms of the density matrix is translated into an equivalent picture in terms of a real state vector $\mathbf{x} \in \mathbb{R}^d$ with dimension $d = 4^N$. Accordingly, the quantum dynamics of the system corresponds to a trajectory in the space $\mathbb{R}^d$. The covariance dimension is determined by the following procedure: Let the signal $\mathbf{X} = (\mathbf{x}_0, \mathbf{x}_1, \ldots)$ be the matrix with columns $\mathbf{x}_i$ representing points in $\mathbb{R}^d$ along the trajectory of the system sampled at time intervals Δ_t/V. We choose an index point $\mathbf{x}_{i_0}$ randomly and determine a cluster of at least $d+1$ nearest neighbors that are combined into a matrix $\mathbf{X}_{i_0}$. The covariance matrix of the cluster is then given by

$$C_{\mathbf{X}_{i_0}} = \frac{1}{N_d - 1}\left(\mathbf{X}_{i_0} - \bar{\mathbf{X}}_{i_0}\right)\left(\mathbf{X}_{i_0} - \bar{\mathbf{X}}_{i_0}\right)^{\intercal}, \tag{D1}$$

where N_d is the number of points in the cluster and $\bar{\mathbf{X}}_{i_0}$ indicates the mean of the cluster. The covariance dimension $d_c(i_0)$ of each particular cluster is found by performing a principal components analysis (PCA), where we sort the principal components c_i in decreasing order $c_0 > c_1 > \cdots$. The single-cluster covariance dimension $d_c(i_0)$ is defined as the number of principal components c_i of $C_{\mathbf{X}_{i_0}}$ that pass the cutoff threshold $c_i/c_0 > \varepsilon_c$ with c_0 being the largest principle component. In general, the cutoff threshold represents the signal-to-noise ratio that we assume in the reservoir dynamics and that has an influence on the detectable covariance dimension. Here, we choose a value of $\varepsilon_c = 10^{-6}$. The covariance dimension of the whole signal is found by averaging over many random index points i_0, i.e.,

$$D_c = \frac{1}{N_I}\sum_{i_0 \in I} d_c(i_0), \tag{D2}$$

where I is the set of random indices and N_I is the number of elements in this set. The procedure of determining the covariance dimension is illustrated in Fig. 3 in the main text for a reservoir signal confined to a Möbius strip as an example. In this example the signal (cyan dots) is embedded in a three-dimensional space while the principal component analysis (PCA) of each cluster (magenta) would reveal the two dimensions of the submanifold the signal is confined to.

To provide more insight to the underlying statistics, we show typical exemplary data for these processes in Fig. 6. To obtain the data, we let a three-qubit QRC solve an arbitrary STM task and recorded the state trajectory during training and validation. In Fig. 6(a), the distribution of the principal components of a typical cluster of nearest neighbors in the trajectory is shown. The cutoff threshold is shown as a magenta vertical line. It represents the division below which we neglect the contributions of the cluster to the covariance dimension. As mentioned above, in an experimental realization the

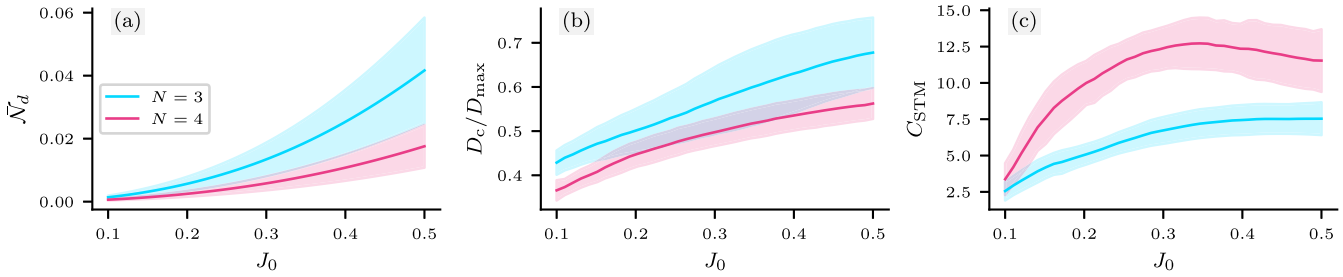

FIG. 7. Alternative representation of the data given in Fig. 2 of the main text. Characteristic QRC properties as a function of coupling strength J_0 for $N = 3, 4$ qubits averaged over 20 random QRC realizations with standard deviations given by the shaded areas. (a) Mean negativity $\bar{\mathcal{N}}_d$, (b) covariance dimension D_c (given as a fraction of the theoretically available dimension D_{max}), and (c) linear short-term memory capacity C_{STM}.

cutoff threshold is related to the signal-to-noise ratio. Statistical noise in the reservoir signal will generally lead to an increased background in the principal components. Thus, an appropriate threshold to filter out the background leading to a meaningful value of the covariance dimension will depend on the specific QRC implementation.

Determining the covariance dimension D_c of the whole trajectory $\mathbf{X}$ is done by averaging over many single-cluster covariance dimensions d_c, the distribution of which is shown in Fig. 6(b) for 1000 clusters. The small width of the distribution of d_c for individual clusters shows the consistency of our approach.

APPENDIX E: ALTERNATIVE REPRESENTATION OF QUANTUM RESERVOIR COMPUTING PROPERTIES

In Fig. 7 we provide an alternative representation of the results shown in Fig. 2 in the main text. Clearly visible is the near-linear dependence of the covariance dimension D_c when increasing the coupling strength J_0.

[1] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse *et al.*, Language models are few-shot learners, in Advances in Neural Information Processing Systems (2020), Vol. 33, pp. 1877–1901.

[2] J. Gupta and D. Koppad, Artificial neural network hardware implementation: Recent trends and applications, in *Computational Vision and Bio-Inspired Computing* (Springer International Publishing, Cham, 2020), pp. 345–354.

[3] H. Jaeger and H. Haas, Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication, Science **304**, 78 (2004).

[4] D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, Parallel photonic information processing at gigabyte per second data rates using transient states, Nat. Commun. **4**, 1364 (2013).

[5] M. Akai-Kasaya, Y. Takeshima, S. Kan, K. Nakajima, T. Oya, and T. Asai, Performance of reservoir computing in a random network of single-walled carbon nanotubes complexed with polyoxometalate, Neuromorphic Comput. Eng. **2**, 014003 (2022).

[6] K. Fujii and K. Nakajima, Harnessing disordered-ensemble quantum dynamics for machine learning, Phys. Rev. Appl. **8**, 024030 (2017).

[7] K. Nakajima, K. Fujii, M. Negoro, K. Mitarai, and M. Kitagawa, Boosting computational power through spatial multiplexing in quantum reservoir computing, Phys. Rev. Appl. **11**, 034021 (2019).

[8] D. Marković and J. Grollier, Quantum neuromorphic computing, Appl. Phys. Lett. **117**, 150501 (2020).

[9] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, Parameterized quantum circuits as machine learning models, Quantum Sci. Technol. **4**, 043001 (2019).

[10] S. Ghosh, A. Opala, M. Matuszewski, T. Paterek, and T. C. H. Liew, Quantum reservoir processing, npj Quantum Inf. **5**, 35 (2019).

[11] P. Mujal, R. Martínez-Peña, J. Nokkala, J. García-Beni, G. L. Giorgi, M. C. Soriano, and R. Zambrini, Opportunities in quantum reservoir computing and extreme learning machines, Adv. Quantum Technol. **4**, 2100027 (2021).

[12] S. A. Khan, F. Hu, G. Angelatos, and H. E. Türeci, Physical reservoir computing using finitely-sampled quantum systems, arXiv:2110.13849.

[13] R. A. Bravo, K. Najafi, X. Gao, and S. F. Yelin, Quantum reservoir computing using arrays of Rydberg atoms, PRX Quantum **3**, 030325 (2022).

[14] L. C. G. Govia, G. J. Ribeill, G. E. Rowlands, H. K. Krovi, and T. A. Ohki, Quantum reservoir computing with a single nonlinear oscillator, Phys. Rev. Res. **3**, 013077 (2021).

[15] W. D. Kalfus, G. J. Ribeill, G. E. Rowlands, H. K. Krovi, T. A. Ohki, and L. C. G. Govia, Hilbert space as a computational resource in reservoir computing, Phys. Rev. Res. **4**, 033007 (2022).

[16] J. Nokkala, R. Martínez-Peña, R. Zambrini, and M. C. Soriano, High-performance reservoir computing with fluctuations in linear networks, IEEE Trans. Neural Netw. Learn. Syst. **33**, 2664 (2022).

[17] L. C. G. Govia, G. J. Ribeill, G. E. Rowlands, and T. A. Ohki, Nonlinear input transformations are ubiquitous in quantum reservoir computing, Neuromorphic Comput. Eng. **2**, 014008 (2022).

[18] L. Innocenti, S. Lorenzo, I. Palmisano, A. Ferraro, M. Paternostro, and G. M. Palma, On the potential and limitations of quantum extreme learning machines, Comm. Phys. **6**, 118 (2023).

[19] Y. Suzuki, Q. Gao, K. C. Pradel, K. Yasuoka, and N. Yamamoto, Natural quantum reservoir computing for temporal information processing, Sci. Rep. **12**, 1353 (2022).

[20] R. B. Stinchcombe, Ising model in a transverse field. I. Basic theory, J. Phys. C: Solid State Phys. **6**, 2459 (1973).

[21] P. Pfeuty and R. J. Elliott, The Ising model with a transverse field. II. Ground state properties, J. Phys. C: Solid State Phys. **4**, 2370 (1971).

[22] K. Życzkowski, P. Horodecki, A. Sanpera, and M. Lewenstein, Volume of the set of separable states, Phys. Rev. A **58**, 883 (1998).

[23] J. Chen, H. I. Nurdin, and N. Yamamoto, Temporal information processing on noisy quantum computers, Phys. Rev. Appl. **14**, 024065 (2020).

[24] J. García-Beni, G. L. Giorgi, M. C. Soriano, and R. Zambrini, Scalable photonic platform for real-time quantum reservoir computing, Phys. Rev. Appl. **20**, 014051 (2023).

[25] P. Mujal, R. Martínez-Peña, G. L. Giorgi, M. C. Soriano, and R. Zambrini, Time series quantum reservoir computing with weak and projective measurements, Npj Quantum Inf. **9**, 16 (2023).

[26] R. Martínez-Peña, G. L. Giorgi, J. Nokkala, M. C. Soriano, and R. Zambrini, Dynamical phase transitions in quantum reservoir computing, Phys. Rev. Lett. **127**, 100502 (2021).

[27] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, Recent advances in physical reservoir computing: A review, Neural Networks **115**, 100 (2019).

[28] G. Vidal and R. F. Werner, Computable measure of entanglement, Phys. Rev. A **65**, 032314 (2002).

[29] M. B. Plenio, Logarithmic negativity: A full entanglement monotone that is not convex, Phys. Rev. Lett. **95**, 090503 (2005).

[30] M. Horodecki, P. Horodecki, and R. Horodecki, Separability of mixed states: Necessary and sufficient conditions, Phys. Lett. A **223**, 1 (1996).

[31] A. Peres, Separability criterion for density matrices, Phys. Rev. Lett. **77**, 1413 (1996).

[32] T. L. Carroll, Dimension of reservoir computers, Chaos **30**, 013102 (2020).

[33] T. L. Carroll, Low dimensional manifolds in reservoir computers, Chaos **31**, 043113 (2021).

[34] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, The analysis of observed chaotic data in physical systems, Rev. Mod. Phys. **65**, 1331 (1993).

[35] T. L. Carroll, Optimizing memory in reservoir computers, Chaos **32**, 023123 (2022).

[36] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, Information processing capacity of dynamical systems, Sci. Rep. **2**, 514 (2012).

[37] O. Gamel, Entangled Bloch spheres: Bloch matrix and two-qubit state space, Phys. Rev. A **93**, 062320 (2016).

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