



OPEN

# Online quantum time series processing with random oscillator networks

Johannes Nokkala

Reservoir computing is a powerful machine learning paradigm for online time series processing. It has reached state-of-the-art performance in tasks such as chaotic time series prediction and continuous speech recognition thanks to its unique combination of high computational power and low training cost which sets it aside from alternatives such as traditionally trained recurrent neural networks, and furthermore is amenable to implementations in dedicated hardware, potentially leading to extremely compact and efficient reservoir computers. Recently the use of random quantum systems has been proposed, leveraging the complexity of quantum dynamics for classical time series processing. Extracting the output from a quantum system without disturbing its state too much is problematic however, and can be expected to become a bottleneck in such approaches. Here we propose a reservoir computing inspired approach to online processing of time series consisting of quantum information, sidestepping the measurement problem. We illustrate its power by generalizing two paradigmatic benchmark tasks from classical reservoir computing to quantum information and introducing a task without a classical analogue where a random system is trained to both create and distribute entanglement between systems that never directly interact. Finally, we discuss partial generalizations where only the input or only the output time series is quantum.

Tasks where one time series need to be transformed into another include time series forecasting<sup>1,2</sup>, pattern generation<sup>3,4</sup> and pattern recognition<sup>5–8</sup>. In online time series processing both the given data and desired transformed data are functions of time, which separates it from approaches such as first recording the data and later processing it. Instead, the objective is to realize the time dependent function which for a given timestep and input time series up to that step returns the corresponding element of the output time series. Such tasks are also known as temporal tasks. When successful, online time series processing facilitates, e.g., the processing of arbitrarily long sequences of data since the inputs are continuously processed into outputs. This is possible in particular for tasks that can be solved by so called fading memory functions, which are functions well approximated by continuous functions of only a finite number of past inputs<sup>9</sup>. Under typically mild conditions the input time series can be used to drive random dynamical systems such that their internal variables become such fading memory functions, which can then be combined to approximate the desired output by training a simple, even linear readout function. This is known as reservoir computing (RC), which is a powerful approach to solving temporal tasks thanks to a remarkably low training cost<sup>10,11</sup> combined with state-of-the-art performance<sup>12</sup>. Furthermore, classical or quantum physical systems are also amenable to be used as the dynamical system<sup>13–15</sup>, paving the way to harvesting computational power from essentially random physical systems with fading memory and complex dynamics. In RC such systems are usually called reservoirs.

After recent seminal works investigating the suitability of the transverse-field Ising model for RC purposes<sup>16,17</sup>, there has been a surge of interest in the quantum case in particular. Indeed, the initial model has been refined and analyzed in several ways<sup>18–21</sup>, whereas new proposals have introduced RC based on quantum circuits<sup>22,23</sup>, nuclear magnetic resonance (NMR) systems<sup>24</sup> and continuous variable quantum systems<sup>25</sup>. The results have been promising, suggesting that both in the discrete and continuous variable case quantum reservoirs may have an advantage over their classical counterparts in terms of how rapidly the potential reservoir performance improves with size<sup>16,20,25</sup>. One of the biggest hurdles is in fact the extraction of the classical output from the quantum systems, since not only does a single measurement reveal only a tiny amount of information about a quantum system in an unknown state, it also alters the state and therefore competes with the inputs in driving reservoir dynamics. For certain special systems, such as NMR systems, an enormous amount of copies of the reservoir

<sup>1</sup>Department for Physics and Astronomy, University of Turku, 20014 Turun Yliopisto, Finland. <sup>2</sup>IFISC, Instituto de Física Interdisciplinar y Sistemas Complejos (UIB-CSIC) UIB Campus, 07122 Palma de Mallorca, Spain. email: jsinok@utu.fi

are naturally available which has been proposed to allow one to bypass the measurement back-action problem when collective input injections and measurements can be carried out<sup>24</sup>. In general, repeatedly initializing and subsequently measuring a quantum system extracts classical information out of it—for example, a value of an observable—however such an approach is far from ideal for time series processing for two reasons. Firstly, carrying out the repetitions anew for every element in the output time series one wishes to learn introduces severe overhead, and secondly, it is hard to imagine how such a protocol can run in an online mode, continuously producing elements of the output time series. All in all, output extraction is a major challenge in exploiting the potential of quantum reservoirs for classical time series processing.

Here we lay down an alternative, RC inspired path that can fully harness the quantumness of the reservoir while largely sidestepping the measurement problem. Namely, we introduce online time series processing with random fading memory quantum systems where both the input and desired output time series consist of quantum information. The main advantages over previous related work in quantum RC are two-fold: the role of quantumness is clearer as the tasks are by construction impossible for a classical reservoir; no measurements are required after the reservoir has been trained since the output can remain quantum. That being said the proposed scheme considers fundamentally different tasks and as such cannot replace quantum RC, therefore not solving the measurement problem of quantum RC either. Specifically, we consider random networks of interacting quantum harmonic oscillators as the reservoir. Taking inspiration from RC, we train only the interaction Hamiltonian between the network and the carriers of input information. The main difference with RC is how the output is formed; in the former it is a trained function of reservoir observables, here the output is imprinted directly on the quantum systems acting as carriers of data. This affects also the training process as will be seen.

We illustrate the possibilities of the proposed model with three different temporal tasks. The short term quantum memory (STQM) task is the quantum analog of the short term memory task<sup>26</sup> commonly used as a benchmark task in classical RC—the objective is to recall past inputs that are no longer available using the memory of the reservoir. Another common task is channel equalization task<sup>27</sup>, where the input time series is transmitted through a noisy nonlinear channel that also mixes the time series with various echoes of itself, and the objective is to recover the original time series from the distorted one. Here we generalize it to inverting the transformation caused by a quantum channel. Despite being generalizations of a classical task it will be seen that the quantum cases have notable differences. Furthermore, we introduce a task without a classical counterpart which we call the entangler. Here the objective is to create entanglement between different initially uncorrelated systems by letting each of them in turn interact once with the reservoir but never between each other. We find that all these tasks are possible to solve using random untrained networks of interacting quantum harmonic oscillators; remarkably, not even the network initial state needs to be controlled. Finally, we briefly discuss partial generalizations, i.e. cases where only the input or only the output time series is quantum.

## Results

**The model.** In RC a reservoir is a dynamical system that can be steered by an input time series to a trajectory in its state space determined by the inputs alone, i.e. its internal variables become completely determined by the input history at the limit of many inputs. If the variables can be monitored then the response of the reservoir at different timesteps can be post-processed to achieve a desired transformation from the input time series to an output time series. Importantly, for sufficiently complex reservoirs nontrivial transformations can be achieved by cheap post-processing, such as a linear combination of the variables.

Here the reservoir is a network of  $N$  unit mass quantum harmonic oscillators interacting with springlike couplings. Such units are used that  $\hbar = 1$  and  $k_B = 1$ . Let  $\mathbf{p}^\top = \{p_1, p_2, \dots, p_N\}$  and  $\mathbf{q}^\top = \{q_1, q_2, \dots, q_N\}$  be the vectors of momentum and position operators of the oscillators. The reservoir Hamiltonian  $H_R$  is

$$H_R = \frac{\mathbf{p}^\top \mathbf{p}}{2} + \frac{\mathbf{q}^\top (\Delta_\omega^2 + \mathbf{L}) \mathbf{q}}{2}, \quad (1)$$

where the diagonal matrix  $\Delta_\omega$  holds the oscillator frequencies  $\omega^\top = \{\omega_1, \omega_2, \dots, \omega_N\}$  and the symmetric matrix  $\mathbf{L}$  has elements  $L_{ij} = \delta_{ij} \sum_k g_{ik} - (1 - \delta_{ij}) g_{ij}$ . Here  $g_{ij} \geq 0$  are interaction strengths between the reservoir oscillators. Aside from oscillator frequencies there is a one-to-one correspondence between  $H_R$  and weighted simple graphs. Indeed,  $\mathbf{L}$  can be interpreted as the Laplace matrix of such a graph, and a given graph with a Laplace matrix  $\mathbf{L}$  defines  $H_R$  through Eq. (1).

We consider temporal quantum tasks (analogous to temporal tasks in RC), which we define in this work as follows. The input time series  $\mathbf{s} = \{\dots, \rho_{m-1}^I, \rho_m^I, \rho_{m+1}^I, \dots\}$  consists of quantum states  $\rho_m^I$  where  $m$  indicates the timestep. In general these can be states of multimode continuous variable quantum systems, however we assume that apart from their states the systems are identical, i.e. each system has the same Hamiltonian  $H_S$ . The input time series is processed by the reservoir into output time series  $\mathbf{o} = \{\dots, \rho_{m-1}^O, \rho_m^O, \rho_{m+1}^O, \dots\}$  by letting each system in turn interact with the reservoir for some time  $\Delta t$  according to an interaction Hamiltonian  $H_I$  coupling every reservoir oscillator to every subsystem. The order of interactions is given by the timesteps. Consequently  $\mathbf{o}$  is the image of  $\mathbf{s}$  and reservoir initial conditions under a transformation induced by the full Hamiltonian  $H = H_R + H_S + H_I$  and the interaction time  $\Delta t$ . In a temporal quantum task we attempt to realize a given transformation from  $\mathbf{s}$  to  $\mathbf{o}$  in this way. Besides the uncorrelated case one may also consider correlations between the systems at different timesteps, and we will return to this point later.

In the special case where  $H_S$  consist of  $M$  unit mass quantum harmonic oscillators and the interactions in  $H_I$  are springlike couplings,  $H$  has the same general form as  $H_R$ . The transformation induced by  $H$  and  $\Delta t$  on the operators of the reservoir and input system is now linear and can be given in terms of a symplectic matrix  $\mathbf{S}$ . Let  $\mathbf{x}_k^R$  be the form of the reservoir operators after  $k$ -th input has been processed, let  $\mathbf{x}_k^I$  be the operators of the  $k$ -th input and let  $\mathbf{x}_k^O$  be the operators of the  $k$ -th output. Now

$$\begin{pmatrix} \mathbf{x}_{k+1}^R \\ \mathbf{o}_{k+1} \\ \mathbf{x}_{k+1}^I \end{pmatrix} = \mathbf{S} \begin{pmatrix} \mathbf{x}_k^R \\ \mathbf{x}_k^I \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{x}_k^R \\ \mathbf{x}_k^I \end{pmatrix}, \quad (2)$$

where the symplectic matrix has been divided into blocks such that  $\mathbf{A}$  is  $2N \times 2N$  and  $\mathbf{D}$  is  $2M \times 2M$ . By iterating this equation we immediately get the form of both the reservoir and input modes for some timestep  $m$ :

$$\begin{cases} \mathbf{x}_m^R = \mathbf{A}^m \mathbf{x}_0^R + \sum_{k=1}^m \mathbf{A}^{m-k} \mathbf{B} \mathbf{x}_k^I, \\ \mathbf{x}_m^O = \mathbf{C} \mathbf{x}_{m-1}^R + \mathbf{D} \mathbf{x}_m^I, \end{cases} \quad (3)$$

where  $\mathbf{x}_0^R$  is the initial form of the reservoir modes. The form of the output modes for some timestep  $m$  as a function of  $\mathbf{x}_0^R$  and input history is then

$$\begin{aligned} \mathbf{x}_m^O &= \mathbf{C} \mathbf{A}^{m-1} \mathbf{x}_0^R + \mathbf{D} \mathbf{x}_m^I + \mathbf{C} \sum_{k=1}^{m-1} \mathbf{A}^{m-k-1} \mathbf{B} \mathbf{x}_k^I \\ &\approx \mathbf{D} \mathbf{x}_m^I + \mathbf{C} \sum_{k=1}^{m-1} \mathbf{A}^{m-k-1} \mathbf{B} \mathbf{x}_k^I \quad \text{when } \rho(\mathbf{A}) < 1 \text{ and } m \gg 1, \end{aligned} \quad (4)$$

where the first line is exact and the second line an approximation which holds when the spectral radius  $\rho(\mathbf{A})$ —not to be confused with quantum states—is less than 1 and enough inputs have been processed. Given that the equations of motion are conveniently expressed in terms of the operators, in the rest of this manuscript we will simply write  $\mathbf{s} = \{\dots, \mathbf{x}_{m-1}^I, \mathbf{x}_m^I, \mathbf{x}_{m+1}^I, \dots\}$  and  $\mathbf{o} = \{\dots, \mathbf{x}_{m-1}^O, \mathbf{x}_m^O, \mathbf{x}_{m+1}^O, \dots\}$ .

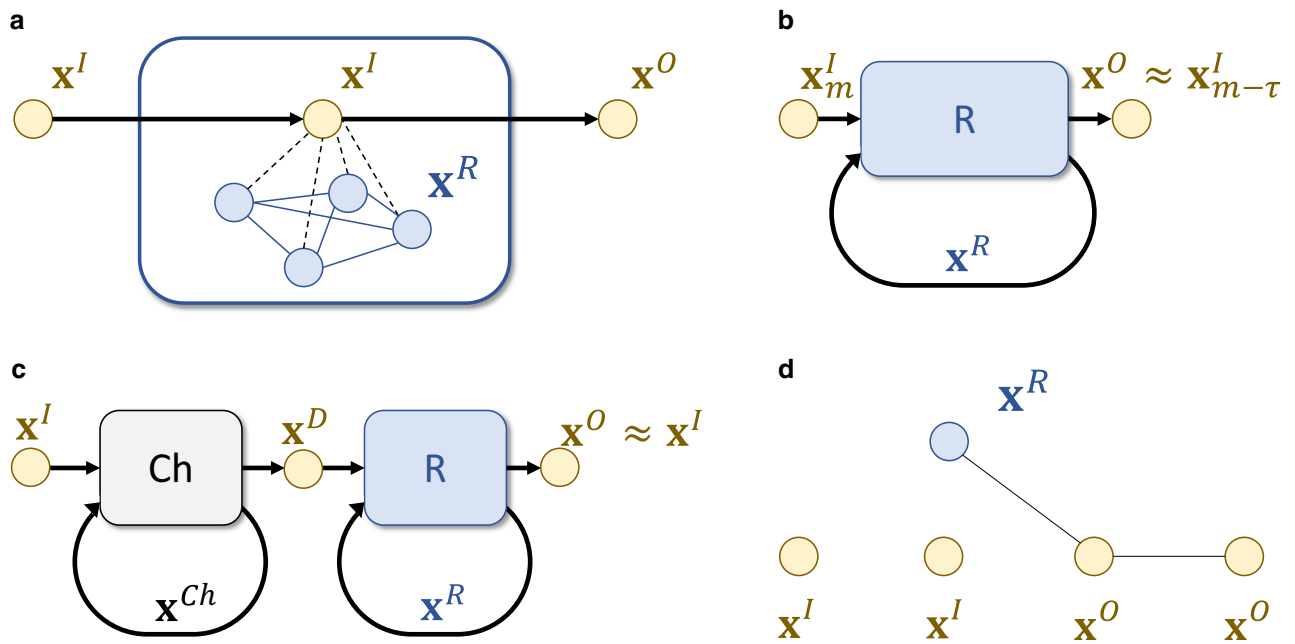
When  $\rho(\mathbf{A}) < 1$  the output time series  $\mathbf{o}$  becomes independent of the initial conditions at the limit of infinitely long input history, which is known as the echo state property<sup>1</sup> in the RC literature. In fact, it can be shown<sup>25</sup> that satisfying the spectral radius condition gives the reservoir also the so-called fading memory property<sup>9</sup>, which guarantees that  $\mathbf{x}_m^R$  and therefore  $\mathbf{x}_m^O$  become well-approximated by a continuous function of only a finite number of past inputs at the limit  $m \gg 1$ . This not only ensures that the initial conditions can be ignored but also prevents any physical quantities from diverging: the reservoir state never leaves the state space as long as all input states are physical, not even in the limit  $m \rightarrow \infty$ . At variance, if  $\rho(\mathbf{A}) \geq 1$  then, e.g., reservoir excitations may diverge. Finally, in the special case where  $\mathbf{A}$  is nilpotent for some index  $n$  there is sudden death of reservoir memory where  $\mathbf{x}_m^R$  is a function of exactly  $n$  previous inputs. As the index of a nilpotent matrix is always at most its order<sup>28</sup>,  $n \leq 2N$ .

**RC inspired quantum time series processing.** The system given in Eq. (3) can be harnessed for RC by considering only  $\mathbf{x}_m^R$  when it is assumed that the states in  $\mathbf{s}$  are in fact functions of the elements of a classical time series<sup>25</sup>. When  $\rho(\mathbf{A}) < 1$  there is fading memory and the observables of the reservoir become well-approximated by continuous functions of only a finite number of past inputs. Different transformations to a classical output time series can then be accomplished by training a simple function of the reservoir observables such as first moments, second moments or covariances of  $\mathbf{x}_m^R$ . The resulting RC can be analyzed with contemporary RC theory since the latter is agnostic to the mechanism that creates the functions of the input. Importantly,  $H_R$  is not trained and can be random since  $\rho(\mathbf{A}) < 1$  can typically be achieved by just tuning  $\Delta t$ .

The scheme for using the reservoir to process temporal quantum information instead is shown in Fig. 1a. Here  $\mathbf{s}$  itself will be the input while  $\mathbf{o}$  consists of quantum information encoded in  $\mathbf{x}_m^O$ . Thanks to fading memory, each  $\mathbf{x}_m^O$  is completely determined by  $\mathbf{x}_i^I$  where  $i \leq m$  and the initial reservoir state can be ignored. To achieve different transformations  $\mathbf{s} \mapsto \mathbf{o}$ , the matrix  $\mathbf{S}$  induced by  $H = H_R + H_S + H_I$  and  $\Delta t$  must be changed while preserving  $\rho(\mathbf{A}) < 1$ . Whereas the number of parameters in  $H$  is proportional to  $(N + M)^2$ , we take an approach inspired by RC and attempt to achieve online quantum time series processing by only training the  $NM$  interaction terms in  $H_I$ , leaving both  $H_R$  and  $H_S$  fixed. We will provide strong numerical evidence that this is in practice enough to succeed in many different temporal quantum tasks, even with random  $H_R$ .

The similarity of this result with the fact that in RC with fading memory systems it is enough to only train the final weights in the readout layer is uncanny, since the situations are quite different. Indeed, unlike in RC here reservoir dynamics cannot be separated from training since tuning  $H_I$  changes the dynamics. We are in fact not aware of any theoretical results that could be used to explain the phenomenon. In the following we provide a brief overview of the general purpose training process. See “Methods” for further details.

First the input time series  $\mathbf{s}$  is divided into three phases: the preparation, training and test phases. The role of the preparation phase is to get rid of the influence of the reservoir initial state. During the training phase the performance is checked using a cost or objective function, which varies depending on the task but for a fixed input and reservoir is completely determined by  $H_I$ . The specific forms will be introduced along with the respective tasks. The interaction Hamiltonian is varied to optimize the function using a simple stochastic function optimizer. Spectral radius condition is enforced by providing the optimizer as initial points only such  $H_I$  that the condition is satisfied; during minimization points that violate the condition can be expected to perform worse and are therefore discarded. Unless the time between inputs  $\Delta t$  is fixed by the task, training may be repeated for different choices of  $\Delta t$ . In the test phase trained  $H_I$  and best  $\Delta t$  are used and reservoir output is collected to check the performance using a task dependent figure of merit. In this way the trained reservoir is exposed to new input and must be able to generalize beyond the specific inputs in the training phase to succeed. These are the results shown in the figures.



**Figure 1.** (a) A time series of quantum systems with operators  $\mathbf{x}^I$  is processed into another time series where the transformed operators are indicated by  $\mathbf{x}^O$ . Each system interacts one after another with a random network of oscillators—later called the reservoir—with operators  $\mathbf{x}^R$ . In general the reservoir state depends on the states of systems it has interacted with, in turn making  $\mathbf{x}^O$  a function of all previous  $\mathbf{x}^I$ . Different transformations can be achieved by tuning only the interaction terms, indicated by dashed black lines. (b) In the short term quantum memory task the objective is to transform  $\mathbf{x}^I$  at timestep  $m$  into  $\mathbf{x}^I$  at timestep  $m - \tau$  where  $\tau \geq 0$  is a delay. This is possible when the sought state can be distilled out of the reservoir memory, stored in  $\mathbf{x}^R$ . (c) In the quantum channel equalization task the input is transformed by a random, uncontrollable system with operators  $\mathbf{x}^{Ch}$  into the distorted input  $\mathbf{x}^D$ . The reservoir is to recover the original input from the distorted one. (d) In the entangler task the reservoir is to entangle the systems in the time series. Like before the systems never directly interact and only one of them interacts with the reservoir at any given time. Here there is only one reservoir oscillator and only a part of the input and output time series is shown for simplicity, whereas entanglement is indicated by solid black lines.

**Examples of temporal quantum tasks.** *Parameter values used in numerical experiments.* In all numerical experiments reported throughout the section we consider as the reservoir random completely connected networks of  $N$  identical oscillators with a bare frequency  $\omega_0 = 0.25$ . Each coupling strength between the reservoir oscillators is  $g_{ij} \in [0, 0.2]$ , chosen uniformly at random. The  $M$  input modes are also such oscillators. Although in principle having  $\rho(\mathbf{A}) < 1$  is enough for fading memory, in practice we impose the limit  $\rho(\mathbf{A}) < 0.99$  to avoid issues with finite precision numerics. The influence of the initial state of the reservoir is washed out during the preparation phase and is therefore irrelevant, however in practice we use the ground state of  $H_R$ . For each different case we show results of 100 random realizations of the reservoir and when applicable, other quantities such as the input time series  $\mathbf{s}$ . In all cases the lengths of preparation, training and test phases are 40, 80 and 40, respectively. The values of the time between inputs  $\Delta t$  and the input states themselves vary and will be reported along the tasks.

**Short term quantum memory task.** The short term memory task is a paradigmatic task in classical RC where the input  $s_k$  at some timestep  $k$  is a real scalar or vector, and the target is  $s_{k-\tau}$  where  $\tau$  is the delay. Checking the performance in this task for different delays is often used to gauge how much linear memory a reservoir has. The short term quantum memory (STQM) task is its direct generalization where  $s_k$  is a state of a quantum system. Specifically,

$$\begin{cases} \mathbf{x}_k^I, & \text{(input at timestep } k) \\ \mathbf{x}_k^O \approx \mathbf{x}_{k-\tau}^I. & \text{(target at timestep } k) \end{cases} \quad (5)$$

The objective is to achieve the transformation  $\mathbf{s} \mapsto \mathbf{o}$  defined by Eq. (5) by training  $\Delta t$  and Hamiltonian  $H_I$ .  $H_R$  is assumed to be random but fixed and  $\mathbf{x}_0^R$  can be arbitrary. Unlike in the classical case where the amount of information grows linearly with input size, here the growth is more rapid as the reservoir must be able to delay also the correlations and entanglement between different input modes in  $\mathbf{x}_k^I$ . We mention in passing that the process tomography of such a delay map in the discrete variable case was considered in Ref.<sup>29</sup>.

It should be pointed out that for single-mode input states the task can in principle be done exactly for any delay  $\tau$  by concatenating  $N = \tau$  single oscillator reservoirs, which can be compared to deep RC. For  $\tau = 1$  the

states of the reservoir and input must be swapped; the required interaction strength and time can be solved analytically. Clearly using a single reservoir oscillator with double the interaction time solves the task for  $\tau = 0$  since the states are swapped twice, while using a sequence of swaps with  $N$  different single mode reservoirs achieves a delay  $\tau = N$ . For  $M > 1$  and  $\tau > 0$  the task can be done by using  $N = M\tau$  non-interacting reservoir oscillators, provided that the input systems are likewise non-interacting. Here we show that the task can in fact be solved by using a single random and fixed reservoir and letting the input interact only once with it, which can be compared to ordinary (shallow) RC.

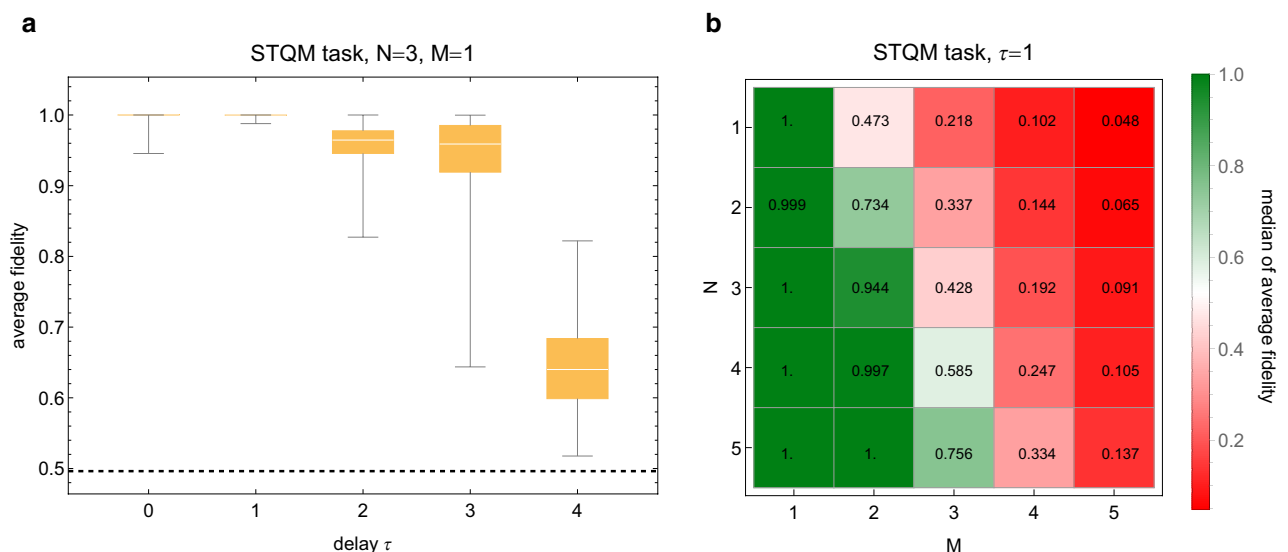
The simplicity of this task makes it amenable to a special purpose training procedure. Indeed, from Eq. (4) the following conditions to solve the task can be observed:

$$\begin{cases} \mathbf{D} \approx \mathbf{I}, \mathbf{CA}^t \mathbf{B} \approx \mathbf{0} \forall t \geq 0 & \text{if } \tau = 0, \\ \mathbf{CA}^{\tau-1} \mathbf{B} \approx \mathbf{I}, \mathbf{D} \approx \mathbf{CA}^{t \neq \tau-1} \mathbf{B} \approx \mathbf{0} & \text{if } \tau > 0. \end{cases} \quad (6)$$

When satisfied the contributions from the incorrect timesteps are suppressed while the contribution from the correct one is enhanced. The conditions in Eq. (6) for the full symplectic matrix induced by all three Hamiltonians must be satisfied by training only one of them,  $H_I$ . In practice the cost function given by Eq. (11) in “Methods” is minimized to train the reservoir. Additionally, training of  $H_I$  is repeated for  $\Delta t = 2\pi/\omega_0, 4\pi/\omega_0, \dots, 8\pi/\omega_0$  and the best value is chosen for testing phase. Notably, training is input state independent; although the training phase in  $\mathbf{s}$  could therefore be omitted, it is kept for the sake of consistency. It should be stressed that since the task is linear in the involved modes for any state, if training is successful the reservoir can delay also non-Gaussian  $M$ -mode states.

Results of numerical experiments are shown in Fig. 2. The input time series  $\mathbf{s}$  consists of random zero mean  $M$ -mode Gaussian states generated by acting with a random symplectic matrix on a thermal state of non-interacting oscillators. For details and parameter values used see “Methods”; it should be stressed that in general this creates correlated states. The reservoir parameters are as specified previously. Different values of  $N$ ,  $M$  and delay  $\tau$  are considered, whereas the figure of merit is the (Uhlmann) fidelity averaged over the test phase, possible to calculate in closed form for arbitrary Gaussian states applying, e.g., results of Ref.<sup>30</sup>.

In panel a reservoir size  $N$  and input size  $M$  have been fixed and the delay  $\tau$  is varied. The performance is excellent especially for small delays  $\tau = 0, 1, 2$  where the median values are  $\bar{F} > 0.99999$ ,  $\bar{F} > 0.9999$  and  $\bar{F} > 0.964$ , respectively. Even the worst possible performance is decent for small delays, and it is conceivable that if in their case training was repeated with a different seed for the random number generator the performance might be improved significantly. In panel b the delay is fixed and  $N$  and  $M$  are varied; the diagonal part of the array shows the case  $N = M$ . As expected, the reservoir struggles significantly with the task when  $N < M$  whereas for  $N > M$  the performance is mostly good. Although one might expect the diagonal where  $N = M$  to remain roughly constant in practice the performance is found to deteriorate as these parameters increase. This could be because the training method is not guaranteed to find the global optimum whereas the optimization problem becomes increasingly challenging as the number of parameters increases. Remarkably, a random reservoir can be trained to delay also multimode states by only tuning  $H_I$  and  $\Delta t$ .



**Figure 2.** Results for STQM task. Different reservoir sizes  $N$ , input sizes  $M$  and delays  $\tau$  are considered, and for each different set of values 100 random reservoirs were used. **(a)** Delay  $\tau$  is varied for fixed  $N$  and  $M$  and all results are shown with a box plot. The box plot shows the minimum (lower whisker), maximum (upper whisker), median (line between boxes) and the first and third quartiles (beginning of lower box and end of the upper box, respectively). The fidelity achieved by random guessing is indicated by the dashed horizontal line. **(b)** Delay  $\tau$  is fixed while both  $N$  and  $M$  are varied. Here only the median value is shown. See text for details.



**Quantum channel equalization.** This task is inspired by the channel equalization task where the input time series is distorted by a transmission through a classical channel with fading memory and the objective is to invert the transformation by the channel and restore the original time series. The quantum counterpart is presented in Fig. 1c, where the original time series consists of states of quantum systems with operators  $\mathbf{x}^I$  that are transformed by an interaction with some fixed but random system with operators  $\mathbf{x}^{Ch}$ , which constitutes the channel. We assume the channel to induce a linear transformation of the modes given by some symplectic matrix, i.e. the relevant Hamiltonians are quadratic, and furthermore that it has fading memory, which implies that  $\mathbf{x}^{Ch}$  is well-approximated by a function of a finite number of past inputs. The transformed operators of the input system are denoted by  $\mathbf{x}^D$ . A reservoir with operators  $\mathbf{x}^R$  is trained to recover the original time series from  $\mathbf{x}^D$ . The equations of motion can be recast in the same general form as before as explained in “Methods”.

It should be stressed that simply training the reservoir to perform the inverse of the channel symplectic matrix will not work since the channel acts on  $\mathbf{x}^{Ch}$  and  $\mathbf{x}^I$  but the reservoir symplectic matrix  $\mathbf{S}$  acts on  $\mathbf{x}^R$  and  $\mathbf{x}^D$ . The transformed input  $\mathbf{x}_k^D$  for some timestep  $k$  in general does not contain full information of any of the previous inputs because part of the information is in  $\mathbf{x}^{Ch}$  and in general also in the correlations between the channel and the distorted input. Since it is well known that unknown quantum states can neither be cloned nor amplified the task as given is in fact impossible—in stark contrast with its classical counterpart. To have any hope of success, the task must be modified.

Here we will do this using techniques inspired by classical RC. Instead of a single copy of the input state we will transmit a product state of  $s$  copies, but still require the reservoir to distill only a single copy of the original input, thus giving the reservoir additional quantum information to work with. We call this spatial multiplexing at order  $s$ . Additionally, we transmit the same product state  $m$  times, one copy after another, requiring the reservoir to output the original input only after the  $m$ -th copy. We call this temporal multiplexing at order  $m$ . More formally,

$$\left\{ \begin{array}{l} \bigoplus_{i=1}^s \mathbf{x}_k^I, \text{ } m \text{ copies sequentially (inputs to channel at timestep } k) \\ \mathbf{x}_{k,1}^D, \mathbf{x}_{k,2}^D, \dots, \mathbf{x}_{k,m}^D \text{ sequentially (inputs to reservoir at timestep } k) \\ \mathbf{x}_k^O \approx \mathbf{x}_k^I, \text{ (target at timestep } k) \end{array} \right. \quad (7)$$

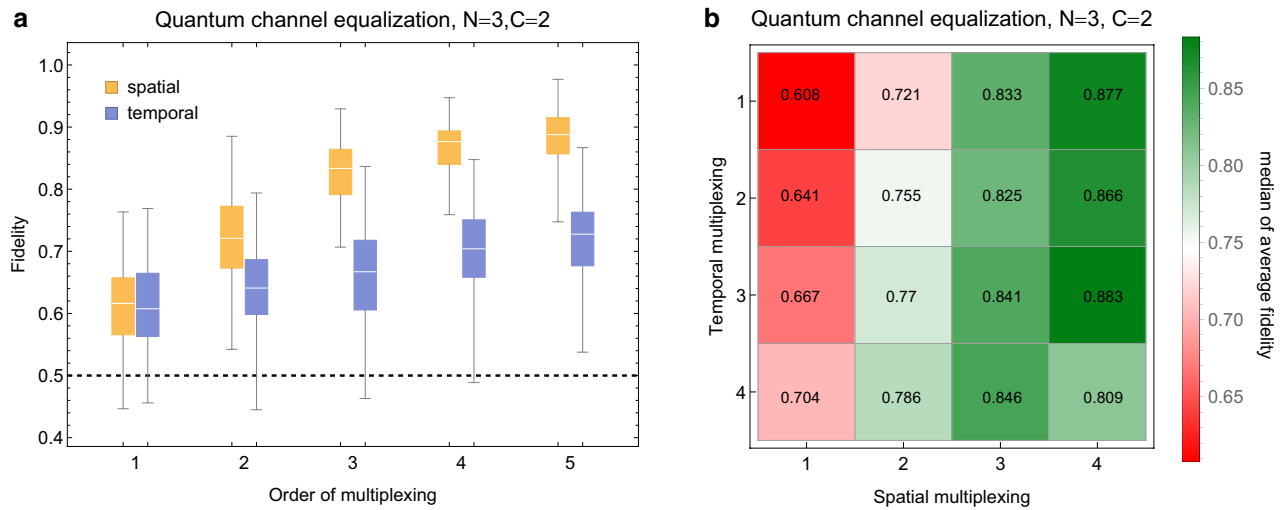
Here the timesteps are to be understood as the points where we switch from one set of identical copies to another, as determined by the original unaltered input time series  $s$ . As a final remark before moving on, if  $s$  is given but unknown neither spatial nor temporal multiplexing can be used, making the task again impossible. It must be assumed that there is a source that directly generates states according to some  $s$  and some  $m$ , or alternatively that the states are known to the sender who may then prepare the copies. It may be asked if some other modifications could help solve the task even for unknown  $s$ , but this is outside the scope of present work.

With the analysis complete for now, we move on to numerical experiments. The input consists of single mode zero mean Gaussian states. The channel has  $C = 2$  oscillators and the Hamiltonian has the same general form as the reservoir Hamiltonian. This system is taken to interact with the inputs such that the spectral radius of the relevant block in the symplectic matrix is at most 0.95. For the reservoir  $N = 3$ . The input time series consists of random zero mean Gaussian states like before in the STQM task, however for simplicity we focus only on the single mode case where  $M = 1$ . Unlike before, the time between inputs is taken to be fixed by the channel and is therefore kept at a constant value which was chosen to be  $\Delta t = 1.5\pi/\omega_0$ . Results are shown in Fig. 3.

In panel a spatial and temporal multiplexing are considered separately. At  $s = m = 1$  both reduce to the original formulation of the task, which was already concluded to be unsolvable. Still, performance exceeds that of random guessing. Performance increases quickly with  $s$  and slowly with  $m$ . Indeed, already at  $s = 2$  results tend to be better than at  $m = 5$ . That being said, unlike increasing  $m$  increasing  $s$  increases the number of terms in  $H_I$  which, e.g., makes training slower. In panel b spatial and temporal multiplexing are considered together. Curiously, performance does not always increase with  $m$  and  $s$ . For example, both  $m = 3, s = 4$  and  $m = 4, s = 3$  lead to a better performance than  $m = 4, s = 4$ . Moreover, spatial multiplexing alone achieves a performance not too far off from the best case. A possible interpretation is that since at every timestep some quantum information permanently leaves the reservoir temporal multiplexing also introduces a mechanism which hinders the task and is therefore not always beneficial.

**Entangler.** In this task the inputs  $\mathbf{s}_k$  are taken to be uncorrelated single mode systems in the vacuum state whereas the target time series consists of entangled states. Although more complicated patterns can be envisioned, here we focus on entangling systems with a fixed delay  $\tau$ . That is to say the goal is to entangle  $\mathbf{s}_k$  with  $\mathbf{s}_{k-\tau}$  by training  $H_I$ . For  $\tau = 1$  the target is then a chain of systems with nearest neighbor connections where the connections are entanglement, for  $\tau = 2$  a chain with next nearest neighbors connected and so on. Importantly, we assume that only one input interacts with the reservoir at any given timestep and all the others are unavailable, and furthermore assume that there are never any direct interactions between the systems in  $s$ . In fact, if we imagine that the systems are periodically generated by a source of vacuum states then the input  $\mathbf{s}_{k+1}$  does not even exist at some timestep  $k$ . A system or a device that can solve the task can turn the source of uncorrelated states into one of entangled states.

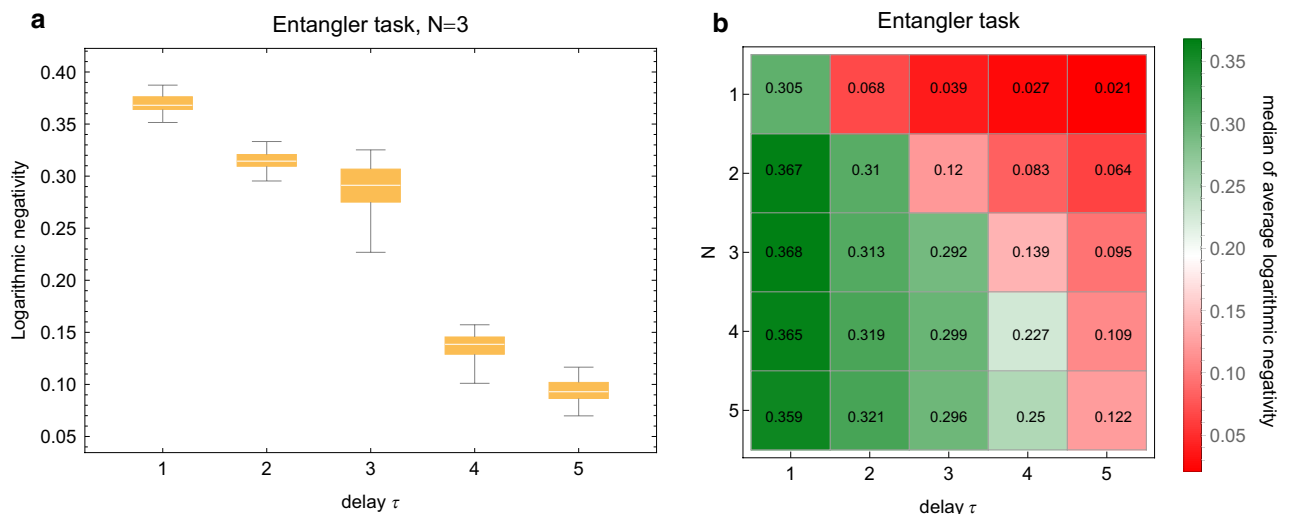
Much like in the STQM task, in certain special cases and allowing for time-dependent  $H_I$  the task can be achieved analytically and exactly, as depicted in Fig. 1d where only the case  $\tau = 1$  is considered for simplicity. At each timestep the reservoir and the ancilla are first entangled using an interaction Hamiltonian of one form and then their states are swapped using a different interaction Hamiltonian; in both cases one can analytically solve what the precise form of the interactions and interaction times must be. In fact, since the operations commute the order does not matter. Every application of the entangling gate creates a link in Fig. 1d, which is later



**Figure 3.** Results for the quantum channel equalization task. Reservoir, input and channel sizes are fixed to  $N = 3, M = 1$  and  $C = 2$ , respectively, while the orders of spatial and temporal multiplexing of the input are varied. In the former product states of identical copies of the input are used instead of single mode states, and in the latter, identical copies of the input are injected sequentially. (a) The two multiplexings are used separately with results shown with a box plot as in Fig. 2. The fidelity achieved by random guessing is indicated by the dashed horizontal line. (b) The two multiplexings are used together. In all cases results of 100 random realizations of the reservoir, input and channel have been used.

swapped to the next input system. The role of the reservoir oscillator is to provide short term quantum memory. Without it, the task becomes impossible.

Here we solve the task with the RC inspired approach by training the time-independent  $H_I$  to maximize the entanglement, as quantified by logarithmic negativity<sup>31</sup> between  $s_k$  and  $s_{k-\tau}$ . Like in the STQM task, training is repeated for  $\Delta t = 2\pi/\omega_0, 4\pi/\omega_0, \dots, 8\pi/\omega_0$  and the best value is chosen for testing phase. To succeed the reservoir must simultaneously create entanglement and re-distribute the quantum information correctly as explained previously. Results are shown in Fig. 4. In panel a the reservoir size is fixed to  $N = 3$  and the delay  $\tau$  is varied, and the logarithmic negativity averaged over the systems in the test phase is shown. The logarithmic negativity achieved for the shortest delay is between 0.35 and 0.4, which corresponds to that of a twin beam state with two-mode squeezing parameter  $0.175 \leq s \leq 0.2$ . Performance decreases slowly up to  $N = \tau$ , but then collapses for delays  $\tau > N$ , bearing a striking similarity with panel a of Fig. 2. One may interpret this as the reservoir being able to remember up to  $N$  single mode states before running out of memory. The same behaviour can be observed also in panel b where both  $N$  and delay  $\tau$  are varied and median performances are shown.



**Figure 4.** Results for entangler task. For each different case 100 random networks were created, whereas the figure of merit is average logarithmic negativity. (a) Delay is varied for fixed  $N$  and all results are shown with a box plot as in Fig. 2. Delay 1 corresponds to nearest neighbors, 2 to next nearest neighbors and so on. (b) Median logarithmic negativity when  $N$  and delay are varied. See text for details.

**Partial generalizations.** In all previously introduced tasks both the input and the output time series consist of quantum information, however one may consider partial generalizations where one of them is still classical. Here we briefly illustrate the possibilities with two simple examples.

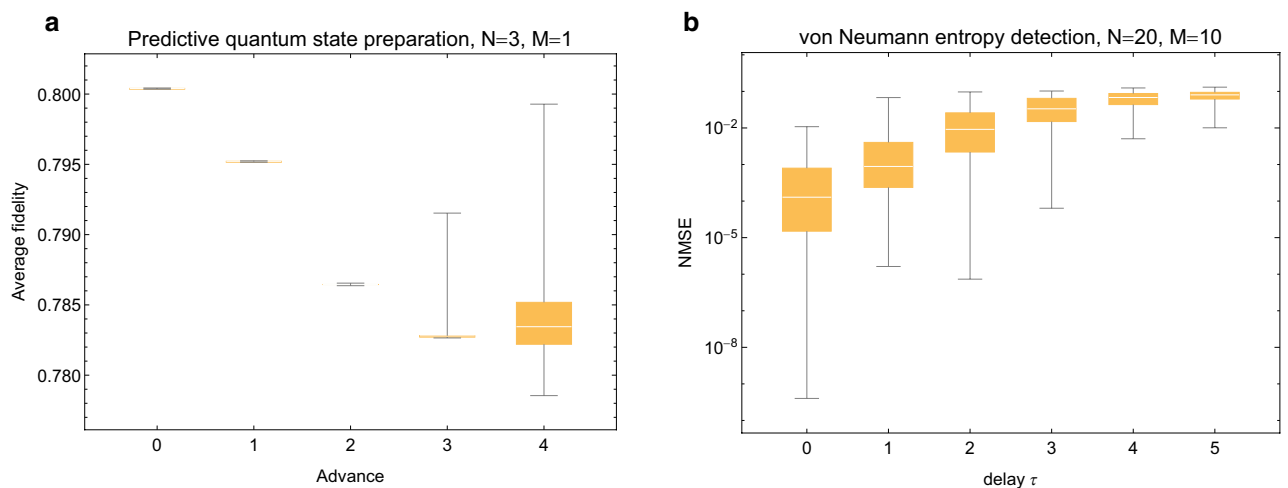
If the output is quantum but  $\mathbf{s}$  is classical, say, a time series of systems in thermal states, one may follow the framework used previously. As an example task we consider predictive quantum state preparation where the reservoir is trained to prepare a given quantum state—here, squeezed vacuum—based on future classical inputs. For arbitrary  $\mathbf{s}$  this is of course impossible, but if  $\mathbf{s}$  is at least approximately predictable then the task can be in principle solved. Here we consider the Santa Fe chaotic time series, a dataset recorded from a far-infrared laser in a chaotic state<sup>32,33</sup> often used to benchmark the predictive power of classical reservoirs. Specifically, we normalize the Santa Fe time series and consider as  $\mathbf{s}$  single mode thermal states such that the number of thermal excitations  $(n_{\text{th}})_k$  follows the normalized time series and the target is a squeezed vacuum state with a squeezing parameter  $r_k = (n_{\text{th}})_{k+a}$  where  $a \geq 0$  is the advance, or the number of timesteps in the future the reservoir must be able to predict.

Results are shown in Fig. 5a, where the average fidelity between the target state and the actual output state is shown for different values of the advance. There is very little spread for most values since only the reservoir is randomized between different realizations. Interestingly, there is an abrupt change in behavior when the advance  $a$  exceeds the number of reservoir oscillators  $N$ . Even then the fidelity remains decent but there is considerably more spread in the performance.

In the opposite case where output is classical information, say, about the properties of the states carried by the input systems in  $\mathbf{s}$ , the approach where the classical output is formed from reservoir observables can be used, as outlined previously in “The model” section. Let  $\sigma(\mathbf{x}_k^R)$  be the reservoir covariance matrix at some timestep  $k$ . It can be shown<sup>25</sup> that

$$\begin{aligned}\sigma(\mathbf{x}_m^R) &= \mathbf{A}^m \sigma(\mathbf{x}_0^R) (\mathbf{A}^\top)^m + \sum_{k=1}^m \mathbf{A}^{m-k} \mathbf{B} \sigma(\mathbf{x}_k^I) \mathbf{B}^\top (\mathbf{A}^\top)^{m-k} \\ &\approx \sum_{k=1}^m \mathbf{A}^{m-k} \mathbf{B} \sigma(\mathbf{x}_k^I) \mathbf{B}^\top (\mathbf{A}^\top)^{m-k} \quad \text{when } \rho(\mathbf{A}) < 1 \text{ and } m \gg 1,\end{aligned}\tag{8}$$

which holds for any number  $M$  of input modes. In principle, the elements of  $\sigma(\mathbf{x}_k^R)$  can be estimated by performing measurements on multiple copies of the reservoir that has processed identical inputs  $\mathbf{s}$ , which however introduces substantial overhead. Analyzing just how much overhead is incurred is beyond the scope of this work but has recently been done in Ref.<sup>34</sup> proposing also a possible implementation; in what follows, it is assumed that the exact values of the elements of  $\sigma(\mathbf{x}_k^R)$  are available. That being said, since the target is some function of reservoir observables the full Hamiltonian can remain constant, decoupling training from the dynamics. Indeed, once the elements are available multiple trained functions can be used to estimate a number of different features of  $\mathbf{s}$ .



**Figure 5.** Partial generalizations. (a) Results for predictive quantum state preparation where the input is classical but target output is quantum. The reservoir is trained to prepare the quantum state according to future classical inputs which it must deduce from previous inputs. Specifically, the inputs are thermal states where the number of thermal excitations follows the well known Santa Fe time series, and the targets are squeezed vacuum states where the squeezing parameter is to coincide with the number of thermal excitations of a future input. (b) Results for von Neumann entropy detection where the input is quantum but the target output is classical. Here the reservoir is trained to estimate the determinants of input covariance matrices—which completely determines, e.g., the von Neumann entropy—for different delays. Random product states of  $M = 10$  identical single mode states are used as input. If reservoir observables are available for a single timestep, then the von Neumann entropies of multiple previous inputs can be estimated with a very low error.



As an example task, we consider as input random single mode Gaussian states and as target  $\det(\sigma(\mathbf{x}_{k-\tau}^I))$ , or the determinant of the single mode covariance matrix for some delay  $\tau \geq 0$ . This is an important quantity that, e.g., completely determines the purity, amount of thermal excitations and the von Neumann entropy of the state. Below we give an overview of the conditions under which this task was simulated; for full details, see “Methods”.

We consider a reservoir of size  $N = 20$  and consider  $M = 10$  input modes such that the input is in a random product state of identical single mode states. Furthermore, the Hamiltonian is such that only two reservoir oscillators interact with a single input mode, and there are no interactions between these triplets of two reservoir oscillators and a single input oscillator. The interaction strengths are random but fixed and the spectral radius condition is satisfied by tuning  $\Delta t$ . One may observe from Eq. (8) that  $\sigma(\mathbf{x}_k^R)$  is linear in  $\sigma(\mathbf{x}_{k-\tau}^I)$  for any delay  $\tau$  unlike the determinant, however this problem can be overcome by considering trained linear combinations of products of pairs of elements of  $\sigma(\mathbf{x}_k^R)$ . Here the output is a trained linear combination of products of distinct pairs of the first row of  $\sigma(\mathbf{x}_k^R)$ , with training carried out as in Ref.<sup>25</sup>. Finally, unlike elsewhere, we consider preparation, training and test phases of length 500, 2000 and 500, respectively.

Results are shown in Fig. 5b, where the the normalized mean squared error (NMSE) between the actual von Neumann entropy to that computed from the determinants estimated by the reservoir is shown. As can be seen, the NMSE is very small for all considered delays, suggesting an excellent agreement between the actual and predicted value.

## Discussion

In this work we have introduced a RC inspired model for online processing of time series consisting of quantum information. Importantly, we have found that just with a judicious choice of the interaction Hamiltonian random instances of the model starting from any initial state can solve a variety of different tasks with high performance. The scheme might be further developed by considering, e.g., how training also the time between inputs can affect the performance or considering the case of non-Gaussian states or operations, which can be expected to lead to nonlinear memory<sup>35</sup> where  $\mathbf{x}_m^O$  can be nonlinear in  $\mathbf{x}_i^I$  for  $i \leq m$ . One may also consider the prospects of a proof-of-principle experimental implementation since the general form of the reservoir Hamiltonian can in principle be realized in a multimode optics platform<sup>36</sup>. We have also briefly illustrated the possibilities of partial generalizations of classical temporal tasks to cases where either the input or the target time series remains classical. Looking at the bigger picture, it is interesting to compare and contrast two distinct situations: when the output time series is to be quantum, and when it is to be classical.

In the former case the output extraction problem hindering previous related work vanishes but engineering freedom is preserved: control of only a small subset of all parameters is sufficient for high performance. Furthermore, if the reservoir Hamiltonian is random but known then measurements are not needed even in the training stage provided one can simulate the dynamics. For the considered model in particular an unknown Hamiltonian can be probed first<sup>37</sup>. It can be imagined that a classical RC augmented with a state preparation mechanism could emulate the case where input is classical, but otherwise there is genuine quantumness: in the single shot case it is clear that no classical RC can emulate its quantum counterpart since the input data would first have to be transformed to classical information. That being said, training the interaction Hamiltonian is in general somewhat costly even when the dynamics can be simulated. If simulation is not possible the cost or objective function must be estimated with measurements, which should be expected to be a very challenging optimization problem in its own right by comparing with, e.g., variational quantum algorithms<sup>38–42</sup>. Finally, another advantage of RC is lost in multitasking where the same reservoir processing the same input can simultaneously solve many different tasks by using differently trained readout functions. In the quantum case any attempts to multitask will inevitably affect performance because there is only so much quantum information for forming the output.

When instead the output is classical, multitasking is possible and the training cost is minimal. The challenges are two-fold: the output extraction problem and pinpointing what role exactly quantumness plays aside from providing a larger state space. Moreover, even if the output extraction problem can be solved, the specific way it is solved may dictate what quantum systems are ultimately suitable. As we demonstrate here with von Neumann entropy detection, the case where input is quantum might be of particular interest however, since thanks to the memory and multitasking a plethora of information concerning multiple past input states can be distilled even if the reservoir observables are known only for one or few timesteps. This may be compared to recent proposals where information of only a single quantum state is extracted with the help of a larger quantum system and supervised machine learning<sup>43–45</sup>.

Indeed, the results have created a fertile ground for further work in the direction where at least one of the time series is quantum. To the best of our knowledge there is currently little work on such temporal tasks, however the inverse problem of performing tomography of an unknown temporal quantum map has been recently considered in a spin system<sup>29</sup>. Comparisons may also be made with a recent proposal to train quantum system to induce quantum gates between qubits<sup>46</sup>; its temporal generalization might consider gates between inputs at different timesteps, for example. Indeed, temporal quantum tasks could be tackled also in the discrete variable case with e.g. spins or superconducting qubits. Since the Hilbert space dimension grows exponentially with the number of systems it is conceivable that they might permit a wider range of tasks than Gaussian states also in the case of quantum time series processing, and investigating the matter is an interesting avenue of further research. In particular, single nitrogen vacancy center spins in diamond have long coherence times whereas coherent information exchange could be mediated by polaritons as recently proposed<sup>47</sup>. One may also consider hybrid quantum systems exhibiting phenomena with potential applications in quantum information processing such as coherent perfect absorption<sup>48</sup>, higher-order exceptional points<sup>49</sup> and Kerr nonlinearity<sup>50</sup>, or quasiparticle systems such as magnons for which a proposal for a well-controlled system with tunable parameters has been made<sup>51</sup>.

## Methods

**Generation of random zero mean Gaussian states.** In the single mode case where  $M = 1$  the states may be parameterized in terms of the thermal excitations  $n_{\text{th}}$ , magnitude of squeezing  $r$  and phase of squeezing  $\varphi$ . For displacement we consistently use  $\alpha = 0$ , leading to the input first moments to vanish—the state is now completely characterized by its covariance matrix, which reads

$$\sigma(\mathbf{x}^J) = \frac{2n_{\text{th}} + 1}{2} \begin{pmatrix} (\cosh(2r) + \cos(\varphi) \sinh(2r))/\omega & \sin(\varphi) \sinh(2r) \\ \sin(\varphi) \sinh(2r) & (\cosh(2r) - \cos(\varphi) \sinh(2r))\omega \end{pmatrix}, \quad (9)$$

which is a covariance matrix of a squeezed thermal state.

In the multimode case where  $M > 1$  the state is essentially parameterized by  $M$  thermal excitations, each independently and uniformly distributed, and  $M$  squeezing parameters, also independently and uniformly distributed, in the following way. We begin from the product state of  $M$  single mode thermal states, each with their own thermal excitations. Then we act with a random basis change, apply single mode squeezing of the position to all the modes with random magnitudes, and finally act on the resulting state with another random basis change. The random basis changes are built from Haar random  $M \times M$  unitary matrices; let such a matrix be  $\mathbf{U}$ . Then by construction

$$\mathbf{O} = \begin{pmatrix} \text{Re}(\mathbf{U}) & \text{Im}(\mathbf{U}) \\ -\text{Im}(\mathbf{U}) & \text{Re}(\mathbf{U}) \end{pmatrix} \quad (10)$$

is orthogonal and also a symplectic matrix w.r.t. the chosen ordering of operators.

For both STQM and channel equalization tasks we have chosen the intervals to be  $n_{\text{th}} \in [0, 10]$  and  $r \in [0, 1]$ . In the single mode case  $\varphi \in [0, 2\pi]$ . These intervals also apply to the input states used for von Neumann entropy detection task shown in Fig. 5b. In the entangler task the input states are always single mode vacuum states.

**Training.** *Cost function minimization.* A simple stochastic function optimizer called differential evolution (DE) is used. It treats the cost function as a black box, allowing it to, e.g., attempt to optimize functions where gradients (1st derivatives) or Hessians (2nd derivatives) either do not exist or are not practical to calculate. Specifically, the implementation of Wolfram Mathematica 11.2 is used, which is described in Ref.<sup>52</sup>. Here we give an overview of the method and the parameter values used; for full details consult the reference.

DE iterates a population of points  $\{x_1, x_2, \dots, x_d\}$ . At each iteration a new population is created from the old one as follows. For each  $x_j$  in the old population, three other old points  $x_w, x_u$  and  $x_v$  are chosen randomly and a point  $x_s = x_w + s(x_u - x_v)$  is formed where  $s \in \mathbb{R}$  is a parameter called scaling factor. Then a new point  $x_j^{\text{new}}$  is created by taking each element either from  $x_j$  or  $x_s$  with probabilities  $p$  and  $1 - p$ , respectively, where the parameter  $p$  is called cross probability. Finally, the new point  $x_j^{\text{new}}$  replaces  $x_j$  if  $f(x_j^{\text{new}})$  is better than  $f(x_j)$ , where  $f$  is a given cost or objective function. The stopping criterion is met when both  $|f(x_j^{\text{new}}) - f(x_j)|$  and  $\|x_j^{\text{new}} - x_j\|$  are sufficiently small.

We initialize the population by generating  $30NM$  points, each corresponding to different interaction Hamiltonian  $H_I$  where each interaction strength  $g_{nm}$  between some reservoir oscillator  $n \in \{1, \dots, N\}$  and some input mode  $m \in \{1, \dots, M\}$  is uniformly and independently distributed in  $g_{nm} \in [0, 0.2]$  such that the spectral radius condition  $\rho(\mathbf{A}) \leq 0.99$  is satisfied. In the event that some point does not satisfy  $\rho(\mathbf{A}) \leq 0.99$  it is generated anew.

We consistently use a scaling factor of  $s = 0.05$  and a cross probability  $p = 0.4$ , i.e. rather small shifts are used to create the shifted points  $x_s$  and when forming  $x_j^{\text{new}}$  the elements are slightly more likely to be picked from  $x_s$ . We settled for these values through a simple lattice search. All other settings use the default values listed in Ref.<sup>52</sup>.

*Cost function of the STQM task.* The cost function is

$$\begin{cases} f(H_I, \Delta t) = \|\mathbf{D} - \mathbf{I}\| + 1/\|H_I\|_{\infty} & \text{if } \tau = 0, \\ f(H_I, \Delta t) = 0.5\|\mathbf{D}\| + 5\|\mathbf{C}\mathbf{A}^{\tau-1}\mathbf{B} - \mathbf{I}\| & \text{if } \tau > 0, \end{cases} \quad (11)$$

where  $\|\cdot\|$  is the Frobenius norm and where with a slight abuse of notation we have indicated by  $\|H_I\|_{\infty}$  the maximum coupling strength between a reservoir oscillator and the input oscillator(s). The point of the term  $1/\|H_I\|_{\infty}$  is to prevent the training to converge to the trivial solution  $H_I = \mathbf{0}$ . The factors 0.5 and 5 control the relative importance of minimizing the norm of  $\mathbf{D}$  and achieving  $\mathbf{C}\mathbf{A}^{\tau-1}\mathbf{B} \approx \mathbf{I}$ ; these values were chosen after some trial and error. While the function does not feature all of the relevant terms in Eq. (6), numerical experiments suggest that including more terms leads to worse results.

*Objective functions of the quantum channel equalization and entangler tasks.* Unlike in the relatively simple STQM task, in these tasks there is no obvious way to derive conditions on the reservoir symplectic matrix. This is why the objective function is the task dependent figure of merit—fidelity between reservoir output and original input in channel equalization and the logarithmic negativity in entangler—during training phase.

**Additional details about the quantum channel equalization task.** Let us write down the transformations caused by the channel and the reservoir at some timestep  $k$ . The interaction between input and channel modes induces a symplectic matrix  $S'$ . Its action on all of the relevant modes reads

$$\begin{pmatrix} \mathbf{x}_{k+1}^{Ch} \\ \mathbf{x}_k^R \\ \mathbf{x}_{k+1}^D \end{pmatrix} = \begin{pmatrix} \mathbf{A}' & \mathbf{0} & \mathbf{B}' \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{C}' & \mathbf{0} & \mathbf{D}' \end{pmatrix} \begin{pmatrix} \mathbf{x}_k^{Ch} \\ \mathbf{x}_k^R \\ \mathbf{x}_{k+1}^I \end{pmatrix}, \quad (12)$$

where  $\mathbf{S}'$  has already been divided into blocks such that  $\mathbf{A}'$  is  $C \times C$  and  $\mathbf{D}'$  is  $M \times M$ . Nothing happens to the reservoir modes since there is no interaction between the reservoir and the channel. The reservoir processes  $\mathbf{x}_{k+1}^D$  according to

$$\begin{pmatrix} \mathbf{x}_{k+1}^{Ch} \\ \mathbf{x}_{k+1}^R \\ \mathbf{x}_{k+1}^O \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{C} & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{k+1}^{Ch} \\ \mathbf{x}_k^R \\ \mathbf{x}_{k+1}^D \end{pmatrix}. \quad (13)$$

Combining these two transformations we get

$$\begin{pmatrix} \mathbf{x}_{k+1}^{Ch} \\ \mathbf{x}_{k+1}^R \\ \mathbf{x}_{k+1}^O \end{pmatrix} = \begin{pmatrix} \mathbf{A}' & \mathbf{0} & \mathbf{B}' \\ \mathbf{BC}' & \mathbf{A} & \mathbf{BD}' \\ \mathbf{DC}' & \mathbf{C} & \mathbf{DD}' \end{pmatrix} \begin{pmatrix} \mathbf{x}_k^{Ch} \\ \mathbf{x}_k^R \\ \mathbf{x}_{k+1}^I \end{pmatrix}, \quad (14)$$

where the intermediate form  $\mathbf{x}^D$  of the input modes has been eliminated. The dynamics now follows Eqs. (2)–(4) with the replacements

$$\mathbf{x}_k^R \mapsto \mathbf{x}_k^{Ch} \oplus \mathbf{x}_k^R, \quad \mathbf{A} \mapsto \begin{pmatrix} \mathbf{A}' & \mathbf{0} \\ \mathbf{BC}' & \mathbf{A} \end{pmatrix}, \quad \mathbf{B} \mapsto \begin{pmatrix} \mathbf{B}' \\ \mathbf{BD}' \end{pmatrix}, \quad \mathbf{C} \mapsto (\mathbf{DC}' \ \mathbf{C}), \quad \mathbf{D} \mapsto \mathbf{DD}', \quad (15)$$

that is to say the channel and the reservoir may be treated together as if they formed a new, larger reservoir. This simplifies the equations of motion and the simulation of the dynamics. Although one may now consider Eq. (6) to solve the task, in practice the performance is very poor because only the reservoir blocks are controllable, hence the modifications of Eq. (7).

**Additional details about the von Neumann entropy detection task.** Let  $\rho$  be a single mode Gaussian state. Then its von Neumann entropy is defined as  $S_V(\rho) = -\text{Tr}(\rho \ln(\rho))$ . It can be shown<sup>53</sup> that

$$S_V(\rho) = n_{\text{th}} \ln \left( \frac{n_{\text{th}} + 1}{n_{\text{th}}} \right) + \ln(n_{\text{th}} + 1) \quad (16)$$

where  $n_{\text{th}}$  is the amount of thermal excitations of the state  $\rho$ . This quantity in turn is connected to the determinant of the associated covariance matrix  $\sigma$  through

$$\text{Det}(\sigma) = (0.5 + n_{\text{th}})^2, \quad (17)$$

which can be seen by direct calculation starting from, e.g., Eq. (9). In Fig. 5b the actual  $S_V(\rho)$  is compared to that computed from the estimated determinant of the input covariance matrix using Eqs. (16) and (17). The minimum value of the determinant is 0.25, reached by pure states. To enforce physical values and for convenience, any estimated  $\text{Det}(\sigma) \leq 0.251$  is set to  $\text{Det}(\sigma) = 0.251$ .

## Data availability

Data is available from the corresponding author upon reasonable request. The code to reproduce the figures is available at <https://github.com/jsinok/onlinequantumtimeseries>.

Received: 4 August 2021; Accepted: 8 May 2023

Published online: 11 May 2023

## References

1. Jaeger, H. The, “echo state” approach to analysing and training recurrent neural networks—with an erratum note. *Bonn Ger. Ger. Natl. Res. Cent. Inf. Technol. GMD Tech. Rep.* **148**, 13 (2001).
2. Jaeger, H. Adaptive nonlinear system identification with echo state networks. *Adv. Neural. Inf. Process. Syst.* **15**, 609–616 (2002).
3. Jaeger, H. *Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the “echo state network” approach*, vol. 5 (GMD-Forschungszentrum Informationstechnik Bonn, 2002).
4. Hauser, H., Ijspeert, A. J., Fuchslin, R. M., Pfeifer, R. & Maass, W. The role of feedback in morphological computation with compliant bodies. *Biol. Cybern.* **106**, 595–613 (2012).
5. Verstraeten, D., Schrauwen, B., Stroobandt, D. & Van Campenhout, J. Isolated word recognition with the liquid state machine: a case study. *Inf. Process. Lett.* **95**, 521–528 (2005).
6. Soh, H. & Demiris, Y. Iterative temporal learning and prediction with the sparse online echo state gaussian process. In *The 2012 international joint conference on neural networks (IJCNN)*, 1–8 (IEEE, 2012).
7. Paquot, Y. *et al.* Optoelectronic reservoir computing. *Sci. Rep.* **2**, 1–6 (2012).
8. Jalalvand, A., Van Wallendael, G. & Van de Walle, R. Real-time reservoir computing network-based systems for detection tasks on visual contents. In *2015 7th International Conference on Computational Intelligence, Communication Systems and Networks*, 146–151 (IEEE, 2015).

9. Boyd, S. & Chua, L. Fading memory and the problem of approximating nonlinear operators with volterra series. *IEEE Trans. Circ. Syst.* **32**, 1150–1161 (1985).
10. Lukoševičius, M. A practical guide to applying echo state networks. In *Neural networks: Tricks of the trade*, 659–686 (Springer, 2012).
11. Butcher, J. B., Verstraeten, D., Schrauwen, B., Day, C. R. & Haycock, P. W. Reservoir computing and extreme learning machines for non-linear time-series data analysis. *Neural Netw.* **38**, 76–89. <https://doi.org/10.1016/j.neunet.2012.11.011> (2013).
12. Schrauwen, B., Verstraeten, D. & Van Campenhout, J. An overview of reservoir computing: theory, applications and implementations. In *Proceedings of the 15th european symposium on artificial neural networks*, p. 471–482 2007, 471–482 (2007).
13. Tanaka, G. *et al.* Recent advances in physical reservoir computing: A review. *Neural Netw.* **115**, 100–123 (2019).
14. Mujal, P. *et al.* Opportunities in quantum reservoir computing and extreme learning machines. *Adv. Quant. Technol.* 2100027. <https://doi.org/10.1002/qute.202100027> (2021).
15. Nakajima, K. & Fischer, I. *Reservoir Computing* (Springer, Singapore, 2021).
16. Fujii, K. & Nakajima, K. Harnessing disordered-ensemble quantum dynamics for machine learning. *Phys. Rev. Appl.* **8**, 024030. <https://doi.org/10.1103/PhysRevApplied.8.024030> (2017).
17. Chen, J. & Nurdin, H. I. Learning nonlinear input-output maps with dissipative quantum systems. *Quantum Inf. Process.* **18**, 198. <https://doi.org/10.1007/s11128-019-2311-9> (2019).
18. Nakajima, K., Fujii, K., Negoro, M., Mitarai, K. & Kitagawa, M. Boosting computational power through spatial multiplexing in quantum reservoir computing. *Phys. Rev. Appl.* **11**, 034021. <https://doi.org/10.1103/PhysRevApplied.11.034021> (2019).
19. Kutvonen, A., Fujii, K. & Sagawa, T. Optimizing a quantum reservoir computer for time series prediction. *Sci. Rep.* **10**, 14687. <https://doi.org/10.1038/s41598-020-71673-9> (2020).
20. Martínez-Peña, R., Nokkala, J., Giorgi, G. L., Zambrini, R. & Soriano, M. C. Information processing capacity of spin-based quantum reservoir computing systems. *Cognit. Comput.* 1–12. <https://doi.org/10.1007/s12559-020-09772-y> (2020).
21. Martínez-Peña, R., Giorgi, G. L., Nokkala, J., Soriano, M. C. & Zambrini, R. Dynamical phase transitions in quantum reservoir computing. *Phys. Rev. Lett.* **127**, 100502 (2021).
22. Chen, J., Nurdin, H. I. & Yamamoto, N. Towards single-input single-output nonlinear system identification and signal processing on near-term quantum computers. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, 401–406. <https://doi.org/10.1109/CDC40024.2019.9029180> (2019).
23. Chen, J., Nurdin, H. I. & Yamamoto, N. Temporal information processing on noisy quantum computers. *Phys. Rev. Appl.* **14**, 024065. <https://doi.org/10.1103/PhysRevApplied.14.024065> (2020).
24. Negoro, M., Mitarai, K., Fujii, K., Nakajima, K. & Kitagawa, M. Machine learning with controllable quantum dynamics of a nuclear spin ensemble in a solid. [arXiv:1806.10910](https://arxiv.org/abs/1806.10910) (2018).
25. Nokkala, J. *et al.* Gaussian states of continuous-variable quantum systems provide universal and versatile reservoir computing. *Comm. Phys.* **4**, 1–11 (2021).
26. Jaeger, H. Short term memory in echo state networks. gmd-report 152. In *GMD-German National Research Institute for Computer Science* (2002), <http://www.faculty.jacobs-university.de/hjaeger/pubs/STMEchoStatesTechRep.pdf> (Citeseer, 2002).
27. Mathews, V. J. & Lee, J. Adaptive algorithms for bilinear filtering. In *Advanced Signal Processing: Algorithms, Architectures, and Implementations V*, vol. 2296, 317–327 (International Society for Optics and Photonics, 1994).
28. Zhang, F. *Matrix theory: basic results and techniques* (Springer, Berlin, 2011).
29. Tran, Q. H. & Nakajima, K. Learning temporal quantum tomography. *Phys. Rev. Lett.* **127**, 260401 (2021).
30. Banchi, L., Braunstein, S. L. & Pirandola, S. Quantum fidelity for arbitrary gaussian states. *Phys. Rev. Lett.* **115**, 260501 (2015).
31. Vidal, G. & Werner, R. F. Computable measure of entanglement. *Phys. Rev. A* **65**, 032314 (2002).
32. Huebner, U., Abraham, N. & Weiss, C. Dimensions and entropies of chaotic intensity pulsations in a single-mode far-infrared NH<sub>3</sub> laser. *Phys. Rev. A* **40**, 6354 (1989).
33. Weigend, A. S. & Geršhenfeld, N. A. Results of the time series prediction competition at the santa fe institute. In *IEEE international conference on neural networks*, 1786–1793 (IEEE, 1993).
34. Garcia-Beni, J., Giorgi, G. L., Soriano, M. C. & Zambrini, R. Scalable photonic platform for real-time quantum reservoir computing. [arXiv:2207.14031](https://arxiv.org/abs/2207.14031) (2022).
35. Braunstein, S. L. & Van Loock, P. Quantum information with continuous variables. *Rev. Mod. Phys.* **77**, 513 (2005).
36. Nokkala, J. *et al.* Reconfigurable optical implementation of quantum complex networks. *New J. Phys.* **20**, 053024 (2018).
37. Nokkala, J., Galve, F., Zambrini, R., Mascalco, S. & Piilo, J. Complex quantum networks as structured environments: Engineering and probing. *Sci. Rep.* **6**, 1–7 (2016).
38. McClean, J. R., Babbush, R., Love, P. J. & Aspuru-Guzik, A. Exploiting locality in quantum computation for quantum chemistry. *J. Phys. Chem. Lett.* **5**, 4368–4380 (2014).
39. Wecker, D., Hastings, M. B. & Troyer, M. Progress towards practical quantum variational algorithms. *Phys. Rev. A* **92**, 042303 (2015).
40. Babbush, R. *et al.* Low-depth quantum simulation of materials. *Phys. Rev. X* **8**, 011044 (2018).
41. Cai, Z. Resource estimation for quantum variational simulations of the hubbard model. *Phys. Rev. Appl.* **14**, 014059 (2020).
42. Garcia-Pérez, G. *et al.* Learning to measure: Adaptive informationally complete generalized measurements for quantum algorithms. *Prx Quant.* **2**, 040342 (2021).
43. Ghosh, S., Opala, A., Matuszewski, M., Paterek, T. & Liew, T. C. Quantum reservoir processing. *NPJ Quant. Inf.* **5**, 1–6 (2019).
44. Ghosh, S., Opala, A., Matuszewski, M., Paterek, T. & Liew, T. C. Reconstructing quantum states with quantum reservoir networks. *IEEE Trans. Neural Netw. Learn. Syst.* (2020).
45. Angelatos, G., Khan, S. A. & Türeci, H. E. Reservoir computing approach to quantum state measurement. *Phys. Rev. X* **11**, 041062 (2021).
46. Ghosh, S., Krisnanda, T., Paterek, T. & Liew, T. C. Realising and compressing quantum circuits with quantum reservoir computing. *Commun. Phys.* **4**, 1–7 (2021).
47. Xiong, W. *et al.* Strong tunable spin-spin interaction in a weakly coupled nitrogen vacancy spin-cavity electromechanical system. *Phys. Rev. B* **103**, 174106 (2021).
48. Xiong, W., Chen, J., Fang, B., Lam, C.-H. & You, J. Coherent perfect absorption in a weakly coupled atom-cavity system. *Phys. Rev. A* **101**, 063822 (2020).
49. Xiong, W. *et al.* Higher-order exceptional point in a pseudo-hermitian cavity optomechanical system. *Phys. Rev. A* **104**, 063508 (2021).
50. Chen, J. *et al.* Strong single-photon optomechanical coupling in a hybrid quantum system. *Opt. Express* **29**, 32639–32648 (2021).
51. Zhang, G.-Q., Chen, Z., Xiong, W., Lam, C.-H. & You, J. Parity-symmetry-breaking quantum phase transition via parametric drive in a cavity magnonic system. *Phys. Rev. B* **104**, 064423 (2021).
52. Wolfram Research. Numerical Nonlinear Global Optimization. <https://reference.wolfram.com/language/tutorial/ConstrainedOptimizationGlobalNumerical.html> (2021). [Online; accessed 27-July-2021].
53. Agarwal, G. Entropy, the wigner distribution function, and the approach to equilibrium of a system of coupled harmonic oscillators. *Phys. Rev. A* **3**, 828 (1971).

## Acknowledgements

The author acknowledges the Spanish State Research Agency, through the Severo Ochoa and María de Maeztu Program for Centers and Units of Excellence in R &D (MDM-2017-0711) and through the QUARESC project (PID2019-109094GB-C21 and -C22/ AEI / 10.13039/501100011033). The author also acknowledges funding by CAIB through the QUAREC project (PRD2018/47). The author acknowledges financial support from the Turku Collegium for Science, Medicine and Technology as well as the Academy of Finland under project no. 348854. Finally, the author would like to thank Roberta Zambrini, Gian Luca Giorgi and Miguel C. Soriano for helpful discussion and comments.

## Competing interests

The author declares no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to J.N.

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

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023