Quantum-classical hybrid information processing via a single quantum system

Quoc Hoan Tran [1,*] Sanjib Ghosh,[2,†] and Kohei Nakajima [1,3,‡]

[1]*Next Generation Artificial Intelligence Research Center, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*

[2]*Beijing Academy of Quantum Information Sciences, Beijing 100193, China*

[3]*Department of Mechano-Informatics, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*

(Received 30 November 2022; accepted 5 October 2023; published 7 November 2023)

Current quantum technologies bring a new integration of quantum data with classical data for hybrid processing. However, the frameworks of these technologies are restricted to a single classical or quantum task, which limits their flexibility in near-term applications. We propose a quantum reservoir processor to harness quantum dynamics in computational tasks requiring both classical and quantum inputs. This analog processor comprises a quantum network in which quantum data are incident to the network and classical data are encoded via a coherent field exciting the network. We perform a multitasking application of quantum tomography and nonlinear equalization of classical channels. Interestingly, the tomography can be performed in a closed-loop manner via the feedback control of classical data. Therefore, if the classical input comes from a dynamical system, embedding this system in a closed loop enables hybrid processing even if access to the external classical input is interrupted. Finally, we demonstrate preparing quantum depolarizing channels as a quantum machine learning technique for quantum data processing.

DOI: 10.1103/PhysRevResearch.5.043127

I. INTRODUCTION

Recent advances in machine learning (ML) and quantum computing have revolutionized the methodology of processing complex and large-scale data. While merging these fields, classical or quantum systems can generate a massive amount of time series data, such as sensing data or quantum states that flow through multiple quantum channels in a network of quantum devices [1–3]. This context leads to the requirement of a novel learning paradigm to process these data efficiently, such as the easy manipulation used in training and deployment, while maintaining rich representation capability. Currently, algorithms are being designed on specific homogeneous data, such as quantum-native or classical-native data. However, most quantum devices rely on classical controls [4,5], such as temperature or signals from electronic controllers [6,7]. The outputs of these devices are not simply derived from quantum channels and are also considered a function of classical controls and quantum input. A representative example is a quantum switch with classical control, which simulates the indefinite causal order between two operations [8–12] [Fig. 1(a)]. Therefore, the research on hybrid quantum and classical data processing can lead to broader and near-term applicability for quantum devices. For example, we can use the same resource to learn the tomography of devices receiving both classical and quantum data without doing it separately for each control setting. Naturally, if the quantum or classical data are fixed, one could opt for either a fully classical or fully quantum approach when resources allow. However, when faced with limitations in computational resources, time constraints, or physical components, employing a single system for hybrid data becomes a more resource-efficient alternative compared to maintaining separate systems for classical and quantum tasks. Moreover, the key advantage lies in its capability to capture shared patterns with diverse types of inputs, proving to be a more suitable candidate than either a fully classical or a fully quantum approach for representing a function of hybrid quantum-classical data.

We propose the concept of quantum-classical hybrid information processing via the reservoir computing (RC) framework. RC is based on modeling the transformation of input sequences that employs nonlinear transformations to emulate nonlinear functions [13–16]. RC consists of three main parts: the input part to store the input, the reservoir, and the readout part. The reservoir is a dynamical system driven by input to encode recurrent relations and nonlinear dynamics of the input history via sufficiently complex and high-dimensional trajectories. The readout can extract accessible observables in the reservoir, which are useful features retained for emulating the target sequence. In general, the connections between the input and the reservoir, and between the reservoir and the output, are linear mappings. While the reservoir can be viewed as a black box, these mappings are controllable so that we can train them without interfering

*Present address: Quantum Laboratory, Fujitsu Research, Fujitsu Limited, Tokyo, Japan; k09tranhoan@gmail.com

†sanjibghosh@baqis.ac.cn

‡k-nakajima@isi.imi.i.u-tokyo.ac.jp

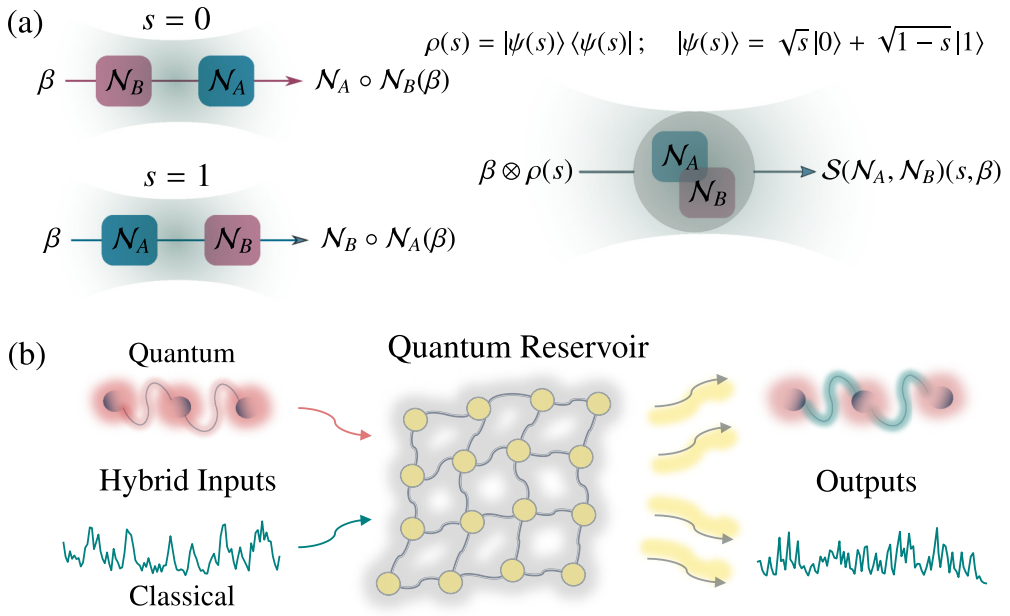

FIG. 1. A quantum reservoir (QR) processor for quantum-classical hybrid data processing. (a) An example of a quantum device with hybrid inputs. We consider a quantum switch that includes two quantum channels $\mathcal{N}_A$ and $\mathcal{N}_B$ and an independent switch state $\rho(s)$ controlled by a classical signal s. This quantum switch can be considered a function of the hybrid input (s, β) as $\mathcal{S}(\mathcal{N}_A, \mathcal{N}_B)(s, \beta)$. Given a quantum state β, $\mathcal{S}(\mathcal{N}_A, \mathcal{N}_B)$ returns the state $[\mathcal{N}_A \circ \mathcal{N}_B(\beta)] \otimes |0\rangle\langle 0|$ if $s = 0$ and $\rho(0) = |0\rangle\langle 0|$ and $[\mathcal{N}_B \circ \mathcal{N}_A(\beta)] \otimes |1\rangle\langle 1|$ if $s = 1$ and $\rho(1) = |1\rangle\langle 1|$. When $\rho(s)$ is in a superposition of $|0\rangle$ and $|1\rangle$, such as $\rho(s) = |\psi(s)\rangle\langle\psi(s)|$ with $|\psi(s)\rangle = \sqrt{s}|0\rangle + \sqrt{1-s}|1\rangle$ $(0 < s < 1)$, the output becomes a correlated state as a result of $\mathcal{N}_A$ and $\mathcal{N}_B$ acting on β in a superposition of two alternative orders. (b) Our QR is a network of quantum dots that can receive both quantum and classical data as input. Quantum inputs are incident via optical fields, and classical inputs are encoded in experimental control fields. The appropriate readout after a time evolution on QR can provide a high-dimensional transformation for both classical and quantum inputs, which can be used in learning tasks.

with the reservoir's internal parameters. This property makes RC particularly suitable for physical implementations where a physical system with rich dynamics can be a good candidate for a learning system, which defines the framework of physical reservoir computing (PRC) [17,18].

The success and efficiency of PRC rely on good physical realizations of the reservoir, which has attracted considerable interest from diverse research fields. The seminal work Ref. [19] uses a disordered ensemble quantum dynamics system as a quantum reservoir (QR) to process classical data, with the possibility of having a large number of degrees of freedom. QRs have been developed in various platforms, such as nuclear magnetic resonance systems [19–21], superconducting quantum processors [22,23], fermions and bosonic models [24–27], quantum oscillators [28–30], arrays of Rydberg atoms [31], and photonic quantum memristors [32]. Several studies have focused on the processing of data in the form of quantum states [24–26,33–35], which provide certain advantages over classical ML methods. However, a QR is yet to be treated as a homogeneous data-driven model because it lacks the ability to deal with hybrid forms of quantum-classical data. Therefore, a unified architecture for hybrid quantum-classical processing is required from theoretical and applied perspectives.

In this paper, we establish a framework that considers a QR as an analog processor to process hybrid quantum-classical data. Inspired by Refs. [24,25], our QR is a network of quantum dots with random intersite couplings. While classical and quantum inputs can be combined into a single quantum state to be processed by a QR, this encoding method is generally costly and can limit the flexibility of the QR due to the constraints of the combination scheme. Our method aims to provide a more flexible hybrid QR, where the pipelines for quantum input and classical input can be separated. Specifically, the classical input is encoded into the uniform excitation coherent field strength, which can be generated from a single laser source to excite each site in the QR [36]. The input quantum state is represented by the optical modes incident on the QR, which couple for a short time with all sites. Therefore, the incident state merges with the QR and influences its dynamics. For temporal processing, each quantum input interacts with the QR for a short duration before being replaced by another input. The time evolution of the interactions provides a high-dimensional nonlinear mapping of the input via the correlations in the QR, which can be extracted by classical or quantum readouts on accessible nodes. This enables us to learn the function of the input sequence, leading to diverse applications in classical and quantum data processing.

II. RESULTS

In this section, we describe our proposed QR for quantum-classical hybrid information processing (Sec. II A). We then present three main applications for reconstructing the function of hybrid quantum-classical inputs. First, we demonstrate that our framework can be used as a single computational resource for multitasking by performing both the tomography of a quantum channel and the equalization of a classical

nonlinear channel (Sec. II A). While both equalization and tomography tasks have been studied in classical [37] and quantum RC approaches [25,34], this paper examines the combination of classical and quantum input sequences, leading to a solution via our framework. Second, we show that quantum tomography can be performed in a closed loop of classical control, enabling us to embed classical dynamics via quantum dynamics and utilize it in hybrid processing (Sec. II C). While closed-loop control is a common task in classical RC approaches with classical data, our paper defines it with the combination of the quantum task, paving the way for using classical dynamics to generate interesting patterns of quantum data. Finally, we demonstrate that the QR can be used to prepare quantum channels with classical control (Sec. II D). This final application is inspired by the concept of using quantum reservoir computing (QRC) to realize and compress quantum circuits in Ref. [38]. However, the key factor in our paper is the consideration of hybrid quantum-classical input information, which was not previously explored.

A. Quantum-classical hybrid information processing

When we describe a quantum device processing quantum data in a realistic scenario, we must incorporate classical control into the model. In this case, a quantum device is in fact a function of quantum input β and classical control u as $\mathcal{F}(u, \beta)$, where we consider the scalar u for ease of explanation. The QR can be used for both temporal and nontemporal processing tasks of hybrid inputs. In temporal tasks, we are given an input sequence $(u_1, \beta_1), (u_2, \beta_2), \ldots, (u_L, \beta_L)$ to predict the target sequence $\hat{\boldsymbol{y}}_1, \ldots, \hat{\boldsymbol{y}}_L$. Here, the target $\hat{\boldsymbol{y}}_l$ is described via the temporal map $\mathcal{F}(\{(u_l, \beta_l)\})$, which is a function of input history [34] and reflects the temporal dependency between the output and input sequences. To solve this task, the QR system needs an input-driven map that retains the information of the input history. To reconstruct the target $\hat{\boldsymbol{y}}_l$, the system includes a readout map h, where the output signal $\boldsymbol{y}_l$ is obtained from

$$\boldsymbol{y}_l = h(u_l, \beta_l, \boldsymbol{\mu}_{l-1}), \tag{1}$$

such that $\boldsymbol{y}_l \approx \hat{\boldsymbol{y}}_l$ $(l = 1, \ldots, L)$. Here, $\boldsymbol{\mu}_{l-1}$ denotes the QR's internal quantum state before injecting the input (u_l, β_l). We note that $\boldsymbol{\mu}_{l-1}$ also depends on the input (u_{l-1}, β_{l-1}), forming the recurrent relation from Eq. (1) and reflecting the dependency between $\boldsymbol{y}_l$ and the input history. In nontemporal tasks, we only consider a specific input (u_l, β_l), and the target is not a sequence but a specific value corresponding to the given input data.

The proposed framework contains three main parts: an input part containing input modes to receive the data, a QR processor to interact with inputs in a quantum evolution, and a readout for further processing [Fig. 1(b)]. We consider the QR processor as a two-dimensional lattice of N quantum dots, represented by the Hamiltonian

$$\hat{H} = \sum_i E_i \hat{c}_i^\dagger \hat{c}_i + \sum_{\langle i,j \rangle} h_{ij} (\hat{c}_i^\dagger \hat{c}_j + \hat{c}_j^\dagger \hat{c}_i) + \sum_i Q_i \hat{c}_i^\dagger \hat{c}_i^\dagger \hat{c}_i \hat{c}_i + P(t) \sum_i (\hat{c}_i^\dagger + \hat{c}_i), \tag{2}$$

where $\hat{c}_i$, E_i, h_{ij}, Q_i, and $P(t)$ are the field operators, onsite energies, hopping amplitudes between the nearest neighbor sites, nonlinearity strengths, and uniform time-dependent coherent field strengths, respectively. $P(t)$ can be used to encode the classical input $u(t)$ as $P(t) = P + Wu(t)$, where P and W are the constant coefficient and input scaling, respectively.

The dynamics of the combined quantum state ρ of the QR as well as the input modes can be described by the quantum master equation (we use the unit where Plank constant $\hbar = 1$):

$$\dot{\rho} = -i[\hat{H}, \rho] + \gamma \sum_j \mathcal{L}(\hat{c}_j)\rho + \Omega(t - t_{\text{init}})\hat{A}\rho, \tag{3}$$

where $\Omega(t) = 1$ for $t \geqslant 0$ and 0 otherwise. Here, $\hat{A}\rho = \sum_k \frac{\gamma_k}{\gamma} \mathcal{L}(\hat{a}_k)\rho + \sum_{k,j} W_{jk}^{\text{in}}([\hat{a}_k \rho, \hat{c}_j^\dagger] + [\hat{c}_j, \rho \hat{a}_k^\dagger])$ represents the cascade coupling between the input modes $\hat{a}_k$ and the QR [39]. The Lindblad superoperator $\mathcal{L}(\hat{x})$ is defined for any arbitrary operator $\hat{x}$ by $\mathcal{L}(\hat{x})\rho = \hat{x}\rho\hat{x}^\dagger - \frac{1}{2}(\hat{x}^\dagger\hat{x}\rho + \rho\hat{x}^\dagger\hat{x})$. In our numerical simulation, we evaluate ρ by numerically solving Eq. (3). The density matrix ρ at each time t is obtained by numerically integrating Eq. (3) starting from a vacuum state. To represent the particle creation operator $\hat{c}_j$, we use a finite cutoff in the particle number, which yields a finite-dimensional Hilbert space. In the Fock space, the density matrix can be represented in matrix form, and Eq. (3) becomes a first-order differential equation. We solve this equation using the JULIA package "DifferentialEquations" [40].

We explain quantum-classical hybrid processing using the proposed platform. First, the QR is excited only with the uniform P for $0 \leqslant t < t_{\text{init}}$ and no incident quantum inputs. We choose t_{init} such that the QR at time t_{init} reaches a steady state. This setting ensures the echo state property [13] for the reproducible computation, where the response to the same input sequence is independent of the QR's initial state. Then, the quantum input β (described by the input modes $\hat{a}_k$) is incident to the reservoir, and the classical input $u(t) = u$ is activated at the same time. At time $t_1 = t_{\text{init}} + \tau$ for time interval τ, an appropriate and practical readout from the QR is performed for nontrivial transformations of input data. We consider two readout schemes: a linear combination of measurement results on the accessible observables (classical readout) and the other with a linear combination of quantum modes (quantum readout). The former is associated with a measurement process, while the latter has been considered in a quantum neuromorphic platform for quantum state preparation [26].

For a nontemporal processing task, we repeat the above procedure for every hybrid data instance (u, β). For a temporal processing task, at $t_l = t_{\text{init}} + (l-1)\tau$ $(l = 1, 2, \ldots)$, the classical input is switched to $u(t) = u_l$, and the quantum state β_l replaces the partial state in the input modes. Since the input information is transferred into the QR during the interaction, this scheme enables the memory ability, which is required in temporal processing tasks.

In the classical readout, measuring the expectation values of the occupation numbers $n_j = \langle \hat{c}_j^\dagger \hat{c}_j \rangle$ can extract the information from the QR to reconstruct $\mathcal{F}$. A representative application is quantum tomography, which reconstructs the density matrix output of $\mathcal{F}$ via the linear regression model: $W^{\text{out}}\boldsymbol{n} + \boldsymbol{b} \approx \boldsymbol{Y}_{\mathcal{F}}$ [25,34]. Here, $\boldsymbol{n} = (n_1, \ldots, n_K)^\top$ is the

K-dimensional reservoir state for readout, $\boldsymbol{Y}_{\mathcal{F}}$ is the real vector form to stack the real and imaginary elements of $\mathcal{F}$, and W^{out} and $\boldsymbol{b}$ are the weight and bias parameters to be optimized via the training (see Appendix A). In the classical readout, multitasking is possible since the training cost is minimal for independent training with different W^{out} for different tasks. If the measurement is performed after an interaction time τ for the current input and right before the next input, the dimensionality K is equal to the number of quantum dots N. One can increase this dimensionality by performing measurements at different timings in the interval τ, which is known as the temporal multiplexing technique. Between two inputs, we perform measurements at equal interval τ/V, forming the dimensionality $K = NV$. Here, V is called the measurement multiplexity. Another technique to increase the dimensionality K is spatial multiplexing [20], where readout reservoir states in different QRs are combined to learn the target.

In the quantum readout, the standard toolbox of linear optical elements [41] enables us to generate M quantum output modes $\hat{C}_m = \sum_j o_{mj}\hat{c}_j$ via the unitary transformation with complex coefficients o_{mj}. The output modes must satisfy the commutation relations $[\hat{C}_m, \hat{C}_n^\dagger] = \delta_{mn}$, which impose the condition $\sum_j o_{mj}o_{nj}^* = \delta_{mn}$. Since the target is the quantum state, the training process is not as simple as the one used for linear regression on the accessible observables in the classical readout. Consider the separation of nonadjustable and adjustable parameters in PRC, we assume that the parameters of Hamiltonian in Eq. (2) are random and not trainable. Instead, we train interaction (W_{jk}^{in}) and readout ($\{o_{mj}\}$) coefficients such that the quantum state described via $\{\hat{C}_m\}$ becomes the same as the output of $\mathcal{F}$ (see Appendix I).

In the following numerical simulations, we consider $E_i/\gamma = 0$, $\gamma_k/\gamma = \sum_j (W_{jk}^{\text{in}})^2$, with W_{jk}^{in} being the input weights randomly chosen from the interval $[0.0, \gamma]$. The hopping amplitudes h_{ij} are randomly taken from a uniform distribution in the interval $[-\gamma, \gamma]$ and are then normalized such that the spectral radius, which is the largest modulus of the eigenvalues of the matrix (h_{ij}), is equal to 1.0. If we consider a large τ, the occupation numbers come back to the steady state before the next input. If we perform the readout measurements at this timing, we cannot obtain sufficient information from the input. Therefore, we choose τ such that $n_j(t)$ are largely away from the steady values at the timing before the next input is incident on the system. In our experiments, we consider $t_{\text{init}} = 5.0/\gamma$ and $\tau = 1.0/\gamma$ (see Appendix B). For the tomography tasks via the training of the classical readout (Secs. II B and II C), we assume that there is no interaction between reservoir sites, i.e., $Q_i = 0$. For instance, in the case where $\hat{c}_j$ are fermionic fields, such a nonlinear effect can occur. However, even without this effect, the fermionic nature of the QR causes the occupation numbers n_j to be nonlinear functions of the input state. It is important to note that the physical measurement process that produces the nonlinear effect is already included in the classical readout transformation in the QR. Hence, we can obtain a nonlinear function of the input by processing the readout values n_j in a linear manner. This approach still achieves the main goal of using a reservoir. We allow for the interactions ($Q_i/\gamma = 1.0$) in training the quantum readout, to enable the nonlinearity in the QR without relying solely on the measurement process. Here, the QR can be considered as a nonlinear bosonic lattice with an interacting scheme. (Sec. II D).

B. Quantum tomography and the channel equalizer

We present an application of QR to hybrid tasks in which quantum tomography and noise-free reconstruction of classical data are performed simultaneously. Consider a temporal map $\mathcal{F}\{(s_l, \beta_l)\}$ where $\{s_l\}$ and $\{\beta_l\}$ are the sequences of classical controls and quantum inputs, respectively. The tomography task learns the relation between $\mathcal{F}_l$ and $\{(s_l, \beta_l)\}$ for $l \leqslant L$ and reconstructs $\mathcal{F}_l$ with $l > L$.

We assume that the density matrix of each $\mathcal{F}_l$ is a $D \times D$-dimensional matrix. In standard quantum tomography, a naive approach would be to perform state tomography of $\mathcal{F}_l$ for every l. However, this would require many repetition experiments on copies of $\mathcal{F}_l$, resulting in an $O(D^4)$ number of measurements and the inversion of a huge $(D^4 \times D^4)$ linear system for every l. By leveraging the temporal dependency in $\mathcal{F}$ with the hybrid input $\{(s_l, \beta_l)\}$, our approach aims to streamline the experimental protocol and minimize the implementation cost of this task. Specifically, we assume that the target state $\mathcal{F}_l = \mathcal{F}\{(s_l, \beta_l)\}$ is accessible at $l = 1, \ldots, L$ for training. It is worth noting that accessible full tomography of the target is a common and necessary condition in implementing supervised learning methods in quantum tomography during the training phase. However, after the training process is completed, we can reconstruct the target quantum state using a single process of measurement in the QR, without the need for full tomography with multiple identical copies of the quantum state. Moreover, the QR can be applied to different types of hybrid quantum maps $\mathcal{F}$ without the need to reconfigure the measurement bases. Obviously, the QR cannot learn this hybrid task without the information contained in $\{s_l\}$. Therefore, we further assume that the classical control data are also accessible, although only in a distorted form of a nonlinear transformation $s_l \to u_l$. Since multitasking is feasible in the classical readout, we can also reconstruct $\{s_l\}$ from $\{u_l\}$.

In the following example, we consider $\mathcal{F}$ as a quantum switch with classical control [Fig. 1(a)]. Technically, a quantum switch includes two quantum channels $\mathcal{N}_A$ and $\mathcal{N}_B$ representing the operations by Alice and Bob, respectively, and an independent switch state $\rho(s) = |\psi(s)\rangle\langle\psi(s)|$ with $|\psi(s)\rangle = \sqrt{s}\,|0\rangle + \sqrt{1-s}\,|1\rangle$ $(0 \leqslant s \leqslant 1)$. Signal communication between Alice and Bob is only restricted to a partial order. However, the quantum switch can send the information under the indefinite causal order of quantum channels [8–12]. Given a state β on which these channels act, the quantum switch $\mathcal{S}(\mathcal{N}_A, \mathcal{N}_B)$ can be considered a function of hybrid input (s, β). Here, the quantum switch produces an output $[\mathcal{N}_A \circ \mathcal{N}_B(\beta)] \otimes |0\rangle\langle 0|$ if $s = 0$ and $[\mathcal{N}_A \circ \mathcal{N}_B(\beta)] \otimes |1\rangle\langle 1|$ if $s = 1$. When the switch state $\rho(s)$ is in a superposition of $|0\rangle$ and $|1\rangle$ $(0 < s < 1)$, the output becomes a quantum superposition of two alternative orders $\mathcal{N}_A \circ \mathcal{N}_B(\beta)$ and $\mathcal{N}_B \circ \mathcal{N}_A(\beta)$.

We use our QR to mimic the behavior of the quantum switch applied to the input sequence. Given a delay d, we demonstrate that the QR with current inputs β_l

and u_l can utilize memory effects to reconstruct $\sigma_l = \mathcal{S}(\mathcal{N}_A, \mathcal{N}_B)(s_{l-d}, \beta_{l-d})$ and s_{l-d}. We consider $\mathcal{N}_A$ and $\mathcal{N}_B$ as two depolarizing quantum channels (see Appendix C) and the reconstruction of $\{s_{l-d}\}$ from $\{u_l\}$ as the nonlinear channel equalization task. In wireless communication, signals sent from the antenna of a transmitter are transmitted to a receiver by following various paths while being reflected by structures such as buildings. Consequently, the transmitted signal is received with distortion due to the influence of noise added during transmission and the difference in transmission time depending on the path. Since this distortion depends on the frequency (channel), it is necessary to remove the distortion using an equalizer to demodulate the signal at the receiver. This process is called channel equalization.

The idea of using RC to learn the channel equalization task is presented in Ref. [37]. We further explain the essentials of channel equalization in this paper, which is also used in our paper. A sender wants to transmit a symbol sequence s_l, which is first converted into an analog envelope signal q_l. Afterward, the signal is modulated onto a high-frequency carrier signal and transmitted. Once received, the signal is demodulated into an analog signal u_l, which is often corrupted by various factors such as thermal noise, interfering signals, multipath propagation, and nonlinear distortion. Nonlinear distortion arises from operating the sender's power amplifier in the high-gain region and is mitigated by running the power amplification below the maximum level, resulting in significant energy loss. To overcome signal corruption, the equalizing filter is applied to the corrupted signal u_l, generating an output y_l that restores s_l as closely as possible. Finally, the equalized signal y_l is converted back into a symbol sequence $\hat{s}_l$. The quality of the entire process is measured by the symbol error rate (SER), which is the fraction of incorrect symbols. We further introduce a delay d to demonstrate that our task requires nonlinear effects and a memory to reconstruct s_{l-d}.

In our numerical experiments, $\{\beta_l\}$ is an independent and identically distributed (i.i.d.) sequence of one-qubit density matrices, and $\{s_l\}$ is an i.i.d. discrete sequence of symbols, where s_l are symbols selected from $\{0, \frac{1}{3}, \frac{2}{3}, 1\}$ with equal probability. We assume that the QR can only access $\{\beta_l\}$ and the distorted input $\{u_l\}$, which is transformed from $\{s_l\}$. We use the linear and nonlinear channels from Refs. [37,42] for the distortion:

$$e_l = 6s_l - 3, \tag{4}$$

$$\begin{aligned} q_l = {} & 0.08e_{l+2} - 0.12e_{l+1} + e_l + 0.18e_{l-1} - 0.1e_{l-2} \\ & + 0.09e_{l-3} - 0.05e_{l-4} + 0.04e_{l-5} \\ & + 0.03e_{l-6} + 0.01e_{l-7}, \end{aligned} \tag{5}$$

$$u_l = q_l + 0.036q_l^2 - 0.011q_l^3 + \nu_l. \tag{6}$$

Here, ν_l is an i.i.d. Gaussian noise with zero mean adjusted in power to yield a signal-to-noise ratio as 24 dB.

In the training stage, we are given an input sequence $\{(u_1, \beta_1), \ldots, (u_L, \beta_L)\}$ and the target sequence $\{(\hat{y}_1, \hat{\boldsymbol{Y}}_1), \ldots, (\hat{y}_L, \hat{\boldsymbol{Y}}_L)\}$ where $\hat{y}_l = s_{l-d}$ and $\hat{\boldsymbol{Y}}_l$ is the real vector form to stack the real and imaginary elements of σ_l. Here, we consider the size of the training data set as $T = 800$. The QR's output is divided into two parts: the tomography result $\boldsymbol{Y}_l$ in the real vector form and the equalized result y_l. The training is performed to minimize the mean-square error (MSE) between y_l and $\hat{y}_l$, and $\boldsymbol{Y}_l$ and $\hat{\boldsymbol{Y}}_l$ over $l = 1, \ldots, L$:

$$\mathrm{MSE} = \frac{1}{L}\sum_{l=1}^{L}\left(|y_l - \hat{y}_l|^2 + \|\boldsymbol{Y}_l - \hat{\boldsymbol{Y}}_l\|_2^2\right). \tag{7}$$

Here, we assume the access of some data at nonpositive index $s_0, s_{-1}, \ldots$ and $\beta_0, \beta_{-1}, \ldots$ for meaningful expression in Eq. (7). Since the QR's output is constructed from the linear combination of the occupation numbers $n_1, \ldots, n_K$ obtained from the QR's dynamics, the minimization of Eq. (7) can reduce to the minimization of $\frac{1}{L}\sum_{l=1}^{L}|y_l - \hat{y}_l|^2$ and $\frac{1}{L}\sum_{l=1}^{L}\|\boldsymbol{Y}_l - \hat{\boldsymbol{Y}}_l\|_2^2$, independently. This multitasking property is one of the advantages of the RC framework as we can use the same reservoir resource to solve multitasks simultaneously.

In the testing stage, we are given an input sequence $\{(u_{L+1}, \beta_{L+1}), \ldots, (u_{L+T}, \beta_{L+T})\}$ to produce the output sequence $\{(y_{L+1}, \boldsymbol{Y}_{L+1}), \ldots, (y_{L+T}, \boldsymbol{Y}_{L+T})\}$, where $T = 200$ is the size of the testing data set. Each y_l is converted back into a nearest symbol $\hat{s}_l \in \{0, \frac{1}{3}, \frac{2}{3}, 1\}$, and $\boldsymbol{Y}_l$ is transformed in the density matrix form $\hat{\sigma}_l$ with the consideration of a projection to obtain a positive semidefinite matrix (see Appendix D).

The tomography testing performance is evaluated via the root mean square of fidelities (RMSF):

$$\mathrm{RMSF} = \sqrt{(1/T)\sum_{l=L+1}^{l=L+T} F^2(\sigma_l, \hat{\sigma}_l)}, \tag{8}$$

where $F(\rho, \sigma) = \mathrm{Tr}[\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}]$. To evaluate RMSF, we assume that we have access to full tomography for the corresponding target states σ_l. Here, RMSF is only calculated to determine the appropriate values of model parameters and is not necessary once the model parameters are fixed. The equalization testing performance is evaluated via the SER:

$$\mathrm{SER} = \mathrm{card}(\{l \mid \hat{s}_l \neq s_{l-d}\})/T. \tag{9}$$

Figure 2(a) illustrates a sequence of one-qubit quantum inputs in the evaluation phase (upper panel) and a result for the channel equalizer (bottom panel) with delay $d = 1$. Here, the prediction and target sequences for the reconstruction of classical symbols $\{s_l\}$ are overlapped at almost all time steps. The density matrix at each time step is represented as a real vector by stacking the real and imaginary parts. Figure 2(b) depicts that the quantum target sequence can be reconstructed well. Here, in Fig. 2, we use $N = 3$ reservoir sites with $P/\gamma = 0.1$, $W/\gamma = 1.0$, and the measurement multiplexity $V = 8$.

Figure 3 displays the RMSF of the tomography task and the SER of the channel equalization task when we increase the value of the delay d with different number of sites in the QR. The performances drop significantly when $d \geqslant 3$, implying the short-term memory (STM) property in the QR. We further systematically evaluate the performance via the RMSF (left axis) and SER (right axis) in Fig. 4(a) for different N and W. A large value of W compared with h_{ij} and W_{jk}^{in} leads to nonergodic behavior in the QR, i.e., a strong and qualitative dependence on the initial state at t_{init}. In addition, in Appendix H, we further investigate the effects of the

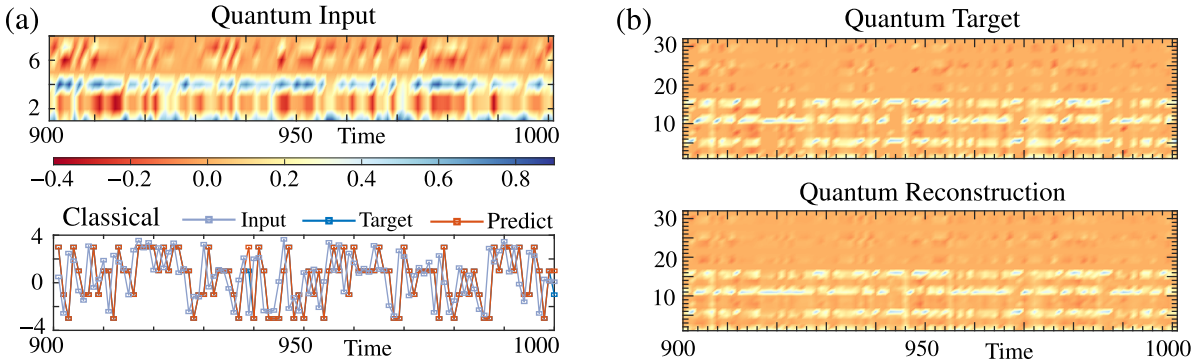

FIG. 2. Demonstration of the quantum tomography and the classical channel equalizer. (a) The upper panel shows a sequence of one-qubit quantum inputs, and the bottom panel displays the result for the channel equalizer in the evaluation phase. The horizontal axes represent time indices. In the upper panel, each quantum state at each time point is represented as a real vector, with the real and imaginary parts of the density matrix stacked together. The range of values is indicated by the color bar. (b) The target and reconstructed tomography with $N = 3$ reservoir sites, $P/\gamma = 0.1$, $W/\gamma = 1.0$, and the measurement multiplexity $V = 8$. The vectorization of the quantum state at each time point and the color bar are the same as in (a).

classical input in the reconstruction of the quantum input. With a large W, the input state is incident with weak coupling $[|W^{\text{in}}_{jk}| \ll |P(t) = P + Wu(t)|]$ under a strong effect of the classical input to the QR's dynamics, which means that not much information regarding quantum inputs can be retained in the QR. In contrast, a small W reduces the memory effect in reconstructing the previous classical input. This explains the existence of a region of W for an optimal performance $(W/\gamma \approx 1.0)$.

The left panel of Fig. 4(b) displays the RMSF of the tomography task when we increase the number N of reservoir sites. In the right panel of Fig. 4(b), we further compare the performance in the channel equalization task with the echo state network (ESN) in classical reservoir computing under the condition of the same number of computational nodes (see Appendix A). Here, we set the input scaling as $W/\gamma = 1.0$ and use the QR with the measurement multiplexity $V = 8$; therefore, the QR containing 16, 24, 32, 40, and 48 computational nodes corresponds to $N = 2, 3, 4, 5,$ and 6 sites in the reservoir. We confirm that with an appropriate setting of the constant coherent field P, we can obtain almost the same performance with the ESN in the channel equalization task. The crucial factor here is that, using the same computational resources, we can simultaneously perform both quantum and classical tasks without having to carry them out separately for each experimental setting. Our approach sheds light on the design of unconventional computing frameworks that can leverage a single resource to multitask with various types of input data.

At the end of this section, we should note that there is no direct classical access to the input quantum and output states of the QR. Only the classical readout relies on the measurements of the occupation numbers in the QR, which are then used to construct the tomography of the target quantum state. Therefore, we cannot compare our framework with other classical RC approaches in the context of quantum tomography. However, assuming classical access to the quantum input, we can perform a comparison with a classical baseline method. We implemented a linear regression approach that directly processes the input sequence of density matrices without using the reservoir dynamics. In this approach, we assume that we have access to full tomography of the input states, which is very costly and not required in our method. The reservoir state is constructed directly from the vector form of the hybrid input by stacking the real and imaginary parts of the density matrix of the quantum input and the value of the classical input. The training process is the same as our method. Under the same setting of input sequences in Fig. 4(b), we obtained an average RMSF over ten trials of 0.692 for the classical baseline method. This is significantly lower than the RMSF achieved by our method in Fig. 4(b). This result demonstrates that our QR represents a high-dimensional and nonlinear

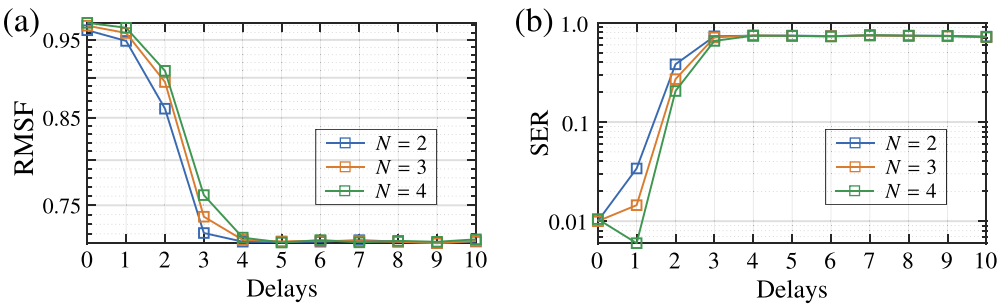

FIG. 3. The average RMSF of the tomography task (a) and the average SER of the channel equalization task (b) over ten trials when we increase the value of the delay d. Here, we consider the QR with $N = 2, 3, 4$ sites in the reservoir.

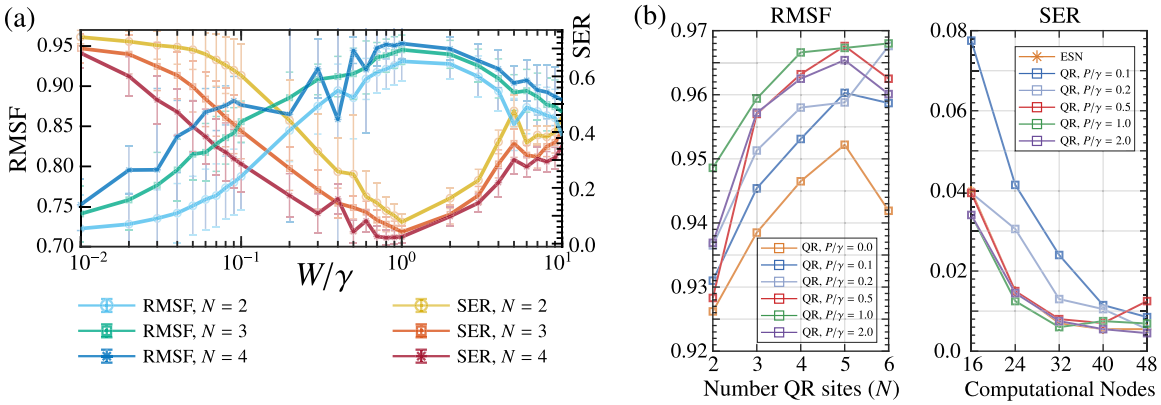

FIG. 4. Performance in the quantum tomography and classical channel equalizer. (a) The average RMSF and the average SER with shaded error bars over ten trials. (b) Left: The average RMSF in the tomography task when we increase the number N of reservoir sites in the QR. Right: Comparison between the average SER in the ESN and in our QR for the same number of computational nodes. In (b), we set the input scaling as $W/\gamma = 1.0$ and the measurement multiplexity as $V = 8$ for numerical experiments; therefore, the QR with number of computational nodes 16, 24, 32, 40, 48 corresponds to $N = 2, 3, 4, 5, 6$ sites in the reservoir.

transformation for hybrid input and outperforms the classical baseline approach based on the linear transformation of the input data.

C. Continuous variable quantum tomography and the closed loop

We modify the situation in the tomography task where, after the training phase, we were unable to access the information from the classical control s_l. Surprisingly, owing to the advantages of multitasking, our QR can autonomously generate s_l in a closed-loop manner while performing the tomography task with the hybrid input. In the training phase, s_l is learned in an open loop where we predict the next step s_{l+1} given the input $u_l = s_l$. After training, the prediction is used as the classical input for the next step, forming a closed-loop control without any external interventions. This model-free prediction is well established in classical reservoir computing, for example, to predict low-dimensional chaotic systems [37] or large spatiotemporally chaotic systems [43]. Our demonstration combines the closed-loop setting with the quantum tomography task, which is only effective in the QR setting.

We consider the quantum tomography of continuous variable states. The target is to reconstruct the output $\mathcal{F}_l = \mathcal{F}\{(s_l, \beta_l)\}$ in the Wigner function form $\mathcal{W}(\mathcal{F}_l; x_i, p_j)$ defined on a grid of continuous variables x_i and p_j (Appendix D). We use 300 randomly generated one-mode thermal states β_l and the periodic signals $s_l = 0.5 + 0.5\sin(\frac{l\pi f}{510})$ in the training phase. The target $\mathcal{F}_l$ is created by applying the one-mode squeezing operator to β_l as

$$\hat{S}(\xi_l) = \exp(\xi_l \hat{a}^\dagger \hat{a}^\dagger - \xi_l^* \hat{a}\hat{a}), \tag{10}$$

where $\xi_l = s_l e^{i\pi/4}$. Here, we consider that the cutoff Fock space dimension (the effective dimension) of these continuous variables states is $D_{\text{eff}} = 9$.

Figure 5(a) shows examples of the control signals in the training and closed-loop phase for $f = 60$. With $W/\gamma = 0.8$ and $N = 3$ sites, the control signal is almost reconstructed perfectly for all time steps in the closed-loop phase. This QR can efficiently reconstruct the Wigner function even without accessing the control signal [Fig. 5(b)]. We further investigate the stability of the closed-loop trajectories plotted in the (s_l, s_{l+10}) plane [Fig. 5(c)]. The QR presents a stable embedding of sinusoidal classical inputs if the trajectory can return to the target after adding a small perturbation (green line) into a predicted value, suggesting that our system successfully learned the target attractor. We observe an appropriate setting of input scaling W to obtain stable closed loops ($W/\gamma \approx 0.8$). Intriguingly, if we increase W/γ, for example to $W/\gamma = 1.8$, the closed loop fails to reconstruct the trajectory of the sinusoidal input in the evaluation stage but can produce chaoticlike behavior in the embedding space. In this case, the generated trajectory is not elliptical as the trajectory of sinusoidal inputs but still robust with respect to a small perturbation. We also observe the dependency of the performance of closed-loop control and the production of chaoticlike behavior on time scales f of the control signals, which are investigated in Appendix F.

D. Quantum readout and the depolarizing channel

We present an application using the quantum readout scheme to output quantum states. We use the QR to prepare a depolarizing quantum channel $\mathcal{F}\{(s_l, \beta_l)\} = s_l I/D + (1 - s_l)\beta_l$, where $\{\beta_l\}$ are randomly generated in a D-dimensional Hilbert space and $\{s_l\}$ is a random sequence in [0, 1].

First, we consider a sequence of 200 one-qubit quantum states for the training and 100 states for the evaluation. The baseline is computed when we set the output as the same as the input. We use the Nelder-Mead simplex algorithm [44] to minimize the fidelity error (see Appendix I for more details):

$$\text{EF} = \sqrt{\frac{1}{L}\sum_{l=1}^{L}[1 - F(\sigma_l, \hat{\sigma}_l)]^2}, \tag{11}$$

where σ_l and $\hat{\sigma}_l$ are the target and preparing quantum states, respectively. In Fig. 6(a), the evaluated fidelity errors with different readout and training configurations are presented for the QR with $N = 2$ sites, $P = 1.0$, and $W = 2.0$. The interquartile range is contained within the box, and the 5th

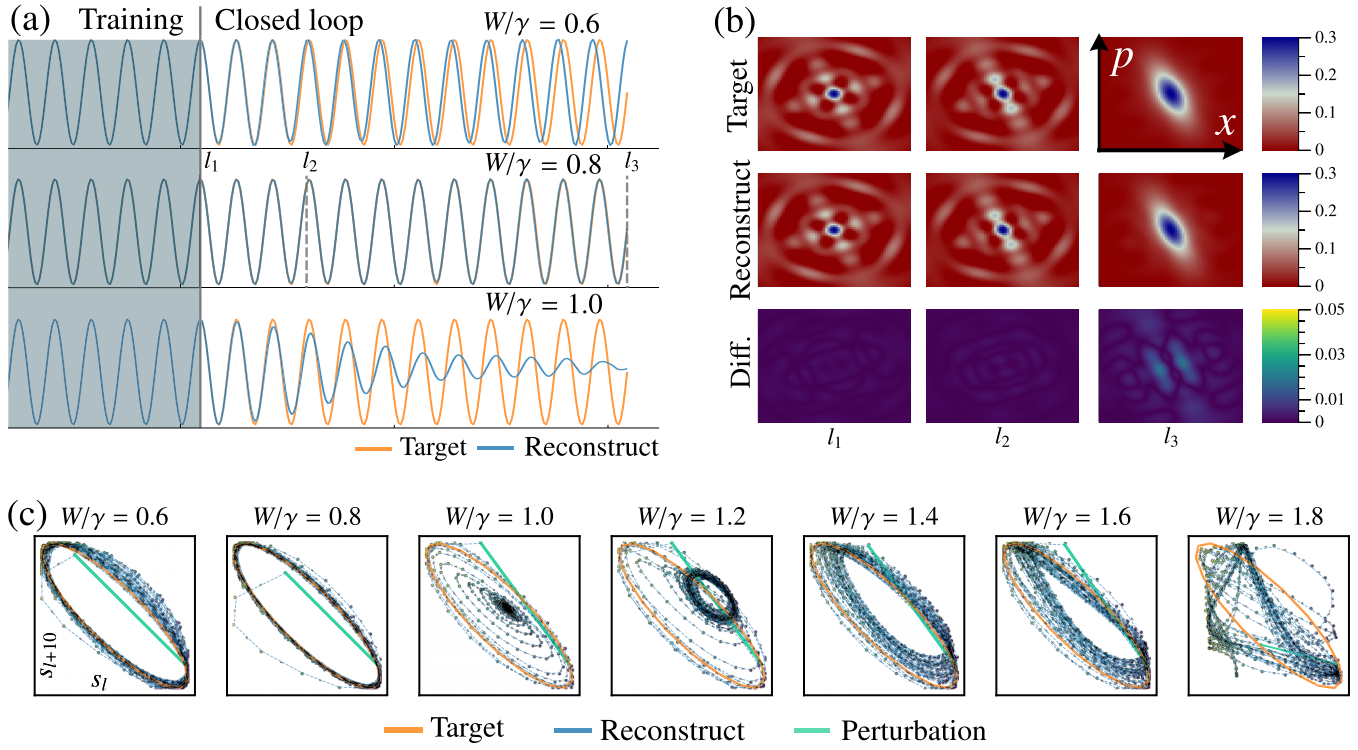

FIG. 5. Continuous variable tomography and closed-loop control of periodic classical signals. (a) Closed-loop control of classical signals with $N = 3$, $V = 10$, $P/\gamma = 1.0$, and $W/\gamma = 0.6, 0.8$, and 1.0. (b) Continuous variable tomography at typical time steps in the closed loop with $W/\gamma = 0.8$. The last panel displays the absolute difference between the target and reconstructed Wigner functions. (c) Stability after adding a small perturbation to the trajectory for different input scaling W/γ.

and 95th percentiles are marked by whiskers. The median is the line across the box, and the outliers are located outside the whiskers of each box plot. Here, IN, RV, and ALL correspond to the setting where only input modes $\hat{a}_k$, only reservoir modes $\hat{c}_j$, or both of them are considered as the readout nodes, respectively. Wo and Wio correspond to the situation where only readout weights or both readout weights and interaction coefficients W_{jk}^{in} are considered as the training parameters, respectively. The result implies that the consideration of both input and reservoir modes as N_R readout modes and both interaction coefficients and readout weights for training leads to the best performance. Under this setting, we display the variation in fidelity errors EF with the input scaling W/γ and N_R in Fig. 6(b). Even with a small QR ($N_R = 3, 4$) we can prepare the target channel with a relatively low error (<2%), which is significantly better compared with the baseline (≈8%). Furthermore, increasing W/γ basically leads to a better performance where more information regarding the classical input is integrated.

Finally, we prepare the depolarizing channel using the input quantum states as random squeezed thermal states in the continuous variable form. We minimize the cost function

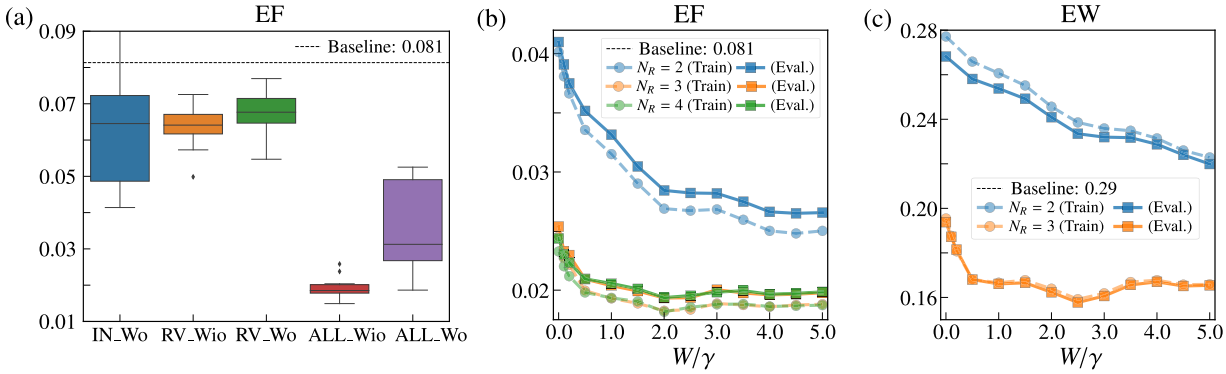

FIG. 6. The error in training and evaluating the quantum readout to prepare the depolarizing quantum channel. (a) Combinations of readout nodes and training parameters, where only input modes (IN), only reservoir modes (RV), or both of them (ALL) are considered as N_R readout nodes. The training parameters are readout weights (Wo) or both readout weights and interaction coefficients (Wio). (b), (c) Fidelity error with one-qubit input states (b) and error taken in the Wigner representation with continuous variable states (c) varying with input scaling W/γ.

taken in the Wigner representation as follows:

$$\mathrm{EW} = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\frac{\sum_{i,j}[\mathcal{W}(\sigma_l;x_i,p_j)-\mathcal{W}(\hat{\sigma}_l;x_i,p_j)]^2}{\sum_{i,j}[\mathcal{W}(\sigma_l;x_i,p_j)+\mathcal{W}(\hat{\sigma}_l;x_i,p_j)]^2}}. \quad (12)$$

Owing to the scale limitation, we only simulate the continuous variable states of the effective dimension $D_{\mathrm{eff}} = 3$, where $D = D_{\mathrm{eff}}^2 = 9$. Figure 6(c) presents the errors in 50 training and 50 evaluating data varying with W. We can observe a similar trend in Fig. 6(b); that is, with sufficient classical information ($W/\gamma \geqslant 1.0$), the error EW ($\approx$0.16) with $N_R = 3$ readout nodes is significantly lower than the baseline's error ($\approx$0.29). This result is still below a considerably good preparation (EW < 0.1), but it demonstrates that hybrid inputs can be effectively considered for training the quantum readout.

III. DISCUSSION

We proposed a framework for an analog QR processor with hybrid inputs and classical and quantum readouts for learning heterogeneous quantum-classical data. This aligns well with scenarios where one wishes to model a quantum device to process quantum input but must rely on classical control signals in physical experiments. Our framework, therefore, has the potential to be physically implemented in quantum network systems where classical control and quantum sources can interact with nonlinear quantum systems to form a quantum channel. It can help realize quantum adaptive systems capable of quantum information processing. These agents can be used to interpret and memorize both classical and quantum signals from their environment and to respond accordingly to the actions of their surroundings [45].

Processing hybrid quantum and classical data is a promising idea to facilitate future innovative use cases for quantum computers. This concept aims to leverage the advantages of quantum mechanics in ML with an unconventional computing framework and intriguing applications. It is not limited to the conventional discussion on practical quantum advantages, such as the "beating speedup" of quantum to classical ML methods [46]. For example, classical readouts lead to interesting applications of multitasking where quantum data can be processed in a closed loop of the classical control. Furthermore, adding this closed-loop mechanism allows us to utilize the unique coherence properties of quantum systems to generate unique classical dynamics. We consider the quantum readout to avoid the measurement process of preparing the quantum output. However, optimization is challenging and requires improvement, since we need to simulate or drive the quantum system and evaluate the cost function for a wide range of parameters.

A further enticing discussion would be the case of the correlation between the processing of quantum data and classical data in a hybrid setting of the QR. We can consider a QR to simultaneously process quantum data and classical data as separate tasks. An intriguing research question arises: Can this multitasking mechanism induce positive or negative effects on information processing? For example, if we repeatedly modify the coherent field strengths of the QR via a classical input with a large magnitude, it can limit the short-term memory properties of quantum data processing (see Appendices G and H). However, one can also expect positive effects and not only negative ones. There may exist a situation where simultaneously processing different models of data can actually create an optimal regime rather than solely solving a single task. We can start by investigating relations between hybrid input protocols with the dynamics of the QR, such as the fact that the classical input may induce the dynamical phase transition in the QR [47]. We can also study how classical and quantum data are processed via the QR's dynamics, such as by decomposing the readout reservoir states in terms of basis polynomials for input history [48,49]. Along this research line, one can refer to a recent study demonstrating that quantum noise in real quantum processors can induce the information processing capability when using classical data [50].

Finally, while our framework is general and applicable to different types of quantum maps, its performance is only evaluated numerically. An intriguing theoretical research direction pertains to a quantum version of the universal approximation property (UAP) in RC [51]. The UAP refers to the ability to find elements within the RC class (such as the ESN class) that can closely approximate any fading memory map with arbitrary precision. Here, a fading memory map is a continuous function that depends on a finite number of past inputs. In the context of QRC with quantum or hybrid inputs, a similar and intriguing question arises: Can we find elements within a QRC class that can reconstruct arbitrary quantum channels? We consider this to be an interesting and open question for future research.

ACKNOWLEDGMENTS

The authors acknowledge Shumpei Kobayashi for fruitful discussions. This work is supported by the Ministry of Education, Culture, Sports, Science, and Technology Quantum Leap Flagship Program (Grants No. JPMXS0118067394 and No. JPMXS0120319794).

APPENDIX A: RESERVOIR COMPUTING

Mathematically, RC is described by the input-driven map $g : \boldsymbol{U} \times \mathcal{X} \to \mathcal{X} \subset \mathbb{R}^K$, where $\boldsymbol{U}$ and $\mathcal{X}$ are the input and the reservoir's state space, respectively. Here, K is considered the dimension of the reservoir's state. If we feed a discrete-time input sequence $\{\ldots, \boldsymbol{u}_{-1}, \boldsymbol{u}_0, \boldsymbol{u}_1, \ldots\}$ into the reservoir, the readout reservoir state $\boldsymbol{x}_l$ is represented by the following recurrent relation:

$$\boldsymbol{x}_l = g(\boldsymbol{u}_l, \boldsymbol{x}_{l-1}). \quad (A1)$$

In temporal supervised learning tasks, we are given an input sequence $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_L\}$ and the corresponding target sequence $\hat{\boldsymbol{y}} = \{\hat{\boldsymbol{y}}_1, \ldots, \hat{\boldsymbol{y}}_L\}$, where $\hat{\boldsymbol{y}}_k \in \mathbb{R}^d$ with d the output dimension. We consider a parametrized readout map $h_{\boldsymbol{w}} : \mathcal{X} \to \mathbb{R}^d$, where the output signal is $\boldsymbol{y}_l = h_{\boldsymbol{w}}(\boldsymbol{x}_l)$. The readout map is often taken as a linear combination of the readout reservoir states as $\boldsymbol{y}_l = h_{\boldsymbol{w}}(\boldsymbol{x}_l) = \boldsymbol{w}^\top \boldsymbol{x}_l$. Here, $\boldsymbol{w}$ is the trainable parameter obtained by minimizing the mean squared error

$$\mathrm{MSE} = \frac{1}{L}\sum_{l=1}^{L} \|\boldsymbol{y}_l - \hat{\boldsymbol{y}}_l\|_2^2, \quad (A2)$$

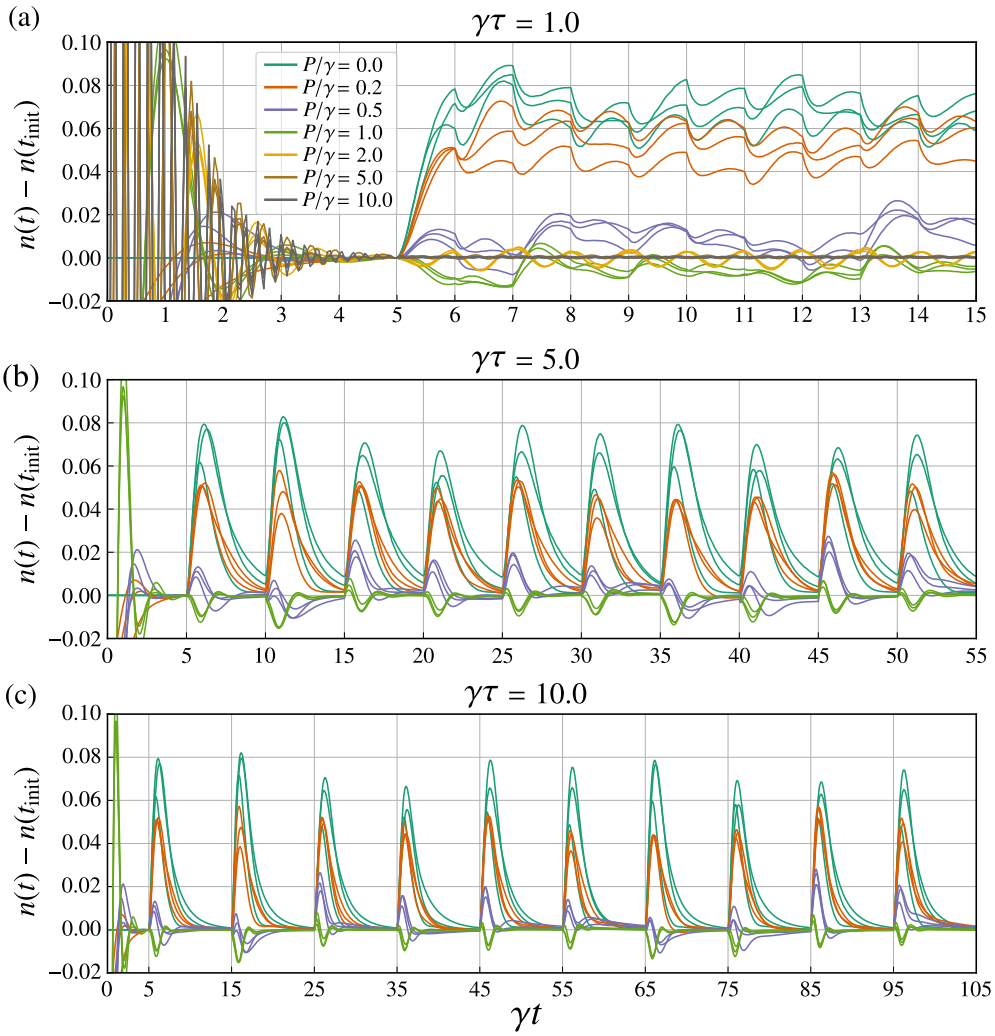

FIG. 7. The typical dynamics of the occupation numbers $n_j(t)$ in the QR, represented by the amount of the input photons $n_j(t) - n_j(t_{\rm init})$ entering the QR. The dynamics starts from $t = 0$ to $t_{\rm init} = 5.0/\gamma$, where one-qubit quantum inputs are incident at $t_l = t_{\rm init} + (l-1)\tau$.

where $\|\cdot\|_2$ denotes the Euclidean norm between two vectors in $\mathbb{R}^d$. For training, we add a constant bias term $x_{l,K+1} = 1$ to the readout reservoir state $\boldsymbol{x}_l$ and optimize $\boldsymbol{w}$ via the linear regression $\hat{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{w}$. Here, $\hat{\boldsymbol{Y}} = [\hat{\boldsymbol{y}}_1 \quad \ldots \quad \hat{\boldsymbol{y}}_L]^\top$ is the $L \times d$ target matrix and $\boldsymbol{X}$ is the $L \times (K+1)$ matrix that combines the readout reservoir states $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_L$ of the training data. The optimal value of $\boldsymbol{w}$ is obtained via the Ridge regression in the matrix form $\hat{\boldsymbol{w}}^\intercal = (\boldsymbol{X}^\top\boldsymbol{X} + \eta\boldsymbol{I})^{-1}\boldsymbol{X}^\top\hat{\boldsymbol{Y}}$, where η is a positive constant for the regularization.

An ESN is a realization of the input-driven map in RC. It belongs to the concept of an artificial recurrent neural network, as we have a large network with randomly fixed inner and recurrent connections. Considering an ESN with N computational nodes and a discrete-time input sequence $\{\boldsymbol{u}_l\}$, the reservoir state $\boldsymbol{x}_l$ at time step l is described by

$$\boldsymbol{x}_{l+1} = \mathbf{tanh}(\boldsymbol{W}^{\rm in}\boldsymbol{u}_{l+1} + \boldsymbol{W}\boldsymbol{x}_l),$$

where $\mathbf{tanh}(\cdot)$ is the activation function applied on vector $\boldsymbol{x}=(x_1, x_2, \ldots, x_K)^\top$ as $\mathbf{tanh}(\boldsymbol{x})=(\tanh(x_1), \ldots, \tanh(x_K))^\top$. Here, $\boldsymbol{W}^{\rm in}$ and $\boldsymbol{W}$ are the input weight matrix and recurrent weight matrix, respectively. In the channel equalization task, the input weight matrix $\boldsymbol{W}^{\rm in}$ is generated from random uniform distribution in $[-1, 1]$. We also set the connection probability and the spectral radius of the recurrent weight matrix $\boldsymbol{W}$ to 0.1 and 0.9, respectively.

APPENDIX B: DYNAMICS OF THE RESERVOIR'S RESPONSE

The assumption that the QR reaches a steady state before processing the input sequence is based on the assumption of ergodicity, which means that the reservoir dynamics will eventually relax to a unique steady state regardless of its initial conditions. This assumption is difficult to validate theoretically, especially if the specifics of the reservoir dynamics are unknown. In our paper, we validate the ergodicity assumption by measuring the reservoir's response and selecting a large enough initial time $t_{\rm init}$ such that the QR at time $t_{\rm init}$ reaches a steady state. For example, in Fig. 7(a), we show the dynamics of the occupation numbers $n_1(t)$, $n_2(t)$, and $n_3(t)$ compared to the corresponding numbers at time $t_{\rm init}$ for $N = 3$ reservoir sites with a constant classical input ($W/\gamma = 0$) and randomized quantum inputs of one-qubit quantum states. Here, we see that at $t_{\rm init} = 5.0/\gamma$ the QR relaxes to the steady state

for a wide range of P values. The dynamics from $t = 0$ to $t_1 = 5.0/\gamma$ are solely driven by P, and the system reaches an initial state before t_{init}. The first quantum input is incident at $t_1 = t_{\text{init}}$, where we can see that the occupation numbers deviate from the steady states.

Furthermore, it is further important to set the interval τ. We fix $t_1 = 5.0/\gamma$ and investigate the reservoir's response with different input intervals τ and different incoherent excitation P. For large values of τ [Figs. 7(b) and 7(c)], the occupation numbers return to their steady-state values before the next input is incident on the system. If we perform the readout measurements at this timing, we cannot obtain sufficient information from the input. Therefore, we choose τ such that $n_j(t)$ deviates significantly from their steady values at the timing before the next input is incident on the system [Fig. 7(a)]. In our experiments, we consider $\gamma\tau = 1.0$.

APPENDIX C: QUANTUM SWITCH

In the classical counterpart, a switch is an operation of control that can route a target system through two classical channels C_A and C_B in series following one causal order (C_A then C_B) or the other (C_B then C_A). The quantum switch is different in that it induces entirely new quantum trajectories where the order of the two operators is indefinite. Technically, a quantum switch includes two quantum channels $\mathcal{N}_A$ and $\mathcal{N}_B$ to create a new channel $\mathcal{S}(\mathcal{N}_A, \mathcal{N}_B)$, which uses the channels $\mathcal{N}_A$ and $\mathcal{N}_B$ in an order that is entangled with an independent switch quantum state $\rho(s)$. Given a quantum input β, the channel $\mathcal{S}(\mathcal{N}_A, \mathcal{N}_B)$ returns the state $[\mathcal{N}_A \circ \mathcal{N}_B(\beta)] \otimes |0\rangle\langle 0|$ if $s = 0$ and $\rho(0) = |0\rangle\langle 0|$ and $[\mathcal{N}_B \circ \mathcal{N}_A(\beta)] \otimes |1\rangle\langle 1|$ if $s = 1$ and $\rho(1) = |1\rangle\langle 1|$. When $\rho(s)$ is in a superposition of $|0\rangle$ and $|1\rangle$, the channel returns a correlated state as a result of $\mathcal{N}_A$ and $\mathcal{N}_B$ acting on β in a superposition of two alternative orders.

To describe $\mathcal{S}(\mathcal{N}_A, \mathcal{N}_B)$, we denote the Kraus operators of channels $\mathcal{N}_A$ and $\mathcal{N}_B$ as $\{K_i^{(A)}\}$ and $\{K_j^{(B)}\}$, respectively, where $\mathcal{N}_A(\beta) = \sum_i K_i^{(A)} \beta K_i^{(1)\dagger}$ and $\mathcal{N}_B(\beta) = \sum_j K_j^{(B)} \beta K_j^{(2)\dagger}$. The Kraus operators of $\mathcal{S}(\mathcal{N}_A, \mathcal{N}_B)$ are defined as

$$W_{ij} = K_i^{(A)} K_j^{(B)} \otimes |0\rangle\langle 0| + K_j^{(B)} K_i^{(A)} \otimes |1\rangle\langle 1| . \tag{C1}$$

The action of the quantum switch is given by

$$\mathcal{S}(\mathcal{N}_A, \mathcal{N}_B)[\beta \otimes \rho(s)] = \sum_{i,j} W_{ij} [\beta \otimes \rho(s)] W_{ij}^\dagger . \tag{C2}$$

In our paper, we consider $\mathcal{N}_A$ and $\mathcal{N}_B$ as two depolarizing channels with parameters q_A and q_B, which are given by

$$\mathcal{N}_A(\beta) = (1 - q_A)\beta + q_A \frac{I}{D} = \frac{q_A}{D^2} \sum_{i=0}^{D^2} U_i \beta U_i^\dagger , \tag{C3}$$

$$\mathcal{N}_B(\beta) = (1 - q_B)\beta + q_B \frac{I}{D} = \frac{q_B}{D^2} \sum_{j=0}^{D^2} V_j \beta V_j^\dagger , \tag{C4}$$

where $D \times D$ is the dimension of β and $\{U_i\}_{i=1}^{D^2}$ and $\{V_j\}_{j=1}^{D^2}$ are orthonormal bases of the space of $D \times D$ matrices. Here, we introduce the notation $U_0 = \frac{D\sqrt{1-q_A}}{\sqrt{q_A}} I$ and $V_0 = \frac{D\sqrt{1-q_B}}{\sqrt{q_B}} I$. We define the extension Kraus operators for $\mathcal{N}_A$ and $\mathcal{N}_B$ as $K_i^{(A)} = \frac{\sqrt{q_A}}{D} U_i$ for $i = 0, 1, \ldots, D^2$ and $K_j^{(B)} = \frac{\sqrt{q_B}}{D} V_j$ for $j = 0, 1, \ldots, D^2$, respectively. We can express the Kraus operators of $\mathcal{S}(\mathcal{N}_A, \mathcal{N}_B)$ as

$$W_{ij} = \frac{\sqrt{q_A q_B}}{D^2} (U_i V_j \otimes |0\rangle\langle 0| + V_j U_i \otimes |1\rangle\langle 1|). \tag{C5}$$

We consider the control state $\rho(s) = |\psi(s)\rangle\langle\psi(s)|$, where $\psi(s) = \sqrt{s}\,|0\rangle + \sqrt{1-s}\,|1\rangle$ $(0 \leqslant s \leqslant 1)$. The output of the quantum switch is given by

$$\mathcal{S}(\mathcal{N}_A, \mathcal{N}_B)[\beta \otimes \rho(s)] = A^{00} \otimes |0\rangle\langle 0| + A^{01} \otimes |0\rangle\langle 1| + A^{10} \otimes |1\rangle\langle 0| + A^{11} \otimes |1\rangle\langle 1| , \tag{C6}$$

where

$$A^{00} = s \frac{q_A q_B}{D^4} \sum_{i=0}^{D^2} \sum_{j=0}^{D^2} U_i V_j \beta V_j^\dagger U_i^\dagger = s \mathcal{N}_A \mathcal{N}_B(\beta), \tag{C7}$$

$$A^{01} = A^{10} = \sqrt{s(1-s)} \frac{q_A q_B}{D^4} \sum_{i=0}^{D^2} \sum_{j=0}^{D^2} U_i V_j \beta U_i^\dagger V_j^\dagger \tag{C8}$$

$$= \sqrt{s(1-s)} \left(\frac{q_A q_B}{D^2} \beta + q_A (1 - q_B) \frac{I}{D} + q_B (1 - q_A) \frac{I}{D} + (1 - q_A)(1 - q_B)\beta \right), \tag{C9}$$

$$A^{11} = (1 - s) \frac{q_A q_B}{D^4} \sum_{i=0}^{D^2} \sum_{j=0}^{D^2} V_j U_i \beta U_i^\dagger V_j^\dagger = (1 - s) \mathcal{N}_B \mathcal{N}_A(\beta). \tag{C10}$$

APPENDIX D: LEARNING QUANTUM TOMOGRAPHY

A quantum device can be described by a function of quantum input β and classical control u as $\mathcal{F}(u, \beta)$, where we consider the scalar u for ease of notation. Given a sequence of hybrid inputs (u_1, β_1), (u_2, β_2), …, and a quantum device with a time-dependent behavior, we can describe it using the temporal map $\mathcal{F}_l = \mathcal{F}(\{(u_i, \beta_i)\}_{i=1:l})$ of the current and past inputs [34]. We assume that we have full tomography for the corresponding output states of $\mathcal{F}$ in the training, where we are given a hybrid input sequence $\{(u_1, \beta_1), \ldots, (u_L, \beta_L)\}$ and the corresponding target sequence $\hat{\mathbf{y}} = \{\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_L\}$. Here, $\hat{\mathbf{y}}_l$ is the real vector form of $\mathcal{F}_l$. If $\mathcal{F}_l$ is described by the

density matrix, $\hat{\boldsymbol{y}}_l$ is formed by stacking the real and imaginary elements of $\mathcal{F}(\beta_l)$. In the evaluation stage, we are given an input sequence $\{(u_{L+1}, \beta_{L+1}), \ldots, (u_{L+T}, \beta_{L+T})\}$ with the corresponding target $\{\hat{\sigma}_{L+1}, \ldots, \hat{\sigma}_{L+T}\}$, where $\hat{\sigma}_i = \mathcal{F}(u_i, \beta_i)$. The output sequence is $\{\boldsymbol{y}_{L+1}, \ldots, \boldsymbol{y}_{L+T}\}$, which is rearranged in the matrix form $\{\sigma'_{L+1}, \ldots, \sigma'_{L+T}\}$. To obtain the final positive semidefinite matrix σ_i, we project σ'_i onto the spectrahedron such that the trace of σ_i is equal to 1 and the Frobenius norm of $\sigma_i - \sigma'_i$ is minimized [52,53].

Tomography learning can be performed with other forms of $\mathcal{F}_l$, for example, in the Wigner function representation of continuous variable states. Given a density matrix σ, the continuous variable states associated with σ can be described by the Wigner function

$$W(\sigma; x_i, p_j) = \int \frac{dy}{2\pi} \left\langle x_i + \frac{y}{2} \middle| \sigma \middle| x_i - \frac{y}{2} \right\rangle e^{-iyp_j}, \tag{D1}$$

where the integral is evaluated in the whole space and the states $|x_i \pm \frac{y}{2}\rangle$ represent continuous position bases. We evaluate Wigner functions on a 61×61 grid of x_i and p_j, where we divide the interval $[-3, 3]$ into 60 equal intervals for the range of x_i and p_j. The target of continuous variable tomography is to reconstruct these Wigner functions, i.e., the real 61×61 dimensional matrices.

APPENDIX E: NONTEMPORAL CONTINUOUS VARIABLE QUANTUM TOMOGRAPHY

In this demonstration, we consider the quantum tomography of continuous variable states in a nontemporal setting. We evaluate the tomography for three settings $\mathcal{F}_{\text{cv-amp}}$, $\mathcal{F}_{\text{cv-phase}}$, and $\mathcal{F}_{\text{cv-sw}}$, of the quantum map $\mathcal{F}(s, \beta)$, given a one-mode quantum input β and a classical input s.

For $\mathcal{F}_{\text{cv-amp}}$ and $\mathcal{F}_{\text{cv-phase}}$, we consider a random sequence in $[0, 1]$ of $\{s_l\}_{l=1:200}$ and a random sequence of one-mode thermal states $\{\beta_l\}_{l=1:200}$ and take the index of $l = 1, \ldots, 100$ for the training and $l = 101, \ldots, 200$ for the evaluation. Here, we consider one-mode thermal states β_l as Gaussian continuous variable states with the density matrices

$$\beta_l = \sigma_l \quad \text{for} \quad \sigma_l = \frac{1}{1+\overline{v}_l} \sum_{n=0}^{\infty} \left(\frac{\overline{v}_l}{1+\overline{v}_l} \right)^n |n\rangle \langle n|, \tag{E1}$$

where $|n\rangle$ represents the state corresponding to n photon numbers, and $\overline{v}_l$ is the expectation value of the photon number in the state. We consider $\overline{v}_l = [r_l \cos(\phi_l)]^2$, where r_l and ϕ_l are taken randomly in $[0.0, 0.3]$ and $[0.0, \pi]$, respectively. The quantum maps $\mathcal{F}_{\text{cv-amp}}$ and $\mathcal{F}_{\text{cv-phase}}$ are defined as

$$\mathcal{F}_{\text{cv-amp}}(s, \beta) = \hat{\mathcal{S}}[\xi_{\text{amp}}(s)] \beta \hat{\mathcal{S}}[\xi_{\text{amp}}(s)]^\dagger, \tag{E2}$$

$$\mathcal{F}_{\text{cv-phase}}(s, \beta) = \hat{\mathcal{S}}[\xi_{\text{phase}}(s)] \beta \hat{\mathcal{S}}[\xi_{\text{phase}}(s)]^\dagger, \tag{E3}$$

where $\mathcal{S}(\xi)$ is the one-mode squeezing operator, defined as $\mathcal{S}(\xi) = \exp(\xi \hat{a}^\dagger \hat{a}^\dagger - \xi^* \hat{a}\hat{a})$. Here, we consider ξ as functions of classical data s defined as

$$\xi_{\text{amp}}(s) = s \exp(i\pi/4), \quad \xi_{\text{phase}}(s) = 0.3 \exp(i2\pi s). \tag{E4}$$

For the quantum map $\mathcal{F}_{\text{cv-sw}}$, we consider the same $\{s_l\}$ but random one-mode squeezed-thermal states

$$\beta_l = \hat{\mathcal{S}}(\xi_l) \sigma_l \hat{\mathcal{S}}(\xi_l)^\dagger, \tag{E5}$$

where σ_l is defined as in Eq. (E1) and $\xi_l = r_l \sin(\phi_l)$. The quantum map $\mathcal{F}_{\text{cv-sw}}$ is defined as the quantum switch with the input β_l and the switch state

$$|\psi(s_l)\rangle = \sqrt{s_l}\, |\alpha\rangle + \sqrt{1 - s_l}\, |-\alpha\rangle, \tag{E6}$$

where $\alpha = 2.5$ with the following coherent states:

$$|\pm\alpha\rangle = \exp\left(-\frac{|\alpha|^2}{2}\right) \sum_{n=0}^{\infty} \frac{(\pm\alpha)^n}{\sqrt{n!}} |n\rangle. \tag{E7}$$

Figure 8(a) shows several examples of input, target, and reconstruct Wigner functions for $\mathcal{F}_{\text{cv-phase}}$ (upper panel), $\mathcal{F}_{\text{cv-amp}}$ (middle panel), and $\mathcal{F}_{\text{cv-sw}}$ (lower panel). Here, we use $N =$ 4-site QR with a measurement multiplexity $V = 10$ and a constant coherent field strength $P/\gamma = 1.0$. The columns labeled "(Hybrid)" and "(Quantum only)" describe the results when we consider the input scaling $W/\gamma = 1.0$ (both classical and quantum inputs) and $W/\gamma = 0.0$ (no classical input), respectively. We observe that if both of quantum and classical inputs are included in the QR, the reconstructed states are very similar to the target states of the quantum maps with hybrid quantum-classical input.

To further evaluate the performance systematically, we calculate the error based on the Wigner representation

$$\text{EW} = \sqrt{\frac{1}{T} \sum_{l=L+1}^{L+T} \frac{\sum_{i,j} [\mathcal{W}(\mathcal{F}_l; x_i, p_j) - \hat{\mathcal{W}}(\mathcal{F}_l; x_i, p_j)]^2}{\sum_{i,j} [\mathcal{W}(\mathcal{F}_l; x_i, p_j) + \hat{\mathcal{W}}(\mathcal{F}_l; x_i, p_j)]^2}}, \tag{E8}$$

where $\mathcal{W}(\mathcal{F}_l; x_i, p_j)$ and $\hat{\mathcal{W}}(\mathcal{F}_l; x_i, p_j)$ are the target and reconstructed Wigner functions of the state $\mathcal{F}_l = \mathcal{F}(s_l, \beta_l)$, respectively. The error metric is evaluated in the evaluation phase with $L = 100$ and $T = 100$.

Figure 8(b) depicts the errors to reconstruct the quantum maps $\mathcal{F}_{\text{cv-phase}}$ (orange lines) and $\mathcal{F}_{\text{cv-amp}}$ (blue lines) at different input scaling W with $N = 2$ (upper plot) and $N = 3$ (lower plot). The errors are calculated over ten trials of data and QR configurations, with the solid lines depicting the average values associated with error bars. We observe an optimal range of input scaling W for optimal performance in each task. We note that setting a too small value of W limits the effect of the classical input into the QR. In contrast, setting a too large value of W will impose the localization in the quantum dots and may lead to nonergodic behavior in the QR. In this case, when the input state β_l is incident on the QR with weak coupling ($|W^{\text{in}}_{jk}| \ll |P + W s_l|$), sufficient information regarding β_l cannot be extracted from the QR.

Figure 8(c) depicts the errors EW (blue lines labeled "Full-train") to reconstruct the quantum map $\mathcal{F}_{\text{cv-sw}}$. We can observe that the effect of input scaling W is not significant as in other tasks. We further present an intriguing setting by limiting the variety of the classical input in the training phase while keeping the same data in the evaluating phase. In the results labeled "Bin-train," we only consider binary classical data in the training phase, i.e., $s_l \in \{0.0, 1.0\}$ for $l = 1, \ldots, 100$. The performance is worse since there is no superposition of the switch state in the training phase to help the learning of the quantum switch. However, for trivalues of $s_l \in \{0.0, 0.5, 1.0\}$ (labeled "Tri-train") in the training phase, only one pattern of

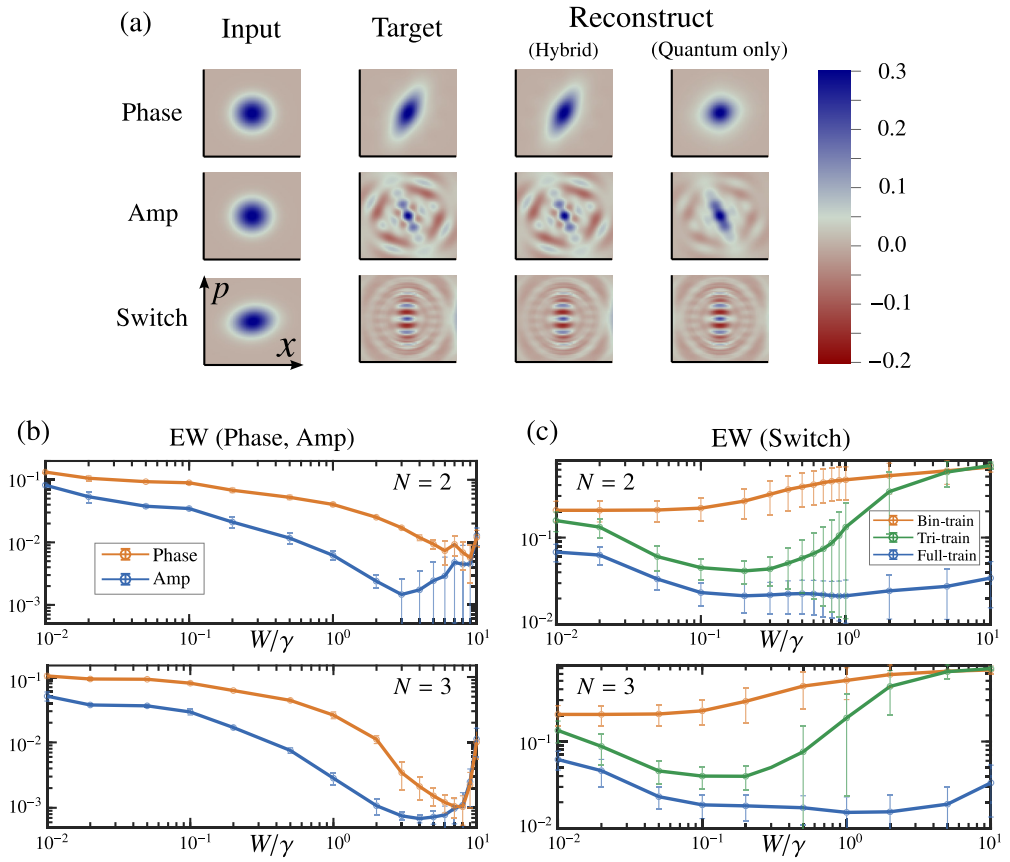

FIG. 8. (a) Examples of input, target, and reconstruct Wigner functions in the nontemporal continuous variable tomography for $\mathcal{F}_{\text{cv-phase}}$ (upper panel), $\mathcal{F}_{\text{cv-amp}}$ (middle panel), and $\mathcal{F}_{\text{cv-sw}}$ (lower panel) by $N = 4$-site QR with a measurement multiplexity $V = 10$ and a constant coherent field strength $P/\gamma = 1.0$. The columns labeled "(Hybrid)" and "(Quantum only)" describe the results with the input scaling $W/\gamma = 1.0$ (both classical and quantum inputs) and $W/\gamma = 0.0$ (no classical input), respectively. (b) Variation in the tomography error EW for $\mathcal{F}_{\text{cv-phase}}$ (orange line) and $\mathcal{F}_{\text{cv-amp}}$ (blue line) with the input scaling W and $N = 2, 3$-site QR. (c) Variation in the tomography error EW for $\mathcal{F}_{\text{cv-sw}}$ and $N = 2, 3$-site QR. In the results labeled as "Full-train," we consider a random sequence in $[0, 1]$ of $\{s_l\}_{l=1:200}$ and a random sequence of one-mode thermal states $\{\beta_l\}_{l=1:200}$ with the index of $l = 1, \ldots, 100$ for the training and $l = 101, \ldots, 200$ for the evaluation. In the results labeled "Bin-train" and "Tri-train," only binary or trivalues classical data in the training phase are considered, i.e., $s_l \in \{0.0, 1.0\}$ (for "Bin-train") and $s_l \in \{0.0, 0.5, 1.0\}$ (for "Tri-train") for $l = 1, \ldots, 100$.

the superposition switch state is considered in the training; we can obtain a relatively low error with a suitable range of input scaling W. For example, the performance at $W = 0.2$ is comparable with the performance of "Full-train" at $W = 0.02$. These results demonstrate that tomography for the quantum switch can be performed with limited patterns of training data.

APPENDIX F: TIME SCALES OF THE CLASSICAL INPUT

In Sec. II C, we demonstrated an intriguing problem in predicting the future evolution of the continuous variable quantum tomography of hybrid inputs The performance depends on the prediction of the next step of the classical control signals, forming a closed-loop control without any external interventions.

We notice the dependency of the performance on the time scales of the classical control signals, where the classical inputs with higher frequencies f (lower timescales) basically lead to better performance. We analyze the effect of perturbation to investigate the stability of the embedded classical trajectories. Figure 9 shows the output dynamics of both the target and perturbed prediction trajectories in the closed-loop phase plotted in the (s_l, s_{l+10}) plane for different values of f and W/γ. After the training phase, we add a small perturbation into the predicted value, which results in an extra drift in the (s_l, s_{l+10}) plane (green line). The reservoir presents a stable embedding of sinusoidal classical inputs if the trajectory can return to the target one after the addition of the perturbation. We observe appropriate ranges of input scaling W/γ and f to obtain stable closed loops. Furthermore, if we increase the input scaling W/γ, the closed loop fails to reconstruct the trajectory of sinusoidal inputs but can produce chaoticlike behavior in the embedding space. Intriguingly, the generated trajectory is not elliptical as the trajectory of sinusoidal inputs but is still robust with respect to a small perturbation.

To characterize the activity of the QR, we observe the dynamics of the average occupation numbers $\bar{n}(t)$ over reservoir sites at different f and input scaling W [left panels in Fig. 10(a)]. In the presence of periodic classical inputs $(W/\gamma > 0.0)$, an oscillatory response is superposed on the intrinsic dynamics of the quantum input without the classical input $(W/\gamma = 0.0)$. We further calculate the autocorrelation

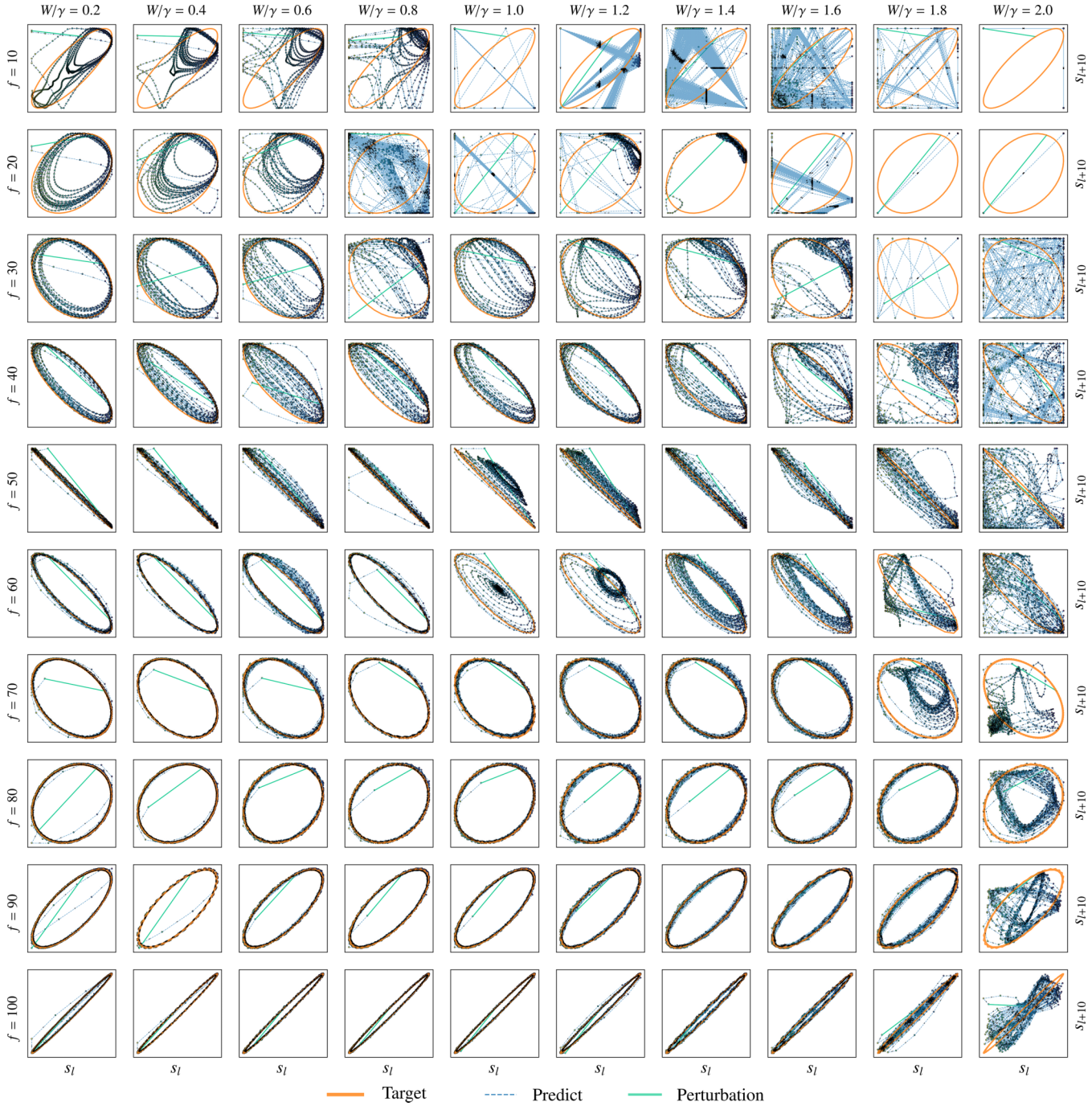

FIG. 9. Stability after adding a small perturbation to the trajectory. The QR presents a stable embedding of sinusoidal classical inputs if the trajectory can return to the target after adding a small perturbation (green line) into the predicted value. We observe appropriate ranges of input scaling W and f to obtain stable closed loops. There is an intriguing observation that if we increase the input scaling W/γ the closed loop fails to reconstruct the trajectory of sinusoidal inputs but can produce chaoticlike behavior in the embedding space.

function of each frequency f averaged across all the reservoir sites:

$$C(\tau_c) = \frac{1}{N}\sum_{j=1}^{N}\langle(n_j(t) - \langle n_j(t)\rangle)[n_j(t+\tau_c) - \langle n_j(t)\rangle]\rangle, \tag{F1}$$

where the angular brackets denote a time average. Here, $C(0)$ depicts the total variance in the fluctuations of the occupation numbers in the reservoir sites, whereas $C(\tau_c)$ with $\tau_c > 0$ provides information about the temporal structure of the reservoir activity. In the right panels of Fig. 10(a), we plot the autocorrelation for one trial of the QR's configuration and data at different f and input scaling W. With no classical input ($W/\gamma = 0.0$), the autocorrelation function decays to the values around zero as τ_c increases. This implies that temporal fluctuations are uncorrelated over large time intervals, which is due to the effect of random quantum inputs and disordered

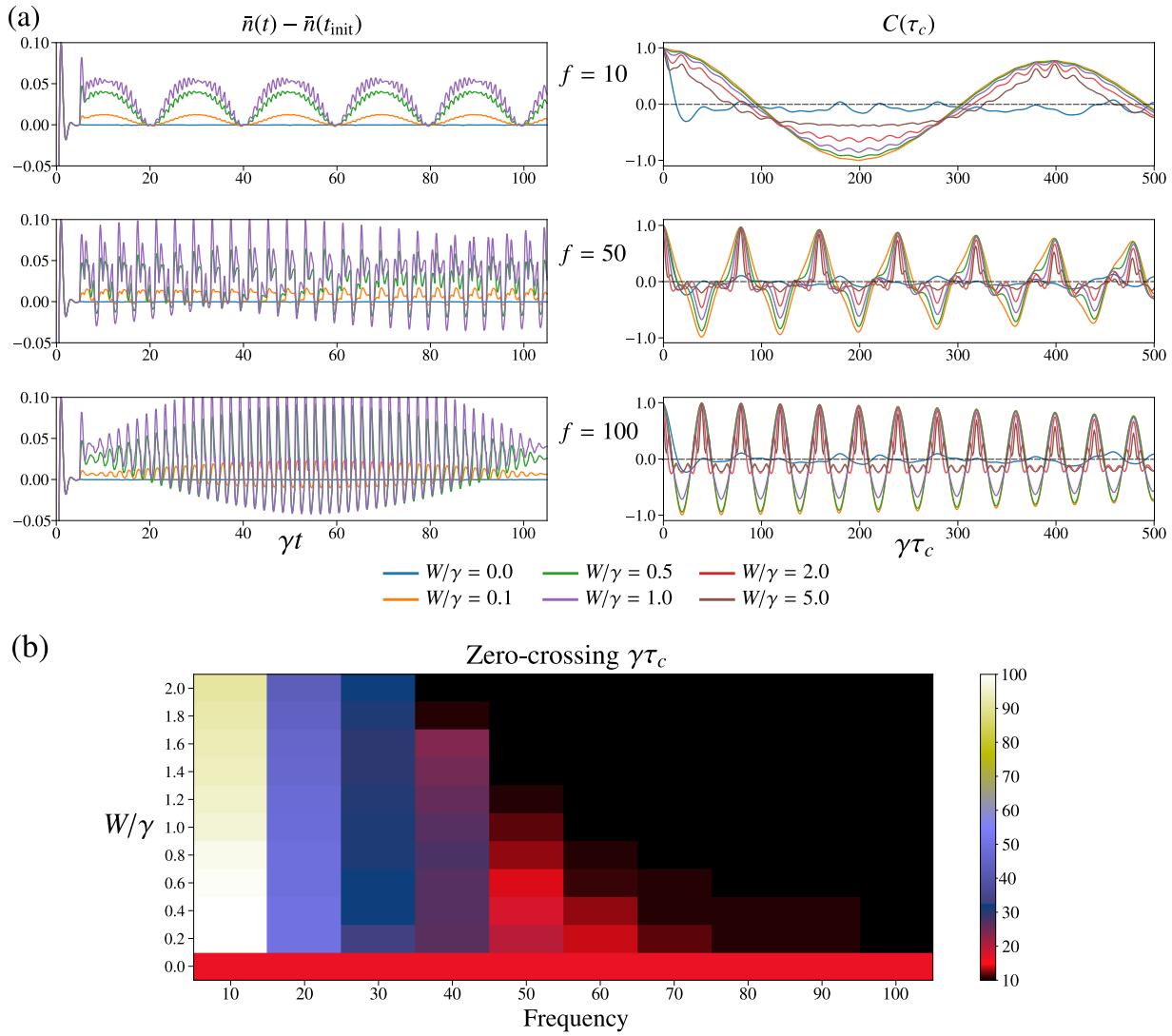

FIG. 10. (a) The dynamics of the average occupation numbers $\bar{n}(t)$ over $N = 3$ reservoir sites (left panel) and the autocorrelation for one trial of the QR's configuration and data (right panel) at different f and input scaling W/γ. (b) The average values of autocorrelation zero-crossing times over ten trials of the QR's configurations and input data for each combination of $W/\gamma \in \{0.0, 0.2, \ldots, 1.8, 2.0\}$ and $f \in \{10, 20, \ldots, 90, 100\}$.

dynamics in the QR. When the QR is driven by sinusoidal classical inputs, we observe that the periodic activity induced by these inputs is superposed on the background of the quantum inputs.

We define the timescale of the QR as the first τ_c such that $C(\tau_c)$ crosses the zero line, which can be understood as the first time interval where the temporal temporal fluctuations are uncorrelated. This zero-crossing time depends on the spontaneous activity of the QR and the timescale of the external classical input. We plot in Fig. 10(b) the average values of zero-crossing times over ten trials of the QR's configurations and input data for each combination of $W/\gamma \in \{0.0, 0.2, \ldots, 1.8, 2.0\}$ and $f \in \{10, 20, \ldots, 90, 100\}$. In the presence of external classical inputs ($W/\gamma > 0.0$), if the zero-crossing times are larger than those of no classical inputs ($W/\gamma = 0.0$), we observe bad performance in reconstructing the classical trajectories (Fig. 9). These results imply that the timescales of the QRs without classical inputs should be larger than the timescales induced by classical inputs. These timescales can be modified by adjusting the constant coherent field P and the input scaling W.

APPENDIX G: QUANTUM MEMORY CAPACITY

In Sec. II B, we present the results of tomography for the quantum switch, which requires the information of previous input signals. The performance of this task depends on the amount of memory from previous inputs that the classical readout can retrieve from the reservoir.

In the conventional RC, we evaluate the STM property of the reservoir via the delay-reconstruction task for reconstructing the previous input. Given a time delay $d \geqslant 0$ and a uniform random input sequence $\{u_n\}$, the target of this task is to produce the output sequence $\{y_n\}$ such that $\{y_n\}$ can approximate the target sequence $\{\hat{y}_n = u_{n-d}\}$. For each delay time step d, the readout part is trained to remember the input sequences at delayed d-time steps. The performance is evaluated by the square of the correlation coefficient $\mathcal{C}(d)$ between

the output and delayed input sequences [14] as follows:

$$\mathcal{C}^2(d) = \frac{\mathrm{cov}^2(\{y_n\}, \{u_{n-d}\})}{\mathrm{var}(\{y_n\})\mathrm{var}(\{u_n\})}. \tag{G1}$$

Here, $\mathrm{cov}(\cdot)$ and $\mathrm{var}(\cdot)$ denote the covariance and variance function, respectively. The STM property represents that this $\mathcal{C}^2(d)$ is sufficiently small at large values of the delay d. $\mathcal{C}^2(d)$ is defined as the *memory function* to characterize the memory profile of the reservoir. Furthermore, the *memory capacity* (MC) of the reservoir is given by

$$\mathrm{MC} = \sum_{d=0}^{\infty} \mathcal{C}^2(d). \tag{G2}$$

In our paper, we consider the concept of quantum memory capacity (QMC) [34] to measure the ability of the QR to reconstruct the previous quantum inputs via the classical readout. We investigate the quantum version of STM in the QR via the quantum version of the delay-reconstruction task $\mathcal{F}(u_n = 0, \beta_n) = \beta_{n-d}$ given the delay d, where classical inputs are zero. Since the input and output are quantum states, the capacity to reconstruct the previous d steps of the input states is evaluated via the square of the distance correlation [54] between the output $\{\sigma_n\}$ (obtained via the training of the classical readout) and the target $\{\hat{\sigma}_n\} = \{\beta_{n-d}\}$:

$$\mathcal{R}^2(d) = \frac{\mathcal{V}^2(\{\sigma_n\}, \{\beta_{n-d}\})}{\sqrt{\mathcal{V}^2(\{\sigma_n\}, \{\sigma_n\})\mathcal{V}^2(\{\beta_n\}, \{\beta_n\})}}. \tag{G3}$$

Here, $\mathcal{V}^2(\{\rho_n\}, \{\sigma_n\})$ represents the squared distance covariance of random sequences of density matrices $\{\rho_n\}$ and $\{\sigma_n\}$. The squared distance covariance $\mathcal{V}^2(\{\rho_n\}, \{\sigma_n\})$ is calculated from all pairwise distances $A(\rho_j, \rho_k)$ and $A(\sigma_j, \sigma_k)$ for $j, k = 1, 2, \ldots, n$, where the distance $A(\rho, \sigma) = \arccos F(\rho, \sigma)$ for given density matrices ρ and σ is defined as the angle induced from the fidelity $F(\rho, \sigma) = \mathrm{Tr}[\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}]$. We construct the distance matrices for $\{\rho_n\}$ and $\{\sigma_n\}$ as (R_{jk}) and (S_{jk}) with the elements $R_{jk} = A(\rho_j, \rho_k)$ and $S_{jk} = A(\sigma_j, \sigma_k)$. We take all double centered distances

$$r_{j,k} = R_{j,k} - \bar{R}_{j\cdot} - \bar{R}_{\cdot k} + \bar{R}_{\cdot\cdot}, \tag{G4}$$

$$s_{j,k} = S_{j,k} - \bar{S}_{j\cdot} - \bar{S}_{\cdot k} + \bar{S}_{\cdot\cdot}, \tag{G5}$$

where $\bar{R}_{j\cdot}$ and $\bar{R}_{\cdot k}$ are the jth row mean and the kth column mean, respectively, and $\bar{R}_{\cdot\cdot}$ is the grand mean of the distance matrix (R_{jk}) (the same notations for S). The squared distance covariance is the arithmetic average

$$\mathcal{V}^2(\{\rho_n\}, \{\sigma_n\}) = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n} r_{j,k}s_{j,k}. \tag{G6}$$

$\mathcal{R}^2(d)$ gives information about the serial dependence between $\{\sigma_n\}$ and $\{\hat{\sigma}_n\} = \{\beta_{n-d}\}$. Here, $0 \leqslant \mathcal{R}^2(d) \leqslant 1$ and $\mathcal{R}^2(d) = 1$ if we can find some linear transformation from the output sequence $\{\sigma_n\}$ to the target sequence $\{\hat{\sigma}_n\}$. In contrast, $\mathcal{R}^2(d) = 0$ implies that the system cannot reconstruct the previous d steps of the inputs because the output and the target sequences are completely independent. We define $\mathcal{R}^2(d)$ as the *quantum memory function* of the QR via tomography learning with the classical readout. Consequently, the QMC is defined as

$$\mathrm{QMC} = \sum_{d=0}^{\infty} \mathcal{R}^2(d). \tag{G7}$$

Figure 11 presents a demonstration for the quantum memory function and quantum memory capacity, where we consider $\{\beta_n\}$ as a random sequence of one-qubit input states with 400 time steps for the training and 100 time steps for the evaluation. Figure 11(a) displays the values of RMSF, and Fig. 11(b) displays the quantum memory function $\mathcal{R}^2(d)$ for $N = 3$-site QR at several values of uniform excitation $P(t) = P$. We perform the tomography task with $V = 5$ measurement multiplexity, which means that the dimension of the reservoir state is $VN = 15$. The STM property depends on the value of P, where the memories at $d < 5$ dominate all the regions and converge to almost the same value at a sufficiently large d. This value is nonzero owing to the effect of the finite data length.

We further plot the dependency of QMC on the coherent field strength P in Fig. 11(c) at $N = 2$, 3-site QR. Here, Eq. (G7) is calculated until the maximum delay $d_{\max} = 40$. We observe an optimal region of P ($2 \leqslant P < 10$), where the QMC is favorable. To explain this behavior, we further analyze the dynamics of the occupation numbers $n_j(t)$ in the reservoir sites. Figure 11(d) plots the absolute difference $|\bar{n}(t) - \bar{n}(t_{\mathrm{init}})|$ between the average element $\bar{n}(t)$ of the reservoir states at initial time $t_{\mathrm{init}} = 5/\gamma$ and an arbitrary t ($0 \leqslant \gamma t \leqslant 15$). This difference approaches zero as t approaches t_{init}. The quantum input is incident to the QR at $\gamma t = 5, 6, 7, \ldots, 14, 15$, which increases the value $|\bar{n}(t) - \bar{n}(t_{\mathrm{init}})|$ before decreasing it until the next input. We anticipate that increasing the magnitude of the coherent field strength P compared with h_{ij} in Eq. (2) may lead to nonergodic behavior in the QR, i.e., a strong and qualitative dependence of expectation values on the initial state at t_{init}. Furthermore, the input state β_l is incident to the QR with weak coupling ($|W^{\mathrm{in}}_{jk}| \ll |P|$) in this case. Therefore, sufficient information regarding β_l cannot be extracted from the QR as $|\bar{n}(t) - \bar{n}(t_{\mathrm{init}})| \approx 0$. In contrast, a small $P(t)$ strongly drives the system from the steady state at the input-injecting timing but reduces the memory effect of the QR in reconstructing past information since the old information is replaced very quickly.

APPENDIX H: EFFECTS OF CLASSICAL INPUT ON THE RECONSTRUCTION OF QUANTUM INPUT

In the main text, we considered the target function, which is a function $\mathcal{F}$ of hybrid inputs (u, β). In this case, information of both the classical input $u(t)$ and quantum input $\beta(t)$ is retained in the reservoir states. Since the classical input $u(t)$ is encoded into the strength of the coherent field $P(t) = P + Wu(t)$, the classical input and input scaling W/γ have a strong effect on the dynamics of the QR. If the target function $\mathcal{F}$ does not depend on the classical input $u(t)$, then the injection of $u(t)$ into the QR may affect the reconstruction of $\mathcal{F}$.

In this section, we verify this observation by investigating the memory capacity. Given a sequence of hybrid inputs $\{u_n, \beta_n\}$, we use our QR in a multitask setting with the

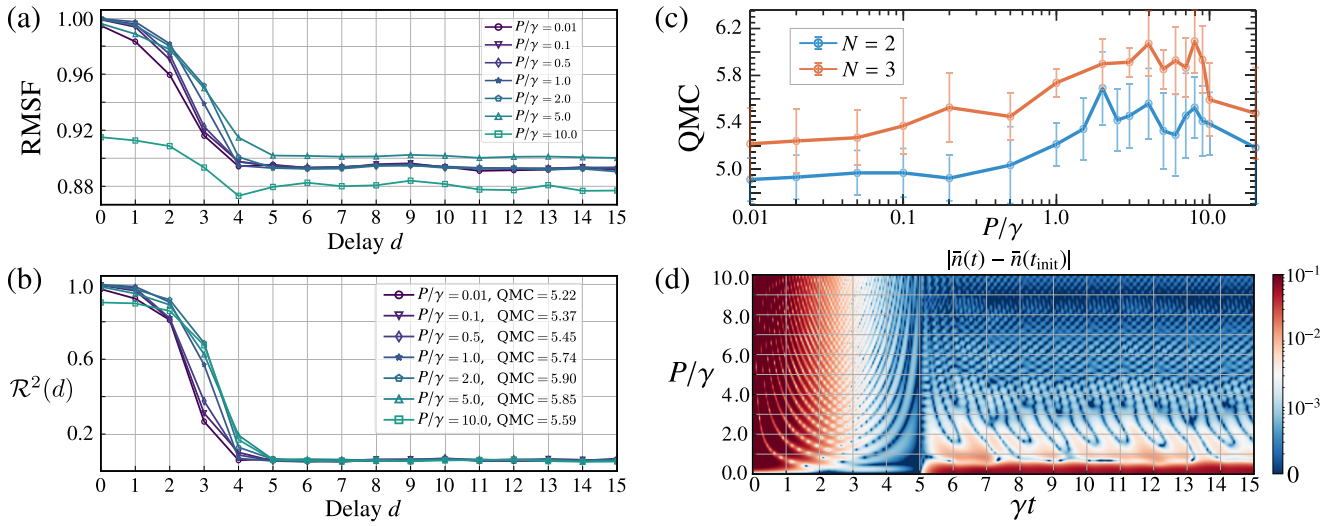

FIG. 11. (a) The RMSF of the STM task varying by the delay d and (b) the quantum memory function $\mathcal{R}^2(d)$ for $N = 3$-site QR at several values of uniform excitation $P(t) = P$. The tomography task is performed with $V = 5$ measurement multiplexity and a random sequence of one-qubit input states with 400 time steps for the training and 100 time steps for the evaluation. (c) The dependency of QMC on the coherent field strength P at $N = 2, 3$-site QR. The QMC is averaged over ten random trials with displayed error bars. (d) The absolute difference $|\bar{n}(t) - \bar{n}(t_{\text{init}})|$ between the average occupation numbers $\bar{n}(t)$ at initial time $t_{\text{init}} = 5/\gamma$ and an arbitrary t $(0 \leqslant \gamma t \leqslant 15)$ varying by P/γ. The quantum input is incident on the QR at $\gamma t = 5, 6, 7, \ldots, 14, 15$, which increases the value $|\bar{n}(t) - \bar{n}(t_{\text{init}})|$ before decreasing it until the next input.

classical and quantum delay-reconstruction tasks mentioned in the previous section. Given a delay d, we consider the delay reconstruction of the classical input $\{u_{n-d}\}$ in the classical task and the delay reconstruction of the quantum input $\{\beta_{n-d}\}$ in the quantum task. We compute the corresponding MC and QMC for the classical and quantum tasks, respectively.

Figure 12 displays the dependency of MC and QMC on the input scaling W/γ for $N = 3, 4$-site QR with a constant coherent field strength $P/\gamma = 1.0$. Here, we consider the random uniform $\{u_n\}$ for classical inputs and $\{\beta_n\}$ as a random sequence of one-qubit states for quantum inputs. To attain the same setting that was used in the task described in Fig. 2 in the main text, we perform the tomography with $V = 8$ measurement multiplexity and use 800 time steps for the training and 200 time steps for the evaluation. We compute the MC and QMC until the maximum delay $d_{\max} = 20$. The result demonstrates that the QMC is reduced when the random classical input is introduced into the QR with increasing input scaling $W/\gamma > 0$. For a relatively large $W/\gamma > 1.0$, both MC and QMC decrease owing to the localization effect with a large strength of the coherent field $P(t) = P + Wu(t)$. However, at $W/\gamma \leqslant 1.0$, we observe a tradeoff relation between MC and QMC. Here, increasing W/γ from zero can help improve the MC but reduce the QMC. This observation implies that this QR may not perform well if the target function does not depend on the classical input, and the fluctuation of classical inputs has a strong effect on the QR dynamics. However, if the target function is the function of the classical and quantum input, we can use the tradeoff of MC and QMC to adjust W/γ for an optimal performance.

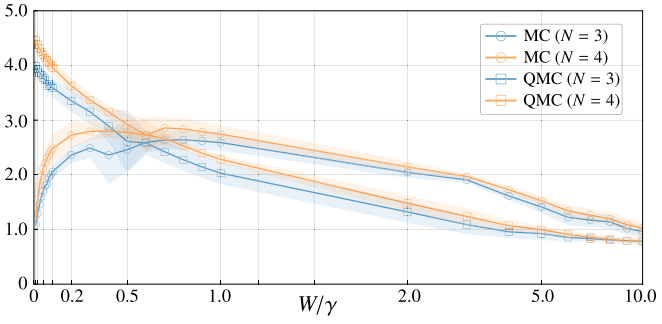

FIG. 12. The dependency of MC and QMC on the input scaling W/γ for $N = 3, 4$-site QR with a constant coherent field strength $P/\gamma = 1.0$. We consider the random uniform $\{u_n\}$ for classical inputs and $\{\beta_n\}$ as a random sequence of one-qubit states for quantum inputs with 800 time steps for the training and 200 time steps for the evaluation. The tomography task is performed with $V = 8$ measurement multiplexity, and MC and QMC are calculated until the maximum delay $d_{\max} = 20$. The solid lines and the shaded areas indicate the median values and the confidence intervals (one standard deviation) calculated in the ensemble of ten random trials of the input sequence and the QR's configuration, respectively.

APPENDIX I: TRAINING THE QUANTUM READOUT

In the classical readout, the training process is simply a linear regression of measurement results to target data, such as the tomography of the quantum state. However, it is more complicated in the quantum readout since the target is the physical quantum state. We can keep the inner parameter fixed and train readout parameters and the interaction between the reservoir and the input. We rely on the fact that any unitary matrix that describes the mixing between optical modes can be implemented with linear optics devices such as phase shifters and beam splitters [41]. Therefore, we can implement the combination of the transmitted fields to

generate M quantum output modes $\hat{C}_m = \sum_j o_{mj}\hat{c}_j$ with complex coefficients o_{mj}. The output modes must satisfy the commutation relations $[\hat{C}_m, \hat{C}_n^\dagger] = \delta_{mn}$, which impose the condition $\sum_j o_{mj}o_{nj}^* = \delta_{mn}$.

Let us consider $\boldsymbol{\theta}$ as the vector of transformed real parameters for the interaction coefficients $\{W_{jk}^{\text{in}}\}$ and readout coefficients $\{o_{mj}\}$. Given L training data with hybrid inputs $\{(u_l, \beta_l)\}$ and target quantum states $\hat{\sigma}_l$, the cost function $\mathcal{L}_{\boldsymbol{\theta}}(\{(u_l, \beta_l), \hat{\sigma}_l\}_{l=1}^L)$ evaluates the difference between the quantum states $\{\sigma_l\}_{l=1}^L$ described via $\{\hat{C}_m\}$ and the target quantum states $\{\hat{\sigma}_l\}_{l=1}^L$. In our numerical simulations, $\mathcal{L}_{\boldsymbol{\theta}}$ is defined via the fidelity error [Eq. (11)] or the error taken in the Wigner representation [Eq. (12)]. Here, $\mathcal{L}_{\boldsymbol{\theta}}$ becomes a nonlinear function of parameters $\boldsymbol{\theta}$. We find the optimal $\boldsymbol{\theta}$ such that $\mathcal{L}_{\boldsymbol{\theta}}$ is minimized. Several methods can be used for this nonlinear optimization problem, and we use the Nelder-Mead simplex algorithm [44], which is fast and effective for problems with a large number of parameters. The algorithm starts from an initial guess for the parameters and generates a simplex in the multidimensional parameter space. In each iteration, the cost function is evaluated at each point in the simplex. Under a selecting and replacing procedure, the points in the simplex with the worst value of the cost function are reconstructed for each step until a convergence condition is satisfied. In our simulation, we use the JULIA framework [40] for simulating the quantum master equation and the built-in function with default parameters in the OPTIM package for the Nelder-Mead algorithm.

[1] H. J. Kimble, The quantum internet, Nature (London) **453**, 1023 (2008).

[2] C. Simon, Towards a global quantum network, Nat. Photon. **11**, 678 (2017).

[3] S. Wehner, D. Elkouss, and R. Hanson, Quantum internet: A vision for the road ahead, Science **362**, eaam9288 (2018).

[4] L. Viola, E. Knill, and S. Lloyd, Dynamical decoupling of open quantum systems, Phys. Rev. Lett. **82**, 2417 (1999).

[5] D. Dong and I. R. Petersen, Quantum control theory and applications: A survey, IET Control Theory Appl. **4**, 2651 (2010).

[6] J. P. G. van Dijk, E. Kawakami, R. N. Schouten, M. Veldhorst, L. M. K. Vandersypen, M. Babaie, E. Charbon, and F. Sebastiano, Impact of classical control electronics on qubit fidelity, Phys. Rev. Appl. **12**, 044054 (2019).

[7] D. Rist, S. Fallek, B. Donovan, and T. A. Ohki, Microwave techniques for quantum computers, IEEE Microwave Mag. **21**, 60 (2020).

[8] G. Chiribella, Perfect discrimination of no-signalling channels via quantum superposition of causal structures, Phys. Rev. A **86**, 040301(R) (2012).

[9] L. M. Procopio, A. Moqanaki, M. Araújo, F. Costa, I. A. Calafell, E. G. Dowd, D. R. Hamel, L. A. Rozema, Č. Brukner, and P. Walther, Experimental superposition of orders of quantum gates, Nat. Commun. **6**, 7913 (2015).

[10] G. Rubino, L. A. Rozema, A. Feix, M. Araújo, J. M. Zeuner, L. M. Procopio, Č. Brukner, and P. Walther, Experimental verification of an indefinite causal order, Sci. Adv. **3**, e1602589 (2017).

[11] K. Goswami, C. Giarmatzi, M. Kewming, F. Costa, C. Branciard, J. Romero, and A. G. White, Indefinite causal order in a quantum switch, Phys. Rev. Lett. **121**, 090503 (2018).

[12] K. Wei, N. Tischler, S.-R. Zhao, Y.-H. Li, J. M. Arrazola, Y. Liu, W. Zhang, H. Li, L. You, Z. Wang, Y.-A. Chen, B. C. Sanders, Q. Zhang, G. J. Pryde, F. Xu, and J.-W. Pan, Experimental quantum switching for exponentially superior quantum communication complexity, Phys. Rev. Lett. **122**, 120504 (2019).

[13] 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 (2001), Vol. 148, p. 13.

[14] H. Jaeger, in *Short Term Memory in Echo State Networks* (GMD-Forschungszentrum Informationstechnik, 2001), Vol. 5, p. 60.

[15] 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).

[16] M. Lukoševičius and H. Jaeger, Reservoir computing approaches to recurrent neural network training, Comput. Sci. Rev. **3**, 127 (2009).

[17] K. Nakajima, Physical reservoir computing—an introductory perspective, Jpn. J. Appl. Phys. **59**, 060501 (2020).

[18] *Reservoir Computing: Theory, Physical Implementations, and Applications*, edited by K. Nakajima and I. Fischer (Springer, New York, 2021).

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

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

[21] Q. H. Tran and K. Nakajima, Higher-order quantum reservoir computing, arXiv:2006.08999.

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

[23] 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).

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

[25] S. Ghosh, A. Opala, M. Matuszewski, T. Paterek, and T. C. H. Liew, Reconstructing quantum states with quantum reservoir networks, IEEE Trans. Neural Networks Learn. Syst. **32**, 3148 (2020).

[26] S. Ghosh, T. Paterek, and T. C. H. Liew, Quantum neuromorphic platform for quantum state preparation, Phys. Rev. Lett. **123**, 260404 (2019).

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

[28] J. Nokkala, R. Martínez-Peña, G. L. Giorgi, V. Parigi, M. C. Soriano, and R. Zambrini, Gaussian states of

continuous-variable quantum systems provide universal and versatile reservoir computing, Commun. Phys. **4**, 53 (2021).

[29] G. Angelatos, S. A. Khan, and H. E. Türeci, Reservoir computing approach to quantum state measurement, Phys. Rev. X **11**, 041062 (2021).

[30] 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).

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

[32] M. Spagnolo, J. Morris, S. Piacentini, M. Antesberger, F. Massa, A. Crespi, F. Ceccarelli, R. Osellame, and P. Walther, Experimental photonic quantum memristor, Nat. Photon. **16**, 318 (2022).

[33] S. Ghosh, K. Nakajima, T. Krisnanda, K. Fujii, and T. C. H. Liew, Quantum neuromorphic computing with reservoir computing networks, Adv. Quantum Technol. **4**, 2100053 (2021).

[34] Q. H. Tran and K. Nakajima, Learning temporal quantum tomography, Phys. Rev. Lett. **127**, 260401 (2021).

[35] J. Nokkala, Online quantum time series processing with random oscillator networks, Sci. Rep. **13**, 7694 (2023).

[36] J. C. Lopez Carreño and F. P. Laussy, Excitation with quantum light. I. Exciting a harmonic oscillator, Phys. Rev. A **94**, 063825 (2016).

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

[38] S. Ghosh, T. Krisnanda, T. Paterek, and T. C. H. Liew, Realising and compressing quantum circuits with quantum reservoir computing, Commun. Phys. **4**, 105 (2021).

[39] C. W. Gardiner, Driving a quantum system with the output field from another driven quantum system, Phys. Rev. Lett. **70**, 2269 (1993).

[40] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, Julia: A fresh approach to numerical computing, SIAM Rev. **59**, 65 (2017).

[41] S. L. Braunstein and P. van Loock, Quantum information with continuous variables, Rev. Mod. Phys. **77**, 513 (2005).

[42] V. J. Mathews and J. Lee, Adaptive algorithms for bilinear filtering, in *Advanced Signal Processing: Algorithms, Architectures, and Implementations V*, edited by F. T. Luk (SPIE, Bellingham, WA, 1994).

[43] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach, Phys. Rev. Lett. **120**, 024102 (2018).

[44] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, Convergence properties of the Nelder–Mead simplex method in low dimensions, SIAM J. Optim. **9**, 112 (1998).

[45] T. J. Elliott, M. Gu, A. J. P. Garner, and J. Thompson, Quantum adaptive agents with efficient long-term memories, Phys. Rev. X **12**, 011007 (2022).

[46] M. Schuld and N. Killoran, Is quantum advantage the right goal for quantum machine learning? PRX Quantum, **3**, 030101 (2022).

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

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

[49] T. Kubota, H. Takahashi, and K. Nakajima, Unifying framework for information processing in stochastically driven dynamical systems, Phys. Rev. Res. **3**, 043135 (2021).

[50] T. Kubota, Y. Suzuki, S. Kobayashi, Q. H. Tran, N. Yamamoto, and K. Nakajima, Temporal information processing induced by quantum noise, Phys. Rev. Res. **5**, 023057 (2023).

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

[52] Y. Chen and X. Ye, Projection onto a simplex, arXiv:1101.6081.

[53] B. I. Bantysh, A. Y. Chernyavskiy, and Y. I. Bogdanov, Quantum tomography benchmarking, Quantum Inf. Process. **20**, 339 (2021).

[54] G. J. Székely, M. L. Rizzo, and N. K. Bakirov, Measuring and testing dependence by correlation of distances, Ann. Stat. **35**, 2769 (2007).