Engineering quantum reservoirs through Krylov complexity, expressivity, and observability

Saud Čindrak,* Lina Jaurigue, and Kathy Lüdge
Institute of Physics, Technische Universität Ilmenau, Ilmenau, Germany

(Received 13 February 2025; accepted 22 August 2025; published 20 November 2025)

This study employs Krylov-based information measures to understand task performance in quantum reservoir computing, a subfield of quantum machine learning. In our study, we show that fidelity and spread complexity can only explain the task performance for short time evolutions of the quantum systems. We then discuss two measures, Krylov expressivity and Krylov observability, and compare them to task performance and the information processing capacity. Our results show that Krylov observability exhibits almost identical behavior to information processing capacity, while being three orders of magnitudes faster to compute. In the case when the system is undersampled, Krylov observability best captures the behavior of the task performance.

DOI: 10.1103/84f3-63mz

I. INTRODUCTION

Quantum machine learning (QML) models attempt to exploit the exponential scaling of Hilbert space for computation. A subfield of QML that has garnered significant attention in physics is quantum reservoir computing (QRC). QRC is inspired by classical reservoir computing, where physical systems are used to solve machine learning tasks [1–7]. QRC research, particularly with the Ising model, has been explored in Refs. [8–15], with initial implementations on IBM's quantum processor [16–19]. It has been shown that quantum reservoir computing can utilize noise and dissipation as resources [20–22]. To address the time complexity problem for time-series tasks, weak measurements [23,24] and reinitialization schemes [25] have been proposed, reducing complexity and improving performance. Other approaches include reintroducing measured outputs [26,27] and continuous measurements [28], while Ref. [29] utilizes techniques from parametrized quantum circuits to define expressive limits in QRC. However, understanding in quantum machine learning and quantum reservoir computing remains limited. In classical machine learning, this understanding arises from expressivity and explainability measures, which is still an active field and is mostly unresolved in quantum machine learning. A key challenge is comparing quantum systems described by unitaries or Hamiltonians based on their expressivity within the Hilbert space. An ideal information measure would provide insights into task performance while offering a physical interpretation of the system, making it physically interpretable.

In this work, we discuss various measures regarding task performance in quantum reservoir computing. We train four distinct quantum reservoirs on a chaotic time series and larger reservoirs with random interspin couplings. As a benchmark, the Lorenz63 system is used [30]. We compare fidelity, information processing capacity [31], and spread complexity [32,33] as established methods. Additionally, we discuss Krylov expressivity and Krylov observability, two measures designed to capture the effective phase-space dimension of the Hilbert space [34,35]. We demonstrate that Krylov observability and information processing capacity perform best in explaining the behavior of task performance.

Spread and operator complexity have successfully captured various effects of time evolution in relation to spin chains, Sachdev–Ye–Kitaev models, chaotic quantum systems, driven quantum systems, and others [32,33,36–85]. A detailed explanation on the methods is given in Ref. [86]. However, a major challenge in calculating spread and operator complexity is the necessity of classical simulation for constructing the Krylov spaces. In Ref. [35], we address this by introducing quantum-mechanically measurable Krylov spaces for the computation of spread complexity and propose *Krylov expressivity*, a measurable expressivity measure of the effective phase-space dimension. In Ref. [34], we extend this research by proposing *Krylov observability*, a measure to quantify the effective phase-space dimension of various observables measured multiple times. Our results show that Krylov observability effectively captures how well a quantum system can retain and map macroscopic data nonlinearly, thus highlighting the importance of Krylov spaces in quantum reservoir computing.

The authors in Ref. [87] first explored the insights that spread complexity can provide in the field of quantum reservoir computing. The authors found that quantum reservoirs with larger mean spread complexity tended to perform better.

In Ref. [88], the authors discuss how state estimation in a quantum extreme learning machine can occur beyond the scrambling time, which challenges the common belief that information cannot be retrieved after this point. A similar pattern is observed in quantum reservoir computing, where task performance initially increases and then saturates [89]. We show that fidelity and spread complexity can only explain the initial change in task performance, but fail to explain the

*Contact author: saud.cindrak@tu-ilmenau.de

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

saturation in task performance. Our work explains this saturation using the concepts of Krylov expressivity and Krylov observability. The initial state or data are mapped onto the Krylov space, where expressivity and observability increase before reaching a saturation point. Beyond this point, further increases in time do not enhance the system's expressivity, resulting in stable task performance.

Among all measures, Krylov observability and information processing capacity explain the general trend of task performance best and show almost identical behavior for larger readout dimensions. For small readout dimensions, we show that Krylov observability performs better than information processing capacity in capturing the behavior of the Lorenz task. This is due to the fact that information processing capacity is upper bounded by the readout dimension, while Krylov observability adapts better to a smaller number of measurements. Our results suggest that Krylov expressivity is not as insightful as Krylov observability in quantum reservoir computing. This is because Krylov expressivity only considers how input states are mapped onto the Krylov space. However, if the initial state is sampled from an encoding unitary, Krylov expressivity can further the understanding of different encoding strategies. Two subfields of quantum machine learning with various encoding strategies are variational quantum machine learning and quantum extreme learning [90–114]. Krylov observability, on the other hand, can provide insights into how much of the Krylov space of a variational quantum algorithm can be sampled. Current research on the explainability of quantum machine learning is sparse and primarily employs techniques from classical machine learning in QML [115–118], while Ref. [119] explores the possibilities of designing explanation techniques for parametrized quantum circuits. We further give a brief discussion of the number of matrix operations required for the computation of Krylov observability (N_{obs}) and for the computation of the state matrix ($N_{\mathbf{u}}$), which is required for the information processing capacity (IPC). The ratio of these two scales with $r = N_{\text{obs}}/N_{\text{sys}} \in \mathcal{O}(N_R/N_{\mathbf{u}})$, where $N_{\mathbf{u}}$ is the length of the time series and N_R is the readout dimension. In typical RC approaches, $N_u \gg N_R$ is given, which results in $r \ll 1$ and in our experiments $r = 0.000\,75$. Similar to the discussion about the grade of the Krylov state space presented in Ref. [35], this work introduces a similar identity for Krylov operator spaces, where the grade of the Krylov operator space can be derived using only the spectral properties of the Hamiltonian and the operator. This is of interest because numerical errors during the orthonormalization can artificially increase the apparent dimension of the Krylov space. Knowing the Krylov grade in advance allows for a cutoff based on physical rather than numerical considerations. This understanding is then used to explain why the operator Krylov space can be smaller for certain Hamiltonians, even when their corresponding Krylov state spaces are larger than those of others.

This work is organized as follows. Section II introduces quantum reservoir computing, information processing capacity, and the Lorenz task. Section III discusses Krylov spaces in quantum mechanics. Here, we discuss that any time-evolved state lies within a Krylov space and explain measurable Krylov spaces, leading to the introduction of Krylov expressivity. This concept is extended to operator complexity, along with a brief discussion of Krylov observability as proposed in Ref. [34]. In Sec. IV, we demonstrate the limitations of fidelity and spread complexity in explaining task performance due to their oscillatory behavior, even though Lorenz task performance remains mostly constant. We show that Hamiltonians with higher expressivity can exhibit worse performance when Krylov observability is lower. We conclude our work with a discussion of the measures presented.

II. QUANTUM RESERVOIR COMPUTING

Quantum reservoir computing makes use of a quantum system as a reservoir for classical or quantum tasks. Figure 1 shows a sketch of a quantum reservoir scheme and is explained in the following:

(1) *Initialization*. The reservoir is first initialized with an initial time series, in our case of length $N_{\text{In}} = 10\,000$.

(2) *Input layer (green) and encoding (red)*. The encoding of the nth input $u_n \in \{u_1, \ldots, u_{N_u}\}$ into the first qubit leads to the encoded state

$$|\Psi_n\rangle = \sqrt{\frac{1-u_n}{2}}|0\rangle + \sqrt{\frac{1+u_n}{2}}|1\rangle. \tag{1}$$

After encoding, the state of the system is described by

$$\rho_{\text{enc},n} = |\Psi_n\rangle\langle\Psi_n| \otimes \text{Tr}_1(\rho_n). \tag{2}$$

The overwriting of the first qubit at each time step results in loss of information about previous time steps. This is called fading memory property and is a crucial requirement for reservoir computing, as reservoirs exhibit a memory-nonlinearity trade-off [1,2,120–122]. In Ref. [25], this memory-nonlinearity trade-off was shown in quantum reservoir computing.

(3) *Hidden layer (blue)*. The reservoir evolves for a clock cycle T by the unitary evolution

$$U_R = \exp(-iHT).$$

A discrete set of observables $\{Z_k\}_{k=1,\ldots,K}$ is measured to construct the expectation values. The expectation values at time t_j after the nth input are given by

$$\langle Z_k(t_j, u_n)\rangle = \text{Tr}(Z_k e^{-iHt_j}\rho_{\text{enc},n}e^{iHt_j}). \tag{3}$$

A sketch of the evolution of an observable with respect to different inputs is shown above. To increase the readout dimension, each observable is sampled V times for each input u_n at times $t_j = nT + (j+1)\tau$, with $\tau = T/V$. The sketch in Fig. 1 shows a quantum reservoir with four virtual nodes $V = 4$. This results in $N_R = VN_S$ readout nodes for each input u_n, where N_S represents the number of observables measured. Writing the readouts for each input into a row results in the state matrix $\mathbf{S} \in \mathbb{R}^{N_u, N_R}$.

(4) *Output layer (yellow)*. The state matrix is multiplied with readout weights $\mathbf{W}^{\text{out}}$ resulting in the output

$$\mathbf{Y} = \mathbf{S}\mathbf{W}^{\text{out}}. \tag{4}$$

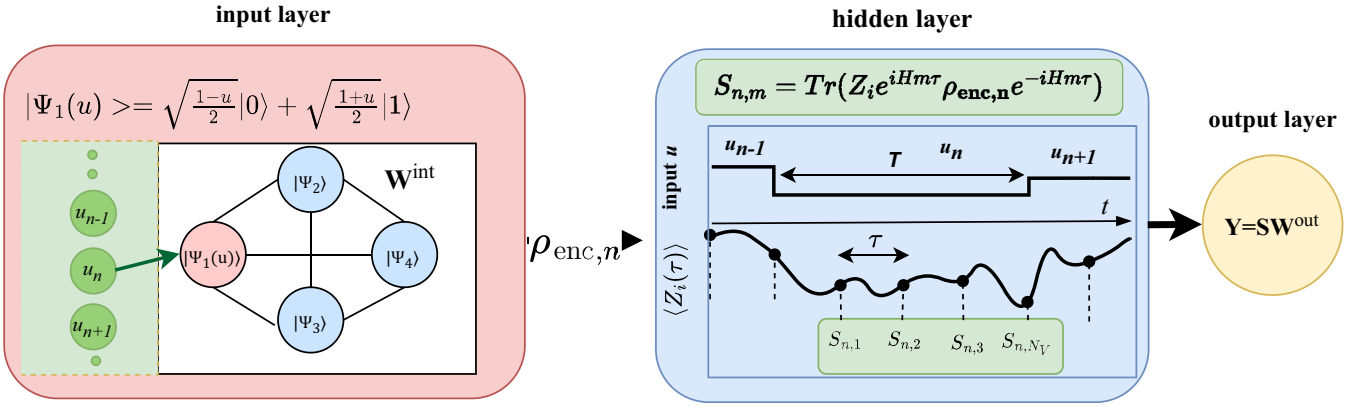


FIG. 1. A time series u_n is encoded into a qubit of a quantum system, described by a Hamiltonian H. The system state is described by the density matrix ρ_{enc} evolving over time t. Measurements of some observables Z_i at times $m\tau$ are taken to construct the state matrix $\mathbf{S}$. The total number of measurements per observable is V and the clock cycle is defined as $V\tau = T$. In the output layer, the state matrix $\mathbf{S}$ is multiplied by the readout weights $\mathbf{W}^{\text{out}}$ to construct the output vector $\mathbf{Y} = \mathbf{S}\mathbf{W}^{\text{out}}$.

(5) *Training*. For training, the data are separated into an initialization set u_{In}, a training set u_{Tr}, and a testing set u_{Te} of lengths N_{In}, N_{Tr}, and N_{Te}, respectively. A buffer of length N_b is added between the sets. We first initialize the system with N_{In} steps; afterwards, the weights are optimized in regard to the training set u_{Tr}, resulting in the trained weights $\mathbf{W}^{\text{out}}$. The testing set is used to check whether the reservoir can successfully generalize to data. The weights are optimized to minimize the loss

$$L := (\mathbf{Y} - \mathbf{Y}^{\text{targ}})^2, \tag{5}$$

where $\mathbf{Y}^{\text{targ}}$ is the target vector. To consider shot noise of the measurement, Gaussian distributed noise $\mathcal{N} \in \mathbb{R}^{N_u, N_R}$ will be added on the state matrix

$$\mathbf{S} \leftarrow \mathbf{S} + \eta\mathcal{N}. \tag{6}$$

The noise term $\eta\mathcal{N}$ can be used as a regularization parameter when computing the state matrix; this is typically addressed as regularization by noise. The readout weights $\mathbf{W}^{\text{out}}$ are then computed by

$$\mathbf{W}^{\text{out}} = \left(\mathbf{S}_{\text{Tr}}^T \mathbf{S}_{\text{Tr}}\right)^{-1} \mathbf{S}_{\text{Tr}}^T \mathbf{Y}^{\text{targ}}. \tag{7}$$

A. Transverse field Ising Hamiltonian

Four transverse field Ising Hamiltonians with different interspin couplings will be used as quantum reservoirs and analyzed in regard to spread complexity, fidelity, Krylov observability, Krylov expressivity, and task performance.

$$H_{I\alpha} = \sum_{i=1, j>i}^{N_S} J_{ij} X_i X_j + \sum_{i=1}^{N_S} h Z_i, \tag{8}$$

where $\alpha \in \{1, 2, 3, 4\}$ is the index referencing each of the four Hamiltonians, N_S is the number of qubits, X_i, Y_i, and Z_i are, respectively, the σ_x, σ_y, and σ_z Pauli matrices of the ith qubit given by

$$X_i, Y_i, Z_i = \left(\bigotimes_{k=1}^{i-1} I_2\right) \otimes \sigma_{x,y,z} \otimes \left(\bigotimes_{k=i+1}^{N_S} I_2\right), \tag{9}$$

and $h = 0.5$ is set for all Hamiltonians. The number of pairwise distinct eigenvalues d and the interspin couplings $J_{i,j}$ are shown in Table I.

B. Lorenz task

The Lorenz63 system is commonly used to construct time series prediction benchmark tasks [30]. The dynamics of the system are governed by Eq. (10):

$$\begin{aligned} \dot{x} &= a(y - x), \\ \dot{y} &= x(b - z) - y, \\ \dot{z} &= xy - cz. \end{aligned} \tag{10}$$

With the parameters set to $a = 10$, $b = 28$, and $c = 8/3$, the Lorenz63 system exhibits chaotic dynamics. We simulate Eq. (10) with an integration time step $dt = 0.001$. To construct the input and target data for the task, we down sample to a discretized step of $\Delta t = 0.02$, which results in the time series $x_n = x(n\Delta t)$, $y_n = y(n\Delta t)$, and $z_n = z(n\Delta t)$. We consider two time series prediction tasks. The first considers the time series x_n as input and tries to predict the pth future step x_{n+p}, referred to as the LXX task. In our study, we choose to

TABLE I. The number of pairwise distinct eigenvalues d and the interspin couplings $J_{i,j}$ for the different Ising reservoirs are shown.

	H_{I1}	H_{I2}	H_{I3}	H_{I4}
No. of eigenvalues d	9	16	15	16
$J_{1,2}$	0.50	0.40	0.35	0.35
$J_{1,3}$	0.50	0.50	0.40	0.40
$J_{1,4}$	0.50	0.50	0.45	0.45
$J_{2,3}$	0.50	0.50	0.50	0.50
$J_{2,4}$	0.50	0.50	0.55	0.55
$J_{3,4}$	0.50	0.50	0.60	0.65

predict $\Delta t = 0.1$ into the future, which results in a $p = 5$-step ahead prediction. The second task takes x_n as input but tries to predict z_n. This task is referred to as the LXZ task or cross-prediction task.

C. Error measure

The normalized root mean squared error (NRMSE) between output Y of the reservoir and target Y^{targ} is defined as

$$\text{NRMSE} = \sqrt{\frac{(\mathbf{Y} - \mathbf{Y}^{\text{targ}})^2}{N\text{var}(\mathbf{Y}^{\text{targ}})}} = \sqrt{1 - C(\mathbf{Y}, \mathbf{Y}^{\text{targ}})}, \tag{11}$$

where the capacity or Pearson correlation is defined as

$$C(\mathbf{Y}, \mathbf{Y}^{\text{targ}}) = \frac{\text{cov}(\mathbf{Y}, \mathbf{Y}^{\text{targ}})}{\sigma^2(\mathbf{Y})\sigma^2(\mathbf{Y}^{\text{targ}})}. \tag{12}$$

$\text{cov}(\cdot)$ and $\sigma(\cdot)$ are the covariance and standard deviation, respectively.

D. Information processing capacity

The IPC is a measure of how well a reservoir can generalize prior data onto a set of orthogonal functions [31]. In this study, Legendre polynomials are used as target functions, which requires the input series to be sampled from a uniform distribution, i.e., $u_n \in \mathcal{U}([-1, 1])$ [123]. The IPC does not predict task performance, but is a measure to gain understanding of the reservoir and the underlying behavior of reservoir computers. Reference [25] showed that the memory-nonlinearity trade-off can be used to increase the nonlinear term of the IPC, which is proved to be beneficial for chaotic time-series prediction tasks.

The first-order IPC, denoted by IPC_1, is also known as linear IPC, memory capacity, or linear short-term memory. It quantifies how well the system can recall previous data [124]. Let $\mathbf{u} = (u_{n_1}, u_{n_1+1}, \ldots, u_{n_2})$ with $n_2 > n_1$. The target is to reconstruct the time series m steps into the past:

$$\mathbf{Y}^{\text{targ}}(-m) = (u_{n_1-m}, u_{n_1+1-m}, \ldots, u_{n_2-m}).$$

The first-order IPC is then defined as

$$\text{IPC}_1 = \sum_m C(\mathbf{Y}, \mathbf{Y}^{\text{targ}}(-m)), \tag{13}$$

where $\mathbf{Y}$ is the output of the reservoir.

In practice, the target is given by

$$\mathbf{Y}^{\text{targ}}(-m) = (l_1(u_{n_1-m}), l_1(u_{n_1+1-m}), \ldots, l_1(u_{n_2-m})),$$

where l_1 is the first-order Legendre polynomial, defined as $l_1(x) = x$. Therefore, the first-order IPC_1 is equivalent to the memory capacity. The Legendre polynomials are defined as

$$l_0(x) = 1, \qquad l_1(x) = x, \qquad l_2(x) = \frac{1}{2}(3x^2 - 1),$$
$$l_n(x) = \frac{2n+1}{n+1} x l_n(x) - \frac{n}{n+1} l_{n-1}(x), \quad \text{for } n \geqslant 2. \tag{14}$$

For readability, we define

$$\mathbf{l}_k(-m) = (l_k(u_{n_1-m}), l_k(u_{n_1+1-m}), \ldots, l_k(u_{n_2-m})).$$

The IPC can be generalized to higher orders. For second order, IPC_2 is computed by

$$\text{IPC}_{2,1} = \sum_m C(\mathbf{Y}, \mathbf{l}_2(-m)),$$
$$\text{IPC}_{2,2} = \sum_{\substack{m_1,m_2 \\ m_1>m_2}} C(\mathbf{Y}, \mathbf{l}_1(-m_1)\mathbf{l}_1(-m_2)),$$
$$\text{IPC}_2 = \text{IPC}_{2,1} + \text{IPC}_{2,2}. \tag{15}$$

We note that there are two second-order contributions. In the case of $\text{IPC}_{2,2}$, the constraint $m_1 > m_2$ is necessary because the case $m_1 = m_2$ is already covered by $\text{IPC}_{2,1}$, and because the product $l_1(m_1)l_1(m_2)$ is commutative, making $m_1 > m_2$ sufficient to represent all unique combinations.

This idea can be extended to higher-order IPCs. The third-order IPC consists of three types of hypertasks, involving

$$l_3(m), \quad l_1(m_1)l_2(m_2), \quad \text{and} \quad l_1(m_1)l_1(m_2)l_1(m_3).$$

We see that third-order IPC involves three indices m_1, m_2, and m_3, and all valid combinations must be considered. Each combination requires training the reservoir. If only the past 15 inputs are considered, there are already $\binom{15}{3} = 455$ distinct target combinations for m_1, m_2, and m_3. This highlights the computational cost of IPC, as it requires training the reservoir thousands of times if even higher IPCs have to be computed. Higher-order IPC_i are defined analogously, as discussed in detail in Refs. [1,25]. The total IPC is given by the sum of all individual orders:

$$\text{IPC} = \sum_i \text{IPC}_i. \tag{16}$$

While IPC is often referred to as memory capacity in the literature [125–127], in this work we use the term memory capacity specifically for the first-order IPC_1. One important identity is that the IPC is upper bounded by the readout dimension N_R, i.e., $\text{IPC} \leqslant N_R$. The trends of IPC and task performance often resemble each other when the readout dimension is sufficiently large. In Refs. [128,129], analytical evidence is provided showing the influence of eigenvalues on IPC in classical reservoir computing. Furthermore, it has been demonstrated that IPC can help predict task performance in reservoir computing [130,131]. The concept has also been extended to quantum reservoir computing in Refs. [25,89].

The computation of the IPC requires simulating the state matrix $\mathbf{S}$. For this reason, we will briefly discuss the number of matrix multiplications needed to construct $\mathbf{S}$. Each reservoir evolution requires the following computational steps:

$$|\Phi_n\rangle = \sqrt{\frac{1-u_n}{2}}|0\rangle + \sqrt{\frac{1+u_n}{2}}|1\rangle,$$
$$\rho_n = e^{-iHt}(|\Phi_n\rangle\langle\Phi_n| \otimes \text{Tr}_1(\rho_{n-1}))e^{iHt},$$
$$\langle Z_i(t)\rangle = \text{Tr}(Z_i\rho_n). \tag{17}$$

The computation of the trace requires four matrix multiplications, with an additional two from the exponentials, one from the encoded state, and another K for the number of observables measured. This results in a total of $7 + K$ matrix multiplications per evolution step. Let $N_{\mathbf{u}} = N_{\text{Tr}} + N_{\text{Te}}$, where N_{Tr} and N_{Te} represent the number of training and testing

TABLE II. QRC simulation parameters.

Parameter	Value
N_{In} (input length)	10 000
N_{Tr} (training length)	25 000
N_{Te} (test length)	5000
Number of spins N_S	4
Regularization η	10^{-5}
Buffer N_b	100

steps, respectively, and let V be the number of multiplexing operations. Then, the total number of matrix multiplications is given by

$$N_{\text{state}} = (7 + K)N_{\mathbf{u}}V. \tag{18}$$

E. Simulation parameters

In Table II, the simulation parameters for the computation of the results are listed. The information processing capacity is computed up to the fourth order.

III. KRYLOV SPACES IN QUANTUM MECHANICS

This chapter introduces Krylov spaces in relation to quantum evolution. We first discuss spread complexity [32], a measure that quantifies the spread over the Krylov basis. Krylov expressivity is defined as a quantum-mechanically measurable quantity that quantifies the effective phase-space dimension of evolved quantum states [35]. Next, operator complexity is explored, defining a measure of how operators spread over a Krylov operator basis [33]. Building on this, Krylov observability is introduced as a measure that effectively captures the phase-space dimension of a set of operators $O_1, \ldots, O_K$, considering the number of measurements involved in constructing the effective Krylov space. In Ref. [34], we demonstrate that Krylov observability effectively captures the behavior of the IPC and the generalization properties of the quantum reservoir.

A. Krylov spaces for state evolution

The observation of the time-evolution operator being a map onto a Krylov space was first discussed in Refs. [32,33]. There, the proof for Eq. (22) was given, which we will be going over briefly.

1. Spread complexity

The Schrödinger equation for the time-independent Hamiltonian H with initial condition $|\Psi(0)\rangle$ is given by

$$\begin{aligned} \partial_t|\Psi(t)\rangle &= -iH|\Psi(t)\rangle, \\ |\Psi(0)\rangle &:= |\Psi_0\rangle. \end{aligned} \tag{19}$$

The solution to this equation is

$$|\Psi(t)\rangle = e^{-iHt}|\Psi_0\rangle = \sum_{k=0}^{\infty} \frac{(-iHt)^k}{k!}|\Psi_0\rangle, \tag{20}$$

where the series representation of the matrix exponent was used. Defining the linear function $f(|\Psi_0\rangle) := -iH$ leads to

$$\begin{aligned} |\Psi(t)\rangle &= e^{-iHt}|\Psi_0\rangle, \\ &= \sum_{k=0}^{\infty}(-iH)^k\frac{t^k}{k!}|\Psi_0\rangle = \sum_{k=0}^{\infty} f^k(|\Psi_0\rangle)\frac{t^k}{k!}. \end{aligned} \tag{21}$$

For any time $t \in \mathbb{R}$, it holds that

$$|\Psi(t)\rangle \in \text{Span}\left\{f^0(|\Psi_0\rangle), f^1(|\Psi_0\rangle)t, f^2(|\Psi_0\rangle)\frac{t^2}{2!}, \ldots\right\}.$$

Since $t^k/k! > 0$ for $t > 0$, it follows that

$$|\Psi(t)\rangle \in \text{Span}\{f^0(|\Psi_0\rangle), f^1(|\Psi_0\rangle), \ldots\} := K_\infty.$$

f being a linear function and considering the Krylov space property implies that there exists an $m \leqslant N$, such that

$$\begin{aligned} K_\infty &= \text{Span}\{f^0(|\Psi_0\rangle), f^1(|\Psi_0\rangle), f^2(|\Psi_0\rangle), \ldots\} \\ &= \text{Span}\{f^0(|\Psi_0\rangle), f^1(|\Psi_0\rangle), \ldots, f^{m-1}(|\Psi_0\rangle)\} \\ &= K_m. \end{aligned} \tag{22}$$

m is called the grade of $|\Psi_0\rangle$ in regard to $f := -iH$ [32,33]. Building upon this, Ref. [32] proposed the spread complexity, which is a measure of how the state spreads over the Krylov basis of K_m. This is achieved by orthonormalization through the Lanczos algorithm of the vectors $f^j(|\Psi_0\rangle)$, which results in

$$K_m = \text{Span}\{|k_0\rangle, \ldots, |k_{m-1}\rangle\}. \tag{23}$$

The time-evolved state is reconstructed through the basis representation $|k_i\rangle$ as

$$|\Psi(t)\rangle = \sum_{n=0}^{m-1}\langle k_n|\Psi(t)\rangle|k_n\rangle. \tag{24}$$

With $\alpha_n(t) = \langle k_n|\Psi(t)\rangle$, the spread complexity is defined as

$$\mathcal{K}_S(t) = \sum_{n=0}^{m-1}(n+1)|\alpha_n(t)|^2. \tag{25}$$

Spread complexity characterizes how an initial state evolves over a basis constructed from different powers of the Hamiltonian. This metric provides valuable insights and helps in the understanding of the time-evolution operator, but faces some challenges as an expressivity measure for quantum machine learning and quantum reservoir computing. First, the requirement for knowledge of the system's Hamiltonian poses a challenge because the quantum machine learning network is typically not described by a Hamiltonian but by a series of quantum circuits. Additionally, the different powers of the Hamiltonian must be computed classically, which becomes increasingly difficult as the number of qubits grows. To address these challenges, we demonstrated that time-evolved states can be used to construct a basis from which the complexity measure can be defined, exhibiting the same behavior in the researched systems [35].

2. Krylov expressivity

In Ref. [35], we demonstrated that instead of using the Krylov space

$$K_m = \mathrm{Span}(|\Psi_0\rangle, H|\Psi_0\rangle, \ldots, H^{m-1}|\Psi_0\rangle),$$

a set of time-evolved states can be employed to construct the space

$$G_m := \mathrm{Span}(e^{-iHt_0}|\Psi_0\rangle, \ldots, e^{-iHt_{m-1}}|\Psi_0\rangle).$$

The basis of G_m is evolved states and therefore quantum-mechanically measurable. The global phase does not change the space dimension and detailed proofs are given in Ref. [35]. In Ref. [35], we showed that the grade m of the Krylov space K_m is equal to the number of pairwise distinct eigenvalues of the Hamiltonian d, i.e., $m = d$. Building on this, Krylov expressivity is defined, which is upper bounded by the grade m.

The computation of *Krylov expressivity* requires the knowledge of the grade m. For this, time-evolved states $|g_i\rangle = \exp(-iHt_i)|\Psi_0\rangle$ are computed, where some T_K is picked and $t_i = (i+1)T_K/(N+1)$. For exactly the grade m, it will hold that $|g_m\rangle \in \mathrm{Span}(|g_1\rangle, |g_2\rangle, ..., |g_{m-1}\rangle) = G_m$. We then simulate the time-evolved state until some time T and calculate the sampling times $\tau_i = (i+1)T/m$ with $i = 0, \ldots, m-1$. For small T, it is expected that the vectors $|g_i\rangle$ are close to each other; i.e., for all i, $|g_i\rangle \approx |g_{i+1}\rangle$ would hold. Analytically speaking, $|g_i\rangle$ are independent, but numerically almost equal. For the computation of the Krylov expressivity, a measure of similarity between two time-evolved states $|g_i\rangle$ and $|g_{i+1}\rangle$ is required. Here, we proposed the computation of the fidelity for pure states $F(|g_i\rangle, |g_{i+1}\rangle)$ but other measures can also be utilized.

$$\lambda_i := F(|g_i\rangle\langle g_i|, |g_{i+1}\rangle\langle g_{i+1}|) = |\langle g_i|g_{i+1}\rangle|. \tag{26}$$

The *Krylov expressivity* $\mathcal{E}_K$ for λ, where $\lambda \in [0, 1]$ holds, is given by

$$m_{\mathrm{eff}_i} = \begin{cases} 1, & \text{if } \lambda_i < \lambda, \\ 1 - \frac{1}{1-\lambda} \cdot (\lambda_i - \lambda), & \text{if } \lambda_i \geqslant \lambda, \end{cases}$$

$$\mathcal{E}_K = 1 + \sum_{i=1}^{m-1} m_{\mathrm{eff}_i}. \tag{27}$$

The first vector $|g_1\rangle$ adds a dimension of 1. For $\lambda_i < \lambda$, we say that the two vectors are linearly independent and the Krylov expressivity is increased by 1. In the region $\lambda_i \geqslant \lambda$, the Krylov expressivity m_{eff_i} is interpolated to consider the difference between analytical independence and the independence that is needed for computation. We will explore the Krylov expressivity $\mathcal{E}_K$ depending on the timescale T and which influence this has on the performance in quantum reservoir computing. For the following discussion, we set $\lambda = \frac{1}{\sqrt{2}}$, inspired by a 3 dB falloff.

In Krylov complexity, the Krylov basis

$$K_m = \mathrm{Span}\{|\psi_0\rangle, H|\psi_0\rangle, H^2|\psi_0\rangle, \ldots, H^{m-1}|\psi_0\rangle\} \tag{28}$$

is computed, and after normalization, the state at a given time $|\psi(T)\rangle$ is represented in the orthonormalized Krylov space. The complexity is then defined by how the state evolves within this Krylov space.

Krylov expressivity, on the other hand, defines a time-dependent Krylov space for each time T, given by

$$G_m(T) = \mathrm{Span}\{|\psi_0\rangle, e^{iH\frac{T}{m}}|\psi_0\rangle, \ldots, e^{iHT}|\psi_0\rangle\}. \tag{29}$$

Krylov expressivity is then defined as the effective phase-space dimension of $G_m(T)$ and is therefore time-dependent. The same interpretation applies to Krylov operator complexity and Krylov observability, which will both be introduced in the following.

B. Krylov spaces for operator evolution

Where spread complexity tries to quantify the spread of states, operator complexity quantifies the spread of time-evolved operators [33] and is defined in the following section.

1. Operator complexity

The evolution of operators is governed by

$$\partial_t O(t) = i[H, O(t)]. \tag{30}$$

The solution to this equation is

$$O(t) = e^{iHt}Oe^{-iHt} = \sum_{k=0}^{\infty} \frac{(it)^k}{k!}\mathcal{L}^k(O), \tag{31}$$

where $\mathcal{L}(O) = HO - OH$ is the Liouvillian superoperator and linear in O, i.e.,

$$\begin{aligned} \mathcal{L}(\lambda O_1 + O_2) &= H(\lambda O_1 + O_2) - (\lambda O_1 + O_2)H \\ &= \lambda HO_1 - \lambda HO_1 + HO_2 - HO_2 \\ &= \lambda\mathcal{L}(O_1) + \mathcal{L}(O_2). \end{aligned} \tag{32}$$

Any time-evolved operator $O(t)$ has therefore to be in the span of the powers of the Liouvillian $\mathcal{L}^n$:

$$O(t) \in \mathrm{Span}\{\mathcal{L}^0(O), \mathcal{L}^1(O), \mathcal{L}^2(O), ...\}. \tag{33}$$

By using the linear property of $\mathcal{L}$ and the Krylov space property, authors of Ref. [33] showed that there exists an $M \in \mathbb{N}$ such that

$$O(t) \in \mathrm{Span}\{\mathcal{L}^0(O), \mathcal{L}^1(O), ..., \mathcal{L}^{M-1}(O), ...\} = \mathcal{L}_M \tag{34}$$

holds true. Similar to the Krylov state complexity, the authors perform an orthonormalization procedure to construct the space

$$\begin{aligned} \mathcal{L}_M &:= \mathrm{Span}\{\mathcal{L}^0(O), \mathcal{L}^1(O), ..., \mathcal{L}^{M-1}(O)\} \\ &= \mathrm{Span}\{\mathcal{W}_0, \mathcal{W}_1, ..., \mathcal{W}_{M-1}\}, \end{aligned} \tag{35}$$

where $\{\mathcal{W}_i\}_{i=...,M-1}$ is the orthonormal basis constructed through the Lanczos algorithm. In the next step, the time-evolved operator is expressed in the Krylov basis as

$$O(t) = \sum_{n=0}^{M-1} i^n \beta_n(t)\mathcal{W}_n, \tag{36}$$

where $\beta_n(t) = (O(t), \mathcal{W}_n)$ is a scalar product defined in the operator space. Operator complexity $\mathcal{K}_O$ is then defined as

$$\mathcal{K}_O(t) = \sum_{n=0}^{M-1} (n+1)|\beta_n(t)|^2. \tag{37}$$

2. Krylov observability

One aspect that operator complexity lacks is a way to define an effective space dimension and a method to consider multiple operators. This is especially crucial for quantum machine learning, where a set of observables is measured, and quantum reservoir computing, where multiplexing is used to increase the readout dimension. In Ref. [34], we show that instead of using the Krylov space of the different powers of the Liouvillian $\mathcal{L}^k$,

$$O(t) \in \mathcal{L}_M := \mathrm{Span}\{\mathcal{L}^0(O), \mathcal{L}^1(O), ..., \mathcal{L}^{M-1}(O)\},$$

we can use time-evolved observables to construct an equivalent space:

$$O(t) \in \mathcal{F}_M = \mathrm{Span}\{\tilde{O}(t_0), \tilde{O}(t_1), ..., \tilde{O}(t_{M-1})\} = \mathcal{L}_M. \quad (38)$$

Building on this, we present an Algorithm 1 that enables the computation of a minimal space $\mathcal{F}$ such that any time-evolved operator $O_i(t) \in \mathcal{F}$ holds for all $O_i \in \{O_1, O_2, \ldots, O_K\}$ and all t [34]. This algorithm further returns the contribution of each operator O_i as a space $\mathcal{F}_i$ with the following properties:

$$\mathcal{F} = \bigcup_{k=1}^{K} \mathcal{F}_k = \bigcup_{k=1}^{K} \tilde{\mathcal{F}}_k,$$

$$\dim\left(\bigcup_{j=1}^{l} \tilde{\mathcal{F}}_j\right) = \sum_{j=1}^{l} \dim(\tilde{\mathcal{F}}_j).$$

The computation of Krylov observability considers K observables $O_1, \ldots, O_K$. For each observable O_k, the spaces $\mathcal{F}_k$ are computed with $\dim(\mathcal{F}_k) = M_k$. With $R_k = \min(V, M_k)$ and $\tau_k = T/M_k$, we can then define

$$\mathcal{G}_k = \{O_k(\tau_1), O_k(\tau_2), \ldots, O_k(\tau_{R_k})\},$$

where V is the number of measurements. The observability of the kth observable O_k is defined as

$$\kappa_k(T) = 1 + \sum_{j=1}^{R_k-1} (1 - \mathrm{F}(O_k(\tau_j), O_k(\tau_{j+1}))), \quad (39)$$

where F is the normalized fidelity between the two time-evolved operators given by

$$\mathrm{F}(A, B) = \left|\mathrm{Tr}\left(\frac{A^\dagger B}{\|A\|\|B\|}\right)\right|. \quad (40)$$

The *Krylov observability* $\mathcal{O}_K(T)$ of V multiplexed observables $O_1, \ldots, O_K$ is defined as [34]

$$\mathcal{O}_K(T) = \sum_{k=1}^{K} \kappa_k(T). \quad (41)$$

To compute the Krylov observability, we require the orthonormalization of the space $\mathcal{F}$. To construct a linearly independent basis, an algorithm such as the Gram-Schmidt procedure may be applied. In that case, each time-evolved operator at time t_n must be orthonormalized with respect to all previous basis elements, requiring a total of $(n-1)$ matrix multiplications for each $O(t_n)$. Given $V \cdot K$ matrices, the total number of matrix multiplications due to Gram-Schmidt is

$$\sum_{n=1}^{VK} (n-1) = \frac{VK(VK-1)}{2}. \quad (42)$$

Next, we require the computation of the time-evolved operators via $O(t_n) = e^{iHt} O e^{-iHt}$, which requires at most VK matrix multiplications. Finally, for the computation of the overlaps F [Eq. (40)], other VK matrix multiplications are required. The resulting number of matrix multiplications required is then given by

$$N_{\mathrm{obs}} = \frac{VK(VK+3)}{2}. \quad (43)$$

The ratio r of the number of matrix multiplications for Krylov complexity to the construction of the state matrix is

$$r = \frac{N_{\mathrm{obs}}}{N_{\mathrm{state}}} = \frac{V^2K^2 + 3VK}{2(7+K)N_{\mathbf{u}}V} \in O\left(\frac{VK}{N_{\mathbf{u}}}\right). \quad (44)$$

In typical reservoir computing approaches, the number of inputs $N_{\mathbf{u}}$ is much larger than the readout dimension $N_R = VK$; i.e., $N_{\mathbf{u}} \gg VK$, thus resulting in $r \ll 1$.

C. Dimension of Krylov spaces

The construction of Krylov spaces is typically performed until a cutoff is observed in the algorithm, as discussed in Appendix A. Due to numerical errors, the actual dimension considered may increase. Therefore, it is necessary to understand the grades m and M of the Krylov spaces

$$K_m = \mathrm{Span}\{|\psi_0\rangle, H|\psi_0\rangle, \ldots, H^{m-1}|\psi_0\rangle\},$$
$$\mathcal{L}_M := \mathrm{Span}\{\mathcal{L}^0(O), \mathcal{L}^1(O), \ldots, \mathcal{L}^{M-1}(O)\}. \quad (45)$$

It is also important to determine what these grades should be without relying on iterative algorithms like Lanczos, which are prone to numerical errors. In Ref. [25], we showed that the grade m is upper bounded by the number of pairwise distinct eigenenergies and stated the following theorem.

Theorem 1. Let $H \in \mathbb{C}^{N\times N}$ be a Hermitian Hamiltonian with d pairwise distinct eigenvalues $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{d-1}$, and let $\{|\phi_j\rangle\}$ be an orthonormal eigenbasis of H, satisfying

$$H|\phi_j\rangle = \varepsilon_j|\phi_j\rangle. \quad (46)$$

Then, the time-evolved state $|\Psi(t)\rangle = e^{-iHt}|\Psi_0\rangle$ lies in a d-dimensional subspace $E_d \subseteq \mathbb{C}^N$, i.e.,

$$|\Psi(t)\rangle \in E_d := \mathrm{Span}\{|\xi_0\rangle, |\xi_1\rangle, \ldots, |\xi_{d-1}\rangle\}, \quad (47)$$

where the vectors $|\xi_p\rangle$ are defined by

$$|\xi_p\rangle := \frac{1}{\sqrt{|J_p|}} \sum_{j\in J_p} \alpha_j|\phi_j\rangle, \quad \text{with } \alpha_j := \langle\phi_j|\Psi_0\rangle \quad (48)$$

and $J_p := \{j \mid \varepsilon_j = \varepsilon_p\}$ denotes the set of indices corresponding to the degenerate eigenspace of eigenvalue ε_p. The normalization factor $|J_p|$ is the cardinality of the set J_p. The starting state $|\psi_0\rangle$ can be represented in the basis $\{|\xi_p\rangle\}_p$, with $\gamma_p = \langle\xi_p|\psi_0\rangle$, as

$$|\psi_0\rangle = \sum_{p=0}^{d-1} \gamma_p|\xi_p\rangle. \quad (49)$$

Let n_1 denote the number of coefficients for which $\gamma_p = 0$. Then, the number of linearly independent vectors is reduced to $d - n_1$. It also holds that for the Krylov state space $K_m = \mathrm{Span}\{|\psi_0\rangle, H|\psi_0\rangle, \ldots, H^{m-1}|\psi_0\rangle\}$,

$$m = d - n_1 \tag{50}$$

holds.

Proof. See Appendix C. ■

The following theorem shows a similar identity for operator spaces $\mathcal{L}_M$.

Theorem 2. Let $H \in \mathbb{C}^{N\times N}$ be a Hamiltonian with eigenbasis $\{|\phi_j\rangle\}$ and corresponding eigenvalues ε_j, and let O be an operator on the same Hilbert space. Define the Liouvillian Krylov space

$$\mathcal{L}_M := \mathrm{Span}\{\mathcal{L}^0(O), \mathcal{L}^1(O), \ldots, \mathcal{L}^{M-1}(O)\},$$

where $\mathcal{L}(O) = [H, O]$ is the Liouvillian superoperator.

Define the transition frequencies $\omega_{mn} := \varepsilon_m - \varepsilon_n$, and let $\{\omega_P\}_{P=0}^{N_\omega - 1}$ be the set of all pairwise distinct values taken by ω_{mn}. For each ω_P, define the index set

$$J_P := \{(m, n) \,|\, \omega_{mn} = \omega_P\}$$

and the corresponding matrix

$$\sigma_P := \sum_{(m,n)\in J_P} \langle\phi_m|O|\phi_n\rangle |\phi_m\rangle\langle\phi_n|.$$

Let N_1 be the number of vanishing contributions $\sigma_P = 0$, then the time-evolved operator is given by

$$O(t) = \sum_{P\in S} e^{i\omega_P t}\sigma_P,$$

with $S = \{P \mid \sigma_P \neq 0\} = \{s_0, s_1, \ldots, s_{N_\omega - N_1 - 1}\}$. The operator lies in the span

$$O(t) \in \mathrm{Span}\{\sigma_{s_0}, \sigma_{s_1}, \ldots, \sigma_{s_{N_\omega - N_1 - 1}}\} = \mathcal{P}_{N_\omega - N_1}.$$

Further, the grade M of the Krylov space $\mathcal{L}_M$ is given by

$$M = N_\omega - N_1.$$

Proof. See Appendix D. ■

IV. RESULTS

The aim of this work is to utilize and compare expressivity measures to deepen the understanding of quantum reservoir computing. To begin, we simulate spread complexity and fidelity to investigate whether task performance can be explained by these measures. Figure 2(a) illustrates fidelity and spread complexity averaged over 20 random starting states. We see that spread complexity $\mathcal{K}_S$ increases, while fidelity F decreases with larger clock cycles. Additionally, we observe an antiproportional relationship between the two measures. When fidelity reaches a minimum at $T \approx 36$, spread complexity exhibits a maximum, and when fidelity increases, spread complexity decreases, and vice versa.

The oscillations of F and $\mathcal{K}_S$ arise due to the periodic nature of the quantum system, where the state evolves as

$$|\Psi(t)\rangle := \sum_{n=0}^{N-1} e^{-i\varepsilon_n t}|\phi_n\rangle\langle\phi_n|\psi_0\rangle, \tag{51}$$

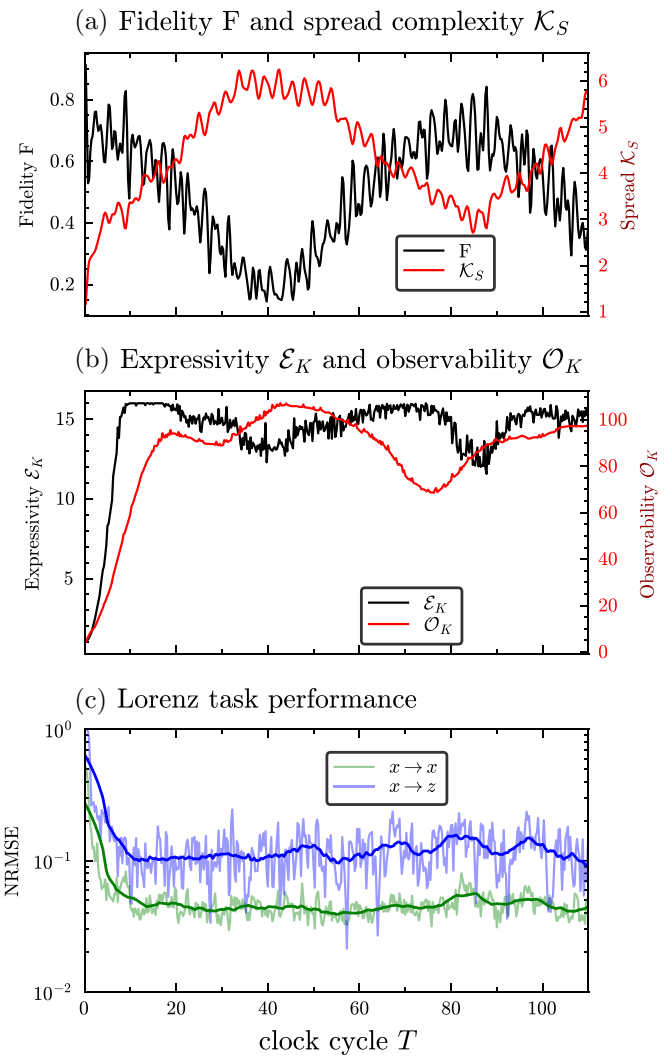


FIG. 2. Fidelity F (black) and spread complexity $\mathcal{K}_S$ (red) in panel (a), Krylov observability $\mathcal{O}_K$ (red) and Krylov expressivity $\mathcal{E}_K$ in panel (b), and the Lorenz task performance for the five-step ahead prediction ($\Delta t = 0.1$) of the x variable (green) and the cross prediction of the z variable (blue) in panel (c) depending on the clock cycle T, when measuring one site of the H_{I4} Hamiltonian. The bold line in panel (c) shows the first-order Savitzky-Golay filtered NRMSE, used to better illustrate the saturation in task performance (computed using `scipy.signal.savgol_filter`).

where $H|\phi_n\rangle = \varepsilon_n|\phi_n\rangle$ is the eigenvalue equation and $N = \dim(\mathcal{H})$ is the dimension of the Hilbert space. This is a superposition of N periodic functions and is therefore periodic (or quasiperiodic) in nature.

The fidelity is given by $F = |\langle\Psi(t)|\Psi_0\rangle|$. Due to the quasiperiodic structure, there exist times τ for which $|\Psi(\tau)\rangle \approx |\Psi_0\rangle$. At such times, the fidelity $F = |\langle\Psi(\tau)|\Psi_0\rangle|$ will be close to 1, which is observed at $\tau \approx 80$.

On the other hand, Krylov spread complexity $\mathcal{K}_S$ measures the spread over the Krylov basis. The first basis state represents the initial state, $|k_0\rangle = |\Psi_0\rangle$. In the construction of Krylov spread complexity [Eq. (25)], the amplitude with respect to $|k_0\rangle$ contributes with the weakest weight of 1 and is given by $|\alpha_0(t)| = |\langle\Psi(t)|\Psi_0\rangle| = F$. Since the

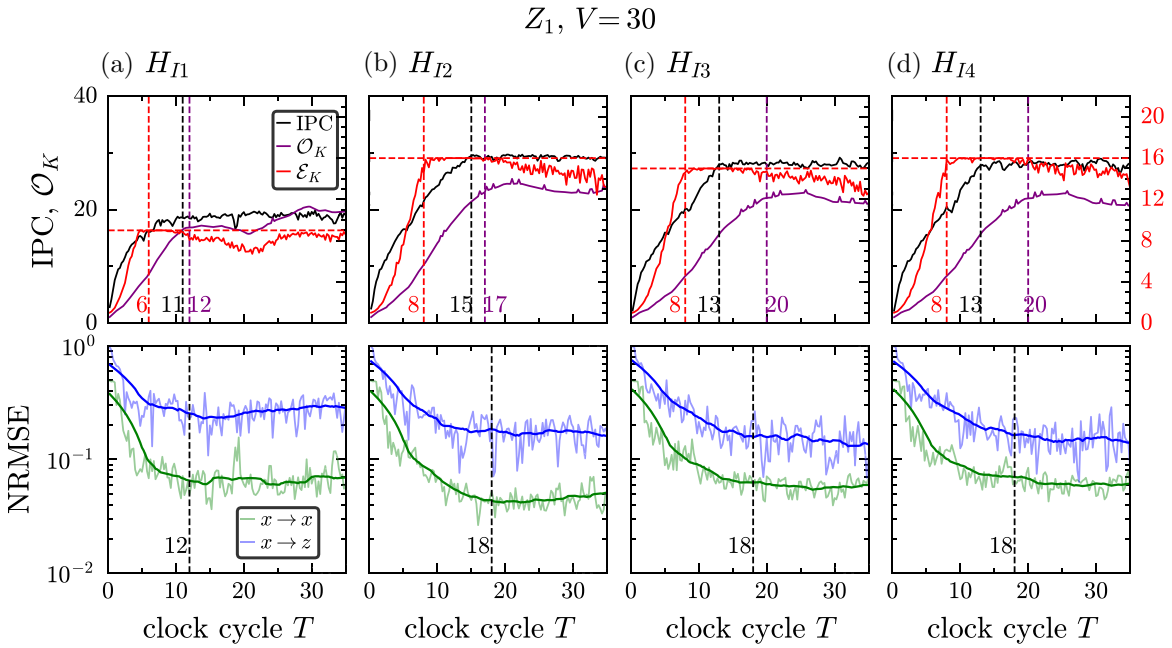


FIG. 3. The first row shows the information processing capacity IPC (black), the Krylov expressivity $\mathcal{E}_K$ (red), and the Krylov observability $\mathcal{O}_K$ (purple), and the second row shows the Lorenz task performance for the five-step ahead prediction ($\Delta t = 0.1$) of the x variable (green) and the cross prediction of the z variable (blue) depending on the clock cycle T for the different quantum reservoirs (a) H_{I1}, (b) H_{I2}, (c) H_{I3}, and (d) H_{I4}. The state matrix is constructed by measuring the Z_1 observable $V = 30$ times. The vertical dotted lines in the first row represent when IPC (black), $\mathcal{O}_K$ (purple), and $\mathcal{E}_K$ (red) saturate and the time is given next to the line. The vertical black line in the second row indicates when the Lorenz task performance saturates. The bold line in panel (c) shows the first-order Savitzky-Golay filtered NRMSE, used to better illustrate the saturation in task performance (computed using `scipy.signal.savgol_filter`).

normalization condition $\sum_n |\alpha_n(t)|^2 = 1$ holds, a large value of $|\alpha_0(t)|$ implies that the contributions from the other $|\alpha_n(t)|$ must be smaller. This, in turn, implies that the amplitudes of the higher-weight contributions are reduced, thereby lowering $\mathcal{K}_S$. This explains the dip in Krylov spread complexity, which varies inversely with the fidelity.

In Fig. 2(c), the NRMSE values of the Lorenz tasks for the five-step prediction of the x variable (green) and the cross prediction of the z variable (blue) are presented. We observe a reduction in error, followed by a saturation in task performance. The initial decrease in NRMSE can be explained using fidelity, which captures how far the state evolves from the initial state. This, however, only tells us that the system has to evolve, to perform any operations as at $T = 0$ the reservoir evolution is simply the identity. Meanwhile, spread complexity quantifies how the time-evolved state spreads over time in the Krylov basis. Higher spread complexity therefore implies a broader spread within the Krylov basis at the corresponding clock cycle. However, neither fidelity nor spread complexity alone can explain the saturation of NRMSE. The oscillatory behavior observed in fidelity and spread complexity [Fig. 2(a)] is not reflected in task performance [Fig. 2(c)]. Figure 2(b) shows Krylov expressivity $\mathcal{E}_K$ and Krylov observability $\mathcal{O}_K$. Both measures show an increase followed by saturation, around which they oscillate, similar to the task performance in Fig. 2(c). This shows that Krylov expressivity and Krylov observability capture the saturation in task performance, while only limited insights can be gained from fidelity and spread complexity.

A. Specific Hamiltonians

Figure 3 shows the results when only the first observable Z_1 is measured, while in Fig. 4, measurements of the observables Z_1, Z_2, Z_3, and Z_4 are considered. In all cases, the observables are multiplexed $V = 30$ times. A summary of the results is provided in Appendix B. The Krylov expressivity $\mathcal{E}_K$ of each Hamiltonian H_{I1}, H_{I2}, H_{I3}, and H_{I4} is upper bounded by 9, 16, 15, and 16, respectively. This upper bound corresponds to the grade of the Krylov space or the number of pairwise distinct eigenvalues. Figure 3(a) shows the result for H_{I1}. We observe that NRMSE of the Lorenz task, IPC, and $\mathcal{O}_K$ all saturate around $T = 12$, while $\mathcal{E}_K$ reaches saturation at $T = 6$. Increasing the complexity of the Hamiltonian slightly by modifying one of the interspin couplings $J_{i,j}$ results in H_{I2}. This Hamiltonian reaches a higher IPC, Krylov expressivity $\mathcal{E}_K$, and Krylov observability $\mathcal{O}_K$. Furthermore, the Lorenz task errors are slightly decreased. When using pairwise distinct interspin couplings $J_{i,j}$ with equal spacing, the resulting Hamiltonian is H_{I3} [Fig. 3(c)]. Here, $\mathcal{E}_K$, IPC, and $\mathcal{O}_K$ each saturate at $T_{\text{sat}} \approx 8$, 13, and 20, respectively. The Lorenz task performance, however, reaches saturation at $T_{\text{sat}} \approx 18$, which best aligns with the behavior of Krylov observability $\mathcal{O}_K$.

For H_{I1} and H_{I2}, IPC and $\mathcal{O}_K$ show similar behavior, but a discrepancy appears in H_{I3}. This occurs because IPC is upper bounded by the readout dimension of $N_R = 30$, to which the IPC then saturates, IPC $\leqslant N_R$. In this case, Krylov observability $\mathcal{O}_K$ best explains the behavior of task performance. The worse task performance of H_{I1} compared to the other

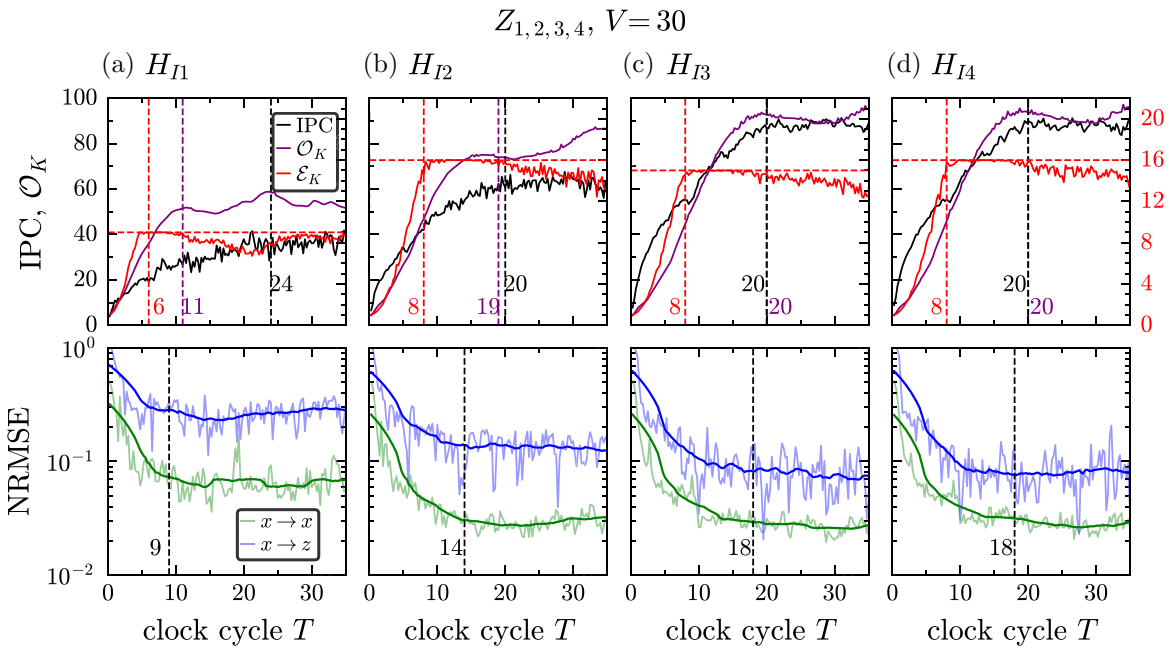


FIG. 4. Same as in Fig. 3, except that the state matrix is constructed by measuring the Z_1, Z_2, Z_3, and Z_4 observables $V = 30$ times.

Hamiltonians is explained by the fact that metrics IPC, $\mathcal{E}_K$, and $\mathcal{O}_K$ are all smaller.

When comparing H_{I3} and H_{I2}, we note that $\mathcal{E}_K(H_{I3}) < \mathcal{E}_K(H_{I2})$. At the same time, we observe that the task performance for the cross-prediction task is better for H_{I3}, while it is worse for the prediction of the x variable. To further investigate this discrepancy, we introduce H_{I4}. This Hamiltonian is nearly identical to H_{I3}, differing only in one interspin coupling, which results in a Krylov expressivity equal to that of H_{I2}. Despite this adjustment, the IPC, $\mathcal{O}_K$, $\mathcal{E}_K$, and Lorenz task errors for H_{I4} remain nearly identical to those of H_{I3}.

To discuss this further, we consider a typical quantum reservoir where each observable is measured, as shown in Fig. 4. This results in a readout dimension of $N_R = 120$. In this setup, H_{I1} again exhibits largest errors. While the Lorenz task performance remains unchanged, the information processing capacity increases from IPC = 20 (Fig. 3) to IPC = 40 in Fig. 4(a). For H_{I2}, H_{I3}, and H_{I4} [Figs. 4(b)–4(d)], we observe that IPC and $\mathcal{O}_K$ both perform similarly well in explaining the saturation of the NRMSE of the Lorenz task.

The improved performance of IPC in Fig. 4 compared to Fig. 3 is due to the system being sampled with $N_R = 120$ readout nodes. In this case, the IPC is upper bounded by IPC $\leqslant N_R = 120$, which is sufficiently large to not observe the upper bound by the readout dimension.

Table III shows the NRMSE of the Lorenz task at saturation, when only Z_1 is measured and when all four sites are measured. For H_{I1}, we observe that the number of sites measured does not change task behavior. For H_{I2}, we observe a slightly smaller NRMSE when all sites are measured. H_{I3} and H_{I4} exhibit identical task performance, with the error of both tasks halved when more observables are measured. To discuss the differing behavior of expressivity and observability between H_{I2} and H_{I3}, we take a closer look at task performance when all sites are measured (Fig. 4). We observe that the cross-prediction task for H_{I3} yields an error of NRMSE = 0.08, while for H_{I2} the error is higher, at NRMSE = 0.15. The Krylov expressivity for H_{I2}, with $\mathcal{E}_K(H_{I2}) = 16$, is slightly higher than that of H_{I3}, which has $\mathcal{E}_K(H_{I3}) = 15$.

At first glance, this may seem counterintuitive: H_{I2} maps input data into a larger Krylov space, suggesting that more information should be accessible. However, although the input is projected into a higher-dimensional space, information can only be extracted through the set of measured observables. In this case, the number of independent measurements for H_{I2} is actually smaller than for H_{I3}, which is mirrored by the smaller Krylov observability $\mathcal{O}_K(H_{I2}) = 60$ versus $\mathcal{O}_K(H_{I3}) = 85$.

To further probe this behavior, the IPC and $\mathcal{O}_K$ are computed for various clock cycles, virtual nodes, and numbers of readout observables for H_{I2} (Fig. 5) and H_{I3} (Fig. 6). In the first and second rows, only the first site and the first two sites are measured for the construction of the state matrix, respectively, while the third row shows the results when all sites are measured. For H_{I2} (Fig. 5), we observe IPC < 85 and Krylov observability $\mathcal{O}_K < 85$. In contrast, H_{I3} shows a maximum of 105 (Fig. 6), indicating that the larger IPC is due to the increased Krylov observability $\mathcal{O}_K$. Furthermore, an increase in the number of observables measured shifts the

TABLE III. NRMSE for the Lorenz task of the four Hamiltonians H_{I1}, H_{I2}, H_{I3}, and H_{I4}, when measuring one site Z_1 (rows 1 and 3) and when measuring all sites (rows 2 and 4).

	Z_1 $x \to x$ NRMSE$_{\text{sat}}$	All sites $x \to x$ NRMSE$_{\text{sat}}$	Z_1 $x \to z$ NRMSE$_{\text{sat}}$	All sites $x \to z$ NRMSE$_{\text{sat}}$
H_1	0.08	0.08	0.30	0.30
H_2	0.04	0.03	0.20	0.15
H_3	0.06	0.03	0.16	0.08
H_4	0.06	0.03	0.16	0.08

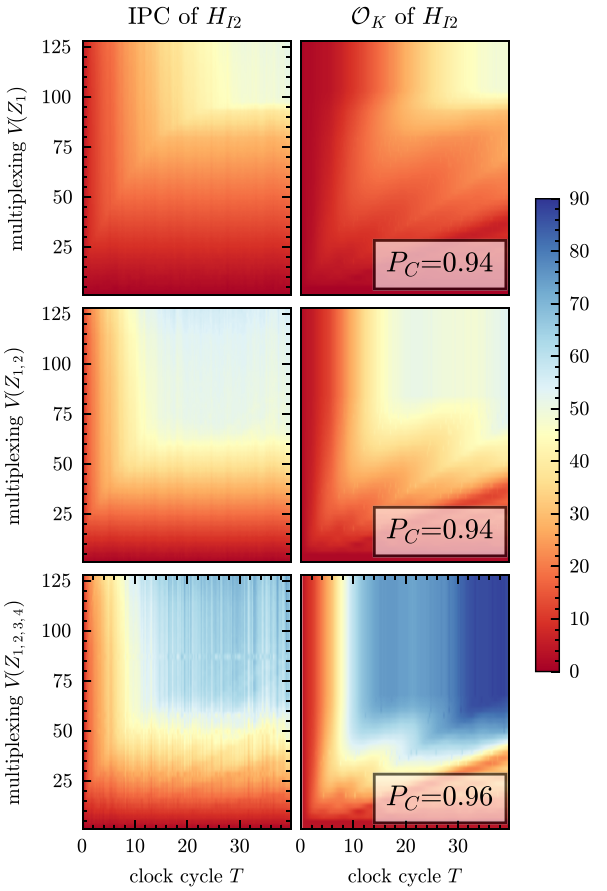


FIG. 5. Observability $\mathcal{O}_K$ (right column) and IPC (left column) are color coded according the the color bar, depending on multiplexing V and clock cycle T for H_{I2}. The Pearson correlation factor P_C between the two images in each row is calculated, indicating an almost identical behavior between $\mathcal{O}_K$ and IPC.

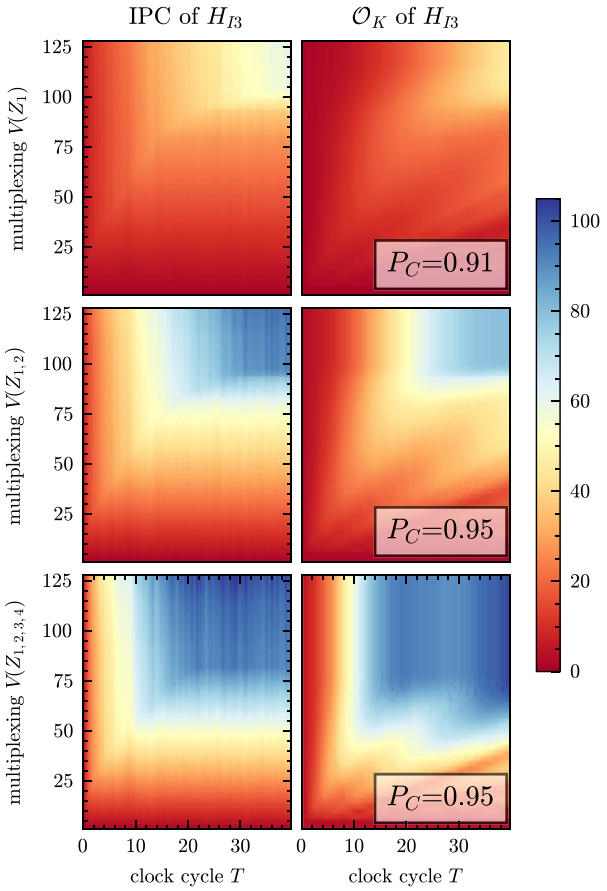


FIG. 6. Same as in Fig. 5 for H_{I3}.

best performance from larger clock cycles and a larger number of measurements to smaller clock cycles and fewer measurements. Each plot includes the Pearson correlation coefficient P_C [Eq. (12)] between IPC and $\mathcal{O}_K$. For H_{I2}, the correlation coefficients are $P_C = 0.94$ when one or two observables are measured, and $P_C = 0.96$ when all observables are considered. This indicates a strong and consistent correlation across all configurations. The corresponding correlation coefficients for H_{I3} are $P_C = 0.91$ for one observable and $P_C = 0.95$ for both two and four observables.

While Krylov observability captures the discrepancy between H_{I2} and H_{I3}, a deeper explanation is provided by the analysis in Sec. III C. There, the Krylov grades of $\mathcal{L}_M$ and K_m are defined in terms of the number of distinct eigenvalues d, transition frequencies N_ω, and vanishing contributions N_1, where $\sigma_P = 0$. The rank of the Liouvillian Krylov space is given by $M = N_\omega - N_1$. We compute these quantities for all observables Z_1–Z_4 and summarize the results in Table IV. Although H_{I2} has more distinct eigenvalues ($d = 16$) and frequencies ($N_\omega = 237$), it also exhibits significantly more vanishing contributions: $N_1 = 176$ for Z_1 and Z_2 and $N_1 = 158$ for Z_3 and Z_4. In contrast, H_{I3} shows fewer zero terms with $N_1 = 112$ and $N_\omega = 211$ across all observables.

The resulting Krylov grades are $M(H_{I2}) = 61$ for Z_1 and Z_2, and $M = 79$ for Z_3 and Z_4, while H_{I3} consistently yields $M = 99$ across all observables. This supports the higher expressivity of H_{I3}, as reflected by its larger IPC values, and highlights how differences in Krylov space dimensions influence performance. Results for the remaining Hamiltonians are provided in Appendix E.

In a final experiment, we compute the IPC and $\mathcal{O}_K$ for an Ising Hamiltonian with five sites and pairwise distinct couplings. The results are presented in Fig. 7, where one site (first

TABLE IV. Spectral and dynamical statistics for each (Hamiltonian, observable) pair. d is the number of pairwise distinct eigenvalues, d^2 is the number of eigenvalue pairs, N_ω is the number of distinct transition frequencies, N_1 is the number of zero σ_P contributions, and M is the resulting Krylov dimension.

Hamiltonian	Observable	d	d^2	N_ω	N_1	M
H_{I2}	Z_1	16	256	237	176	61
H_{I2}	Z_2	16	256	237	176	61
H_{I2}	Z_3	16	256	237	158	79
H_{I2}	Z_4	16	256	237	158	79
H_{I3}	Z_1	15	225	211	112	99
H_{I3}	Z_2	15	225	211	112	99
H_{I3}	Z_3	15	225	211	112	99
H_{I3}	Z_4	15	225	211	112	99

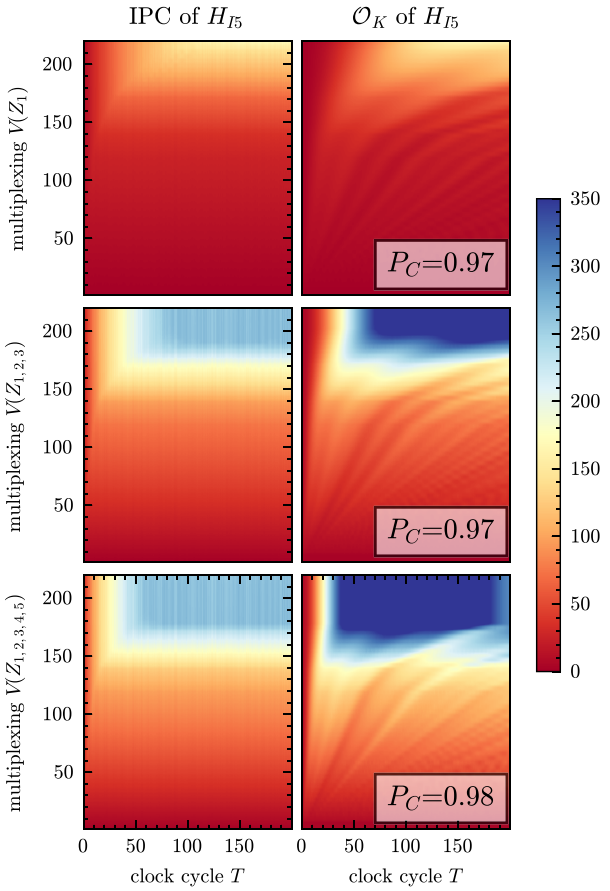


FIG. 7. Same as in Fig. 6 for a reservoir with five sites, $N_S = 5$.

row), three sites (second row), and all five sites (third row) are measured in the z direction. Simulations are performed up to $V = 220$ virtual nodes and clock cycles up to $T = 200$. Computing the IPC for each parameter configuration requires hundreds of hours for the five-site system and tens of hours for the four-site system, which limits our ability to simulate even larger systems. Once again, we observe almost identical behavior between IPC and $\mathcal{O}_K$, as indicated by a correlation factor of $P_C = 0.97$.

B. Hamiltonian with random coupling

Until now, specific Hamiltonians have been analyzed with respect to task performance, typically restricted to very small systems in order to describe behavior across the full parameter space. However, most research simulates larger spin systems with random interspin coupling. The interspin couplings of such quantum reservoirs can be sampled from distributions to compute statistics and reveal general trends in task performance.

In this regard, we now analyze the behavior of the transverse field Ising model in Eq. (8), where the interspin couplings are sampled uniformly in the interval $[0.25, 0.75]$, i.e., $J_{ij} \in \mathcal{U}([0.25, 0.75])$, with $h = 0.5$. These parameters are commonly used for QRC purposes [8]. The state matrix is constructed by measuring all spin sites of a system with $N_S = 6$ sites, repeated V times.

Figure 8 presents the results of simulating the transverse field Ising model with random interspin couplings. Three configurations for the number of measurements are shown: $V = 10$ (a), $V = 30$ (b), and $V = 50$ (c). For each case, ten Hamiltonians were randomly sampled, and the average (bold line) and standard deviation (shaded fill) were computed. The top row in each subfigure displays the IPC in black and the Krylov observability $\mathcal{O}_K$ in purple.

The bottom row shows task performance for the Lorenz forecasting task. The green curve represents the five-step ahead prediction of the x variable with a prediction step size of $\Delta t = 0.1$, while the blue curve shows the cross prediction of the z variable from the measured x dynamics.

With $V = 10$ measurements [Fig. 8(a)], both Krylov observability $\mathcal{O}_K$ and the information processing capacity IPC exhibit a rapid increase and reach visible saturation at $T = 6$ and $T = 2$, respectively. Task performance also decreases rapidly and reaches a saturation region around $T = 6$, similarly to Krylov observability.

Increasing the number of measurements to $V = 30$ yields the results shown in the second row [Fig. 8(b)]. Here, IPC reaches a maximum of approximately 170 at $T = 8$, likely due to the increased number of readout nodes. $\mathcal{O}_K$ saturates around $T = 11$, which better corresponds to the saturation of Lorenz task performance observed around $T = 15$.

Finally, the system is measured $V = 50$ times, with results shown in the third row [Fig. 8(c)]. In this case, IPC and $\mathcal{O}_K$ reach saturation at $T = 16$ and $T = 26$, respectively, while visible saturation in Lorenz task performance occurs at $T = 23$. Once again, Krylov observability better captures the trend in task performance compared to IPC, especially when the system is undersampled, thereby showcasing the generality of these results.

C. Computational cost

We conclude this section with a brief discussion on the computational cost associated with constructing the state matrix $\mathbf{S}$ for the computation of IPC and Krylov observability, as discussed in Secs. II and III.

In our case, with $N_{\mathrm{Tr}} = 25\,000$, $N_{\mathrm{Te}} = 5000$, $V = 30$, and four measured observables ($K = 4$), the total number of matrix multiplications required for constructing the state matrix is given by Eq. (18) to

$$N_{\mathrm{state}} = (7 + K)N_{\mathbf{u}}V = (7 + 4) \times 30\,000 \times 30 = 9.9 \times 10^6. \tag{52}$$

The number of matrix multiplications required for computing Krylov observability by Eq. (43) is

$$N_{\mathrm{obs}} = \frac{VK(VK + 3)}{2} = \frac{30 \times 4(30 \times 4 + 3)}{2} = 7380. \tag{53}$$

The computation of Krylov observability therefore requires only

$$r = \frac{N_{\mathrm{obs}}}{N_{\mathrm{state}}} = 0.75\%$$

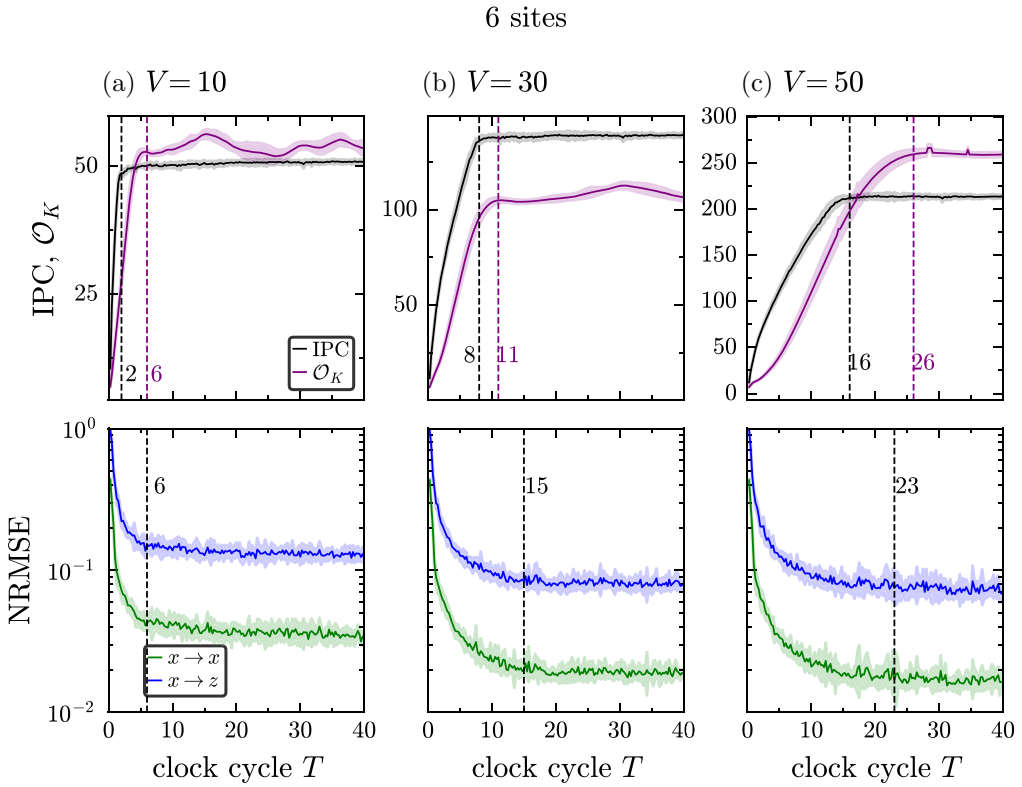


FIG. 8. The first row shows the IPC in black and the Krylov observability $\mathcal{O}_K$ in purple. The second row shows the Lorenz task performance: the five-step ahead prediction ($\Delta t = 0.1$) of the x variable (green) and the cross prediction of the z variable (blue), both plotted against the clock cycle T, for different numbers of measurements: (a) $V = 10$, (b) $V = 30$, and (C) $V = 50$. The state matrix is constructed by measuring all sites in the Pauli-z direction. The interspin couplings are sampled uniformly with $J_{ij} \in \mathcal{U}([0.25, 0.75])$. The bold lines in this plot represent the average over ten different Hamiltonians, and the shaded regions indicate the standard deviation of each curve.

of the matrix multiplications required for constructing the state matrix, demonstrating a three-orders-of-magnitude speedup. It is important to note that this estimate does not include the actual training across thousands of tasks, which involves inversion of the state matrix and optimization of readout weights. These steps are typically more computationally intensive than the construction of the state matrix itself, further amplifying the efficiency gains offered by the Krylov operator complexity compared to IPC.

V. DISCUSSION AND CONCLUSION

This work investigates Krylov-based information measures to explain and predict task performance in quantum reservoir computing, using the Lorenz63 system for chaotic time-series prediction. Initially, fidelity and spread complexity are computed, but they fail to explain task performance saturation (Fig. 2).

We then extended our research to Krylov expressivity $\mathcal{E}_K$ [35] and Krylov observability $\mathcal{O}_K$ [34] and show that both measures show saturation for larger clock cycles. The measures are compared to the information processing capacity IPC, which captures how well the system can retain and map data nonlinearly. We test four quantum reservoirs in an undersampled regime, where Krylov observability $\mathcal{O}_K$ outperforms IPC in explaining the trend in task performance. In this case, IPC reaches a maximum, while the error for the Lorenz tasks continues to decrease with larger clock cycles T.

The quantum reservoirs are then simulated when all sites are measured, as shown in Fig. 4. The first observation from this experiment is that IPC and $\mathcal{O}_K$ exhibit almost identical behavior. Another significant result is the superior Lorenz task performance and larger IPC of H_{I3} compared to H_{I2}, despite H_{I3} having a smaller Krylov expressivity. While Krylov expressivity provides insights into the space onto which the data are mapped, the amount of information extracted from that space is smaller for H_{I2}, as evidenced by the smaller Krylov observability $\mathcal{O}_K$ of H_{I2} compared to H_{I3}. To better understand this, we discuss the grade of Krylov state spaces in accordance with Ref. [35] in Theorem 1 and introduce a similar identity for Krylov operator spaces in Theorem 2. This is then used to show that the smaller Krylov operator space of H_{I2} is due to an increased number of zero contributions compared to H_{I3}, which explains the overall larger Krylov observability for H_{I3}.

To confirm that our measures are not dependent on system size, we computed the information processing capacity and Krylov observability for a five-qubit system over various clock cycles and number of measurements, as shown in Fig. 7, achieving correlation factors between IPC and Krylov observability of $P_C \geqslant 0.97$. We then simulate a six-site Ising model with random interspin couplings, as is commonly done in quantum reservoir computing. This allows for statistical analysis of these measures, where we show that the results

match the behavior observed in the four- and five-site systems, in which Krylov observability better explains the trend in task behavior when the system is undersampled (Fig. 8). Since Krylov expressivity considers how input states are mapped onto the Krylov space, information about the input encoding can be gained. In quantum reservoir computing, this might not be of much importance, because the state evolves over time. In quantum machine learning, however, where input encoding is one of the main parts to be optimized, Krylov expressivity can be used to effectively understand and compare various encoding strategies. One approach would be to encode the input data x_i through a unitary circuit $U_E(x_i)$ into the quantum reservoir or quantum machine learning network. The initial states in the computation of Krylov expressivity can then be sampled from the set of encoded states, i.e., $|x\rangle \in \{U_E(x_i)|s\rangle \mid x_i \in \mathcal{X}\}$, where $\mathcal{X}$ is the set of inputs. The utility of Krylov observability might be used to further gain understanding in quantum machine learning, such as barren plateaus or as a quantum-mechanical information measure that can be utilized in the understanding of quantum dynamics, further advancing knowledge in chaos, decoherence, or thermalization.

ACKNOWLEDGMENTS

We acknowledge support for the publication costs by the Open Access Publication Fund of the Technische Universitt Ilmenau.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX A: ALGORITHM FOR THE CONSTRUCTION OF THE SPACES

Consider the basis $\mathcal{F}^B$ and $\mathcal{F}_i^B$ of the spaces $\mathcal{F}$ and $\mathcal{F}_i$, i.e., $\mathcal{F} = \mathrm{Span}(\mathcal{F}^B)$ and $\mathcal{F}_i = \mathrm{Span}(\mathcal{F}_i^B)$. Then, this Appendix constructs spaces such that the properties given in Eq. (39) are defined.

ALGORITHM 1. Construction of observability spaces.

$\mathcal{I}_O = \{1, 2, \ldots, K\}$
$\mathcal{T} = \{t_1, t_2, \ldots, t_R\}$
$\mathcal{F}^{(B)} = \emptyset$
$\tilde{\mathcal{F}}_1^{(B)}, \tilde{\mathcal{F}}_2^{(B)}, \ldots, \tilde{\mathcal{F}}_K^{(B)} = \emptyset, \emptyset, \ldots, \emptyset$
while $t_j \in \mathcal{T}$ **do**
  for $k \in \mathcal{I}_O$ **do**
    if $O_k(t_j) \notin \mathcal{F}$ **then**
      $\mathcal{F}^{(B)} = \mathcal{F}^{(B)} \cup O_k(t_j)$
      $\tilde{\mathcal{F}}_k^{(B)} = \tilde{\mathcal{F}}_k^{(B)} \cup O_k(t_j)$
    end if
  end for
end while
$\mathcal{F}^{(B)} \leftarrow \mathrm{ONB}(\mathcal{F}^{(B)})$

APPENDIX B: SATURATION

Table V gives information when each measure exhibits a saturation point.

APPENDIX C: PROOFS FOR KRYLOV STATE SPACES AND THEOREM 1

Theorem 1 (Repeated). Let $H \in \mathbb{C}^{N\times N}$ be a Hermitian Hamiltonian with d pairwise distinct eigenvalues $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{d-1}$, and let $\{|\phi_j\rangle\}$ be an orthonormal eigenbasis of H, satisfying

$$H|\phi_j\rangle = \varepsilon_j|\phi_j\rangle. \tag{C1}$$

Then, the time-evolved state $|\Psi(t)\rangle = e^{-iHt}|\Psi_0\rangle$ lies in a d-dimensional subspace $E_d \subseteq \mathbb{C}^N$, i.e.,

$$|\Psi(t)\rangle \in \mathrm{E}_d := \mathrm{Span}\{|\xi_0\rangle, |\xi_1\rangle, \ldots, |\xi_{d-1}\rangle\}, \tag{C2}$$

where the vectors $|\xi_p\rangle$ are defined by

$$|\xi_p\rangle := \frac{1}{\sqrt{|J_p|}} \sum_{j\in J_p} \alpha_j|\phi_j\rangle, \quad \text{with } \alpha_j := \langle\phi_j|\Psi_0\rangle \tag{C3}$$

and $J_p := \{j \mid \varepsilon_j = \varepsilon_p\}$ denotes the set of indices corresponding to the degenerate eigenspace of eigenvalue ε_p. The normalization factor $|J_p|$ is the cardinality of the set J_p. The starting state $|\psi_0\rangle$ can be represented in the basis $\{|\xi_p\rangle\}_p$, with $\gamma_p = \langle\xi_p|\psi_0\rangle$, as

$$|\psi_0\rangle = \sum_{p=0}^{d-1} \gamma_p|\xi_p\rangle. \tag{C4}$$

Let n_1 denote the number of coefficients for which $\gamma_p = 0$. Then, the number of linearly independent vectors is reduced to $d - n_1$. It also holds that for the Krylov state space $K_m = \mathrm{Span}\{|\psi_0\rangle, H|\psi_0\rangle, \ldots, H^{m-1}|\psi_0\rangle\}$,

$$m = d - n_1 \tag{C5}$$

holds.

Proof. Starting with the initial state $|\Psi_0\rangle \in \mathbb{C}^N$, expand it in the eigenbasis of H:

$$|\Psi_0\rangle = \sum_j \alpha_j|\phi_j\rangle, \alpha_j = \langle\phi_j|\Psi_0\rangle. \tag{C6}$$

The time-evolved state is then given by

$$|\Psi(t)\rangle = e^{-iHt}|\Psi_0\rangle = \sum_j e^{-i\varepsilon_j t}\alpha_j|\phi_j\rangle. \tag{C7}$$

Since the eigenvalues ε_j take only d distinct values, we partition the index set $\{0, 1, \ldots, N-1\}$ into disjoint subsets J_p, where

$$J_p := \{j \in \{0, \ldots, N-1\} \mid \varepsilon_j = \varepsilon_p\}, \text{ for } p = 0, \ldots, d-1.$$

TABLE V. Information about the saturation behavior, saturation time, and saturation points of the different Hamiltonians.

	$\mathcal{E}_K$ T_{sat}	$\mathcal{O}_K$ T_{sat}	IPC T_{sat}	IPC saturation point	$x \to x$ T_{sat}	$x \to z$ T_{sat}	$x \to x$, $\text{NRMSE}_{\text{sat}}$	$x \to z$ $\text{NRMSE}_{\text{sat}}$
H_1, Z_1	6	12	12	20	12	12	0.08	0.3
H_1, all Z_i	6	10	34	40	12	12	0.08	0.3
H_2, Z_1	8	17	15	30	18	18	0.04	0.2
H_2, all Z_i	8	20	22	60	18	18	0.03	0.15
H_3, Z_1	8	20	13	30	18	18	0.06	0.16
H_3, all Z_i	8	20	20	87	20	20	0.03	0.08
H_4, Z_1	8	20	13	30	18	18	0.06	0.16
H_4, all Z_i	8	20	20	87	20	20	0.03	0.08

This lets us rewrite the time-evolved state as

$$|\Psi(t)\rangle = \sum_j e^{-i\varepsilon_j t}\alpha_j|\phi_j\rangle = \sum_{p=0}^{d-1}\sum_{j\in J_p} e^{-i\varepsilon_p t}\alpha_j|\phi_j\rangle$$
$$= \sum_{p=0}^{d-1} e^{-i\varepsilon_p t}\left(\sum_{j\in J_p}\alpha_j|\phi_j\rangle\right). \tag{C8}$$

Define the time-independent vectors

$$|\xi_p\rangle := \frac{1}{\sqrt{|J_p|}}\sum_{j\in J_p}\alpha_j|\phi_j\rangle, \tag{C9}$$

which represent normalized superpositions. Then, the time-evolved state becomes

$$|\Psi(t)\rangle = \sum_{p=0}^{d-1} e^{-i\varepsilon_p t}\sqrt{|J_p|}|\xi_p\rangle$$
$$\Rightarrow |\Psi(t)\rangle \in \text{Span}\{|\xi_0\rangle, \ldots, |\xi_{d-1}\rangle\} = E_d. \tag{C10}$$

Thus, $|\Psi(t)\rangle$ lies in the d-dimensional subspace E_d for all $t \in \mathbb{R}$. Given the starting state $|\psi_0\rangle$, it can be represented in the basis $\{|\xi_p\rangle\}_p$, with $\gamma_p = \langle\xi_p|\psi_0\rangle$, as

$$|\psi_0\rangle = \sum_{p=0}^{d-1}\gamma_p|\xi_p\rangle. \tag{C11}$$

This is a superposition of d linearly independent vectors. Let n_1 denote the number of coefficients for which $\gamma_p = 0$. Then, the number of linearly independent vectors is reduced to $d - n_1$. Since $|\xi_p\rangle$ is given by

$$|\xi_p\rangle = \sum_{j\in J_p}\alpha_j|\phi_j\rangle, \tag{C12}$$

it follows that $\gamma_p = 0$ only if all $\alpha_j = \langle\phi_j|\Psi_0\rangle = 0$ for all $j \in J_p$. Since Lemmas C1 and C2 show that both K_m and E_{d-n_1} consist of the smallest possible number of basis states, it follows that the dimension of both must be equal, i.e., $m = d - n_1$. ■

Lemma C1 (E_m consists of the minimum number of basis states.). Given an initial state $|\Psi_0\rangle$, a Hamiltonian H, and the corresponding space of eigenstates E_d, such that $|\Psi(t)\rangle \in K_m$ for all $t \in \mathbb{R}$, there exists no basis $\mathcal{B}$ with $\dim(\mathcal{B}) < d$ such that any time-evolved state is in the span of $\mathcal{B}$.

Proof. Assume times $t_0 < t_1 < \cdots < t_{d-1}$ and the states evolved at those times are given by

$$|\Psi(t_j)\rangle = e^{-iHt_j}|\Psi_0\rangle = \sum_{k=0}^{N} e^{-i\varepsilon_k t_j}|\phi_k\rangle\langle\phi_k|\Psi_0\rangle$$
$$= \sum_{p=0}^{d-1} e^{-i\varepsilon_p t_j}\sum_{j\in J_p}\alpha_j|\phi_j\rangle = \sum_{p=0}^{d-1} e^{-i\varepsilon_p t_j}|\xi_p\rangle$$
$$= \begin{pmatrix}|\xi_0\rangle & |\xi_1\rangle & \cdots & |\xi_{d-1}\rangle\end{pmatrix}\begin{pmatrix} e^{-i\varepsilon_0 t_j} \\ e^{-i\varepsilon_1 t_j} \\ \vdots \\ e^{-i\varepsilon_{d-1} t_j}\end{pmatrix}. \tag{C13}$$

Here, $|\xi_p\rangle = \sum_{j\in J_p}\alpha_j|\phi_j\rangle$ and $\alpha_j = \langle\phi_j|\Psi_0\rangle$ is used. We do not normalize $|\xi_i\rangle$, since it does not change anything but allows better readability. Writing the d states in terms of the $|\xi_p\rangle$ basis results in

$$(|\Psi(t_0)\rangle \cdots |\Psi(t_{d-1})\rangle) = (|\xi_0\rangle \cdots |\xi_{d-1}\rangle)\Sigma,$$
$$\Sigma = \begin{pmatrix} e^{-i\varepsilon_0 t_0} & e^{-i\varepsilon_0 t_1} & \cdots & e^{-i\varepsilon_0 t_{d-1}} \\ e^{-i\varepsilon_1 t_0} & e^{-i\varepsilon_1 t_1} & \cdots & e^{-i\varepsilon_1 t_{d-1}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-i\varepsilon_{d-1} t_0} & e^{-i\varepsilon_{d-1} t_1} & \cdots & e^{-i\varepsilon_{d-1} t_{d-1}}\end{pmatrix}. \tag{C14}$$

If the matrix Σ is invertible, then it follows that any d time-evolved states span E_d. If $|\varepsilon_j| < \pi$, we can write $x_j^t = (e^{-i\varepsilon_j})^t$. If this is not the case, the Hamiltonian can be rescaled as $H \leftarrow H/|\varepsilon_{\max}|$ and time as $t \leftarrow t \cdot |\varepsilon_{\max}|$, which results in the same dynamics. With this, the matrix in Eq. (C14) can be rewritten as

$$(|\Psi(t_0)\rangle \cdots \quad |\Psi(t_{d-1})\rangle) = (|\xi_0\rangle \cdots \quad |\xi_{d-1}\rangle)\Sigma, \tag{C15}$$

where

$$\Sigma = \begin{pmatrix} x_0^{t_0} & x_0^{t_1} & \cdots & x_0^{t_{d-1}} \\ x_1^{t_0} & x_1^{t_1} & \cdots & x_1^{t_{d-1}} \\ \vdots & \vdots & \ddots & \vdots \\ x_{d-1}^{t_0} & x_{d-1}^{t_1} & \cdots & x_{d-1}^{t_{d-1}}\end{pmatrix}. \tag{C16}$$

For pairwise distinct times, the columns are linearly independent if $t_{d-1} < T_P$, where T_P is the period of the system. Σ is a generalized Vandermonde matrix, for which an inverse exists. Therefore,

$$(|\Psi(t_0)\rangle \cdots \quad |\Psi(t_{d-1})\rangle)\Sigma^{-1} = (|\xi_0\rangle \cdots \quad |\xi_{d-1}\rangle). \tag{C17}$$

Since all time-evolved states lie in E_d, and since these d time-evolved states span the space, it follows that E_d consists of the minimum number of basis vectors. ■

Lemma C2 (K_m consists of the minimum number of basis states.). Assume $|\Psi(t)\rangle$ is a time-evolved state under Hamiltonian H and assume that the corresponding Krylov space is given by K_m, such that $|\Psi(t)\rangle \in K_m$ for all $t \in \mathbb{R}$. Then, it holds that there exists no basis $\mathcal{B}$ with $\dim(\mathcal{B}) < m$ such that $|\Psi(t)\rangle \in \mathrm{Span}\{\mathcal{B}\}$.

Proof. The proof proceeds as follows. Take $L > m-1$ time-evolved states $|\Psi(t_j)\rangle = e^{-iHt_j}|\Psi_0\rangle$ at times $0 = t_0 < t_1 < t_2 < \cdots < t_L < T_P$. Pick L very large, such that

$$|\Psi(t_j)\rangle = \sum_{k=0}^{\infty}(-iH)^k\frac{t_j^k}{k!}|\Psi_0\rangle = \sum_{k=0}^{L}(-iH)^k\frac{t_j^k}{k!}|\Psi_0\rangle + \varepsilon \tag{C18}$$

holds with $\varepsilon \to 0$ for $L \to \infty$. Introduce the substitution

$$h_j(|\Psi_0\rangle) := \sum_{k=0}^{L} f^k(|\Psi_0\rangle)\frac{t_j^k}{k!}, \tag{C19}$$

with

$$f^k(|\Psi_0\rangle) := (-iH)^k|\Psi_0\rangle.$$

For readability, the dependence on $|\Psi_0\rangle$ is ignored in the following calculations. The vectors h_j can be written as

$$h_i = \sum_{j=0}^{n} f^j\frac{t_i^j}{j!} = (f^0 \;\; f^1 \;\; \cdots \;\; f^n)\begin{pmatrix}1\\ t_i/1!\\ t_i^2/2!\\ \vdots\\ t_i^n/n!\end{pmatrix}. \tag{C20}$$

Writing all vectors $h_0, \ldots, h_n$ yields

$$(h_0 \;\; h_1 \;\; \cdots \;\; h_n) = (f^0 \;\; f^1 \;\; \cdots \;\; f^n)\Theta,$$
$$\Theta = \begin{pmatrix}1 & 1 & \cdots & 1\\ t_1/1! & t_2/1! & \cdots & t_n/1!\\ t_1^2/2! & t_2^2/2! & \cdots & t_n^2/2!\\ \vdots & \vdots & \ddots & \vdots\\ t_1^n/n! & t_2^n/n! & \cdots & t_n^n/n!\end{pmatrix}. \tag{C21}$$

Since all times t_i are pairwise distinct by definition, the columns of Θ are linearly independent. Therefore, Θ is invertible, and its inverse Θ^{-1} exists:

$$(h_0 \;\; h_1 \;\; \cdots \;\; h_n)\Theta^{-1} = (f^0 \;\; f^1 \;\; \cdots \;\; f^n). \tag{C22}$$

Since all vectors $\{f^0, f^1, \ldots, f^L\}$ can be represented as linear combinations of vectors $\{h_0, h_1, \ldots, h_L\}$, the spans of both sides are equal:

$$\begin{aligned}H_L^L &= \mathrm{Span}\{h_0, h_1, \ldots, h_n\}\\ &= \mathrm{Span}\{f^0, f^1, \ldots, f^L\} = K_L.\end{aligned} \tag{C23}$$

Since K_L contains only m linearly independent vectors, H_L^L must also contain only m linearly independent vectors ($H_m^L = K_m$). This implies that there cannot exist a basis $\mathcal{B}$ with $\dim(\mathcal{B}) < m$ such that $|\Psi(t)\rangle \in \mathrm{Span}\{\mathcal{B}\}$. If such a basis existed, then the L vectors h_j would not be able to represent the vectors f^j, and the matrix Θ would not be invertible, which would be contradicting the fact that Θ is indeed invertible. ■

APPENDIX D: PROOFS FOR OPERATOR SPACES AND THEOREM 2

Theorem 2 (Repeated). Let $H \in \mathbb{C}^{N\times N}$ be a Hamiltonian with eigenbasis $\{|\phi_j\rangle\}$ and corresponding eigenvalues ε_j, and let O be an operator on the same Hilbert space. Define the Liouvillian Krylov space

$$\mathcal{L}_M := \mathrm{Span}\{\mathcal{L}^0(O), \mathcal{L}^1(O), \ldots, \mathcal{L}^{M-1}(O)\},$$

where $\mathcal{L}(O) = [H, O]$ is the Liouvillian superoperator.

Define the transition frequencies $\omega_{mn} := \varepsilon_m - \varepsilon_n$, and let $\{\omega_P\}_{P=0}^{N_\omega-1}$ be the set of all pairwise distinct values taken by ω_{mn}. For each ω_P, define the index set

$$J_P := \{(m,n) \,|\, \omega_{mn} = \omega_P\}$$

and the corresponding matrix

$$\sigma_P := \sum_{(m,n)\in J_P}\langle\phi_m|O|\phi_n\rangle|\phi_m\rangle\langle\phi_n|.$$

Let N_1 be the number of vanishing contributions $\sigma_P = 0$, then the time-evolved operator is given by

$$O(t) = \sum_{P\in S} e^{i\omega_P t}\sigma_P,$$

with $S = \{P \mid \sigma_P \neq 0\} = \{s_0, s_1, \ldots, s_{N_\omega - N_1 - 1}\}$. The operator lies in the span

$$O(t) \in \mathrm{Span}\{\sigma_{s_0}, \sigma_{s_1}, \ldots, \sigma_{s_{N_\omega-N_1-1}}\} = \mathcal{P}_{N_\omega - N_1}.$$

Further, the grade M of the Krylov space $\mathcal{L}_M$ is given by

$$M = N_\omega - N_1.$$

Proof. The time-evolved operator is given by

$$O(t) = e^{iHt}Oe^{-iHt} = \sum_{m,n} e^{i(\varepsilon_m-\varepsilon_n)t}O_{mn}|\phi_m\rangle\langle\phi_n|,$$

with matrix elements $O_{mn} = \langle\phi_m|O|\phi_n\rangle$. Define the transition frequencies $\omega_{mn} := \varepsilon_m - \varepsilon_n$, and let $\{\omega_P\}_{P=0}^{N_\omega-1}$ be the set of all pairwise distinct values taken by ω_{mn}. For each ω_P, define the index set

$$J_P := \{(m,n) \mid \omega_{mn} = \omega_P\}.$$

Then, the operator can be grouped as

$$\begin{aligned}O(t) &= \sum_{P=0}^{N_\omega-1} e^{i\omega_P t}\sum_{(m,n)\in J_P} O_{mn}|\phi_m\rangle\langle\phi_n|\\ &= \sum_{P=0}^{N_\omega-1} e^{i\omega_P t}\sigma_P,\end{aligned}$$

where we define

$$\sigma_P := \sum_{(m,n)\in J_P} O_{mn}|\phi_m\rangle\langle\phi_n|.$$

The functions $e^{i\omega_P t}$ are linearly independent over $\mathbb{R}$ when the frequencies ω_P are pairwise distinct. Due to the Hermiticity of H, the vectors $|\phi_m\rangle$ are linearly independent, which

implies the linear independence of σ_P. Let N_1 denote the number of such zero contributions, i.e., $\sigma_P = 0$. Then, the sum reduces to

$$O(t) = \sum_{P\in S} e^{i\omega_P t}\sigma_P,$$

where $S = \{P \mid \sigma_P \neq 0\} = \{s_0, s_1, \ldots, s_{N_\omega - N_1 - 1}\}$ and the operator lies in the span

$$O(t) \in \mathrm{Span}\{\sigma_{s_0}, \sigma_{s_1}, \ldots, \sigma_{s_{N_\omega - N_1 - 1}}\} = \mathcal{P}_{N_\omega - N_1}.$$

By Lemma D1, it holds that this space is minimal; i.e, there exists no basis $\mathcal{B}$ with $\dim(\mathcal{B}) < N_\omega - N_1$ such that $O(t) \in \mathrm{Span}(\mathcal{B})$. On the other hand, the time-evolved operator lies in the Liouvillian Krylov space,

$$O(t) \in \mathcal{L}_M = \mathrm{Span}\{\mathcal{L}^0(O), \mathcal{L}^1(O), \ldots, \mathcal{L}^{M-1}(O)\}.$$

Since due to Lemma D2 it holds $\mathcal{L}^{M-1}(O)\}$ is a minimal space as well, we conclude that

$$M = N_\omega - N_1.$$

■

Lemma D1 ($\mathcal{P}_{N_\omega - N_1}$ consists of the minimum number of basis operators.). Given an operator $O \in \mathbb{C}^{N\times N}$, a Hamiltonian H with eigenvalues ε_j, and the associated decomposition of the time-evolved operator $O(t) = e^{iHt}Oe^{-iHt}$, let $\mathcal{P}_{N_\omega - N_1} = \mathrm{Span}\{\sigma_0, \ldots, \sigma_{N_\omega - N_1 - 1}\}$ denote the space generated by the $N_\omega - N_1$ nonzero frequency components in the Liouvillian decomposition as discussed in Theorem 2. Then, there exists no smaller set of operators that spans all $O(t)$ for arbitrary time $t \in \mathbb{R}$.

Proof. Assume times $t_0 < t_1 < \cdots < t_{N_\omega - N_1 - 1}$ and write the time-evolved operators in the basis discussed in Theorem 2 at those times:

$$\begin{aligned} O(t_j) &= \sum_{P=0}^{N_\omega - N_1 - 1} e^{i\omega_P t_j}\sigma_P \\ &= \begin{pmatrix}\sigma_0 & \sigma_1 & \cdots & \sigma_{N_\omega - N_1 - 1}\end{pmatrix}\begin{pmatrix} e^{i\omega_0 t_j} \\ e^{i\omega_1 t_j} \\ \vdots \\ e^{i\omega_{N_\omega - N_1 - 1} t_j}\end{pmatrix}. \end{aligned} \tag{D1}$$

Stacking the $N_\omega - N_1$ operators into a matrix gives

$$\begin{pmatrix}O(t_0) & \cdots & O(t_{N_\omega - N_1 - 1})\end{pmatrix} = \begin{pmatrix}\sigma_0 & \cdots & \sigma_{N_\omega - N_1 - 1}\end{pmatrix}\Sigma, \tag{D2}$$

where

$$\Sigma = \begin{pmatrix} e^{i\omega_0 t_0} & \cdots & e^{i\omega_0 t_{N_\omega - N_1 - 1}} \\ e^{i\omega_1 t_0} & \cdots & e^{i\omega_1 t_{N_\omega - N_1 - 1}} \\ \vdots & \ddots & \vdots \\ e^{i\omega_{N_\omega - N_1 - 1} t_0} & \cdots & e^{i\omega_{N_\omega - N_1 - 1} t_{N_\omega - N_1 - 1}}\end{pmatrix}. \tag{D3}$$

This matrix Σ is a generalized Vandermonde matrix in the variables $x_P := e^{i\omega_P}$. If the frequencies ω_P are distinct and the time points t_j are also distinct (and less than the system period T_P), then Σ is invertible. This holds due to the construction of $\mathcal{P}_{N_\omega - N_1}$ in Theorem 2.

Therefore, we can recover σ_P as

$$\begin{pmatrix}\sigma_0 & \cdots & \sigma_{N_\omega - N_1 - 1}\end{pmatrix} = \begin{pmatrix}O(t_0) & \cdots & O(t_{N_\omega - N_1 - 1})\end{pmatrix}\Sigma^{-1}. \tag{D4}$$

Hence, the $N_\omega - N_1$ time-evolved operators span the space $\mathcal{P}_{N_\omega - N_1}$, and no smaller set of operators can generate the time evolution. Therefore, $\mathcal{P}_{N_\omega - N_1}$ is minimal. ■

Lemma D2 ($\mathcal{L}_M$ Consists of the Minimum Number of Basis States). Let O be an operator and H the Hamiltonian under which O evolves. Then, there exists no space W such that $O(t) \in W$ for all t and $\dim(W) < \dim(\mathcal{L}_M)$.

Proof. Let $0 = t_0 < t_1 < \cdots < t_Q$ be a time discretization with $t_Q < T_P$, where T_P is the period of the system. Define

$$\tilde{O}(t_a) = \sum_{k=0}^{Q} \frac{(it_a)^k}{k!}\mathcal{L}^k(O)$$

as an approximation of the time-evolved operator at time t_a. Then,

$$\tilde{O}(t_a) = (i^0\mathcal{L}^0(O), i^1\mathcal{L}^1(O), \ldots, i^Q\mathcal{L}^Q(O))\begin{pmatrix}1 \\ t_a \\ \vdots \\ t_a^Q/Q!\end{pmatrix}.$$

Writing all $\tilde{O}(t_a)$ with pairwise distinct t_a gives

$$(\tilde{O}(t_0), \ldots, \tilde{O}(t_Q)) = (i^0\mathcal{L}^0(O), \ldots, i^Q\mathcal{L}^Q(O))\Theta, \tag{D5}$$

with

$$\Theta = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_0 & t_1 & \cdots & t_Q \\ t_0^2/2! & t_1^2/2! & \cdots & t_Q^2/2! \\ \vdots & \vdots & \ddots & \vdots \\ t_0^Q/Q! & t_1^Q/Q! & \cdots & t_Q^Q/Q!\end{pmatrix}. \tag{D6}$$

This is the same representation as in Lemma C2, where Θ is a generalized Vandermonde matrix and is thus invertible. Therefore,

$$(\tilde{O}(t_0), \ldots, \tilde{O}(t_Q))\Theta^{-1} = (i^0\mathcal{L}^0(O), \ldots, i^Q\mathcal{L}^Q(O)). \tag{D7}$$

This shows that the time-evolved operators can reconstruct the powers of the Liouvillian and since for all time-evolved operators $O(t) \in \mathcal{L}_M$ holds, and each Liouvillian power can be expressed using the time-evolved operators, it follows that $\mathcal{L}_M$ has the minimum possible number of basis elements. ■

APPENDIX E: KRYLOV SPACE DIMENSION FOR THE HAMILTONIANS

This appendix lists spectral and dynamical statistics for pairs of Hamiltonians and observables. Table VI reports the number of eigenvalues d, eigenvalue pairs d^2, transition frequencies N_ω, vanishing contributions N_1, and the Krylov dimension M.

TABLE VI. Spectral and dynamical statistics for each (Hamiltonian, observable) pair. d is the number of pairwise distinct eigenvalues, d^2 is the number of eigenvalue pairs, n_ω is the number of distinct transition frequencies, N_1 is the number of zero σ_P contributions, and M is the resulting Krylov dimension.

Hamiltonian	Observable	d	d^2	N_ω	N_1	M
H_{I1}	Z_1	9	81	71	40	31
H_{I1}	Z_2	9	81	71	40	31
H_{I1}	Z_3	9	81	71	40	31
H_{I1}	Z_4	9	81	71	40	31
H_{I2}	Z_1	16	256	237	176	61
H_{I2}	Z_2	16	256	237	176	61
H_{I2}	Z_3	16	256	237	158	79
H_{I2}	Z_4	16	256	237	158	79
H_{I3}	Z_1	15	225	211	112	99
H_{I3}	Z_2	15	225	211	112	99
H_{I3}	Z_3	15	225	211	112	99
H_{I3}	Z_4	15	225	211	112	99
H_{I4}	Z_1	16	256	241	128	113
H_{I4}	Z_2	16	256	241	128	113
H_{I4}	Z_3	16	256	241	128	113
H_{I4}	Z_4	16	256	241	128	113

[1] H. Jaeger, The "echo state" approach to analysing and training recurrent neural networks, GMD Report 148, GMD Forschungszentrum Informationstechnik, 2001, https://doi.org/10.24406/publica-fhg-291111.

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

[3] D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt, An experimental unification of reservoir computing methods, Neural Netw. **20**, 391 (2007).

[4] L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, Information processing using a single dynamical node as complex system, Nat. Commun. **2**, 468 (2011).

[5] L. Mühlnickel, J. A. Jaurigue, L. C. Jaurigue, and K. Lüdge, The influence of timescales and data injection schemes for reservoir computing using spin-VCSELs, Commun. Phys. **7**, 370 (2024).

[6] L. C. Jaurigue and K. Lüdge, Reducing reservoir computer hyperparameter dependence by external timescale tailoring, Neuromorph. Comput. Eng. **4**, 014001 (2024).

[7] J. A. Jaurigue, J. Robertson, A. Hurtado, L. C. Jaurigue, and K. Lüdge, Post-processing methods for delay embedding and feature scaling of reservoir computers, Commun. Eng. **4**, 10 (2025).

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

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

[10] A. Kutvonen, K. Fujii, and T. Sagawa, Optimizing a quantum reservoir computer for time series prediction, Sci. Rep. **10**, 14687 (2020).

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

[12] W. Xia, J. Zou, X. Qiu, and X. Li, The reservoir learning power across quantum many-body localization transition, Front. Phys. **17**, 33506 (2022).

[13] N. Götting, F. Lohof, and C. Gies, Exploring quantumness in quantum reservoir computing, Phys. Rev. A **108**, 052427 (2023).

[14] 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).

[15] A. H. Abbas, H. Abdel-Ghani, and I. S. Maksymov, Classical and quantum physical reservoir computing for onboard artificial intelligence systems: A perspective, arXiv:2407.04717v1.

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

[17] Y. Suzuki, Q. Gao, K. C. Pradel, K. Yasuoka, and N. Yamamoto, Natural quantum reservoir computing for temporal information processing, Sci. Rep. **12**, 1353 (2022).

[18] P. Pfeffer, F. Heyder, and J. Schumacher, Hybrid quantum-classical reservoir computing of thermal convection flow, Phys. Rev. Res. **4**, 033176 (2022).

[19] 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.

[20] A. Sannia, R. Martínez-Peña, M. C. Soriano, G. L. Giorgi, and R. Zambrini, Dissipation as a resource for quantum reservoir computing, Quantum **8**, 1291 (2024).

[21] D. Fry, A. Deshmukh, S. Y. Chen, V. Rastunkov, and V. Markov, Optimizing quantum noise-induced reservoir computing for nonlinear and chaotic time series prediction, Sci. Rep. **13**, 19326 (2023).

[22] L. Domingo, G. Carlo, and F. Borondo, Taking advantage of noise in quantum reservoir computing, Sci. Rep. **13**, 8790 (2023).

[23] 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).

[24] G. Franceschetto, M. Płodzień, M. Lewenstein, A. Acín, and P. Mujal, Harnessing quantum back-action for time-series processing, arXiv:2411.03979.

[25] S. Čindrak, B. Donvil, K. Lüdge, and L. C. Jaurigue, Enhancing the performance of quantum reservoir computing and solving the time-complexity problem by artificial memory restriction, Phys. Rev. Res. **6**, 013051 (2024).

[26] K. Kobayashi, K. Fujii, and N. Yamamoto, Feedback-driven quantum reservoir computing for time-series analysis, PRX Quantum **5**, 040325 (2024).

[27] O. Ahmed, F. Tennie, and L. Magri, Prediction of chaotic dynamics and extreme events: A recurrence-free quantum reservoir computing approach, Phys. Rev. Res. **6**, 043082 (2024).

[28] C. Zhu, P. J. Ehlers, H. I. Nurdin, and D. Soh, Practical and scalable quantum reservoir computing, Phys. Rev. Res. **7**, 023290 (2025).

[29] N. Schütte, N. Götting, H. Müntinga, M. List, and C. Gies, Expressive limits of quantum reservoir computing, arXiv:2501.15528.

[30] E. N. Lorenz, Deterministic nonperiodic flow, J. Atmos. Sci. **20**, 130 (1963).

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

[32] V. Balasubramanian, P. Caputa, J. M. Magan, and Q. Wu, Quantum chaos and the complexity of spread of states, Phys. Rev. D **106**, 046007 (2022).

[33] D. E. Parker, X. Cao, A. Avdoshkin, T. Scaffidi, and E. Altman, A universal operator growth hypothesis, Phys. Rev. X **9**, 041017 (2019).

[34] S. Čindrak, K. Lüdge, and L. C. Jaurigue, companion paper, From Krylov complexity to observability: Capturing phase space dimension with applications in quantum reservoir computing, Phys. Rev. Res. **7**, L042039 (2025).

[35] S. Čindrak, A. Paschke, L. C. Jaurigue, and K. Lüdge, Measurable Krylov spaces and eigenenergy count in quantum state dynamics, J. High Energy Phys. 10 (2024) 083.

[36] M. Alishahiha and S. Banerjee, A universal approach to Krylov state and operator complexities, SciPost Phys. **15**, 080 (2023).

[37] A. Bhattacharya, P. Nandy, P. P. Nath, and H. Sahu, Operator growth and Krylov construction in dissipative open quantum systems, J. High Energy Phys. 12 (2022) 081.

[38] M. Afrasiar, J. K. Basak, B. Dey, and K. Pal, Time evolution of spread complexity in quenched Lipkin–Meshkov–Glick model, J. Stat. Mech. (2023) 103101.

[39] T. Anegawa, N. Iizuka, and M. Nishida, Krylov complexity as an order parameter for deconfinement phase transitions at large N, J. High Energy Phys. 04 (2024) 119.

[40] S. Baek, Krylov complexity in inverted harmonic oscillator, arXiv:2210.06815.

[41] J. L. F. Barbón, E. Rabinovici, R. Shir, and R. Sinha, On the evolution of operator complexity beyond scrambling, J. High Energy Phys. 10 (2019) 264.

[42] B. Bhattacharjee, P. Nandy, and T. Pathak, Operator dynamics in Lindbladian SYK: A Krylov complexity perspective, J. High Energy Phys. 01 (2024) 094.

[43] H. A. Camargo, V. Jahnke, K. Kim, and M. Nishida, Krylov complexity in free and interacting scalar field theories with bounded power spectrum, J. High Energy Phys. 05 (2023) 226.

[44] X. Cao, A statistical mechanism for operator growth, J. Phys. A: Math. Theor. **54**, 144001 (2021).

[45] P. Caputa and S. Datta, Operator growth in 2d CFT, J. High Energy Phys. 12 (2021) 188.

[46] A. Chattopadhyay, A. Mitra, and H. J. R. Van Zyl, Spread complexity as classical dilaton solutions, Phys. Rev. D **108**, 025013 (2023).

[47] A. Dymarsky and M. Smolkin, Krylov complexity in conformal field theory, Phys. Rev. D **104**, L081702 (2021).

[48] Z. Fan, Universal relation for operator complexity, Phys. Rev. A **105**, 062210 (2022).

[49] S. Guo, Operator growth in SU(2) Yang-Mills theory, arXiv:2208.13362.

[50] K. Hashimoto, K. Murata, N. Tanahashi, and R. Watanabe, Krylov complexity and chaos in quantum mechanics, J. High Energy Phys. 11 (2023) 040.

[51] S. He, P. H. C. Lau, Z. Xian, and L. Zhao, Quantum chaos, scrambling and operator growth in $T\overline{T}$ deformed SYK models, J. High Energy Phys. 12 (2022) 070.

[52] R. Heveling, J. Wang, and J. Gemmer, Numerically probing the universal operator growth hypothesis, Phys. Rev. E **106**, 014152 (2022).

[53] N. Iizuka and M. Nishida, Krylov complexity in the IP matrix model, J. High Energy Phys. 11 (2023) 065.

[54] S. Jian, B. Swingle, and Z. Xian, Complexity growth of operators in the SYK model and in JT gravity, J. High Energy Phys. 03 (2021) 014.

[55] J. Kim, J. Murugan, J. Olle, and D. Rosa, Operator delocalization in quantum networks, Phys. Rev. A **105**, L010201 (2022).

[56] X. Li, Q. Zhu, C. Zhao, X. Duan, B. Zhao, X. Zhang, H. Ma, J. Sun, and W. Lin, Higher-order granger reservoir computing: Simultaneously achieving scalable complex structures inference and accurate dynamics prediction, Nat. Commun. **15**, 2506 (2024).

[57] C. Liu, H. Tang, and H. Zhai, Krylov complexity in open quantum systems, Phys. Rev. Res. **5**, 033085 (2023).

[58] J. M. Magan and J. Simón, On operator growth and emergent Poincaré symmetries, J. High Energy Phys. 05 (2020) 071.

[59] W. Mück and Y. Yang, Krylov complexity and orthogonal polynomials, Nucl. Phys. B **984**, 115948 (2022).

[60] A. A. Nizami and A. W. Shrestha, Krylov construction and complexity for driven quantum systems, Phys. Rev. E **108**, 054222 (2023).

[61] D. Patramanis, Probing the entanglement of operator growth, Prog. Theor. Phys. **2022**, 063A01 (2022).

[62] E. Rabinovici, A. Sánchez-Garrido, R. Shir, and J. Sonner, Krylov localization and suppression of complexity, J. High Energy Phys. 03 (2022) 211.

[63] E. Rabinovici, A. Sánchez-Garrido, R. Shir, and J. Sonner, Krylov complexity from integrability to chaos, J. High Energy Phys. 07 (2022) 151.

[64] M. J. Vasli, K. B. Velni, M. R. M. Mozaffar, A. Mollabashi, and M. Alishahiha, Krylov complexity in Lifshitz-type scalar field theories, Eur. Phys. J. C **84**, 235 (2024).

[65] P. Caputa, H. Jeong, S. Yi, J. F. Pedraza, and L. Qu, Krylov complexity of density matrix operators, J. High Energy Phys. 05 (2024) 337.

[66] V. Balasubramanian, J. M. Magan, and Q. Wu, Tridiagonalizing random matrices, Phys. Rev. D **107**, 126001 (2023).

[67] B. Craps, O. Evnin, and G. Pascuzzi, A relation between Krylov and Nielsen complexity, Phys. Rev. Lett. **132**, 160402 (2024).

[68] S. E. Aguilar-Gutierrez and A. Rolph, Krylov complexity is not a measure of distance between states or operators, Phys. Rev. D **109** L081701 (2024).

[69] P. Caputa and S. Liu, Quantum complexity and topological phases of matter, Phys. Rev. B **106**, 195125 (2022).

[70] P. Caputa, J. M. Magan, and D. Patramanis, Geometry of Krylov complexity, Phys. Rev. Res. **4**, 013041 (2022).

[71] J. Erdmenger, S. Jian, and Z. Xian, Universal chaotic dynamics from Krylov space, J. High Energy Phys. 08 (2023) 176.

[72] W. Gilpin, Model scale versus domain knowledge in statistical forecasting of chaotic systems, Phys. Rev. Res. **5**, 043252 (2023).

[73] S. Nandy, B. Mukherjee, A. Bhattacharyya, and A. Banerjee, Quantum state complexity meets many-body scars, J. Phys.: Condens. Matter **36**, 155601 (2024).

[74] K. Pal, A. Gill, and T. Sarkar, Time evolution of spread complexity and statistics of work done in quantum quenches, Phys. Rev. B **108**, 104311 (2023).

[75] B. Bhattacharjee, S. Sur, and P. Nandy, Probing quantum scars and weak ergodicity breaking through quantum complexity, Phys. Rev. B **106**, 205150 (2022).

[76] M. Gautam, K. Pal, K. Pal, A. Gill, N. Jaiswal, and T. Sarkar, Spread complexity evolution in quenched interacting quantum systems, Phys. Rev. B **109** 014312 (2024).

[77] M. A. Nielsen, M. R. Dowling, M. Gu, and A. C. Doherty, Quantum computation as geometry, Science **311**, 1133 (2006).

[78] A. Bhattacharya, R. N. Das, B. Dey, and J. Erdmenger, Spread complexity for measurement-induced non-unitary dynamics and Zeno effect, J. High Energy Phys. 03 (2024) 179.

[79] V. Balasubramanian, J. M. Magan, and Q. Wu, Quantum chaos, integrability, and late times in the Krylov basis, Phys. Rev. E **111** 014218 (2025).

[80] H. W. Lin, The bulk Hilbert space of double scaled SYK, J. High Energy Phys. 11 (2022) 060.

[81] E. Rabinovici, A. Sánchez-Garrido, R. Shir, and J. Sonner, A bulk manifestation of Krylov complexity, J. High Energy Phys. 08 (2023) 213.

[82] K. Huh, H. Jeong, and J. F. Pedraza, Spread complexity in saddle-dominated scrambling, J. High Energy Phys. 05 (2024) 137.

[83] Y. Zhou, W. Xia, L. Li, and W. Li, Diagnosing quantum many-body Chaos in non-Hermitian quantum spin chain via Krylov complexity, Phys. Rev. Res. **7** 033281 (2025).

[84] B. Craps, O. Evnin, and G. Pascuzzi, Multiseed Krylov complexity, Phys. Rev. Lett. **134**, 050402 (2025).

[85] P. Suchsland, R. Moessner, and P. W. Claeys, Krylov complexity and Trotter transitions in unitary circuit dynamics, Phys. Rev. B **111**, 014309 (2025).

[86] P. Nandy, A. S. Matsoukas-Roubeas, P. Martínez-Azcona, A. Dymarsky, and A. del Campo, Quantum dynamics in Krylov space: Methods and applications, Phys. Rep. **1125**, 1 (2025).

[87] L. Domingo, F. Borondo, G. Scialchi, A. J. Roncaglia, G. G. Carlo, and D. A. Wisniacki, Quantum reservoir complexity by the Krylov evolution approach, Phys. Rev. A **110**, 022446 (2024).

[88] M. Vetrano, G. Lo Monaco, L. Innocenti, S. Lorenzo, and G. M. Palma, State estimation with quantum extreme learning machines beyond the scrambling time, npj Quantum Inf. **11**, 20 (2025).

[89] R. Martínez-Peña, J. Nokkala, G. L. Giorgi, R. Zambrini, and M. C. Soriano, Information processing capacity of spin-based quantum reservoir computing systems, Cogn. Comput. **15**, 1440 (2023).

[90] M. Schuld and F. Petruccione, *Supervised Learning with Quantum Computers*, *Quantum Science and Technology* (Springer International Publishing, Cham, 2018).

[91] B. Romeira, J. M. L. Figueiredo, and J. Javaloyes, Delay dynamics of neuromorphic optoelectronic nanoscale resonators: Perspectives and applications, Chaos **27**, 114323 (2017).

[92] V. Dunjko and H. J. Briegel, Machine learning & artificial intelligence in the quantum domain: A review of recent progress, Rep. Prog. Phys. **81**, 074001 (2018).

[93] N. Killoran, T. R. Bromley, J. M. Arrazola, M. Schuld, N. Quesada, and S. Lloyd, Continuous-variable quantum neural networks, Phys. Rev. Res. **1**, 033063 (2019).

[94] E. Schöll, J. Lehnert, A. Keane, T. Dahms, and P. Hövel, Control of desynchronization transitions in delay-coupled networks of type-I and type-II excitable systems, in *Proceedings of the International Symposium, Hanse institute of advanced studies, delmenhorst, 13–16 November, 2012*, Self-Organization in Complex Systems: The Past, Present, and Future of Synergetics, edited by A. Pelster and G. Wunner (Springer, Berlin, 2016), pp. 25–42.

[95] E. Farhi and H. Neven, Classification with quantum neural networks on near term processors, arXiv:1802.06002.

[96] D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, Parallel photonic information processing at gigabyte per second data rates using transient states, Nat. Commun. **4**, 1364 (2013).

[97] Z. Li, X. Liu, N. Xu, and J. Du, Experimental realization of a quantum support vector machine, Phys. Rev. Lett. **114**, 140504 (2015).

[98] Y. Li, G. W. Holloway, S. C. Benjamin, G. A. D. Briggs, J. Baugh, and J. A. Mol, Double quantum dot memristor, Phys. Rev. B **96**, 075446 (2017).

[99] V. Saggio, B. E. Asenbeck, A. Hamann, T. Strömberg, P. Schiansky, V. Dunjko, N. Friis, N. C. Harris, M. Hochberg, D. Englund, S. Wölk, H. J. Briegel, and P. Walther, Experimental quantum speed-up in reinforcement learning agents, Nature (London) **591**, 229 (2021).

[100] M. Schuld, R. Sweke, and J. J. Meyer, Effect of data encoding on the expressive power of variational quantum-machine-learning models, Phys. Rev. A **103**, 032430 (2021).

[101] J. Liu, M. Liu, L. P. Liu, Z. Ye, Y. Wang, Y. Alexeev, J. Eisert, and L. Jiang, Towards provably efficient quantum algorithms for large-scale machine-learning models, Nat. Commun. **15**, 434 (2024).

[102] G. B. Huang, Q. Y. Zhu, and C. K. Siew, Extreme learning machine: A new learning scheme of feedforward neural networks, in *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)* (IEEE, 2004), Vol. 2, pp. 985–990.

[103] S. Ding, H. Zhao, Y. Zhang, X. Xu, and R. Nie, Extreme learning machine: Algorithm, theory and applications, Artif. Intell. Rev. **44**, 103 (2015).

[104] T. Wang, Y. D. Yang, Y. Z. Hao, Z. N. Zhang, Y. Zhu, J. L. Xiao, and Y. Z. Huang, Nonlinear dynamics of a semiconductor microcavity laser subject to frequency comb injection, Opt. Express **30**, 45459 (2022).

[105] G. B. Huang, What are extreme learning machines? Filling the gap between Frank Rosenblatt's dream and John von Neumann's puzzle, Cogn. Comput. **7**, 263 (2015).

[106] G. Huang, Q. Zhu, and C. Siew, Extreme learning machine: Theory and applications, Neurocomputing, **70**, 489 (2006).

[107] G. Huang, H. Zhou, X. Ding, and R. Zhang, Extreme learning machine for regression and multiclass classification, IEEE Trans. Syst. Man. Cybern. B Cybern. **42**, 513 (2012).

[108] P. Mujal, J. Nokkala, R. Martínez-Peña, G. L. Giorgi, M. C. Soriano, and R. Zambrini, Analytical evidence of nonlinearity in qubits and continuous-variable quantum reservoir computing, J. Phys. Complex. **2**, 045008 (2021).

[109] L. Innocenti, S. Lorenzo, I. Palmisano, A. Ferraro, M. Paternostro, and G. M. Palma, Potential and limitations of quantum extreme learning machines, Commun. Phys. **6**, 118 (2023).

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

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

[112] W. Xiong, G. Facelli, M. Sahebi, O. Agnel, T. Chotibut, S. Thanasilp, and Z. Holmes, On fundamental aspects of quantum extreme learning machines, Quantum Mach. Intell. **7**, 20 (2025).

[113] A. Suprano, D. Zia, L. Innocenti, S. Lorenzo, V. Cimini, T. Giordani, I. Palmisano, E. Polino, N. Spagnolo, F. Sciarrino, G. M. Palma, A. Ferraro, and M. Paternostro, Experimental property reconstruction in a photonic quantum extreme learning machine, Phys. Rev. Lett. **132**, 160802 (2024).

[114] H. Qi, X. Liu, A. Gani, and C. Gong, Quantum particle Swarm optimized extreme learning machine for intrusion detection, J. Supercomput. **80**, 14622 (2024).

[115] P. Steinmüller, T. Schulz, F. Graf, and D. Herr, eXplainable AI for Quantum Machine Learning, arXiv:2211.01441.

[116] L. Pira and C. Ferrie, On the interpretability of quantum neural networks, Quantum Mach. Intell. **6**, 52 (2024).

[117] R. Lifshitz, Quantum deep dreaming: A novel approach for quantum circuit design, arXiv:2211.04343.

[118] R. Heese, T. Gerlach, S. Mücke, S. Müller, M. Jakobs, and N. Piatkowski, Explaining quantum circuits with shapley values: Towards explainable quantum machine learning, Quantum Mach. Intell. **7**, 27 (2025).

[119] E. Gil-Fuster, J. R. Naujoks, G. Montavon, T. Wiegand, W. Samek, and J. Eisert, Opportunities and limitations of explaining quantum machine learning, arXiv:2412.14753.

[120] D. Verstraeten, J. Dambre, X. Dutoit, and B. Schrauwen, Memory versus non-linearity in reservoirs, in *The 2010 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2010), pp. 1–8.

[121] J. B. Butcher, D. Verstraeten, B. Schrauwen, C. R. Day, and P. W. Haycock, Reservoir computing and extreme learning machines for non-linear time-series data analysis, Neural Netw. **38**, 76 (2013).

[122] M. Inubushi and K. Yoshimura, Reservoir computing beyond memory-nonlinearity trade-off, Sci. Rep. **7**, 10199 (2017).

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

[124] H. Jaeger, Short term memory in echo state networks, GMD Report 152, GMD—Forschungszentrum Informationstechnik GmbH, 2002, https://doi.org/10.24406/publica-fhg-291107.

[125] M. Goldmann, F. Köster, K. Lüdge, and S. Yanchuk, Deep time-delay reservoir computing: Dynamics and memory capacity, Chaos **30**, 093124 (2020).

[126] F. Köster, D. Ehlert, and K. Lüdge, Limitations of the recall capabilities in delay based reservoir computing systems, Cogn. Comput. **15**, 1419 (2023).

[127] I. Bauwens, G. Van der Sande, P. Bienstman, and G. Verschaffelt, Using photonic reservoirs as preprocessors for deep neural networks, Front. Phys. **10**, 1051941 (2022).

[128] F. Köster, S. Yanchuk, and K. Lüdge, Insight into delay based reservoir computing via eigenvalue analysis, J. Phys. Photonics **3**, 024011 (2021).

[129] F. Köster, S. Yanchuk, and K. Lüdge, Master memory function for delay-based reservoir computers with single-variable dynamics, IEEE Trans. Neural Netw. Learn. Syst. **35**, 7712 (2024).

[130] T. Hülser, F. Köster, L. C. Jaurigue, and K. Lüdge, Role of delay-times in delay-based photonic reservoir computing, Opt. Mater. Express **12**, 1214 (2022).

[131] T. Hülser, F. Köster, K. Lüdge, and L. C. Jaurigue, Deriving task specific performance from the information processing capacity of a reservoir computer, Nanophotonics **12**, 937 (2023).