

Prediction of chaotic dynamics and extreme events: A recurrence-free quantum reservoir computing approach

*Original*

Prediction of chaotic dynamics and extreme events: A recurrence-free quantum reservoir computing approach / Ahmed, Osama; Tennie, Felix; Magri, Luca. - In: PHYSICAL REVIEW RESEARCH. - ISSN 2643-1564. - 6:4(2024), pp. 1-19. [10.1103/PhysRevResearch.6.043082]

*Availability:*

This version is available at: 11583/2995089 since: 2024-12-07T20:24:08Z

*Publisher:*

American Physical Society - APS

*Published*

DOI:10.1103/PhysRevResearch.6.043082

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

## Prediction of chaotic dynamics and extreme events: A recurrence-free quantum reservoir computing approach

Osama Ahmed<sup>1,\*</sup>, Felix Tennie<sup>1</sup> and Luca Magri<sup>1,2,3,†</sup>

<sup>1</sup>*Imperial College London, Department of Aeronautics, Exhibition Road, London SW7 2BX, United Kingdom*

<sup>2</sup>*The Alan Turing Institute, London NW1 2DB, United Kingdom*

<sup>3</sup>*Politecnico di Torino, DIMEAS, Corso Duca degli Abruzzi, 24 10129 Torino, Italy*



(Received 7 May 2024; accepted 7 October 2024; published 1 November 2024)

In chaotic dynamical systems, extreme events manifest in time series as unpredictable large-amplitude peaks. Although deterministic, extreme events appear seemingly randomly, which makes their forecasting difficult. By learning the dynamics from observables (data), reservoir computers can time accurately predict extreme events and chaotic dynamics, but they may require many degrees of freedom (large reservoirs). In this paper, by exploiting quantum-computer ansätze and entanglement, we design reservoir computers with compact reservoirs and accurate prediction capabilities. First, we propose the recurrence-free quantum reservoir computer (RF-QRC) architecture. By developing *ad hoc* quantum feature maps and removing recurrent connections, the RF-QRC has quantum circuits with smaller depths. This allows the RF-QRC to scale well with higher-dimensional chaotic systems, which makes it suitable for hardware implementation. Second, we forecast the temporal chaotic dynamics and their long-term statistics of low- and higher-dimensional dynamical systems. We find that RF-QRC requires smaller reservoirs than classical reservoir computers for higher-dimensional systems and the same predictability. Third, we apply the RF-QRC to the time prediction of extreme events in a model of a turbulent shear flow with turbulent bursts. We find that the RF-QRC has longer predictability than the classical reservoir computer for extreme events forecasting. The results and analyses indicate that quantum-computer ansätze offers nonlinear expressivity and computational scalability, which are useful for forecasting chaotic dynamics and extreme events. This work opens new opportunities for using quantum machine learning on near-term quantum computers.

DOI: [10.1103/PhysRevResearch.6.043082](https://doi.org/10.1103/PhysRevResearch.6.043082)

### I. INTRODUCTION

Data-driven prediction of chaotic dynamical systems has gained significant interest in the last decade [1,2]. A data-driven model learns the solution of a dynamical system from data with the goal of predicting its temporal evolution. Weather and climate predictions [3,4], financial time-series forecasting [5], thermoacoustics [6], turbulence [7,8], among many others, are examples of chaotic dynamical systems. The use of data-driven methods for predicting nonlinear systems has been motivated mainly by two reasons: (a) the availability of large amounts of data for these systems, and (b) the difficulty of predicting temporal dynamics due to chaos. In chaotic systems, a small change in the initial conditions can drastically change the dynamical system's solution.

Because chaotic dynamics is a sequential-data problem, recurrent neural networks (RNNs) are a natural choice for

chaotic forecasting. RNNs are machine learning models that introduce recurrence in conventional neural network architectures. RNNs are employed in time-series forecasting, modeling nonlinear dynamical systems and chaos, and extreme event predictions [7,9–13]. On the one hand, the computational capabilities of RNNs are excellent. On the other hand, RNNs are difficult to train as they require backpropagation through time at each time step [14]. Because RNNs are designed to learn correlations using internal hidden states and long-lasting time dependences, the training through backpropagation can be difficult for long time-series forecasting [9].

Reservoir computing (RC) bypasses the problem of backpropagation by introducing a static recurrence—the reservoir—and training with ridge regression. This avoids backpropagation at each time step. Reservoir computing is a unified computing framework that was introduced as Echo State Networks [15] and Liquid State Machines [16]. In this study, we work with echo state networks, which are a type of reservoir computing approach, because of their potential in forecasting chaotic dynamics [9,11,17], extreme events predictions [8,18,19] and the stability properties [6,12]. (For brevity, we will refer to echo state networks as reservoir computing in this work). Despite their excellent forecasting abilities, the prediction capabilities of reservoir computers are sensitive to the choice of hyperparameters and the reservoir

\*Contact author: o.ahmed22@imperial.ac.uk

†Contact author: l.magri@imperial.ac.uk

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

size, which is constrained by the computational resources and memory of current classical computers.

Improving the performance of reservoir computing typically requires higher computational resources. Quantum computers hold the potential to provide exponential speed-up over classical computers for certain computational tasks [20]. This speed-up is also known as quantum advantage. There are already algorithms that have theoretically proven this concept. For example, the prime factorization problem, which is the basis of most modern encryption and is NP-hard (nondeterministic polynomial time hard) classically, can be solved in polynomial time using quantum computing principles [21]. The second example is the matrix inversion problem for solving linear systems of equations, which is a key step in many numerical solvers. Proposed quantum algorithms for matrix inversion have the potential to provide quadratic speed-up over conventional classical methods [22,23]. The underlying phenomena that enable the quantum advantage are entanglement, superposition, and interference effects in quantum computers. Recently, quantum computing has been applied in the area of machine learning. The current quantum hardware, however, is limited by the maximum number of qubits in the order of  $\sim 10^2 - 10^3$  with environmental noise and decoherence. In the current noisy intermediate scale quantum (NISQ)-era [24], hybrid quantum-classical methods are one of the leading candidates with prospects of achieving a quantum advantage by employing methods such as variational quantum algorithms and quantum circuit learning [25,26]. The hybrid quantum-classical methods that can be used for the prediction of chaotic dynamics and extreme events are inspired by classical machine learning architectures such as RNNs.

Quantum reservoir computing (QRC) is a type of quantum machine learning model that combines both frameworks, including quantum computing and reservoir computing. The idea of QRC is to use quantum dynamics to enhance the reservoir implementation [27–29]. Some numerical implementations of QRC have shown that the quantum systems of 5–7 qubits have comparable computational capabilities to the classical reservoir sizes of 100–500 [27]. Previous works have proposed hybrid QRC architectures for time series predictions and chaos modeling [30–33]. In hybrid QRC, the input data is encoded in the form of Bloch-sphere rotation angles in individual qubits along with reservoir states, which are the probability amplitudes. The resulting quantum states undergo a unitary operation at each time step [30]. In another approach, a reservoir state is associated with the density operator of the encoded quantum state, and reservoir states correspond to measured expectation values for each qubit [32,34]. Although it has been conjectured that quantum reservoir computers are potential candidates for providing a quantum advantage in the near-term NISQ era, only a few studies have been conducted on analyzing higher-dimensional chaotic systems [35], and long-term statistical predictions using QRC. The performance of classical reservoir computers is often limited by classical computers' memory [9], which constrains the predictability of these models. Reservoir computers require a high dimensional feature space, which increases with the nonlinearity and the dimensions of the physical system. Consequently, using QRC to predict high dimensional nonlinear dynamical systems that

require a large reservoir size is key to realizing a quantum advantage of QRC in predicting chaos.

In this work, we build up on the concept of hybrid QRC that encodes reservoir states as probability amplitudes of the quantum system [30]. We use gate-based quantum computing to implement this architecture and use it for high-dimensional chaotic time-series forecasting and extreme-event predictions. More specifically, we investigate the three-dimensional Lorenz-63, ten-dimensional Lorenz-96 systems [36] and the MFE (Moehlis, Faisst, and Eckhardt) [37] model for time-accurate, statistical predictions, and extreme events forecasting. Extreme events are sudden and unmitigated changes in the observables of chaotic flows, while long-term statistical prediction is also an important metric in forecasting chaos.

We benchmark our classical and quantum reservoir networks by studying them for low-order models and then extend our analysis to incorporate higher-dimensional dynamics. Both classical and quantum networks are used for extreme event predictions in a reduced order model of a shear flow between infinite plates subjected to a sinusoidal body forcing, also referred to as Moehlis, Faisst, and Eckhardt (MFE) [37]. We assess the performance using multiple performance metrics: first, by considering short-term time series prediction capabilities that are quantified using the predictability horizon (PH) and valid prediction time (VPT) [9,38], and, second, by comparison of long-term statistics that are quantified using statistical measures such as probability density function (PDF) and F-Score [12,13].

This paper is structured as follows. In Secs. II and III, we present the classical and proposed quantum reservoir architectures and their comparison. Different quantum reservoir architectures refer to the different ansätze chosen for forming a reservoir. Second, we apply both classical and quantum reservoir architectures on three-dimensional Lorenz-63 and ten-dimensional Lorenz-96 systems. We increase the dimensionality and complexity of the dynamical systems progressively for the analysis. Third, we compare the results of the best quantum feature map with the best classical reservoir applied to the MFE model to analyze the extreme-event prediction capabilities of both architectures. The numerical models of different dynamical systems and the results for the reservoir predictions are presented in Sec. IV. Finally, in Sec. V, we conclude our findings by highlighting the potential of QRC in chaotic time series forecasting along with its associated challenges and the direction of future works.

## II. BACKGROUND

In this section, we explain and compare classical and quantum reservoir computing algorithms and their schematic representations. To be self-contained, we also outline the basic mathematical framework and building blocks of quantum computers.

### A. Classical reservoir computing (CRC)

Unlike other RNNs, reservoir computing is a type of RNN framework that does not require backpropagation through time, which represents a major computational advantage [13,15]. As a result, the training cost is substantially reduced and the objective function is minimized through

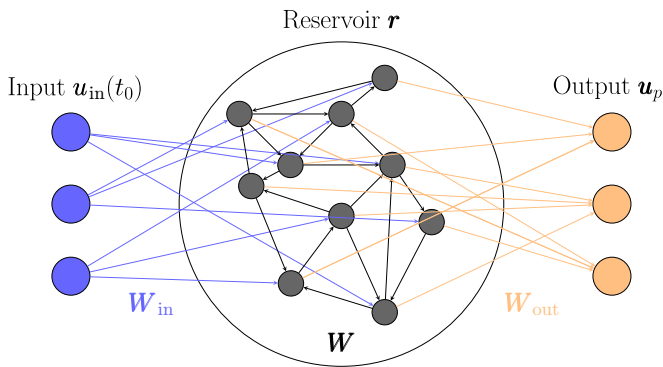


FIG. 1. Schematic representation of a classical reservoir computer [15]. The input data  $\mathbf{u}_{in}$  is mapped to the reservoir matrix via  $\mathbf{W}_{in}$ . The reservoir neuron connections governed by  $\mathbf{W}$  matrix allow the flow of information between neurons. The linear readout layer using the trained  $\mathbf{W}_{out}$  matrix is used to make output predictions  $\mathbf{u}_p$ .

a simple linear ridge regression. The schematic representation of classical reservoir computer is shown in Fig. 1. Let  $\{\hat{\mathbf{u}}(t_0), \hat{\mathbf{u}}(t_1), \dots, \hat{\mathbf{u}}(t_{N_r})\}$  be a training data set of a dynamical state  $\hat{\mathbf{u}} \in \mathbb{R}^{N_u}$ , which is known at  $N_r + 1$  steps in time. Generally, it is recommended to choose a set of variables that is (re-)scaled by the range component-wise, which is indicated by  $()$  on the training data set [39]. In reservoir computing, these states are mapped onto higher-dimensional reservoir states of an  $N_r$ -dimensional vector space. For a randomly chosen initial reservoir state  $\mathbf{r}(t_0)$ , we recursively compute a set of  $N_r$  reservoir vectors  $\mathbf{r}(t_1), \dots, \mathbf{r}(t_{N_r})$  as

$$\hat{\mathbf{r}}(t_{i+1}) = \tanh(\mathbf{W}_{in}\hat{\mathbf{u}}_{in}(t_i) + \mathbf{W}\mathbf{r}(t_i)), \quad (1)$$

$$\mathbf{r}(t_{i+1}) = (1 - \varepsilon)\mathbf{r}(t_i) + \varepsilon\hat{\mathbf{r}}(t_{i+1}). \quad (2)$$

Here, the state  $\hat{\mathbf{r}} \in \mathbb{R}^{N_r}$  represent a state-vector containing reservoir activation states,  $\mathbf{W}_{in} \in \mathbb{R}^{N_r \times N_u}$  is the input matrix and  $\mathbf{W} \in \mathbb{R}^{N_r \times N_r}$  is the reservoir weight matrix [15]. The tanh function is applied componentwise to provide a nonlinear activation. The matrices  $\mathbf{W}_{in}$  and  $\mathbf{W}$  are pseudorandomly generated and constant throughout the training and prediction phase. The elements of  $\mathbf{W}_{in}$  are sampled from a uniform distribution in  $[-\sigma_{in}, \sigma_{in}]$ , where  $\sigma_{in}$  is the input scaling. The reservoir weight matrix  $\mathbf{W}$  is an Erdős-Rényi matrix (Erdős-Rényi is a model for generating random graphs or the evolution of a random network in the field of graph theory), with average connectivity between each neuron varied by input reservoir density  $D$ . The spectral radius of the reservoir state matrix  $\rho$  is given by its maximal eigenvalue. The matrix  $\mathbf{W}$  is rescaled such that  $\rho \leq 1$  [15,39]. Effectively, this guarantees a rapid decay of temporal correlations between successive reservoir states and helps to avoid overfitting. Whereas Eq. (1) is the nonlinear activation step, Eq. (2) combines the linear memory of the reservoir with the nonlinear activation, as parametrized by the leaking rate  $\varepsilon$ . This additional postprocessing step is known as memory nonlinearity tradeoff [39].

Reservoir computing networks can be run in either open-loop or closed-loop configurations (Figs. 2 and 3). In the open-loop, we use the input data at each time step and compute the corresponding reservoir dynamics  $\mathbf{r}(t_i)$  according to

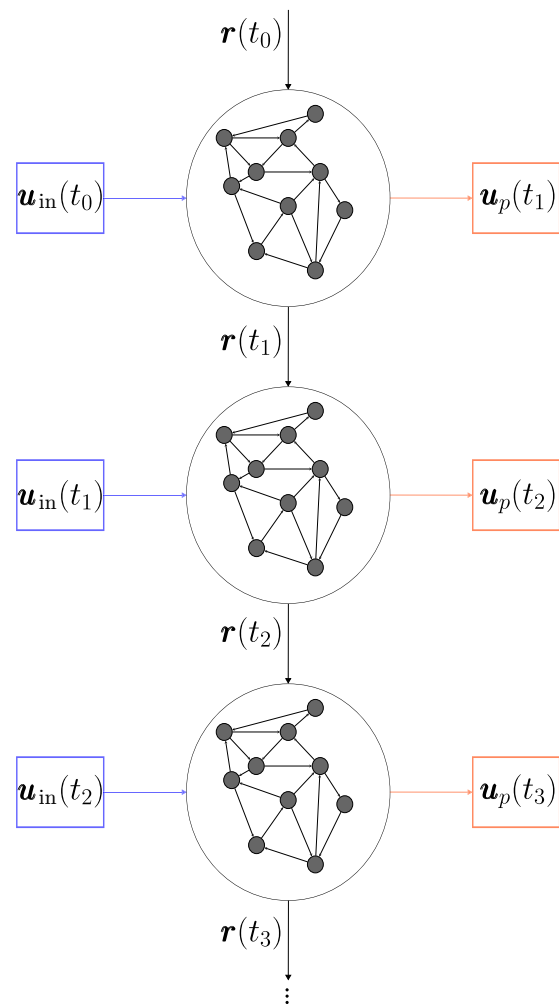


FIG. 2. Schematic representation of reservoir computers. Open-loop training phase.

Eqs. (1) and (2). Because of the initialization process with a randomly chosen reservoir state, it is necessary to discard a small number  $N_w$  of the initial transients to satisfy the echo state property (ESP) [15], which leads to dynamics that are less sensitive to the initial random choice. The process of discarding the initial transients is also called the washout phase. After the washout interval, reservoir dynamics at each time step  $\mathbf{r}(t_i)$  are collected to form a reservoir state matrix  $\mathbf{R} \in \mathbb{R}^{N_r \times (N_r - N_w)}$ . Reservoir computing in the closed-loop prediction mode requires an output matrix  $\mathbf{W}_{out}$  for forecasting the dynamics of the learned chaotic system. The training of the output matrix  $\mathbf{W}_{out}$  involves minimizing the mean square error between input and output data over the training data set, consisting of  $N_{tr}$  number of training steps. The simplicity of the reservoir network allows us to achieve this by solving a linear ridge regression problem

$$(\mathbf{R}\mathbf{R}^T + \beta\mathbf{I})\mathbf{W}_{out} = \mathbf{R}\mathbf{U}_d^T, \quad (3)$$

where  $\beta$  is a user-defined Tikhonov regularization parameter to prevent overfitting,  $\mathbf{I}$  is the identity matrix, and  $\mathbf{U}_d^T \in \mathbb{R}^{N_u \times N_r}$  is the horizontal concatenation of the output data. Once the output matrix has been computed from Eq. (3), it can be used to predict the evolution of dynamical

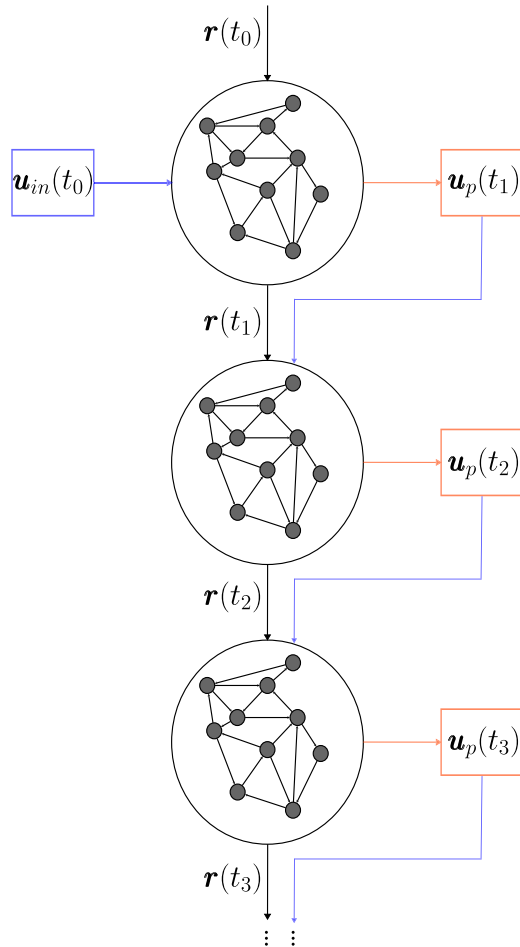


FIG. 3. Schematic representation of reservoir computers. Closed-loop prediction phase.

variables by

$$\mathbf{u}_p(t_{i+1}) = [\mathbf{r}(t_{i+1})]^T \mathbf{W}_{\text{out}}. \quad (4)$$

In the closed-loop configuration, we recursively predict the system's dynamics from Eq. (4) without additional samples from a training data set. Given an initial condition, this allows for an autonomous prediction on the unseen data set. Because reservoir computers have a symmetry [Eq. (1)], we add a constant output bias of 1 to the reservoir activation states to break the inherent symmetry of the reservoir architecture [40,41], effectively replacing  $\mathbf{r} \rightarrow [\mathbf{r}, 1]$ . A summary of the reservoir computing procedure is shown in Fig. 4, and compared with its quantum counterpart, which will be introduced in Sec. II C. The performance of reservoir computers critically depends on the set of chosen hyper-parameters. We, therefore, use grid search and Bayesian optimization to tune the hyperparameters [42]. We have used the *scikit-learn* library [43] and recycle validation techniques [13] for hyperparameter tuning. The hyperparameters and performance metrics for different chaotic systems are reported in Sec. IV.

### B. Qubits and quantum states

To be self-contained, we provide a brief outline of the key elements of quantum computers. A detailed introduction can

be found in Ref. [44]. In quantum computers, the internal state of the machine can be represented by a complex vector in a Hilbert space. The time evolution during computation is described by a matrix multiplication of this vector with a unitary operator.

In most quantum hardware architectures, the elementary building blocks, known as qubits, are microscopic systems that can be described by a complex-valued, normalized, two-dimensional vector in a Hilbert space, in which the computational basis states  $|0\rangle$  and  $|1\rangle$  (kets) can be associated with the Cartesian basis vectors, and the complex-valued amplitudes  $\alpha$  and  $\beta$  are normalized as  $|\alpha|^2 + |\beta|^2 = 1$

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \hat{=} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \in \mathbb{C}_2. \quad (5)$$

States of  $n$  qubits occupy the tensor product Hilbert space of the individual two-dimensional Hilbert spaces

$$|\psi_n\rangle = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{2^n} \end{pmatrix} \in \mathbb{C}_2^{\otimes n}, \quad (6)$$

where the computational basis states are tensor products of the single qubit basis states. The state-vector dimension scales exponentially with the number of qubits  $n$ . Unentangled pure states can be written as a single product of individual qubit states

$$|\phi_n\rangle = (\alpha_0|0_0\rangle + \beta_0|1_0\rangle) \otimes (\alpha_1|0_1\rangle + \beta_1|1_1\rangle) \otimes \dots \otimes (\alpha_n|0_n\rangle + \beta_n|1_n\rangle). \quad (7)$$

Most quantum states [Eq. (6)] are entangled and therefore cannot be decomposed into such a product. Unitary evolution amounts to a matrix multiplication, i.e., the final quantum state after a computation is

$$|\psi_f\rangle = U|\psi_i\rangle. \quad (8)$$

Quantum hardware architectures provide elementary sets of unitary operators, known as gates that can be used to assemble any arbitrary desired unitary operation. The final output of the computation is revealed by measurements of the quantum state, which provide estimates of the modular squared amplitudes ( $|a_1|^2, |a_2|^2, \dots$ ).

### C. Hybrid quantum-classical reservoir computing

Hybrid quantum-classical reservoir computing [30] is based on principles that are similar to classical reservoir computing introduced in Sec. II A. In QRC, the exponential size of  $n$ -qubit Hilbert spaces is utilized to encode and process classical reservoir state vectors,  $\mathbf{r}(t_i)$ . Thus, the classical reservoir is replaced by a quantum counterpart. More specifically, within a single loop of the training phase, the reservoir state  $\mathbf{r}(t_i)$  is mapped onto an  $n$ -qubit unitary operator  $P(\mathbf{r}(t_i))$ , that takes  $|0\rangle^{\otimes n}$  to the state  $|\psi(t_i)\rangle = P(\mathbf{r}(t_i))|0\rangle^{\otimes n}$ . The explicit form of  $P$  is given by the quantum circuit architecture that characterizes the specific QRC framework. Circuits are generally parametrized by single-qubit rotation angles and a number of entangling CNOT gates (Appendix A).

ALGORITHM 1: Classical Reservoir Computer (CRC)	ALGORITHM 2: Quantum Reservoir Computer (QRC)
1: Pre-process the input data time-series $\mathbf{u}_{\text{in}}(t_i)$ .	1: Pre-process the input data time-series $\mathbf{u}_{\text{in}}(t_i)$ and encode it into the quantum circuit using rotational encoding.
2: Random and sparse input weight matrix $\mathbf{W}_{\text{in}}$ maps the input to high dimensional reservoir. $\triangleright \mathbf{W}_{\text{in}} \mathbf{u}_{\text{in}}(t_i)$	2: Use the specified quantum feature maps to map the input data to the high-dimensional Hilbert space. Here, $U$ represents unitary evolution of the quantum state. $\triangleright U(\mathbf{u}_{\text{in}}(t_i))$
3: Random reservoir weight matrix $\mathbf{W}$ with specified connectivity provides connections and the flow of information between different neurons in the reservoir.	3: Apply an additional unitary $V(\alpha)$ sampling from a uniform random distribution to provide randomization in the quantum circuit.
4: Run reservoir update to calculate reservoir activation states $\mathbf{r}(t_{i+1})$ from the input $\mathbf{u}_{\text{in}}(t_i)$ . $\triangleright \tanh(\mathbf{W}_{\text{in}} \mathbf{u}_{\text{in}}(t_i) + \mathbf{W} \mathbf{r}(t_i))$	4: Quantum reservoir state update $ \psi(t_{i+1})\rangle$ is achieved by running the quantum circuit with the encoded input data $\mathbf{u}_{\text{in}}(t_i)$ , previous reservoir states $ \psi(t_i)\rangle$ and the additional random unitary. $\triangleright V(\alpha) \Phi(\mathbf{u}_{\text{in}}(t_i)) P(\mathbf{r}(t_i))  0\rangle$ .
5: Compute linear readout weight $\mathbf{W}_{\text{out}}$ by minimizing the MSE between predicted and true output, using linear ridge regression.	5: Measure the quantum circuit in a Pauli-Z basis to convert quantum reservoir state $ \psi(t_{i+1})\rangle$ to classical reservoir state $\mathbf{r}(t_{i+1})$ . Re-encode the measured $\mathbf{r}(t_{i+1})$ to $P$ unitary for the next time-step.
6: Make predictions using the trained $\mathbf{W}_{\text{out}}$ matrix and the reservoir activation states $\mathbf{r}(t_{i+1})$ . Use open-loop configuration for one-step-ahead predictions and the closed-loop configuration for the autonomous predictions from the input data.	6: Repeat steps 5 and 6 similarly to a classical reservoir computer, to train the network and make predictions by repeating the quantum reservoir state updates.

FIG. 4. Classical and quantum reservoir computing algorithms.

For a training data set  $\{\hat{\mathbf{u}}(t_0), \hat{\mathbf{u}}(t_1), \dots, \hat{\mathbf{u}}(t_{N_r})\}$ , the input time series is encoded in the quantum circuit using single-qubit rotation angles. With this encoding, a second unitary operator  $\Phi(\mathbf{u}_{\text{in}}(t_i))$  is applied to the quantum state  $|\psi(t_i)\rangle$ . Thereafter, a third unitary operator  $V(\alpha)$  given by a random parameterized circuit with  $n$  parameters  $\alpha$  is applied. The combined action for each time step is in Fig. 6:

$$|\psi(t_{i+1})\rangle = V(\alpha) \Phi(\mathbf{u}_{\text{in}}(t_i)) P(\mathbf{r}(t_i)) |0\rangle^{\otimes n}. \quad (9)$$

Similarly to classical reservoir computing, we derive a preprocessed reservoir state vector  $\hat{\mathbf{r}}(t_{i+1})$  from measured probabilities of  $|\psi(t_{i+1})\rangle$  in the computational basis. This requires sampling and measuring the state  $|\psi(t_{i+1})\rangle$  multiple times for each time stepping.

Following the postprocessing of  $\hat{\mathbf{r}}(t_{i+1})$  to  $\mathbf{r}(t_{i+1})$  [Eq. (2)], the reservoir state matrix  $\mathbf{R}$  is formed by concatenating the reservoir state vectors. As in classical reservoir computing, the initial reservoir state is chosen randomly, which in turn, necessitates the washout of a few initial time-stepped reservoir vectors. In the training phase, QRC is run in an open-loop and the output matrix  $\mathbf{W}_{\text{out}}$  is computed by a linear ridge regression [Eq. (3)]. When operated in a closed-loop, QRC is used to predict the system dynamics [Eq. (4)]. We compare the elementary steps of classical and quantum reservoir computing algorithms in Fig. 4. The open-loop and closed-loop configurations of the hybrid quantum-classical architecture

are presented in Fig. 5. Each reservoir update  $\mathbf{r}(t_i)$  in both the open-loop and closed-loop requires the input of the previous reservoir state  $\mathbf{r}(t_{i-1})$ , similarly to the classical reservoir update (dotted feedback loop in Fig. 5). In Sec. III, we propose a quantum-classical architecture, which is independent of this feedback loop, thereby allowing for parallelization of the training phase of QRC.

Although QRC exploits the exponential size of the state space of the quantum register, there is a practical limitation due to the limited connectivity of the employed quantum processors (e.g., neighbor-neighbour interactions only on 2D superconducting qubit processors). The performance of QRC generally depends on the degree of the entanglement present in the circuit [45]. Therefore, we investigate the performance of structurally different architectures.

### III. RECURRENCE-FREE QUANTUM RESERVOIR COMPUTING (RF-QRC)

The gate-based quantum reservoir computing framework of [30] (Sec. II C) is in principle equivalent to its classical counterpart. It has been used to study thermal convection flows and the turbulent Rayleigh-Bernard flow [30,46]. In this section, we propose an ansätze for the circuit architecture in QRC. To do so, a specific quantum feature map must be chosen and we therefore start by briefly reviewing a number

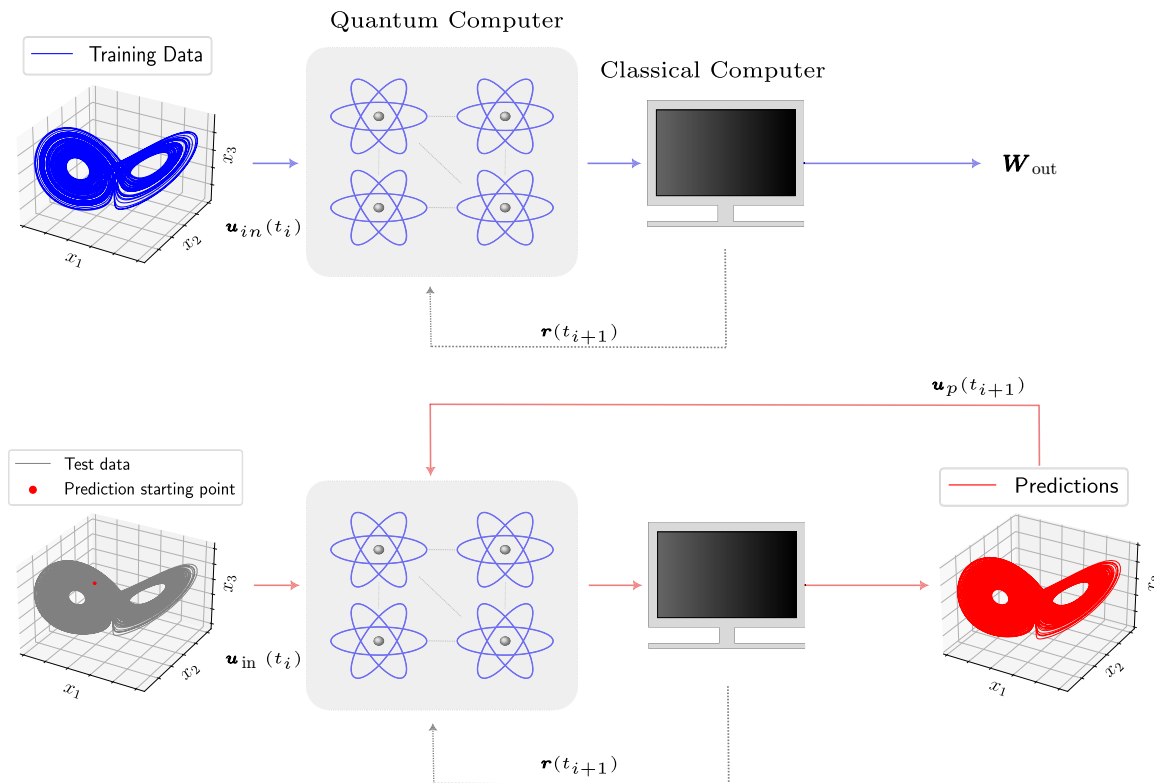


FIG. 5. Quantum-classical reservoir architecture schematic representation (a) open-loop training to produce  $W_{out}$  matrix (b) closed-loop autonomous predictions starting from an arbitrary point in the unseen data set. The dotted line in both figures indicates the recurrence involved in conventional reservoir architectures. In RF-QRC architecture, we remove this recurrent feedback layer of reservoir states.

of common choices. Following that, we present the RF-QRC architecture, which will be employed to study low and high-dimensional chaotic flows in Sec. IV.

We consider four different feature maps for QRC. These are the (a) Linearly entangling feature map, used in Ref. [30] for QRC applications, (b) Product state feature map, which was proposed in Ref. [47], and has been shown to offer a high expressivity [48], (c) Fully entangling feature map, in which CNOT gates entangle pairs of qubits, (d) Symmetric fully entangling feature map, which only differs from (c) by an additional data encoding layer following the CNOT gates. These feature maps do not require anywhere-to-anywhere connectivity (which would increase the computational overhead on current superconducting quantum processors). Schematic circuit diagrams of all four feature maps are shown in Appendix A. When the number  $\Theta$  of encoded parameters of a string of data (e.g., for  $r$ ,  $\Theta = 2^n$ ), exceeds the number of qubits, the feature map encoding is applied multiple times. In particular, this leads to exponentially growing circuit depths of the unitary  $P$  and motivates the exploration of new efficient QRC architectures, as displayed in Table I.

An example comparison of circuit depths corresponding to the five architectures is shown in Fig. 7. Removing the recurrence, i.e., by setting  $P$  to the identity, and by applying  $\Phi$  twice for stimulating the reservoir dynamics, yield circuits depths that are independent of the reservoir size. The suitability of this choice is justified *a posteriori* by an improved performance of QRC in a range of prototypical model systems. Recurrence-free QRC is different than quantum extreme

learning machines [49], because of the use of a linear combination of the memory and nonlinearity of previous reservoir activation states, as expressed by Eq. (2) (also known as the leaky-integral reservoir computing approach [50]).

#### IV. NUMERICAL RESULTS

In this section, we analyze the dynamics of a number of chaotic systems, such as Lorenz-63, Lorenz-96, and MFE using CRC and QRC. Thereby, we can assess the prediction capabilities of various architectures. More specifically, we evaluate the short-term time-accurate and long-term statistical predictions with performance measures. The sets of training, validation, and test data are generated by a fourth-order Runge-Kutta numerical scheme. To emulate the quantum circuits, we have used the QISKIT [51] software package to compute the noise-free evolution of the quantum register.

In a physical implementation, at each time step between  $10^4$  to  $10^5$  shots are required to reconstruct the reservoir activation signal. A number of shots between  $10^4$  to  $10^6$  is common in variational algorithms and other *quantum machine learning* applications [52]. Mitigating the requirement of a large number of shots in quantum computing is an area of active research, which is beyond the scope of this paper. (QRC and RF-QRC could be combined with the concept of weak and projective measurements [52], classical shadows [53], resolvable expressive capacity [54], or artificial memory restriction [55] to potentially reduce the number of required shots.)

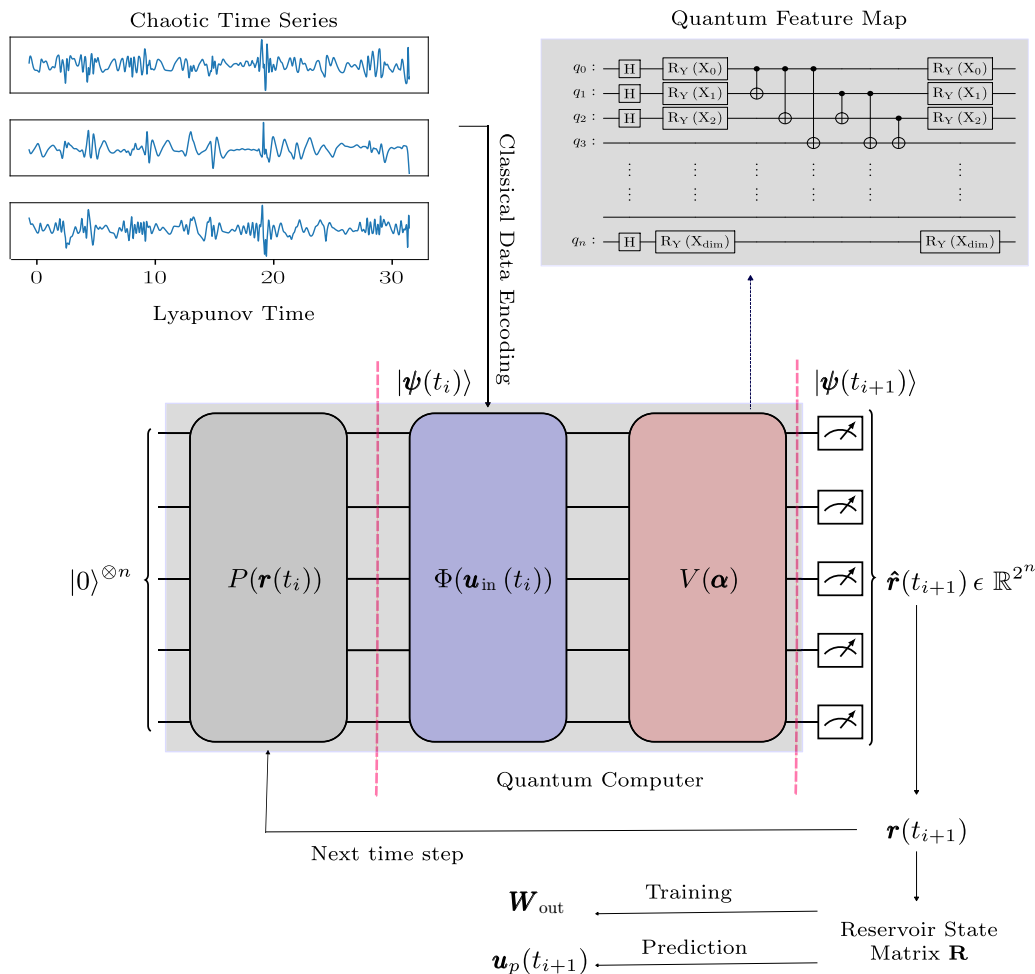


FIG. 6. Schematic representation of gate-based quantum reservoir computer. The first two unitaries are used to encode previous reservoir states  $r(t_i)$  and input time series  $u_{in}(t_i)$ , the third additional unitary provide randomization. Each unitary consists of a specific feature map. The quantum circuit execution starts from the ground state  $|0\rangle$  and measures the reservoir state activation for the next time step  $\hat{r}(t_{i+1})$ , which is postprocessed classically before feeding the next time step. Each reservoir activation state  $r(t_{i+1})$  is saved classically to use for training and predictions. In RF-QRC, we remove the feedback loop and the first unitary  $P$ .

**A. Performance measures**

The dynamics of chaotic systems can be characterized by the leading Lyapunov exponent  $\Lambda_1$  [38]. This exponent corresponds to the average exponential rate of divergence for initially nearby trajectories. The Lyapunov exponent also provides a time scale to assess the time-accurate prediction of the chaotic systems. We have rescaled our time units to the inverse of the Lyapunov exponent ( $\Lambda_1$ ), which is called the Lyapunov time ( $1LT = \Lambda_1^{-1}$ ).

In this work, we choose the valid prediction time (VPT) [9] as a performance measure of short-term time-accurate predictions. For a given threshold value  $\epsilon$ , VPT is defined as the time for which the normalized root mean square error (NRMSE) between the predicted and true values is less than  $\epsilon$ . In this work we take  $\epsilon = 0.5$ , as in Ref. [9]:

$$NRMSE = \sqrt{\left| \frac{y_t - \hat{y}_t}{\sigma^2} \right|}, \tag{10}$$

TABLE I. Quantum reservoir computer - Different ansätze designs.

	Reservoir states ( $P$ )	Input ( $\Phi$ )	Variation ( $V$ )
QRC-C1	Linearly entangled	Fully entangled	Fully entangled symmetric
QRC-C2	Linearly entangled	Linearly entangled	Linearly entangled
QRC-C3	–	Linearly entangled (x2)	Linearly entangled
QRC-C4 (RF-QRC)	–	Fully entangled (x2)	Fully entangled symmetric
QRC-C5	–	Product states (x2)	Linearly entangled

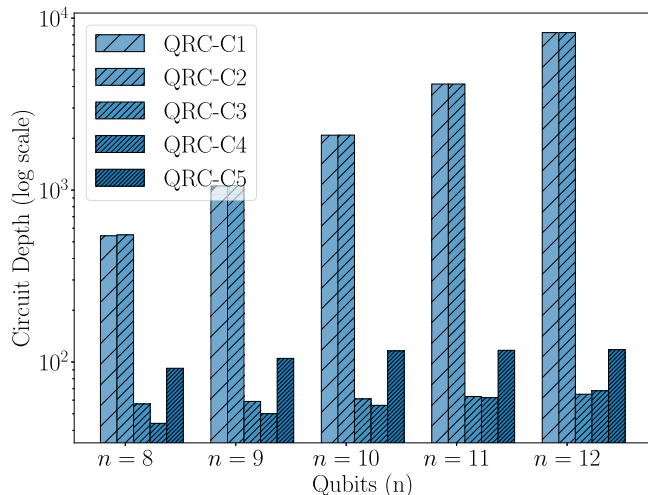


FIG. 7. Circuit depth of different ansätze for the ten-dimensional Lorenz-96 input.

$$VPT = \frac{1}{\Lambda_1} \operatorname{argmax} [t_f | \text{NRMSE} < \varepsilon = 0.5, \forall t \leq t_f], \quad (11)$$

where  $y_t$  is the true value and  $\hat{y}_t$  is the corresponding predicted value, and  $\sigma$  is the standard deviation of the time series,  $t_f$  is the largest value of the time step at which NRMSE is smaller than the threshold  $\varepsilon$ . High VPT values indicate a high predictability. To assess the long-term statistical prediction capabilities of our reservoir networks, we evolve the trained networks autonomously in a closed-loop configuration from an arbitrary point on the attractor and compare the state variable distributions with the actual distributions. For each configuration, we run simulations with five different sets of optimal hyperparameters, which corresponds to different random seeds.

### B. Three-dimensional Lorenz-63 model

As a first test case, we start with the analysis of the three-dimensional Lorenz-63 system [36]. The Lorenz-63 system is a reduced-order model of a thermal convection flow, in which the fluid is heated uniformly from below and cooled from the top. This model is defined mathematically by

$$\begin{aligned} \frac{dx_1}{dt} &= \sigma(x_2 - x_1), & \frac{dx_2}{dt} &= x_1(\rho - x_3) - x_2, \\ \frac{dx_3}{dt} &= x_1x_2 - \beta x_3, \end{aligned} \quad (12)$$

where  $[\sigma, \rho, \beta] = [10, 28, 8/3]$  results in a chaotic behavior of the system. For the Lorenz-63 system,  $\Lambda = 0.9$  and  $1LT = 1/0.9$ , [36]. The time series data set is derived by a Runge-Kutta method for  $dt = 0.01s$ . The training time series comprises data points over a total time of 20 LT for both the classical and quantum reservoir networks. For the training, both networks are executed in an open loop to evolve the reservoir states and calculate the  $\mathbf{W}_{\text{out}}$  matrix. For predictions, both classical and quantum networks evolve dynamically in the closed-loop configuration from ensembles of points in

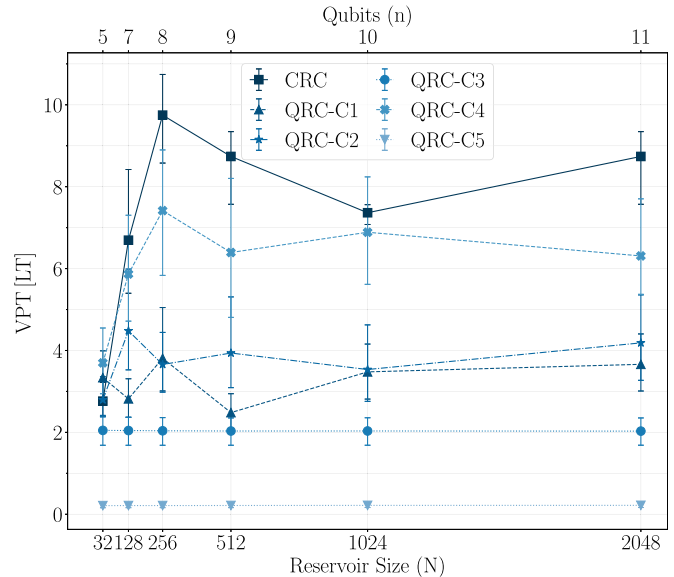


FIG. 8. Lorenz-63 system. VPT vs reservoir sizes for both classical and quantum reservoir configurations.

phase space, sampled on the attractor randomly, to quantify the time accuracy. These points correspond to the points in the unseen test data set.

In Table II, we compare the hyperparameters of both quantum and classical reservoirs. Values inside square brackets indicate that the parameters are optimized within this range, multiple values indicate that we repeated the hyperparameter search for each of these values and then selected the best VPT out of them. The random rotation angles ( $\alpha$ ) in the third unitary block can also be treated as a hyperparameter. Alternatively, one could derive them from a uniform random distribution, which we will do in this paper. For each ensemble, we sample  $V(\alpha) \in \mathbb{R}^n$  from a uniform distribution interval  $[0, 4\pi]$  with a predefined seed, which we keep fixed throughout the training and prediction of a particular realization. This is similar to the classical input  $\mathbf{W}_{\text{in}}$  and reservoir weight matrices  $\mathbf{W}$ , which are also pseudo-randomly generated and fixed for any specific realization, as explained in Sec. II A.

#### 1. Time-accurate predictions

In Fig. 8, we present the VPT values derived with classical reservoir computing and compare them with different configurations of quantum reservoir computers from Table I. The error bars at each value indicate the variation arising from different random seeds. We have tested five different seeds for each value. These seeds correspond to randomly generated weight matrices  $\mathbf{W}$  and  $\mathbf{W}_{\text{in}}$  for the classical reservoir and variation of  $V(\alpha)$  values for quantum reservoir architectures. For each of these choices, after training, we initialize the closed-loop predictions from 20 different initial points for quantifying VPT and its mean value.

The performance of different reservoir networks varies with the reservoir sizes. We find that CRC outperforms all the QRC architectures, for the same reservoir sizes. The reservoir size corresponds to the state-vector dimension, which

TABLE II. Parameters for three-dimensional Lorenz-63 system.

Parameters	Symbol	Classical RC	Quantum RC
Time step	$dt$	0.01s	0.01s
Input scaling	$\sigma_{in}$	[0, 1]	-
Spectral radius	$\rho$	[0.1, 1]	1
Tikhonov regularization	$\beta$	$1 \times 10^{-6}, 1 \times 10^{-9}, 1 \times 10^{-12}$	$1 \times 10^{-6}, 1 \times 10^{-9}, 1 \times 10^{-12}$
Leak rate	$\epsilon$	[0.05, 1]	[0.05, 0.3]
Reservoir density	$D$	0.1,0.6,0.9	Configurations C1-C5

scales exponentially with the number of qubits ( $n$ ). The best-performing quantum reservoir architecture, QRC-C4 with 9-11 qubits ( $n$ ), can predict VPT values similar to the classical reservoir sizes of 500-2000.

We emphasize that the three QRC configurations C3-C5 do not have a feedback loop. These configurations evolve only from the input time series and the information on reservoir activation states from previous time steps is provided via the classical update Eq. 2. In contrast, architectures QRC-C1 and QRC-C2 involve an additional active feedback loop that encodes reservoir states to the quantum circuit at each time step, which leads to an additional overhead in the quantum-classical layer.

The prediction performance of the two RF-QRC configurations, QRC-C3 and QRC-C5, do not improve with the reservoir sizes. The QRC-C4 architecture, however, shows comparable performance to the reservoirs with feedback loops (QRC-C1 and QRC-C2). These results highlight the importance of carefully choosing feature maps for QRC.

In Fig. 9, the VPT values normalized by the circuit depth are compared between different QRC architectures. We find that the proposed QRC-C4 architecture outperforms all other architectures by demonstrating the best prediction capabilities while requiring a smaller circuit depth.

Time-accurate predictions for the Lorenz-63 system are shown in Fig. 10. We display the best-performing CRC

outcomes (reservoir size of 512), as well as predictions made by the QRC-C4 architecture using 9 qubits. Notice that both the quantum and classical reservoirs can accurately predict up to  $\sim 8.5$  LTs. Beyond that, the prediction diverges from the true solution due to the inherent chaotic nature of the system.

### 2. Long-term statistical predictions

We select the best-performing networks for which the short-term predictions are shown in Fig. 10. We evolve the networks autonomously in a closed-loop for 250 LT, starting from an arbitrary point. The resulting statistics are shown in Fig. 11 for each state variable. Both quantum and classical reservoir architectures are able to predict long-term statistics accurately.

### C. Higher-dimensional Lorenz-96 model

In this section, we extend our comparison of classical and quantum reservoirs to higher-dimensional chaotic systems. We follow the same procedure defined in Sec. IV B and analyze the Lorenz-96 model [36]:

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \quad i = 1, \dots, m, \quad (13)$$

where  $F$  is the external body forcing term which we set to  $F = 8$ , to ensure chaotic behavior [9]. We apply periodic boundary conditions, i.e.,  $x_1 = x_{m+1}$ , and study the reduced order model of Lorenz-96 with ten dimensions ( $m = 10$ ). The training set covers an evolution time of 200 LT for the Lorenz-96 system with a leading Lyapunov exponent value of  $\Lambda_1 = 1.2$ . In Table III, we present the hyperparameters for the ten-dimensional Lorenz-96 system. The set of hyperparameters is derived by using the same procedure as discussed in Sec. II.

#### 1. Time-accurate predictions

We analyze the time-accurate prediction capabilities of CRC and different QRC architectures for a ten-dimensional reduced-order model of Lorenz-96. Fig. 12 shows the VPT values of both networks. The QRC architectures C1, C2, and C4 outperform the classical reservoir for increasing reservoir sizes. To study the robustness of the model, the results are averaged over five different sets of optimal hyperparameters and 20 different starting points. For a particular realization of hyperparameters and a starting point, at reservoir size 4096 the VPT value for CRC can be as large as  $\sim 10$  LT. However, the performance is not robust and degrades quickly for other hyperparameter choices and different initial points. Improved

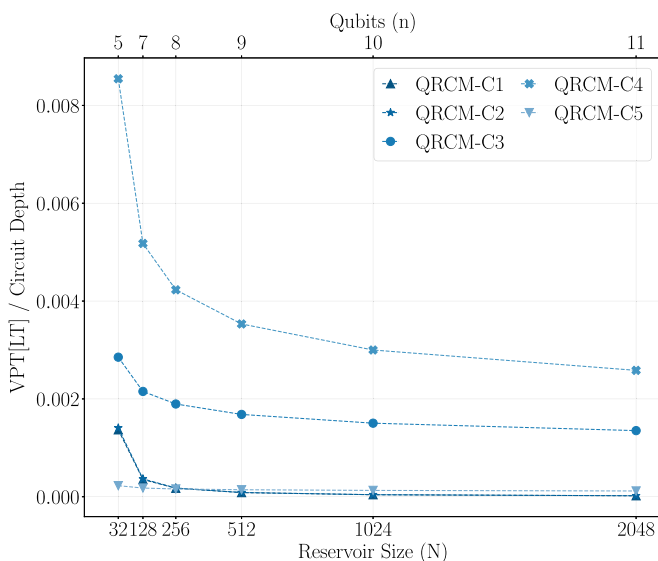


FIG. 9. Lorenz-63 system. VPT/circuit depth vs reservoir sizes for different quantum reservoir configurations.

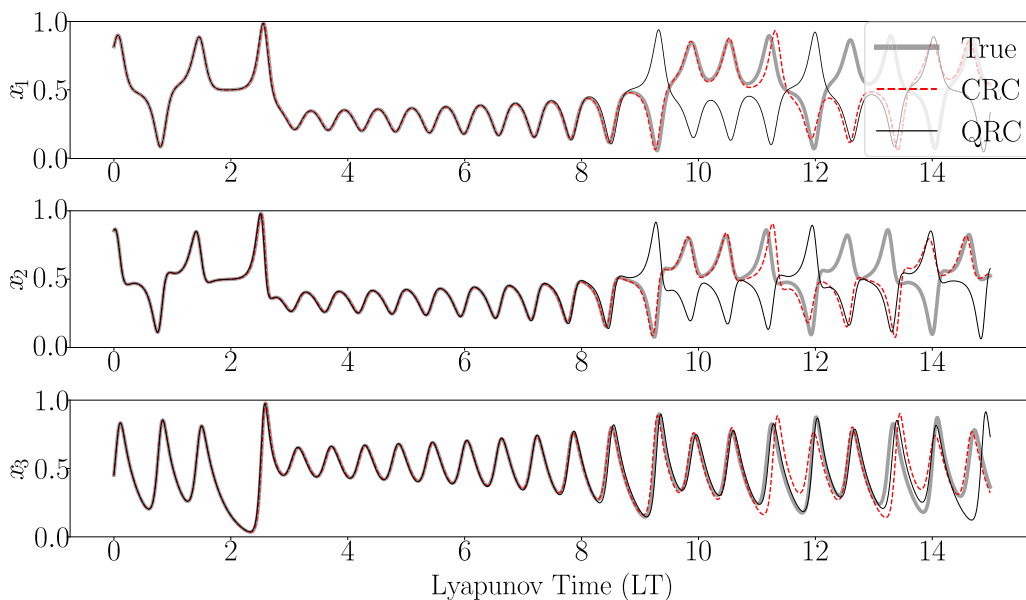


FIG. 10. Lorenz-63 system. Closed-loop time-series predictions. Comparison of true predictions with classical reservoir computing (CRC) and emulated quantum reservoir computing (QRC).

performance scaling and robustness indicate a potential quantum advantage in quantum reservoir computers.

Similar to the results in the Lorenz-63 system, the choice of a feature map and quantum reservoir architectures is critical for the performance. In Fig. 12, QRC-C4 and C1 have comparable VPT values at different reservoir sizes while QRC-C2 underperforms slightly. The performance of QRC-C3 and C5 do not improve with the reservoir size because there is no feedback loop, and because the linear input feature map only utilises a fraction of state space. By contrast, QRC-C4 (RF-QRC), which also does not have a recurrent feedback loop, has a fully connected input feature map that enriches the reservoir dynamics and has a comparable VPT to QRC-C1 and C2.

As compared to QRC with recurrent connections, QRC-C4 achieves the highest VPT to circuit depth ratio for all of the reservoir sizes as shown in Fig. 13. This is due to a lack of recurrence. In comparison, QRC-C1 and C2 require exponentially higher circuit depths at increasing reservoir sizes to achieve similar performance.

2. Long-term statistical predictions

In Figs. 14 and 15, we compare the long-term statistical predictions. The displayed results show the best-performing classical reservoir and quantum reservoir network (QRC-C4) outputs. We evolve these networks for 400 LT in a closed-loop, for the reservoir size of 1024. Both quantum and

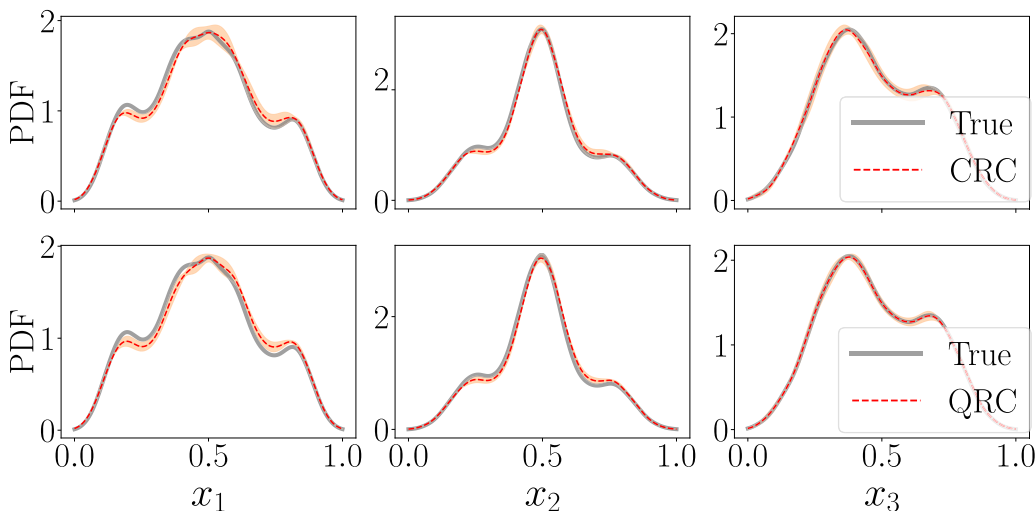


FIG. 11. Lorenz-63 system. Closed-loop statistical predictions. The top panels present the result of the true and classical (CRC) network. The bottom panels compare the true with the QRC predictions. The shaded region in both figures indicates the uncertainty associated with different random seeds in both CRC and QRC.

TABLE III. Parameters for ten-dimensional Lorenz-96 system.

Parameters	Symbol	Classical RC	Quantum RC
Time step	$dt$	0.01s	0.01s
Input scaling	$\sigma_{in}$	[0, 1]	-
Spectral radius	$\rho$	[0.1, 1]	1
Tikhonov regularization	$\beta$	$1 \times 10^{-6}, 1 \times 10^{-9}, 1 \times 10^{-12}$	$1 \times 10^{-6}, 1 \times 10^{-9}, 1 \times 10^{-12}$
Leak rate	$\epsilon$	[0.05, 1]	[0.05, 0.3]
Reservoir density	$D$	0.1,0.6,0.9	Configurations C1-C5

classical reservoir networks can recover long-term statistics of the chaotic model. The shaded region highlights the uncertainty of different realizations. We find that QRC-C4 predicts the statistical variables with smaller uncertainties when compared to the classical reservoir with equal reservoir sizes.

**D. Chaotic shear flow model for extreme event forecasting**

In fluid mechanics, extreme events are sudden and unmitigated changes of observables. The forecasting of extreme events is the first step towards the control and suppression of these violent bursts. RNN's and RC's are used to study these events qualitatively and quantitatively [7,8,18]. We consider a qualitative low-order model of turbulent shear flows, which is based on Fourier modes and describes a self-sustained turbulent process. This model is also known as the (MFE) model. The MFE model is nonlinear and it captures the relaminarization and turbulent bursts [37]. Owing to the nonlinear nature of this model, the MFE model has been employed to study turbulence transitions and chaos predictability [7,19]. Mathematically, the MFE system can be described by the nondimensional Navier-Stokes equations for forced incompressible flow

$$\frac{\partial \mathbf{v}}{\partial t} = -(\mathbf{v} \cdot \nabla) \mathbf{v} - \nabla p + \frac{1}{Re} \Delta \mathbf{v} + \mathbf{F}(y), \quad \nabla \cdot \mathbf{v} = 0, \quad (14)$$

where  $\mathbf{v} = (u, v, w)$  is the three-dimensional velocity vector,  $p$  is the pressure,  $Re$  is the Reynolds number,  $\nabla$  is the gradient, and  $\Delta$  is the Laplacian operator.  $\mathbf{F}(y)$  on the right-hand side is the sinusoidal body forcing term and equals  $\mathbf{F}(y) = \sqrt{2}\pi^2/(4Re) \sin(\pi y/2) \mathbf{e}_x$ . The body forcing term is applied along the  $x, y$  direction of the shear between the plates. We consider a three-dimensional domain of length  $L_x, L_y, L_z = [4\pi, 2, 2\pi]$  and apply free slip boundary conditions at  $y = L_y/2$ , periodic boundary conditions at  $x = [0; L_x]$  and  $z = [0; L_z]$ . The set of PDEs can be converted into ODEs by projecting the velocities onto Fourier modes as given by Eq. (15)

$$\mathbf{v}(\mathbf{x}, t) = \sum_{i=1}^9 a_i(t) \hat{\mathbf{v}}_i(\mathbf{x}). \quad (15)$$

These nine decompositions for the amplitudes  $a_i(t)$  are substituted into Eq. (14) to yield a set of nine ordinary differential equations as in Ref. [37]. The MFE system displays a chaotic transient, which in the long term converges to a stable laminar solution. We want to predict the turbulent burst of kinetic energy and chaotic transients, which are extreme events. Figure 16 shows the evolution of kinetic energy  $[k(t) = \frac{1}{2} \sum_{i=1}^9 a_i^2(t)]$  and the associated extreme event, and the long-term statistical distribution of the kinetic energy. An extreme event occurs when the kinetic energy  $k(t)$  value exceeds the

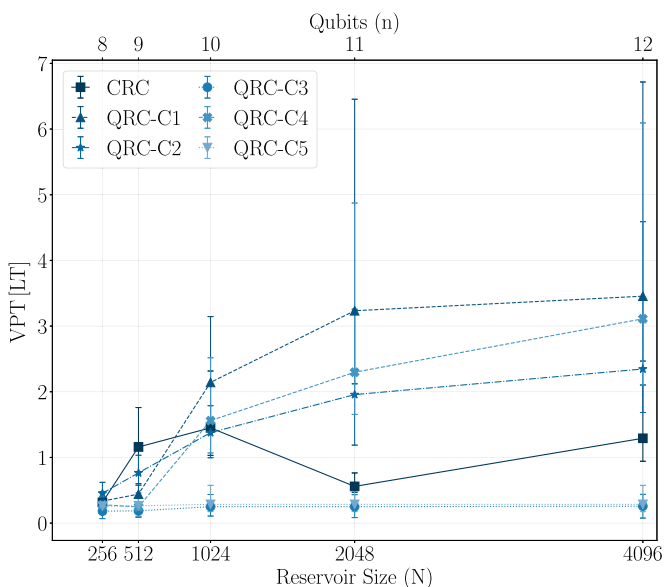


FIG. 12. Lorenz-96 system with ten dimensions. VPT vs reservoir sizes for both classical and quantum reservoir configurations.

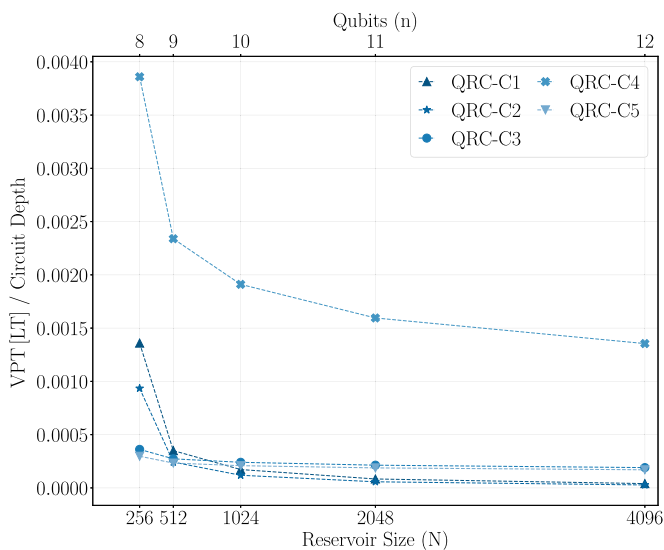


FIG. 13. Lorenz-96 system with ten dimensions. VPT/circuit depth vs reservoir sizes for different quantum reservoir configurations.

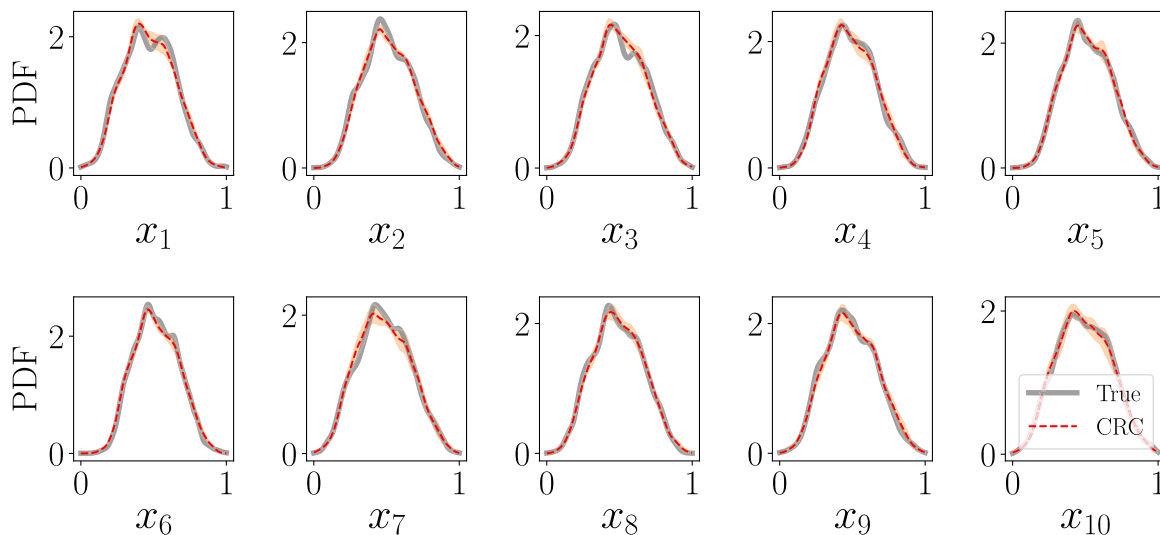


FIG. 14. Lorenz-96 system with ten dimensions. Closed-loop statistical predictions using CRC. The shaded region indicates the uncertainty associated with different random seeds in CRC.

threshold ( $k(t) \geq k_e$ ). In this work, we take  $k_e = 0.1$ , as in Ref. [18].

We solve the MFE system ODEs using an RK4 solver with  $dt = 0.25s$ . The leading Lyapunov exponent is  $\Lambda = 0.0163$  for MFE model [18]. Unlike the Lorenz-63 and Lorenz-96 models, generating a single long-time series for washout, training, and test sets is not feasible for MFE because a single long-time series eventually laminarizes and gives limited information about the chaotic transients. To address this issue, first, we generate an ensemble of 2000 time series from different random initial points. We take the length of a single time series equal to 65 LT or 16000 time steps. Second, we discard the time series whose maximum kinetic energy is larger than the laminarization threshold  $k_l = 0.48$ . This threshold is selected to be close to the (asymptotic) laminarization value of  $k = 0.5$ . Out of the 2000 time series, 27% of the series laminarizes. We divide the remaining 1441 time series into

washout, training, and test sets. We train our network on a 25-time series, each of length  $20LT$ . The test set consists of 500 time series. In Table IV, we present the range of hyperparameters for the MFE system. The set of hyperparameters are derived by using the same procedure as discussed previously in Sec. II. In Secs. IVD 1 and IVD 2, we compare the time-accurate and statistical predictions for the classical reservoir method against the best-performing quantum reservoir QRC4.

### 1. Time-accurate predictions

In the MFE model, we are concerned about quantifying the extreme event prediction capabilities of our reservoir networks. These capabilities are not accurately quantified by VPT, because extreme events are difficult to predict due to the chaotic nature of the attractor. This means that the network can show good predictability in low kinetic energy regions while

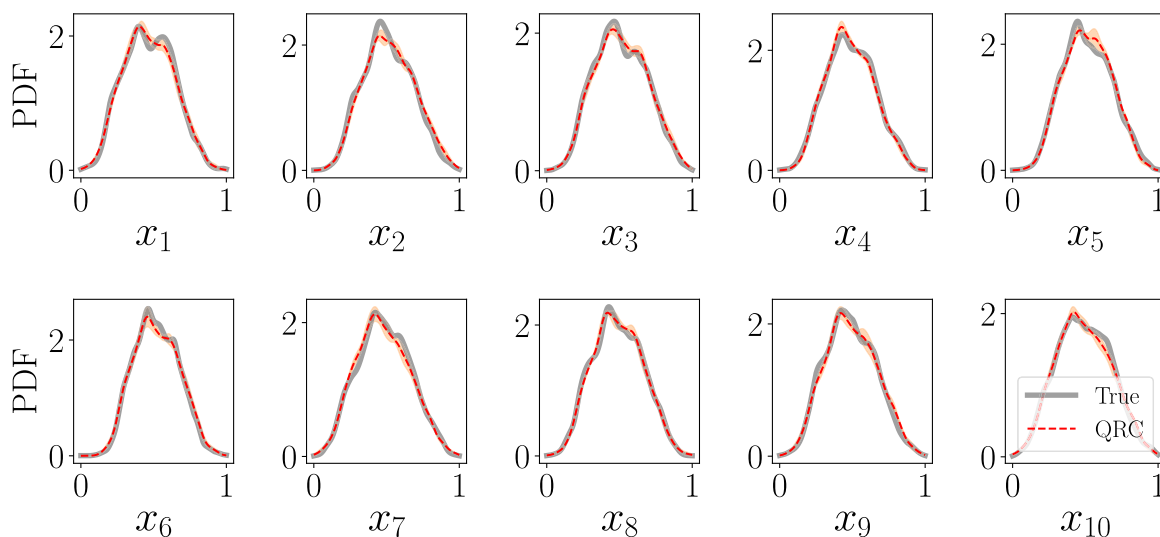


FIG. 15. Lorenz-96 system with ten dimensions. Closed-loop statistical predictions using QRC. The shaded region indicates the uncertainty associated with different random seeds in QRC.

TABLE IV. Parameters for the MFE system.

Parameters	Symbol	Classical RC	Quantum RC
Time step	$dt$	0.25s	0.25s
Input scaling	$\sigma_{in}$	[0, 1]	-
Spectral radius	$\rho$	[0.1, 1]	1
Tikhonov regularization	$\beta$	$1 \times 10^{-6}, 1 \times 10^{-9}, 1 \times 10^{-12}$	$1 \times 10^{-6}, 1 \times 10^{-9}, 1 \times 10^{-12}$
Leak rate	$\epsilon$	[0.05, 1]	[0.05, 0.3]
Reservoir density	$D$	0.1,0.6,0.9	Configuration C4 (RF-QRC)

showing smaller VPT values near extreme events. This may produce an inaccurate representation of the predictability of our reservoir networks.

To quantify extreme event prediction capabilities, we use the predictability horizon (PH) [18,19,56], which is defined as the time interval during which the predicted kinetic energy  $k_{pred}$  and true kinetic energy  $k_{true}$  are bounded by

$$\left| \frac{k_{pred}(t) - k_{true}(t)}{k_e - \bar{k}} \right| < 0.2, \quad (16)$$

where  $k_e$  is the extreme event threshold, which is equal to 0.1 here as in Ref. [18].  $\bar{k}$  is the time average of the kinetic energy, and 0.2 is the user-defined error threshold [18]. To quantify the PH, we sample 100 different extreme events from the unseen test set data. For each extreme event, we start the

predictions 12 LT before the extreme event and discard the transients in the washout phase for 2 LT by running reservoir networks in an open loop. This gives  $\delta_{k_e}$  of 10 LT. After this, we start autonomous closed-loop predictions with  $\delta_{k_e} = 10$ . The value of  $\delta_{k_e}$  represents the time difference from the start of closed-loop predictions to the extreme event. If the PH value is greater than  $\delta_{k_e}$  we take PH equal to  $\delta_{k_e}$ . If the prediction diverges before the extreme event occurs, or the PH value is less than  $\delta_{k_e}$ , we decrease the  $\delta_{k_e}$  value by a factor of  $\tau_e$ . We take  $\tau_e = 0.5$  LT and the new  $\delta_{k_e}$  value equals to 9.5 LT. For the next step, we repeat the closed-loop predictions for the extreme event occurring at 9.5 LT, i.e.,  $\delta_{k_e} = 9.5$  LT. We repeat the same method decreasing  $\delta_{k_e}$  by  $\tau_e$  until the PH is greater than  $\delta_{k_e}$  (Fig. 17). That is, the extreme event is predicted correctly by the reservoir computers by the value of PH in the future.

We calculate the PH for a set of reservoir sizes for both quantum and classical reservoir networks. For a classical reservoir, the number of neurons is varied from  $N = 256$  to  $N = 2048$ . We tune the hyperparameter for each reservoir and calculate the corresponding PH value. We also perform the same analysis for the quantum reservoir with qubits  $n = 8$  to  $n = 11$ , which, in principle, corresponds to the classical reservoirs of  $N = 256$  to  $N = 2048$ . The performance of the quantum reservoir also improves with the reservoir size. In Fig. 18, the comparison of classical and quantum reservoirs for various reservoir sizes is shown. Both reservoir networks have comparable performances for a reservoir sizes  $N = \{256, 512\}$ , with the classical reservoir outperforming quantum reservoir by a small margin. However, the PH value of the classical reservoir converges with a median value of around 5.5 LT. This convergence of PH value with increasing reservoir sizes is due to performance saturation in CRC and it is consistent with the results in Ref. [18]. In contrast, the quantum reservoir outperforms the classical reservoir for increasing reservoir sizes and has a median PH value of 8.09LT with 11 qubits.

A second statistical measure to assess the extreme event prediction capabilities is the  $F$ -Score [57], which measures the model's accuracy by combining metrics such as precision and recall. To calculate the  $F$  score, we measure three different occurrences: (a) an extreme event occurs in the true data set and the model also predicts the extreme event (True Positive, TP) (b) there is no extreme event in the true data but the model predicts an extreme event (False Positive, FP) (c) an extreme event occurs in the true data set but the model does not predict an extreme event (False Negative, FN) [18]. These three metrics are combined to calculate precision ( $p$ ), recall ( $r$ ), and  $F$  score ( $F$ ) from 1000 different starting points,

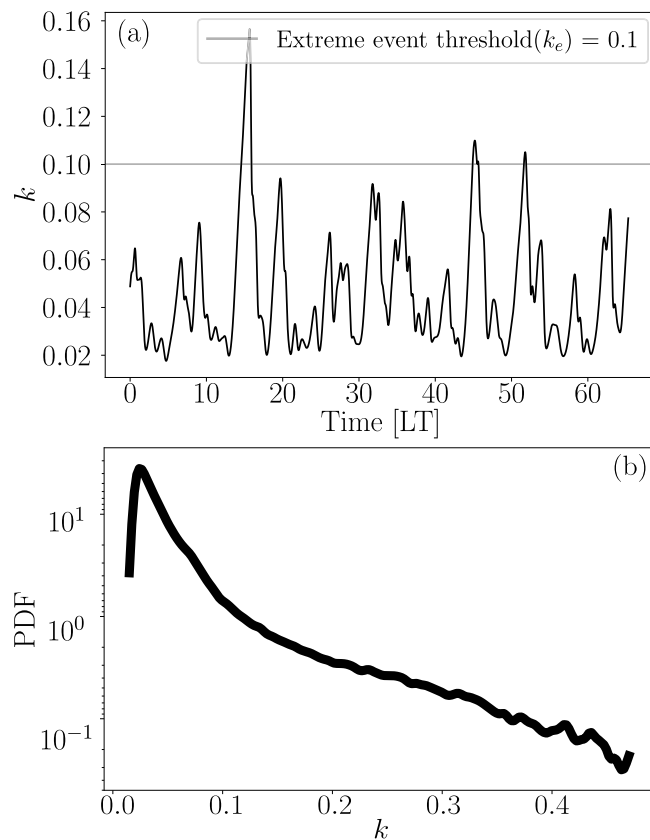


FIG. 16. MFE time series kinetic energy and probability density function (PDF) (a) single time series in which the kinetic energy exceeds the extreme event threshold  $k_e$ , highlighting the presence of an extreme event (b) probability distribution of kinetic energy calculated over an ensemble of time series.

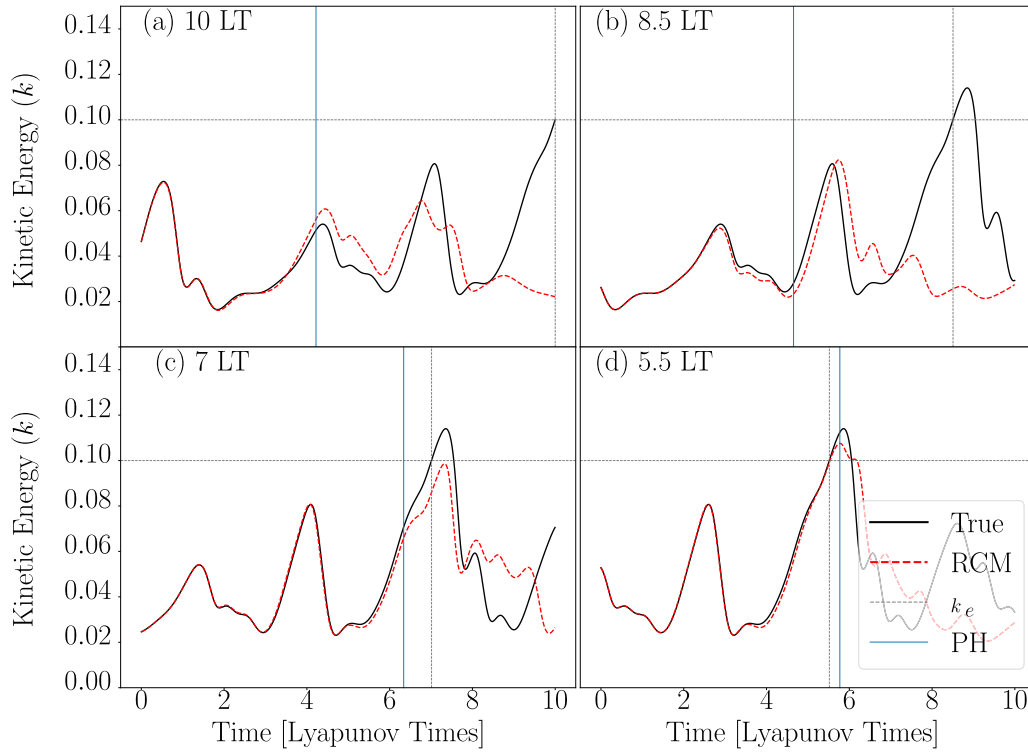


FIG. 17. Visualization of the prediction of extreme events for both reservoir computing methods (RCMs), i.e., both CRC and QRC (PH=5.89LT).

which span 20 different time series:

$$p = \frac{TP}{TP + FP}, \quad r = \frac{TP}{TP + FN}, \quad F = \frac{2}{p^{-1} + r^{-1}}. \quad (17)$$

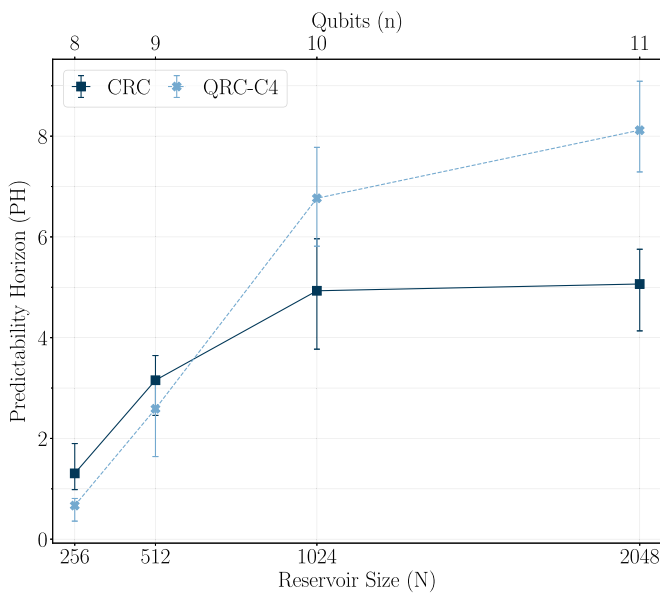


FIG. 18. MFE system. Predictability horizon (PH) for predictions with different reservoir sizes for the best performing CRC and QRC-C4 (RF-QRC) networks.

We compare quantum and classical reservoirs for two different cases. First, we fix the reservoir size for  $N = 1024$ , which is equals to 10 qubits for the quantum reservoir and varies prediction time (PT). The PT is defined as the time interval between the extreme event and the start of the prediction. The PT value of 0 LT indicates that the prediction interval is between 0 and 1 LT, and  $PT = 3$  indicates the prediction interval of 3–4 LT. The lower PT values have an  $F$ -score value closer to 1 because the network predicts nearly all the extreme events accurately, which are located within 1 LT of the start of predictions. As expected, the performance deteriorates with increasing the PT value due to the inherent chaotic nature of the model. The proposed quantum reservoir architecture (QRC-C4) achieves a higher  $F$  score as compared to the CRC for different PT intervals as shown in Fig. 19.

Second, we fix the prediction time to 3 LT ( $PT = 3LT$ ), which indicates the model is predicting extreme events between 3 and 4 LT. After fixing the PT value, we vary the reservoir sizes from  $N = 256$  to  $N = 2048$  and compare the performance metrics. The performances are similar for both quantum and classical reservoirs with reservoir sizes of {256, 512}. Similar to the results shown for the predictability horizon previously, the quantum architecture has a higher  $F$ -score value for 10,11 qubits as compared to the classical reservoirs of the same sizes (Fig. 19). All the results are computed for five different ensembles of networks and for 1000 different starting points.

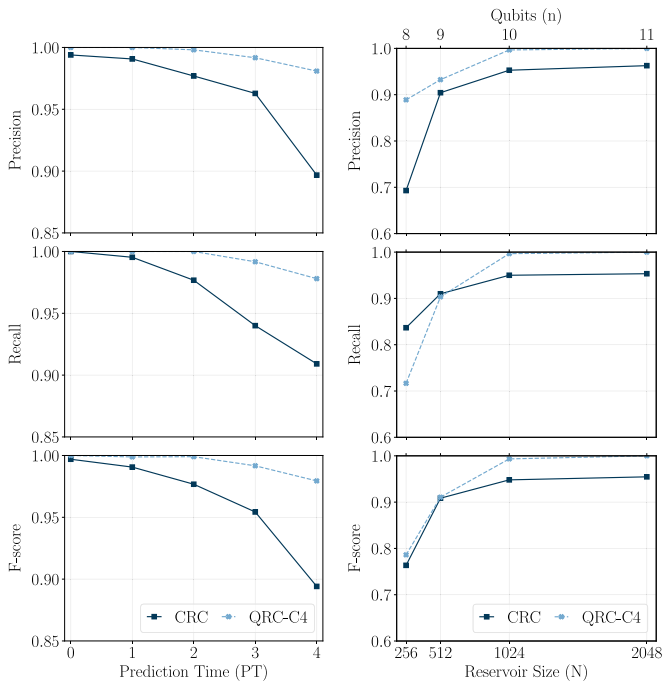


FIG. 19. MFE system. 75<sup>th</sup> percentile of Precision, Recall, F-score vs on the left side, prediction time (PT) for a constant reservoir size ( $N = 1024$ ). On the right side, reservoir size for a constant prediction time ( $PT = 3LT$ ).

### 2. Long-term statistical prediction

For the classical reservoir, we chose the reservoir size of  $N = 512$ . For the quantum reservoir, we chose a reservoir that corresponds to  $n = 9$  qubits. We evolve the networks from different starting points in the test set of 500 time series, in which each starting point corresponds to a different time series. As previously mentioned, we perform washout for each time series and discard the time series whose kinetic energy exceeds the laminarization threshold  $k_l > 0.48$ . Finally, we calculate the PDF.

Figure 20 presents the results for the statistical predictions against the true value. Both reservoirs can predict long-term statistics with high accuracy for the smaller kinetic energy

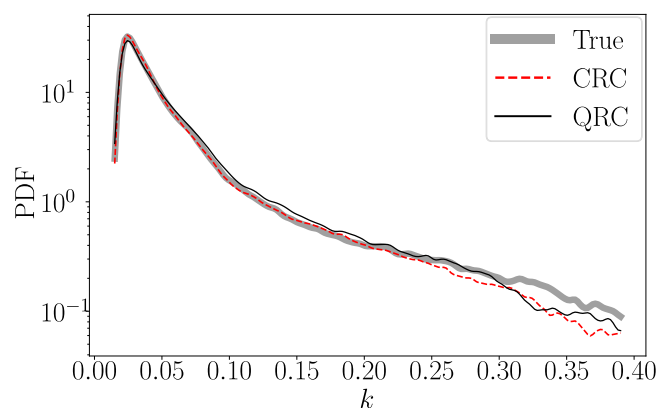


FIG. 20. MFE system. Long-term statistical predictions with classical and quantum reservoir computing.

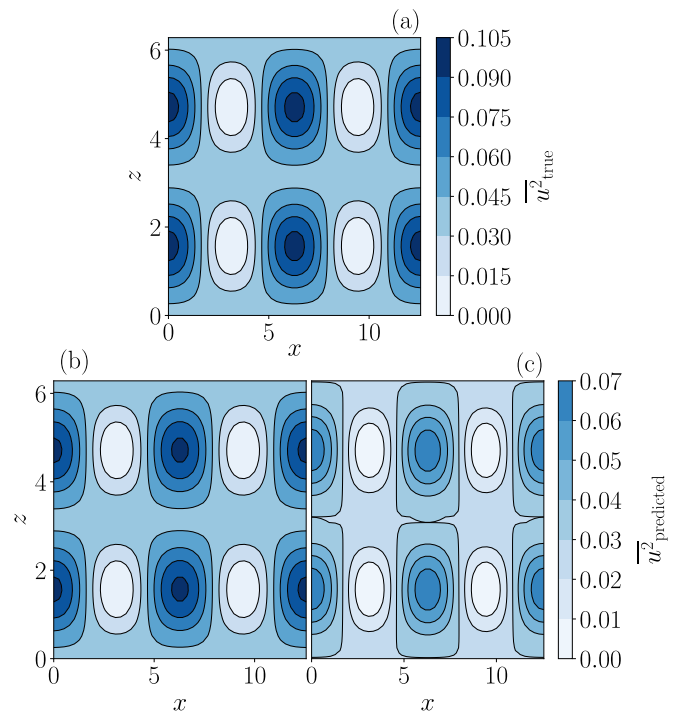


FIG. 21. MFE system. The time average of the square of the midplane velocity along  $x$ . (a) True data (b) QRC-C4 predictions (c) CRC predictions for the best-performing networks with a reservoir size of 512 (9 qubits).

values. Particularly, for  $k < 0.15$  the prediction agrees with the true value of the statistics. This analysis provides a statistical measure of the prediction of extreme events. Both quantum and classical reservoirs can predict most of the extreme events ( $k < 0.10$ ) in the test set. For the higher kinetic energy values, the log scale amplifies the difference between true value and predictions. The accuracy in the prediction of extreme events could be improved by considering a larger reservoir size, which scales better for QRC as compared to CRC, as shown previously in Figs. 18 and 19.

To further visualize the statistical predictions and compare CRC and RF-QRC, we predicted the velocity components of the flow fields from the predicted Fourier modes. In Fig. 21 the results of the average streamwise velocities are shown for the true data, and compared against classical and quantum reservoir predictions on the test data set. The predictions of RF-QRC approximate the true flow field better than the predictions from CRC, which have larger deviations. We conjecture that RF-QRC can be used to reconstruct flow fields with better accuracy than CRC with similar reservoir sizes. The prediction of flow fields with a better accuracy also allows the RF-QRC model to predict turbulent bursts of relaminarization and thus kinetic energies (extreme events) with better prediction horizon as previously shown in Figs. 18 and 19.

## V. CONCLUSION

By exploiting quantum-computer ansätze and entanglement, we analyze and design reservoir computers to predict chaotic dynamics and extreme events from data. First, we

show that the design of a feature map, which encapsulates entanglement and data-encoding layers, is key to the reservoir’s performance. Second, we design the RF-QRC, which has a small circuit depth that does not scale with the number of qubits (reservoir size), unlike previous proposals of QRC. RF-QRC also does not have recurrent connections at each time step. This enables the training of the network without an additional classical feedback loop, which makes the implementation suitable to current noisy intermediate-scale quantum devices. Third, we numerically analyze quantum reservoir computers and the proposed RF-QRC on prototypical chaotic dynamical systems, from low- to higher-dimensional. We show that (a) the forecasting capability of state-of-the-art classical and quantum reservoir computers increases with the reservoir size until saturation; (b) for low-dimensional chaotic systems, classical reservoir computers have larger predictability than quantum reservoir computers for the same degrees of freedom; (c) for higher-dimensional chaotic systems, RF-QRC has larger predictability than classical reservoir computers for the same degrees of freedom; and (d) RF-QRC can predict extreme events while scaling better than state-of-the-art reservoir computers. The RF-QRC requires smaller reservoir sizes as compared to classical reservoir computers for the same performance. The RF-QRC can be used to encode classical and quantum data on a quantum computer to make time-series predictions. This work opens new opportunities for the prediction of chaotic dynamics and extreme events with quantum reservoir computing. Current work is focused on the analysis of hardware and environmental noise, and finite-sampling errors.

**ACKNOWLEDGMENTS**

The authors acknowledge financial support from the UKRI New Horizon Grant No. EP/X017249/1. L.M. is grateful for the support from the ERC Starting Grant No. PhyCo 949388, and the Grant No. EU-PNRRYoungResearcher TWIN ERC-PI 0000005. F.T. acknowledges support from the UKRI AI for Net Zero Grant No. EP/Y005619/1.

**APPENDIX A: QUANTUM CIRCUITS FOR DIFFERENT ANSÄTZE**

Figures 22–25 represent the quantum circuits for different ansätze presented in Table I. The rotation angles  $X_\Theta$  represent the mapped classical data rescaled to the interval  $[0, 2\pi]$ .

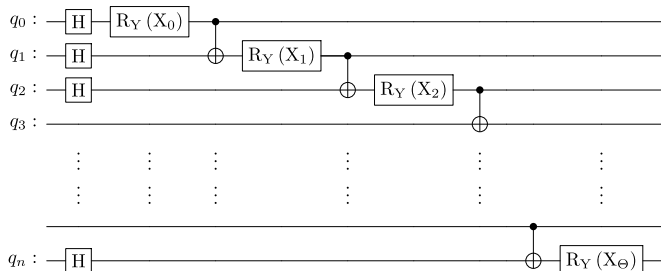


FIG. 22. Quantum circuit for linearly entangled qubits.

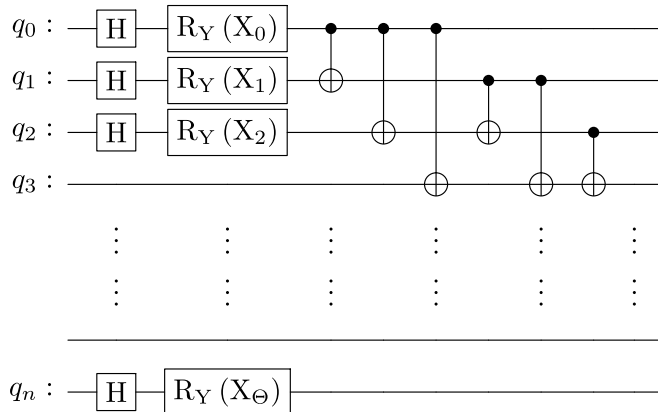


FIG. 23. Quantum circuit for fully entangled qubits.

**APPENDIX B: MEMORY OF THE RESERVOIR NETWORKS**

In this section, we discuss the empirical metrics including fading memory and linear memory capacity (MC) [58–62] for different reservoir architectures. In order to efficiently reconstruct the reservoir dynamics, reservoir computers must satisfy the ESP [15,39]. In classical reservoir networks, ESP is enforced by scaling the hyperparameter  $\rho$  (Sec. II A). In quantum reservoir networks, unitary evolution is norm-preserving, i.e.,  $(U^\dagger U = I)$ , and therefore no tuneable hyperparameter  $\rho$  exists. In both, quantum and classical reservoir computing, we perform a washout interval by discarding a limited number of initial reservoir states.

The short-term MC is a performance measure that is used to characterize the linear memory capacity of reservoir computers [62]. MC quantifies the maximally possible linear correlations between the current reservoir states and previous input data,  $\mathbf{u}(t - d)$ , and can be computed as

$$MC = \sum_{d=1}^{d_{\max}} MF_d, \quad MF_d = \frac{cov^2(\mathbf{u}_d(t), \mathbf{u}(t - d))}{\sigma^2(\mathbf{u}_d(t))\sigma^2(\mathbf{u}(t))} \quad (B1)$$

where  $cov(\cdot, \cdot)$  is the covariance,  $\sigma^2(\cdot)$  is the variance,  $d$  is the delay,  $\mathbf{u}(t - d)$  is the delayed prediction for a particular

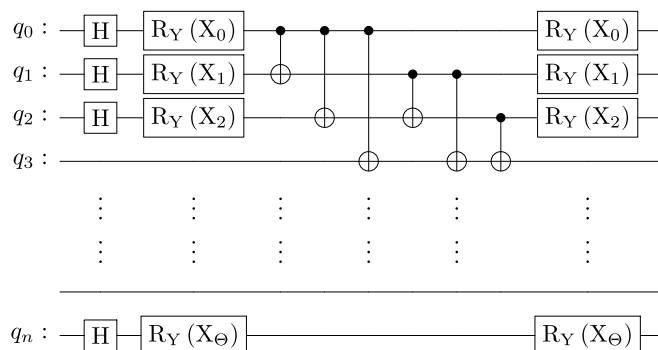


FIG. 24. Quantum circuit for fully entangled symmetric qubits.

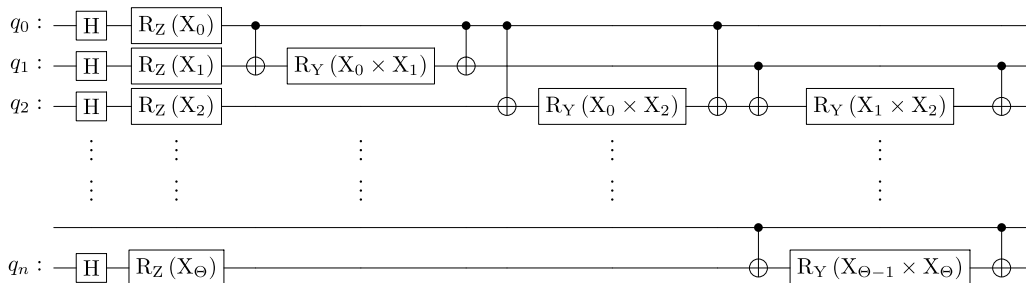


FIG. 25. Quantum circuit for product feature map qubits.

$d$  value and  $u_d(t)$  is the current true input value. Although the total MC comprises infinitely many terms ( $d_{\max} = \infty$ ), for practical reasons it is sufficient to truncate the sum at a finite  $d_{\max}$ . For a comparison of the memory capacities of CRC and QRC we choose  $d_{\max} = 25$ . While the MC quantifies the linear memory of the reservoir, the dynamical system itself has some correlations that can increase the MC. To quantify the memory of the reservoir itself, we train our quantum and classical reservoir networks on statistically independent (i.i.d) uniform distributions. In Figs. 26 and 27, we compare the linear memory capacity of the classical reservoir with QRC-C1 (architecture with recurrence) and QRC-C4 (recurrence-free architecture). The global MC is bounded by the size of the reservoir and is sensitive to the choice of hyperparameters [59]. We compute our MC for various hyperparameter values and reservoir sizes ( $N = 32, 512$ ).

The results in Figs. 26 and 27 demonstrate that CRC exhibits the highest MC value followed by QRC-C1 and then QRC-C4. A high memory capacity in CRC highlights that the reservoir contains linear information about the input time series for a larger delay value  $d$ . In contrast, QRC-C4 does not have any recurrent or feedback loop. Due to the removal of the recurrence reservoir memory is lost. We emphasize that the higher MC value is not indicative of better predictability. This is because of two reasons: (a) The short-term memory capacity quantifies only the linear memory of the reservoir. (b) Maximizing the linear memory and nonlinear functional approximation are mutually exclusive and are referred to as the memory nonlinearity trade-off [63]. We find comparable performances in short and long-term prediction tasks, irrespective of the MC value, for both classical and quantum reservoirs.

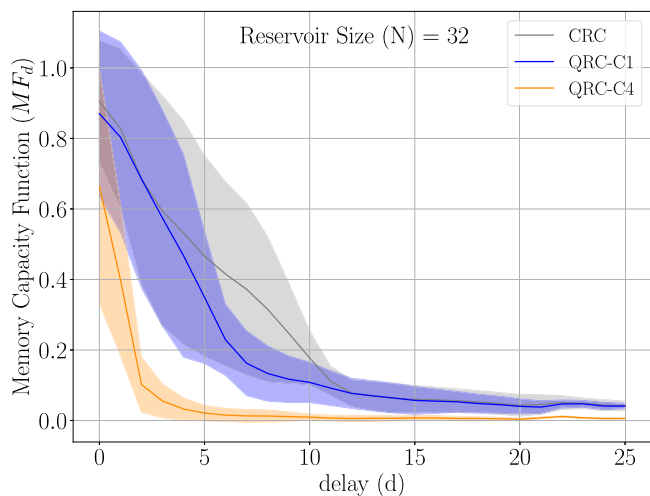


FIG. 26. Linear memory capacity (MC) of reservoir networks, mean and standard deviation with hyperparameters. For a summary of hyperparameters cf. Table II. For a reservoir size ( $N$ ) = 32.

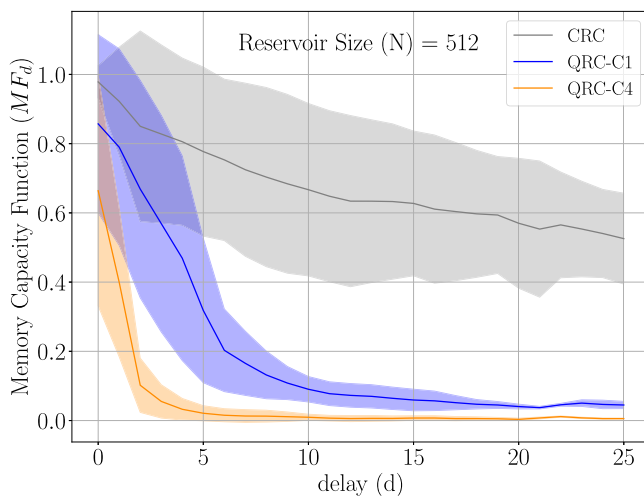


FIG. 27. Linear memory capacity (MC) of reservoir networks, mean and standard deviation with hyperparameters. For a summary of hyperparameters cf. Table II. For a reservoir size ( $N$ ) = 512.

- [1] C. Cheng, A. Sa-Ngasoongsong, O. Beyca, T. Le, H. Yang, Z. Kong, and S. T. Bukkapatnam, Time series forecasting for non-linear and non-stationary processes: A review and comparative study, *IIE Trans.* **47**, 1053 (2015).
- [2] A. Ghadami and B. I. Epureanu, Data-driven prediction in dynamical systems: Recent developments, *Philos. Trans. R. Soc. A* **380**, 20210213 (2022).
- [3] S. Dadhich, V. Pathak, R. Mittal, and R. Doshi, Chapter 10 Machine learning for weather forecasting, *Machine Learning for Sustainable Development*, edited by K. K. Hiran, D. Khazanchi, A. K. Vyas, and S. Padmanaban (De Gruyter, Berlin, Boston, 2021), pp. 161–174.
- [4] D. Rolnick, P. L. Donti, L. H. Kaack, K. Kochanski, A. Lacoste, K. Sankaran, A. S. Ross, N. Milojevic-Dupont, N. Jaques, A. Waldman-Brown *et al.*, Tackling climate change with machine learning, *ACM Comput. Surv. (CSUR)* **55**, 1 (2023).
- [5] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, Financial time series forecasting with deep learning : A systematic literature review: 2005–2019, *Appl. Soft Comput.* **90**, 106181 (2020).
- [6] F. Huhn and L. Magri, Stability, sensitivity and optimisation of chaotic acoustic oscillations, *J. Fluid Mech.* **882**, A24 (2020).
- [7] P. A. Srinivasan, L. Guastoni, H. Azizpour, P. Schlatter, and R. Vinuesa, Predictions of turbulent shear flows using deep neural networks, *Phys. Rev. Fluids* **4**, 054603 (2019).
- [8] N. A. K. Doan, W. Polifke, and L. Magri, A physics-aware machine to predict extreme events in turbulence, [arXiv:1912.10994](https://arxiv.org/abs/1912.10994).
- [9] P. R. Vlachas, J. Pathak, B. R. Hunt, T. P. Sapsis, M. Girvan, E. Ott, and P. Koumoutsakos, Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics, *Neural Networks* **126**, 191(2020).
- [10] A. Racca and L. Magri, Statistical prediction of extreme events from small datasets, in *Computational Science – ICCS 2022*, edited by D. Groen, C. de Mulatier, M. Paszynski, V. V. Krzhizhanovskaya, J. J. Dongarra, and P. M. A. Sloot (Springer International Publishing, Cham, 2022), pp. 707–713.
- [11] N. A. K. Doan, W. Polifke, and L. Magri, Physics-informed echo state networks for chaotic systems forecasting, in *Computational Science – ICCS 2019*, edited by J. M. F. Rodrigues, P. J. S. Cardoso, J. Monteiro, R. Lam, V. V. Krzhizhanovskaya, M. H. Lees, J. J. Dongarra, and P. M. Sloot (Springer International Publishing, Cham, 2019), pp. 192–198.
- [12] G. Margazoglou and L. Magri, Stability analysis of chaotic systems from data, *Nonlinear Dynamics* **111**, 8799 (2023).
- [13] A. Racca and L. Magri, Robust optimization and validation of echo state networks for learning chaotic dynamics, *Neural Networks* **142**, 252 (2021).
- [14] P. Werbos, Backpropagation through time: What it does and how to do it, *Proc. IEEE* **78**, 1550 (1990).
- [15] H. Jaeger, The “echo state” approach to analysing and training recurrent neural networks-with an erratum note, *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report* **148**, 13 (2001).
- [16] W. Maass, T. Natschläger, and H. Markram, Real-time computing without stable states: A new framework for neural computation based on perturbations, *Neural Comput.* **14**, 2531 (2002).
- [17] J. Boedecker, O. Obst, J. T. Lizier, N. M. Mayer, and M. Asada, Information processing in echo state networks at the edge of chaos, *Theory Biosci.* **131**, 205 (2012).
- [18] A. Racca and L. Magri, Data-driven prediction and control of extreme events in a chaotic flow, *Phys. Rev. Fluids* **7**, 104402 (2022).
- [19] N. A. K. Doan, W. Polifke, and L. Magri, Short-and long-term predictions of chaotic flows and extreme events: A physics-constrained reservoir computing approach, *Proc. R. Soc. A* **477**, 20210135 (2021).
- [20] R. P. Feynman, Simulating physics with computers, in *Feynman and Computation* (CRC Press, Boca Raton, FL, 2018), pp. 133–153.
- [21] P. Shor, Algorithms for quantum computation: Discrete logarithms and factoring, in *Proceedings 35th Annual Symposium on Foundations of Computer Science* (Publisher, City, 1994), pp. 124–134.
- [22] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum algorithm for linear systems of equations, *Phys. Rev. Lett.* **103**, 150502 (2009).
- [23] C. Bravo-Prieto, R. LaRose, M. Cerezo, Y. Subasi, L. Cincio, and P. J. Coles, Variational quantum linear solver, *Quantum* **7**, 1188 (2023).
- [24] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke *et al.*, Noisy intermediate-scale quantum algorithms, *Rev. Mod. Phys.* **94**, 015004 (2022).
- [25] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).
- [26] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, *Phys. Rev. A* **98**, 032309 (2018).
- [27] K. Fujii and K. Nakajima, Harnessing disordered-ensemble quantum dynamics for machine learning, *Phys. Rev. Appl.* **8**, 024030 (2017).
- [28] K. Fujii and K. Nakajima, Quantum reservoir computing: A reservoir approach toward quantum machine learning on near-term quantum devices, in *Reservoir Computing: Theory, Physical Implementations, and Applications*, edited by K. Nakajima and I. Fischer (Springer, Singapore, 2021), pp. 423–450.
- [29] S. Y.-C. Chen, D. Fry, A. Deshmukh, V. Rastunkov, and C. Stefanski, Reservoir computing via quantum recurrent neural networks, [arXiv:2211.02612](https://arxiv.org/abs/2211.02612).
- [30] P. Pfeffer, F. Heyder, and J. Schumacher, Hybrid quantum-classical reservoir computing of thermal convection flow, *Phys. Rev. Res.* **4**, 033176 (2022).
- [31] 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).
- [32] D. Fry, A. Deshmukh, S. Y.-C. Chen, V. Rastunkov, and V. Markov, Optimizing quantum noise-induced reservoir computing for nonlinear and chaotic time series prediction, *Sci. Rep.* **13**, 19326 (2023).
- [33] F. Wudarski, D. O’Connor, S. Geaney, A. A. Asanjan, M. Wilson, E. Strbac, P. A. Lott, and D. Venturelli, Hybrid quantum-classical reservoir computing for simulating chaotic systems, [arXiv:2311.14105](https://arxiv.org/abs/2311.14105).

- [34] J. Dudas, B. Carles, E. Plouet, F. A. Mizrahi, J. Grollier, and D. Marković, Quantum reservoir computing implementation on coherently coupled quantum oscillators, *npj Quantum Inf.* **9**, 64 (2023).
- [35] Q. H. Tran and K. Nakajima, Higher-Order Quantum Reservoir Computing, [arXiv:2006.08999](https://arxiv.org/abs/2006.08999).
- [36] E. Lorenz, Predictability: A problem partly solved, Ph.D. thesis, Shinfield Park, Reading, 1995.
- [37] J. Moehlis, H. Faisst, and B. Eckhardt, A low-dimensional model for turbulent shear flows, *New J. Phys.* **6**, 56 (2004).
- [38] G. Boffetta, M. Cencini, M. Falcioni, and A. Vulpiani, Predictability: A way to characterize complexity, *Phys. Rep.* **356**, 367 (2002).
- [39] M. Lukosevicius, A Practical Guide to Applying Echo State Networks, in *Neural Networks: Tricks of the trade*, Lecture Notes in Computer Science, 2nd ed., edited by G. Montavon, G. B. Orr, and K.-R. Müller (Springer, Berlin, 2012), pp. 659–686.
- [40] F. Huhn and L. Magri, Learning ergodic averages in chaotic systems, in *Computational Science – ICCS 2020*, edited by V. V. Krzhizhanovskaya, G. Závodszy, M. H. Lees, J. J. Dongarra, P. M. A. Sloot, S. Brissos, and J. Teixeira (Springer International Publishing, Cham, 2020), pp. 124–132.
- [41] J. Herteux and C. Răth, Breaking symmetries of the reservoir equations in echo state networks, *Chaos* **30**, 123142 (2020).
- [42] J. Snoek, H. Larochelle, and R. P. Adams, Practical Bayesian optimization of machine learning algorithms, *Advances in Neural Information Processing Systems*, edited by F. Pereira, C. J. Burges, L. Bottou, and K. Q. Weinberger, Vol. 25 (Curran Associates, Inc., 2012).
- [43] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.* **12**, 2825 (2011).
- [44] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, 2011).
- [45] N. Götting, F. Lohof, and C. Gies, Exploring quantumness in quantum reservoir computing, *Phys. Rev. A* **108**, 052427 (2023).
- [46] P. Pfeffer, F. Heyder, and J. Schumacher, Reduced-order modeling of two-dimensional turbulent Rayleigh-Bénard flow by hybrid quantum-classical reservoir computing, *Phys. Rev. Res.* **5**, 043242 (2023).
- [47] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, *Nature (London)* **567**, 209 (2019).
- [48] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, The power of quantum neural networks, *Nat. Comput. Sci.* **1**, 403 (2021).
- [49] 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).
- [50] H. Jaeger, Mantas, D. Popovici, and U. Siewert, Optimization and applications of echo state networks with leaky-integrator neurons, *Neural Networks* **20**, 335 (2007).
- [51] Qiskit contributors, Qiskit: An open-source framework for quantum computing (2023).
- [52] 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).
- [53] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
- [54] F. Hu, G. Angelatos, S. A. Khan, M. Vives, E. Türeci, L. Bello, G. E. Rowlands, G. J. Ribeill, and H. E. Türeci, Tackling sampling noise in physical systems for machine learning applications: Fundamental limits and eigentasks, *Phys. Rev. X* **13**, 041020 (2023).
- [55] S. Čindrak, B. Donvil, K. Lüdge, and L. Jaurigue, Enhancing the performance of quantum reservoir computing and solving the time-complexity problem by artificial memory restriction, *Phys. Rev. Res.* **6**, 013051 (2024).
- [56] J. Pathak, A. Wikner, R. Fussell, S. Chandra, B. R. Hunt, M. Girvan, and E. Ott, Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model, *Chaos* **28**, 041101 (2018).
- [57] C. Goutte and E. Gaussier, A probabilistic interpretation of precision, recall and f-score, with implication for evaluation, in *Advances in Information Retrieval*, edited by D. E. Losada and J. M. Fernández-Luna (Springer, Berlin, 2005), pp. 345–359.
- [58] S. Boyd and L. Chua, Fading memory and the problem of approximating nonlinear operators with Volterra series, *IEEE Trans. Circuit Syst.* **32**, 1150 (1985).
- [59] T. L. Carroll, Optimizing Memory in reservoir computers, *Chaos* **32**, 023123 (2022).
- [60] I. Farkaš, R. Bosák, and P. Gergel', Computational analysis of memory capacity in echo state networks, *Neural Networks* **83**, 109 (2016).
- [61] X. Han, Y. Zhao, and M. Small, Revisiting the memory capacity in reservoir computing of directed acyclic network, *Chaos* **31**, 033106 (2021).
- [62] H. Jaeger, *Short Term Memory in Echo State Networks* (GMD Forschungszentrum Informationstechnik, 2001).
- [63] L. Storm, K. Gustavsson, and B. Mehlig, Constraints on parameter choices for successful time-series prediction with echo-state networks, *Mach. Learn.: Sci. Technol.* **3**, 045021 (2022).