Temporal information processing on noisy quantum computers

Jiayin Chen,[1, 2] Hendra I. Nurdin,[1, *] and Naoki Yamamoto[2, 3]

[1]*School of Electrical Engineering and Telecommunications,*
The University of New South Wales (UNSW), Sydney NSW 2052, Australia.
[2]*Quantum Computing Center, Keio University, Hiyoshi 3-14-1, Kohoku, Yokohama 223-8522, Japan.*
[3]*Department of Applied Physics and Physico-Informatics,*
Keio University, Hiyoshi 3-14-1, Kohoku, Yokohama 223-8522, Japan.

The combination of machine learning and quantum computing has emerged as a promising approach for addressing previously untenable problems. Reservoir computing is a state-of-the-art machine learning paradigm that utilizes nonlinear dynamical systems for temporal information processing, whose state-space dimension plays a key role in the performance. Here we propose a quantum reservoir system that harnesses complex dissipative quantum dynamics and the exponentially large quantum state-space. Our proposal is readily implementable on available noisy gate-model quantum processors and possesses universal computational power for approximating nonlinear short-term memory maps, important in applications such as neural modeling, speech recognition and natural language processing. We experimentally demonstrate on superconducting quantum computers that small and noisy quantum reservoirs can tackle high-order nonlinear temporal tasks. Our theoretical and experimental results pave the way for attractive temporal processing applications of near-term gate-model quantum computers of increasing fidelity but without quantum error correction, signifying the potential of these devices for wider applications beyond static classification and regression tasks in interdisciplinary areas.

I. INTRODUCTION

The ingenious use of quantum effects has led to a significant number of quantum machine learning algorithms that offer computational speed-ups [1, 2]. While awaiting the demonstration of these quantum algorithms on full-fledge quantum computers equipped with quantum error correction, quantum computing has transitioned from theoretical ideas to the noisy intermediate-scale quantum (NISQ) technology era [3]. Hybrid quantum-classical algorithms using short-depth circuits are particularly suitable for implementation on NISQ devices. Many notable experimental demonstrations of NISQ devices employ hybrid algorithms for data classification [4] and quantum chemistry [5]. An on-going quest is to find interesting applications on quantum computers with increasingly lower noise profile but not reaching a low enough threshold to enable continuous quantum error correction.

Here we propose a hybrid quantum-classical algorithm that *utilizes dissipative quantum dynamics* for temporal information processing on gate-model NISQ quantum processors. Our approach exploits dissipative quantum systems as *universal approximators* for nonlinear maps with short-term or fading memory, important in a broad class of real-world problems including spoken digit recognition [6], neural modeling [7] and machine learning tasks (e.g., speech processing and natural language processing) [8, 9]. This is a quantum analogue of the universal function approximation property neural networks enjoy [10], but for nonlinear mappings from sequential input to sequential output data [11–13].

If a map has short-term memory, it must eventually forget its initial conditions. Recurrent neural networks are popular approximators for short-term memory maps [14, 15]. These are artificial neural networks (NNs), a neuromorphic computing scheme inspired by the human brain. As the number of "neurons" responsible for learning increases, NNs with appropriate connections can approximate certain classes of maps [16], at the cost of being increasingly difficult to train [17, 18]. Reservoir computing (RC) circumvents this training cost by using a fixed but randomly generated dynamical system, the "reservoir", to nonlinearly map sequential inputs into a reservoir state. The connectivity of artificial neurons in the reservoir is random and never requires training. Only a simple linear regression algorithm is required to optimize parameters of a readout function to approximate target outputs. The use of simple linear read-out has connections to the biological concept of mixed selectivity, as demonstrated in monkeys [19]. The ease of RC implementation has brought forward many successful hardware implementations of classical (i.e., non-quantum) RC schemes [6, 20, 21]. Experiments suggest that the reservoir state-space dimension contributes importantly to the performance improvement for some tasks [21, 22]. Recently, a class of quantum reservoirs (QR) has been proposed to harness the exponentially large quantum state dimension for temporal information processing [23, 24]. This class of QRs is suitable for ensemble quantum systems and a static (non-temporal) version of [23] has been demonstrated in NMR for approximating static maps [25]. Chen and Nurdin further proposed a different class of QRs that is provably universal for approximating nonlinear short-term memory maps [26]. However, realizing these previous proposals in the quantum gate-model remains challenging due to the large number of gates re-

* h.nurdin@unsw.edu.au

quired to implement the dynamics via Trotterization.

In this Letter, we propose and experimentally demonstrate the first universal class of QRs that is readily implementable on noisy gate-model quantum computers for temporal signal processing. The attractiveness of our proposal is that random and noisy quantum circuits of arbitrary depths can be exploited for temporal information processing. Our proof-of-principle demonstration is performed on superconducting quantum processors, showing that quantum reservoirs with a small number of noisy qubits can tackle complex nonlinear temporal tasks, even in the absence of readout and process error mitigation techniques. This work serves as the first theoretical and experimental realization of applying near-term gate-model quantum computers to nonlinear temporal information processing tasks, opening an avenue for time series modeling and signal processing applications of these devices.

II. TEMPORAL INFORMATION PROCESSING TASKS

Two challenging temporal information processing problems are posed. The first is the multi-step ahead prediction problem, where we are given a length L input sequence $u = \{u_1, \ldots, u_L\}$ and the corresponding output sequence of the target map $y = M(u, \mathbf{x}_0) = \{y_1, \ldots, y_L\}$ that depends on the initial condition $\mathbf{x}_0$. The first $L_T < L$ input-output data pair $(u|_{1:L_T}, y|_{1:L_T})$, where $u|_{1:L_T} = \{u_1, \ldots, u_{L_T}\}$ is the train data (in the sequel, we use the input-output train data during $l = 5, \ldots, L_T$; see the next section on "Experimental demonstration" for a discussion. The goal is to use the train data to optimize the parameters $\mathbf{w}$ of another map $\overline{M}_{\mathbf{w}}(\cdot, \overline{\mathbf{x}}_0)$, so that the outputs $\overline{y}|_{L_T+1:L} = \overline{M}_{\mathbf{w}}(u|_{L_T+1:L}, \overline{\mathbf{x}}_0)$ approximate $y|_{L_T+1:L}$. The second problem is the map emulation problem, that is to optimize $\mathbf{w}$ of $\overline{M}_{\mathbf{w}}(\cdot, \overline{\mathbf{x}}_0)$ to emulate $M(\cdot, \mathbf{x}_0)$ using K different input-output train data pairs (u^k, y^k). The input u^k and output $y^k = M(u^k, \mathbf{x}_0)$ sequences have the same length L' for $k = 1, \ldots, K$, so that the total number of train data is KL' (we will again use train data during $l = 5, \ldots, l = L'$). When given a previously unseen input sequence u^{K+1}, the task is for $\overline{y}^{K+1} = \overline{M}_{\mathbf{w}}(u^{K+1}, \overline{\mathbf{x}}_0)$ to approximate the entire output sequence $y^{K+1} = M(u^{K+1}, \mathbf{x}_0)$. To obtain each input-output data pair, the initial conditions $\mathbf{x}_0$ and $\overline{\mathbf{x}}_0$ of the two maps $M(\cdot, \mathbf{x}_0)$ and $\overline{M}_{\mathbf{w}}(\cdot, \overline{\mathbf{x}}_0)$ are reset.

III. RESERVOIR COMPUTING

To say that M has short-term memory means that the output y_L becomes less dependent of initial condition $\mathbf{x}_0$ as L is increased and the output at time k becomes increasingly less dependent on input samples from much earlier times $l \ll k$; see Appendix Sec. A 2. To approximate this class of maps, RC exploits nonlinear
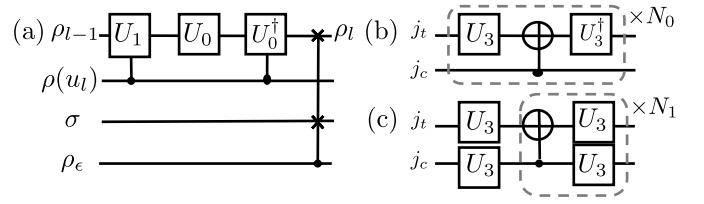

FIG. 1. (a) Quantum circuit interpretation of the QR dynamics Eq. (1), where $\rho(u_l) = u_l|0\rangle\langle0| + (1-u_l)|1\rangle\langle1|$ and $\rho_\epsilon = (1-\epsilon)|0\rangle\langle0| + \epsilon|1\rangle\langle1|$. (b) and (c) show the circuit schematics for U_0 and U_1, respectively.

dynamical systems to project the input u_l into a reservoir state $\mathbf{x}_l$ at time l. If the reservoir eventually forgets its initial conditions, the state $\mathbf{x}_l$ contains short-term memory of the sequential inputs. The dynamics of the reservoir is random and never requires training. The tunable parameters $\mathbf{w}$ appear in a readout function $h_{\mathbf{w}}$, which linearly combines the elements of $\mathbf{x}_l$ into an output $\overline{y}_l = h_{\mathbf{w}}(\mathbf{x}_l) = \mathbf{w}^T\mathbf{x}_l + \mathbf{w}_c$, with a bias term $\mathbf{w}_c$. The parameters $\mathbf{w}$ are optimized using linear regression to minimize an empirical mean squared-error between $y_{1:L_T}$ and $\overline{y}_{1:L_T}$. The combination of the reservoir and the readout function produces an input-output map $\overline{M}_{\mathbf{w}}$.

Echo-state networks, one of the pioneering classical RC schemes, have been numerically demonstrated to achieve state-of-the-art performance in chaotic system modeling [27]. Subsequent hardware realizations of classical RC proposals suggest empirically that for certain tasks, such as spoken digit recognition, the reservoir state dimension plays a role in the RC's task performance [6, 21, 22].

IV. UNIVERSAL QUANTUM RESERVOIR COMPUTERS

We propose to use a QR, with a view towards taking advantage of its exponentially large state space. For an n-qubit QR, we define its state ρ_l's dissipative evolution, with initial condition ρ_0, as

$$\rho_l = T(u_l)\rho_{l-1} = (1-\epsilon)(u_l T_0 + (1-u_l)T_1)\rho_{l-1} + \epsilon\sigma, \quad (1)$$

where $0 < \epsilon \leq 1$, $0 \leq u_l \leq 1$ and σ is an arbitrary but fixed density matrix. Here T_0 and T_1 are two random but fixed completely positive trace-preserving (CPTP) maps. In particular, we can choose $T_j(\rho_{l-1}) = U_j\rho_{l-1}U_j^\dagger$ ($j = 0, 1$) with arbitrary unitaries U_j implemented by native quantum gates of NISQ devices. High fidelity for the unitaries U_j is not strictly required; if the noise in the system is such that it acts to replace $U_j\rho U_j^\dagger$ with another CPTP map $T_j'(\rho)$, the QR dynamics is again of the form Eq. (1). This form of QR dynamics has a natural quantum circuit interpretation, see Fig. 1(a) and Appendix Sec. D for an explanation. The input u_l is encoded as the probability of applying U_0 or U_1. The term ϵ controls the speed at which the QR forgets its initial condition (i.e., the "memory" of the QR, with $\epsilon =$

1 being memoryless). Since ρ_l grows exponentially in size with the number of qubits, we obtain partial information about ρ_l by measuring each qubit in the Pauli Z basis to obtain $\text{Tr}(\rho_l Z^{(i)})$ for $i = 1, \ldots, n$, where $Z^{(i)}$ is acting on qubit i. We define a linear readout function

$$\overline{y}_l = h_{\mathbf{w}}(\rho_l) = \sum_{i=1}^{n} \mathbf{w}_i \text{Tr}(\rho_l Z^{(i)}) + \mathbf{w}_c. \qquad (2)$$

Eqs. (1) and (2) define a QR that implements an input-output map $\overline{M}_{\mathbf{w}}$. More generally, our QR proposal consists of multiple non-interacting subsystems with differing numbers of qubits, similar to the proposals in [24, 26], and a more general form of the readout function can be used. Our class of QRs possesses the universal computational power for approximating nonlinear short-term memory maps. Furthermore, this universality property is invariant under stationary Markovian noise and time-invariant readout error. See Appendix Sec. A, B and C for both proofs. This is supported by experiments, demonstrating that the proposed QRs can tackle nonlinear tasks under the gate and readout error levels achievable by available superconducting NISQ hardware.

V. EXPERIMENTAL DEMONSTRATION

Four nonlinear tasks are chosen to carefully test different computational aspects of the QR proposal. Tasks I and II test the QR's ability to learn high-dimensional and highly nonlinear maps. Task III tests the short-term memory ability and Task IV is a long-term memory map for testing the capability of the QR beyond its theoretical guarantee. For all experimental and numerical details, see Appendix Sec. E. We implement four distinct QRs on three IBM superconducting quantum processors [28]. A 4-qubit and a 10-qubit QRs are implemented on the 20-qubit Boeblingen device; qubits with lower gate errors and longer coherence times are chosen. The 5-qubit Ourense and Vigo devices are used for two distinct 5-qubit QRs. These 5-qubit quantum devices admit simpler qubit couplings but lower gate errors than the 20-qubit Boeblingen device. Through comparison among the four QRs, we can investigate the impact of the size of QRs, the complexity of quantum circuits implementing the QR dynamics and the intrinsic hardware noise on the QRs' approximation performance.

We require the QRs to forget initial conditions for approximating short-term memory maps. Traditionally, initial conditions are washed-out with a sufficiently long input sequence until reaching a steady state. Here we bypass the washout by choosing $\sigma = (|0\rangle\langle0|)^n$ and U_0 so that $|0\rangle^n$ is the steady state of Eq. (1) under $u_l = 1$, meaning that we can initialize the QR circuits in $|0\rangle^n$. Furthermore, U_0 and U_1 should be different and hardware-efficient but sufficiently complex to produce nontrivial quantum dynamics. We choose $U_0(\boldsymbol{\theta}) = \prod_{j=1}^{N_0} \left(U_3^{(j_t)}(\boldsymbol{\theta}_{j_t}) \text{CX}_{j_c j_t} U_3^{(j_t)}(\boldsymbol{\theta}_{j_t})^\dagger \right)$ and

$U_1(\boldsymbol{\phi}) = \bigotimes_{i=1}^{n} U_3^{(i)}(\boldsymbol{\phi}_{0_i}) \prod_{j=1}^{N_1} \left(\bigotimes_{i=1}^{n} U_3^{(i)}(\boldsymbol{\phi}_{j_i}) \text{CX}_{j_c j_t} \right)$,

where $U_3^{(i)}$ is an arbitrary rotation on single qubit i [29] and $\text{CX}_{j_c j_t}$ is the CNOT gate with control qubit j_c and target qubit j_t; see Fig. 1(b) and (c) for the circuit schematics. The gate parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are uniformly randomly sampled from $[-2\pi, 2\pi]$. The numbers of layers N_0 and N_1 are sufficiently large to couple all qubits linearly while respecting the coherence limits of these devices. Owing to the more flexible qubit couplings in the Boeblingen device, circuits implementing the 4-qubit and 10-qubit QRs have more gate and random parameters than the 5-qubit QRs'.

When restricted to pure state preparation, instead of realizing Fig. 1(a), we efficiently implement QRs through Monte Carlo sampling as in [30]. At each time step l, we draw N_m random circuits, each circuit implements U_0 and U_1 with probabilities $(1 - \epsilon)u_l$ and $(1 - \epsilon)(1 - u_l)$, respectively; otherwise the circuit is reset in $|0\rangle^n$ with probability ϵ. We choose a sufficiently large $N_m = 1024$ and $\epsilon = 0.1$ for a moderate short-term memory. To estimate $\text{Tr}(\rho_l Z^{(i)})$, circuits implementing the QRs on the Boeblingen device and the 5-qubit QRs are run for $S = 1024$ and $S = 8192$, respectively. These are chosen according to circuit execution times of the devices.

We apply the four QRs to the four nonlinear tasks on the multi-step ahead prediction and map emulation problems. To implement the same washout as for the QRs for each target map, we inject a constant input sequence $u_l = 1$ of length 50 followed by train and test inputs uniformly randomly sampled from $u_l \in [0, 1]$. This change in the input statistics leads to a transitory target output response. We remove the associated transients by discarding the first four target input-output data and the corresponding QR experimental data, see Appendix Sec. E 2 for all data. For the multi-step ahead problem, train and test time steps run from $l = 5$ to $L_T = 23$ and $L_T + 1 = 24$ to $L = 30$, respectively. For the map emulation problem, $K = 2$ train input-output pairs running from $l = 5$ to $L' = 24$ are used, followed by one unseen test input-output pair with the same time steps. To harness the flexibility of the QR approach, a multi-tasking technique is used, in which the four QRs are evolved and the estimates of $\text{Tr}(\rho_l Z^{(i)})$ for all time steps are recorded once, whereas the readout parameters $\mathbf{w}$ are optimized independently for each task. That is a fixed QR dynamics, with fixed gate parameter values, is exploited for multiple tasks simultaneously. We evaluate and compare the task performance of QRs using the normalized mean-squared error, NMSE $= \sum_{l=L_T+1}^{L} |y_l - \overline{y}_l|^2 / \Delta_y^2$, where $\mu = \frac{1}{L - L_T} \sum_{l=L_T+1}^{L} y_l$ and $\Delta_y^2 = \sum_{l=L_T+1}^{L} (y_l - \mu)^2$ between prediction $\overline{y}_{L_T+1:L}$ and target $y_{L_T+1:L}$. While the success of experimental demonstration of hybrid quantum-classical algorithms often requires error mitigation techniques to reduce the effect of decoherence [31, 32], we remark that our results are obtained without any process or readout error mitigation.

As the number of qubits increases, the 10-qubit Boe-
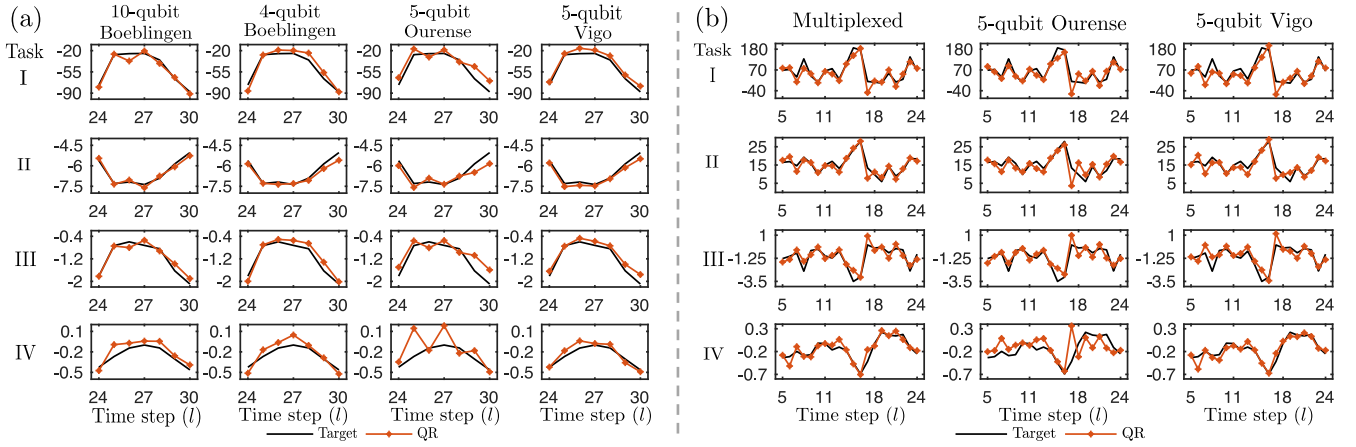
iv

FIG. 2. (a) Shows the QRs' predicted outputs for the multi-step prediction problem, rows and columns correspond to different tasks and QRs, respectively. (d) Shows the QRs' predicted outputs for the map emulation problem, first column corresponds to the multiplexed QR.

blingen QR is expected to perform better than other QRs. For the multi-step ahead prediction problem, we observe that two qubits in the 10-qubit Boeblingen QR experienced significant time-varying deviations between the experimental data and simulation results on the Qiskit simulator; see Appendix Sec. E 4 for a discussion. To remedy this issue, we set the corresponding elements of **w** to be zeros. The resulting 10-qubit Boeblingen QR (with NMSE<0.08) outperforms other QRs with a smaller number of qubits on the first three tasks, and achieves an almost two-fold performance improvement on Task II; see Table I for all NMSEs. The 10-qubit Boeblingen QR predicted outputs follow the target outputs relatively closely as shown in Fig. 2(a). The 5-qubit Ourense QR admits very simple dynamics–with U_0 consisting only of CNOTs and $U_1(\phi) = \bigotimes_{i=1}^{5} U_3^{(i)}(\phi_i)$ whereas the 5-qubit Vigo QR has more gate operations and random gate parameters. The 5-qubit Ourense QR is outperformed by the 5-qubit Vigo QR in all tasks. Considering that the Ourense and Vigo devices have similar noise characteristics and the same qubit coupling map, this suggests that the QR performance can be improved by choosing a more complex quantum circuit, in the sense of having a longer gate sequence.
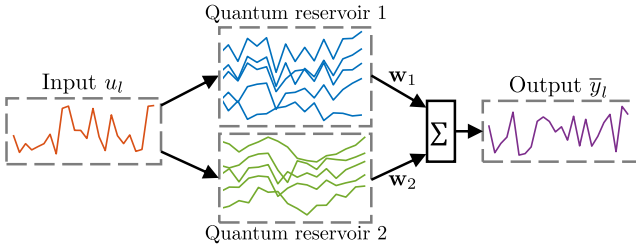
The 10-qubit Boeblingen QR performs better on all tasks than the 5-qubit QRs except on Task IV. This could be due to the impact of the higher noise level in the Boeblingen device and the fact that the output sequence is generated by a map that is not known to have short-term memory, see Appendix Sec. E 5 for the hardware specifications. Our universal class of QRs can exploit the property of spatial multiplexing as initially proposed in Ref. [24]; see also [26] and Fig. 3 for an illustration. Outputs of distinct and non-interacting 5-qubit QRs can be combined linearly to harness the computational features of both members. Since the combined Ourense and Vigo devices have 10 qubits overall as with the 10-qubit Boeblingen QR but with lower noise levels, it would be meaningful to combine the 5-qubit Vigo and Ourense QRs via spatial multiplexing on the map emulation problem. The results of this multiplexing is summarized in Table II.

TABLE I. NMSEs on the multi-step ahead prediction problem.

Task	10-qubit Boeblingen	4-qubit Boeblingen	5-qubit Ourense	5-qubit Vigo
I	0.051	0.088	0.24	0.070
II	0.043	0.10	0.25	0.081
III	0.079	0.092	0.34	0.11
IV	0.47	0.41	2.3	0.20

TABLE II. NMSEs on the map emulation problem.

Task	Multiplexed QR	5-qubit Ourense	5-qubit Vigo
I	0.20	0.26	0.32
II	0.16	0.46	0.26
III	0.25	0.30	0.36
IV	0.20	1.1	0.17

FIG. 3. The spatial multiplexing schematic. The same input sequence is injected into two distinct 5-qubit QRs. The internal states $\mathrm{Tr}(\rho_l Z^{(i)})$ of the two QRs are linearly combined to form a single output.

The combination of two 5-qubit QRs as discussed

above achieves NMSE $= 0.20, 0.16, 0.25, 0.20$ for the four tasks without any readout or process error mitigation. The predicted multiplexed QR outputs corresponding to the unseen inputs follow the target outputs relatively closely as shown in Fig. 2(b). Without spatial multiplexing, the 5-qubit Ourense or the 5-qubit Vigo QR show a worse performance in the first three tasks; see Table II. The spatial multiplexed 5-qubit QR combines computational features from the constituent QRs and can achieve comparable performance to the individual members as well as gaining an almost two-fold performance boost on Task II. We anticipate that spatial multiplexing of QRs with more complex circuit structures and a larger number of qubits can lead to further performance improvements.

VI. CONCLUSION

We propose the first universal class of quantum reservoir computers that utilizes dissipative quantum dynamics and is readily implementable on available noisy gate-model quantum hardware for temporal information processing. Our approach harnesses noisy quantum circuits of arbitrary depths, signifying the potential of these devices beyond binary data classification or static regression applications. The theoretical analysis is supported by proof-of-concept experiments on superconducting quantum processors, demonstrating that small-scale noisy quantum reservoirs can perform complex nonlinear tasks in the absence of readout and process error mitigation techniques.

Our approach is scalable in the number of qubits by offloading exponentially costly computations to noisy quantum systems and utilizing classical algorithms with linear cost to process temporal data. Moreover, when implemented on NISQ devices, the micro-second timescale for the evolution of the quantum reservoir suggests its potential for real-time fast signal processing tasks. Guided by our theory, we applied the spatial multiplexing technique initially proposed in [24], and demonstrate experimentally that exploiting distinct computational features of multiple small noisy quantum reservoirs can lead to a computational boost. As NISQ hardware becomes increasingly accessible and the noise level is continually reduced,we anticipate that the quantum reservoir approach will find useful applications in a broad range of scientific disciplines that employ time series modeling and analysis. We are also optimistic for useful applications to be possible even for a noise level above the threshold for continuous quantum error correction.

VII. ACKNOWLEDGMENTS

Appendix A: Universality for nonlinear fading memory maps

We first define notation for the rest of this section. Let $K([0,1])$ be the set of infinite sequences $u = \{\ldots, u_{-1}, u_0, u_1, \ldots\}$ such that $u_l \in [0,1]$ for all $l \in \mathbb{Z}$. Let $K^+([0,1])$ and $K^-([0,1])$ be subsets of $K([0,1])$ for which the indices are restricted to $\mathbb{Z}^+ = \{1, 2, \ldots\}$ and $\mathbb{Z}^- = \{\ldots, -2, -1, 0\}$, respectively. For any complex matrix A, $\|A\|_p = \text{Tr}(\sqrt{A^\dagger A}^p)^{1/p}$ is the Schatten p-norm for some $p \in [1, \infty)$. For any operator T, the induced operator norm is $\|T\|_{p-p} = \sup_{A \in \mathbb{C}^{n \times n}, \|A\|_p = 1} \|T(A)\|_p$. Let $\mathcal{D}(2^n)$ denotes the set of $2^n \times 2^n$ density operators.

Consider a map M that maps an infinite input sequence $u \in K([0,1])$ to a real infinite output sequence $y \in K(\mathbb{R})$. We say that M is w-fading memory if there exists a decreasing sequence $w = \{w_0, w_1, \ldots\}$ with $\lim_{l \to \infty} w_l = 0$, such that for any $u, v \in K^-([0,1])$, we have $|M(u)_0 - M(v)_0| \to 0$ whenever $\sup_{l \in \mathbb{Z}^-} |w_{-l}(u_l - v_l)| \to 0$. Here $M(u)_l = y_l$ is the output sequence at time l. We also require M to be causal and time-invariant as in Ref. [26], meaning that the output of M at time l only depends the input up to and including that time, and its outputs are invariant under time-shifts. Now we are interested in approximating M with a time-invariant fading memory map $\overline{M}$ produced by a quantum reservoir computer.

1. The convergence property

Since M is fading memory, the map $\overline{M}$ must also forget its initial condition ρ_0. This is the convergence property [33] or the echo-state property [27]. We now show that the QR dynamics given by Eq. (1) in the main article is convergent with respect to any $u \in K^+([0,1])$. For any $\rho, \sigma \in \mathcal{D}(2^n)$, $u_l \in [0,1]$ and $\epsilon \in (0,1]$,

$$\|T(u_l)(\rho - \sigma)\|_1 = (1 - \epsilon) \|(u_l T_0 + (1 - u_l)T_1)(\rho - \sigma)\|_1 \le (1 - \epsilon)\|\rho - \sigma\|_1, \tag{A1}$$

where the last inequality follows from [34, Theorem 9.2] and the fact that the convex combination $u_l T_0 + (1 - u_l)T_1$ of CPTP maps is again a CPTP map. Now let $\rho_{1,0}$ and $\rho_{2,0}$ be two arbitrary initial density operators, using the

inequality Eq. (A1) k times, we have

$$\|\rho_{1,k} - \rho_{2,k}\|_1 = \left\|\left(\overleftarrow{\prod}_{l=1}^k T(u_l)\right)(\rho_{1,0} - \rho_{2,0})\right\|_1 \leq (1-\epsilon)^k \|\rho_{1,0} - \rho_{2,0}\|_1 \leq 2(1-\epsilon)^k,$$

where $\overleftarrow{\prod}_{l=1}^k T(u_l)$ is the time-composition of $T(u_l)$ from right to left. This implies that there exists a steady state ρ_* for T, depending only on the input sequence $u \in K^+([0,1])$, such that for any initial condition ρ_0, we have

$$\lim_{k\to\infty} \|\rho_k - \rho_*\|_1 = \lim_{k\to\infty} \left\|\left(\overleftarrow{\prod}_{l=1}^k T(u_l)\right)\rho_0 - \rho_*\right\|_1 = 0.$$

We now define a general form of the quantum reservoir (QR) dynamics, in which each QR consists of N non-interacting subsystems initialized in a product state of the subsystems, where each subsystem j evolves according to the completely positive trace-preserving map (CPTP) $T^{(j)}$ with the same form as Eq. (1) in the main article. That is, at any time l the density operator ρ_l of the QR is governed by the dynamics,

$$\rho_l = \bigotimes_{j=1}^N \rho_l^{(j)} = T(u_l)\rho_{l-1} = \bigotimes_{j=1}^N T^{(j)}(u_l)\rho_{l-1}^{(j)}. \tag{A2}$$

Here the j-th subsystem with n_j qubits undergoes the dissipative evolution,

$$\rho_l^{(j)} = T^{(j)}(u_l)\rho_{l-1}^{(j)} = (1-\epsilon_j)\left(u_l T_0^{(j)} + (1-u_l)T_1^{(j)}\right)\rho_{l-1}^{(j)} + \epsilon_j K_{\sigma_j}, \tag{A3}$$

where $\epsilon_j \in (0,1]$, $\rho_l^{(j)}$ is the j-th subsystem density operator at time l and σ_j is an arbitrary but fixed density operator. For any $X \in \mathbb{C}^{2^{n_j} \times 2^{n_j}}$ we have $K_{\sigma_j}(X) = \text{Tr}(X)\sigma_j$, therefore the CPTP map K_{σ_j} maps density operators to the constant density matrix σ_j. The CPTP maps $T_0^{(j)}$ and $T_1^{(j)}$ are arbitrary but fixed and input-independent. To show that $T(u_l) = \bigotimes_{j=1}^N T^{(j)}(u_l)$ is again convergent when the subsystems are initialized in a product state as in the above, we can apply the same argument as for [35, Lemma 5].

Consider two CPTP maps $T^{(1)}(u_l)$ and $T^{(2)}(u_l)$ of the form Eq. (A2). Let $\rho_{1,0}\otimes\sigma_{1,0}$ and $\rho_{2,0}\otimes\sigma_{2,0}$ be two arbitrary initial product states. Then,

$$\|\rho_{1,k} \otimes \sigma_{1,k} - \rho_{2,k} \otimes \sigma_{2,k}\|_1$$
$$= \left\|\left(\overleftarrow{\prod}_{l=1}^k T^{(1)}(u_l) \otimes T^{(2)}(u_l)\right)(\rho_{1,0}\otimes\sigma_{1,0} - \rho_{2,0}\otimes\sigma_{2,0})\right\|_1$$
$$\leq \left\|\left(\overleftarrow{\prod}_{l=1}^k T^{(1)}(u_l) \otimes T^{(2)}(u_l)\right)(\rho_{1,0}\otimes\sigma_{1,0} - \rho_{2,0}\otimes\sigma_{1,0})\right\|_1 + \left\|\left(\overleftarrow{\prod}_{l=1}^k T^{(1)}(u_l) \otimes T^{(2)}(u_l)\right)(\rho_{2,0}\otimes\sigma_{1,0} - \rho_{2,0}\otimes\sigma_{2,0})\right\|_1$$
$$= \left\|\left(\overleftarrow{\prod}_{l=1}^k T^{(1)}(u_l)\right)(\rho_{1,0} - \rho_{2,0}) \otimes \left(\overleftarrow{\prod}_{l=1}^k T^{(2)}(u_l)\right)\sigma_{1,0}\right\|_1 + \left\|\left(\overleftarrow{\prod}_{l=1}^k T^{(1)}(u_l)\right)\rho_{2,0} \otimes \left(\overleftarrow{\prod}_{l=1}^k T^{(2)}(u_l)\right)(\sigma_{1,0} - \sigma_{2,0})\right\|_1$$
$$= \left\|\left(\overleftarrow{\prod}_{l=1}^k T^{(1)}(u_l)\right)(\rho_{1,0} - \rho_{2,0})\right\|_1 \|\sigma_{1,k}\|_1 + \left\|\left(\overleftarrow{\prod}_{l=1}^k T^{(2)}(u_l)\right)(\sigma_{1,0} - \sigma_{2,0})\right\|_1 \|\rho_{2,k}\|_1$$
$$\leq 2(1-\epsilon_1)^k + 2(1-\epsilon_2)^k.$$

Therefore, $T^{(1)}(u_l) \otimes T^{(2)}(u_l)$ is again convergent with respect to all $u \in K^+([0,1])$.

2. The fading memory property

Now we associate a readout function to Eq. (A2). Let $n = \sum_{j=1}^N n_j$ be the total number of qubits, define

$$\overline{y}_l = h_{\mathbf{w}}(\rho_l) = \sum_{d=1}^R \sum_{i_1=1}^n \sum_{i_2=i_1+1}^n \cdots \sum_{i_n=i_{n-1}+1}^n \sum_{r_{i_1}+\cdots+r_{i_n}=d} \mathbf{w}_{i_1,\ldots,i_n}^{r_{i_1},\ldots,r_{i_n}} \langle Z^{(i_1)}\rangle_l^{r_{i_1}} \cdots \langle Z^{(i_n)}\rangle_l^{r_{i_n}} + \mathbf{w}_c, \tag{A4}$$

where $\langle Z^{(i_j)}\rangle_l = \text{Tr}(Z^{(i_j)}\rho_l)$. This readout function is a multivariate polynomial in variables $\langle Z^{(i_j)}\rangle_l$ and R is its degree. When $R = 1$ we have a linear readout function, which is used in all experiments.

The quantum reservoir dynamics T in Eq. (A2) and readout function $h_{\mathbf{w}}$ in Eq. (A4) define a unique functional $\overline{M}_{(T,h_{\mathbf{w}})}$ depending on T and $h_{\mathbf{w}}$; see [11, 12, 26] for a detailed discussion. For any $u \in K^-([0,1])$ and any initial condition $\rho_{-\infty}$, the functional is given by

$$\overline{M}_{(T,h_{\mathbf{w}})}(u) = h_{\mathbf{w}}\left(\left(\overrightarrow{\prod}_{j=0}^{\infty} T(u_{-j})\right)\rho_{-\infty}\right),$$

where $\overrightarrow{\prod}_{j=0}^{\infty} T(u_{-j}) = \lim_{N\to\infty} T(u_0)\cdots T(u_{l-N})$ and the limit is point-wise. We can restate the fading memory property in terms of continuity of $\overline{M}_{(T,h_{\mathbf{w}})}$ with respect to a certain norm. Given a decreasing sequence w with $\lim_{l\to\infty} w_l = 0$ and any $u \in K^-([0,1])$, define a weighted norm $\|u\|_w = \sup_{l\in\mathbb{Z}^-}|u_l|w_{-l}$. The map $\overline{M}_{(T,h_{\mathbf{w}})}$ is w-fading memory if it is continuous in $(K^-([0,1]), \|\cdot\|_w)$.

We now show that $\overline{M}_{(T,h_{\mathbf{w}})}$ is w-fading memory for any decreasing sequence w. Using the same argument in [26, Lemma 3], it follows that $\overline{M}_{(T,h_{\mathbf{w}})}$ is w-fading memory if $T^{(j)}(u_l)$ is continuous with respect to the inputs $u_l \in [0,1]$ for all $j = 1, \ldots, N$. If fact, we show that $T^{(j)}(u_l)$ is uniformly continuous. Let $x, y \in [0,1]$ and $A \in \mathbb{C}^{2^{n_j} \times 2^{n_j}}$,

$$\|T^{(j)}(x) - T^{(j)}(y)\|_{1-1} = \sup_{A\in\mathbb{C}^{2^{n_j}\times 2^{n_j}}, \|A\|_1=1}\left\|\left(T^{(j)}(x) - T^{(j)}(y)\right)A\right\|_1$$

$$= (1-\epsilon_j)|x-y| \sup_{A\in\mathbb{C}^{2^{n_j}\times 2^{n_j}}, \|A\|_1=1}\left\|T_0^{(j)}(A) - T_1^{(j)}(A)\right\|_1$$

$$\leq (1-\epsilon_j)|x-y|\left(\left\|T_0^{(j)}\right\|_{1-1} + \left\|T_1^{(j)}\right\|_{1-1}\right)$$

$$\leq 2(1-\epsilon_j)|x-y|,$$

where the last inequality follows from [36, Theorem 2.1]. We remark that [26, Lemma 3] is stated with respect to the Schatten $p = 2$ norm, but the same argument holds for $p = 1$.

3. The universality property

Now consider the family $\mathcal{M}$ of maps $\overline{M}_{(T,h_{\mathbf{w}})}$ arising from different numbers of qubits, different dynamics $T(\cdot)$ and different readout parameters $\mathbf{w}$ and degree R in $h_{\mathbf{w}}$. To show that $\mathcal{M}$ is universal for approximating nonlinear w-fading memory with any decreasing sequence w, we apply the Stone-Weierstrass Theorem [37, Theorem 7.3.1] to show that $\mathcal{M}$ is dense in the set of all continuous functions $C(K^-([0,1]), \|\cdot\|_w)$ defined on $(K^-([0,1]), \|\cdot\|_w)$. It has been shown that the space $(K^-([0,1]), \|\cdot\|_w)$ is a compact metric space [12, Lemma 2]. We now state the Stone-Weierstrass Theorem.

Theorem 1 (Stone-Weierstrass) *Let E be a compact metric space and $C(E)$ be the set of real-valued continuous functions defined on E. If a subalgebra A of $C(E)$ contains the constant functions and separates points of E, then A is dense in $C(E)$.*

The family $\mathcal{M}$ forms a polynomial algebra follows from [26, Lemma 5] and the observation that for any QR dynamics $T_1(u_l) = \bigotimes_{j=1}^{N_1} T_1^{(j)}(u_l)$ and $T_2(u_l) = \bigotimes_{j=1}^{N_2} T_2^{(j)}(u_l)$ of two (non-interacting subsystems) indexed by the superscript k of the form Eq. (A2), we have $T(u_l)(\rho_1 \otimes \rho_2) = T_1(u_l)\rho_1 \otimes T_2(u_l)\rho_2$ is again of the form Eq. (A2) for any density operator ρ_k of subsystem k. Furthermore, $T(u_l) = T_1(u_l) \otimes T_2(u_l)$ is again convergent when initialized in a product state of the subsystems. Therefore, the family $\mathcal{M}$ forms a polynomial algebra consisting of w-fading memory maps.

Lastly, constant functions can be obtained by setting $\mathbf{w}_{i_1,\ldots,i_n}^{r_{i_1},\ldots,r_{i_n}} = 0$. It remains to show that $\mathcal{M}$ separates points in $K^-([0,1])$. That is, for any distinct $u,v \in K^-([0,1])$ with $u_l \neq v_l$ for at least one l, we need to find a map $\overline{M}_{(T,h_{\mathbf{w}})} \in \mathcal{M}$ such that $\overline{M}_{(T,h_{\mathbf{w}})}(u) \neq \overline{M}_{(T,h_{\mathbf{w}})}(v)$. We show that we can construct a single-qubit quantum reservoir with this property.

Consider a single-qubit quantum reservoir with a linear readout function ($n = 1, R = 1, N = 1$). For the rest of this proof, we drop the subsystem index. This quantum reservoir consists of one system qubit and one ancilla qubit denoted as ρ_a. Choose the dynamics to be

$$\rho_l = T(u_l)\rho_{l-1} = (1-\epsilon)\left(u_l\text{Tr}_a\left(e^{-iH}(\rho_{l-1}\otimes\rho_a^0)e^{iH}\right) + (1-u_l)\text{Tr}_a\left(e^{-iH}(\rho_{l-1}\otimes\rho_a^1)e^{iH}\right)\right) + \epsilon K_{\frac{I}{2}}, \quad \text{(A5)}$$

where $\rho_a^j = |j\rangle\langle j|$ for $j = 0, 1$, Tr_a denotes the partial trace over ancilla ρ_a and $\epsilon \in (0,1)$. The map $K_{\frac{I}{2}}$ is a CPTP map defined as $K_{\frac{I}{2}}(X) = \text{Tr}(X)\frac{I}{2}$ for any $X \in \mathbb{C}^{2\times 2}$. The Hamiltonian H is of the Ising type $H = J(X^{(0)}X^{(1)} +$

$Y^{(0)}Y^{(1)}) + \alpha \sum_{j=0}^{1} Z^{(j)}$, where $X^{(j)}, Y^{(j)}$ and $Z^{(j)}$ are the Pauli X, Y and Z operators on qubit j, with $j = 0$ being the ancilla qubit.

We order an orthogonal basis for $\mathbb{C}^{2\times 2}$ as $\{I, Z, X, Y\}$. The matrix representation of the CPTP map Eq. (A5) is

$$\overline{T}(u_l) = |00\rangle\langle 00| + (1 - \epsilon)\begin{pmatrix} 0 & 0 & 0 & 0 \\ \sin^2(2J)(2u_l - 1) & \cos^2(2J) & 0 & 0 \\ 0 & 0 & \cos(2J)\cos(2\alpha) & -\cos(2J)\sin(2\alpha) \\ 0 & 0 & \cos(2J)\sin(2\alpha) & \cos(2J)\cos(2\alpha) \end{pmatrix}.$$

Since Eq. (A5) is convergent, we can choose any initial condition $\rho_{-\infty} = |0\rangle\langle 0|$ with the corresponding vector representation $\overline{\rho}_\infty = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix}$. Taking a linear readout function, for $u \in K^-([0,1])$, the quantum reservoir implements a functional

$$\overline{M}_{(T,h_\mathbf{w})}(u) = 2\mathbf{w}_1\left[\left(\overrightarrow{\prod}_{j=0}^{\infty}\overline{T}(u_{-j})\right)\overline{\rho}_{-\infty}\right]_2 + \mathbf{w}_c,$$

where $[\cdot]_2$ is the second element of the vector corresponding to $\mathrm{Tr}(Z\rho_0)/2$. Now given two distinct inputs $u, v \in K^-([0,1])$, suppose that $u_0 \neq v_0$. Then choose J such that $\cos^2(2J) = 0$ and therefore,

$$\overline{M}_{(T,h_\mathbf{w})}(u) - \overline{M}_{(T,h_\mathbf{w})}(v) = 2\mathbf{w}_1(1 - \epsilon)(u_0 - v_0) \neq 0.$$

Suppose $u_0 = v_0$, note that in general

$$\overline{M}_{(T,h_\mathbf{w})}(u) = \mathbf{w}_1\sin^2(2J)(1 - \epsilon)\sum_{j=0}^{\infty}\left((1 - \epsilon)\cos^2(2J)\right)^j(2u_{-j} - 1).$$

Choose $\epsilon \in (0,1)$ and J such that $(1 - \epsilon)\cos^2(2J) \in (0, 1 - \epsilon)$. Then the above is a convergent power series and the subtraction is well-defined:

$$\overline{M}_{(T,h_\mathbf{w})}(u) - \overline{M}_{(T,h_\mathbf{w})}(v) = 2\mathbf{w}_1\sin^2(2J)(1 - \epsilon)\sum_{j=0}^{\infty}\left((1 - \epsilon)\cos^2(2J)\right)^j(u_{-j} - v_{-j}).$$

The above is a power series of the form

$$f(\theta) = 2\mathbf{w}_1\sin^2(2J)(1 - \epsilon)\sum_{j=0}^{\infty}\theta^j(u_{-j} - v_{-j}),$$

where $f(\theta)$ has a nonzero radius of convergence and is non-constant since $\theta = (1 - \epsilon)\cos^2(2J) \in (0, 1 - \epsilon)$ and $(1 - \epsilon)\sin^2(2J) \in (0, 1 - \epsilon)$. Furthermore, since we assume that $u_0 = v_0$, we have $f(0) = 0$. Invoking [38, Theorem 3.2], there exists $\beta > 0$ such that $f(\theta) \neq 0$ for all $|\theta| \leq \beta, \theta \neq 0$. This concludes the proof for separation of points. The universality of $\mathcal{M}$ now follows from the Stone-Weierstrass Theorem.

Appendix B: Robustness against stationary Markovian noise

The quantum reservoir model proposed is robust against stationary Markovian noise. For the j-th subsystem in the QR dynamics Eq. (A2), a stationary Markovian noise process during some time interval $\tau(l - 1) \leq t \leq \tau l$, where l is the time step and $\tau > 0$, can be modeled as a CPTP map $\mathcal{T}^{(j)}$ for all $l \in \mathbb{Z}^+$. The j-th subsystem in the QR dynamics Eq. (A2) under this noise is

$$\rho_l^{(j)} = \left(\mathcal{T}^{(j)} \circ T^{(j)}(u_l)\right)\rho_{l-1}^{(j)} = (1 - \epsilon_j)\left(u_l\mathcal{T}^{(j)} \circ T_0^{(j)} + (1 - u_l)\mathcal{T}^{(j)} \circ T_1^{(j)}\right)\rho_{l-1}^{(j)} + \epsilon_j\mathcal{T}^{(j)}(\sigma_j),$$

where $\mathcal{T}^{(j)} \circ T_k^{(j)}$ is again a CPTP for $k = 0, 1$ and $\mathcal{T}^{(j)}(\sigma_j) = \sigma_j'$ is again a density operator. That is, the universal family $\mathcal{M}$ is *invariant and remains universal* under stationary Markovian noise.

Appendix C: Robustness against time-invariant readout error

The universal class of QRs is robust against time-invariant readout error whenever a linear readout function is used. Let $\mathcal{B} = \{|i\rangle\}$ be the computational basis for an n-qubit system, with $i = 1, \ldots, 2^n$. The readout error is characterized by a measurement calibration matrix A whose i,j-th element $A_{i,j} = \Pr(i|j)$ is the probability of measuring the state $|i\rangle \in \mathcal{B}$ given that the state is prepared in the state $|j\rangle \in \mathcal{B}$.

We employ the readout error correction method described in Ref. [4]. For an n-qubit QR, at each time step l, we execute 2^n calibration circuits with each circuit initialized in one of the 2^n computational basis elements. The outcomes are used to create the measurement calibration matrix A_l. The readout error at time step l is corrected by applying the pseudo-inverse of A_l to the measured outcomes from the experiments.

For all experiments, the measurement outcomes are stored as the count of measuring each basis elements in $\mathcal{B}$. Let $\mathbf{v}_l = \begin{pmatrix} \mathbf{v}_l^1 & \cdots & \mathbf{v}_l^{2^n} & 1 \end{pmatrix}$, where $\mathbf{v}_l^i$ is the count of measuring $|i\rangle \in \mathcal{B}$ at time step l. Let $\mathbf{z}_l = \begin{pmatrix} \overline{\langle Z^{(1)}\rangle}_l & \cdots & \overline{\langle Z^{(n)}\rangle}_l & 1 \end{pmatrix}$, where $\overline{\langle Z^{(i)}\rangle}_l$ is the finite-sampled approximation of $\langle Z^{(i)}\rangle_l$ for $i = 1, \ldots, n$. Then we have $\mathbf{z}_l = \mathbf{v}_l B$, where B is a linear transformation. After applying readout error correction, we have $\mathbf{z}'_l = \mathbf{v}_l A_l^+ B$, where A_l^+ is the pseudo-inverse of A_l. To optimize the readout function parameters $\mathbf{w}$, collect all measurement data in a matrix $\mathbf{v} = \begin{pmatrix} \mathbf{v}_1^T & \cdots & \mathbf{v}_L^T \end{pmatrix}^T$ so that $\mathbf{z} = \begin{pmatrix} \mathbf{z}_1^T & \cdots & \mathbf{z}_L^T \end{pmatrix}^T = \mathbf{v}B$, where L is the sequence length. The linear output of the quantum reservoir computer is $\overline{\mathbf{y}} = \mathbf{v}B\mathbf{w}$, where $\mathbf{w}$ includes the bias term $\mathbf{w}_c$. Append a corresponding row and column to A_l to account for the bias term. Suppose the readout error is time-invariant, then $A^+ = A_l^+$ for $l = 1, \ldots, L$. The quantum reservoir computer output after readout error correction is $\overline{\mathbf{y}'} = \mathbf{v}A^+B\mathbf{w}'$. Assume that A^+ has all rows linearly independent, then ordinary least squares yields $B\mathbf{w}' = AB\mathbf{w}$. Now given test data with readout error correction,

$$\mathbf{v}_{\text{test}} A^+ B \mathbf{w}' = \mathbf{v}_{\text{test}} A^+ A B \mathbf{w} = \mathbf{v}_{\text{test}} B \mathbf{w}.$$

Therefore, the QR predicted output and thus the universal family $\mathcal{M}$ are *invariant* under time-invariant readout error.

Appendix D: Quantum circuit interpretation of the quantum reservoir dynamics

We provide an explanation for the quantum circuit interpretation of the QR dynamics given by Eq. (1) and Fig. 1(b) in the main article, with $T_0(\rho_{l-1}) = U_0 \rho_{l-1} U_0^\dagger$ and $T_1(\rho_{l-1}) = U_1 \rho_{l-1} U_1^\dagger$. The state $\rho(u_l)$ encodes the input u_l as a classical mixture $\rho(u_l) = u_l|0\rangle\langle 0| + (1 - u_l)|1\rangle\langle 1|$, meaning that with probability u_l, we apply $U_0 \rho_{l-1} U_0^\dagger$ and with probability $1 - u_l$, we apply $U_0^\dagger U_0 U_1 \rho_{l-1} U_1^\dagger U_0^\dagger U_0 = U_1 \rho_{l-1} U_1^\dagger$. Let ρ'_{l-1} denotes the quantum reservoir state after these operations. The state ρ_ϵ is a classical mixture $\rho_\epsilon = (1 - \epsilon)|0\rangle\langle 0| + \epsilon|1\rangle\langle 1|$ that encodes the rate at which the QR forgets its initial conditions. That is, with probability ϵ, the states ρ'_{l-1} and σ are exchanged, equivalent to resetting the reservoir state ρ'_{l-1} to the fixed state σ; otherwise the state ρ'_{l-1} is unchanged with probability $1 - \epsilon$.

Appendix E: Experimental and numerical details

1. Nonlinear temporal processing tasks

We give detailed descriptions of the four nonlinear temporal processing tasks and the circuits implementing the four QRs. Task I is a linear reservoir with polynomial readout [11, 12], described by

$$\begin{cases} \mathbf{x}_l = A\mathbf{x}_{l-1} + cu_l \\ y_l = h(\mathbf{x}_l), \end{cases}$$

where $A \in \mathbb{R}^{2000 \times 2000}$ and $c \in \mathbb{R}$. The readout function h is a degree 2 polynomial in the state elements. Task II is a recently proposed classical reservoir computing model that achieves good performance in chaotic system modeling [12], described by

$$\begin{cases} \mathbf{x}_l = p(u_l)\mathbf{x}_{l-1} + q(u_l) \\ y_l = \mathbf{w}^T \mathbf{x}_l, \end{cases}$$

where $p(u_l) = \sum_{j=0}^{4} A_j u_l^j$ and $q(u_l) = \sum_{j=0}^{2} B_j u_l^j$ are matrix-valued polynomials in the input u_l, $A_j \in \mathbb{R}^{700 \times 700} \bigoplus \mathbb{R}^{700 \times 700}$ and $B_j \in \mathbb{R}^{700 \times 1} \bigoplus \mathbb{R}^{700 \times 1}$. For both Task I and Task II, We rescale the maximum singular value $\sigma_{\max}(A) = 0.5$ and $\sigma_{\max}(A_j) < \frac{1}{3}$ for all j so that both tasks have short-term memory. Task III is a Volterra series with kernel order 5 and memory 2, commonly applied to model responses of nonlinear systems in control engineering [11],

$$y_l = \mathbf{w}_c + \sum_{i=1}^{5} \sum_{j_1,\ldots,j_i=0}^{2} \mathbf{w}_i^{j_1,\ldots,j_i} \prod_{k=1}^{i} u_{l-j_k}.$$

For the first two tasks, elements of A, A_j, B and $\mathbf{w}$ are uniformly randomly sampled from $[-1, 1]$. The constant c and coefficients of readout function h are also sampled from the same distribution. The same applies to the kernel coefficients $\mathbf{w}_i^{j_1,\ldots,j_i}$ and $\mathbf{w}_c$ in Task III.

Task IV is a missile moving with a constant velocity in the horizontal plane, a continuous-time long-term memory nonlinear map [39] described by

$$\begin{cases} \dot{\mathbf{x}}_1 = \mathbf{x}_2 - 0.1\cos(\mathbf{x}_1)(5\mathbf{x}_1 - 4\mathbf{x}_1^3 + \mathbf{x}_1^5) - 0.5\cos(\mathbf{x}_1)u \\ \dot{\mathbf{x}}_2 = -65\mathbf{x}_1 + 50\mathbf{x}_1^3 - 15\mathbf{x}_1^5 - \mathbf{x}_2 - 100u, \end{cases}$$

with $y = \mathbf{x}_2$. This missile dynamics is simulated using the $(4, 5)$ Runge-Kutta formula in MATLAB, with a sampling time of $\tau = 1/80$ for 1 second.

2. Full input-output sequential data

Since we bypass the washout for QRs by initializing them in the state $|0\rangle^n$, this is equivalent to washing out their initial conditions with a length L_w constant input sequence $u_l = 1$. The same washout has been applied to all nonlinear tasks. We have checked that $L_w = 50$ is enough for all tasks to reach steady states given the same initialization $\mathbf{x}_0 = 0$. Particular caution has been taken to washout Task IV, in which we set $u_l = 1$ for $l = -2, -1$. For each target map, we discard the first four input-output sequence data points, and the corresponding QR experimental data, to remove the transitory output response due to the change in input statistics. In Fig. 4, we show the full washout, train and test input-output target sequences for both the multi-step ahead prediction and the map emulation problems. Fig. 5 plots the full target output sequences, the train and test QR outputs on the multi-step ahead prediction problem. Fig. 6 plots the full target output sequences, the train and test QR outputs on the map emulation problem. In all figures, the transitory responses are indicated by dotted lines.

3. Quantum circuits for the QRs

We specify the quantum circuits for each QR implemented in the experiments. For the 4-qubit and 10-qubit QRs on the Boeblingen device, we choose

$$U_0(\boldsymbol{\theta}) = \prod_{j=1}^{N_0} \left(U_3^{(j_t)}(\boldsymbol{\theta}_{j_t}) \mathrm{CX}_{j_c j_t} U_3^{(j_t)}(\boldsymbol{\theta}_{j_t})^\dagger \right), \qquad U_1(\boldsymbol{\phi}) = \bigotimes_{i=1}^{n} U_3^{(i)}(\boldsymbol{\phi}_{0_i}) \prod_{j=1}^{N_1} \left(\mathrm{CX}_{j_c j_t} \bigotimes_{i=1}^{n} U_3^{(i)}(\boldsymbol{\phi}_{j_i}) \right).$$

The number of layers in U_0 and U_1 is $N_0 = N_1 = 5$ for both the 4-qubit and 10-qubit Boeblingen QRs. Here $U_3^{(j_t)}(\boldsymbol{\theta}_{j_t})$ is an arbitrary single-qubit rotation defined in Ref. [29] with three parameters $\boldsymbol{\theta}_{j_t} = (\boldsymbol{\theta}_{j_t}^0, \boldsymbol{\theta}_{j_t}^1, \boldsymbol{\theta}_{j_t}^2)$, and $U_3^{(j_t)}(\boldsymbol{\theta}_{j_t})^\dagger$ is its inverse. Each parameter is uniformly randomly generated from $[-2\pi, 2\pi]$. The same applies to the parameters $\boldsymbol{\phi}_{j_i} = (\boldsymbol{\phi}_{j_i}^0, \boldsymbol{\phi}_{j_i}^1, \boldsymbol{\phi}_{j_i}^2)$ for $j = 0, \ldots, N_1$ in $U_1(\boldsymbol{\phi})$. The inverse operation $U_3^{(j_t)}(\boldsymbol{\theta}_{j_t})^\dagger$ is implemented by permuting and negating its parameters so that $U_3^{(j_t)}(\boldsymbol{\theta}_{j_t})^\dagger = U_3^{(j_t)}(-\boldsymbol{\theta}_{j_t}^0, -\boldsymbol{\theta}_{j_t}^2, -\boldsymbol{\theta}_{j_t}^1)$. The control and target qubits are chosen such that they are natively coupled, meaning that a CNOT gate can be directly applied to the qubit pair. See Fig. 7(a) for the Boeblingen qubit coupling map. Qubits $Q = 0, 1, 2, 3$ in the Boeblingen device are used for the 4-qubit Boeblingen QR and $Q = 0, 1, 2, 3, 5, 6, 7, 8, 10, 12$ for the 10-qubit Boeblingen QR. These qubits are chosen due to their shorter gate durations, lower gate errors and longer coherence times, see Sec. E 5 for hardware specifications and Fig. 8 for the quantum circuits.
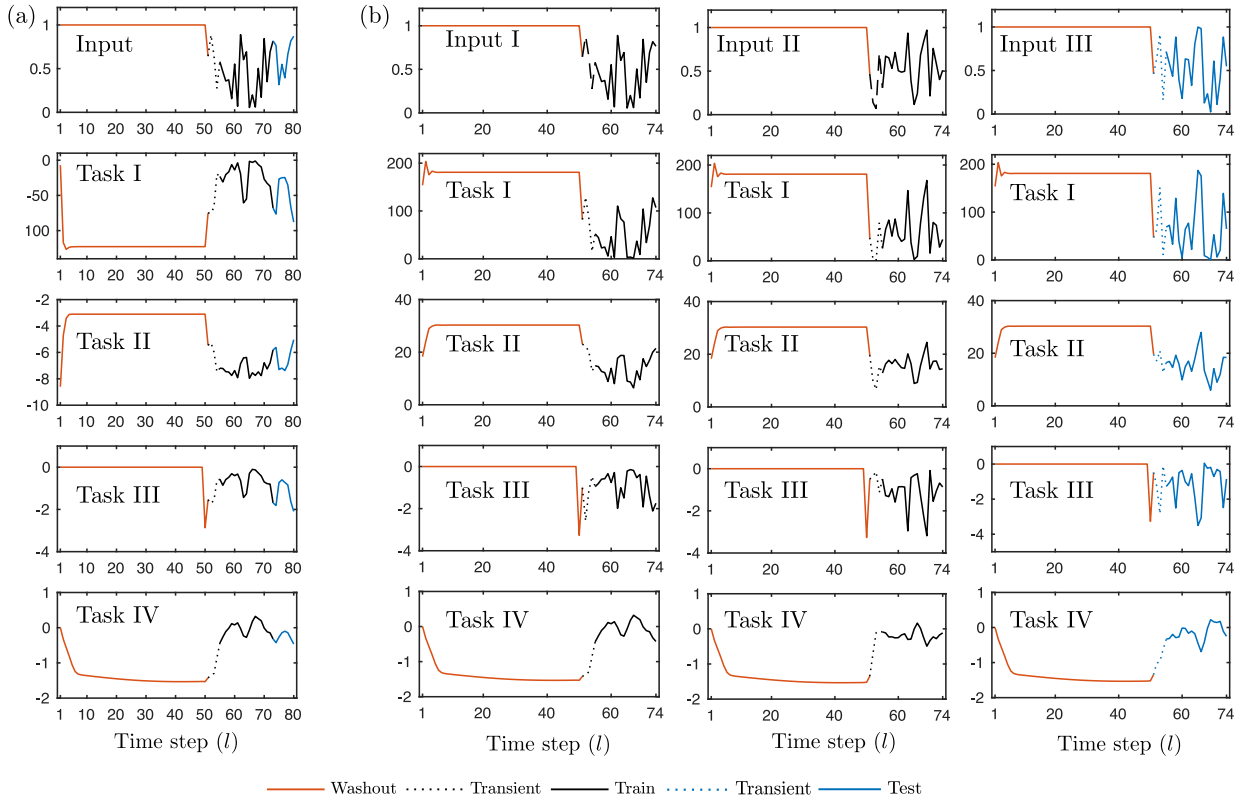
FIG. 4. Full washout, train and test input-output sequences for (a) The multi-step ahead prediction problem and (b) The map emulation problem.

For the 5-qubit QR on the Ourense device, with the qubit coupling map shown in Fig. 7(b), we choose

$$U_0 = \prod_{j=1}^{4} \mathrm{CX}_{j_c j_t}, \quad U_1(\boldsymbol{\phi}) = \bigotimes_{i=1}^{5} U_3^{(i)}(\boldsymbol{\phi}_i).$$

For the 5-qubit QR on the Vigo device,

$$U_0(\boldsymbol{\theta}) = \prod_{j=1}^{3} \left(R_y^{(j_t)}(\boldsymbol{\theta}_{j_t}) \mathrm{CX}_{j_c j_t} R_y^{(j_t)}(\boldsymbol{\theta}_{j_t})^{\dagger} \right), \qquad U_1(\boldsymbol{\phi}) = \bigotimes_{i=1}^{5} R_x^{(i)}(\boldsymbol{\phi}_i).$$

Here $R_y^{(i)}$ and $R_x^{(i)}$ are rotational Y and X gates on qubit i, respectively. Both gates have one gate parameter and $R_y^{(j_t)}(\boldsymbol{\theta}_{j_t})^{\dagger} = R_y^{(j_t)}(-\boldsymbol{\theta}_{j_t})$. See Fig. 9 for the quantum circuits.

4. Measurement and simulation data

We simulate the four QRs using the IBM Qiskit simulator under ideal and noisy conditions. The noise models used are obtained from the device calibration data. We fetched the updated device calibration data each time a job was executed on the hardware. The circuits simulated are the same as the circuits employed for the experiments and so is the number of shots. For the multi-step ahead prediction problem, the 10-qubit Boeblingen QR experienced a significant deviation from simulated results on qubits $Q = 1, 8$ (see Fig. 10), resulting in larger NMSE $= 0.26, 0.068, 0.15, 6.1$ for the four tasks. After setting the readout parameters $\mathbf{w}_1 = \mathbf{w}_8 = 0$ for $Q = 1, 8$, this issue was circumvented at the cost of using a fewer number of computational features. The resulting 10-qubit Boeblingen QR still achieves performance improvement over other QRs with a smaller number of qubits on the multi-step ahead prediction problem in the first three tasks. A time-invariant readout error in qubit i linearly transforms the expectation $\langle Z^{(i)} \rangle_l$. The QR predicted outputs are invariant under time-invariant readout errors when using linear regression to optimize $\mathbf{w}, \mathbf{w}_c$ as derived
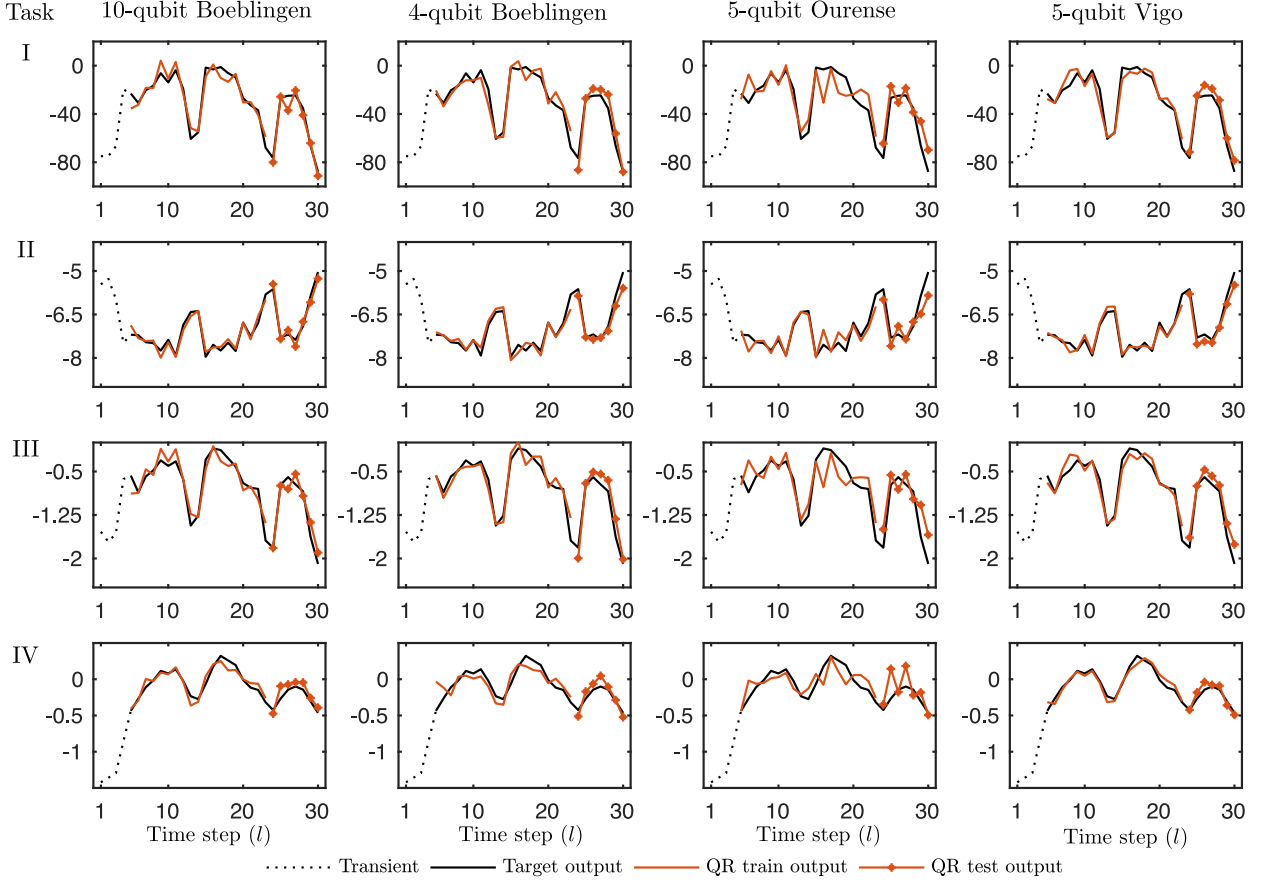
FIG. 5. The full target output sequences, the train and test output sequences of the four QRs for each task on the multi-step ahead prediction problem. Each column corresponds to each n-qubit QR outputs. Each row corresponds to each task.

in Sec. C. However, for the 10-qubit Boeblingen QR, the deviations in qubits $Q = 1, 8$ were time-varying as shown in Fig. 10. On the other hand, the 5-qubit Vigo device experienced almost time-invariant deviations in qubit $Q = 0$ as shown in Fig. 11, but this does not affect the performance of this QR noticeably. The experimental results of the 5-qubit Ourense QR follow the noisy simulation results closely. For the map emulation problem, the experimental results of both 5-qubit QRs follow the simulated results closely, with an almost time-invariant shift in $Q = 0$ for the 5-qubit Vigo QR.

5. Hardware specifications

The experiments were conducted on the IBM 20-qubit Boeblingen (version 1.0.0), 5-qubit Ourense (version 1.0.0) and 5-qubit Vigo (version 1.0.0) superconducting quantum processors [28]. The gate duration for an arbitrary single-qubit rotation gate U_3 [29] is $\tau_{U_3} \approx 71.1$ ns for all qubits whereas the CNOT gate durations differ for different qubits.

See Fig. 4 for the 4-qubit and 10-qubit Boeblingen QR quantum circuits. The circuits are chosen such that both QRs have the same number of layers in U_0 and U_1. In this setting, the maximum duration of a circuit executed on the Boeblingen device is the same for both QRs. As stated in the main article, the chosen qubits for the 4-qubit QR and the 10-qubit QR on the Boeblingen device are $Q = 0, 1, 2, 3$ and $Q = 0, 1, 2, 3, 5, 6, 7, 8, 10, 12$. These qubits were chosen due to their longer coherence times, shorter CNOT gate durations, smaller gate and readout errors. During the experiment, the maximum readout error was 10^{-2} and the maximum U_3 gate error implemented was 10^{-3}. The maximum CNOT gate error implemented was 4.3×10^{-2} and the maximum CNOT gate duration was $\tau_{\text{CNOT}} \approx 427$ ns. We assume that commuting gates can be executed in parallel. We choose $N_0 = N_1 = 5$ numbers of layers for U_0 and U_1 in the 4-qubit and 10-qubit Boeblingen QRs. The maximum length of any input sequence (including the transient) for the multi-step ahead prediction and the map emulation problems is $L = 30$. Therefore, the maximum
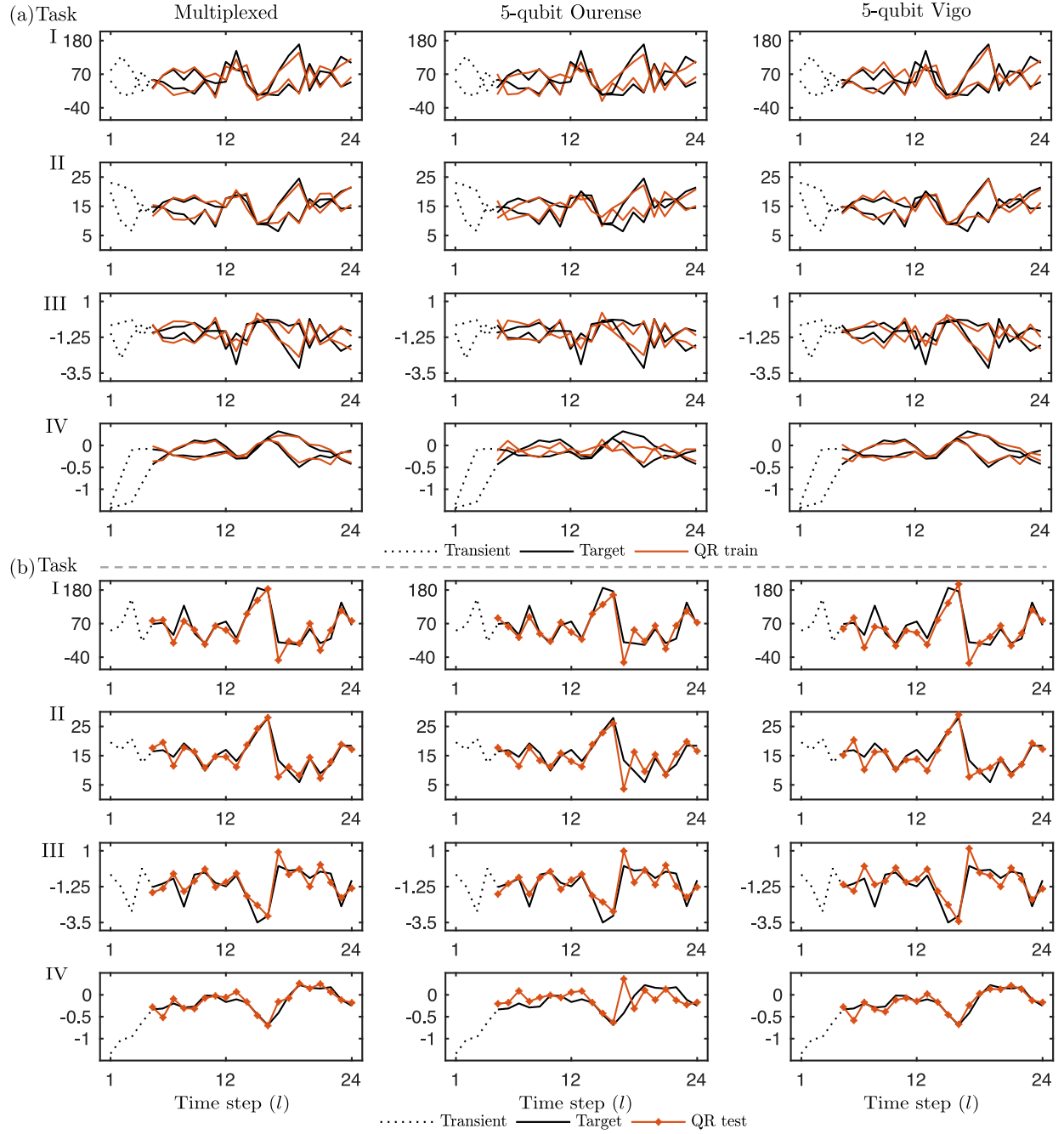
FIG. 6. The full target output sequences, the train and test output sequences of the QRs for each task on the map emulation problem. (a) Shows the two train output sequences. (b) Shows the test output sequence. The columns correspond to the multiplexed 5-qubit QRs, 5-qubit Ourense QR and the 5-qubit Vigo QR from the left to the right. Each row corresponds to each task.

numbers of U_3 gate executions and CNOT gate executions is $5L = 5 \times 30 = 150$. The maximum duration of a circuit executed on the Boeblingen device was $150 \times (\tau_{U_3} + \tau_{\text{CNOT}}) \approx 150 \times (71.1 + 427) = 74.7$ μs, within the coherence times (T_1, T_2) for most qubits chosen.

Fig. 5 shows the quantum circuits for the 5-qubit Ourense and 5-qubit Vigo QRs. Owing to the more restricted qubit couplings in these 5-qubit devices, the circuits for the 5-qubit QRs are simpler than that of the 4-qubit and 10-qubit Boeblingen QRs. To combine different computational features for the spatial multiplexing technique, we choose circuits that are sufficiently different for these two 5-qubit QRs. In particular, the 5-qubit Vigo QR consists of single-qubit rotational Y gates in U_0 and single-qubit rotational X gates in U_1. On the other hand, the 5-qubit
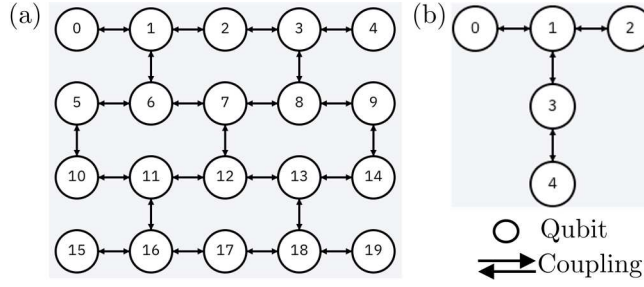
FIG. 7. Qubit coupling maps of the IBM superconducting quantum processors. (a) The 20-qubit Boeblingen device. (b) Both the 5-qubit Ourense and Vigo devices.
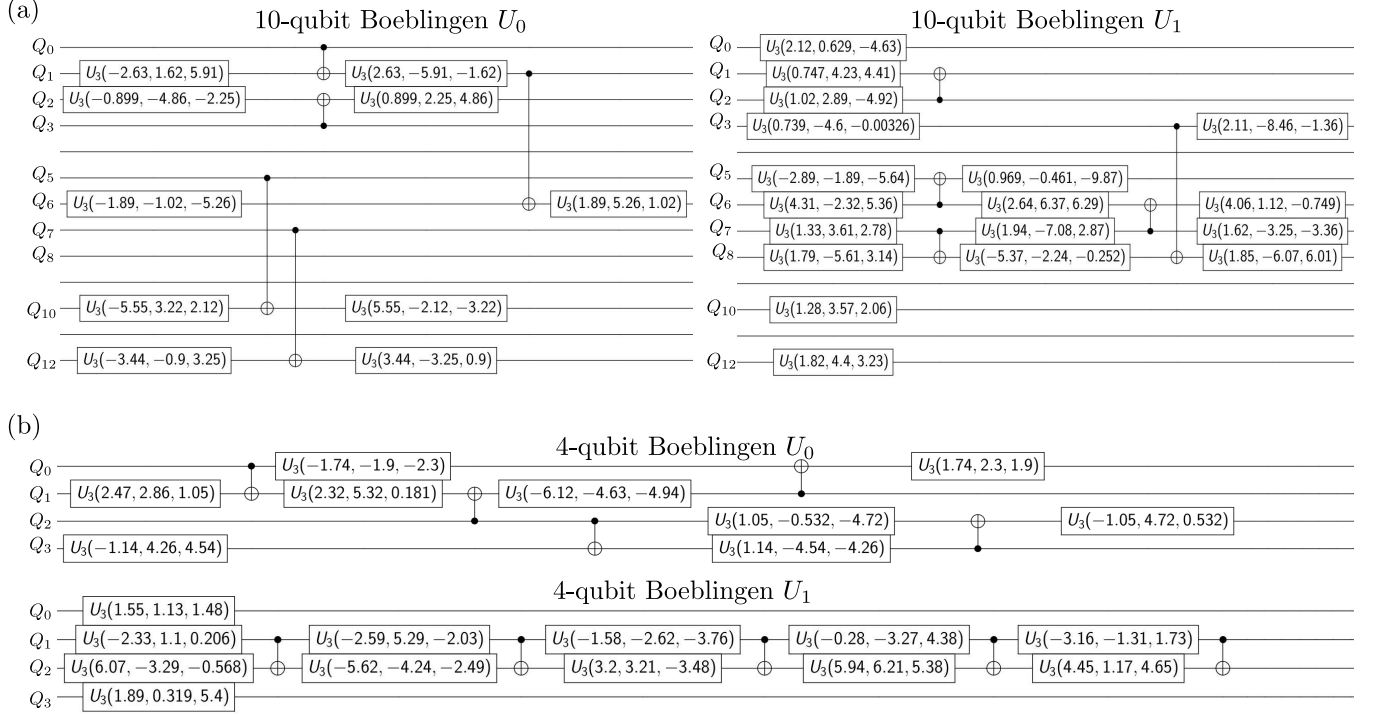

FIG. 8. Quantum circuits for the (a) 10-qubit QR and (b) 4-qubit QR on the Boeblingen device.
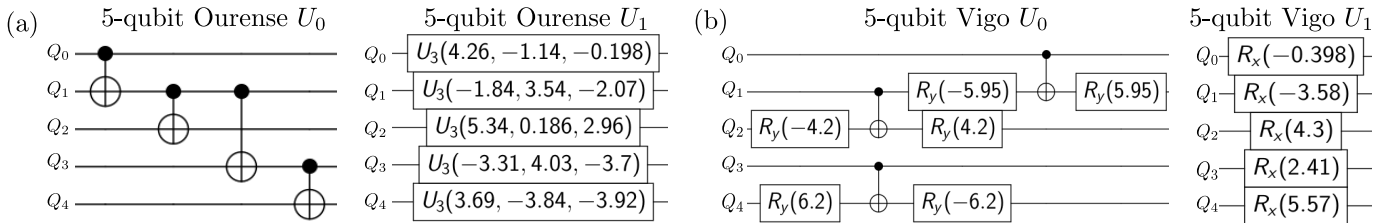

FIG. 9. Quantum circuits for the (a) 5-qubit Ourense QR and (b) 5-qubit Vigo QR.

Ourense QR uses arbitrary single-qubit rotational gates U_3 only in the circuit implementing U_1.

The 5-qubit Ourense device achieves the same order of magnitude in readout errors, coherence times and CNOT gate durations as the 20-qubit Boeblingen device, but lower CNOT gate errors. For the Ourense device, the maximum U_3 gate error and readout error implemented were 0.9×10^{-3} and 4.1×10^{-2}, and the maximum CNOT gate error implemented was 8×10^{-3}, a lower error compared to the Boeblingen device. The maximum CNOT gate duration implemented was $\tau_{\text{CNOT}} \approx 576$ ns. For the 5-qubit Ourense QR, the circuit implementing U_0 is longer than that for U_1. The U_0 circuit consists of four CNOT gates, and the maximum duration of a circuit executed for the 5-qubit
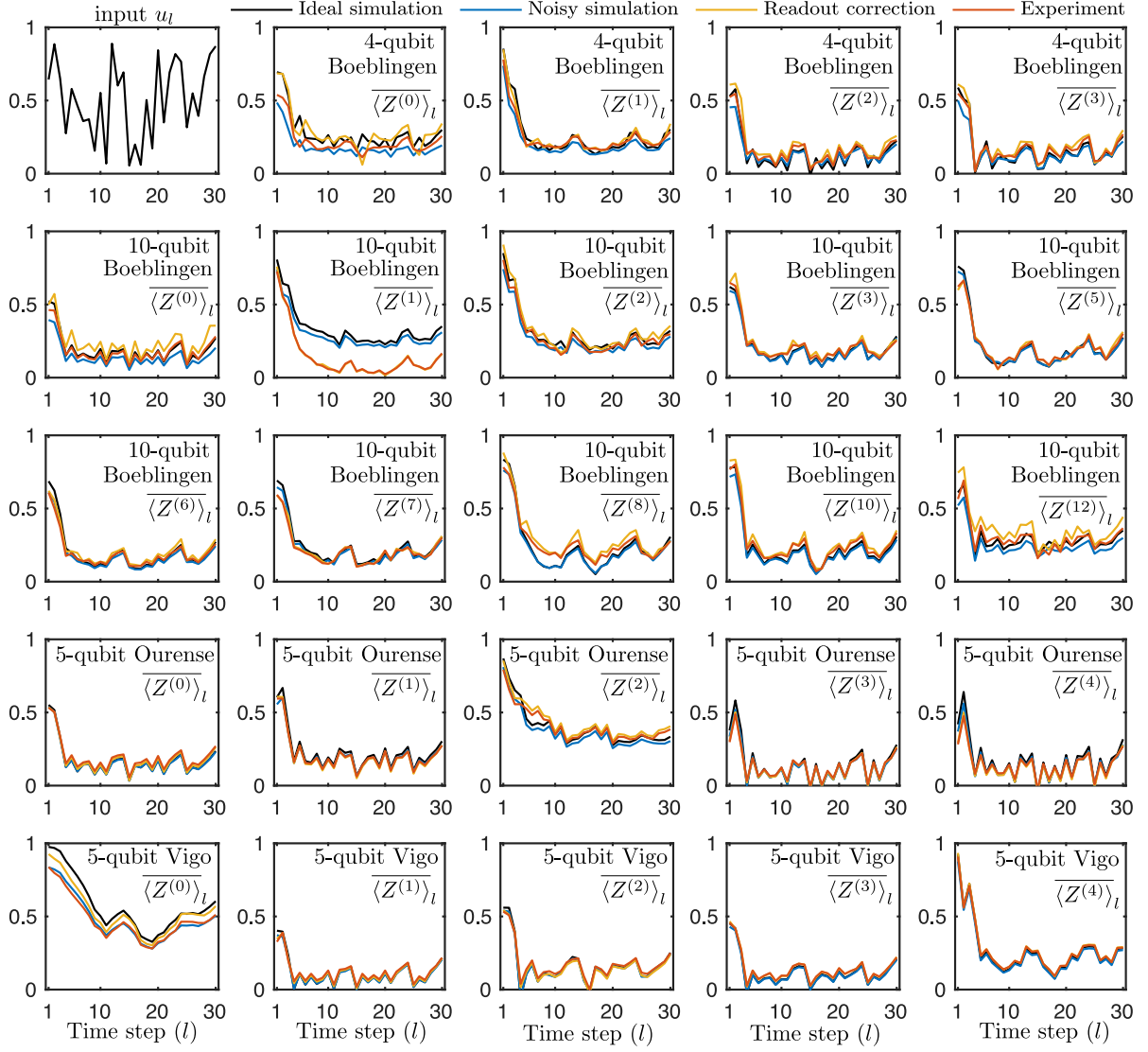
FIG. 10. Input sequence, experimental and simulation results for each qubit of the four QRs at each time step $l = 1, \ldots, 30$, for the multi-step ahead prediction problem.

Ourense QR was $4L \times \tau_{\mathrm{CNOT}} \approx 70 \, \mu\mathrm{s}$, also within the coherence limits of most qubits.

The 5-qubit Vigo device is similar to the 5-qubit Ourense device. They have the same qubit couplings and share similar noise profile and hardware specifications. Rotational X and Y gates were used on this device, with gate duration $\tau = 35.5$ ns. The maximum single-qubit gate error implemented was 0.8×10^{-3} and the maximum readout error implemented was 7.8×10^{-2}. The maximum CNOT gate error and gate duration implemented was 1.3×10^{-2} and $\tau_{\mathrm{CNOT}} \approx 462.2$ ns, respectively. For this QR, U_0 is the longer circuit consisting of three layers of single-qubit rotation Y gates and two layers of CNOT gates. Therefore, the maximum duration of a circuit implemented was $(3\tau + 2\tau_{\mathrm{CNOT}})L = (3 \times 35.5 + 2 \times 462.2) \times 30 \approx 31 \, \mu\mathrm{s}$.
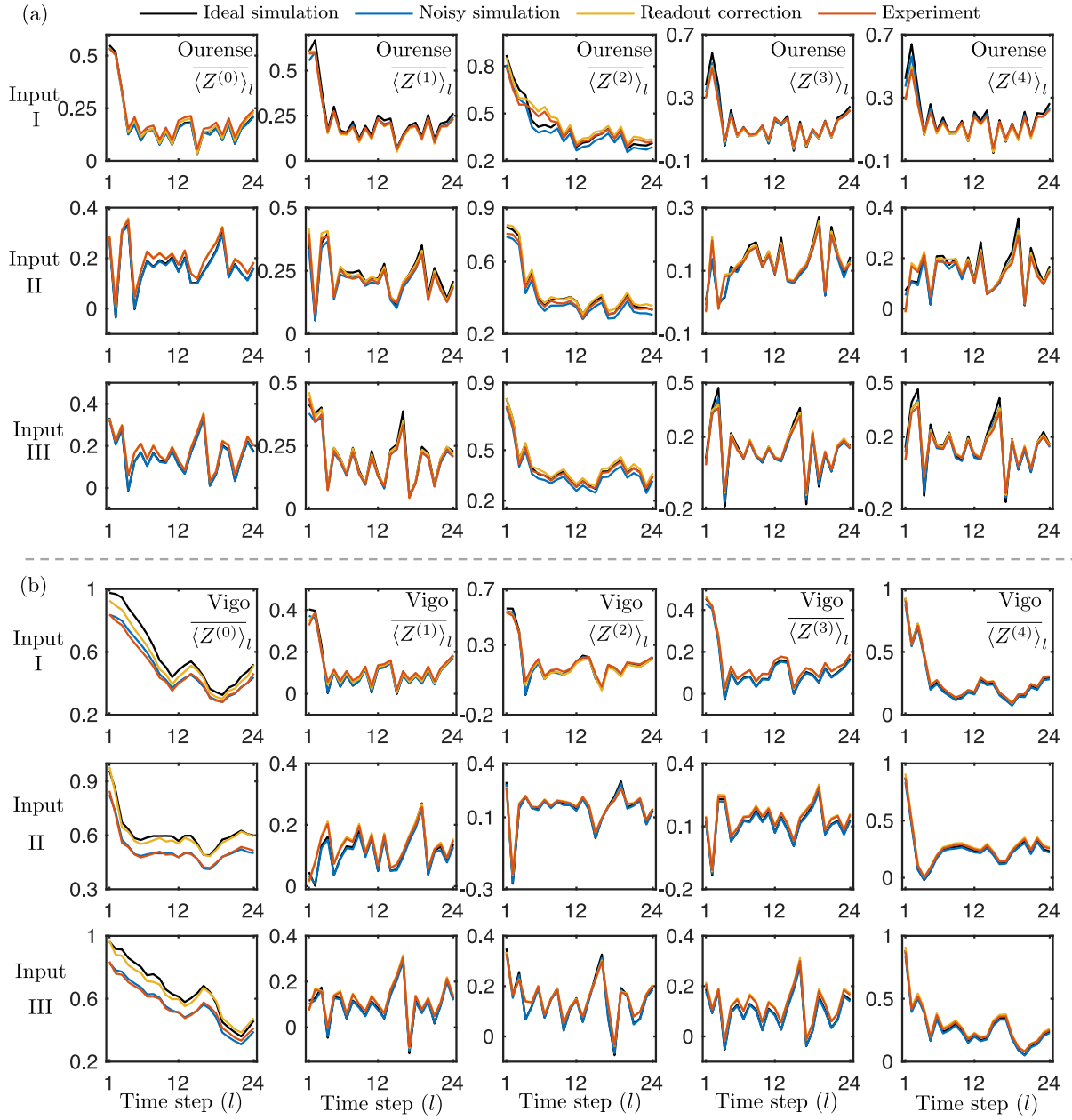
FIG. 11. Experimental and simulation results for each qubit $i = 0, \ldots, 4$ and each time step $l = 1, \ldots, 24$, for the map emulation problem. Three input sequences are used in this problem, labeled as inputs I, II and III. Row i in a sub-figure corresponds to the experimental data for the i-th input sequence. Column j corresponds to the experimental data for the j-th qubit. (a) Shows the experimental data for the 5-qubit Ourense QR. (b) Shows the experimental data for the 5-qubit Vigo QR.

[1] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature **549**, 195 (2017).

[2] C. Ciliberto, M. Herbster, A. D. Ialongo, M. Pontil, A. Rocchetto, S. Severini, and L. Wossnig, Quantum machine learning: a classical perspective, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **474**, 20170551 (2018).

[3] J. Preskill, Quantum Computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[4] 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 **567**, 209 (2019).

[5] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature **549**, 242 (2017).

[6] J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima, *et al.*, Neuromorphic computing with nanoscale spintronic oscillators, Nature **547**, 428 (2017).

[7] W. Maass, T. Natschläger, and H. Markram, Real-time computing without stable states: A new framework for neural computation based on perturbations, Neural Computation **14**, 2531 (2002).

[8] J. Miller and M. Hardt, Stable recurrent models, in *Proceedings of the 2019 International Conference on Learning Representations (ICLR)* (2019).

[9] J. Hanson and M. Raginsky, Universal approximation of input-output maps by temporal convolution nets, in *Proceedings of the 33rd annual conference on Neural Information Processing Systems (NeurIPS)* (2019).

[10] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, Neural Networks **2**, 359 (1989).

[11] S. Boyd and L. Chua, Fading memory and the problem of approximating nonlinear operators with Volterra series, IEEE Trans. Circuits Syst. **32**, 1150 (1985).

[12] L. Grigoryeva and J.-P. Ortega, Universal discrete-time reservoir computers with stochastic inputs and linear readouts using non-homogeneous state-affine systems, The Journal of Machine Learning Research **19**, 892 (2018).

[13] L. Grigoryeva and J.-P. Ortega, Echo state networks are universal, Neural Networks **108**, 495 (2018).

[14] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Networks **61**, 85 (2015).

[15] S. Hochreiter and J. Schmidhuber, Long short-term memory, Neural Computation **9**, 1735 (1997).

[16] B. Irie and S. Miyake, Capabilities of three-layered perceptrons, in *IEEE International Conference on Neural Networks*, Vol. 1 (1988) p. 218.

[17] X. Glorot and Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (2010) pp. 249–256.

[18] R. Pascanu, T. Mikolov, and Y. Bengio, On the difficulty of training recurrent neural networks, in *International conference on machine learning* (2013) pp. 1310–1318.

[19] M. Rigotti *et al.*, The importance of mixed selectivity in complex cognitive tasks, Nature **497**, 585 (2013).

[20] C. Du, F. Cai, M. A. Zidan, W. Ma, S. H. Lee, and W. D. Lu, Reservoir computing using dynamic memristors for temporal information processing, Nature Communications **8**, 2204 (2017).

[21] K. Vandoorne, P. Mechet, T. Van Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, Experimental demonstration of reservoir computing on a silicon photonics chip, Nature Communications **5**, 3541 (2014).

[22] Q. Vinckier, F. Duport, A. Smerieri, K. Vandoorne, P. Bienstman, M. Haelterman, and S. Massar, High-performance photonic reservoir computer based on a coherently driven passive cavity, Optica **2**, 438 (2015).

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

[24] K. Nakajima, K. Fujii, M. Negoro, K. Mitarai, and M. Kitagawa, Boosting computational power through spatial multiplexing in quantum reservoir computing, Physical Review Applied **11**, 034021 (2019).

[25] M. Negoro, K. Mitarai, K. Fujii, K. Nakajima, and M. Kitagawa, Machine learning with controllable quantum dynamics of a nuclear spin ensemble in a solid, arXiv preprint arXiv:1806.10910 (2018).

[26] J. Chen and H. I. Nurdin, Learning nonlinear input–output maps with dissipative quantum systems, Quantum Information Processing **18**, 198 (2019).

[27] H. Jaeger and H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communications, Science **304**, 5667 (2004).

[28] IBM Quantum Experience, https://www.ibm.com/quantum-computing/.

[29] A. W. Cross, L. S. Bishop, J. A. Smolin, and J. M. Gambetta, Open quantum assembly language, arXiv preprint arXiv:1707.03429 (2017).

[30] J. Chen, H. I. Nurdin, and N. Yamamoto, Single-input single-output nonlinear system identification and signal processing on near-term quantum computers, in *Proceedings of the 2019 IEEE Conference on Decision and Control (CDC)* (2019) pp. 401–406.

[31] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, Nature **567**, 491 (2019).

[32] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, Physical Review X **7**, 021050 (2017).

[33] A. Pavlov, N. van de Wouw, and H. Nijmeijer, Convergent systems: Analysis and synthesis, in *Control and Observer Design for Nonlinear Finite and Infinite Dimensional Systems*, Lecture Notes in Control and Information Science, Vol. 322, edited by T. Meurer, K. Graichen, and E. D. Gilles (Springer, 2005) pp. 131–146.

[34] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (New York: Cambridge University Press, 2010).

[35] J. Chen and H. I. Nurdin, Correction to: Learning nonlinear input–output maps with dissipative quantum systems, Quantum Information Processing **18**, 354 (2019).

[36] D. Perez-Garcia, M. M. Wolf, D. Petz, and M. B. Ruskai, Contractivity of positive and trace-preserving maps under l_p norms, Journal of Mathematical Physics **47**, 083506 (2006).

[37] J. Dieudonné, *Foundations of Modern Analysis* (Read Books Ltd, 2013).

[38] S. Lang, *Complex Analysis*, Graduate Texts in Mathematics (Springer-Verlag, 1985).

[39] X. Ni, M. Verhaegen, A. J. Krijgsman, and H. B. Verbruggen, A new method for identification and control of nonlinear dynamic systems, Engineering Applications of Artificial Intelligence **9**, 231 (1996).