

Quantum Wiener architecture for quantum reservoir computing

Alessio Benavoli ^{*}

*School of Computer Science and Statistics, Trinity College Dublin, and Trinity Quantum Alliance,
Unit 16, Trinity Technology and Enterprise Centre, Pearse Street, Dublin 2, Ireland*

Felix C. Binder [†]

*School of Physics, Trinity College Dublin, and Trinity Quantum Alliance,
Unit 16, Trinity Technology and Enterprise Centre, Pearse Street, Dublin 2, Ireland*



(Received 12 January 2026; revised 30 May 2026; accepted 22 June 2026; published 14 July 2026)

This work focuses on quantum reservoir computing and, in particular, on quantum Wiener architectures (qWiener), consisting of quantum linear dynamic networks with weak continuous measurements and classical nonlinear static readouts. We provide the first rigorous proof that qWiener systems retain the fading-memory property and universality of classical Wiener architectures, despite quantum constraints on linear dynamics and measurement backaction. Furthermore, we develop a kernel-theoretic interpretation showing that qWiener reservoirs naturally induce deep kernels, providing a principled framework for analyzing their expressivity. We further characterize the simplest qWiener instantiation, consisting of concatenated quantum harmonic oscillators, and show the difference with respect to the classical case. Finally, we empirically evaluate the architecture on standard reservoir computing benchmarks, demonstrating systematic performance gains over prior classical and quantum reservoir computing models.

DOI: [10.1103/kt8-l2tw](https://doi.org/10.1103/kt8-l2tw)

I. INTRODUCTION

Reservoir computing (RC) [1,2] is a computational framework in which temporal information is processed by driving a dynamical system with input signals while training only the output layer. Despite this strong restriction, RC architectures such as liquid state machines [1] and echo-state networks [2] are provably universal for approximating fading-memory systems, i.e., dynamical systems depending only on a finite past history of inputs.

Recent efforts have extended RC into the quantum regime [3–12], motivated both by the high dimensionality of quantum systems and by the prospect of exploiting near-term noisy intermediate-scale quantum platforms. Quantum reservoir computing (QRC) aims to process temporal data streams through quantum dynamics and extract classical information through suitable measurements. QRC has achieved promising results in addressing temporal and classification tasks [13–19].

In Ref. [7], it was shown that even linear quantum networks restricted to Gaussian states can, in principle, provide universality for RC.

Remark 1. We recall [20] that the universality of an RC system is defined as its ability to approximate, to arbitrary

accuracy, any time-invariant input-output system with the fading memory property. An input-output system is said to exhibit fading memory if two input signals which are close in the recent past, but not necessarily close in the remote past, yield present outputs which are close. ■

In this paper, we will refer to the above concept of universality as *RC-universality*. An important result about RC universality was established by Boyd and Chua [21], who, using Volterra series [22], showed that a system defined by the interconnection of parallel single-input, single-output linear time-invariant (LTI) dynamical systems whose outputs are fed into a polynomial nonlinearity is universal, see Fig. 1 [23]. We have provided details about this result in Appendix A. This architecture, known as the classical Wiener architecture [24], is depicted in Fig. 1. By leveraging the result in Ref. [21] and appropriately combining the outputs of a finite number of quantum linear networks with a polynomial readout, Ref. [7] claimed that universality can be achieved within QRC framework. Here we refer to this specific subclass as the *quantum Wiener architecture (qWiener)*. It comprises parallel quantum linear dynamical systems with linear input encodings, weak continuous measurements, and a nonlinear readout. It is important to stress that Ref. [7] considers a slightly more general setting than qWiener architectures, as it also includes a (nonlinear) encoding of the input. However, the claim of RC universality in Ref. [7] is stated first for the case of no-input encoding (or, equivalently, linear encoding). This claim is particularly interesting, as continuous-variable quantum optical systems provide efficient platforms for its realization [7,25,26]. Moreover, by employing weak measurements, this architecture mitigates the loss of past input memory [7,27,28] that typically results from the measurement process.

^{*}Contact author: alessio.benavoli@tcd.ie

[†]Contact author: felix.binder@tcd.ie

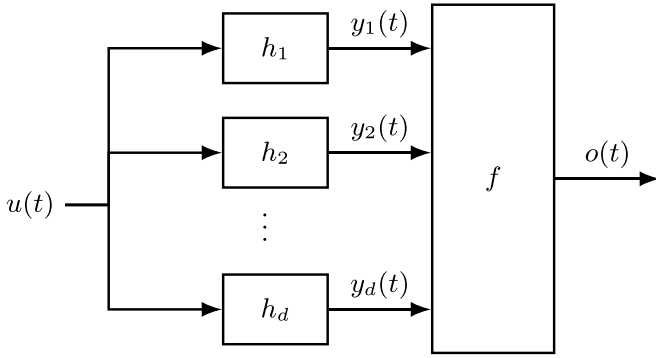


FIG. 1. Wiener architecture: Connection of d parallel linear SISO dynamical systems (h_i) with a memoryless nonlinear function f . $u(t)$, $o(t)$ are the input and output signal, respectively.

However, there are fundamental differences between a classical and a quantum Wiener architecture:

(1) Unlike classical LTI systems, quantum linear networks must preserve the canonical commutation relations, which impose stringent constraints (e.g., unitarity) on their parameters and their dynamics.

(2) Weak measurements also perturb the quantum system, introducing backaction that may compromise both fading-memory and universality properties.

These important aspects were not explicitly addressed in Ref. [7]. In this work, our contributions are as follows.

(1) We rigorously prove the universality of the q Wiener architecture: A system of quantum linear networks with weak continuous measurements, whose outcomes are processed through classical polynomial readouts. We prove that, despite quantum constraints in their dynamics and measurement backaction, q Wiener systems preserve the *RC-universality* approximation capabilities of classical Wiener architectures.

The concept of RC universality [21] and our proof above means that any arbitrary time-invariant input-output system with the fading memory property can be approximated arbitrarily well by a Wiener [21] or q Wiener (our proof) architecture. In practice, in RC Wiener or q Wiener architectures, the reservoir (which is an LTI system in classical Wiener architectures or a quantum linear network in q Wiener architectures) is not chosen arbitrarily: Its parameters are typically generated randomly from specific probability distributions. Different distributions give rise to different reservoir properties, which in turn determine their *expressivity*. We recall that, in statistics and machine learning, the *expressivity* of a model is determined by the space of functions the model can represent, see for instance [29]. In the case of (q)Wiener architectures, this expressivity is thus directly determined by the choice of probability distributions used to generate the reservoir parameters. To characterize expressivity, we do the following:

(1) We develop a kernel-theoretic perspective on classical Wiener architectures for RC, linking their expressive power to kernel methods and Gaussian processes [30]. We show that the reservoir naturally induces kernel functions (covariance functions) capturing correlations between input sequences. This leads to a deep-kernel [31] interpretation of classical Wiener architectures, where the “deep” structure arises from

combining the reservoir layer (the randomized LTI dynamical system) with the readout layer. Similarly, we show that, with some important differences (due to the need of preserving the canonical commutation relations), just as randomized LTI dynamical systems with nonlinear readouts yield deep kernels in classical RC, q Wiener architectures combining randomized quantum linear networks with classical nonlinear readouts also generate deep kernels. This allows us to bridge QRC with modern kernel methods in machine learning, providing an approach to assess their expressivity, that is, through the deep kernel they generate.

(2) Leveraging this kernel-based viewpoint, we analyze the expressive power of the simplest q Wiener instantiation, where the quantum linear block is the concatenation of n_c quantum harmonic oscillators, each linearly coupled to an external field. This allows us to highlight the difference between the classical and quantum case. We further show that the resulting system in this simple instantiation is completely passive and, therefore, implementable using only passive optical components (this case is in contrast to the more general architecture that we use to prove the RC universality of q Wiener which is not completely passive). Moreover, we prove that the weak measurement does not disturb its dynamics but merely adds classical Wiener noise to the output. This result exploits the quantum nature of the proposed QRC architecture, in particular the fact that an observed quantum linear network evolves according to the quantum Kalman filter equations [32].

(3) We evaluate the simplest q Wiener architecture against the model proposed in Ref. [7], as well as a classical echo-state network (used as a baseline), on three benchmark reservoir computing tasks: parity-check, NARMA10, and delay. Our results demonstrate that the proposed approach consistently outperforms the method in Ref. [7]. In particular, we show empirically (consistently with our theoretical derivation of the corresponding deep kernel) that the performance of our model systematically improves as the number of harmonic oscillators in the linear network increases.

(4) Finally, we exploit our universality result for the q Wiener architecture to show that it can be designed to have a prescribed long-term memory capacity, and a computational basis related to the Legendre polynomials, using a result derived in Ref. [33] for classical reservoir computing.

II. BACKGROUND AND NOVEL PRELIMINARIES

In this section, we introduce the notation and present both established and, to the best of our knowledge, novel background results that form the foundation of our main contributions. While extensive, this material is essential, as these results will be directly employed in developing our core findings in the subsequent sections.

A. Classical Wiener architecture and connection to kernel methods

A classical Wiener architecture consists of multiple parallel classical single input–single output (SISO) LTI systems whose outputs are fed into a static nonlinearity. When this architecture is used in RC, only the final (readout) layer is

trained. Consequently, the parameters of the parallel SISO LTI systems must be selected beforehand. There are two main approaches for choosing these parameters: (1) random generation from specific probability distributions and (2) deliberate design to achieve a desired dynamic behavior of the LTI system [33]. In the first approach, the main challenge is to select distributions that yield a system with sufficient *expressivity*. This *expressivity* can be assessed by examining the statistical properties of the nonlinear basis functions obtained through combinations of randomly generated LTI systems. In this paper, we concentrate on this approach, although we will provide an example of the second approach in the numerical experiments section.

A unifying perspective for analyzing the first approach is offered by kernel theory and its Bayesian counterpart, Gaussian processes [30]. We recall the following:

Definition 1 (Gaussian Process (GP)). A GP is a collection of random variables, any finite number of which have a joint Gaussian distribution. Formally, a real-valued function $f(\mathbf{x})$ with $\mathbf{x} \in \mathbb{R}^n$ is said to be GP distributed, denoted as $f \sim \mathcal{GP}(\mu, k)$, if for any finite set of inputs $\mathbf{x}_1, \dots, \mathbf{x}_n$, the corresponding function values $f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)$ follow a multivariate normal distribution:

$$f(\mathbf{x}) \sim \mathcal{N}\left(\begin{pmatrix} \mu_\theta(\mathbf{x}_1) \\ \vdots \\ \mu_\theta(\mathbf{x}_n) \end{pmatrix}, \begin{pmatrix} K_\theta(\mathbf{x}_1, \mathbf{x}_1) & \cdots & K_\theta(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ K_\theta(\mathbf{x}_n, \mathbf{x}_1) & \cdots & K_\theta(\mathbf{x}_n, \mathbf{x}_n) \end{pmatrix}\right),$$

where $\mu_\theta(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$ is the mean function and $K_\theta(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - \mu(\mathbf{x}))(f(\mathbf{x}') - \mu(\mathbf{x}'))]$ is the covariance function (or kernel) for any $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n$. The mean and covariance functions fully define a GP. They usually depend on a vector of hyperparameters $\theta \in \mathbb{R}^{n_p}$. ■

In the paper, we use $\mathcal{N}(\mu, \Sigma)$ to denote the multivariate Gaussian distribution with mean μ and covariance Σ . Generally, the mean function μ_θ of a GP is assumed to be zero, so its covariance function entirely characterizes the properties and behavior of the process.

1. Linear dynamical systems

Now consider a deterministic SISO LTI model:

$$d\mathbf{x}(t) = A\mathbf{x}(t)dt + Bu(t)dt, \quad (1)$$

$$y(t) = C\mathbf{x}(t) + Du(t), \quad (2)$$

where $t \geq 0$ is time, $u(t) \in \mathbb{R}$ is the input signal, $y(t) \in \mathbb{R}$ is the output signal, and $\mathbf{x}(t) \in \mathbb{R}^{n_x}$ is the state. The matrices are $A \in \mathbb{R}^{n_x \times n_x}$, $B \in \mathbb{R}^{n_x \times 1}$, $C \in \mathbb{R}^{1 \times n_x}$, and $D \in \mathbb{R}$. We assume the system is stable, that is the eigenvalues of A have strictly negative real parts. The output of a SISO LTI system can be written as

$$y(t) = \int_0^t h(t - \tau)u(\tau)d\tau, \quad (3)$$

where h is the impulse response [we have assumed that $u(t) = 0$ for $t < 0$] [34]. The impulse response of the above system depends on the matrix parameters as follows: $h(t) = Ce^{At}B + D\delta(t)$, where e^M denotes the *matrix exponential* of M and $\delta(t)$ is the Dirac delta. It is well known that the term $Ce^{At}B$ can be written as the linear combination of decaying sinusoidal

functions [35]:

$$h(t) = \sum_{i=1}^{n_c} e^{-\alpha_i t} (\beta_i \cos(\omega_i t) - \gamma_i \sin(\omega_i t)) + D\delta(t), \quad (4)$$

where $\alpha_i > 0$ and $\omega_i \in \mathbb{R}$ are, respectively, the decay rate and the angular frequency of the i th damped oscillation and $\beta_i, \gamma_i \in \mathbb{R}$ are scaling parameters. The number of components is $n_c = n_x/2$, where we have assumed that $n_x \geq 2$ is an even number. For simplicity, in this section we assume that $D = 0$.

We now consider the case where the parameters $\alpha_i, \omega_i, \beta_i, \gamma_i$ are randomly generated from some specific probability distributions. We report results on dynamic system identification [36], which characterize how specific distributions over these parameters generate, in the limit $n_c \rightarrow \infty$, a kernel function.

Lemma 1 (Ref. [36]). For $i = 1, \dots, n_c$, assume that the parameters α_i, ω_i of the impulse response (4) are jointly sampled from a PDF $\phi_\theta(\alpha, \omega)$ (depending on some hyperparameters θ), and $\beta_i, \gamma_i \sim \mathcal{N}(0, 1/n_c)$ independently. Then, given the parameters $\{\alpha_i, \omega_i\}_{i=1}^{n_c}$, the impulse response $h(t)$ is GP distributed with zero-mean function and covariance function,

$$E[h(t)h(s)] = \frac{1}{n_c} \sum_{i=1}^{n_c} e^{-\alpha_i(t+s)} \cos(\omega_i(t-s)), \quad (5)$$

for all time instants t, s . Then, in the limit as $n_c \rightarrow \infty$, $E[h(t)h(s)]$ tends to

$$K_\theta(t, s) = \int_0^\infty \int_{-\infty}^\infty \phi_\theta(\alpha, \omega) e^{-\alpha(t+s)} \cos(\omega(t-s)) d\alpha d\omega. \quad (6)$$

Moreover, in the limit, $h(t) \sim \mathcal{GP}(0, K_\theta(t, s))$. ■

Note that $\phi_\theta(\alpha, \omega)$ is a generalized power spectral density [that is, a non-negative function satisfying $\phi_\theta(\alpha, \omega) = \phi_\theta(\alpha, -\omega)$], which has been normalized to integrate to 1. We can then interpret it as a PDF and, therefore, the limit in (6) follows by the strong law of large numbers [37]

Proposition 1 (Ref. [36]). Assume that the impulse response $h(t)$ is GP distributed with zero mean function and covariance function (6). Depending on the choice of the probability density function $\phi_\theta(\alpha, \omega)$, from (6) we can derive that

(1) if $\phi_\theta(\alpha, \omega) = \delta(\alpha)\mathcal{N}(\omega; \omega_0, \frac{1}{\ell^2})$ then

$$K_\theta^{(\text{SE})}(t, s) = e^{-\frac{(t-s)^2}{2\ell^2}} \cos(\omega_0(t-s)); \quad (7)$$

(2) if $\phi_\theta(\alpha, \omega) = \frac{\alpha}{\pi(\omega^2 + \alpha^2)} \text{Unif}(\alpha; \frac{\alpha_m}{2}, \frac{\alpha_M}{2})$ then the kernel is

$$K_\theta^{(\text{TC})}(t, s) = \frac{e^{-\alpha_m \max(t,s)} - e^{-\alpha_M \max(t,s)}}{2 \max(t,s)(\alpha_M - \alpha_m)}; \quad (8)$$

where $\delta(\alpha)$ is the Dirac's delta. The vector θ includes the hyperparameters, such as, for instance, $\ell > 0$ (lengthscale), $\omega_0 \in [0, 2\pi)$, and $0 < \alpha_m < \alpha_M$. ■

Observe that, for $\omega_0 = 0$, the first kernel function coincides with the squared-exponential (SE) kernel, which is a standard kernel in machine learning [30]. For $\omega_0 \neq 0$, one obtains one component of a spectral mixture kernel [38], which generalizes the SE kernel. The second kernel is known as the integrated tuned-correlated kernel [39,40], which is specif-

ically used in system identification to estimate the impulse response of linear systems.

Remark 2. We recall that a reproducing kernel Hilbert space (RKHS) [41] is a Hilbert space $\mathcal{H}$ of functions in which point evaluation is a continuous linear functional. Equivalently, a Hilbert space $\mathcal{H}$ of functions on a set $\mathcal{Z}$ is an RKHS if for every $z \in \mathcal{Z}$ there exists an element $K(z, \cdot) \in \mathcal{H}$ (indexed by z) such that, for all $f \in \mathcal{H}$,

$$\langle f, K(z, \cdot) \rangle = f(z). \quad (9)$$

The function $K(z, \cdot)$ is then called the reproducing kernel at z , and it reproduces the value of f at z via the inner product. For any positive definite kernel K [such as the kernels $K_\theta^{(\text{SE})}(t, \cdot)$, $K_\theta^{(\text{TC})}(t, \cdot)$ above, here $z = t$], there exists a unique RKHS of real valued functions defined over $\mathcal{Z}$ whose reproducing kernel is K . From (9), it follows that, in a RKHS, every function is a linear combination of a (possibly infinite) number of kernel sections, that is $f(z) = \sum_i a_i K(z, z_i)$ for some real coefficients a_i . This fact has an important consequence in statistics and machine learning: Instead of specifying a whole set of basis functions, it suffices to choose a single positive definite kernel function that encodes the desired properties of the function to be modelled. Consider, for instance, the kernel $K_\theta^{(\text{TC})}$ in (8) defined on $\mathcal{Z} = [0, T]$ (time interval), then all the functions in the corresponding RKHS are (infinite) sums of damped decaying exponential. When we perform inference in GPs, we are simply Bayesian learning the coefficients a_i in the expression $f(z) = \sum_i a_i K(z, z_i)$ from data $\{(z_i, y_i)\}_{i=1}^n$. ■

Remark 3. From the above remark, it follows that the choice of the distribution $\phi_\theta(\alpha, \omega)$, which determines the kernel via (6), plays a fundamental role in determining the *expressivity* of the randomized LTI system which constitutes the reservoir in a Wiener architecture. For instance, the SE kernel, which is ubiquitous in machine learning, is not a good choice [[42], Sec. 13] if we are interested in learning a stable linear dynamical system (equivalently, we aim to learn its impulse response h), because it does not model the fact that h decays to zero as time increases. Instead, the SE assumes a constant variability $E[h(t)h(t)] = K_\theta^{(\text{SE})}(t, t) = e^{-\frac{(t-t)^2}{2t^2}} = 1$ at each time t . Note that the integrated TC kernel is indeed a much better choice, because $E[h(t)h(t)] = K_\theta^{(\text{TC})}(t, t) \rightarrow 0$ for $t \rightarrow \infty$. Although both the SE and TC kernels can be used to approximate any impulse response h with arbitrary precision (given sufficiently many data points n), the TC kernel provides a better approximation for the same n , see Fig. 7 in Ref. [42]. ■

The previous results have focused on the impulse response h of a stable linear dynamical system. The response to a generic input signal u is given by $y(t) = \int_0^t h(t-\tau)u(\tau)d\tau$. Since we have modelled $h(t) \sim \mathcal{GP}(0, K_\theta(t, s))$, from $y(t) = \int_0^t h(t-\tau)u(\tau)d\tau$, we have that

$$y(t) \sim \mathcal{GP}\left(0, \int_0^t \int_0^{t'} K_\theta(t-\tau_1, t'-\tau_2)u(\tau_1)u(\tau_2)d\tau_1d\tau_2\right), \quad (10)$$

which follows by the fact that the integral is a linear operator and, therefore, preserves Gaussianity.

While in the previous proposition we considered the limiting case $n_c \rightarrow \infty$, in the context of reservoir computing we

focus on the finite case, where Ref. [43]. In this regime, the impulse response can be *physically realized* by a state-space model (i.e., by a physical system), since

$$h(t) = \sum_{i=1}^{n_c} e^{-\alpha_i t} (\beta_i \cos(\omega_i t) - \gamma_i \sin(\omega_i t)) = C e^{At} B,$$

where the matrices A , B , and C include the parameters $\{\alpha_i, \beta_i, \gamma_i, \omega_i\}_{i=1}^{n_c}$.

Proposition 2. Assume the parameters $(\alpha_i, \omega_i)_{i=1}^{n_c}$ are sampled from $\phi_\theta(\alpha, \omega)$. Define

$$h_j(t) = \sum_{i=1}^{n_c} e^{-\alpha_i t} (\beta_i^{(j)} \cos(\omega_i t) - \gamma_i^{(j)} \sin(\omega_i t)), \quad (11)$$

where $\beta_i^{(j)}, \gamma_i^{(j)} \sim \mathcal{N}(0, 1/n_c)$ independently for $i = 1, \dots, n_c$ and $j = 1, \dots, d$. Denote with $y_j(t) = \int_0^t h_j(t-\tau)u(\tau)d\tau$.

Define $\mathbf{y}(t) = \frac{1}{\sqrt{d}}(y_1(t), \dots, y_d(t))^T$, then we have that for a large d and n_c ,

$$\mathbf{y}^T(t)\mathbf{y}(t') \approx \int_0^t \int_0^{t'} K_\theta(t-\tau_1, t'-\tau_2)u(\tau_1)u(\tau_2)d\tau_1d\tau_2, \quad (12)$$

where $K_\theta(\cdot, \cdot)$ is the covariance defined in (6). ■

The proof of this proposition is provided in Appendix D, which also contains the proofs of all results stated without explicit references. Proposition 2, which follows directly from Lemma 1 combined with (10), shows how we can use d parallel LTI systems to produce a vector $\mathbf{y}(t)$, whose squared euclidean norm approximates the covariance function of the GP in (10).

Remark 4. The rate of convergence of the left-hand side of (12) to the right-hand side is $1/\sqrt{r}$ with $r = \min(d, n_c)$, which follows by the properties of Monte Carlo integration [44]. ■

2. Nonlinear dynamical systems

The use of LTI systems to build a reservoir is not sufficient to achieve RC universality. We must also include the last component of the classical Wiener architecture: a memoryless nonlinearity. Assume we use the vector $\mathbf{y}(t)$ as input for the memoryless function f , as shown in Fig. 1. The output is $o(t) = f(y_1(t), \dots, y_d(t))$ at each time instant t . Since we do not know f we put a GP prior on it, that is, $f \sim \mathcal{GP}(0, K_\theta^o(\mathbf{y}(t), \mathbf{y}(t')))$. Note that, contrary to the previous case, the kernel has the vector $\mathbf{y}(t)$ as input and it is memoryless, because the input only depends on the value of $\mathbf{y}(t)$ at time t .

Proposition 3. Assume $f \sim \mathcal{GP}(0, K_\theta^o(\mathbf{y}(t), \mathbf{y}(t')))$ with either a polynomial kernel [[30], Sec. 4.2]:

$$K_\theta^o(\mathbf{y}(t), \mathbf{y}(t')) = \sigma_f^2(\mathbf{y}(t)^T \mathbf{y}(t') + c)^{\text{deg}}, \quad (13)$$

with $\boldsymbol{\vartheta} = (\sigma_f, c)$ denoting the scale and offset hyperparameter and deg the degree or a squared-exponential kernel [[30], Sec. 4.2]:

$$K_\theta^o(\mathbf{y}(t), \mathbf{y}(t')) = \sigma_f^2 \exp\left(-\frac{1}{2\ell^2} \|\mathbf{y}(t) - \mathbf{y}(t')\|^2\right), \quad (14)$$

with $\boldsymbol{\vartheta} = (\sigma_f, \ell)$ denoting the scale and length-scale hyperparameter. Then, for a large d and n_c , we have that, for the

polynomial kernel,

$$K_{\theta}^o(\mathbf{y}(t), \mathbf{y}(t')) \approx \sigma_f^2 \left(\int_0^t \int_0^{t'} K_{\theta}(t - \tau_1, t' - \tau_2) u(\tau_1) u(\tau_2) d\tau_1 d\tau_2 + c \right)^{\text{deg}}, \quad (15)$$

or, for the squared-exponential kernel,

$$\begin{aligned} K_{\theta}^o(\mathbf{y}(t), \mathbf{y}(t')) \approx \sigma_f^2 \exp \Bigg(& -\frac{1}{2\ell^2} \int_0^t \int_0^{t'} K_{\theta}(t - \tau_1, t - \tau_2) u(\tau_1) u(\tau_2) d\tau_1 d\tau_2 \\ & -\frac{1}{2\ell^2} \int_0^{t'} \int_0^{t'} K_{\theta}(t' - \tau_1, t' - \tau_2) u(\tau_1) u(\tau_2) d\tau_1 d\tau_2 \\ & +\frac{2}{2\ell^2} \int_0^t \int_0^{t'} K_{\theta}(t - \tau_1, t' - \tau_2) u(\tau_1) u(\tau_2) d\tau_1 d\tau_2 \Bigg), \end{aligned} \quad (16)$$

where $K_{\theta}(\cdot, \cdot)$ is the covariance defined in (6). ■

The proof provided in Appendix D specializes the construction given in Ref. [45] to a Wiener architecture.

Remark 5. Proposition 3 shows that an input signal $u(t)$, when passed through d stable, SISO LTI state-space models—each differing only in their sampled parameters $\beta_i^{(j)}, \gamma_i^{(j)} \sim \mathcal{N}(0, 1/\sqrt{n_c})$, independently for $i = 1, \dots, n_x$ and $j = 1, \dots, d$ —produces a feature vector $\mathbf{y}(t) = \frac{1}{\sqrt{d}}(y_1(t), \dots, y_d(t))^{\top}$. When this feature vector is used within either a polynomial or a squared-exponential kernel, the resulting covariance function converges to a deep kernel [31] in the limit as $n_c, d \rightarrow \infty$, thereby providing a principled interpretation of reservoir features in terms of reproducing kernel Hilbert spaces. ■

Note that the polynomial kernel is relevant here because, in Boyd and Chua [21] original proof of RC universality, the final layer involves a polynomial nonlinearity. However, the polynomial kernel can also be replaced with other nonlinear kernels that exhibit better extrapolation properties (such as the squared-exponential one), that is, kernels whose induced function spaces generalize more reliably outside the range of the training data, rather than being dominated by rapidly growing polynomial terms.

In summary, the previous derivations provide two equivalent ways to achieve such approximation. In the first approach, we use input-output observations of the system we aim to approximate

$$\mathcal{D}_m = \{u(k\Delta), o(k\Delta)\}_{k=1}^m,$$

where Δ is the sampling interval. Our goal is to learn the input-output function g from noisy measurements o :

$$o(t_i) = g(u_{0:t_i}) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2), \quad t_i = i\Delta,$$

where $u_{0:t_i}$ denote the input signal from time 0 to time t_i . We can estimate the input-output function g of the system using GP regression with the prior model,

$$\begin{aligned} g \sim \mathcal{GP} \Bigg(& 0, \sigma_f^2 \left(\int_0^t \int_0^{t'} K_{\theta, n_c}(\tau_1, \tau_2) u(t - \tau_1) u(t' - \tau_2) \right. \\ & \left. \times d\tau_1 d\tau_2 + c \right)^{\text{deg}} \Bigg), \end{aligned} \quad (17)$$

where we have for instance used a polynomial readout.

The second (asymptotically equivalent for $d, n_c \rightarrow \infty$) approach is through reservoir computing. In this case, instead of explicitly constructing the above kernel, we set up d stable LTI systems with random parameters (defined as in Proposition 2). These physical LTI systems are driven by the input signal $u(t)$ to produce d outputs $\mathbf{y}(t)$. We then use the dataset

$$\mathcal{D}_m = \{\mathbf{y}(k\Delta), o(k\Delta)\}_{k=1}^m,$$

to learn the mapping $o(t_i) = g(u_{0:t_i}) + \varepsilon_i \approx f(\mathbf{y}(t_i)) + \varepsilon_i$, assuming $f \sim \mathcal{GP}(0, \sigma_f^2(\mathbf{y}(t)^{\top} \mathbf{y}(t') + c)^{\text{deg}})$. This latter case is computationally more efficient as we do not need to explicitly build the kernel in (17), which requires the computation of an integral. This integral is instead implicitly computed by the reservoir. Analytical derivations for the computation of the Gaussian posterior of f and the GP predictions are provided in Appendix B.

B. Quantum systems

In the quantum Wiener architecture we propose, classical SISO LTI systems are replaced with quantum linear dynamical systems. Before introducing these systems in the next section, we first establish some preliminary notation. Consider an n -mode quantum system with annihilation and creation operators $\check{a}_j, \check{a}_j^*$ for $j = 1, 2, \dots, n$, obeying bosonic commutation relations $[\check{a}_j, \check{a}_k^*] = \delta_{jk}$, $[\check{a}_j, \check{a}_k] = [\check{a}_j^*, \check{a}_k^*] = 0$, where $[a, b] = ab - ba$ and δ_{jk} is the Kronecker delta. The operators $\check{a}_j, \check{a}_j^*$ can equivalently be described in terms of the real quadrature components (e.g., position $\check{q}_j$ and momentum $\check{p}_j$):

$$\check{a}_j = \frac{\check{q}_j + \iota \check{p}_j}{\sqrt{2}}, \quad \check{a}_j^* = \frac{\check{q}_j - \iota \check{p}_j}{\sqrt{2}}, \quad (18)$$

where ι is the complex unit, with equivalent commutation relations: $[\check{q}_j, \check{p}_k] = \iota \delta_{jk}$, $[\check{q}_j, \check{q}_k] = [\check{p}_j, \check{p}_k] = 0$. We introduce the vector

$$\check{\mathbf{z}}^{(r)} = (\check{q}_1 \ \check{q}_2 \ \cdots \ \check{q}_n \ \check{p}_1 \ \check{p}_2 \ \cdots \ \check{p}_n)^{\top}, \quad (19)$$

where the superscript r denotes the real quadrature form.

We can then rewrite the commutation relations as

$$[\check{z}_i^{(r)}, \check{z}_j^{(r)}] = \iota (J_n^{(r)})_{ij}, \quad (20)$$

where the $2n \times 2n$ is

$$J_n^{(r)} = \begin{pmatrix} 0_{n \times n} & I_n \\ -I_n & 0_{n \times n} \end{pmatrix}. \quad (21)$$

Note that, in quantum mechanics, any linear transformation of the vector $\check{z}^{(r)}$, that is, $\check{z}^{(r)} = S\check{z}^{(r)}$, must preserve the commutation relations. This means that $SJ_n^{(r)S^\top = J_n^{(r)}$, which implies that the matrix S must be *symplectic*. Finally, observe that $(\check{z}^{(r)}\check{z}^{(r)\top})_{ij} = \frac{1}{2}\{\check{z}_i^{(r)}, \check{z}_j^{(r)}\} + \frac{i}{2}(J_n^{(r)})_{ij}$, where $\{\check{z}_i^{(r)}, \check{z}_j^{(r)}\} = \check{z}_i^{(r)}\check{z}_j^{(r)} + \check{z}_j^{(r)}\check{z}_i^{(r)}$ is the anticommutator. We recall that the

$$Q^{(r)} = \begin{pmatrix} \mathbb{E}(\check{q}_1 - \mathbb{E}(\check{q}_1))^2 & \frac{1}{2}\mathbb{E}(\{\check{q}_1 - \mathbb{E}(\check{q}_1), \check{p}_1 - \mathbb{E}(\check{p}_1)\}) \\ \frac{1}{2}\mathbb{E}(\{\check{q}_1 - \mathbb{E}(\check{q}_1), \check{p}_1 - \mathbb{E}(\check{p}_1)\}) & \mathbb{E}(\check{p}_1 - \mathbb{E}(\check{p}_1))^2 \end{pmatrix}.$$

Then the matrix $Q^{(r)} + \frac{i}{2}J_n^{(r)}$ is positive semidefinite if

$$\begin{aligned} \det(Q^{(r)} + \frac{i}{2}J_n^{(r)}) &= \mathbb{E}(\check{q}_1 - \mathbb{E}(\check{q}_1))^2 \mathbb{E}(\check{p}_1 - \mathbb{E}(\check{p}_1))^2 \\ &\quad - (\frac{1}{2}\mathbb{E}(\{\check{q}_1 - \mathbb{E}(\check{q}_1), \check{p}_1 - \mathbb{E}(\check{p}_1)\}))^2 - \frac{1}{4} \geq 0, \end{aligned}$$

which corresponds to the uncertainty principle. This is a first key difference between classical and quantum systems: The Hermitian matrix $Q^{(r)} + \frac{i}{2}J_n^{(r)}$ must be positive semidefinite; the positive semidefiniteness of $Q^{(r)}$ alone is not sufficient.

1. Linear quantum dynamical systems

Now we will focus on linear quantum dynamical systems. The Hamiltonian of a quantum linear system is at most quadratic in terms:

$$H = \frac{1}{2}\check{z}^{(c)\dagger} \begin{pmatrix} M_1 & M_2 \\ M_2^* & M_1^* \end{pmatrix} \check{z}^{(c)} = \frac{1}{2}\check{z}^{(c)\dagger} M \check{z}^{(c)},$$

where $\check{z}^{(c)} = (\check{a}_1 \ \check{a}_2 \ \cdots \ \check{a}_n \ \check{a}_1^* \ \check{a}_2^* \ \cdots \ \check{a}_n^*)^\top$, which specifies the dynamics of the harmonic oscillators, with $M_1, M_2 \in \mathbb{C}^{n \times n}$ and $M = M^\dagger$. We assume that the harmonic oscillators are coupled to s distinct external quantum boson fields. The coupling between the harmonic oscillators and the input bosonic quantum fields is linear, taking the form $(N_1 \ N_2)\check{z}^{(c)}$, $N_1, N_2 \in \mathbb{C}^{s \times n}$. Prior to interacting with the quantum system, the input fields may be processed by static devices, such as beamsplitters or phase shifters. This is modeled by a unitary scattering matrix $S \in \mathbb{C}^{s \times s}$. In the Heisenberg picture, the joint dynamics of the harmonic oscillators and the quantum fields are described by a system of *quantum stochastic differential equations* (QSDEs) [47]. These can be equivalently represented as a *quantum state space* (QSS) model, expressed in the real-quadrature coordinates $\check{z}^{(r)} = (\check{q}_1, \dots, \check{q}_n, \check{p}_1, \dots, \check{p}_n)^\top$ as [48]

$$d\check{z}^{(r)}(t) = A^{(r)}\check{z}^{(r)}(t)dt + L^{(r)}(u(t)dt + d\check{w}^{(r)}(t)), \quad (22)$$

$$d\check{y}^{(r)}(t) = C^{(r)}\check{z}^{(r)}(t)dt + D^{(r)}(u(t)dt + d\check{w}^{(r)}(t)), \quad (23)$$

expectation value of any observable $\check{o}$ with respect to a density operator $\check{\rho}$ is defined as $\mathbb{E}(\check{o}) = \text{Tr}(\check{\rho}\check{o})$. We can then define the expectation of the vector of observables $\check{z}^{(r)}$ as $\mathbb{E}(\check{z}^{(r)}) = \text{Tr}(\check{\rho}\check{z}^{(r)})$, applied element-wise to the elements of $\check{z}^{(r)}$. Similarly, the covariance matrix is defined element-wise as $\mathbb{E}(\check{z}^{(r)} - \mathbb{E}(\check{z}^{(r)}))(\check{z}^{(r)} - \mathbb{E}(\check{z}^{(r)}))^\top = \frac{1}{2}\text{Tr}(\check{\rho}(\check{z}^{(r)} - \mathbb{E}(\check{z}^{(r)}))(\check{z}^{(r)} - \mathbb{E}(\check{z}^{(r)}))) + \frac{i}{2}\text{Tr}(\check{\rho}J_n^{(r)}) = Q^{(r)} + \frac{i}{2}J_n^{(r)}$. Therefore, this implies that a real symmetric positive-definite $2n \times 2n$ matrix $Q^{(r)}$ is a bona fide (quantum) covariance matrix if and only if the Hermitian matrix $Q^{(r)} + \frac{i}{2}J_n^{(r)}$ is positive semidefinite [46]. For instance, for $n = 1$, we have that

where

$$D^{(r)} = \begin{pmatrix} \text{Re}(S) & -\text{Im}(S) \\ \text{Im}(S) & \text{Re}(S) \end{pmatrix}, \quad (24)$$

$$C^{(r)} = \begin{pmatrix} \text{Re}(N_1 + N_2) & \text{Im}(-N_1 + N_2) \\ \text{Im}(N_1 + N_2) & \text{Re}(N_1 - N_2) \end{pmatrix}, \quad (25)$$

$$L^{(r)} = -C^{(r)\sharp}D^{(r)}, \quad (26)$$

$$\begin{aligned} A^{(r)} &= \begin{pmatrix} \text{Im}(M_1 + M_2) & \text{Re}(M_1 - M_2) \\ -\text{Re}(M_1 + M_2) & -\text{Im}(-M_1 + M_2) \end{pmatrix} \\ &\quad - \frac{1}{2}C^{(r)\sharp}C^{(r)}, \end{aligned} \quad (27)$$

where Re denotes the real-part, Im is the imaginary part, the $\sharp$ -adjoint of a matrix $X \in \mathbb{R}^{2s \times 2n}$ is defined as $X^\sharp = -J_n^{(r)}X^\top J_s^{(r)}$ with $J_n^{(r)}$ defined in (21), $u(t)dt + d\check{w}^{(r)}(t)$ is the vector of the s external input fields with $\check{w}^{(r)}(t)$ being a quantum Wiener process satisfying the Itô relationship, $d\check{w}^{(r)}(t)d\check{w}^{(r)\top}(t) = Q^{(r)} + \frac{i}{2}J_s^{(r)}$, and $\check{y}^{(r)}(t)$ is a vector of the output fields produced by the interaction of quantum harmonic oscillators and the external input fields. It can be proved that the system matrices $A^{(r)}, L^{(r)}, C^{(r)}$ satisfy the following physical realizability conditions:

$$A^{(r)}J_n^{(r)} + J_n^{(r)}A^{(r)\top} + L^{(r)}J_s^{(r)}L^{(r)\top} = 0, \quad (28)$$

$$J_n^{(r)}L^{(r)} + C^{(r)\top}J_s^{(r)}D^{(r)} = 0, \quad (29)$$

$$D^{(r)}J_s^{(r)}D^{(r)\top} - J_s^{(r)} = 0. \quad (30)$$

Contrarily, we can also prove that for any system matrices $A^{(r)}, L^{(r)}, C^{(r)}, D^{(r)}$ satisfying the physical realizability conditions, gives rise to a valid QSDE. That is, the QSS can be physically realized by a genuine quantum-mechanical system. This also ensures that the fundamental commutation relations are preserved during temporal evolution [for instance, the commutation relation $\check{z}^{(r)}(t)\check{z}^{(r)\top}(t) - (\check{z}^{(r)}(t)\check{z}^{(r)\top}(t))^\top = \check{z}^{(r)}(t_0)\check{z}^{(r)\top}(t_0) - (\check{z}^{(r)}(t_0)\check{z}^{(r)\top}(t_0))^\top$ will hold for each $t \geq t_0$, see Ref. [48] for the other preserved commutation relations].

Remark 6. This is a key difference between classical linear dynamical systems and quantum ones. In the latter, the system matrices $A^{(r)}, L^{(r)}, C^{(r)}, D^{(r)}$ cannot be chosen arbitrarily,

which consequently imposes a limit on universality and must be taken into account when designing QRC. ■

Remark 7. In (22)–(23), we could alternatively include the input $\mathbf{u}(t)$ as a classical signal, for example a photocurrent. However, we do not do so here because we want to consider the more general case in which the input is quantum, for instance, the amplitude of an input laser. This ensures that the quantum linear network is fully quantum, while only the readout layer will be classical in our qWiener architecture. ■

Finally, we recall the following:

Definition 2. [47,48] A QSS is said to be completely passive if $M_2 = 0$, $N_2 = 0$, $\text{Re}(M_1) = \text{Re}(M_1)^\top$, and $\text{Im}(M_1) = -\text{Im}(M_1)^\top$. ■

In quantum reservoir computing, understanding the realizability of quantum linear systems is crucial. A *completely passive* system has a quadratic Hamiltonian and linear couplings, generating no energy and requiring no active elements and can be implemented using only passive optical components.

2. Quantum Kalman filter

The temporal evolution of an open quantum system under continuous (weak) measurement is described by the quantum Kalman filter. By denoting the density operator of the measured system with $\check{\rho}_m(t)$ and redefining $\pi_t(\check{\mathbf{z}}^{(r)}) = \text{Tr}(\check{\rho}_m(t)\check{\mathbf{z}}^{(r)})$. Note that, before measurement, $\check{\rho}(0) = \check{\rho}_m(0)$. We consider a homodyne detection of the first quadrature. We can rewrite the stochastic evolution of the observed system (22) and (23) as the quantum Kalman filtering equation [48]:

$$d\pi_t(\check{\mathbf{z}}^{(r)}) = A^{(r)}\pi_t(\check{\mathbf{z}}^{(r)})dt + L^{(r)}\mathbf{u}(t)dt + G_t d\mathbf{v}_q(t), \quad (31)$$

$$d\mathbf{y}_{qm}(t) = C_q^{(r)}\pi_t(\check{\mathbf{z}}^{(r)})dt + D_q^{(r)}\mathbf{u}(t)dt + (D_q^{(r)}D_q^{(r)\top})d\mathbf{v}_q(t), \quad (32)$$

where $\mathbf{y}_{qm}(t)$ is the measurement output vector of the first quadratures of the output field obtained by the homodyne detector, $C_q^{(r)} = (I_s \ 0_s)C^{(r)}$, $D_q^{(r)} = D^{(r)}(I_s \ 0_s)^\top$, $G_t = V_t C_q^{(r)\top} + L^{(r)}Q^{(r)}D_q^{(r)\top}$, and the covariance matrix V_t satisfies the Riccati differential equation:

$$\dot{V}_t = A^{(r)}V_t + V_t A^{(r)\top} + L^{(r)}Q^{(r)}L^{(r)\top} - G_t (D_q^{(r)}D_q^{(r)\top})^{-1} G_t^\top, \quad (33)$$

and

$$\mathbf{v}_q(t) = (D_q^{(r)}D_q^{(r)\top})^{-1} \left(\mathbf{y}_{qm}(t) - \int_0^t (C_q^{(r)}\pi_\tau(\check{\mathbf{z}}^{(r)}) + D_q^{(r)}\mathbf{u}(\tau))d\tau \right)$$

is the so-called *innovation process* of the quantum Kalman filter. Note that $\mathbf{v}_q(t)$ is a classical standard Wiener process, $\mathbb{E}[\mathbf{v}_q(t)\mathbf{v}_q(t')^\top] = \min\{t, t'\}I_{2s}$.

We recall the following definitions. For the system (1) and (2), the matrix pair (A, B) is said to be *stabilizable* if there exists some matrix R such that $A - BR$ has all eigenvalues with strictly negative real parts. The matrix pair (A, C) is said to be *detectable* if there exists some matrix R such that $A - RC$ has all eigenvalues with strictly negative real parts.

Given (31) and (32), if the pair $(A^{(r)}, C_q^{(r)})$ is detectable and $(A^{(r)}, L^{(r)}Q^{1/2(r)})$ is stabilizable, then the quantum Kalman filter converges to a unique steady state [[49], Sec. 8.5], whose covariance V_∞ is given by the solution of

$$0 = A^{(r)}V_\infty + V_\infty A^{(r)\top} + L^{(r)}\tilde{Q}^{(r)}L^{(r)\top} - G_\infty (D_q^{(r)}D_q^{(r)\top})^{-1} G_\infty^\top. \quad (34)$$

Remark 8. (1) It is important to notice that, in (31) and (32), the impulse response between the input $\mathbf{u}(t)$ and the output $\mathbf{y}_{qm}$ is simply the restriction to the first quadrature of the impulse response of the system (22) and (23). Therefore, by filtering out the contribution of the noise $\mathbf{v}_q$ into $\mathbf{y}_{qm}$ (induced by the continuous weak measurement), we reduce to the deterministic part of the system (22) and (23).

(2) It is important to notice the particular form of the Kalman gain $G_\infty = V_\infty C_q^{(r)\top} + L^{(r)}Q^{(r)}D_q^{(r)\top}$, where the second term is due to the common noise in the two equations in (22) and (23) (see for instance Sec. 8.4 in Ref. [49]). This means that the Kalman gain can be zero even when the term $V_\infty C_q^{(r)\top}$ is not zero. ■

We illustrate the second remark above by proving an important case in which $G_\infty = 0$.

Theorem 1. Assume $Q^{(r)} = \frac{1}{2}I_{2s}$, which is the covariance of the vacuum, $D^{(r)} = I_{2s}$, $n = s$, the matrix $A^{(r)}$ has eigenvalues with strictly negative real parts, the pair $(A^{(r)}, C_q^{(r)})$ is detectable and $(A^{(r)}, L^{(r)}Q^{1/2(r)})$ is stabilizable and assume that the QSS (22) and (23) is completely passive. Then the asymptotic Kalman gain $G_\infty = 0$. ■

This case is important because it implies that, for large t , there is no noise in the dynamic equation (31), since $G_t \approx G_\infty = 0$. In other words, the weak measurement does not disturb the dynamics of the QSS. We will exploit this result in the qWiener architecture in the next section.

III. MAIN RESULTS

We derive our QRC framework based on a quantum Wiener architecture. This consists of a layer of d QSS models, each continuously monitored via homodyne detection. The d measured outputs are then fed into a classical GP model that constitutes the final memoryless nonlinear layer of the Wiener architecture.

First, we consider a network of d one-dimensional quantum harmonic oscillators connected together via *concatenation product* [50]. This will allow us to understand the difference between classical and quantum linear dynamical systems in terms of impulse response.

Second, we will show that we can build a quantum linear dynamical system with any predefined impulse response, proving their universality.

A. One-dimensional quantum harmonic oscillator

Consider the generic QSS model in (22) and (23). Assume that $n = 1$, i.e., a single degree of freedom, so that the state reduces to $\check{\mathbf{z}}^{(r)} = (\check{q}_1, \check{p}_1)^\top$. We can then prove the following result.

Lemma 2. The state space matrices of a $n = 1$ degree-of-freedom linear passive QSS with s inputs are

$$A^{(r)} = \begin{pmatrix} -\lambda & \text{Re}(M_1) \\ -\text{Re}(M_1) & -\lambda \end{pmatrix}, \quad (35)$$

$$C^{(r)} = \begin{pmatrix} \text{Re}(N_1) & -\text{Im}(N_1) \\ \text{Im}(N_1) & \text{Re}(N_1) \end{pmatrix}, \quad (36)$$

$$L^{(r)} = -\begin{pmatrix} \text{Re}(N_1)^\top & \text{Im}(N_1)^\top \\ -\text{Im}(N_1)^\top & \text{Re}(N_1)^\top \end{pmatrix} \begin{pmatrix} \text{Re}(S) & -\text{Im}(S) \\ \text{Im}(S) & \text{Re}(S) \end{pmatrix}, \quad (37)$$

$$D^{(r)} = \begin{pmatrix} \text{Re}(S) & -\text{Im}(S) \\ \text{Im}(S) & \text{Re}(S) \end{pmatrix}, \quad (38)$$

where $\lambda = \frac{1}{2}(\text{Re}(N_1)^\top \text{Re}(N_1) + \text{Im}(N_1)^\top \text{Im}(N_1))$ and the impulse response is

$$\mathbf{h}(t) = \left(e^{-\lambda t} \begin{pmatrix} -\cos(\omega t) & -\sin(\omega t) \\ \sin(\omega t) & -\cos(\omega t) \end{pmatrix} \otimes \Lambda \right) \quad (39)$$

$$+ I_{2s} \delta(t) \begin{pmatrix} \text{Re}(S) & -\text{Im}(S) \\ \text{Im}(S) & \text{Re}(S) \end{pmatrix}, \quad (40)$$

with $\Lambda = \text{Re}(N_1)\text{Re}(N_1)^\top + \text{Im}(N_1)\text{Im}(N_1)^\top$, $\omega = \text{Re}(M_1)$, and $\text{Re}(S)^2 + \text{Im}(S)^2 = 1$. ■

The impulse response is a matrix $\mathbf{h}(t) \in \mathbb{R}^{2s \times 2s}$ as the dimension of the input and output is $2s$. We then specialize the result to the case where the dimension of the input quantum field is $s = 1$.

Corollary 1. Assume that the dimension $s = 1$, $\text{Re}(M_1) = -\omega$, $\text{Re}(N_1) = \alpha$, $\text{Im}(N_1) = \beta$, and $S = s_1 + \iota s_2$, with $s_1^2 + s_2^2 = 1$, where α , β , ω , s_1 , and $s_2 \in \mathbb{R}$, and then

$$A^{(r)} = \begin{pmatrix} -\frac{\alpha^2 + \beta^2}{2} & -\omega \\ \omega & -\frac{\alpha^2 + \beta^2}{2} \end{pmatrix}, \quad (41)$$

$$L^{(r)} = \begin{pmatrix} -\alpha & \beta \\ -\beta & -\alpha \end{pmatrix} \begin{pmatrix} s_1 & -s_2 \\ s_2 & s_1 \end{pmatrix}, \quad (42)$$

$$C^{(r)} = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}, \quad (43)$$

$$D^{(r)} = \begin{pmatrix} s_1 & -s_2 \\ s_2 & s_1 \end{pmatrix}, \quad (44)$$

and

$$\mathbf{h}(t) = \left((\alpha^2 + \beta^2) e^{-\frac{1}{2}t(\alpha^2 + \beta^2)} \begin{pmatrix} -\cos(\omega t) & \sin(\omega t) \\ \sin(\omega t) & -\cos(\omega t) \end{pmatrix} + I_2 \delta(t) \right) \begin{pmatrix} s_1 & -s_2 \\ s_2 & s_1 \end{pmatrix}. \quad (45)$$

This is an impulse response for a system with a two-dimensional input $\mathbf{u}(t)$ and output $\mathbf{y}(t)$. Note that, without loss of generality, we can choose $\beta = 0$.

To realize our quantum Wiener architecture, composed of n_c one-dimensional harmonic oscillators, we construct the system by concatenating the harmonic oscillators with state matrices defined in (41)–(44), using the concatenation product [50]. The concatenation product simply assembles the different components together, without making any connections between them. Consider n_c harmonic oscillators with parameters $\{\alpha_i, \omega_i, s_{i1}, s_{i2}\}_{i=1}^{n_c}$, their concatenation product is obtained

by setting $M_1 = \text{diag}(\omega_1, \dots, \omega_{n_c})$, $N_1 = \text{diag}(\alpha_1, \dots, \alpha_{n_c})$, and $S = \text{diag}(s_{11} + \iota s_{12}, \dots, s_{n_c1} + \iota s_{n_c2})$.

Proposition 4. Consider the QSS obtained by the concatenation product of n_c harmonic oscillators with parameters $\{\alpha_i, \omega_i, s_{i1}, s_{i2}\}_{i=1}^{n_c}$. Assume that we measure the first quadratures $\tilde{q}_1, \dots, \tilde{q}_{n_c}$ of the state and then sum the measured outputs. Denote this sum by $\tilde{y}_{qm}(t)$. Moreover, assume that the $2n_c$ -dimensional input $\mathbf{u}(t) = (\mathbf{1}_{n_c} \quad \mathbf{0}_{n_c})^\top u(t)$ for some scalar input $u(t)$, and then the SISO input-output impulse response [between $u(t)$ and $\tilde{y}_{qm}(t)$] is

$$\tilde{h}(t) = \sum_{i=1}^{n_c} \alpha_i^2 e^{-\frac{\alpha_i^2 t}{2}} (-s_{i1} \cos(\omega_i t) + s_{i2} \sin(\omega_i t)) + s_{i1} \delta(t). \quad (46)$$

The feedthrough term $s_{i1} \delta(t)$ can be disregarded, since the input $u(t)$ can always be fed directly into or removed from the input of the readout layer.

Remark 9. By comparing (46) with (4), we can notice two major differences in the quantum case: (i) the multiplying scalar α_i^2 and (ii) the variables s_{i1}, s_{i2} are constrained to have unit norm. These differences follow from the physical realizability constraints (28)–(30). These constraints limit the expressivity of quantum linear networks for reservoir computing, as their parameters cannot be freely chosen. ■

The quantum system in Proposition 4 is completely passive and, therefore, it can be realized using only passive optical components like phase shifters, beam splitters, and mirrors. Since it is completely passive, from Theorem 1 it follows that the dynamics of the measured system is deterministic and the noise only affects the measurement equation (32), which implies that

$$\tilde{y}_{qm}(t) = \int_0^t \tilde{h}(t - \tau) u(\tau) d\tau + \mathbf{1}_{n_c}^\top D^{(r)} D^{(r)\top} \mathbf{1}_{n_c} \mathbf{v}(t), \quad (47)$$

where $\tilde{h}(t)$ is defined (46). This is important because it implies that the weak measurement does not disturb the dynamics of the QSS. Filtering out the static noise in (47) is quite straightforward.

1. Limit kernel

The results in Sec. II A exploit the Gaussianity and independence of the coefficients $\{\beta_i, \gamma_i\}_{i=1}^{n_c}$ (given $\{\alpha_i, \omega_i\}_{i=1}^{n_c}$) to show that $\tilde{h}(t)$ is GP distributed.

However, note that, due to the unit-norm constraint $s_{i1}^2 + s_{i2}^2 = 1$, the coefficients $\{s_{i1}, s_{i2}\}_{i=1}^{n_c}$ in (46) cannot be Gaussian distributed. For instance, if we assume that s_{i1}, s_{i2} are independently Gaussian distributed with variance $1/\kappa$ with $\kappa > 0$, then conditioned on $s_{i1}^2 + s_{i2}^2 = 1$, their distribution is von Mises $(s_{i1}, s_{i2})^\top \sim \text{VM}(\frac{1}{\sqrt{2}}, \kappa I_2)$ (see Appendix C for details about the von Mises distribution).

We will now consider another way to derive results similar to Lemma 1 and Proposition 1 in Sec. II A. By exploiting the central limit theorem (similarly to what was done in Refs. [51, 52] for neural networks), we will show that the corresponding stochastic process converges to a GP in the limit as $n_c \rightarrow \infty$.

Theorem 2. Assume that the parameters α_i^2, ω_i are sampled from $\phi_\theta(\alpha^2, \omega)$ independently for $i = 1, \dots, n_c$ and

$(s_{i1}^{(j)}, s_{i2}^{(j)})^\top \sim \text{VM}(\frac{1}{\sqrt{2}}, \kappa I_2)$ independently for $i = 1, \dots, n_c$ and $j = 1, \dots, d$. Define

$$\tilde{h}_j(t) = \sum_{i=1}^{n_c} \alpha_i^2 e^{-\frac{\alpha_i^2 t}{2}} \left(-(s_{i1}^{(j)} - m_1) \cos(\omega_i t) + (s_{i2}^{(j)} - m_2) \sin(\omega_i t) \right), \quad (48)$$

where $m_n = E[s_n] = r_1/\sqrt{2}$ for $n = 1, 2$. Since $\tilde{h}_j(t)$ is the sum of zero-mean independent random variables [that is, the terms $\alpha_i^2 e^{-\frac{\alpha_i^2 t}{2}} (-(s_{i1}^{(j)} - m_1) \cos(\omega_i t) + (s_{i2}^{(j)} - m_2) \sin(\omega_i t))$ for $i = 1, \dots, n_c$], from the central limit theorem, it follows that, as $n_c \rightarrow \infty$,

$$\sqrt{n_c} \left(\frac{\tilde{h}_j(t)}{n_c} \right) \xrightarrow{\text{dist.}} \mathcal{GP}(0, K_\theta(t, t')), \quad (49)$$

where

$$K_\theta(t, t') = \frac{1+r_2-2r_1^2}{2} \int_{-\infty}^{\infty} \int_0^{\infty} \alpha^4 e^{-\frac{\alpha^2(t+t')}{2}} \cos(\omega(t-t')) \times \phi_\theta(\alpha^2, \omega) d\alpha^2 d\omega, \quad (50)$$

$r_n = \mathcal{I}_n(\kappa)/\mathcal{I}_0(\kappa)$ with $\mathcal{I}_n(\kappa)$ being the modified Bessel function of the first kind of order n . ■

Theorem 2 shows that, for large n_c , each j th term is $\frac{\tilde{h}_j(t)}{\sqrt{n_c}} \sim \mathcal{GP}(0, K_\theta(t, t'))$. This is the equivalent of Lemma 1 for quantum harmonic oscillators.

Now consider the deterministic part of (47), that is, $\tilde{y}_j(t) = \int_0^t \tilde{h}_j(t-\tau) u(\tau) d\tau$, and observe that, for large d ,

$$\begin{aligned} \bar{y}_j(t) &= \tilde{y}_j(t) - \frac{1}{d} \sum_{j=1}^d \tilde{y}_j(t) \\ &\approx \int_0^t \sum_{i=1}^{n_c} \alpha_i^2 e^{-\frac{\alpha_i^2(t-\tau)}{2}} \left((s_{i1}^{(j)} - m_1) \cos(\omega_i(t-\tau)) \right. \\ &\quad \left. - (s_{i2}^{(j)} - m_2) \sin(\omega_i(t-\tau)) \right) u(\tau) d\tau, \end{aligned}$$

where $m_1 = E[s_1] = r_1/\sqrt{2}$ and $m_2 = E[s_2] = r_1/\sqrt{2}$ (see Appendix C for details about the von Mises distribution).

Proposition 5. Define $\bar{\mathbf{y}}(t) = \frac{1}{\sqrt{dn_c}} (\bar{y}_1(t), \dots, \bar{y}_d(t))^\top$, and then we have that for large d and n_c ,

$$\bar{\mathbf{y}}^\top(t) \bar{\mathbf{y}}(t') \approx \int_0^t \int_0^{t'} K_\theta(t-\tau_1, t'-\tau_2) u(\tau_1) u(\tau_2) d\tau_1 d\tau_2, \quad (51)$$

where $K_\theta(\cdot, \cdot)$ is the covariance defined in (50). ■

This result is the analog of Proposition 2. The rate of convergence of the left-hand side of (51) to the right-hand side is $1/\sqrt{r}$ with $r = \min(d, n_c)$, which follows by the properties of Monte Carlo integration.

Corollary 2. Assuming that $\phi_\theta(\alpha^2, \omega) = \frac{\frac{\alpha^2}{2}}{\pi(\omega^2 + \frac{\alpha^4}{4})}$ Unif($\alpha^2; \alpha_m, \alpha_M$), then we have

$$K_\theta(t, t') \propto \frac{(1 + a_m \tau + \frac{a_m^2 \tau^2}{2}) e^{-a_m \tau} - e^{-a_M \tau} (1 + a_M \tau + \frac{a_M^2 \tau^2}{2})}{2\tau^3 (a_M - a_m)}, \quad (52)$$

where $\tau = \max(t, t')$. ■

Remark 10. Corollary 2 extends Proposition 1 to the quantum case. It is important to emphasize the difference between the kernel in (52) for the quantum case and the standard integrated tuned-correlated kernel in (8). The former includes the fractional polynomial terms $(1 + a_m \tau + \frac{a_m^2 \tau^2}{2})/\tau^3$ (similarly for a_M), while the latter only contains the term $1/\tau$. Again this difference follows from the physical realizability constraints (28)–(30), which introduce the term α^2 in (46) and, consequently, α^4 in (50). For large values of τ , the two kernels converge. For small τ , the polynomial term $1 + a_m \tau + \frac{a_m^2 \tau^2}{2}$ behaves similarly to the polynomial term that appears in the impulse response of a system in Jordan form (specifically, in the case of a LTI system with a complex-conjugate pair of eigenvalues of multiplicity three). Therefore, paradoxically, the physical realizability constraints (28)–(30) seem to increase the expressivity of the RC when the distribution of randomized parameters is suitably chosen. We will investigate this aspect in future work. ■

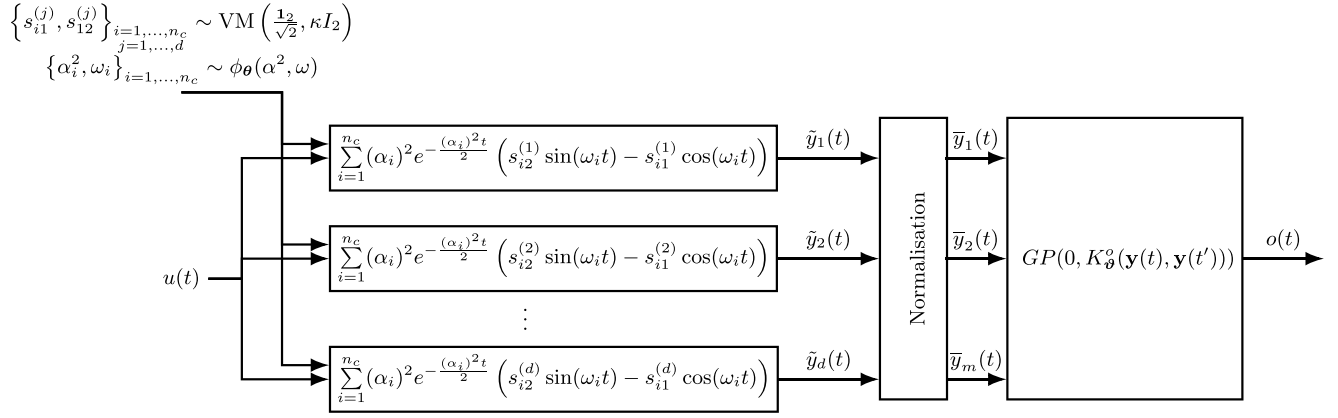
The difference between the kernel (8) and (52) is the main difference between the classical and the quantum Wiener architectures. Indeed, the kernel (52) determines the deep kernel of this qWiener architecture. Therefore, by adding the last component of the qWiener architecture, that is, the classical memoryless nonlinear readout layer, we obtain a deep kernel similar to that derived in Proposition 3, the only difference being that $K_\theta(\cdot, \cdot)$ is now the covariance defined in (52).

The quantum Wiener architecture, obtained by concatenating d quantum harmonic oscillators, performing weak measurements of the first quadratures, and passing the output to the readout layer, is depicted in Fig. 2. To obtain the deterministic component of (47), denoted by $\tilde{y}_j(t)$, we can repeatedly pass the input signal $u(t)$ through the j th quantum harmonic oscillator a total of N_{meas} times and then average the resulting output signals. This averaging is standard in reservoir computing, see for instance Ref. [27]. In Fig. 2, the normalization layer removes the mean $\frac{1}{d} \sum_{j=1}^d \tilde{y}_j(t)$ from each $\tilde{y}_j(t)$ and applies standard scaling to ensure numerical stability during the final layer's fitting process.

B. Universality of quantum Wiener architectures

To prove that $q\text{Wiener}$ architectures can realize general SISO LTI blocks, we need to show that a QSS can implement any impulse response of the form $h(t) = Ce^{At}B + D\delta(t)$ for arbitrary matrices A, B, C, D . Note that, without loss of generality, we can assume $D = 0$ since, in an RC architecture, we can always provide $u(t)$ as an input in the last layer.

It is important to observe that we cannot directly set $A^{(r)} = A$, $L^{(r)} = B$, and $C^{(r)} = C$, because physical realizability constraints (28) and (29) do not allow arbitrary choices of the matrices $A^{(r)}, L^{(r)}, C^{(r)}$. However, these constraints only impose conditions on the matrix elements relative to the state variables $\tilde{q}_i, \tilde{p}_i$, and not, for example, between $\tilde{q}_i$ and $\tilde{q}_j$. Therefore, in the following, we will show that one way to realize any desired impulse response is to embed its realization in the subsystem associated with the components of the state $\tilde{q}_1, \dots, \tilde{q}_n$ (or $\tilde{p}_1, \dots, \tilde{p}_n$).

FIG. 2. Quantum Wiener architecture obtained by concatenating d quantum harmonic oscillators.

Lemma 3. Consider the quantum linear system:

$$A^{(r)} = \begin{pmatrix} A & 0_{n \times n} \\ 0_{n \times n} & Z \end{pmatrix}, \quad (53)$$

$$C^{(r)} = \begin{pmatrix} C_1^{(r)} & 0_{1 \times n} \\ 0_{1 \times n} & C_2^{(r)} \end{pmatrix}, \quad (54)$$

$$L^{(r)} = J_n^{(r)} C^{(r)\top} J_s^{(r)}, \quad (55)$$

$$D^{(r)} = I_{2s}. \quad (56)$$

If $Z = -A^\top - C_1^{(r)\top} C_2^{(r)}$, then the system is physically realisable, that is, it satisfies (28)–(30), and the impulse response is

$$\mathbf{h}(t) = \begin{pmatrix} -C_1^{(r)} e^{At} C_2^{(r)\top} & 0 \\ 0 & -C_2^{(r)} e^{Zt} C_1^{(r)\top} \end{pmatrix} + I_{2s} \delta(t).$$

From the above lemma, it can be noticed that, by setting $C_1^{(r)} = -C$ and $C_2^{(r)} = B^\top$, the restriction of the impulse response to the first component of the input and output leads to the impulse response $Ce^{At}B + \delta(t)$. Therefore, we have shown that any SISO transfer function can be realized. As noted earlier, in RC the feedthrough term $\delta(t)$ can be disregarded, since the input $u(t)$ can always be fed into or removed from the input of the readout layer.

Observe that, in RC, the matrix A is assumed to have all eigenvalues with strictly negative real parts. However, since the matrix $Z = -A^\top + C^\top B^\top$ includes the term $-A^\top$, the matrix Z is not, in general, guaranteed to have eigenvalues with strictly negative real parts. This implies that the resulting QSS could be unstable in general. This is problematic, as the corresponding QRC could then produce an unbounded output, which is physically unrealizable. In what follows, we show that, by suitably choosing the matrices $C_1^{(r)}, C_2^{(r)}$, we can ensure that this does not occur.

Theorem 3. Assume that $C_1^{(r)} = (R^\top \quad -C^\top)^\top$, $C_2^{(r)} = (B \quad 0_{n \times 1})^\top$ and then $Z = -A^\top - R^\top B^\top = (-A - BR)^\top$, where $R \in \mathbb{R}^{1 \times n}$. The impulse response is as follows:

$$\mathbf{h}(t) = \begin{pmatrix} -Re^{At}B & 0 & 0 & 0 \\ Ce^{At}B & 0 & 0 & 0 \\ 0 & 0 & -B^\top e^{Zt}R^\top & B^\top e^{Zt}C^\top \\ 0 & 0 & 0 & 0 \end{pmatrix} + I_4 \delta(t).$$

Assuming that the matrix $(B, AB, A^2B, \dots, A^{n-1}B)$ has rank n , then we can always find R such that Z has eigenvalues with strictly negative real part, so that the QSS (53)–(56) is stable. ■

The restriction of the impulse response to the second component of the output and the first component of the input is $Ce^{At}B + \delta(t)$. In summary, we have shown that a QSS can realize any transfer function. The rank assumption in Theorem 3 is standard in control theory. The matrix R can for instance be determined by solving a linear-quadratic Gaussian control [53] problem with arbitrary weighting matrices.

We note that the quantum linear dynamical systems considered in this work are naturally described within the quantum input-output framework of open quantum systems [54,55]. In this framework, system dynamics are driven by input field operators and give rise to corresponding output fields that encode the system response. The underlying continuous-variable systems can be implemented using common quantum optical components such as cavities, propagating bosonic fields, beam splitters, and homodyne detection and thereby providing a natural route to experimental realization of our framework.

Here it is important to note the following:

Corollary 3. The QSS described by (53)–(56) is not completely passive. ■

This implies that this system requires external driving (e.g., an external pump beam) to implement.

Since the QSS is active, when a weak measurement is performed on it, the system evolves according to (31) and (32). In this case, the measurement backaction introduces stochastic contributions to the system dynamics, which needs to be properly accounted for.

To mitigate the effect of measurement-induced noise in the measured system (31) and (32), we pass the input signal $u(t)$ through the network N_{meas} times, average the outputs, and then feed this averaged result into the readout layer.

IV. CONNECTION TO PREVIOUS WORK

The work [7] proposes a network of interacting quantum harmonic oscillators acting as the reservoir for RC, with springlike interaction strengths g_{ij} . The Hamiltonian of such a system is described in terms of the Laplacian matrix $V = V^\top$ having elements $V_{ij} = \delta_{ij} \sum_k g_{ik} - (1 - \delta_{ij})g_{ij}$, where δ_{ij} is

the Kronecker delta. The resulting Hamiltonian is

$$\frac{\check{\mathbf{q}}^\top (G_\omega + V) \check{\mathbf{q}}}{2} + \frac{\check{\mathbf{p}}^\top \check{\mathbf{p}}}{2},$$

where $\check{\mathbf{p}} = (\check{p}_1, \dots, \check{p}_{n_s})^\top$, $\check{\mathbf{q}} = (\check{q}_1, \dots, \check{q}_{n_s})^\top$, and $G_\omega = \text{diag}(\omega_1, \dots, \omega_{n_s})$ with $\omega_i > 0$. This leads to the differential equations

$$\frac{d\check{\mathbf{q}}}{dt} = \check{\mathbf{p}}, \quad (57)$$

$$\frac{d\check{\mathbf{p}}}{dt} = -(G_\omega + V)\check{\mathbf{q}}. \quad (58)$$

In Ref. [7], the input is injected into the network by resetting at each timestep the state of one of the oscillators. Instead of modeling the input by ancilla resets as in Ref. [7], we use an equivalent linear QSS representation where the input couples through an external field as discussed in Sec. II B 1. Assume without loss of generality, we use the first harmonic oscillator $\check{q}_1, \check{p}_1$ to inject the input. By defining the matrices M_1, M_2, N_1 as follows:

$$M_1 = \frac{1}{2}(G_\omega + I_{n_s} + V), \quad (59)$$

$$M_2 = \frac{1}{2}(G_\omega - I_{n_s} + V), \quad (60)$$

$$N_1 = (a_1, 0, \dots, 0), \quad (61)$$

and $N_2 = 0$, we obtain the QSS matrices:

$$A^{(r)} = \begin{pmatrix} -\frac{a_1^2}{2} \mathbf{e}_1 \mathbf{e}_1^\top & I_{n_s} \\ -G_\omega - V & -\frac{a_1^2}{2} \mathbf{e}_1 \mathbf{e}_1^\top \end{pmatrix}, \quad (62)$$

$$C^{(r)} = a_1 \begin{pmatrix} \mathbf{e}_1 & \mathbf{0}_{n_s} \\ \mathbf{0}_{n_s} & \mathbf{e}_1 \end{pmatrix}^\top, \quad (63)$$

$$D^{(r)} = \begin{pmatrix} s_1 \mathbf{e}_1 & -s_2 \mathbf{e}_1 \\ s_2 \mathbf{e}_1 & s_1 \mathbf{e}_1 \end{pmatrix}, \quad (64)$$

$$L^{(r)} = C^{(r)\dagger} D^{(r)}, \quad (65)$$

where $\mathbf{e}_1$ is the first element of the canonical basis of $\mathbb{R}^{n_s}$ and $s_1^2 + s_2^2 = 1$. Note that, for $a_1 = 0$ (no input), the matrix $A^{(r)}$ coincides with the state matrix defined by (57) and (58). It is interesting to note that this proposed quantum linear system has a very particular structure for its dynamics and we do not know if this structure can represent any linear dynamics, which is necessary for universality of the corresponding qWiener architecture. Conversely, our approach achieves universality as proven for the system in Sec. III B. Additionally, in the numerical experiments, Ref. [7] sets $\omega = 0.25$ and selects random interaction strengths g uniformly distributed in the interval $[0.01, 0.19]$ and argues that this breaks symmetries and favors a spectral radius $\rho(A) < 1$. Reference [7] does not provide an analysis of the statistical properties of the resulting reservoir to justify the choice of the uniform distribution. In the present work, we go beyond such a heuristic for the case of the qWiener model, e.g., by explicitly expressing the limit kernel in Sec. III A 1 and, therefore, justifying the choice of the distribution $\phi_\theta(\alpha^2, \omega) = (\frac{\alpha^2}{2} / (\pi(\omega^2 + \frac{\alpha^4}{4}))) \text{Unif}(\alpha^2; \alpha_m, \alpha_M)$.

Finally, Ref. [7] discusses additional nonlinear capabilities obtained by varying the input encoding of the coherent

ancillary state from amplitude to phase. This type of model—comprising a nonlinear transformation of the input, a bank of SISO LTI systems, and a static nonlinearity in the readout layer—is known as a Hammerstein-Wiener model. In the following numerical experiment, we assume that the amplitude of the ancillary state is equal to the input itself, that is, no nonlinear transformation is applied at this stage. Note that nonlinear transformations of the input can be incorporated both in the model in Ref. [7] and in our proposed models.

V. NUMERICAL EXPERIMENTS

In this section, we compare the Laplacian-based architecture proposed in Ref. [7] and illustrated in Sec. IV, which we refer to as *LqW*, with the qWiener architecture obtained by concatenating harmonic oscillators, as presented in Sec. III A, which we refer to as *HqW*. Differently from Ref. [7], for both *LqW* and *HqW* we use a squared-exponential kernel in the readout layer (instead of a polynomial nonlinearity). This choice eliminates the need to select the order of the polynomial and also provides better extrapolation properties (the results obtained using a polynomial nonlinearity in the readout layer are reported in Appendix E for comparison). For *LqW*, in all the experiments, we select the parameters of the quantum linear network as $\omega = 0.25$, $g_{ij} \sim \text{Unif}(0.01, 0.19)$, and $a_1 = 5$. We tried different combination of the parameters but the performance did not change significantly. For *HqW*, in all experiments, we selected $a_m = 0.01$, $a_M = 20$. For comparison, as a baseline, we also include a classical echo-state network (ESN) [1] consisting of 200 neurons (spectral radius 0.95), that is, a classical RC based on a recurrent neural network. We use this ESN network only as a reference performance benchmark, as our main goal is to compare our qWiener architectures with *LqW*. As benchmark problems, we consider the following three dynamical systems [56].

(A) *Parity check*. The parity check task is a nonlinear temporal benchmark used to assess a model's ability to retain and combine past input symbols in a nonlinear way. Given a binary input sequence $u_t \in \{0, 1\}$, the target output is

$$y_t = \text{mod} \left(\sum_{l=0}^{\tau} u_{t-l}, 2 \right),$$

that is, the parity (even or odd) of the last τ input bits. As input, we use a pseudorandom binary sequences (with probability $p = 0.5$). Note that the parity task was also considered in Ref. [7].

(B) *NARMA10*. The nonlinear autoregressive moving average task is widely used to test nonlinear dynamical modeling and the balance between memory and nonlinearity. For NARMA10, the target output is as follows:

$$y_t = 0.3y_{t-1} + 0.05y_{t-1} \sum_{i=0}^9 y_{t-1-i} + 1.5u_{t-10}u_{t-1} + 0.1,$$

where u_t is drawn from a uniform distribution in $[0, 0.5]$. The task requires both long memory and nonlinear interactions between past inputs and outputs.

(C) *Delay*. The delay task quantifies how much information about past inputs a model can retain. Given an input u_t , the

TABLE I. For each benchmark, in the first row we report the median RMSE for (i) the classical RC ESN, (ii) the QRC LqW, and (iii) our QRC HqW. The second row includes the 25th and 75th RMSE percentiles between square brackets. The number between brackets after the name of the model is relative to the number of harmonic oscillators in the quantum linear network.

RMSE	ESN	LqW(8)	HqW(8)	LqW(16)	HqW(16)	LqW(24)	HqW(24)
Parity($\tau = 2$)	0.056 [0.053,0.059]	0.0024 [0.0016,0.0031]	0.00099 [0.00087,0.00108]	0.0036 [0.0023,0.0051]	0.00093 [0.00070,0.0010]	0.0039 [0.0027,0.0072]	0.00085 [0.00071,0.00091]
Parity($\tau = 4$)	0.67 [0.61,0.84]	0.0109 [0.0084,0.0259]	0.0129 [0.0064,0.0288]	0.0210 [0.0173,0.0462]	0.008 [0.0036,0.0167]	0.0290 [0.0195,0.0388]	0.006 [0.0037,0.0114]
NARMA10	0.042 [0.039,0.043]	0.109 [0.101,0.121]	0.079 [0.070,0.089]	0.109 [0.099,0.120]	0.062 [0.064,0.081]	0.113 [0.111,0.121]	0.060 [0.056,0.067]
Delay($\tau = 2$)	0.026 [0.024,0.031]	0.0032 [0.0027,0.0043]	0.0018 [0.0016,0.0030]	0.0041 [0.0031,0.0053]	0.0017 [0.0014,0.0021]	0.0043 [0.0035,0.0066]	0.0017 [0.0015,0.0019]
Delay($\tau = 4$)	0.096 [0.093,0.099]	0.50 [0.497,0.502]	0.037 [0.022,0.067]	0.50 [0.498,0.502]	0.027 [0.017,0.028]	0.50 [0.499,0.503]	0.017 [0.013,0.024]

target output is a delayed version of it,

$$y_t = u_{t-\tau}.$$

As input, we use a PRRBS sequence (with probability $p = 0.5$).

In all experiments, we selected a sampling step $\Delta = 0.01$. Furthermore, we fixed $Q^{(r)} = \frac{1}{2}I_{2s}$, which is the covariance of the vacuum, and, therefore, to filter out the effect of noise in the measured quantum linear network, we pass the signal through the network N_{meas} times, average the outputs, and then feed this averaged result into the readout layer. This is standard in reservoir computing, where $N_{\text{meas}} \in [10^4, 10^6]$, see for instance Ref. [27]. In this paper, we consider two cases, $N_{\text{meas}} = 1.5 \times 10^3$ (reported in this section) and $N_{\text{meas}} = 500$ (in Appendix E).

Table I reports the median of the root mean square error (RMSE) (averaged over the test set) computed over 20 Monte Carlo repetitions. In each Monte Carlo repetition, a time series of length 1000 is generated for the benchmarks described above, using different realizations of the random input sequence u_t . The first 800 points are used for training, while the remaining 200 points are used to compute the average RMSE. For each benchmark, in the first row, the table shows the median RMSE across the 20 repetitions for three models: (i) the classical RC ESN, (ii) the QRC LqW, and (iii) our QRC HqW. The second row includes the 25th and 75th RMSE percentiles between square brackets. In the top row, the number between brackets after the name of the model is relative to the number of harmonic oscillators in the quantum linear network. In particular, for HqW, we have considered the dimensions $d = n_c \in \{8, 16, 24\}$ using the architecture described in Fig. 2, while, for LqW, $n_s \in \{8, 16, 24\}$ and we measured the first quadratures of the state of the reservoir resulting in $d = n_s$.

It can be seen that our method, HqW, outperforms LqW practically across all benchmarks when using the same number of harmonic oscillators. Moreover, the RMSE of HqW is typically half that of LqW, or even lower. Contrary to what one would expect from a RC architecture, the performance of LqW does not generally improve as the number of harmonic oscillators increases; in fact, it often deteriorates, because the additional oscillators alter the system dynamics in a way that degrades performance (in Ref. [57], which considers a

different architecture, this issue appears to be mitigated). In contrast, a key advantage of our method is that we have proven that its performance improves as the number of oscillators increases, since the covariance converges more closely to the limiting kernel, thus increasing expressivity. Our method also generally outperforms the ESN (with 200 neurons). This is not surprising, as a standard ESN cannot reliably solve tasks that require large values of τ and, in particular, struggles with the parity task [58].

Figure 3 shows the prediction on one of the test sets of the Monte Carlo simulation for the NARMA10 benchmark: The true-signal is in gray, HqW(24) in orange, LqW(24) in green, and ESN(200) in blue.

A. Exploiting universality

In this section, we demonstrate how to exploit the universality of the qWiener architecture (proven in Sec. III B) to implement a QRC with long-term memory. In RC, a key determinant of performance is the *memory capacity*, that is, the ability of the system to retain information about past inputs. Consider the delay task

$$y(t) = u(t - \tau),$$

which requires perfect recall of the input τ time units in the past. This can be written in convolution form as

$$y(t) = \int_{-\infty}^{\infty} \delta(t - s - \tau) u(s) ds,$$

where the impulse response of an ideal delay is the shifted Dirac delta $h(t) = \delta(t - \tau)$.

The Laplace transform of this impulse response is $H(s) = e^{-s\tau}$. Its transfer function $e^{-s\tau}$ is *not* a rational function. Thus, an ideal delay is an infinite-dimensional system and cannot be realized by any finite-dimensional LTI dynamical system. The solution proposed in Ref. [33] is to employ a Padé approximation to find a *rational* approximation of the delay operator by replacing

$$e^{-s\tau} \approx \frac{P_m(s)}{Q_n(s)},$$

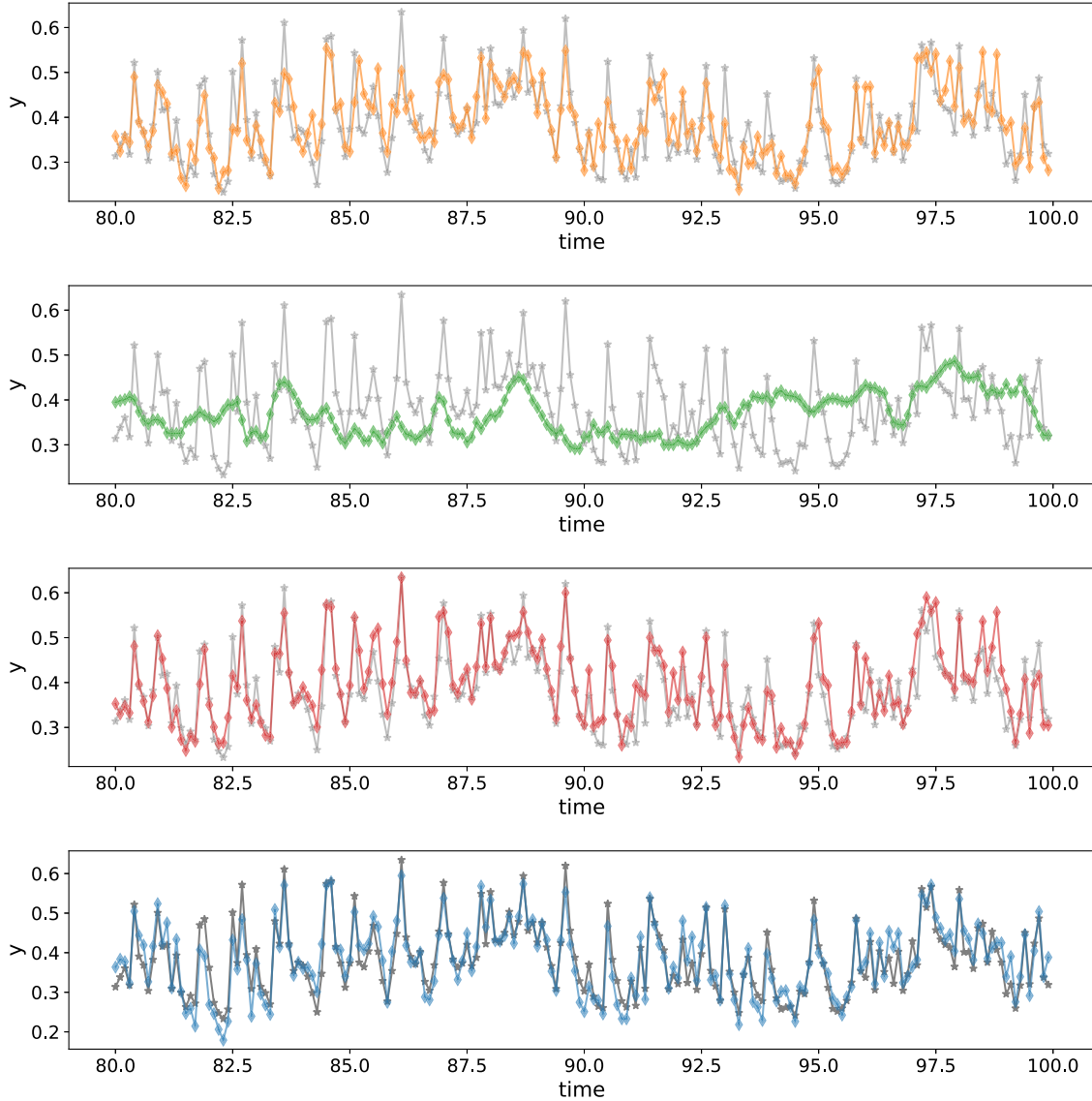


FIG. 3. Prediction on a test set for the NARMA10 benchmark: True signal is in gray, HqW(24) in orange, LqW(24) in green, PadeqW(14) in red, and ESN(200) in blue.

where P_m and Q_n are polynomials of degrees m and n , respectively (with $m \leq n$). This rational approximation can be implemented exactly by an LTI system. Moreover, as a RC, this design realizes a computational basis corresponding to orthogonal Legendre polynomials [33]. This is important because using orthogonal basis functions provides a principled alternative for constructing RCs, compared to the traditional approach of using randomized basis functions (which we used for our HqW).

In this section, we fix $n = 8$ and $m = 7$ and use the controllable canonical form to derive the matrices A , B , and C of a classical LTI system. We then apply Lemma 3 and Theorem 3 to derive the QSS that implements the impulse response Padé approximation of the delay.

We repeat these steps for different delays $\tau = k\Delta$ with $k = 1, 2, \dots, 14$. This allows us to obtain a qWiener architecture with 14 linear networks, which we denote as PadeqW(14). The performance of this QRC are reported in the Table II together with the performance of HqW(24) from the previous

table for comparison. This shows that the system achieves practically zero error on the delay ($\tau = 4$) tasks, as expected since its memory is $\tau = 14$. It also attains a very low RMSE on the NARMA10 task. Figure 4 shows the prediction on the same test set of Fig. 4 for PadeqW(14), showing the improved performance.

TABLE II. Median and [25th,75th] percentile RMSE for our QRC HqW(24) versus our QRC PadeqW(14).

RMSE	HqW(24)	PadeqW(14)
Parity($\tau = 4$)	0.0060 [0.0037,0.0114]	0.0058 [0.0042,0.0129]
NARMA10	0.060 [0.056,0.067]	0.044 [0.041,0.049]
Delay($\tau = 4$)	0.017 [0.013,0.024]	0.0012 [0.0011,0.0013]

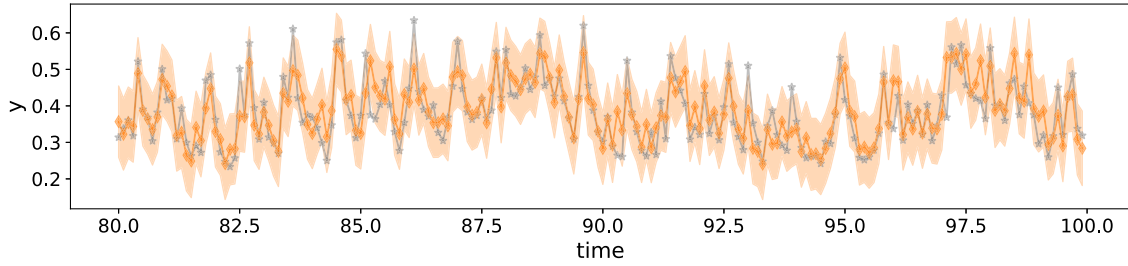


FIG. 4. Prediction on a test set for the NARMA10 benchmark for HqW(24): The mean is the orange line and the orange-shaded region is the 95% credible interval.

B. Implementation

The source code to run the experiments in the manuscript was implemented in Python and is available.

VI. CONCLUSIONS

In this work, we have derived a quantum version of the Wiener architecture for RC. This architecture consists of a quantum linear network followed by a weak measurement and a classical static nonlinear readout layer. We have proved that, although the quantum dynamics impose constraints on the parameters of the quantum linear network, the universality of the architecture can still be guaranteed. Additionally, we have shown that the qWiener architecture can be obtained by concatenating simple harmonic oscillators and using only passive optical components such as phase shifters, beam splitters, and mirrors. To prove universality of the architecture, however, we use an alternative structure which requires active optical components. Finally, through numerical experiments, we have verified that our quantum-RC achieves very high performance on standard benchmarks for RC. ■

In this paper, we have established the universality of the qWiener architecture. However, it remains an open question what advantages it might offer (beyond the inherent high dimensionality of quantum systems) over the classical Wiener architecture, which is also universal. One possible research direction is to explore quantum-RC architectures consisting of additional layers (deep) but where the weak measurement occurs only before the readout layer. This would allow a deeper exploitation of the quantum features of the system and extend the classical deep Wiener architectures proposed in Ref. [59] to the quantum case. Finally, we have not explicitly emphasized that our proposed architecture is Bayesian, due to the Gaussian processes used in the final readout layer. This provides, at no additional implementation cost, a principled quantification of uncertainty, as shown in Fig. 4. We believe this aspect is particularly important for reservoir computing architectures, which can often exhibit larger prediction errors due to their random initialization. As future work, we plan to compare our uncertainty quantification with that obtained from fully trained classical networks, such as those considered in Refs. [60,61].

DATA AVAILABILITY

The data that support the findings of this article are openly available [62].

APPENDIX A: VOLTERRA SERIES

A system can be defined as a map between an input signal $u(t)$ and an output signal $y(t)$, that is, $y(t) = Mx(t)$, where M is a (system) operator from the input to the output function space. The system operator is typically assumed to be time-invariant and continuous. The Volterra series operator (of order m) is defined as

$$y(t) = Vu(t) = H_0 + \sum_{i=1}^m H_i u(t),$$

where H_0 is the constant function and, for $i > 0$,

$$H_i = \int h^{(i)}(\tau_1, \dots, \tau_i) u(t - \tau_1) \cdots u(t - \tau_i) d\tau_1 \cdots d\tau_i,$$

where the function $h^{(i)} : \mathbb{R}^i \rightarrow \mathbb{R}$ is called the i th order Volterra kernel. The integrals are computed over a finite or an infinite time intervals. The support of the Volterra kernel defines the memory of the system, that is, it defines the time interval in which past inputs can influence the current system output. It is evident how the Volterra series generalizes the impulse response of a LTI system, which is given in (3).

If the input signals are nonzero only on a finite time interval $[0, T]$, then it can be shown that any time-invariant, continuous operator M can be uniformly approximated to arbitrary accuracy by a Volterra series operator of sufficient but finite order. Boyd and Chua [21] generalized this result to input signals defined on unbounded domains, but restricting M to have the fading memory property, which is stronger than continuity. Their assumptions on the input signal u are boundedness and Lipschitz continuity. Additionally, Boyd and Chua [21] prove the following result.

Theorem A1. Let $\epsilon > 0$ and u is bounded and Lipschitz continuous, Suppose that M is a time-invariant operator with fading memory. Then there is a finite Volterra series operator V such that for all u (bounded and Lipschitz continuous), it holds that

$$\|Mu(t) - Vu(t)\| \leq \epsilon,$$

where $\|r(t)\| = \sup_{t \in \mathbb{R}} |r(t)|$. Additionally, V is a input-output operator of the dynamical system

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u, y = p(\mathbf{x}),$$

where $\mathbf{A} \in \mathbb{R}^{n_x \times n_x}$ is strictly stable and $p : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$ is a polynomial. ■

This proves that any operator M time-invariant with fading memory, can be approximated by the input-output operator of

a finite-dimensional linear dynamical system with a nonlinear (indeed, polynomial) readout map.

APPENDIX B: COMPUTATION OF THE GAUSSIAN PROCESS POSTERIOR

Assuming the observation noises ε_i are independent, the likelihood of the observed data given the latent function values is as follows:

$$p(\mathcal{D}_m | f(Y)) = \mathcal{N}(\mathbf{o}; f(Y), \sigma^2 I_m), \quad (\text{B1})$$

where $\mathbf{o} = [o(t_1), \dots, o(t_m)]^\top$, $Y = [\mathbf{y}^\top(t_1), \dots, \mathbf{y}^\top(t_m)]^\top$, $t_i = i\Delta$, and I_m is the identity matrix of size m .

(1) *Prior*: We place a Gaussian process prior on the function $f \sim \mathcal{GP}(0, K_\theta^o)$, where K_θ^o is a kernel function defined over the feature vectors $\mathbf{y}(t)$. This implies

$$p(f(Y) | \theta) = \mathcal{N}(f(Y); 0, K_\theta^o(Y, Y)), \quad (\text{B2})$$

where $(K_\theta^o(Y, Y))_{ij} = K_\theta^o(\mathbf{y}(t_i), \mathbf{y}(t_j))$.

(2) *Posterior*: Combining the likelihood (B1) with the prior (B2), the posterior distribution over the function values is as follows:

$$p(f(Y) | \mathcal{D}_m, \theta) = \mathcal{N}(f(Y); \mu_p(Y), K_p(Y, Y)),$$

with posterior mean and covariance given by

$$\mu_p(Y) = K_\theta^o(Y, Y)(K_\theta^o(Y, Y) + \sigma^2 I_m)^{-1} \mathbf{o}, \quad (\text{B3})$$

$$K_p(Y, Y) = K_\theta^o(Y, Y)$$

$$- K_\theta^o(Y, Y)(K_\theta^o(Y, Y) + \sigma^2 I_T)^{-1} K_\theta^o(Y, Y). \quad (\text{B4})$$

This operation is equivalent to linear regression (with nonlinear basis functions) but performed in the dual space of the kernel [[30], Chap. 2].

(3) *Predictive posterior*: For a new input $\mathbf{y}(t_r)$, the predictive distribution is also Gaussian:

$$p(f(\mathbf{y}(t_r)) | \mathcal{D}_m, \theta) = \mathcal{N}(\mu_p(\mathbf{y}(t_r)), K_p(\mathbf{y}(t_r), \mathbf{y}(t_r'))),$$

with

$$\mu_p(\mathbf{y}(t_r)) = K_\theta^o(\mathbf{y}(t_r), Y)(K_\theta^o(Y, Y) + \sigma^2 I_m)^{-1} \mathbf{o}, \quad (\text{B5})$$

$$K_p(\mathbf{y}(t_r), \mathbf{y}(t_r)) = K_\theta^o(\mathbf{y}(t_r), \mathbf{y}(t_r)) - K_\theta^o(\mathbf{y}(t_r), Y) \times (K_\theta^o(Y, Y) + \sigma^2 I_m)^{-1} K_\theta^o(Y, \mathbf{y}(t_r)). \quad (\text{B6})$$

(4) *Hyperparameter selection*. The hyperparameters of the kernel θ and noise variance σ^2 are the only unknowns. They are typically estimated by maximizing the marginal likelihood:

$$p(\mathcal{D}_m | \theta, \sigma^2) = \mathcal{N}(\mathbf{o}; 0, K_\theta^o(Y, Y) + \sigma^2 I_m).$$

APPENDIX C: VON MISES-FISHER DISTRIBUTION

Assume $s_1, s_2 \sim \mathcal{N}(\mathbf{m}, 1/\kappa I_2)$, with $\mathbf{m} \in \mathbb{R}^2$ and $\|\mathbf{m}\| = 1$. Then, conditioned on $\sqrt{s_1^2 + s_2^2} = 1$, the vector $\mathbf{s} = (s_1, s_2)^\top$ is von Mises-Fisher distributed with probability density function

$$f(\mathbf{s}; \mathbf{m}, \kappa) = C_2(\kappa) \exp(\kappa \mathbf{m}^\top \mathbf{s}), \quad (\text{C1})$$

where

$$C_2(\kappa) = \frac{1}{2\pi \mathcal{I}_0(\kappa)}, \quad (\text{C2})$$

with $\mathcal{I}_n(\kappa)$ being the modified Bessel function of the first kind of order n . It then holds [63]:

$$E(\mathbf{s}) = r_1 \mathbf{m}, \quad (\text{C3})$$

$$E(\mathbf{s}\mathbf{s}^\top) = \frac{1-r_2}{2} I_2 + r_2 \mathbf{m}\mathbf{m}^\top, \quad (\text{C4})$$

$$\text{Cov}(\mathbf{s}) = \frac{1-r_2}{2} I_2 + (r_2 - r_1^2) \mathbf{m}\mathbf{m}^\top, \quad (\text{C5})$$

where $r_n = \mathcal{I}_n(\kappa)/\mathcal{I}_0(\kappa)$.

APPENDIX D: PROOFS

This section presents the proofs of the main results of the paper, along with the intermediate lemmas required for their derivation.

(1) *Proof of Proposition 2*. The output of one LTI block is

$$y_j(t) = \int_0^t h_j(t-\tau) u(\tau) d\tau = \int_0^t \sum_{i=1}^{n_c} e^{-\alpha_i(t-\tau)} (\beta_i^{(j)} \cos(\omega_i(t-\tau)) - \gamma_i^{(j)} \sin(\omega_i(t-\tau))) u(\tau) d\tau,$$

and, therefore,

$$\begin{aligned} \mathbf{y}^\top(t) \mathbf{y}(t') &= \frac{1}{d} \sum_{j=1}^d \int_0^t \int_0^{t'} \left(\sum_{i=1}^{n_c} e^{-\alpha_i(t-\tau)} (\beta_i^{(j)} \cos(\omega_i(t-\tau)) - \gamma_i^{(j)} \sin(\omega_i(t-\tau))) u(\tau) d\tau \right) \\ &\quad \times \left(\sum_{i=1}^{n_c} e^{-\alpha_i(t'-\tau')} (\beta_i^{(j)} \cos(\omega_i(t'-\tau')) - \gamma_i^{(j)} \sin(\omega_i(t'-\tau'))) u(\tau') d\tau' \right). \end{aligned}$$

We recall that $\beta_i^{(j)}, \gamma_i^{(j)} \sim \mathcal{N}(0, 1/n_c)$ independently for $i = 1, \dots, n_c$ and $j = 1, \dots, d$. Therefore, for large d and n_c , we have that

$$\begin{aligned} \mathbf{y}^\top(t) \mathbf{y}(t') &\approx \frac{1}{d} \sum_{j=1}^d \int_0^t \int_0^{t'} \left(\sum_{i=1}^{n_c} e^{-\alpha_i(t-\tau) - \alpha_i(t'-\tau')} (\beta_i^{2(j)} \cos(\omega_i(t-\tau)) \cos(\omega_i(t'-\tau')) \right. \\ &\quad \left. + \gamma_i^{2(j)} \sin(\omega_i(t-\tau)) \sin(\omega_i(t'-\tau')) \right) u(\tau) u(\tau') d\tau d\tau' \\ &\approx \frac{1}{n_c} \sum_{i=1}^{n_c} \int_0^t \int_0^{t'} (e^{-\alpha_i(t-\tau) - \alpha_i(t'-\tau')} (\cos(\omega_i(t-\tau)) \cos(\omega_i(t'-\tau')) + \sin(\omega_i(t-\tau)) \sin(\omega_i(t'-\tau'))) u(\tau) u(\tau') d\tau d\tau' \\ &\approx \frac{1}{n_c} \sum_{i=1}^{n_c} \int_0^t \int_0^{t'} (e^{-\alpha_i(t-\tau) - \alpha_i(t'-\tau')} \cos(\omega_i(t-\tau - (t'-\tau'))) u(\tau) u(\tau') d\tau d\tau' \\ &\approx \int_0^t \int_0^{t'} \int_0^\infty \int_{-\infty}^\infty (e^{-\alpha(t-\tau - (t'-\tau'))} \cos(\omega(t-\tau - (t'-\tau'))) \phi_\theta(\alpha, \omega) d\alpha d\omega u(\tau) u(\tau') d\tau d\tau' \\ &= \int_0^t \int_0^{t'} K_{\theta, n_c}(t-\tau, t'-\tau') u(\tau) u(\tau') d\tau d\tau'. \end{aligned}$$

(2) *Proof of Proposition 3.* The result is derived from Proposition 2. For the polynomial kernel, we take advantage of the fact that, for large d and n_c , the term $\mathbf{y}(t)^\top \mathbf{y}(t')$ is approximately equal to (12). For the square-exponential kernel, note that $\|\mathbf{y}(t) - \mathbf{y}(t')\|^2 = \mathbf{y}^\top(t) \mathbf{y}(t) - 2\mathbf{y}^\top(t) \mathbf{y}(t') + \mathbf{y}^\top(t') \mathbf{y}(t')$ and the rest follows similarly to the derivation for the polynomial kernel. ■

(3) *Proof of Theorem 1.* Note that $Q^{(r)} = \frac{1}{2} I_{2s}$ and assume that $V_\infty = Q^{(r)}$, then

$$\begin{aligned} G_\infty &= V_\infty C_q^{(r)\top} + L^{(r)} Q^{(r)} D_q^{(r)\top} = \frac{1}{2} (C_q^{(r)\top} + L^{(r)} D_q^{(r)\top}) = \frac{1}{2} (I_s \quad 0_s) (C^{(r)\top} + L^{(r)} D^{(r)\top}) (I_s \quad 0_s)^\top \\ &= \frac{1}{2} (I_s \quad 0_s) (C^{(r)\top} + L^{(r)}) (I_s \quad 0_s)^\top = 0, \end{aligned}$$

where the last equality follows by the physical realizability constraint (29). Note in fact that, under the assumption $s = n$, we have that

$$J_n^{(r)} L^{(r)} + C^{(r)\top} J_s^{(r)} = J_n^{(r)} (J_n^{(r)} C^{(r)\top} J_n^{(r)}) + C^{(r)\top} J_s^{(r)} = -C^{(r)\top} J_n^{(r)} + C^{(r)\top} J_s^{(r)} = 0.$$

It remains to show that for $V_\infty = \frac{1}{2} I_{2s}$ the following equality holds: $0 = A^{(r)} V_\infty + V_\infty A^{(r)\top} + L^{(r)} \tilde{Q}^{(r)} L^{(r)\top}$. This results follows by

Lemma D1 ([64]). Assume $Q^{(r)} = \frac{1}{2} I_{2s}$, that is, $Q^{(r)}$ is the covariance of the vacuum. If a completely passive QSS is strictly stable, then it must evolve into a vacuum state, that is, $V_\infty = \frac{1}{2} I_n$. ■

(4) *Useful lemma*

Lemma D2. Consider the matrix

$$A = \begin{pmatrix} -\frac{\alpha^2 + \beta^2}{2} & -\omega \\ \omega & -\frac{\alpha^2 + \beta^2}{2} \end{pmatrix},$$

its matrix exponential is

$$e^{At} = e^{-(\alpha^2 + \beta^2)t} \begin{pmatrix} \cos(\omega t) & -\sin(\omega t) \\ \sin(\omega t) & \sin(\omega t) \end{pmatrix}. \quad (\text{D1})$$

Proof. Rewrite

$$\begin{pmatrix} -\frac{\alpha^2 + \beta^2}{2} & -\omega \\ \omega & -\frac{\alpha^2 + \beta^2}{2} \end{pmatrix} = \begin{pmatrix} -\frac{\alpha^2 + \beta^2}{2} & 0 \\ 0 & -\frac{\alpha^2 + \beta^2}{2} \end{pmatrix} + \begin{pmatrix} 0 & -\omega \\ \omega & 0 \end{pmatrix} = X + Y.$$

Since the two matrices are such that $XY = YX$ (they commute), from the properties of the matrix exponential we have that $e^{(X+Y)t} = e^{Xt} e^{Yt}$, where

$$e^{Xt} = e^{-\frac{\alpha^2 + \beta^2}{2}t} I_2 e^{Yt} = \Omega_2^\dagger \begin{pmatrix} e^{i\omega t} & 0 \\ 0 & e^{-i\omega t} \end{pmatrix} \Omega_2 = \Omega_2^\dagger \begin{pmatrix} \cos(\omega t) + i \sin(\omega t) & 0 \\ 0 & \cos(\omega t) - i \sin(\omega t) \end{pmatrix} \Omega_2 = \begin{pmatrix} \cos(\omega t) & -\sin(\omega t) \\ \sin(\omega t) & \sin(\omega t) \end{pmatrix},$$

where

$$\Omega_n = \frac{1}{\sqrt{2}} \begin{pmatrix} I_n & \imath I_n \\ I_n & -\imath I_n \end{pmatrix}, \quad (\text{D2})$$

(5) *Proof of Lemma 2.* Since $n = 1$, and the system is completely passive, then $M_2 = 0$, $N_2 = 0$, $\text{Re}(M_1) = \text{Re}(M_1)^\top$, and $\text{Im}(M_1) = -\text{Im}(M_1)^\top = 0$. Furthermore, since $n = 1$, we have that $M_1, S \in \mathbb{C}$ and $N_1 \in \mathbb{C}^{1 \times s}$. Therefore, from (27), we can derive that

$$\begin{pmatrix} \text{Im}(M_1 + M_2) & \text{Re}(M_1 - M_2) \\ -\text{Re}(M_1 + M_2) & -\text{Im}(-M_1 + M_2) \end{pmatrix} = \begin{pmatrix} 0 & \text{Re}(M_1) \\ -\text{Re}(M_1) & 0 \end{pmatrix}$$

and

$$\begin{aligned} -\frac{1}{2} C^{(r)\dagger} C^{(r)} &= \frac{1}{2} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \text{Re}(N_1)^\top & \text{Im}(N_1)^\top \\ -\text{Im}(N_1)^\top & \text{Re}(N_1)^\top \end{pmatrix} J_s^{(r)} \begin{pmatrix} \text{Re}(N_1) & -\text{Im}(N_1) \\ \text{Im}(N_1) & \text{Re}(N_1) \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \text{Re}(N_1)^\top & \text{Im}(N_1)^\top \\ -\text{Im}(N_1)^\top & \text{Re}(N_1)^\top \end{pmatrix} \begin{pmatrix} 0 & I_s \\ -I_s & 0 \end{pmatrix} \begin{pmatrix} \text{Re}(N_1) & -\text{Im}(N_1) \\ \text{Im}(N_1) & \text{Re}(N_1) \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \text{Re}(N_1)^\top & \text{Im}(N_1)^\top \\ -\text{Im}(N_1)^\top & \text{Re}(N_1)^\top \end{pmatrix} \begin{pmatrix} \text{Im}(N_1) & \text{Re}(N_1) \\ -\text{Re}(N_1) & \text{Im}(N_1) \end{pmatrix} \\ &= \begin{pmatrix} -\lambda & 0 \\ 0 & -\lambda \end{pmatrix}, \end{aligned}$$

where $\lambda = \frac{1}{2} \text{Im}(N_1)^\top \text{Im}(N_1) + \frac{1}{2} \text{Re}(N_1)^\top \text{Re}(N_1)$. Therefore, we have

$$A^{(r)} = \begin{pmatrix} -\lambda & \text{Re}(M_1) \\ -\text{Re}(M_1) & -\lambda \end{pmatrix}.$$

The matrices $C^{(r)}, L^{(r)}$ are

$$C^{(r)} = \begin{pmatrix} \text{Re}(N_1) & -\text{Im}(N_1) \\ \text{Im}(N_1) & \text{Re}(N_1) \end{pmatrix} L^{(r)} = \begin{pmatrix} -\text{Re}(N_1)^\top & -\text{Im}(N_1)^\top \\ \text{Im}(N_1)^\top & -\text{Re}(N_1)^\top \end{pmatrix} \begin{pmatrix} \text{Re}(S) & -\text{Im}(S) \\ \text{Im}(S) & \text{Re}(S) \end{pmatrix}.$$

Therefore, by exploiting (D1), the impulse response is

$$\begin{aligned} C^{(r)} e^{A^{(r)} t} L^{(r)} &= -e^{-\lambda t} \begin{pmatrix} \text{Re}(N_1) & -\text{Im}(N_1) \\ \text{Im}(N_1) & \text{Re}(N_1) \end{pmatrix} \begin{pmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{pmatrix} \begin{pmatrix} \text{Re}(N_1)^\top & \text{Im}(N_1)^\top \\ -\text{Im}(N_1)^\top & \text{Re}(N_1)^\top \end{pmatrix} \begin{pmatrix} \text{Re}(S) & -\text{Im}(S) \\ \text{Im}(S) & \text{Re}(S) \end{pmatrix} \\ &= -e^{-\lambda t} \begin{pmatrix} \text{Re}(N_1) \cos(\omega t) + \text{Im}(N_1) \sin(\omega t) & \text{Re}(N_1) \sin(\omega t) - \text{Im}(N_1) \cos(\omega t) \\ \text{Im}(N_1) \cos(\omega t) - \text{Re}(N_1) \sin(\omega t) & \text{Re}(N_1) \cos(\omega t) + \text{Im}(N_1) \sin(\omega t) \end{pmatrix} \\ &\quad \times \begin{pmatrix} \text{Re}(N_1)^\top & \text{Im}(N_1)^\top \\ -\text{Im}(N_1)^\top & \text{Re}(N_1)^\top \end{pmatrix} \begin{pmatrix} \text{Re}(S) & -\text{Im}(S) \\ \text{Im}(S) & \text{Re}(S) \end{pmatrix} \\ &= -e^{-\lambda t} \begin{pmatrix} f_1 & f_2 \\ f_3 & f_4 \end{pmatrix} \begin{pmatrix} \text{Re}(S) & -\text{Im}(S) \\ \text{Im}(S) & \text{Re}(S) \end{pmatrix}, \end{aligned}$$

where $\omega = \text{Re}(M_1)$ and

$$\begin{aligned}
 f_1 &= \text{Re}(N_1)\text{Re}(N_1)^\top \cos(\omega t) + \text{Im}(N_1)\text{Re}(N_1)^\top \sin(\omega t) \\
 &\quad - \text{Re}(N_1)\text{Im}(N_1)^\top \sin(\omega t) + \text{Im}(N_1)\text{Im}(N_1)^\top \cos(\omega t) \\
 &= (\text{Re}(N_1)\text{Re}(N_1)^\top + \text{Im}(N_1)\text{Im}(N_1)^\top) \cos(\omega t) \\
 f_2 &= \text{Re}(N_1)\text{Im}(N_1)^\top \cos(\omega t) + \text{Im}(N_1)\text{Im}(N_1)^\top \sin(\omega t) \\
 &\quad + \text{Re}(N_1)\text{Re}(N_1)^\top \sin(\omega t) - \text{Im}(N_1)\text{Re}(N_1)^\top \cos(\omega t) \\
 &= (\text{Re}(N_1)\text{Re}(N_1)^\top + \text{Im}(N_1)\text{Im}(N_1)^\top) \sin(\omega t) \\
 f_3 &= \text{Im}(N_1)\text{Re}(N_1)^\top \cos(\omega t) - \text{Re}(N_1)\text{Re}(N_1)^\top \sin(\omega t) \\
 &\quad - \text{Re}(N_1)\text{Im}(N_1)^\top \cos(\omega t) - \text{Im}(N_1)\text{Im}(N_1)^\top \sin(\omega t) \\
 &= -(\text{Re}(N_1)\text{Re}(N_1)^\top + \text{Im}(N_1)\text{Im}(N_1)^\top) \sin(\omega t) \\
 f_4 &= \text{Im}(N_1)\text{Im}(N_1)^\top \cos(\omega t) - \text{Re}(N_1)\text{Im}(N_1)^\top \sin(\omega t) \\
 &\quad + \text{Re}(N_1)\text{Re}(N_1)^\top \cos(\omega t) + \text{Im}(N_1)\text{Re}(N_1)^\top \sin(\omega t) \\
 &= (\text{Re}(N_1)\text{Re}(N_1)^\top + \text{Im}(N_1)\text{Im}(N_1)^\top) \cos(\omega t).
 \end{aligned}$$

Finally, we have that

$$C^{(r)} e^{A^{(r)} t} L^{(r)} = -e^{-\lambda t} \left(\begin{pmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{pmatrix} \otimes \Lambda \right) \begin{pmatrix} \text{Re}(S) & -\text{Im}(S) \\ \text{Im}(S) & \text{Re}(S) \end{pmatrix},$$

where $\Lambda = (\text{Re}(N_1)\text{Re}(N_1)^\top + \text{Im}(N_1)\text{Im}(N_1)^\top)$. ■

(5) *Proof of Corollary 1.* The result follows from Lemma 2 by direct substitution of the parameters. ■

(6) *Proof of Proposition 4.* Without loss of generality we can prove the results for $n_c = 2$ due to the symmetries in the concatenation product,

$$C^{(r)} = \begin{pmatrix} \text{Re}(N_1 + N_2) & \text{Im}(-N_1 + N_2) \\ \text{Im}(N_1 + N_2) & \text{Re}(N_1 - N_2) \end{pmatrix} = \begin{pmatrix} \alpha_1 & 0 & 0 & 0 \\ 0 & \alpha_2 & 0 & 0 \\ 0 & 0 & \alpha_1 & 0 \\ 0 & 0 & 0 & \alpha_2 \end{pmatrix},$$

and, therefore,

$$\begin{aligned}
 C^{(r)\sharp} C^{(r)} &= - \begin{pmatrix} 0_2 & I_2 \\ -I_2 & 0_2 \end{pmatrix} \begin{pmatrix} \alpha_1 & 0 & 0 & 0 \\ 0 & \alpha_2 & 0 & 0 \\ 0 & 0 & \alpha_1 & 0 \\ 0 & 0 & 0 & \alpha_2 \end{pmatrix} \begin{pmatrix} 0_2 & I_2 \\ -I_2 & 0_2 \end{pmatrix} \begin{pmatrix} \alpha_1 & 0 & 0 & 0 \\ 0 & \alpha_2 & 0 & 0 \\ 0 & 0 & \alpha_1 & 0 \\ 0 & 0 & 0 & \alpha_2 \end{pmatrix} \\
 &= - \begin{pmatrix} 0 & 0 & \alpha_1 & 0 \\ 0 & 0 & 0 & \alpha_2 \\ -\alpha_1 & 0 & 0 & 0 \\ 0 & -\alpha_2 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & \alpha_1 & 0 \\ 0 & 0 & 0 & \alpha_2 \\ -\alpha_1 & 0 & 0 & 0 \\ 0 & -\alpha_2 & 0 & 0 \end{pmatrix} \\
 &= - \begin{pmatrix} -\alpha_1^2 & 0 & 0 & 0 \\ 0 & -\alpha_2^2 & 0 & 0 \\ 0 & 0 & -\alpha_1^2 & 0 \\ 0 & 0 & 0 & -\alpha_2^2 \end{pmatrix}. \tag{D3}
 \end{aligned}$$

The matrix $A^{(r)}$ is

$$\begin{aligned}
 A^{(r)} &= \begin{pmatrix} \text{Im}(M_1 + M_2) & \text{Re}(M_1 - M_2) \\ -\text{Re}(M_1 + M_2) & -\text{Im}(-M_1 + M_2) \end{pmatrix} - \frac{1}{2} C^{(r)\sharp} C^{(r)}, \\
 &= \begin{pmatrix} -\frac{\alpha_1^2}{2} & 0 & -\omega_1 & 0 \\ 0 & -\frac{\alpha_2^2}{2} & 0 & -\omega_2 \\ \omega_1 & 0 & -\frac{\alpha_1^2}{2} & 0 \\ 0 & \omega_2 & 0 & -\frac{\alpha_2^2}{2} \end{pmatrix}
 \end{aligned}$$

and

$$L^{(r)} = -C^{(r)\dagger}D^{(r)} = -\begin{pmatrix} \alpha_1 & 0 & 0 & 0 \\ 0 & \alpha_2 & 0 & 0 \\ 0 & 0 & \alpha_1 & 0 \\ 0 & 0 & 0 & \alpha_2 \end{pmatrix} D^{(r)}, \quad D^{(r)} = \begin{pmatrix} s_{11} & 0 & -s_{12} & 0 \\ 0 & s_{21} & 0 & -s_{22} \\ s_{12} & 0 & s_{11} & 0 \\ 0 & s_{22} & 0 & s_{21} \end{pmatrix}.$$

Therefore, by using a derivation similar to the one in Lemma D2, we have that

$$e^{A^{(r)}t} + \begin{pmatrix} e^{-\frac{1}{2}\alpha_1^2 t} \cos(\omega_1 t) & 0 & -e^{-\frac{1}{2}\alpha_1^2 t} \sin(\omega_1 t) & 0 \\ 0 & e^{-\frac{1}{2}\alpha_2^2 t} \cos(\omega_2 t) & 0 & -e^{-\frac{1}{2}\alpha_2^2 t} \sin(\omega_2 t) \\ e^{-\frac{1}{2}\alpha_1^2 t} \sin(\omega_1 t) & 0 & e^{-\frac{1}{2}\alpha_1^2 t} \cos(\omega_1 t) & 0 \\ 0 & e^{-\frac{1}{2}\alpha_2^2 t} \sin(\omega_2 t) & 0 & e^{-\frac{1}{2}\alpha_2^2 t} \cos(\omega_2 t) \end{pmatrix}.$$

The concatenation product simply assembles the different components together, without making any connections between them. Therefore, if we observe the first quadrature for each component, and assume the input $\mathbf{u}(t) = (1 \quad 1 \quad 0 \quad 0)^\top u(t)$, then the impulse response becomes

$$\begin{aligned} \mathbf{h}(t) &= \left(-\begin{pmatrix} \alpha_1 & 0 & 0 & 0 \\ 0 & \alpha_2 & 0 & 0 \end{pmatrix} e^{A^{(r)}t} \begin{pmatrix} \alpha_1 & 0 & 0 & 0 \\ 0 & \alpha_2 & 0 & 0 \\ 0 & 0 & \alpha_1 & 0 \\ 0 & 0 & 0 & \alpha_2 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \delta(t) \right) \\ &\quad \times \begin{pmatrix} s_{11} & 0 & -s_{12} & 0 \\ 0 & s_{21} & 0 & -s_{22} \\ s_{12} & 0 & s_{11} & 0 \\ 0 & s_{22} & 0 & s_{21} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \end{aligned} \quad (\text{D4})$$

$$\begin{aligned} &= \left(-\begin{pmatrix} \alpha_1 & 0 & 0 & 0 \\ 0 & \alpha_2 & 0 & 0 \end{pmatrix} e^{A^{(r)}t} \begin{pmatrix} \alpha_1 & 0 & 0 & 0 \\ 0 & \alpha_2 & 0 & 0 \\ 0 & 0 & \alpha_1 & 0 \\ 0 & 0 & 0 & \alpha_2 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \delta(t) \right) \begin{pmatrix} s_{11} \\ s_{21} \\ s_{12} \\ s_{22} \end{pmatrix} \\ &= \left(-\begin{pmatrix} \alpha_1^2 e^{-\frac{1}{2}\alpha_1^2 t} & 0 \\ 0 & \alpha_2^2 e^{-\frac{1}{2}\alpha_2^2 t} \end{pmatrix} \begin{pmatrix} \cos(\omega_1 t) & 0 & -\sin(\omega_1 t) & 0 \\ 0 & \cos(\omega_2 t) & 0 & -\sin(\omega_2 t) \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \delta(t) \right) \begin{pmatrix} s_{11} \\ s_{21} \\ s_{12} \\ s_{22} \end{pmatrix} \\ &= \begin{pmatrix} -s_{11}\alpha_1^2 e^{-\frac{1}{2}\alpha_1^2 t} \cos(\omega_1 t) + s_{12}\alpha_1^2 e^{-\frac{1}{2}\alpha_1^2 t} \sin(\omega_1 t) \\ -s_{21}\alpha_2^2 e^{-\frac{1}{2}\alpha_2^2 t} \cos(\omega_2 t) + s_{22}\alpha_2^2 e^{-\frac{1}{2}\alpha_2^2 t} \sin(\omega_2 t) \end{pmatrix} + \begin{pmatrix} s_{11} \\ s_{21} \end{pmatrix} \delta(t). \end{aligned} \quad (\text{D5})$$

By summing the components, we obtain (46). ■

(7) *Proof of Theorem 2.* Define $\bar{s}_i = s_i - m_i$ for $i = 1, 2$. For the second moment, we first compute the expectation with respect to $\mathbf{s} = (s_1, s_2)^\top$:

$$\begin{aligned} &E\left(\alpha^2 e^{-\frac{\alpha^2 t}{2}} (\bar{s}_1 \cos(\omega t) - \bar{s}_2 \sin(\omega t)) (\alpha^2 e^{-\frac{\alpha^2 t'}{2}} (\bar{s}_1 \cos(\omega t') - \bar{s}_2 \sin(\omega t'))\right) \\ &= \alpha^4 e^{-\frac{\alpha^2(t+t')}{2}} E(\bar{\mathbf{s}}^\top \mathbf{c}_{\omega t} \mathbf{c}_{\omega t'}^\top \bar{\mathbf{s}}), \\ &= \alpha^4 e^{-\frac{\alpha^2(t+t')}{2}} \text{Tr}(\mathbf{c}_{\omega t} \mathbf{c}_{\omega t'}^\top E(\mathbf{s} \mathbf{s}^\top)), \\ &= \alpha^4 e^{-\frac{\alpha^2(t+t')}{2}} \text{Tr}\left(\mathbf{c}_{\omega t} \mathbf{c}_{\omega t'}^\top \left(\frac{1-r_2}{2} I_2 + (r_2 - r_1^2) \mathbf{1}_2 \mathbf{1}_2^\top\right)\right) \\ &= \alpha^4 e^{-\frac{\alpha^2(t+t')}{2}} \left(\frac{1-r_2}{2} (\cos(\omega t) \cos(\omega t') + \sin(\omega t) \sin(\omega t')) + (r_2 - r_1^2) (\cos(\omega t) - \sin(\omega t)) (\cos(\omega t') - \sin(\omega t'))\right) \\ &= \alpha^4 e^{-\frac{\alpha^2(t+t')}{2}} \left(\frac{1-r_2}{2} \cos(\omega(t-t')) + (r_2 - r_1^2) (\cos(\omega(t-t')) - \sin(\omega(t+t'))\right) \\ &= \alpha^4 e^{-\frac{\alpha^2(t+t')}{2}} \left(\frac{1+r_2-2r_1^2}{2} \cos(\omega(t-t')) - (r_2 - r_1^2) \sin(\omega(t+t'))\right), \end{aligned} \quad (\text{D6})$$

where $\mathbf{c}_{\omega t} = (\cos(\omega t), -\sin(\omega t))^\top$. By computing the expectation with respect to $\phi_\theta(\alpha, \omega)$,

$$\begin{aligned} & \int \alpha^4 e^{-\frac{\alpha^2(t+t')}{2}} \left(\frac{1+r_2-2r_1^2}{2} \cos(\omega(t-t')) - (r_2-r_1^2) \sin(\omega(t+t')) \right) \phi_\theta(\alpha, \omega) d\alpha d\omega \\ &= \int \alpha^4 e^{-\frac{\alpha^2(t+t')}{2}} \frac{1+r_2-2r_1^2}{2} \cos(\omega(t-t')) \phi_\theta(\alpha, \omega) d\alpha d\omega, \end{aligned} \quad (\text{D7})$$

where we have exploited that $\phi_\theta(\alpha, \omega) = \phi_\theta(\alpha, -\omega)$ and $\sin(x) = -\sin(-x)$. Therefore, from the central limit theorem, we have that

$$\sqrt{n_c} \left(\frac{\tilde{h}_j(t)}{n_c} \right) \xrightarrow{\text{dist.}} \mathcal{GP}(0, K_\theta(t, t')). \quad (\text{D8})$$

(8) *Proof of Corollary 2.* Assuming that $\phi_\theta(\alpha^2, \omega) = \frac{\frac{\alpha^2}{2}}{\pi(\omega^2 + \frac{\alpha^2}{4})} \text{Unif}(\alpha^2; \alpha_m, \alpha_M)$ and exploiting the result:

$$\int \cos(\omega\tau) \frac{1}{\pi} \frac{\beta}{\omega^2 + \beta^2} d\omega = e^{-\beta|\tau|}, \quad (\text{D9})$$

then, for $\beta = \alpha^2/2$,

$$\int \alpha^4 e^{-\frac{\alpha^2}{2}(t+s)} e^{-\frac{\alpha^2}{2}|t-s|} \text{Unif}(\alpha^2; \alpha_m, \alpha_M) d\alpha^2 = \frac{(1 + a_m\tau + \frac{1}{2}a_m^2\tau^2)e^{-a_m\tau} - e^{-a_M\tau}(1 + a_M\tau + \frac{1}{2}a_M^2\tau^2)}{2\tau^3(a_M - a_m)}, \quad (\text{D10})$$

where $\tau = \max(s, t)$. ■

(9) *Proof of Proposition 5.* Recall that

$$\tilde{y}_j(t) = \int_0^t \tilde{h}_j(t-\tau) u(\tau) d\tau = - \int_0^t \sum_{i=1}^{n_c} \alpha_i^2 e^{-\frac{\alpha_i^2(t-\tau)}{2}} (s_{i1}^{(j)} \cos(\omega_i(t-\tau)) - s_{i2}^{(j)} \sin(\omega_i(t-\tau))) u(\tau) d\tau$$

and define

$$\bar{y}_j(t) = - \int_0^t \sum_{i=1}^{n_c} \alpha_i^2 e^{-\frac{\alpha_i^2(t-\tau)}{2}} ((s_{i1}^{(j)} - m_1) \cos(\omega_i(t-\tau)) - (s_{i2}^{(j)} - m_2) \sin(\omega_i(t-\tau))) u(\tau) d\tau,$$

where $m_1 = E[s_{11}]$ and $m_2 = E[s_{12}]$. Then, we have that

$$\begin{aligned} \bar{\mathbf{y}}^\top(t) \bar{\mathbf{y}}(t') &= \frac{1}{dn_c} \sum_{j=1}^d \int_0^t \int_0^{t'} \left(\sum_{i=1}^{n_c} \alpha_i^2 e^{-\frac{\alpha_i^2(t-\tau)}{2}} (\bar{s}_{i1}^{(j)} \cos(\omega_i(t-\tau)) - \bar{s}_{i2}^{(j)} \sin(\omega_i(t-\tau))) u(\tau) d\tau \right) \\ &\quad \times \left(\sum_{k=1}^{n_c} \alpha_k^2 e^{-\frac{\alpha_k^2(t'-\tau')}{2}} (\bar{s}_{k1}^{(j)} \cos(\omega_k(t'-\tau')) - \bar{s}_{k2}^{(j)} \sin(\omega_k(t'-\tau'))) u(\tau') d\tau' \right), \end{aligned}$$

where $\bar{s}_{i1} = s_{i1} - m_1$ and $\bar{s}_{i2} = s_{i2} - m_2$. Note that

$$\begin{aligned} & (\bar{s}_{i1}^{(j)} \cos(\omega_i(t-\tau)) - \bar{s}_{i2}^{(j)} \sin(\omega_i(t-\tau))) u(\tau) d\tau \cdot (\bar{s}_{k1}^{(j)} \cos(\omega_k(t'-\tau')) - \bar{s}_{k2}^{(j)} \sin(\omega_k(t'-\tau'))) u(\tau') d\tau' \\ &= \bar{s}_{i1}^{(j)} \bar{s}_{k1}^{(j)} \cos(\omega_i(t-\tau)) \cos(\omega_k(t'-\tau')) + \bar{s}_{i2}^{(j)} \bar{s}_{k2}^{(j)} \sin(\omega_i(t-\tau)) \sin(\omega_k(t'-\tau')) \\ &\quad - \bar{s}_{i1}^{(j)} \bar{s}_{k2}^{(j)} \cos(\omega_i(t-\tau)) \sin(\omega_k(t'-\tau')) - \bar{s}_{i2}^{(j)} \bar{s}_{k1}^{(j)} \sin(\omega_i(t-\tau)) \cos(\omega_k(t'-\tau')). \end{aligned}$$

We can then rewrite

$$\begin{aligned} \bar{\mathbf{y}}^\top(t) \bar{\mathbf{y}}(t') &= \frac{1}{n_c} \int_0^t \int_0^{t'} u(\tau) d\tau u(\tau') d\tau' \sum_{i=1}^{n_c} \sum_{k=1}^{n_c} \alpha_i^2 e^{-\frac{\alpha_i^2(t-\tau)}{2}} \alpha_k^2 e^{-\frac{\alpha_k^2(t'-\tau')}{2}} \frac{1}{d} \sum_{j=1}^d (\bar{s}_{i1}^{(j)} \bar{s}_{k1}^{(j)} \cos(\omega_i(t-\tau)) \cos(\omega_k(t'-\tau')) \\ &\quad + \bar{s}_{i2}^{(j)} \bar{s}_{k2}^{(j)} \sin(\omega_i(t-\tau)) \sin(\omega_k(t'-\tau')) - \bar{s}_{i1}^{(j)} \bar{s}_{k2}^{(j)} \cos(\omega_i(t-\tau)) \sin(\omega_k(t'-\tau')) \\ &\quad - \bar{s}_{i2}^{(j)} \bar{s}_{k1}^{(j)} \sin(\omega_i(t-\tau)) \cos(\omega_k(t'-\tau'))), \end{aligned}$$

which is equal to

$$\begin{aligned}\bar{\mathbf{y}}^\top(t)\bar{\mathbf{y}}(t') &= \frac{1}{n_c} \int_0^t \int_0^{t'} u(\tau) d\tau u(\tau') d\tau' \sum_{i=1}^{n_c} \alpha_i^4 e^{-\frac{\alpha_i^2(t-\tau+t'-\tau')}{2}} \frac{1}{d} \sum_{j=1}^d (\bar{s}_{i1}^{(j)} \bar{s}_{i1}^{(j)} \cos(\omega_i(t-\tau)) \cos(\omega_i(t'-\tau')) \\ &\quad + \bar{s}_{i2}^{(j)} \bar{s}_{i2}^{(j)} \sin(\omega_i(t-\tau)) \sin(\omega_i(t'-\tau')) - \bar{s}_{i1}^{(j)} \bar{s}_{i2}^{(j)} \cos(\omega_i(t-\tau)) \sin(\omega_i(t'-\tau')) \\ &\quad - \bar{s}_{i2}^{(j)} \bar{s}_{i1}^{(j)} \sin(\omega_i(t-\tau)) \cos(\omega_i(t'-\tau')))\end{aligned}$$

as $\frac{1}{d} \sum_{j=1}^d \bar{s}_i^{(j)} \bar{s}_k^{(j)} = 0$ for $i \neq k$ —they are independent variables with zero mean.

Therefore, we have that

$$\begin{aligned}\bar{\mathbf{y}}^\top(t)\bar{\mathbf{y}}(t') &= \frac{1}{n_c} \int_0^t \int_0^{t'} u(\tau) d\tau u(\tau') d\tau' \sum_{i=1}^{n_c} \alpha_i^4 e^{-\frac{\alpha_i^2(t-\tau+t'-\tau')}{2}} \text{Tr} \left(\mathbf{c}_{\omega_i(t-\tau)} \mathbf{c}_{\omega_i(t'-\tau')}^\top \left(\frac{1-r_2}{2} I_2 + (r_2-r_1^2) \mathbf{1}_2 \mathbf{1}_2^\top \right) \right) \\ &= \frac{1}{n_c} \int_0^t \int_0^{t'} u(\tau) d\tau u(\tau') d\tau' \sum_{i=1}^{n_c} \alpha_i^4 e^{-\frac{\alpha_i^2(t-\tau+t'-\tau')}{2}} \left(\frac{1+r_2^2-2r_1^2}{2} \cos(\omega_i(t-\tau-(t'-\tau'))) \right. \\ &\quad \left. - (r_2-r_1^2) \sin(\omega_i(t-\tau+(t'-\tau'))) \right),\end{aligned}$$

where $\mathbf{c}_{\omega t} = (\cos(\omega(t-\tau)), -\sin(\omega(t-\tau)))^\top$. For large n_c , this results in

$$\begin{aligned}\bar{\mathbf{y}}^\top(t)\bar{\mathbf{y}}(t') &\approx \int_0^t \int_0^{t'} u(\tau) d\tau u(\tau') d\tau' \int \alpha^4 e^{-\frac{\alpha^2(t-\tau+t'-\tau')}{2}} \left(\frac{1+r_2^2-2r_1^2}{2} \cos(\omega(t-\tau-(t'-\tau'))) \right. \\ &\quad \left. - (r_2-r_1^2) \sin(\omega(t-\tau+(t'-\tau'))) \right) \phi_\theta(\alpha, \omega) d\alpha d\omega \\ &= \int_0^t \int_0^{t'} u(\tau) d\tau u(\tau') d\tau' \int \int \alpha^4 e^{-\frac{\alpha^2(t-\tau+t'-\tau')}{2}} \frac{1+r_2^2-2r_1^2}{2} \cos(\omega(t-\tau-(t'-\tau'))) \phi_\theta(\alpha, \omega) d\alpha d\omega,\end{aligned}$$

which is $K_\theta(t-\tau, t'-\tau')$.

(10) *Proof of Lemma 3* We need to compute $C^{(r)} e^{A^{(r)} t} L^{(r)}$. Note that

$$\begin{aligned}e^{A^{(r)} t} &= \begin{pmatrix} e^{At} & 0 \\ 0 & e^{Zt} \end{pmatrix} L^{(r)} = J_n^{(r)} C^{(r)\top} J_s^{(r)} \\ &= \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix} \begin{pmatrix} C_1^{(r)} & 0_{1 \times n} \\ 0_{1 \times n} & C_2^{(r)} \end{pmatrix}^\top \begin{pmatrix} 0 & I_s \\ -I_s & 0 \end{pmatrix} \\ &= \begin{pmatrix} -C_2^{(r)\top} & 0_{n \times 1} \\ 0_{n \times 1} & -C_1^{(r)\top} \end{pmatrix}.\end{aligned}$$

Therefore, we have that

$$\begin{aligned}C^{(r)} e^{A^{(r)} t} L^{(r)} &= \begin{pmatrix} C_1^{(r)} & 0_{1 \times n} \\ 0_{1 \times n} & C_2^{(r)} \end{pmatrix} \begin{pmatrix} e^{At} & 0 \\ 0 & e^{Zt} \end{pmatrix} \begin{pmatrix} -C_2^{(r)\top} & 0_{n \times 1} \\ 0_{n \times 1} & -C_1^{(r)\top} \end{pmatrix} \\ &= \begin{pmatrix} -C_1^{(r)} e^{At} C_2^{(r)\top} & 0 \\ 0 & -C_2^{(r)} e^{Zt} C_1^{(r)\top} \end{pmatrix}.\end{aligned}$$

It remains to prove that the QSS is physically realizable. The physical realizability conditions is as follows:

$$\begin{aligned}0 &= A^{(r)} J_n^{(r)} + J_n^{(r)} A^{(r)\top} + L^{(r)} J_s^{(r)} L^{(r)\top} \\ &= \begin{pmatrix} A & 0 \\ 0 & Z \end{pmatrix} \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix} + \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix} \begin{pmatrix} A^\top & 0 \\ 0 & Z^\top \end{pmatrix}\end{aligned}$$

$$\begin{aligned}&+ L^{(r)} \begin{pmatrix} 0 & I_2 \\ -I_2 & 0 \end{pmatrix} L^{(r)\top} \\ &= \begin{pmatrix} 0 & A+Z^\top \\ -Z-A^\top & 0 \end{pmatrix} \\ &\quad + \begin{pmatrix} 0 & -C_2^{(r)\top} \\ C_1^{(r)\top} & 0 \end{pmatrix} \begin{pmatrix} -C_2^{(r)\top} & 0 \\ 0 & -C_1^{(r)\top} \end{pmatrix}^\top \\ &= \begin{pmatrix} 0 & A+Z^\top + C_2^{(r)\top} C_1^{(r)} \\ -Z-A^\top - C_1^{(r)\top} C_2^{(r)} & 0 \end{pmatrix},\end{aligned}$$

which implies that $Z = -A^\top - C_1^{(r)\top} C_2^{(r)}$.

The condition $J_n^{(r)} L^{(r)} + C^{(r)\top} J_s^{(r)} D^{(r)} = 0$ is satisfied as $D^{(r)} = I_4$ and $L^{(r)} = J_n^{(r)} C^{(r)\top} J_s^{(r)}$ and so

$$J_n^{(r)} L^{(r)} + C^{(r)\top} J_s^{(r)} D^{(r)} = -C^{(r)\top} J_s^{(r)} + C^{(r)\top} J_s^{(r)} = 0.$$

Finally, the following condition is satisfied as well: $D^{(r)} J_s^{(r)} D^{(r)\top} - J_s^{(r)} = 0$. ■

(11) *Proof of Theorem 3.* The first part of the proof follows directly from Lemma 3. About the second part, it follows from a standard result in control theory for LTI systems, which states that if the matrix $(B, AB, A^2B, \dots, A^{n-1}B)$ has rank n , then there exists a matrix R such that the closed-loop matrix $A - BR$ is stable [[65], Chap. 5]. A similar result holds for $-A - BR$, since $(B, -AB, (-A)^2B, \dots, (-A)^{n-1}B)$ has also rank n . ■

TABLE III. For each benchmark, in the first row we report the median RMSE for (i) the QRC LqW and (ii) our QRC HqW, using a polynomial nonlinearity (with degree 3) in the readout layer. The second row includes the 25th and 75th RMSE percentile between square brackets. The number between brackets after the name of the model is relative to the number of harmonic oscillators in the quantum linear network.

RMSE	LqW(24)	HqW(24)
Parity($\tau = 2$)	0.12 [0.10,0.15]	0.0031 [0.0025,0.0038]
Parity($\tau = 4$)	0.56 [0.54,0.59]	0.015 [0.011,0.018]
NARMA10	0.13 [0.12,0.15]	0.065 [0.060,0.080]
Delay($\tau = 2$)	0.40 [0.39,0.43]	0.07 [0.048,0.11]
Delay($\tau = 4$)	0.62 [0.57,0.67]	0.25 [0.18,0.31]

(12) *Proof of Corollary 3.* We note that the $C^{(r)}$ matrix is

$$C^{(r)} = \begin{pmatrix} R & 0_{1 \times n} \\ -C & 0_{1 \times n} \\ 0_{1 \times n} & B^T \\ 0_{1 \times n} & 0_{1 \times n} \end{pmatrix}.$$

TABLE IV. $N_{\text{meas}} = 500$. For each benchmark, in the first row we report the median RMSE for (i) the classical RC ESN, (ii) the QRC LqW, and (iii) our QRC HqW. The second row includes the 25th and 75th RMSE percentile between square brackets. The number between brackets after the name of the model is relative to the number of harmonic oscillators in the quantum linear network.

RMSE	HqW(8)	HqW(16)	HqW(24)	PadeqW(14)
Parity($\tau = 2$)	0.0034 [0.0023,0.0040]	0.0027 [0.0023,0.0035]	0.0025 [0.0022,0.0032]	0.0015 [0.0013,0.0017]
Parity($\tau = 4$)	0.049 [0.033,0.090]	0.033 [0.023,0.042]	0.027 [0.022,0.031]	0.015 [0.014,0.020]
NARMA10	0.078 [0.070,0.090]	0.071 [0.068,0.081]	0.066 [0.064,0.070]	0.055 [0.050,0.059]
Delay($\tau = 2$)	0.012 [0.0067,0.020]	0.0075 [0.0059,0.0111]	0.0066 [0.0060,0.0075]	0.0014 [0.0012,0.0017]
Delay($\tau = 4$)	0.17 [0.13,0.21]	0.095 [0.076,0.119]	0.078 [0.066,0.081]	0.0014 [0.0014,0.0020]

We aim to write it as

$$C^{(r)} = \begin{pmatrix} \text{Re}(N_1 + N_2) & \text{Im}(-N_1 + N_2) \\ \text{Im}(N_1 + N_2) & \text{Re}(N_1 - N_2) \end{pmatrix}$$

with $N_1, N_2 \in \mathbb{C}^{2 \times n}$. It can easily be seen that this requires that $\text{Re}(N_2) \neq 0_{2 \times n}$. Therefore, from Definition 2, it follows that the QSS described by (53)–(56) is not completely passive.

APPENDIX E: ADDITIONAL NUMERICAL EXPERIMENTS

Table III reports the comparison between LqW(24) and HqW(24) using a polynomial nonlinearity (with degree 3) in the readout layer instead of the square-exponential kernel (which we instead used in the experiments in the last two columns of Table I). It can be noticed that HqW(24) outperforms LqW(24) also in this case. The performance with the polynomial nonlinearity is generally worse than the one obtained with the squared-exponential kernel in Table I.

Instead, Table IV reports the performance of HqW and PadeqW for $N_{\text{meas}} = 500$. It can be observed that both architectures still achieve good performance even in this higher-noise regime.

- [1] H. Jaeger, The “Echo state” Approach to Analysing and Training Recurrent Neural Networks—With an Erratum Note, GMD Technical Report, German National Research Center for information Technology (2001).
- [2] W. Maass, T. Natschlager, and H. Markram, Real-time computing without stable states: A new framework for neural computation based on perturbations, *Neural Comput.* **14**, 2531 (2002).
- [3] K. Fujii and K. Nakajima, Harnessing disordered-ensemble quantum dynamics for machine learning, *Phys. Rev. Appl.* **8**, 024030 (2017).
- [4] K. Nakajima, K. Fujii, M. Negoro, K. Mitarai, and M. Kitagawa, Boosting computational power through spatial mul-

tiplexing in quantum reservoir computing, *Phys. Rev. Appl.* **11**, 034021 (2019).

- [5] A. Kutvonen, K. Fujii, and T. Sagawa, Optimizing a quantum reservoir computer for time series prediction, *Sci. Rep.* **10**, 14687 (2020).
- [6] J. Chen, H. I. Nurdin, and N. Yamamoto, Temporal information processing on noisy quantum computers, *Phys. Rev. Appl.* **14**, 024065 (2020).
- [7] J. Nokkala, R. Martinez-Pena, 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).

- [8] P. Mujal, R. Martínez-Peña, J. Nokkala, J. García-Beni, G. L. Giorgi, M. C. Soriano, and R. Zambrini, Opportunities in quantum reservoir computing and extreme learning machines, *Adv. Quantum Technol.* **4**, 2100027 (2021).
- [9] R. Martínez-Peña, J. Nokkala, G. L. Giorgi, R. Zambrini, and M. C. Soriano, Information processing capacity of spin-based quantum reservoir computing systems, *Cognit. Comput.* **15**, 1440 (2023).
- [10] F. Hu, S. A. Khan, N. T. Bronn, G. Angelatos, G. E. Rowlands, G. J. Ribeill, and H. E. Türeci, Overcoming the coherence time barrier in quantum machine learning on temporal data, *Nat. Commun.* **15**, 7491 (2024).
- [11] I. Paparelle, J. Henaff, J. García-Beni, É. Gillet, D. Montesinos, G. L. Giorgi, M. C. Soriano, R. Zambrini, and V. Parigi, Experimental memory control in continuous-variable optical quantum reservoir computing, *Nat. Photon.* **20**, 413 (2026).
- [12] M. Selimović, I. Agresti, M. Siemaszko, J. Morris, B. Dakić, R. Albiero, A. Crespi, F. Ceccarelli, R. Osellame, M. Stobińska *et al.*, Experimental neuromorphic computing based on quantum memristor, [arXiv:2504.18694](https://arxiv.org/abs/2504.18694).
- [13] S. Ghosh, A. Opala, M. Matuszewski, T. Paterek, and T. C. Liew, Quantum reservoir processing, *npj Quantum Inf.* **5**, 35 (2019).
- [14] D. Marković and J. Grollier, Quantum neuromorphic computing, *Appl. Phys. Lett.* **117**, 150501 (2020).
- [15] R. A. Bravo, K. Najafi, X. Gao, and S. F. Yelin, Quantum reservoir computing using arrays of Rydberg atoms, *PRX Quantum* **3**, 030325 (2022).
- [16] A. Sakurai, M. P. Estarellas, W. J. Munro, and K. Nemoto, Quantum extreme reservoir computation utilizing scale-free networks, *Phys. Rev. Appl.* **17**, 064044 (2022).
- [17] J. Nokkala, Online quantum time series processing with random oscillator networks, *Sci. Rep.* **13**, 7694 (2023).
- [18] 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).
- [19] A. Hayashi, A. Sakurai, S. Nishio, W. J. Munro, and K. Nemoto, Impact of the form of weighted networks on the quantum extreme reservoir computation, *Phys. Rev. A* **108**, 042609 (2023).
- [20] L. Grigoryeva and J.-P. Ortega, Echo state networks are universal, *Neural Netw.* **108**, 495 (2018).
- [21] S. Boyd and L. Chua, Fading memory and the problem of approximating nonlinear operators with Volterra series, *IEEE Trans. Circuit Syst.* **32**, 1150 (1985).
- [22] V. Volterra and E. Whittaker, *Theory of Functionals and of Integral and Integro-differential Equations* (Dover, New York, 1931).
- [23] The extension to m -dimensional multi-input systems is obtained by combining m such LTI architectures in parallel (one for each input) and feed their outputs to the polynomial readout.
- [24] N. Wiener, *Nonlinear Problems in Random Theory* (MIT Press, Cambridge, MA, 1958).
- [25] J. García-Beni, G. L. Giorgi, M. C. Soriano, and R. Zambrini, Scalable photonic platform for real-time quantum reservoir computing, *Phys. Rev. Appl.* **20**, 014051 (2023).
- [26] J. Nokkala, G. L. Giorgi, and R. Zambrini, Retrieving past quantum features with deep hybrid classical-quantum reservoir computing, *Mach. Learn.: Sci. Technol.* **5**, 035022 (2024).
- [27] P. Mujal, R. Martínez-Peña, G. L. Giorgi, M. C. Soriano, and R. Zambrini, Time-series quantum reservoir computing with weak and projective measurements, *npj Quantum Inf.* **9**, 16 (2023).
- [28] T. Yasuda, Y. Suzuki, T. Kubota, K. Nakajima, Q. Gao, W. Zhang, S. Shimono, H. I. Nurdin, and N. Yamamoto, Quantum reservoir computing with repeated measurements on superconducting devices, [arXiv:2310.06706](https://arxiv.org/abs/2310.06706).
- [29] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang, The expressive power of neural networks: A view from the width, in *Proceedings of the 31st International Conference on Neural Information Processing Systems* (ACM, New York, 2017).
- [30] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, MA, 2006), Vol. 2.
- [31] A. G. Wilson, Z. Hu, R. Salakhutdinov, and E. P. Xing, Deep kernel learning, in *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, edited by A. Gretton and C. C. Robert (PMLR, Cadiz, Spain, 2016), Vol. 51, pp. 370–378.
- [32] V. Belavkin, Quantum filtering of Markov signals with white quantum noise, in *Quantum Communications and Measurement* (Springer, Berlin, 1995), pp. 381–391.
- [33] A. Voelker, I. Kajić, and C. Eliasmith, Legendre memory units: Continuous-time representation in recurrent neural networks, in *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (ACM, New York, 2019).
- [34] Here we considered a deterministic, classical, stable SISO LTI system. For a stochastic LTI system with classical Wiener noise, the output $y(t)$ becomes stochastic. In this case, the convolution $\int_0^t h(t - \tau)u(\tau)d\tau$ coincides with the mean of $y(t)$.
- [35] We do not consider here the case where A has eigenvalues with multiplicity greater than one.
- [36] M. Zorzi and A. Chiuso, The harmonic analysis of kernel functions, *Automatica* **94**, 125 (2018).
- [37] Lemma 1 relies on a different convergence result than that originally used in Ref. [36]. Specifically, we employ the strong law of large numbers of Monte Carlo integration, whereas Ref. [36] relies on the convergence of Riemann sums.
- [38] A. Wilson and R. Adams, Gaussian process kernels for pattern discovery and extrapolation, in *Proceedings of the 30th International Conference on Machine Learning*, edited by S. Dasgupta and D. McAllester (PMLR, Atlanta, Georgia, USA, 2013), Vol. 28, pp. 1067–1075.
- [39] T. Chen, H. Ohlsson, and L. Ljung, On the estimation of transfer functions, regularizations and Gaussian processes—Revisited, *Automatica* **48**, 1525 (2012).
- [40] G. Pillonetto, T. Chen, A. Chiuso, G. De Nicolao, and L. Ljung, Regularized linear system identification using atomic, nuclear and kernel-based norms: The role of the stability constraint, *Automatica* **69**, 137 (2016).
- [41] V. I. Paulsen and M. Raghupathi, *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces* (Cambridge University Press, Cambridge, UK, 2016), Vol. 152.
- [42] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, Kernel methods in system identification, machine learning and function estimation: A survey, *Automatica* **50**, 657 (2014).
- [43] The finite case is also of interest in machine learning as technique in kernel approximation for scalability of kernel methods [66].

- [44] C. P. Robert, G. Casella, and G. Casella, *Monte Carlo Statistical Methods* (Springer, Berlin, 2004), Vol. 2.
- [45] Y. Cho and L. Saul, Kernel methods for deep learning, in *Proceedings of the 23rd International Conference on Neural Information Processing Systems* (ACM, New York, 2009).
- [46] R. Simon, N. Mukunda, and B. Dutta, Quantum-noise matrix for multimode systems: $U(n)$ invariance, squeezing, and normal forms, *Phys. Rev. A* **49**, 1567 (1994).
- [47] H. I. Nurdin and N. Yamamoto, Linear dynamical quantum systems, in *Analysis, Synthesis, and Control* (Springer, Berlin, 2017).
- [48] G. Zhang and Z. Dong, Linear quantum systems: A tutorial, *Annu. Rev. Contr.* **54**, 274 (2022).
- [49] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches* (John Wiley & Sons, New York, 2006).
- [50] J. Gough and M. R. James, The series product and its application to quantum feedforward and feedback networks, *IEEE Trans. Autom. Control* **54**, 2530 (2009).
- [51] C. K. Williams, Computation with infinite neural networks, *Neural Comput.* **10**, 1203 (1998).
- [52] R. M. Neal, *Bayesian Learning for Neural Networks* (Springer Science & Business Media, New York, 2012), Vol. 118.
- [53] M. Green and D. J. Limebeer, *Linear Robust Control* (Courier Corporation, North Chelmsford, MA, 2012).
- [54] C. W. Gardiner and P. Zoller, *Quantum Noise: A Handbook of Markovian and non-Markovian Quantum Stochastic Methods with Applications to Quantum Optics*, 2nd enlarged ed. (Springer, Berlin, 2000).
- [55] H. I. Nurdin, M. R. James, and A. C. Doherty, Network synthesis of linear dynamical quantum stochastic systems, *SIAM J. Control Optim.* **48**, 2686 (2009).
- [56] C. Wringe, M. Trefzer, and S. Stepney, Reservoir computing benchmarks: A tutorial review and critique, *Int. J. Parallel Emerg. Distrib. Syst.* **40**, 313 (2025).
- [57] J. García-Bení, I. Paparella, V. Parigi, G. L. Giorgi, M. C. Soriano, and R. Zambrini, Quantum machine learning via continuous-variable cluster states and teleportation, *EPJ Quantum Technol.* **12**, 63 (2025).
- [58] W. A. Barbosa, A. Griffith, G. E. Rowlands, L. C. Govia, G. J. Ribeill, M.-H. Nguyen, T. A. Ohki, and D. J. Gauthier, Symmetry-aware reservoir computing, *Phys. Rev. E* **104**, 045307 (2021).
- [59] M. Forgione and D. Piga, Dynonet: A neural network architecture for learning dynamical systems, *Int. J. Adapt. Contr. Sign. Process.* **35**, 612 (2021).
- [60] R. de Oliveira Teloli, L. G. Villani, S. da Silva, and M. D. Todd, On the use of the GP-NARX model for predicting hysteresis effects of bolted joint structures, *Mech. Syst. Sig. Process.* **159**, 107751 (2021).
- [61] A. Benavoli, D. Piga, M. Forgione, and M. Zaffalon, Dynogp: Deep Gaussian processes for dynamic system identification, [arXiv:2502.05620](https://arxiv.org/abs/2502.05620).
- [62] <https://github.com/benavoli/qWiener>.
- [63] T. Hillen, K. J. Painter, A. C. Swan, and A. D. Murtha, Moments of von Mises and Fisher distributions and applications, *Math. Biosci. Eng.* **14**, 673 (2017).
- [64] K. Koga and N. Yamamoto, Dissipation-induced pure Gaussian state, *Phys. Rev. A* **85**, 022103 (2012).
- [65] K. J. Åström and R. M. Murray, *Analysis and Design of Feedback Systems* (Princeton University Press, Princeton, NJ, 2004).
- [66] A. Rahimi and B. Recht, Random features for large-scale kernel machines, in *Proceedings of the 21st International Conference on Neural Information Processing Systems* (ACM, New York, 2007).