

Theory and interpretability of quantum extreme learning machines: A Pauli transfer matrix approach

Markus Gross^{✉*} and Hans-Martin Rieser[✉]

Institute for AI Safety and Security, German Aerospace Center (DLR), Sankt Augustin and Ulm, Germany



(Received 8 April 2026; accepted 5 October 2026; published 26 October 2026)

Quantum reservoir computers (QRCs) have emerged as a promising approach to quantum machine learning, since they utilize the natural dynamics of quantum systems for data processing and are simple to train. Here we consider n -qubit quantum extreme learning machines (QELMs) with initial-state encoding and continuous-time reservoir dynamics. Applying the Pauli transfer matrix (PTM) formalism reveals the encoding-generated features and their linear transformation by quantum channels and thereby provides a nonlinear vector autoregression model as an interpretable classical representation of a QELM. We show that the nonlinear information processing capacity of a QELM factorizes into the ability of the encoding to represent a given target function and the decodability of the relevant Pauli features after the effective PTM. Comparing geometric and statistical decodability reveals features that are algebraically hidden yet recoverable through data correlations. We relate the decodability of Pauli features to the spreading of measurement operators under Heisenberg evolution (Krylov growth). Focusing on the unitary channel, we discuss how structured Hamiltonians, when paired with certain observables, can impair model performance, as reflected in a low rank of the readout and the effective PTMs. We trace this effect to Hamiltonian symmetries and derive asymptotic rank estimates for symmetry-resolved observable families. Our decodability analysis disentangles the contributions of encoding, reservoir dynamics, and measurement observables and thus provides interpretable diagnostics that can inform the design of QELMs and reduce reliance on exhaustive hyperparameter search. As a specific application, we consider forecasting nonlinear dynamical systems and show that the readout of a trained QELM closely matches the expression of the underlying flow map at leading orders.

DOI: [10.1103/6tm7-3tn4](https://doi.org/10.1103/6tm7-3tn4)

I. INTRODUCTION

Quantum reservoir computers (QRCs) have enjoyed significant attention since their introduction [1], featuring applications in various domains, such as image classification [2–5], time-series forecasting [1,6–10], fluid dynamics [11,12], or postprocessing of quantum experiments [13–16]. A major advantage of QRCs is their suitability for noisy intermediate-scale quantum hardware, primarily because the training process is restricted to the classical readout layer, thereby rendering the training problem convex and avoiding the barren plateau problem associated with variational quantum algorithms [17,18]. QRCs are often resilient to, and can even benefit from, environmental noise [19–22]. The nonlinear processing of input by QRCs has been analyzed in Refs. [23–26], finding that the encoding is often the main supplier of nonlinearity. Universal function approximation properties have been discussed in Refs. [6,27–31]. Notably, QRCs are closely related to quantum kernel methods [32]—a fact that can be utilized to find optimal measurement

operators [33] and harness insights from the classical domain [34–36]. For general overviews on QRCs, we refer to Refs. [37,38].

A specific QRC variant without any coherently evolving memory is the quantum extreme learning machine (QELM) [2,37–39]. Its classical counterpart, the “extreme learning machine,” is essentially a feedforward network with fixed internal architecture [40], similar to a random-feature model [41]. QELMs are capable of various machine-learning tasks, such as image classification and time-series forecasting. A typical QELM differs from a conventional quantum neural network with fixed internal parameters in two important ways: First, instead of gate-based architectures, it employs the natural dynamics of a physical quantum system (a reservoir). Second, instead of optimizing a measurement output for a certain task, one usually collects multiple unspecific measurements into a large readout vector and feeds it into a classical regression or classification algorithm. The main task in optimizing QELMs thus consists in tuning the quantum properties of the reservoir such that information processing is maximized and relevant features can be properly decoded from the measurements [14,26,42–50]. Identifying possible quantum advantages, e.g., related to the exponentially large Hilbert space into which data is encoded, is an active research topic [3,38,45,48,51,52]. However, the exponential concentration effect emerging for large qubit numbers is a phenomenon conceptually

*Contact author: markus.gross@dlr.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.

related to barren plateaus and can strongly reduce QELM performance [53].

The present study aims at an interpretable framework that can support systematic design of QELMs and reduce the need for expensive training-based hyperparameter tuning. To this end, we will discuss temporal multiplexing [1] as well as information processing capacity [54,55] in the light of the Pauli transfer matrix (PTM) formalism, based on the notion of feature *decodability*. This allows us to systematically address which parts of the model expressivity stem from the reservoir channel (such as the unitary) and which from the encoding. Moreover, our decodability framework can identify features algebraically hidden but still recovered via correlations in the data. We analyze the role of Hamiltonian symmetries for the rank of the readout matrix, which is a crucial measure of expressivity in reservoir computing. These symmetries partition the operator space under Krylov evolution, which allows one to identify suitable observable families for a given Hamiltonian. Chaotic dynamical systems are used as benchmarks due to their beneficial properties for system identification [56]. The present formalism applies to n -qubit QELMs in which the encoding is time-independent and separated from reservoir evolution. Our findings also hold implications for other QML models that fall into this category. However, models with dynamical input injection [14,57] or dissipatively driven reservoirs [21] are not covered by the present formalism.

II. MODEL

We briefly present the formalism of QELMs and the specific implementation used in this work [1,37,38]. Let $\text{Herm}(\mathcal{H})$ denote the space of $d \times d$ Hermitian matrices (real vector space of dimension d^2) on the d -dimensional Hilbert space $\mathcal{H}$, where $d = 2^n$ in terms of the number of qubits n . Given classical inputs $\mathbf{x} \in \mathbb{R}^D$, a QELM is defined by the output function

$$f(\mathbf{x}) = \sum_{k=1}^m w_k \text{tr}[M_k \rho(\mathbf{x})], \quad (2.1)$$

where $\{M_k\}_{k=1}^m$ are a set of measurement operators in $\text{Herm}(\mathcal{H})$ and $\rho(\mathbf{x}) \in \text{Herm}(\mathcal{H})$ is the density matrix of the quantum system, obtained after applying encoding and other quantum channels to an initial state, $\rho(\mathbf{x}) = \mathcal{E}(\dots \mathcal{E}_{\mathbf{x}}^{\text{enc}}(\rho_0) \dots)$. The trainable weights w_k are determined by minimizing the empirical risk:

$$\min_{\mathbf{w} \in \mathbb{R}^m} \sum_{i=1}^P \mathcal{L}(f(\mathbf{x}_i), y_i) + \frac{\lambda}{2} \|\mathbf{w}\|_2^2, \quad (2.2)$$

where $\mathcal{L}$ is a loss function, λ is a regularization parameter, and $\{(\mathbf{x}_i, y_i)\}_{i=1}^P$ are P training data points. Typical loss functions are the least-squares loss $\mathcal{L}(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$ for regression (including time-series prediction), and the hinge loss $\mathcal{L}_{\text{hinge}}(\hat{y}, y) = \max(0, 1 - y\hat{y})$, with $y \in \{-1, +1\}$ for classification. In the case of the least-squares loss, the optimal weights are given by

$$\hat{\mathbf{w}} = (GG^\top + \lambda \mathbb{I}_m)^{-1} G\mathbf{y}, \quad G_{ki} = \text{tr}(M_k \rho(\mathbf{x}_i)), \quad (2.3)$$

where $\mathbf{y} \in \mathbb{R}^P$ is the sample vector and $G \in \mathbb{R}^{m \times P}$ is the readout design matrix.

When optimizing over a complete basis of measurement operators $\{M_k\}_{k=1}^{d^2}$, the QELM takes the form of a quantum kernel model (see Appendix A for a summary). In the present work, we do not use that formalism, but instead work with a preselected observable set and optimize only over the associated weights w_k . Our notation is summarized in Table V.

A. Feature map

We consider pure states, i.e., $\rho(\mathbf{x}) = |\psi(\mathbf{x})\rangle\langle\psi(\mathbf{x})|$, with the quantum state $|\psi(\mathbf{x})\rangle \in \mathcal{H}$ given by

$$|\psi(\mathbf{x})\rangle = U_R(t)U_{\text{enc}}(\mathbf{x})|\Phi_0\rangle. \quad (2.4)$$

Here, $U_{\text{enc}}(\mathbf{x})$ is a data *encoding* unitary, U_R is a *reservoir* unitary, and we take the *initial state* as $|\Phi_0\rangle = |0\rangle^{\otimes n}$. The unitary $U_R(t) = \exp(-itH_R)$ is defined by time evolution under a reservoir Hamiltonian H_R . Specifically, we consider random Hamiltonians H_R as well as two variants of a transverse field Ising model (TFIM)

$$H_{\text{TFIM}}^{\text{zz-x}} = -J \sum_{j=1}^{n-1} Z_j Z_{j+1} - h \sum_{j=1}^n X_j, \quad (2.5a)$$

$$H_{\text{TFIM}}^{\text{xx-z}} = -J \sum_{j=1}^{n-1} X_j X_{j+1} - h \sum_{j=1}^n Z_j, \quad (2.5b)$$

with couplings $J, h \in \mathbb{R}$. Here and in the following, $\{X_j, Y_j, Z_j\}$ denote single-qubit Pauli operators acting on the j th qubit and $Z_j Z_{j+1}$ is a shorthand for the usual n -qubit tensor product (with identity operators acting on the remaining qubits). The two TFIM variants in Eq. (2.5) are related by a global Hadamard transform $U_H = H^{\otimes n}$:

$$\begin{aligned} U_H X_j U_H^\dagger &= Z_j, & U_H Y_j U_H^\dagger &= -Y_j, \\ U_H Z_j U_H^\dagger &= X_j, & U_H H_{\text{TFIM}}^{\text{zz-x}} U_H^\dagger &= H_{\text{TFIM}}^{\text{xx-z}}. \end{aligned} \quad (2.6)$$

B. Encoding schemes

We consider time-independent (parallel) product state encodings with $D = n$, producing the state

$$\begin{aligned} \rho_{\text{enc}}(\mathbf{x}) &= U_{\text{enc}}(\mathbf{x}) \rho_0 U_{\text{enc}}(\mathbf{x})^\dagger = \bigotimes_{j=1}^n |\psi_{\text{enc}}(x_j)\rangle\langle\psi_{\text{enc}}(x_j)|, \\ U_{\text{enc}}(\mathbf{x}) &= \bigotimes_{j=1}^n U_{\text{enc},j}(x_j), \end{aligned} \quad (2.7)$$

where the unitaries $U_{\text{enc},j}$ act on the j th qubit. In our numerical experiments we employ amplitude and rotational encoding schemes [58]. The former comes in two variants, where one either encodes data on half a great circle of the Bloch sphere:

$$|\psi_{\text{enc}}(x)\rangle = \sqrt{x}|0\rangle + \sqrt{1-x}|1\rangle, \quad x \in [0, 1], \quad (2.8)$$

or on the full great circle:

$$|\psi_{\text{enc}}(x)\rangle = x|0\rangle + \sqrt{1-x^2}|1\rangle, \quad x \in [-1, 1]. \quad (2.9)$$

Rotational encoding is defined by the unitary

$$U_{\text{enc},j}(x_j) = \exp(-i\pi x_j \sigma^{p_j}) = R_{p_j}(2\pi x_j), \quad x_j \in [0, 1], \quad (2.10)$$

where $\sigma^p \in \{\mathbb{I}, X, Y, Z\}$ are single-qubit Pauli matrices, $p \in \{0, x, y, z\}$, and R_{p_j} denotes rotation gates. In order to avoid irrelevant phases or trivial states, we exclusively use R_Y operators, which yield the single-qubit state

$$|\psi_{\text{enc}}(x_j)\rangle = R_Y(2\pi x_j)|0\rangle = \cos(\pi x_j)|0\rangle + \sin(\pi x_j)|1\rangle. \quad (2.11)$$

III. PAULI TRANSFER FORMULATION OF A QELM

QML models, including QELMs, have often been analyzed from the Fourier perspective, which is natural for Hamiltonian-based encodings [51–53]. We instead employ the PTM formalism [59–63], which allows a more direct identification of the relevant features for models with initial-state encoding, and has been previously utilized to study QRCs from the state-space perspective [29,37,46,49]. We focus in the following on this classical feature-based representation of a QELM to analyze its readout expressivity and feature decodability.

A. Pauli transfer matrix

In the PTM formalism, we expand states and operators in the n -qubit Pauli basis

$$\{P_k = \sigma^{a_1} \otimes \cdots \otimes \sigma^{a_n} : a_i \in \{0, x, y, z\}\}_{k=0}^{d^2-1}, \quad (3.1)$$

which consists of d^2 mutually orthogonal Pauli strings P_k , with $P_0 = \mathbb{I}$, $d = 2^n$, and normalization $\text{tr}(P_k P_l) = d \delta_{kl}$. Every state admits the expansion

$$\rho(\mathbf{x}) = \frac{1}{2^n} \sum_{k=0}^{d^2-1} \phi_k(\mathbf{x}) P_k, \quad \phi_k(\mathbf{x}) = \text{tr}(P_k \rho(\mathbf{x})), \quad (3.2)$$

giving rise to a vector $\Phi(\mathbf{x}) \in \mathbb{R}^{d^2}$ of *Pauli features*, which are the fundamental input-dependent objects in our study. They have a straightforward expression in the case of product-state encodings (see Sec. III B). Any quantum channel $\mathcal{E}$ acts linearly on this vector, $\Phi' = T_{\mathcal{E}} \Phi$, via the Pauli transfer matrix:

$$(T_{\mathcal{E}})_{kl} = \frac{1}{d} \text{tr}(P_k \mathcal{E}(P_l)). \quad (3.3)$$

Consider the QELM in Eq. (2.1) with a subset $\mathcal{S}$ of measured Pauli operators $\{P_k\}_{k \in \mathcal{S}}$, $m \equiv |\mathcal{S}| \leq d^2$. This selection can be represented by a matrix $S \in \mathbb{R}^{m \times d^2}$ (a subset of rows of the identity matrix) that extracts the corresponding rows from $T_{\mathcal{E}}$. The readout vector $\mathbf{F}(\mathbf{x}) \in \mathbb{R}^m$ contains their expectation values $F_k(\mathbf{x}) = \text{tr}[P_k \mathcal{E}(\rho_{\text{enc}}(\mathbf{x}))]$. Expanding ρ_{enc} as in Eq. (3.2) and using Eq. (3.3) gives the model output function

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{w}^\top \mathbf{F}(\mathbf{x}) = \mathbf{w}^\top T_{\mathcal{E}}^{(m)} \Phi(\mathbf{x}) \\ &= \sum_{k \in \mathcal{S}} w_k \sum_{j=0}^{d^2-1} (ST_{\mathcal{E}})_{kj} \phi_j(\mathbf{x}), \end{aligned} \quad (3.4)$$

where $\mathbf{w} \in \mathbb{R}^m$ are classical weights and $T_{\mathcal{E}}^{(m)} \equiv ST_{\mathcal{E}}$ is the $m \times d^2$ block of $T_{\mathcal{E}}$ restricted to the measured observables. This formulation explicitly shows that $f(\mathbf{x})$ is a linear function of the features $\Phi(\mathbf{x})$ generated solely by the encoding and mixed by the reservoir channel $T_{\mathcal{E}}$.

If the quantum channel $\mathcal{E}$ is completely positive and trace-preserving (CPTP), it admits a Kraus representation

$$\mathcal{E}(\rho) = \sum_{\ell=1}^{\mu} K_{\ell} \rho K_{\ell}^{\dagger} \quad (3.5)$$

with $\sum_{\ell=1}^{\mu} K_{\ell}^{\dagger} K_{\ell} = \mathbb{I}$ and $\mu \leq d^2$ (dependent on the channel, e.g., $\mu = 1$ for a unitary channel, see Appendix E for further discussion). With this representation, we can write

$$(T_{\mathcal{E}}^{(m)})_{kj} = \frac{1}{d} \sum_{\ell=1}^{\mu} \text{tr}(P_k K_{\ell} P_j K_{\ell}^{\dagger}) = \frac{1}{d} \text{tr}(P_k^{(H)} P_j), \quad (3.6)$$

where the last expression introduces the generalized Heisenberg representation of an observable O ,

$$O^{(H)} = \sum_{\ell=1}^{\mu} K_{\ell}^{\dagger} O K_{\ell}. \quad (3.7)$$

Specifically, for the unitary channel, we have $K = U(t)$ ($\mu = 1$), such that $T_{\mathcal{E}}$ becomes an orthogonal matrix $V(t) \in \text{SO}(d^2)$ with elements

$$V_{kj}(t) = \frac{1}{d} \text{tr}[P_k U(t) P_j U(t)^{\dagger}], \quad (3.8)$$

and $V_{kj}(0) = \delta_{kj}$. For general CPTP maps, $T_{\mathcal{E}}$ can also induce contraction and translation of the feature vector (see Appendix E 2).

Temporal multiplexing. In temporal multiplexing [1,64,65], instead of working at a fixed evolution time t and a potentially large set of observables P_k , one selects a small subset $\mathcal{S}$ of size m [see Eq. (3.4)] and evaluates them at several times $t_1, \dots, t_L$. One then stacks the corresponding rows of $V(t)$ into a $mL \times d^2$ observability matrix

$$R_L = \begin{pmatrix} SV(t_1) \\ SV(t_2) \\ \vdots \\ SV(t_L) \end{pmatrix}. \quad (3.9)$$

We will simply call this the *effective* PTM, denoted by R . In the case without temporal multiplexing ($L = 1$), the full unitary PTM $R = V(t)$ is recovered for $S = \mathbb{I}_{d^2}$.

B. Feature encoding

The structure of the Pauli feature vector $\Phi(\mathbf{x})$ can be made explicit for product-state encodings. Using Eq. (2.7) and an initial product state, the encoded state before the reservoir unitary is

$$\begin{aligned} \rho_{\text{enc}}(\mathbf{x}) &= U_{\text{enc}}(\mathbf{x}) |0\rangle\langle 0|^{\otimes n} U_{\text{enc}}(\mathbf{x})^{\dagger} = \bigotimes_{j=1}^n \rho_j(x_j), \\ \rho_j(x_j) &= U_{\text{enc},j}(x_j) |0\rangle\langle 0| U_{\text{enc},j}(x_j)^{\dagger}. \end{aligned} \quad (3.10)$$

It is convenient to introduce the single-qubit Pauli features

$$\begin{aligned} \Phi^{(j)}(x_j) &= (1, \phi_x^{(j)}(x_j), \phi_y^{(j)}(x_j), \phi_z^{(j)}(x_j))^{\top}, \\ \phi_a^{(j)}(x_j) &= \text{tr}[\sigma^a \rho_j(x_j)], \end{aligned} \quad (3.11)$$

with $\phi_0^{(j)} = \text{tr}(\rho_j) = 1$ (normalization), $\sigma^0 \equiv \mathbb{I}$ and $\sigma^{x,y,z} \equiv X, Y, Z$ as before. The three nontrivial components of $\Phi^{(j)}$

TABLE I. Pauli-basis projections $\langle \psi_{\text{enc}} | P | \psi_{\text{enc}} \rangle$ of an encoded product state $\rho_{\text{enc}}(\mathbf{u})$ for amplitude encoding (2.8) or (2.9) as well as rotational encoding (2.10) with R_Y gates. Here $P = \bigotimes_{k=1}^n \sigma^{a_k}$ denotes a Pauli operator ($\sigma^{a_k} \in \{\mathbb{I}, X, Y, Z\}$) and the classical datum entering qubit k is u_k . For convenience, we have included the projection operator $\frac{1}{2}(\mathbb{I} + Z_k)$. One observes that each Pauli factor σ^{a_k} (first column) in an observable P produces the corresponding single-qubit feature $\phi_{a_k}^{(k)}(u_k) := \text{tr}(\sigma^{a_k} \rho_k(u_k))$ (second to fourth columns). For instance, the X_k row gives $\phi_x^{(k)}(u_k)$ and the Z_k row gives $\phi_z^{(k)}(u_k)$.

Factor in observable P	Feature $\phi_{a_k}^{(k)}(u_k)$		
	Amplitude on $[0, 1]$	Amplitude on $[-1, 1]$	Rotational
$\dots \otimes X_k \otimes \dots$	$2\sqrt{(1 - u_k)u_k}$	$2u_k\sqrt{1 - u_k^2}$	$\sin(2\pi u_k)$
$\dots \otimes Y_k \otimes \dots$	0	0	0
$\dots \otimes Z_k \otimes \dots$	$2u_k - 1$	$2u_k^2 - 1$	$\cos(2\pi u_k)$
$\dots \otimes \frac{1}{2}(\mathbb{I} + Z_k) \otimes \dots$	u_k	u_k^2	$\cos^2(\pi u_k)$

define the Bloch vector, which has unit magnitude for pure states (see Appendix E for further discussion). Since both the state and the Pauli basis factorize across qubits, the n -qubit feature vector is

$$\Phi(\mathbf{x}) = \bigotimes_{j=1}^n \Phi^{(j)}(x_j), \quad (3.12)$$

and its components are given by

$$\phi_{a_1 \dots a_n}(\mathbf{x}) = \text{tr} \left[\left(\bigotimes_{j=1}^n \sigma^{a_j} \right) \rho_{\text{enc}}(\mathbf{x}) \right] = \prod_{j=1}^n \phi_{a_j}^{(j)}(x_j), \quad (3.13)$$

$a_j \in \{0, x, y, z\}.$

The subscript $a_1 \dots a_n$ corresponds to the explicit site representation of the index k in Eq. (3.2). The ϕ_k thus run over *all possible combinations* of the single-qubit features in Eq. (3.11): For each qubit one chooses one of the four factors $\{1, \phi_x^{(j)}, \phi_y^{(j)}, \phi_z^{(j)}\}$, yielding a total of at most $d^2 = 4^n$ features. The readout features resulting from some simple encoding schemes are summarized in Table I. Effects of feature mixing by a unitary channel are discussed in the next section.

C. Classical representation of a QELM

Let $q \leq d^2$ be the effective number of features (excluding features identically vanishing due to the encoding), and B be the total number of measurements, e.g., $B = mL$. According to Eq. (3.4), a QELM has an equivalent classical representation as a PTM-weighted nonlinear function library,

$$f(\mathbf{u}) = \mathbf{w}^\top \mathbf{F}(\mathbf{u}), \quad \mathbf{F}(\mathbf{u}) \equiv R\Phi(\mathbf{u}), \quad (3.14)$$

where $R \in \mathbb{R}^{B \times q}$ is the effective PTM (3.9) (including any row-selection and temporal multiplexing) and $F_k = \sum_{j=1}^q R_{kj} \phi_j$ are the Pauli features available at the readout.

If R has full column rank, then it is left-invertible [66] and thus the regressor can always find weights $\mathbf{w} \propto (R^\top)^\top \mathbf{e}_r$ that reconstruct any desired element ϕ_r . Since one usually trains with arbitrary targets $\mathbf{y}$ instead of ϕ_r , the actual weights [Eq. (2.3)] contain projections from the Pauli feature space to the data space:

$$\hat{\mathbf{w}} = (GG^\top + \lambda \mathbb{I}_B)^{-1} G\mathbf{y} = R(CR^\top R + \lambda \mathbb{I}_q)^{-1} \Phi\mathbf{y}, \quad (3.15)$$

where we introduced the feature design and Gram matrices

$$\Phi \equiv [\Phi(\mathbf{u}_1) \dots \Phi(\mathbf{u}_P)] \in \mathbb{R}^{q \times P}, \quad C = \Phi\Phi^\top \in \mathbb{R}^{q \times q}, \quad (3.16)$$

and the corresponding readout design and Gram matrices

$$G \equiv [\mathbf{F}(\mathbf{u}_1) \dots \mathbf{F}(\mathbf{u}_P)] = R\Phi \in \mathbb{R}^{B \times P}, \quad (3.17)$$

$$GG^\top = RCR^\top \in \mathbb{R}^{B \times B}.$$

The latter are typically computed in a QELM implementation, whereas R and Φ are theoretical objects. We will analyze in the following subsections the requirements on the encoding and reservoir unitary to improve trainability and avoid expensive hyperparameter search.

The last equality in (3.15) (which follows from the “push-through” identity) implies that for the full unitary PTM $R = V$ [Eq. (3.8)], the weights are simply rotated versions of the ones without PTM mixing:

$$\hat{\mathbf{w}} = R(C + \lambda \mathbb{I}_q)^{-1} \Phi\mathbf{y}. \quad (3.18)$$

Moreover, since V is orthogonal, it drops out from the predictor (3.14), consistent with the absence of a reservoir unitary in the kernel-representation (A1). Such a QELM is then simply a regressor acting on the full d^2 -dimensional feature library Φ . In the context of time-series processing (especially when using some inputs u_j to encode delay states) this reduces to a “next-generation” reservoir computer (NG-RC) [67] or, equivalently, a nonlinear vector autoregressive process. When R merely has full column rank, it can be interpreted to induce a regressor operating on the mixed feature set $\Phi' = R\Phi$.

A coarse measure for expressivity is the rank of the readout Gram matrix in Eq. (3.15), which can be bounded according to

$$\text{rank}(G) \leq \min(\text{rank}(R), \text{rank}(\Phi)) \leq \min(B, P, q). \quad (3.19)$$

More precisely, $\text{rank}(G) = \text{rank}(R) - \dim(\text{row}(R) \cap \text{null}(\Phi^\top))$. Hence $\text{rank}(G) = \text{rank}(R)$ whenever the sampled feature matrix is nondegenerate on the row space of R . This condition is satisfied, for example, if Φ has full row rank and is expected generically once $\text{rank}(\Phi) \geq \text{rank}(R)$ and the data are sufficiently diverse. Accidental data degeneracies or numerical ill conditioning, since the features ϕ_k are in general not orthogonal, can reduce the actual rank. Accordingly, in

the data-rich generic regime,

$$\text{rank}(G) \simeq \text{rank}(R). \quad (3.20)$$

For random Hamiltonians, $\text{rank}(R)$ typically grows proportionally to the measurement budget B and then saturates near q (see Fig. 4 below). By contrast, structured Hamiltonians like the TFIM are not fully scrambling, resulting in $\text{rank}(R) \ll q$. The achieved ranks indicate the quality of the measurement strategy (see Sec. IV B) and thus of the resulting training accuracy (see also Fig. 13 below).

IV. PTM ANALYSIS FOR A UNITARY QUANTUM CHANNEL

For separable (product) quantum channels, the PTM factorizes as $T = \bigotimes_{j=1}^n T_j$ into single-qubit PTMs T_j . The action of a local unitary is discussed in Appendix E 1. To understand and optimize a QELM with general nonlocal quantum channels, we must analyze the full PTM [Eqs. (3.4) and (3.6)]. We focus in the following on the unitary case. The optimization can be split into several subtasks that define the fundamental reconstruction problem in (quantum) reservoir computing with explicit feature maps [14,26,47,49]:

(1) Decide which types of features $\{\phi_r\}$ are best suited to learn the dataset. For example, non-Markovian time series (including scalar harmonic signals) require delay coordinates [68,69], whereas Markovian dynamical systems can already be modeled by instantaneous monomial features [70]. While commonly used encodings support universal function approximation [6,27–31], they can differ in their approximation errors, generalization behavior [51,71,72], or robustness during autoregressive rollout. As shown in Ref. [70], the training error shows strongest reduction with time resolution when the feature map matches the low-order Lie coefficients of the flow map. This implies that some feature maps are more suitable for learning time series originating from physical models.

(2) Select suitable measurement observables $\{P_k\}$. Ideally, the Heisenberg-evolved observables $P_k^{(H)}(t)$ have strong overlap with the observable P_r generating the desired feature ϕ_r [see Eq. (3.6)]. One possibility to achieve this is to determine an optimal measurement operator M^* via primal or dual optimization, which can then be further expanded into a set of rotated Pauli-Z operators [33]. Another possibility is temporal multiplexing [1,64,65,73,74], discussed further in the PTM framework below (see Sec. IV A).

(3) When the available measurement operations are restricted, one must tune the quantum channel $\mathcal{E}$ and the readout layer such that the desired features ϕ_r become *decodable*. Strict (geometric) decodability means that, for each such ϕ_r , there exists a weight vector $\mathbf{w}_r$ such that $\mathbf{w}_r^\top T_{\mathcal{E}} \Phi \propto \phi_r$ (see Sec. IV B). In practice, decodability in the statistical sense (as a correlation) and within a certain error tolerance is often sufficient, since the features needed for typical training tasks can usually be approximated on limited domains from the available Pauli features (see Sec. IV C). We will show that temporal multiplexing directly enhances decodability by increasing the operator support.

A. Krylov growth and Pauli weight spreading

Temporal multiplexing enables the regressor to “reverse” the information scrambling caused by quantum evolution. The latter is captured by the Krylov growth of an operator under conjugation [47,74,75], which we summarize here in our setting. Specifically, the unitary Heisenberg-evolution [Eq. (3.7)] of a given Pauli observable P_k satisfies

$$\begin{aligned} P_k^{(H)}(t) &\equiv U(t)^\dagger P_k U(t) = e^{iHt} P_k e^{-iHt} = e^{t\mathcal{L}}(P_k) \\ &= \sum_{\ell=0}^{\infty} \frac{t^\ell}{\ell!} \mathcal{L}^\ell(P_k), \end{aligned} \quad (4.1)$$

where $\mathcal{L}^\ell$ denotes ℓ repeated applications of the Liouvillian $\mathcal{L}(O) \equiv i[H, O]$. The time-evolved observable therefore generates the (increasing) Krylov subspace

$$\mathcal{K}_\ell(P_k) = \text{span}\{P_k, \mathcal{L}(P_k), \dots, \mathcal{L}^{\ell-1}(P_k)\}. \quad (4.2)$$

While there is no strict one-to-one identification of the integer Krylov order ℓ with the continuous time t , ℓ can be understood as the depth of a truncation of the nested-commutator series. Due to the suppression by the factorial, a rough estimate is $\ell \sim t \|H\|$ in case of local Hamiltonians. It is useful to note that for the TFIM models in Eq. (2.5), the Hadamard equivalence (2.6) implies that

$$(Z_{j_1} \cdots Z_{j_k})^{(H)}(t)|_{zz-x} = U_H^\dagger (X_{j_1} \cdots X_{j_k})^{(H)}(t)|_{xx-z} U_H, \quad (4.3)$$

i.e., Pauli-Z observables evolve under H_{TFIM}^{zz-x} equivalently to Pauli-X observables under H_{TFIM}^{xx-z} up to a global unitary.

In the PTM approach, Krylov growth can be represented in terms of *operator spreading* in the Pauli basis [76]. Accordingly, expanding the time-evolved observable $P_k^{(H)}(t)$ gives

$$P_k^{(H)}(t) = \sum_{j=0}^{d^2-1} V_{kj}(t) P_j, \quad (4.4)$$

where we used cyclicity of the trace to introduce $V(t)$ from Eq. (3.8). Thus, for non-Clifford $U(t)$, an initial weight fully localized on P_k can spread to other (higher and lower order) Pauli strings as t increases. A convenient diagnostic is the Pauli (Hamming) weight $v(P_j) \in \{0, \dots, n\}$, i.e., the number of nonidentity factors in P_j , and its coefficient-weighted average

$$\bar{v}_k(t) = \sum_{j=0}^{d^2-1} v(P_j) V_{kj}(t)^2. \quad (4.5)$$

For local (including nonintegrable/chaotic ones) Hamiltonians, the Lieb-Robinson bound [50,77–80] heuristically implies that an initially local operator can only develop support within distance $\lesssim v_{\text{LR}} t$ until finite-size saturation:

$$\bar{v}_k(t) \sim \begin{cases} 1 + \text{const} \cdot v_{\text{LR}} t & (\text{intermediate times}) \\ \mathcal{O}(n) & (\text{late times}) \end{cases} \quad (4.6)$$

with a typical late-time value $\bar{v}_k \approx \frac{3}{4}n$, corresponding to a uniform distribution of coefficients V_{kj}^2 over Pauli strings. The late-time limit is to be understood here and in the following

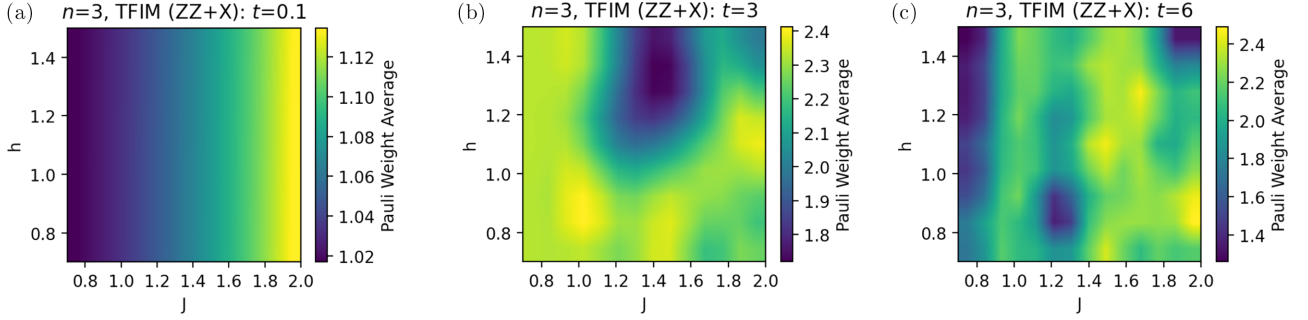


FIG. 1. Time evolution of the Pauli weight average $\bar{v}_k$ [Eq. (4.5)], averaged over all single-site Paulis $P_k \in \{X_j, Y_j, Z_j\}_{j=1}^n$, in the h - J coupling parameter space of the TFIM (2.5a) for $n = 3$ qubits.

in the sense of typical coverage, since finite-dimensional unitary evolution is generally quasiperiodic. From the definition (3.7) and the expansion (4.4), it straightforwardly follows that $\|P_k^{(H)}\|^2 = d = d \sum_{j=0}^{d^2-1} V_{kj}^2$, and thus $\sum_{j=0}^{d^2-1} V_{kj}^2 = 1$ for all t . Since, for a Haar-random unitary, a nonidentity Pauli operator is distributed approximately uniformly over the traceless Pauli subspace of dimension $d^2 - 1$, one expects a typical magnitude

$$\mathbb{E}_U[V_{kj}^2] = \frac{1}{d^2 - 1} \sim 4^{-n} \quad (j, k \neq 0). \quad (4.7)$$

This is the origin of the exponential concentration effect, which makes accessing information about the input infeasible for large qubit numbers $n \rightarrow \infty$ [53,81]. Integrable models (such as the integrable TFIM used here) typically show more structured behavior due to stable quasiparticle modes. An extreme case is a Clifford unitary, for which V_{kj} is a signed permutation matrix (no spreading). By contrast, a Haar-random unitary on the full Hilbert space can generate rapid delocalization [82–86]. The evolution of the Pauli weight average is illustrated in Fig. 1. As shown in Appendix F 1, the interplay of Hamiltonian structure and observables can have significant impact on the Heisenberg evolution and thus on the resulting Krylov spans.

B. Pauli feature decoding

Following Eq. (3.14), we denote the set of Pauli features available at the readout by $F_k = \sum_j R_{kj} \phi_j$ ($k = 1, \dots, B$), which emerge by linear transformations of the input features ϕ_j ($j = 1, \dots, q \leq d^2$) via the effective PTM $R \in \mathbb{R}^{B \times q}$ [Eq. (3.9)], possibly including temporal multiplexing. We now ask for conditions that enable the regressor to reconstruct a given input feature $\phi_r(\mathbf{x})$ by finding weights $\mathbf{w}$ such that

$$\mathbf{w} \cdot \mathbf{F} \simeq \phi_r \Leftrightarrow \mathbf{w}^\top R \simeq \mathbf{e}_r^\top, \quad (4.8)$$

where $\mathbf{e}_r$ is the r th basis vector of $\mathbb{R}^q$. The least-squares solution to this problem is discussed in Appendix B, showing that one can distinguish between *geometric* and *statistical* decodability: the former indicates whether an exact algebraic solution to (4.8) can be found; in the latter, a solution can additionally make use of correlations in the data. We focus in this section on geometric decodability, as it is determined by PTM structure alone, independently of the data. Statistical decodability is then addressed in Secs. IV C and IV D.

If R corresponds to the full unitary channel ($B = q = d^2$), then $R = V(t) \in \mathbb{R}^{q \times q}$ is orthogonal and any feature can be geometrically decoded by

$$\mathbf{w} = V(t) \mathbf{e}_r. \quad (\text{unitary channel}) \quad (4.9)$$

If we select subsets of observables, then Eq. (4.8) can in general *not* be satisfied for arbitrary r since features get mixed under a quantum channel. We thus introduce the geometric decodability score of a Pauli feature (see Appendix B)

$$\gamma_r^2 \equiv (R^+ R)_{rr}, \quad \sum_{r=1}^q \gamma_r^2 = \text{rank}(R) \leq q, \quad (4.10)$$

and its associated decoding weights

$$\mathbf{w} = ((R^+)_{r,:})^\top, \quad (4.11)$$

where R^+ is the Moore-Penrose pseudoinverse. Exact decodability of the r th feature is indicated by $\gamma_r^2 = 1$. A value $\gamma_r^2 < 1$ implies that the feature can only be reconstructed up to an error $\sim |\phi_r - \mathbf{w}_r^\top R \Phi| \leq \sqrt{1 - \gamma_r^2} \|\Phi\|_2$. We take $\gamma_r^2 \gtrsim 0.5$ as a heuristic decodability threshold. The score γ_r^2 is a purely geometric, data-independent notion. Statistical decodability scores formulated in Sec. IV C and Appendix C can exceed it (via correlations), but can also fall below it (for strongly anisotropic feature distributions).

Both decodability score and weights preserve tensor product structure of the general PTM R : If $R = \bigotimes_{i=1}^n R_i$, then $\gamma_r^2 = \prod_{i=1}^n \gamma_{r,i}^2$ and $\mathbf{w} = \bigotimes_{i=1}^n \mathbf{w}_i$. For a single evolution time and before feature clipping ($q = d^2$), $R = SV(t)$ has orthonormal rows and one obtains $\gamma_r^2 = \sum_{k \in \mathcal{S}} V(t)_{kr}^2$, showing that feature r is decodable only to the extent that the “spread” of that Pauli under $V(t)$ lands in the measured subset $\mathcal{S}$ [87]. Specifically, for Haar-random unitaries, typical decodability decays exponentially with qubit number n ,

$$\mathbb{E}_U[\gamma_r^2] \sim |\mathcal{S}| 4^{-n}, \quad (4.12)$$

which follows from Eq. (4.7). Noisy channels can have invertible PTMs [see Appendix E 2] and then do not strictly reduce the accessible feature subspace. If they are contracting, they effectively suppress high-weight Pauli features [see Eq. (E12)], but require large weight norms $\|\mathbf{w}\|$ which degrades conditioning. Such maps are unstable under finite sampling and thus effectively lead to information loss.

Operator spreading under temporal multiplexing is visualized in Fig. 2, where the effective PTM R is plotted for the case of a TFIM and a random unitary. The related projector

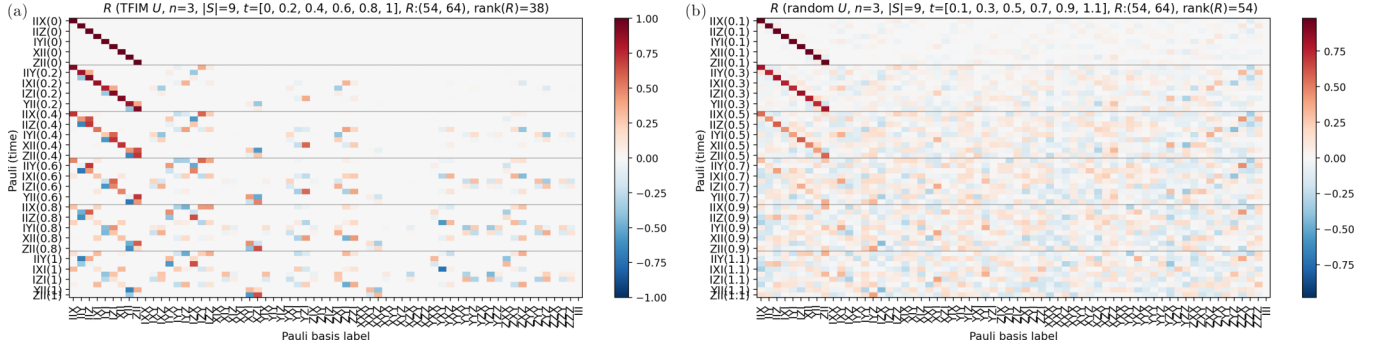


FIG. 2. Visualization of the operator spreading (4.4) via the effective PTM R [Eq. (3.9)] constructed in the temporal multiplexing scheme for the TFIM (a) and random unitary (b). The selected observable set S are all weight-1 Pauli operators of an $n = 3$ qubit system. With increasing evolution time (from top to bottom), these initial operators develop nonzero projection onto the full d^2 -dimensional Pauli basis [Eq. (4.4)]. In the TFIM case, the supporting basis exhibits a sparse pattern, whereas it is dense for a random unitary.

matrices R^+R are shown in Figs. 3(a) and 3(b). When including $t = 0$ in the multiplexing scheme, the first $|S| = 9$ rows of R have 1 on the diagonal (no spreading), leading to perfect decodability of the weight-1 Pauli sector. For the Hamiltonian and parameters in Fig. 3(c), no perfectly decodable feature is observed at the chosen nonzero time.

The evolution with multiplexing steps L of the mean decodability $\bar{\gamma}_r^2$ of each Pauli sector is further illustrated in Fig. 4. Since we now start with $t = 1$, the selected low-weight Paulis are already spread and initial decodability is low on average [88]. Structured Hamiltonians, like the TFIM [Eq. (2.5)], can have strong impact on the achievable decodability score depending on the variant or the specific observables used. Comparing Figs. 4(c) and 4(g), Pauli- Z 's evolving under the Hamiltonian $H_{\text{TFIM}}^{\text{xx-z}}$ render only weak scores compared to the Hamiltonian $H_{\text{TFIM}}^{\text{zz-x}}$ [89].

These differences are a consequence of the underlying symmetries of the Hamiltonian: Specifically, for the TFIM, the Majorana charge of an observable is conserved under the evolution (see Appendix F). Figure 5 shows the symmetry-resolved sector structure of the PTM, obtained by ordering the observables according to the charge convention that makes the corresponding H quadratic in the Majorana operators. As seen in Fig. 5(a), time evolution of Pauli- Z observables under $H_{\text{TFIM}}^{\text{xx-z}}$ leaves encoding Pauli sectors with odd

Majorana degree unpopulated, explaining the low decodability scores in Fig. 4(g). By contrast, the Hamiltonian $H_{\text{TFIM}}^{\text{zz-x}}$ touches all nontrivial encoding sectors except the top one with charge $2n$ [Fig. 5(b)]. The spreading pattern still remains highly structured and oscillatory, leading to slightly undersaturated decodability scores [Fig. 4(c)] compared to a random Hamiltonian [Fig. 4(d)], which does not respect Majorana symmetries.

Furthermore, the analysis based on the Jordan-Wigner (Majorana) formalism (see Appendix F) shows that under $H_{\text{TFIM}}^{\text{xx-z}}$, the Pauli strings generated from an initial Z are mainly consisting of Z strings with X/Y end points, of which there are $\sim \mathcal{O}(n^2)$. By contrast, under $H_{\text{TFIM}}^{\text{zz-x}}$, the evolved Z_j spread over exponentially large odd-degree Majorana sectors, producing a Krylov rank $\sim \mathcal{O}(3^n)$ (up to polynomial prefactors). The implications for the asymptotic rank $r_\infty := \lim_{L \rightarrow \infty} \text{rank}(R_L)$ are summarized in Table II and numerically confirmed in Fig. 6(b). Additionally, Fig. 6(a) illustrates the evolution of $\text{rank}(R_L)$ under temporal multiplexing. These findings can more generally be understood as a consequence of Hamiltonian symmetries and eigenspace degeneracies, as shown in detail in Appendix G.

Random Hamiltonians typically populate the full Pauli basis. Under temporal multiplexing they thus consistently lead to a higher rank of R and a more even decodability over Pauli

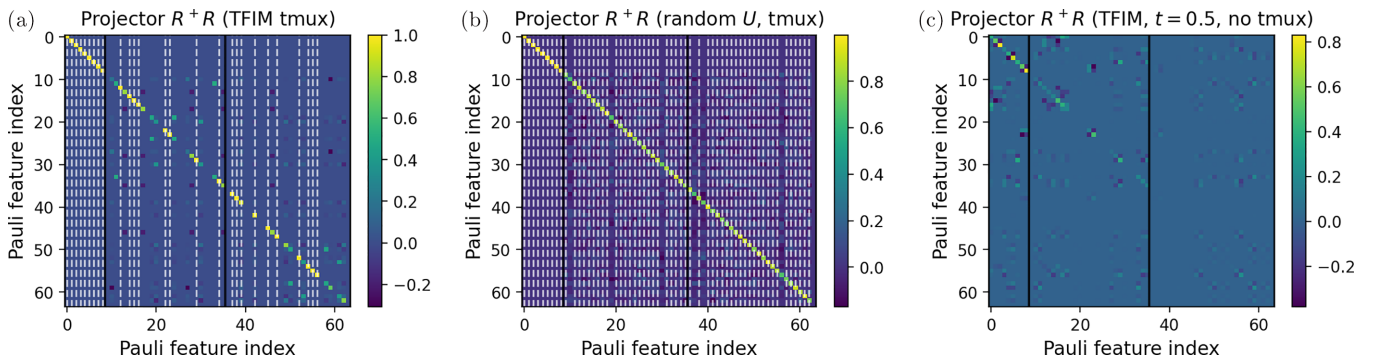


FIG. 3. Projector R^+R of the effective PTM [Eq. (3.9)] for different unitary types and a complete weight-1 Pauli observable set. Panels (a) and (b) directly correspond to the R shown in Fig. 2, which are constructed by temporal multiplexing. Panel (c) shows the projector based on $R = SV(t = 0.5)$ (no temporal multiplexing). The dashed white lines indicate features for which $\gamma_j^2 > 0.5$, while the black lines indicate Pauli weight sector boundaries. Due to operator spreading over a time $t = 0.5$, decodability in (c) has significantly decreased.

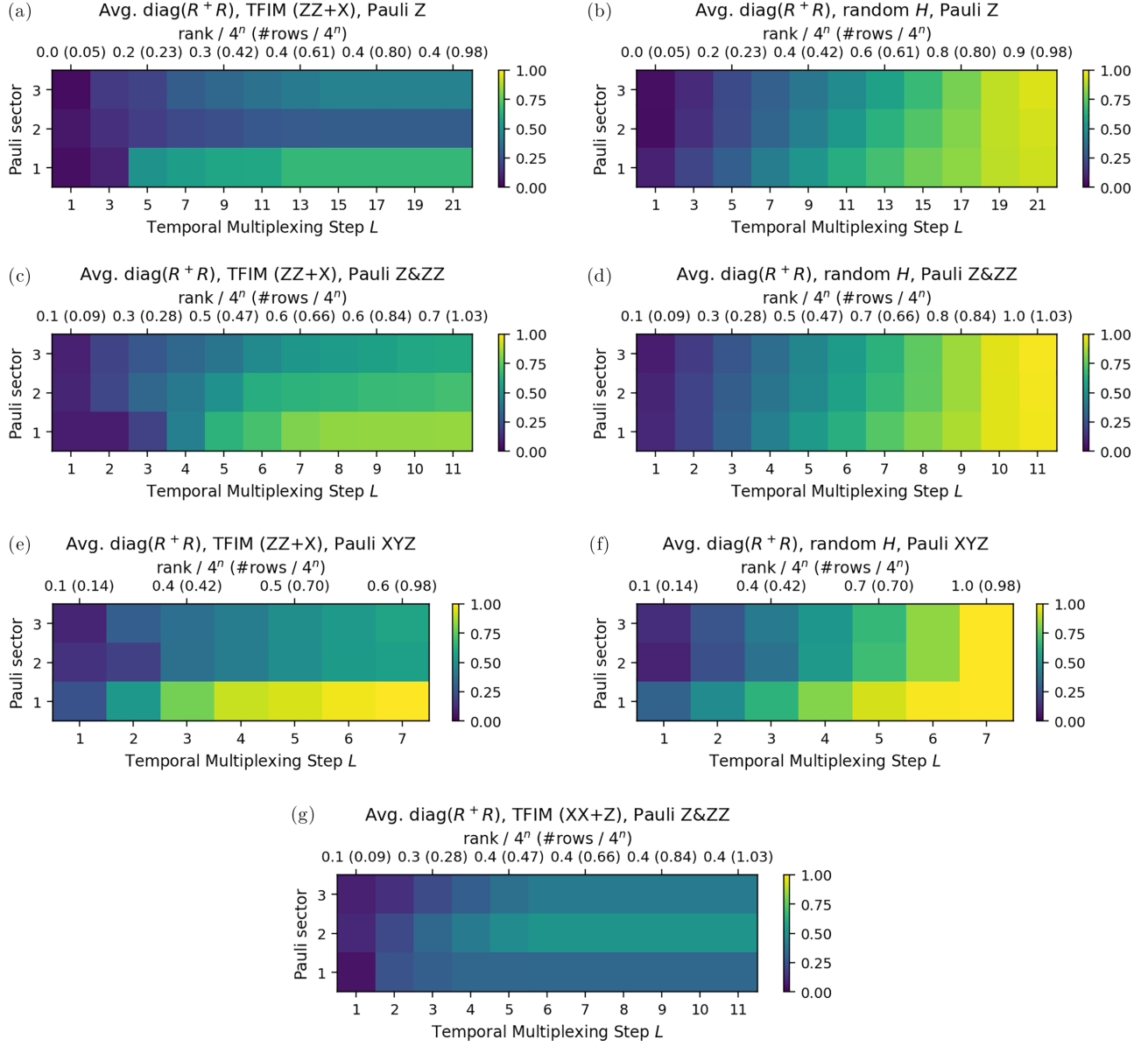


FIG. 4. Feature decodability score $\bar{\gamma}_j^2$ obtained from the effective PTM R [Eq. (4.10)], averaged over each Pauli weight sector, of an $n = 3$ qubit model as a function of temporal multiplexing iteration length L [Eq. (3.9)]. The unitary evolution time is given by $t_L = L$ (i.e., initial time $t_1 = 1$). The corresponding rank and number of rows of R , representing the measurement budget ($B = mL$), are shown on the top axis (as a fraction of the total operator space dimension $d^2 = 64$). We consider the TFIM (2.5a) in (a), (c), and (e), the TFIM (2.5b) in (g), and random Hamiltonians in (b), (d), and (f). The observable set $\mathcal{S}$ is either only Pauli-Z [(a) and (b)], Pauli-Z and Pauli-ZZ [(c), (d), and (g)], or all weight-1 Paulis [(e) and (f)].

sectors for large L . We find that temporal multiplexing with Pauli Z and ZZ observables and random unitaries requires $B \sim d^2$ measurements to achieve high decoding scores in all Pauli sectors, along with a saturated rank $\text{rank}(R) \sim |\mathcal{S}|L \sim d^2$. In practice, however, this exponential measurement budget is rarely needed, since relevant features can be statistically decoded with far fewer measurements (see Figs. 8 and 9 below).

Summary. The decodability framework presented here extends the Krylov observability theory in Refs. [47,74] by resolving geometry on the level of individual Pauli features. We remark that the Krylov grade M discussed in Ref. [47] corresponds to the asymptotic rank of a single observable, $r_\infty(\mathcal{S} = \{O\}) = M$ [Eqs. (F11) and (G23)], while

the effective dimension score O_K can be regarded as a scalar proxy for our multiobservable rank (R_L). The geometric decodability score γ_r^2 of a feature r [Eq. (4.10)] is a data-independent property of the chosen measurement observables and quantum operations after the encoding. For moderate qubit numbers n , full Pauli feature decodability is guaranteed when one either performs a full measurement ($m = d^2$) on an invertible quantum channel, or one employs a temporal multiplexing scheme leading to a full column rank of R , again requiring an exponential measurement budget $mL \sim d^2$. Practically sufficient decodability can already be achieved with modest scores and polynomial-in- n measurement budgets, as typically done in applications [1,65]. Structured Hamiltonians

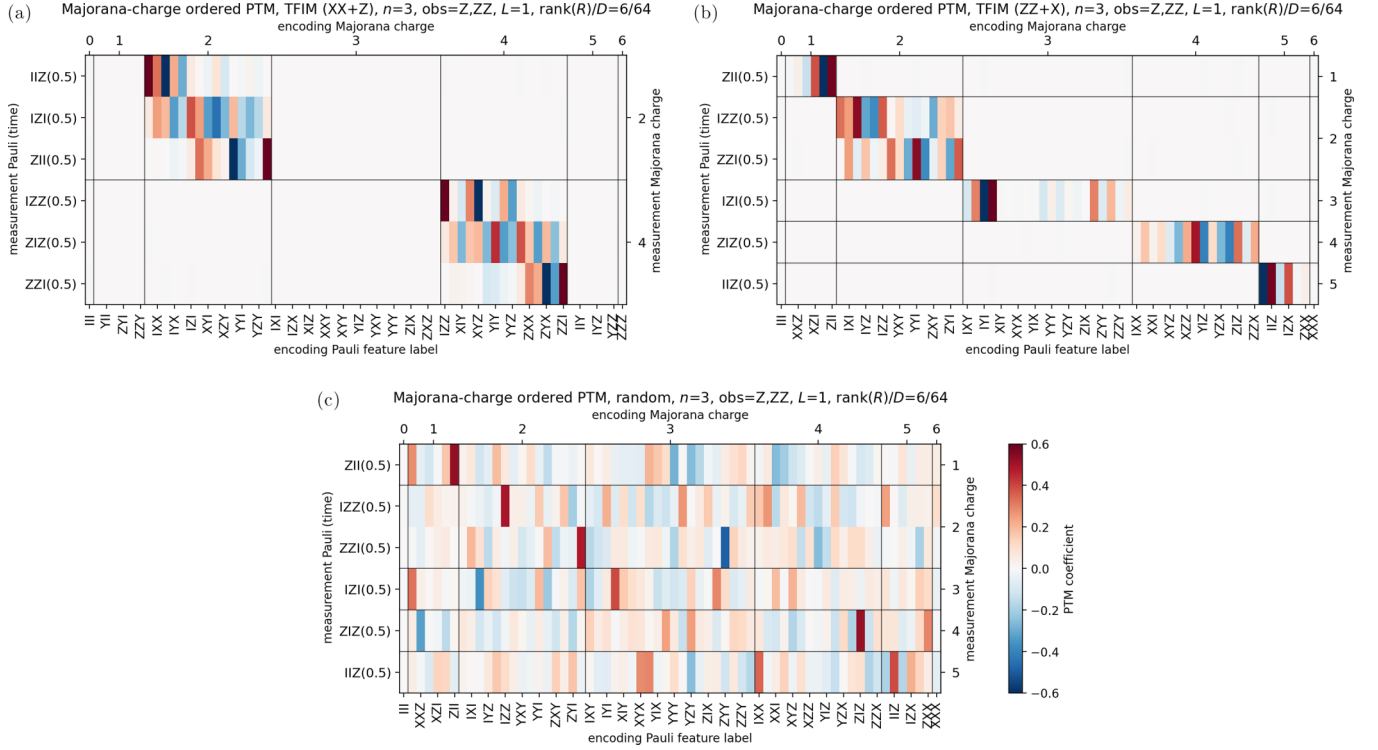


FIG. 5. Majorana-charge ordered effective PTM R for $n = 3$ qubits, the observable set $\mathcal{S} = \{Z_j, Z_i Z_j\}$, and a single evolution time $t = 0.5$. Columns and rows denote encoding Pauli and time-evolved measured Pauli observables, respectively, grouped by their Majorana charge. Panels show time evolution under various Hamiltonians: (a) H_{TFIM}^{xx-z} [Eq. (2.5b)], (b) H_{TFIM}^{zz-x} [Eq. (2.5a)], and (c) a random H . In [(a) and (b)] we use the Majorana charge convention that makes the corresponding H quadratic, revealing a sparse block structure where Z -type observables remain confined to their respective Majorana-charge sectors. By contrast, a random H (c) does not conserve Majorana charge and thus populates the full Pauli basis [for visualization, we use here the same Majorana charge convention as in (b)].

can be advantageous in the large- n limit to avoid exponential concentration issues [53,81]. However, they can negatively impact expressivity of the model under temporal multiplexing of certain observables, as reflected in polynomial or reduced exponential growth of the PTM rank with qubit number n (see Table II and Fig. 6).

C. Statistical decodability of input data

We focused above on the data-agnostic decodability of Pauli features ϕ_j , which are generated by the encoding

TABLE II. Asymptotic saturated rank $r_\infty := \lim_{L \rightarrow \infty} \text{rank}(R_L)$ (see Appendix F2) of the temporally multiplexed PTM (3.9) for the two TFIM variants [Eq. (2.5)] and different observable families $\mathcal{S}$. For temporal multiplexing of pure Pauli Z -type observables, the Hamiltonian H_{TFIM}^{xx-z} leads to exponentially lower expressivity, reflected by polynomial-in- n ranks $\ll 3^n$.

Hamiltonian	Observables	Asymptotic saturated rank r_∞
H_{TFIM}^{xx-z}	Z_j	$\mathcal{O}(n^2)$
H_{TFIM}^{xx-z}	$Z_j, Z_i Z_j$	$\mathcal{O}(n^4)$
H_{TFIM}^{xx-z}	X_j, Y_j, Z_j	$\mathcal{O}(\text{poly}(n)3^n)$
H_{TFIM}^{zz-x}	Z_j	$\mathcal{O}(\text{poly}(n)3^n)$
H_{TFIM}^{zz-x}	$Z_j, Z_i Z_j$	$\mathcal{O}(\text{poly}(n)3^n)$
H_{TFIM}^{zz-x}	X_j, Y_j, Z_j	$\mathcal{O}(\text{poly}(n)3^n)$

and transform linearly under the action of the reservoir. However, they are themselves nonlinear functions of the classical input $\mathbf{u} \in \mathbb{R}^D$. For example, with amplitude encoding (2.8), a single Pauli- Z generates an affine readout $\langle Z \rangle \propto 2u - 1$, a single Pauli- X can introduce significant nonlinearities since $\langle X \rangle = \sqrt{1 - (2u - 1)^2} = 1 - \frac{1}{2}(2u - 1)^2 - \frac{1}{8}(2u - 1)^4 + \dots$ (cf. Table I). It is therefore important to understand the actual nonlinearities in u present at the readout. This is a classical problem in reservoir computing, studied as memory capacity [90] or, more generally, information processing capacity [54,55,64,91]. The latter notion is often called *nonlinear processing capacity* for models like a QELM, where internal memory is absent.

We show in Sec. IV D that the classical nonlinear processing capacity can be split into a feature-representation part and a decodability part. The former describes the nonlinear capacity of the encoding (which is discussed in this section), while the latter accounts for transmission of the features by the PTM (discussed in Sec. IV B).

1. Nonlinear degree at readout

In order to estimate the degree of nonlinearity in u available at the readout, one typically calculates the empirical correlation between the readout features $\mathbf{F}(\mathbf{u})$ [see Eq. (3.4)] and certain nonlinear target functions $y(\mathbf{u})$. Raw monomials like $\{u^j\}$ are strongly correlated under typical distributions of u and thus do not allow for clear identification of the

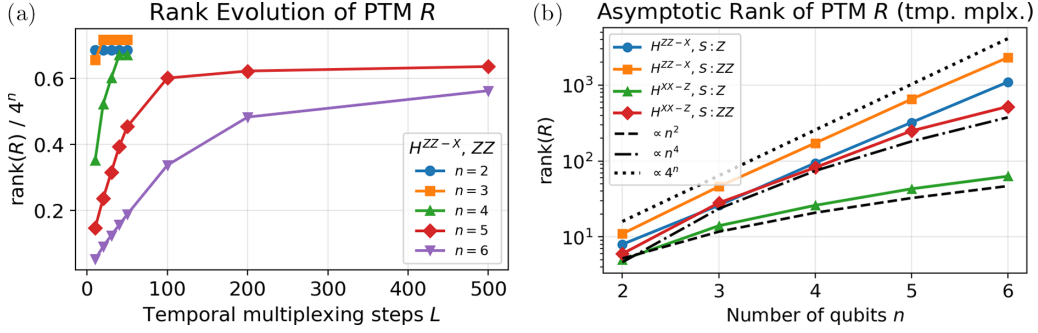


FIG. 6. Rank of the effective PTM R [Eq. (3.9)] obtained by temporal multiplexing an observable set $\mathcal{S}$ (one and two-site Pauli Z) under unitary evolution with a TFIM Hamiltonian [Eq. (2.5)]. (a) Evolution of $\text{rank}(R)$ with temporal multiplexing steps L for the TFIM (2.5a) and $\mathcal{S} = \{Z_i, Z_i Z_j\}$ observables. (b) Asymptotic PTM rank, $\lim_{L \rightarrow \infty} \text{rank}(R)$, obtained numerically (solid curves), compared to the theoretical scaling predictions from Table II (broken lines). The dotted curve follows the operator space dimension 4^n , which upper bounds the exact theoretical result $\text{poly}(n)3^n$.

nonlinear degree within this statistical approach. Following Refs. [54,64], we therefore use orthogonal polynomials $p_j(u_\alpha)$ (degree j) to construct orthogonal degree- k targets as $y_e(\mathbf{u}) = \prod_{\alpha=1}^D p_{e_\alpha}(u_\alpha)$, with $\sum_{\alpha=1}^D e_\alpha = k$. Specifically, for amplitude encoding, we use (shifted) Legendre polynomials, which are orthonormal with respect to the uniform distribution on $[-1, 1]$ or $[0, 1]$; for angle encoding, we use Hermite polynomials, which are orthonormal with respect to the Gaussian distribution $\mathcal{N}(0, 1)$. Here the inputs are encoded using $R_Y(u_\alpha)$, rather than the convention of Eq. (2.10). We generally assume independent inputs u_α .

We take the coefficient of determination [54,55]

$$\mathcal{R}^2(k) = \frac{1}{N_k} \sum_{|e|=k} \mathcal{R}^2[y_e],$$

$$\mathcal{R}^2[y_e] = 1 - \frac{\mathbb{E}_{\mathbf{u}}(y_e(\mathbf{u}) - \hat{f}(\mathbf{u}))^2}{\text{Var}_{\mathbf{u}} y_e(\mathbf{u})}, \quad (4.13)$$

averaged over all $N_k = \binom{D+k-1}{k}$ degree- k orthogonal polynomials, to indicate whether the trained predictor $\hat{f}(\mathbf{u})$ can generate nonlinearities of degree k in the $(D=n)$ -dimensional input [92]. $\mathcal{R}^2$ generalizes the geometric score (4.10) by taking statistical correlations into account (see Appendix C for further discussion). In the special case where the targets are whitened orthonormal features, one has $\mathcal{R}^2[y_r] = \gamma_r^2$ [see Eq. (C10)]. In the case of amplitude encoding (2.8) with complete Pauli- Z -string library as readout, the number of representable degree- k targets is $\binom{D}{k}$, giving an averaged score

$$\mathcal{R}_Z^2(k) = \begin{cases} \binom{D}{k} / \binom{D+k-1}{k} & \text{for } k \leq D \\ 0 & \text{for } k > D, \end{cases} \quad (4.14)$$

which exponentially decreases with k . The capacity calculations use unregularized least squares with a fitted intercept, $\hat{f}(\mathbf{u}) = \hat{b} + \hat{\mathbf{w}}^\top R\phi(\mathbf{u})$.

In Fig. 7, we first display the nonlinear capacity $\mathcal{R}_{\mathbb{I}}^2$ that can be *ideally* achieved from the encoded state without effects of feature mixing, i.e., we take $R = \mathbb{I}$ in (3.14), or $R = S$ when restricting to observable subsets ($m \leq q$). As shown in panel (a), a Pauli- X readout (black symbols with dotted curves) gen-

erates even degree polynomials, behaving similarly to angle encoding with Pauli- Z observables (connected open symbols). In Fig. 7(b) we show the ideal nonlinear capacity score $\mathcal{R}_{\mathbb{I}}^2(k)$ obtained for a complete Pauli observable set. Notably, the presence of non- Z Paulis significantly enhances the nonlinearity present in the readout, as becomes clear by comparison to the dotted curve corresponding to $\mathcal{R}_Z^2(k)$ for $D = n = 5$ [Eq. (4.14)]. Moreover, angle and amplitude encodings show a similar behavior. We attribute the decay of $\mathcal{R}_{\mathbb{I}}^2(k)$ with k mainly to the fact that the product state encoding only generates square-free monomials of the Pauli features ϕ_j , while we test against (the practically relevant case of) all possible degree- k targets [93].

Figure 8 illustrates the nonlinear capacity score $\mathcal{R}^2(k)$ obtained for a QELM, based on the TFIM or a random Hamiltonian, under temporal multiplexing of Pauli- Z and Pauli- ZZ observables. The score follows a similar pattern as the data-independent Pauli decodability γ_j^2 (see Fig. 4) as they both show a delayed growth with increasing nonlinear order k and multiplexing step L . However, $\mathcal{R}^2(k)$ saturates faster than γ_j^2 because the considered nonlinear target functions $y_k(\mathbf{u})$ can be generated already from relatively few Pauli features. The decrease of $\mathcal{R}^2(k)$ with k stems from two facts: (i) the encodings used here can generate only a small subset of all possible degree- k monomials [see Eq. (4.14) and Fig. 7]. A clearer measure of the effectiveness of temporal multiplexing to generate nonlinearities is thus the score $\mathcal{R}^2(k)$ normalized by the score $\mathcal{R}_{\text{all}}^2(k)$ obtained when measuring over the complete Pauli observable set (see Fig. 9). (ii) Higher-order observables can be missing for highly structured Hamiltonians or insufficient multiplexing steps. Hamiltonians with a more global and random structure are thus beneficial when measurements are restricted to Pauli- Z observables but higher-order nonlinearities are desired.

2. Nonlinear monomials at readout

Raw monomials in u_α are often relevant features for practical applications, such as autoregressive learning of dynamical systems. Let thus $\mathcal{B}(\mathbf{u}) = \{1, u_1, \dots, u_1 u_2, \dots, \prod_{\alpha} u_\alpha^r\}$ be the vector of size b of monomials of $\{u_\alpha\}_{1 \leq \alpha \leq D}$ up to degree at most r in each variable. The order- r Taylor expansion

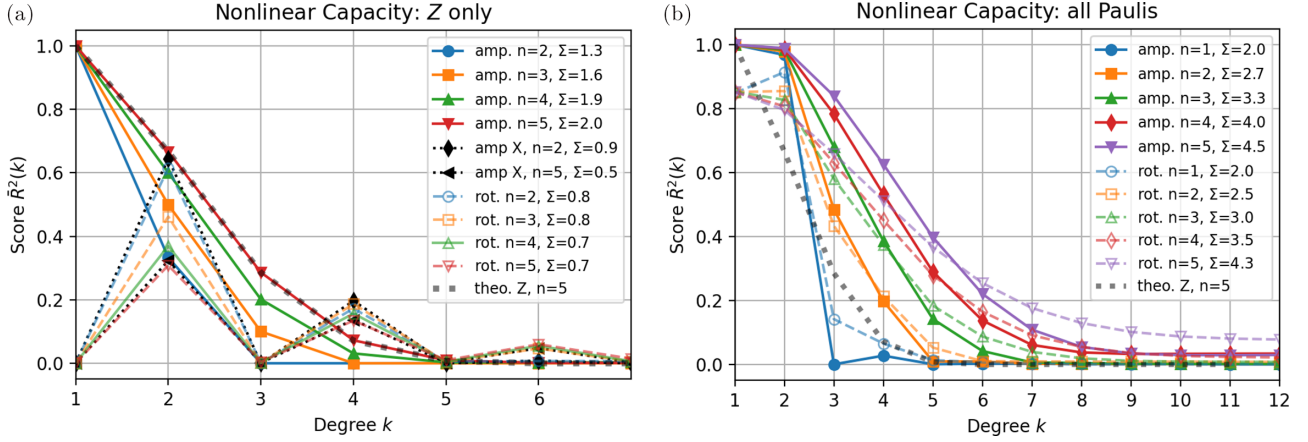


FIG. 7. Ideal nonlinear capacity score $\mathcal{R}_{\mathbb{I}}^2(k)$ [Eq. (4.13)] for amplitude and rotational encodings [Eqs. (2.8) and (2.10)]. This score indicates the ability to generate degree- k polynomials of the input u_α ($\alpha = 1, \dots, D$ and $n = D$) for a QELM $f(\mathbf{u}) = \mathbf{w}^\top \Phi(\mathbf{u})$ without feature mixing [i.e., with $R = \mathbb{I}$ in Eq. (3.14)]. Unless indicated, only Pauli-Z features are measured in (a). Nonlinear capacity is significantly enhanced when measuring over all Paulis (b). This is also demonstrated by the score for amplitude encoding with Pauli-X readout included in (a). The thick dotted curve represents the theoretical score for $\mathcal{R}_{\mathbb{I}}^2(k)$ with amplitude encoding and $n = 5$ input qubits [Eq. (4.14)], which agrees exactly with the numerical results. Σ denotes the integrated capacity over all k .

of the Pauli features $\{\phi_j(\mathbf{u})\}_{1 \leq j \leq q}$ within a small region around some operating point $\mathbf{u}_0$ (e.g., data mean) can then be written as $\Phi(\mathbf{u}) \simeq A(\mathbf{u}_0)B(\mathbf{u} - \mathbf{u}_0)$ with a coefficient matrix $A(\mathbf{u}_0) \in \mathbb{R}^{q \times b}$. In order to measure how accurately a certain monomial in B can be locally constructed from linear combinations of ϕ_j , we define a geometric decodability score

$$\kappa_j^2 \equiv (A^+ A)_{jj}, \quad \sum_{j=1}^b \kappa_j^2 = \text{rank}(A) \leq \min(q, b). \quad (4.15)$$

This score depends on the operating point $\mathbf{u}_0$ and is thus not a global reconstruction indicator such as $\mathcal{R}_{\mathbb{I}}^2[y]$. The score values indicate that construction is possible only within an error $\sim (1 - \kappa_j^2)^{1/2}$ due to off-diagonal terms in the j th row/column of the projector $A^+ A$. A somewhat artificial dependence of

κ_j^2 on the Taylor order r (owing to the rank constraint) can be mitigated by considering large r and b . Indeed, most of the κ_j^2 decrease upon increasing r , except for a few stable ones. These include monomials that arise directly from the tensor product structure of the encoding state. For amplitude encoding (2.8) including all Z strings, these are all square-free monomials up to order n (e.g., for $n = D = 3$: $\mathcal{B}_{\text{stable}} = \{1, x, y, z, xy, xz, yz, xyz\}$).

The typical behavior of κ_j^2 and of the projectors $A^+ A$ (averaged over a range of expansion points) is illustrated in Fig. 10. While the score typically decays with increasing nonlinear degree, rotational encoding tends to have a greater expressivity towards higher-order monomials. This is associated with less leakage into off-diagonal terms of $A^+ A$ [see Fig. 10(c)], and contributes to the robust performance often observed for rotational encoding.

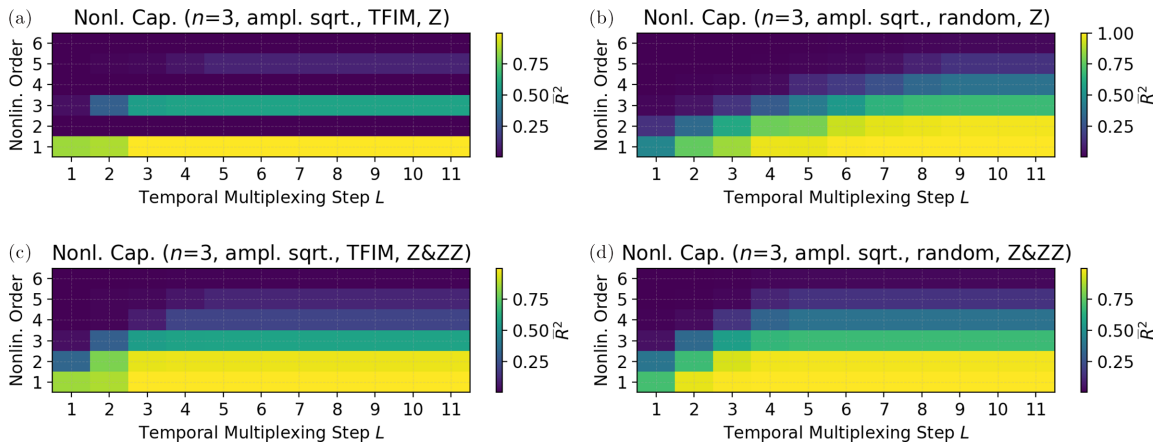


FIG. 8. Nonlinear capacity score $\mathcal{R}^2(k)$ [Eq. (4.13)] as a function of the nonlinear order k and the temporal multiplexing iteration length L [Eq. (3.9)]. The unitary evolution time is given by $t_L = L$. We consider an $n = 3$ qubit model in the same setup as in Fig. 4. The unitary is based on the TFIM (2.5a) [(a) and (c)] or a random Hamiltonian [(b) and (d)], while the observable set $\mathcal{S}$ consists of either only Pauli-Z [(a) and (b)] or Pauli-Z and Pauli-ZZ [(c) and (d)].

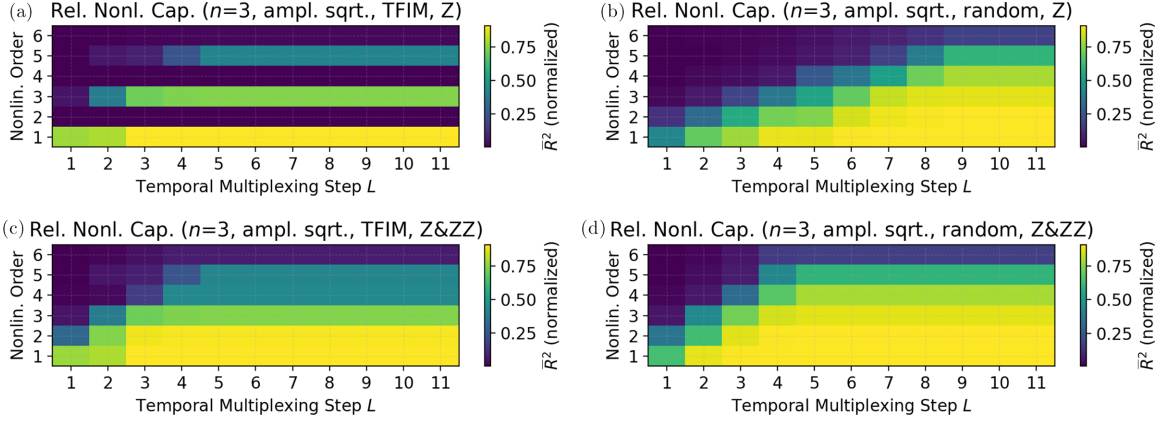


FIG. 9. Relative nonlinear capacity score $\mathcal{R}^2(k)/(\mathcal{R}^2_{\text{all}}(k) + \delta)$ [Eq. (4.13)] as a function of the nonlinear order k and the temporal multiplexing iteration length L [Eq. (3.9)]. $\mathcal{R}^2_{\text{all}}$ is the nonlinear capacity score obtained for a complete Pauli observable set, and we choose $\delta = 0.1$ as a threshold since scores $\mathcal{R}^2_{\text{all}} \lesssim 0.1$ are not considered meaningful. We use the same setup as in Fig. 8. The unitary is based on the TFIM (2.5a) [(a) and (c)] or a random Hamiltonian [(b) and (d)], while the observable set $\mathcal{S}$ consists of either only Pauli-Z [(a) and (b)] or Pauli-Z and Pauli-ZZ [(c) and (d)].

D. Factorization of the nonlinear processing capacity

The diagnostics introduced above probe different but interrelated aspects of the information processing capacity of a QELM: a target $y(\mathbf{u})$ must be approximated by functions con-

structible from the encoding-generated Pauli features $\Phi(\mathbf{u})$ (Secs. IV C 1 and IV C 2), and these features must in turn be decoded from the readout $\mathbf{F} = \mathbf{R}\Phi$ (Sec. IV B). As is shown in Appendix C, the nonlinear processing capacity factorizes

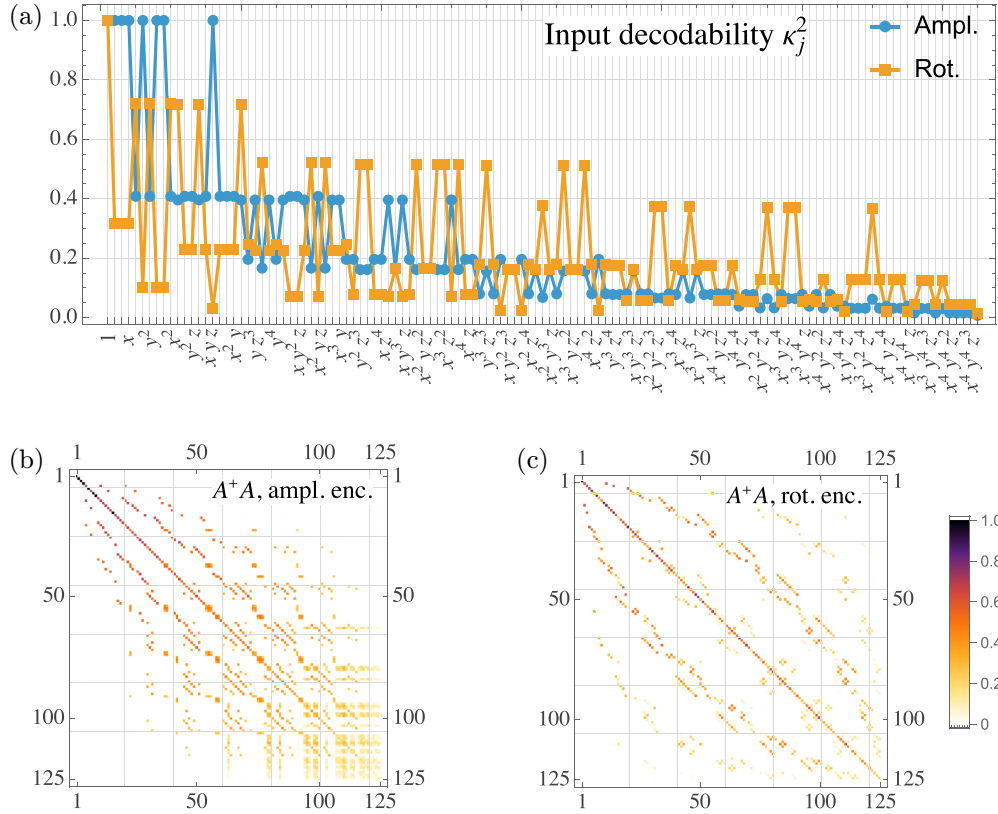


FIG. 10. Ability to construct monomials of the input $\{x, y, z\}$ for amplitude encoding (2.8) and rotational encoding (2.10) with $n = D = 3$. (a) Decodability score κ_j^2 according to Eq. (4.15). [(b) and (c)] Projector matrix A^+A , where nonzero off-diagonals correspond to $\kappa_j^2 < 1$ and indicate decoding is possible only within a certain accuracy. The underlying Taylor expansion is performed up to order $r = 4$ in each $u_j \in \{x, y, z\}$ and changes are mild upon further increasing r . The scores and matrices are averaged over a uniform grid of expansion points $\mathbf{u}_0 \in [0.1, 0.9]^3$ for amplitude encoding (with some margin to the singular boundaries) and $\mathbf{u}_0 \in [0, 1]^3$ for rotational encoding. For readability, monomial labels are incompletely displayed in (a) and replaced by column/row indices in (b) and (c).

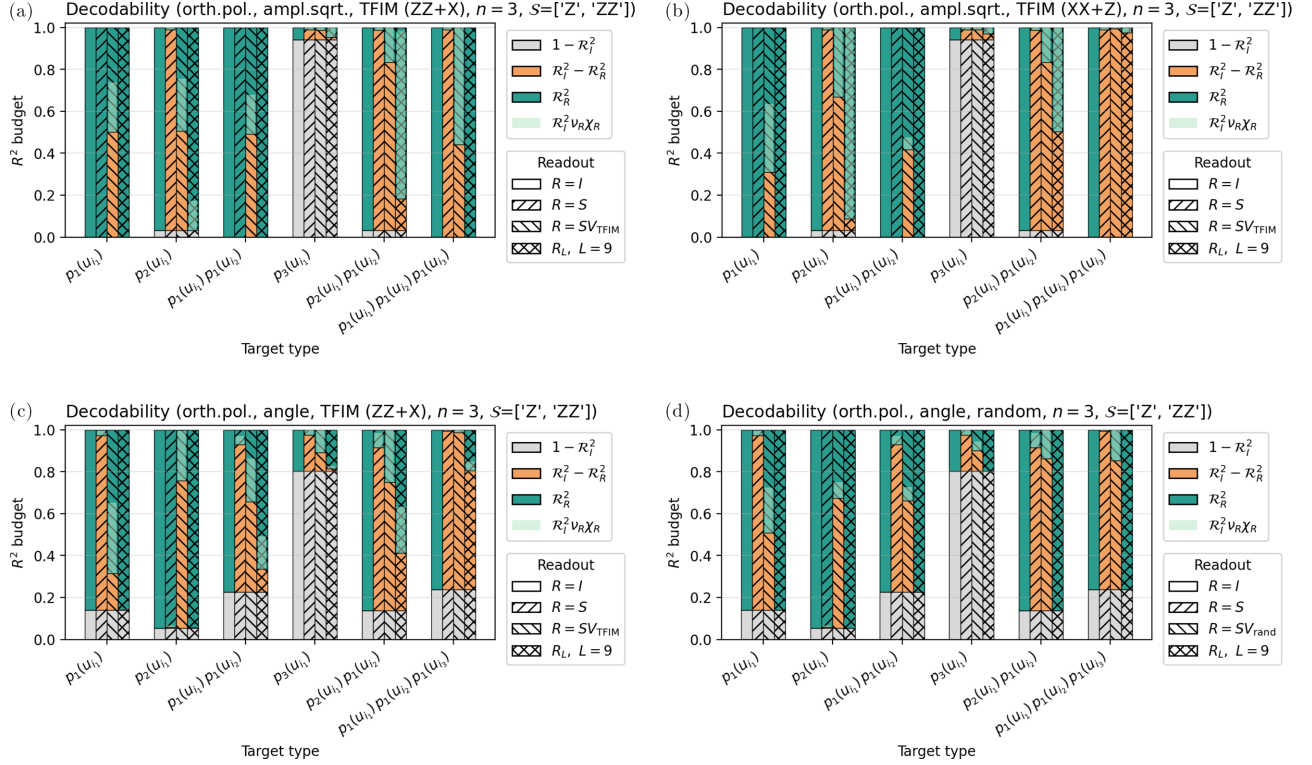


FIG. 11. Decodability scores $\mathcal{R}_R^2$ for families of orthogonal polynomials $p_j(u_{\alpha_j})$. Following Eq. (4.18), the undecodable fraction of the target is decomposed into a part not representable by the encoding ($1 - \mathcal{R}_I^2$, gray) and a part not transmitted through the reservoir (“channel loss”, $\mathcal{R}_I^2 - \mathcal{R}_R^2$, orange). Onto the decodable fraction $\mathcal{R}_R^2$ (green) the gain from geometrically hidden but statistically rescued features $\mathcal{R}_I^2 \nu_R \chi_R$ is overlaid (after clipping and scaling). The readout is constructed by either: measuring all Pauli observables directly after encoding ($R = \mathbb{I}$); just the selected set directly after encoding ($R = S$, here Z and ZZ); the selected set after Hamiltonian evolution ($R = SV$); the selected set after temporal multiplexing with the considered Hamiltonian for L steps (R_L). We consider amplitude or angle encoding [Eqs. (2.8) and (2.10) in the form $R_Y(u)$] and various Hamiltonians, as indicated.

along this chain [see Eq. (C18)]:

$$\mathcal{R}_R^2[y] = \mathcal{R}_I^2[y] \tau_R[y]. \quad (4.16)$$

The ideal factor $\mathcal{R}_I^2[y]$ measures the fraction of the target representable by the encoding-generated Pauli-feature library (Sec. IV C 1), whereas $\tau_R[y]$ measures the fraction of that optimal feature-space representative transmitted through the effective PTM and the selected measurements. Both $\mathcal{R}_I^2$ and τ_R take statistical correlations in the features into account. For a target aligned with a single isotropic or whitened feature coordinate, the channel factor reduces to the corresponding geometric score, $\gamma_r^2 = (R^+ R)_{rr}$ or $\tilde{\gamma}_r^2 = (\Pi_R)_{rr}$ [see Eqs. (C10) and (C12)]. Analogously, the geometric specialization of the encoding factor for the local Taylor map is the monomial score $\kappa_j^2 = (A^+ A)_{jj}$ [see Eqs. (4.15) and (C40)]. Averaging $\mathcal{R}_R^2[y]$ over the degree- k target family gives $\mathcal{R}_R^2(k)$ [see Eq. (4.13)].

The channel transmission τ_R can be refined further by separating geometric visibility from statistical recoverability (see Appendix C):

$$\tau_R[y] = 1 - \nu_R[y](1 - \chi_R[y]), \quad (4.17)$$

where $\nu_R[y]$ quantifies how much of the target lies in the nullspace of the PTM, while $\chi_R[y]$ quantifies how much of the target is predicted from statistical correlations of accessible

features. Thus a large channel transmission score can arise in two distinct ways: The task-relevant representative may be geometrically visible to the PTM or readout, corresponding to small $\nu_R[y]$, or it may be geometrically hidden but statistically recoverable from correlated accessible features, corresponding to large $\chi_R[y]$.

From Eq. (4.16) and the refined channel identity, the normalized target variance can be decomposed as [see Eq. (C34)]

$$1 = \underbrace{1 - \mathcal{R}_I^2[y]}_{\text{not represented by encoding}} + \underbrace{\mathcal{R}_I^2[y] \nu_R[y] (1 - \chi_R[y])}_{\text{hidden and not statistically rescued}} + \underbrace{\mathcal{R}_R^2[y]}_{\text{decoded}}, \quad (4.18)$$

where the middle term is equivalent to $\mathcal{R}_I^2[y] - \mathcal{R}_R^2[y]$. The score $\mathcal{R}_I^2[y] \nu_R[y] \chi_R[y]$ describes the reduction in channel mean-squared error obtained by predicting the geometrically hidden component from the accessible coordinates, relative to the predictor that uses only the directly visible component.

The decodability budget is visualized in Fig. 11 and the stacked bars indicate what the encoding failed to generate, what the encoding generated but the PTM or readout failed to transmit even after statistical rescue, and what is ultimately decodable, either statistically or geometrically. In general, the reservoir cannot enhance decodability in cases where the target cannot already be represented by the encoding alone. As expected, if the reservoir hinders decodability

TABLE III. Summary of the decodability diagnostics. The encoding factor $\mathcal{R}_{\mathbb{I}}^2[y]$ measures how much of a target is represented by the encoding-induced Pauli-feature library, while the channel factor $\tau_R[y]$ measures how much of this representative is transmitted by the effective PTM and readout. The refined decomposition [see Eq. (4.18)] separates the channel loss into the nullspace load $\nu_R[y]$ [Eq. (C27)] and the statistically rescued fraction $\chi_R[y]$ [Eq. (C26)].

	Encoding	Channel	Composite
Geometric, per direction	$\kappa_j^2 = (A^\dagger A)_{jj}$	$\gamma_r^2 = (R^\dagger R)_{rr}$	–
Statistical, per task y	$\mathcal{R}_{\mathbb{I}}^2[y]$	$\tau_R[y]$	$\mathcal{R}_R^2[y] = \mathcal{R}_{\mathbb{I}}^2[y]\tau_R[y]$
Refined channel split	–	$\tau_R[y] = 1 - \nu_R[y](1 - \chi_R[y])$	$\mathcal{R}_R^2[y] = \mathcal{R}_{\mathbb{I}}^2[y][1 - \nu_R[y](1 - \chi_R[y])]$
Statistical rescue	–	$\nu_R[y]\chi_R[y]$	$\mathcal{R}_{\mathbb{I}}^2[y]\nu_R[y]\chi_R[y]$

of a representable feature, this effect can be mitigated by temporal multiplexing. Notably, when measuring with Pauli-Z and -ZZ observables, decodability can grow or decrease when the reservoir unitary is removed from the model. Moreover, comparing Figs. 11(a) and 11(b), the reduced geometric decodability of features for certain structured Hamiltonians observed in Fig. 4 is also reflected by the statistical counterpart (see also Fig. 17 for the monomial case).

E. Summary and discussion

Decodability of Pauli features can be understood as the underlying concept that determines the expressivity of a QELM with initial-state encoding and data-independent reservoir dynamics. The decodability score $\mathcal{R}^2$ corresponds to the (nonlinear) information processing capacity [54] of the QELM, and we have shown here that it separates into two parts: (i) an *encoding part* $\mathcal{R}_{\mathbb{I}}^2[y]$ (geometric proxies: κ_j^2 , A), measuring whether the encoding-generated dictionary of Pauli features can build the monomials and orthogonal polynomials relevant for the task, and (ii) a *channel part* $\tau_R[y]$ (with geometric proxies γ_r^2 , $\tilde{\gamma}_r^2$, $\text{rank } R$), measuring the invertibility of the effective PTM on the task-relevant feature directions, conditioned on the encoding distribution. The channel score receives contributions both from features that are decodable purely algebraically and from features that are algebraically invisible but still decodable via statistical correlations. The different notions of decodability are summarized in Table III.

Nonlinear processing capacity has been found to track Krylov observability very closely [47], which is explained by the present framework in terms of the PTM-induced mixing of the fundamental Pauli features. High decodability scores are associated with high-rank effective PTMs R , which are achieved for random unitaries. Strongly mixing quantum channels can help expressivity and nonlinear processing in QELMs [14,45,50], and are thus often beneficial for training performance apart from concentration effects [53]. By contrast, QRCs with internal memory states often work best at a point where random matrix behavior just begins to dominate (“edge of chaos”) (see, e.g., Refs. [42,94,95]). An alternative to increasing randomness is to devise task-aligned reservoir structures and measurement schemes [2,33,44,50,96]. Knowing the connection between the readout and the encoding-generated features via the PTM framework can thus inform feature engineering strategies to enhance expressivity.

Before proceeding, it is useful to also briefly discuss the *injectivity* property of the feature map [49], which ensures that distinct inputs map to distinct readouts. A sufficient condition for local injectivity of $\mathbf{F}(\mathbf{u}) = R\Phi(\mathbf{u})$ [Eq. (3.14)] is a full-rank Jacobian:

$$\text{rank } RJ_\Phi(\mathbf{u}) = D, \quad J_\Phi(\mathbf{u}) = \frac{\partial \Phi(\mathbf{u})}{\partial \mathbf{u}} \in \mathbb{R}^{q \times D}. \quad (4.19)$$

For the considered product state encodings it is straightforward to check that $\Phi(\mathbf{u})$ is injective on the respective domains in $\mathbb{R}^D$ (possibly after removing one end point). Hence, local injectivity of $\mathbf{F}(\mathbf{u})$ is determined by whether R is injective on the tangent space range $J_\Phi(\mathbf{u})$, i.e., whether

$$\text{null } R \cap \text{im } J_\Phi(\mathbf{u}) = \{0\}. \quad (4.20)$$

In a generic setting, this condition is satisfied (at fixed $\mathbf{u}$) if $\text{rank } R \geq D$, which is the case for adequate measurement schemes and sufficiently scrambling dynamics.

V. INTERPRETABLE LEARNING OF DYNAMICAL SYSTEMS

A. Time-series modeling with QELMs

The classical representation (3.14) allows us to identify the surrogate model a QELM learns during the training process. Specifically, for the product state encodings considered in Sec. II B, the function library $\Phi : \mathbb{R}^D \rightarrow \mathbb{R}^q$ consists of $q = 3^n$ monomials of the single-Pauli features ϕ_k (recall that observables containing Pauli-Y factors have vanishing expectation values, see Table I). We consider the standard problem of learning to forecast a chaotic dynamical system by training on its sampled trajectory $\mathbf{u}_t = \mathbf{u}(t_0 + t \Delta t) \in \mathbb{R}^D$. The continuous and discrete evolutions are given by

$$\text{dynamical system: } \dot{\mathbf{u}} = \mathbf{g}(\mathbf{u}), \quad \mathbf{u}_{t+1} = \Psi_{\mathbf{g}}(\mathbf{u}_t), \quad (5.1)$$

where $\Psi_{\mathbf{g}}(\mathbf{u}) = \mathbf{u} + \Delta t \mathbf{g}(\mathbf{u}) + \mathcal{O}(\Delta t^2)$ is the flow map (propagator) that can be computed from the vector field $\mathbf{g}(\mathbf{u}) \in \mathbb{R}^D$ via a Taylor-Lie series expansion:

$$\Psi_{\mathbf{g}}(\mathbf{u}) = \exp(\Delta t \mathcal{L}_{\mathbf{g}}) \mathbf{u} = \sum_{k=0}^{\infty} \frac{\Delta t^k}{k!} (\mathcal{L}_{\mathbf{g}}^k \mathbf{u}),$$

$$(\mathcal{L}_{\mathbf{g}} f)(\mathbf{x}) = \mathbf{g}(\mathbf{x}) \cdot \nabla f(\mathbf{x}). \quad (5.2)$$

The nonintegrable nature and fractal attractors of chaotic systems make them natural benchmarks for system identification, which addresses the question to which extent the underlying

data-generating process can be uniquely discovered from the data [56,97].

If the dynamics is Markovian, then it can be modeled solely with an NG-RC or QELM based on an instantaneous feature map (see Ref. [33] for examples). Accordingly, we formulate the training problem of the QELM (3.14) as

$$\text{QELM: } \mathbf{u}_{t+1} = \mathbf{f}(\mathbf{u}_t) = W^\top R \Phi(\mathbf{u}_t) \quad (5.3)$$

$$\text{Lorenz-63: } \begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = x(\rho - z) - y \\ \dot{z} = xy - \beta z \end{cases} \quad \text{Halvorsen: } \begin{cases} \dot{x} = -ax - 4y - 4z - y^2 \\ \dot{y} = -ay - 4z - 4x - z^2 \\ \dot{z} = -az - 4x - 4y - x^2 \end{cases}, \quad (5.4)$$

with parameters $\sigma = 10$, $\beta = 8/3$, $\rho = 28$ for Lorenz-63, and $a = 1.4$ for Halvorsen. This setup leads to chaotic behavior with largest Lyapunov exponents $\lambda_{\max}^{\text{Lor63}} \approx 0.89$ and $\lambda_{\max}^{\text{Halv}} \approx 0.76$, respectively. The polynomial form of the vector field in (5.4) then suggests to use amplitude encoding (2.8) with $n = 3$ qubits, leading to a library of the following 27 terms:

$$\begin{aligned} \Phi_{\text{amp}}^{n=3}(\mathbf{u}) = & \{1, 2\sqrt{(1-x)x}, 2x-1, 2\sqrt{(1-y)y}, 2(2x-1)\sqrt{-(y-1)y}, 4\sqrt{(x-1)x(y-1)y}, 2y-1, \\ & 2\sqrt{-(x-1)x(2y-1)}, (2x-1)(2y-1), 2\sqrt{(1-z)z}, 2(2x-1)\sqrt{-(z-1)z}, \\ & 2(2y-1)\sqrt{-(z-1)z}, 2(2x-1)(2y-1)\sqrt{-(z-1)z}, 4\sqrt{(x-1)x(z-1)z}, \\ & 4(2y-1)\sqrt{(x-1)x(z-1)z}, 4\sqrt{(y-1)y(z-1)z}, 4(2x-1)\sqrt{(y-1)y(z-1)z}, \\ & 8\sqrt{-(x-1)x(y-1)y(z-1)z}, 2z-1, 2\sqrt{-(x-1)x(2z-1)}, (2x-1)(2z-1), \\ & 2\sqrt{-(y-1)y(2z-1)}, 2(2x-1)\sqrt{-(y-1)y(2z-1)}, 4(2z-1)\sqrt{(x-1)x(y-1)y}, \\ & (2y-1)(2z-1), 2\sqrt{-(x-1)x(2y-1)(2z-1)}, (2x-1)(2y-1)(2z-1)\}. \end{aligned} \quad (5.5)$$

We will also consider the amplitude encoding scheme (2.9), which has a similar polynomial structure.

In order to confine the trajectory to the interval $\mathbf{u} \in [0, 1]^3$ or $[-1, 1]^3$, as required by the encoding, we apply the transformation $\mathbf{u} \rightarrow \mathbf{u}' = \alpha \odot (\mathbf{u} - \mathbf{m})$ with $\alpha, \mathbf{m} \in \mathbb{R}^3$ to the state (we typically use the same α for all coordinates). This leads to the rescaled Lorenz-63 model:

$$\begin{aligned} \dot{u} &= -\sigma u + \sigma \frac{\alpha_x}{\alpha_y} v + \sigma \alpha_x (m_y - m_x), \\ \dot{v} &= -v + \frac{\alpha_y}{\alpha_x} (\rho - m_z) u - \frac{\alpha_y}{\alpha_z} m_x w - \frac{\alpha_y}{\alpha_x \alpha_z} u w \\ &\quad + \alpha_y m_x (\rho - m_z) - \alpha_y m_y, \\ \dot{w} &= -\beta w + \frac{\alpha_z}{\alpha_x \alpha_y} u v + \frac{\alpha_z}{\alpha_x} m_y u + \frac{\alpha_z}{\alpha_y} m_x v \\ &\quad + \alpha_z (m_x m_y - \beta m_z), \end{aligned} \quad (5.6)$$

and the rescaled Halvorsen model:

$$\begin{aligned} \dot{u} &= -au - 4\frac{\alpha_x}{\alpha_y} v - 4\frac{\alpha_x}{\alpha_z} w - 2\frac{\alpha_x m_y}{\alpha_y} v - \frac{\alpha_x}{\alpha_y^2} v^2 \\ &\quad + \alpha_x (-am_x - 4m_y - 4m_z - m_y^2), \\ \dot{v} &= -av - 4\frac{\alpha_y}{\alpha_z} w - 4\frac{\alpha_y}{\alpha_x} u - 2\frac{\alpha_y m_z}{\alpha_z} w - \frac{\alpha_y}{\alpha_z^2} w^2 \\ &\quad + \alpha_y (-am_y - 4m_z - 4m_x - m_z^2), \\ \dot{w} &= -aw - 4\frac{\alpha_z}{\alpha_x} u - 4\frac{\alpha_z}{\alpha_y} v - 2\frac{\alpha_z m_x}{\alpha_x} u - \frac{\alpha_z}{\alpha_x^2} u^2 \\ &\quad + \alpha_z (-am_z - 4m_x - 4m_y - m_x^2). \end{aligned} \quad (5.7)$$

and seek to learn an approximation to the flow map ψ_g via minimizing the least-squares loss (2.2) on the dataset $\{\mathbf{x} = \mathbf{u}_t, \mathbf{y} = \mathbf{u}_{t+1}\}_{t=1}^P$. The generalization of (2.1) to multiple outputs is straightforward and involves training a separate predictor f_r for each output dimension r via the weight matrix $W = [\mathbf{w}_1, \dots, \mathbf{w}_D] \in \mathbb{R}^{B \times D}$.

We use the Lorenz-63 and Halvorsen models as benchmarks ($D = 3$ dimensions) [98]:

This vector field-level transformation simplifies comparison between the learned coefficients and the true ones.

B. Numerical results

We use here a complete measurement operator set of 4^n Pauli observables. In this setup, the reservoir unitary becomes immaterial [see Eq. (3.18)]. We generate trajectories using a fifth-order Runge-Kutta integrator with a time step of $\Delta t = 0.01$. Reported values of the training error and forecast horizon are averaged over multiple initial conditions. After training the QELM, we obtain a classical surrogate model $\hat{\mathbf{f}}(\mathbf{u})$ by inserting into Eq. (3.14) the learned weights $\hat{\mathbf{w}}$.

The root-mean-square (rms) training error

$$\epsilon_{\text{train}} = \left(\frac{1}{P} \sum_{t=1}^P \|\mathbf{u}_{t+1} - \hat{\mathbf{f}}(\mathbf{u}_t)\|^2 \right)^{1/2} \quad (5.8)$$

measures one-step accuracy of the trained predictor $\hat{\mathbf{f}}$. As a measure of multistep prediction accuracy, the autonomously generated trajectory $\hat{\mathbf{u}}_{t+1} = \hat{\mathbf{f}}(\hat{\mathbf{u}}_t)$ is compared to the true trajectory via the forecast horizon

$$T_{\text{fch}} / \Delta t = \inf \{t : \|\mathbf{u}_t - \hat{\mathbf{u}}_t\| > \varsigma\}, \quad (5.9)$$

where $\bar{\mathbf{u}} = \frac{1}{P} \sum_{t=1}^P \mathbf{u}_t$ and $\varsigma^2 = \frac{1}{P} \sum_{t=1}^P \|\bar{\mathbf{u}} - \mathbf{u}_t\|^2$ are the mean and variance of the trajectory, respectively. We average T_{fch} over multiple starting points $\mathbf{u}_0 \in \mathbb{R}^D$ and express it in units of the Lyapunov time $T_L = 1/\lambda_{\max}$.

TABLE IV. Training rms errors ϵ_{train} [Eq. (5.8)] and forecast horizons T_{fch} [Eq. (5.9)] for the benchmark systems and (unitary-free) QELM setups discussed in the text. We generally use $n = 3$ input qubits (no delay states) and a time resolution of $\Delta t = 0.01$. T_{fch} with its standard deviation (not error of the mean) is given in units of the Lyapunov time T_L . λ denotes the Ridge regularization parameter. The column “type” indicates whether we use the standard [Eq. (5.4)] or rescaled dynamical system. In the former case, we uniformly scale the trajectories to the interval reported in the 3rd column. In the latter case, we solve the dynamical system in Eqs. (5.6) and (5.7) with the rescaling factor α (same for all state coordinates).

System	Type	QELM setup	α	λ	ϵ_{train}	T_{fch}/T_L
Lorenz-63	std.	ampl. enc. [0,0.3], Eq. (2.8)	—	10^{-6}	10^{-5}	4.3 ± 1.7
Lorenz-63	resc.	ampl. enc. [0,1], Eq. (2.8)	0.005	10^{-8}	10^{-6}	6.2 ± 1.4
Halvorsen	std.	ampl. enc. [0.1,0.8], Eq. (2.8)	—	10^{-8}	10^{-4}	2.3 ± 0.8
Halvorsen	std.	$\dots$ + add squares to readout vector	—	10^{-8}	10^{-5}	5.8 ± 1.7
Halvorsen	resc.	ampl. enc. [0,1], Eq. (2.8)	0.027	10^{-7}	10^{-4}	3.4 ± 1.0
Halvorsen	resc.	$\dots$ + add squares to readout vector	0.027	10^{-7}	10^{-6}	6.8 ± 1.7
Halvorsen	std.	ampl. enc. $[-0.01, 0.01]$, Eq. (2.9)	—	10^{-15}	10^{-7}	5.9 ± 1.2
Halvorsen	resc.	ampl. enc. $[-1, 1]$, Eq. (2.9)	0.001	10^{-15}	10^{-7}	5.9 ± 1.5

Results of training a QELM in various setups are summarized in Table IV. We emphasize that our aim here is not to beat previous benchmarks, but to perform controlled interpretability experiments. We thus keep the setup deliberately simple. Significant improvements in training error and forecast horizon can be achieved by further optimizing the encoding and feature-engineering of the readout. Figure 12 demonstrates consistency between the autonomously generated prediction of the trained QELM and the iterated classical model, exemplified for the rescaled *Lorenz-63* model. Both remain close to the true trajectory for a time $T_{\text{fch}}/T_L \approx 6$, when the intrinsic Lyapunov instability of the chaotic system leads to a divergence, triggered by errors in the initial condition,

discretization, and extra terms in the surrogate model (see below) [99].

When repeating the experiments with the *Halvorsen* model (5.7), we obtain slightly lower forecast horizons of $T_{\text{fch}}/T_L \approx 3$ and rms training errors of $\epsilon_{\text{train}} \approx 10^{-4}$. One reason for this reduced performance compared to the Lorenz-63 model lies in the absence of squares of the state variables $\{x, y, z\}$ in the library (5.5). These are crucial for the Halvorsen equations (5.7), whereas the Lorenz-63 model (5.6) depends only on cross terms of the state variables. Indeed, when enhancing the readout vector by its squares, prediction accuracy for the Halvorsen model improves significantly (see Table IV). Alternatively, a similar performance can also be achieved

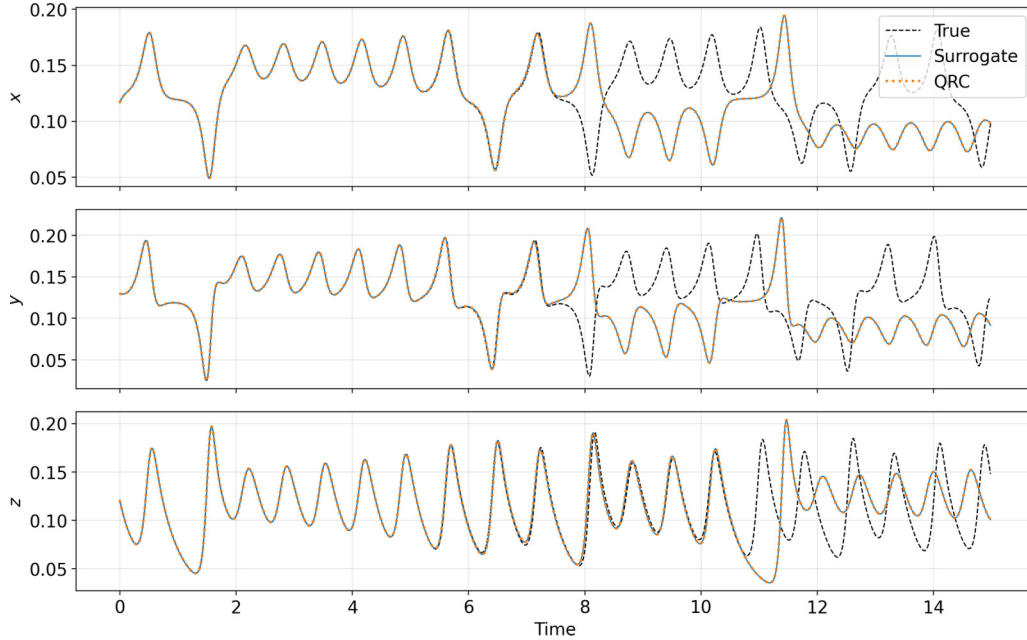


FIG. 12. QELM versus its classical representation: The forecast obtained from a QELM [Eq. (2.1)] trained on the rescaled Lorenz-63 model is matched exactly by the forecast of its iterated classical (surrogate) representation [Eq. (3.14)]. Both eventually diverge from the true trajectory due to the Lyapunov instability. The QELM employs amplitude encoding (2.8) with $n = 3$ qubits. Ground-truth trajectories are generated with a resolution of $\Delta t = 0.01$. Time is given in model time units. For the scaling transformation in Eq. (5.6), we use parameters $\alpha_{x,y,z} = 0.004$ and $\mathbf{m} = [-30, -30, -5]$.

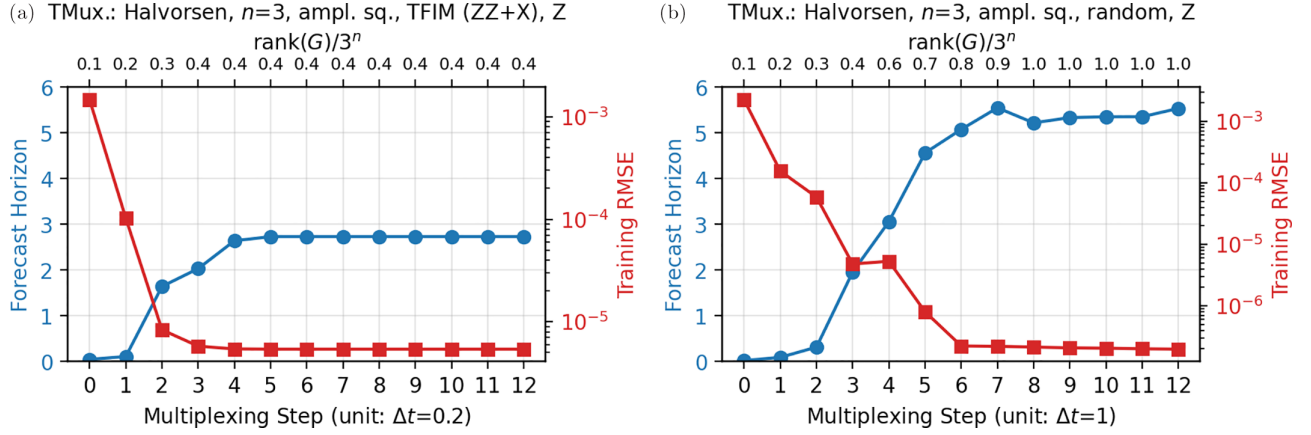


FIG. 13. Prediction performance (forecast horizon in units of the Lyapunov time T_L , and rms training error) under temporal multiplexing of a QELM trained on the Halvorsen model (5.7). We use amplitude encoding (2.9) with $n = 3$ qubits, single-qubit Pauli-Z observables, and (a) the TFIM Hamiltonian $H_{\text{TFIM}}^{\text{ZZ+X}}$ [Eq. (2.5a)] or (b) a random Hamiltonian. The top horizontal axis shows the numerically obtained rank of the readout matrix G [Eq. (3.17)] relative to the maximum achievable rank $q = 3^n$ [Eq. (3.19)]. No additional features are added to the readout. For the TFIM variant $H_{\text{TFIM}}^{\text{XX+Z}}$ [Eq. (2.5b)], we obtain a maximum rank of $\text{rank}(G)/3^n \simeq 0.22$, rms error $\epsilon_{\text{train}} \simeq 7 \times 10^{-4}$, and forecast horizon $T_{\text{fch}} \simeq 0$.

when using the amplitude encoding variant in (2.9), which already involves squares of the state variables. In the context of reservoir computing it is, of course, well known that prediction performance can be enhanced by feature engineering [65,67,100]. The present analysis exemplifies that, within the PTM framework, this optimization can be systematically phrased as an inverse problem, where the structure of the ground-truth model dictates the required terms in the encoding.

When employing *temporal multiplexing*, the rank of the readout matrix G [Eq. (3.17)] can serve as an indicator of the training accuracy, as illustrated in Fig. 13. This rank is upper bounded by the number of nonzero Pauli features q , which, for the encoding (2.9) used here, is $q = 3^n$. When conditioning is controlled and exponential concentration effects are avoided [48,53,81], achieving $\text{rank}(G) \simeq q$ makes essentially all Pauli features available to the readout and can thus improve performance. Indeed, the QELM in Fig. 13(b) achieves training error and forecast horizon essentially identical to the ones reported in Table IV obtained with a complete Pauli observable set. By contrast, Fig. 13(a) shows that performance remains poor when the readout utilizes only a few effective observables. In the present setting, this is a consequence of using a structured Hamiltonian like the TFIM [Eq. (2.5a)], which spreads observables less effectively over the full operator space, so that $\text{rank}(G) \simeq \text{rank}(R) \ll q$ [see Eq. (3.20) and Fig. 4]. This effect is most pronounced when using the specific TFIM variant $H_{\text{TFIM}}^{\text{XX+Z}}$ [Eq. (2.5b)] with Z-type observables. Numerical experiments confirm the close relation (3.20) between the readout and PTM rank.

C. Interpretability

When training an NG-RC with an instantaneous feature library on a Markovian dynamical system, the model learns an approximation of the flow map [67,70,101–103]. Libraries Φ of the form (5.5), however, do not render directly interpretable

models due to the structural mismatch with the ground truth model [104]. In the present teacher-forcing setup, a way to render the model interpretable is thus to express the feature map Φ in terms of an N -dimensional library $\mathcal{B}$ of functions that make up the flow map ψ_g [Eq. (5.1)],

$$\Phi(\mathbf{u}) \simeq K\mathcal{B}(\mathbf{u}), \quad (5.10)$$

and insert this into the learned model (3.14),

$$\hat{\mathbf{f}}(\mathbf{u}) \simeq \hat{W}^\top R K \mathcal{B}(\mathbf{u}) = \Psi_{\text{approx}}(\mathbf{u}). \quad (5.11)$$

This renders an approximation Ψ_{approx} to the flow map via Eq. (5.3), with a transformation matrix $K \in \mathbb{R}^{q \times N}$.

While there are several methods to determine K (such as collocation), we perform here a Taylor expansion of $\Phi(\mathbf{u})$ around some operating point $\bar{\mathbf{u}}$ (data mean). This is natural here, since, for the considered polynomial dynamical systems, we can take $\mathcal{B} = \mathcal{B}_r$ to be the set of *monomials* in u_j up to order r , of which there are $N = \binom{D+r}{r}$. Recall that low-order monomials are well constructible from the library Φ (see Sec. IV C 2). However, due to the tensor product structure of Φ [see Eq. (3.12)] and the fact that the derivatives of the single-qubit feature vector lie in a two-dimensional space (vanishing Pauli-Y), K typically does not have full column rank here when $N < q$. A consequence of this is that only a $\text{rank}(K)$ -dimensional subspace of the library $\mathcal{B}$ is accessible through the encoding features Φ and that hence some monomials in $\mathcal{B}$ are not independently identifiable. Nevertheless, deviations are small for the settings considered here [105].

Figure 14 shows the learned flow map increments $\Psi(\mathbf{u}) - \mathbf{u}$, resulting from Eq. (5.11) after training a QELM with amplitude encoding [Eqs. (2.8) and (2.9)] on the (rescaled) Lorenz-63 and Halvorsen models (no feature engineering in the readout). Comparison to the theoretical expectation [Eq. (5.2)] reveals good agreement especially for low-order coefficients. While deviations can be large for higher-order

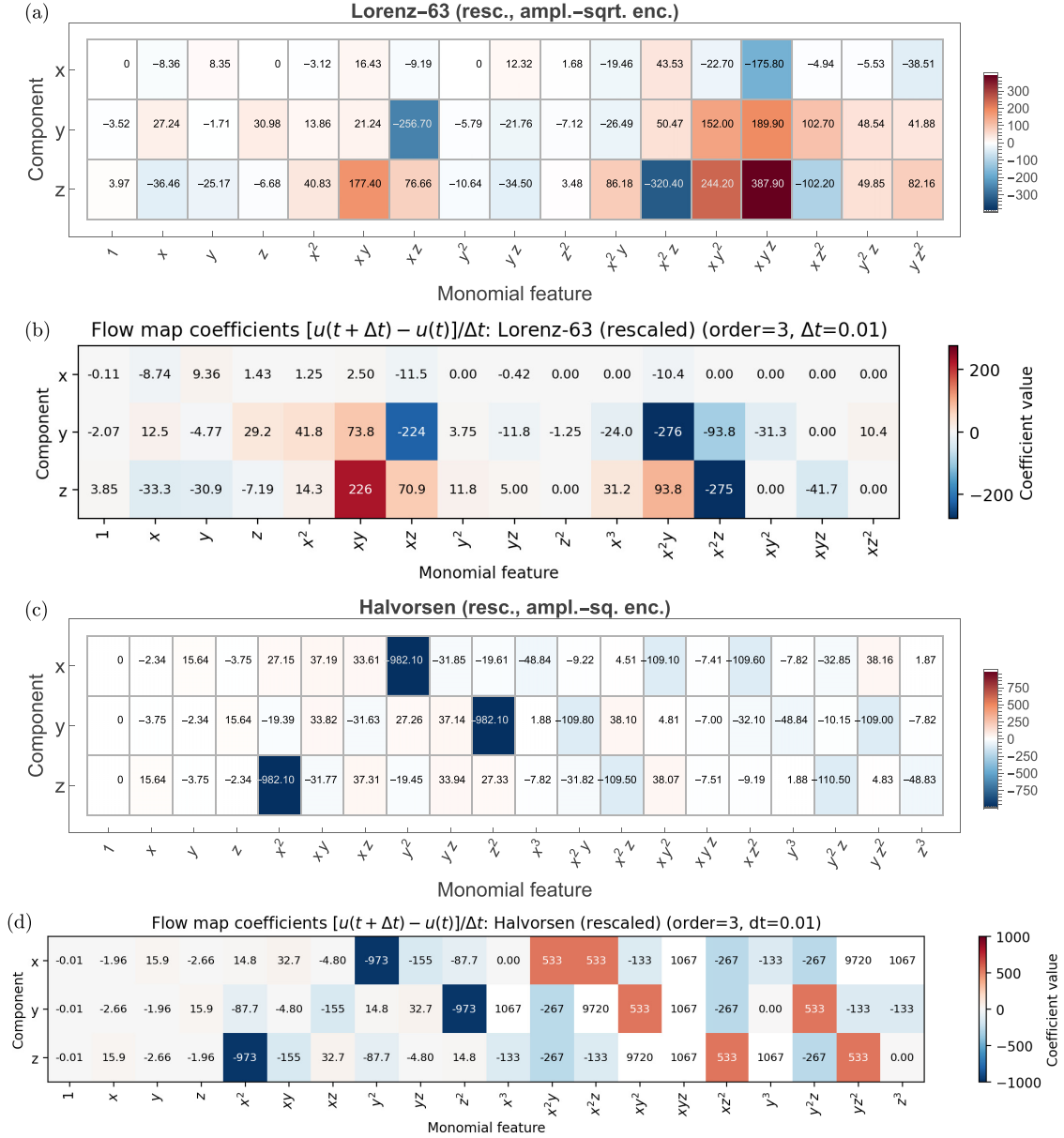


FIG. 14. Illustration of the flow map learned by a QELM, obtained from the Taylor-approximation approach [Eq. (5.11)]. The QELM is trained on (a) the Lorenz-63 model (5.6) and (c) the Halvorsen model (5.7), using the *amplitude* encoding variants (2.8) and (2.9), respectively. The rescaling parameters are here $\mathbf{m} = [-30, -30, -5]$, $\alpha_{x,y,z} = 0.004$ (Lorenz-63) and $\mathbf{m} = [-10, -10, -10]$, $\alpha_{x,y,z} = 0.001$ (Halvorsen). The corresponding true flow map increments $(\psi_g(\mathbf{u}) - \mathbf{u})/\Delta t$ [Eq. (5.1)], expanded to third order in the state variables u_j , are shown in (b) and (d) for comparison (terms identically zero are omitted).

coefficients, their influence on the overall training error is still small.

As illustrated in Fig. 15, also in the case of rotational encoding [Eq. (2.10)] the QELM forms linear combinations of the feature library $\Phi(\mathbf{u})$ so that its Taylor expansion approximately matches the true flow map ψ_g [Eq. (5.2)]. Compared to amplitude encoding, rotational encoding renders a slightly more accurate flow map approximation and lower training error. We find that the lowest training error is achieved when scaling the training trajectories $\mathbf{u}(t)$ to an interval $\sim 10^{-2}$ around their mean.

VI. SUMMARY

We have employed the PTM formalism to describe training of a QELM (a memoryless QRC) as a decoding problem of features mixed by quantum channels, and discussed its implications for system identification and time-series prediction. Writing states and channels in the Pauli basis (or any other orthogonal operator basis) turns the QELM into a sequence of real linear maps $T_{\mathcal{E}_j}$ acting on the vector of Pauli features $\Phi(\mathbf{u})$ representing the encoded state $\rho_{\text{enc}}(\mathbf{u})$ [29,37]. The QELM thus takes the form of a classical regression model (or “next-generation” RC [67]) over random features $T\Phi(\mathbf{u})$,

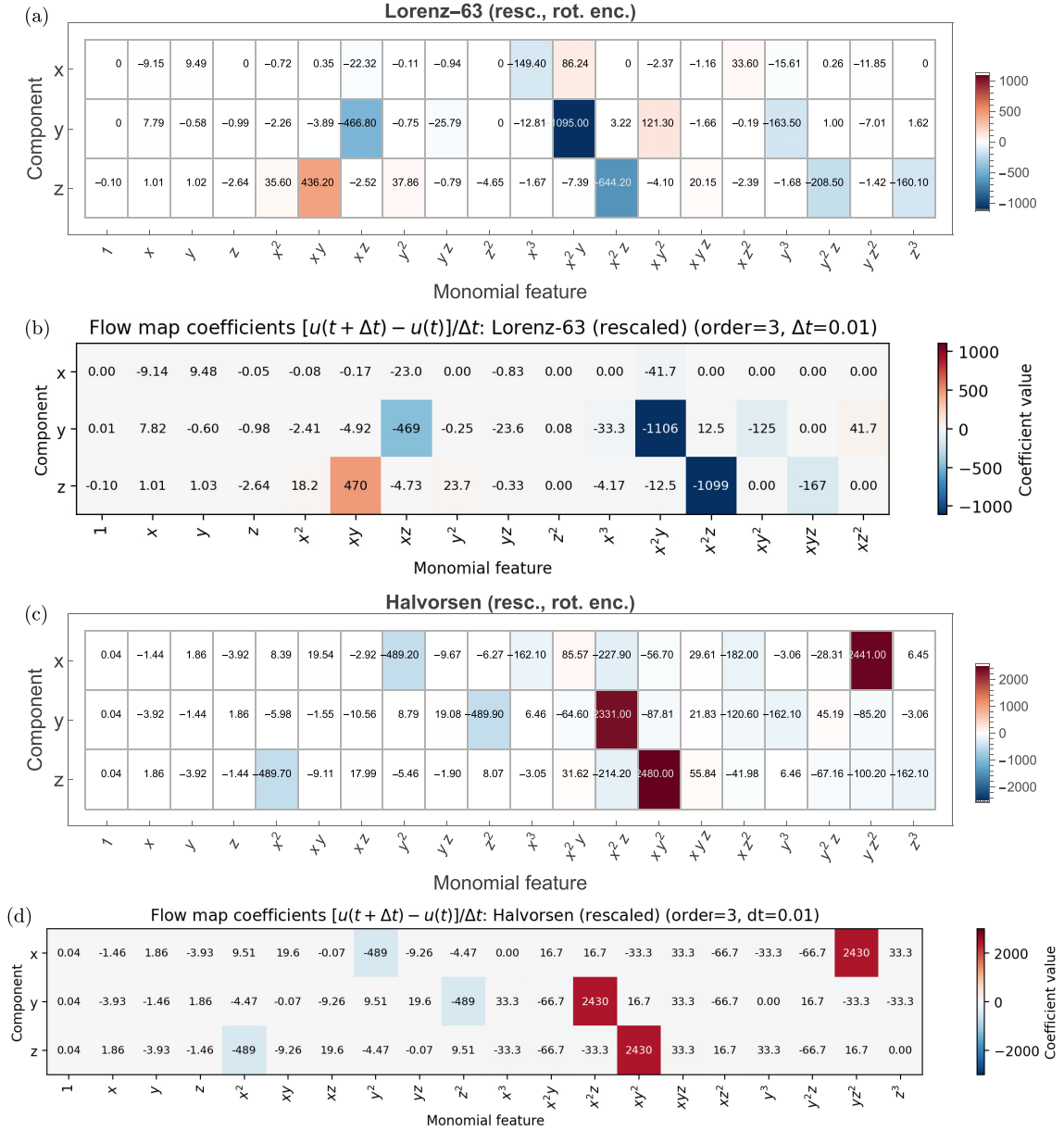


FIG. 15. Illustration of the flow map learned by a QELM, obtained from the Taylor-approximation approach [Eq. (5.11)]. The QELM uses *rotational* encoding [Eq. (2.10)] and is trained on (a) the Lorenz-63 model (5.6) and (c) the Halvorsen model (5.7). The rescaling parameters are here $\mathbf{m} = [1, 1, 20]$, $\alpha_{x,y,z} = 0.002$ (Lorenz-63) and $\mathbf{m} = [-3, -3, -3]$, $\alpha_{x,y,z} = 0.002$ (Halvorsen). The corresponding true flow map increments $(\psi_g(\mathbf{u}) - \mathbf{u})/\Delta t$ [Eq. (5.1)], expanded to third order in the state variables u_j , are shown in (b) and (d) for comparison. The coefficients differ from Fig. 14 due to the different expansion points.

which carry the full dependence of the QELM on the encoded input $\mathbf{u}$ and determine the nonlinearities that the model can possibly express. We remark that classical representations of QML models have been explored in various previous works [50,106–109].

While the PTM-based regression representation of a QELM is essentially a known consequence of the linear nature of a data-independent quantum channel [29,37,46,49], our work provides several new insights into QELMs with initial-state encoding. We have shown that nonlinear information processing capacity of a QELM can be expressed in terms of two factors: (i) how well a target can be approximated in terms of the Pauli features ϕ_k , and (ii) how well these

features are transmitted through the reservoir (see Sec. IV D). Notably, even features obstructed by the reservoir channel can be reconstructed from statistical correlations with visible features. This provides a more fine-grained perspective on the different contributions to the information processing capacity [54] and the training error beyond universal approximation theorems [6,27–31]. In fact, learning of the flow map of a dynamical system benefits from a feature library matching the structure of the vector field [70]. For the encodings used here, the terms in the flow map can be approximated on a restricted domain from the encoded features, which are Fourier-like or of other forms (see Sec. II B). Applying the present insights to construct reduced-order models in cases

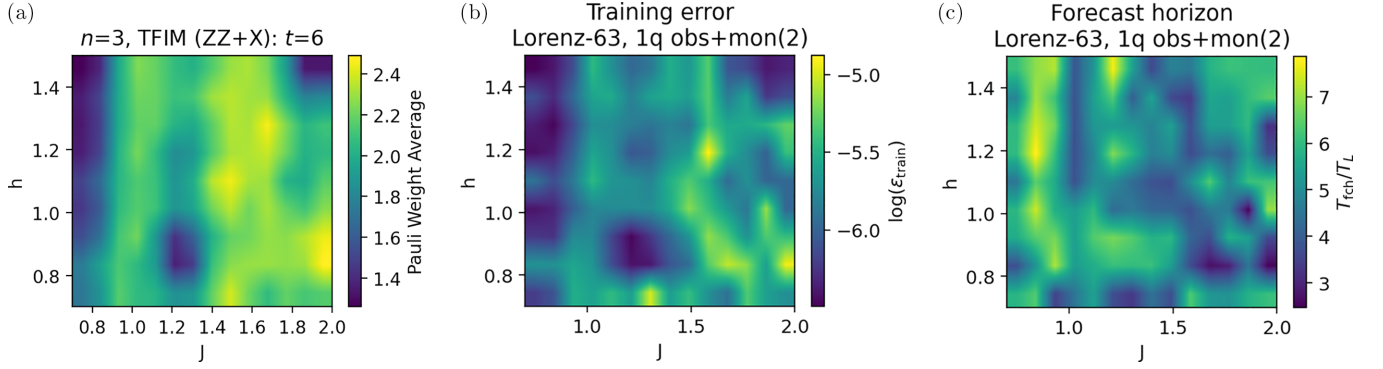


FIG. 16. The Pauli weight average (a) of Heisenberg-evolved measurement observables can predict the training error (b) and forecast horizon (c) of a QELM without temporal multiplexing. We use a TFIM Hamiltonian [Eq. (2.5a), varying coupling parameters J, h , fixed $t = 6$] with $n = 3$ qubits and amplitude encoding (2.8). The QELM in (b) and (c) is trained on the Lorenz-63 model and employs 1-qubit Pauli-Z observables in the readout, enhanced by monomials up to degree 2 to increase sensitivity to nonlinear terms. The Pauli weight average is data independent.

where the exact data-generating process is not known (black-box system identification) is an interesting avenue for future work [97,110,111].

The present work emphasizes Hamiltonian structure as an indicator for QRC performance (see also, e.g., Refs. [44,50,51,96,112]): specifically, for temporal multiplexing, greater reservoir randomness is advantageous under restricted measurement schemes, while highly structured Hamiltonians can lead to low expressivity. We have explained these findings in terms of the underlying Hamiltonian symmetries, which partition the reachable Krylov space (see Appendix G), and provided a theoretical analysis of the effective PTM. As a specific application, we provided asymptotic rank estimates for the transverse field Ising chain based on the Majorana formalism (see Appendix F).

Numerous previous studies noted that simply enhancing the quantum nature of a QML model often does not improve its performance on practical datasets [113–119]. The present study adds further perspectives on this finding: First, as discussed above, the specific mathematical structure of the features generated by a quantum encoding is not necessarily optimal for approximating a classical target function. For instance, the target function may not be exactly representable from the linear span of the features generated by the encoding even before any transformation by a quantum channel (ideal decodability score $\mathcal{R}_{\mathbb{H}}^2 < 1$). This is the classic representability problem in QML [51,120–122]. Second, a reservoir unitary mixes features and reversing this potentially detrimental effect requires complex measurement schemes (e.g., high-order Pauli observables, optimized premeasurement unitaries, or temporal multiplexing schemes). This insight is supported here by comparing the Pauli weight average of single-qubit observables to its training accuracy: as illustrated in Fig. 16, for a minimal QELM without temporal multiplexing trained on a time-series prediction task, large Pauli weight average strongly correlates with high training error and short forecast horizon. Consistent with Refs. [47,74], what matters for task performance is therefore not just how a quantum state evolves, but what part of the Heisenberg-Krylov operator space is actually accessible under the chosen observables and temporal multiplexing. The decodability framework presented here

allows one to separate the impact of the reservoir from that of the encoding on the nonlinear processing capacity of the model (see Appendix C).

The 4^n -dimensional space of Pauli features [123] is analogous to the exponential frequency content of QML models that employ gate-based rotational encoding [51,52,72]. These and similar properties stemming from the exponential Hilbert space dimensionality point to possible quantum advantages, but require specific measurement strategies. Moreover, simply increasing qubit number n does not generally improve QRC performance, as ill-conditioning and exponential concentration effects can become prevalent [12,48,50,53]. In that regime, more sophisticated reservoir design strategies will be required to steer feature propagation through the quantum layers. Finally, using the Pauli basis in the transfer matrix definition is justified by the measurement process, but not the only possibility. For future work, it could thus be interesting to study other orthogonal operator sets and assess the generality of the present findings.

ACKNOWLEDGMENT

This project was made possible by the DLR Quantum Computing Initiative and the Federal Ministry for Research, Technology and Space [124].

DATA AVAILABILITY

The data and code are available from the authors upon reasonable request.

APPENDIX A: KERNEL REPRESENTATION

Training over a Hilbert-Schmidt orthonormal basis of measurement operators $\{M_k\}_{k=1}^{d^2}$ results in a quantum-kernel form of the QELM [33]:

$$f(\mathbf{x}) = \sum_{j=1}^P \alpha_j^* K(\mathbf{x}, \mathbf{x}_j) = \text{tr}(M^* \rho(\mathbf{x})),$$

$$K(\mathbf{x}, \mathbf{x}') = \text{tr}(\rho(\mathbf{x})\rho(\mathbf{x}')). \quad (\text{A1})$$

TABLE V. Summary of notation used in this work. A trained quantity is indicated by a hat (e.g., $\hat{\mathbf{w}}$).

Symbol	Meaning	Defined in
$\mathcal{H}$	Hilbert space	Sec. II
$\text{Herm}(\mathcal{H})$	Space of Hermitian operators on $\mathcal{H}$	Sec. II
n	Number of qubits	Sec. II
d	Hilbert space dimension ($d = 2^n$)	Sec. II
D	Input data dimension	Sec. II
P	Number of training samples	Eq. (2.2)
m	Number of measurement operators	Eq. (2.1)
L	Number of temporal multiplexing steps	Eq. (3.9)
B	Total number of measurements (e.g., $B = mL$)	Eq. (3.14)
q	Number of nonzero Pauli features ($q \leq d^2$)	Eq. (3.14)
$\mathbf{x}, \mathbf{u}$	Input data ($\mathbf{x}, \mathbf{u} \in \mathbb{R}^D$)	Eqs. (2.1) and (3.14)
y	Target value ($y \in \mathbb{R}$)	Eq. (2.2)
$\rho(\mathbf{x})$	Density matrix of the quantum state (feature map)	Eq. (2.1)
M_k	Measurement operators	Eq. (2.1)
P_k	n -qubit Pauli operators ($k = 0, \dots, d^2 - 1$)	Eq. (3.2)
σ^a	Single-site Pauli matrices ($\sigma^0 \equiv \mathbb{I}, X, Y, Z$)	Eq. (3.11)
U_R, U_{enc}	Reservoir and encoding unitary operators	Eq. (2.4)
$T_{\mathcal{E}}$	Pauli transfer matrix (channel $\mathcal{E}$, $T_{\mathcal{E}} \in \mathbb{R}^{d^2 \times d^2}$)	Eq. (3.3)
V	Unitary PTM ($V \in \text{SO}(d^2)$)	Eq. (3.8)
R	Effective PTM (temporal multiplexing) ($R \in \mathbb{R}^{B \times q}$)	Eq. (3.9)
$\mathcal{S}$	Selected observable set ($ \mathcal{S} = m$)	Eq. (3.4)
S	Selector matrix corresponding to $\mathcal{S}$	Eq. (3.4)
$\Phi(\mathbf{x})$	Pauli feature vector ($\Phi(\mathbf{x}) \in \mathbb{R}^q$)	Eq. (3.2)
$\mathbf{F}(\mathbf{x})$	Readout feature vector ($\mathbf{F}(\mathbf{x}) = R\Phi(\mathbf{x})$)	Eq. (3.14)
$\mathbf{w}$	Readout weights ($\mathbf{w} \in \mathbb{R}^B$)	Eq. (3.14)
λ	Regularization parameter	Eq. (2.2)
$f(\mathbf{x})$	QELM output function ($f(\mathbf{x}) = \mathbf{w}^\top \mathbf{F}(\mathbf{x})$)	Eq. (2.1)
ν	Pauli weight	Eq. (4.5)
γ_r	Geometric channel decodability score	Eq. (4.10)
κ_j	Geometric encoding decodability score	Eq. (4.15)
$\mathcal{R}_R^2$	Information processing capacity	Eqs. (4.13) and (4.16)
$\mathcal{R}_{\mathbb{I}}^2$	Information processing capacity for an identity reservoir	Eq. (4.16)
τ_R	Channel transmission score	Eq. (4.17)
ν_R	Nullspace load	Eq. (4.17)
χ_R	Statistical rescue score	Eq. (4.17)
α	Dynamical system rescaling parameter	Eqs. (5.6) and (5.7)

For the least-squares loss, the dual parameters are $\alpha^* = (\mathbf{K} + \lambda \mathbb{I}_P)^{-1} \mathbf{y}$, where $K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$, and the optimal measurement operator is

$$M^* = \sum_{j=1}^P \alpha_j^* \rho(\mathbf{x}_j). \quad (\text{A2})$$

Compared with the equivalent primal formulation in Eq. (2.1), this representation is computationally more efficient when $P \ll d^2$.

If the reservoir is a data-independent unitary U_R , then the resulting state can be written as [see Eq. (2.7)] $\rho_U(\mathbf{x}) = U_R \rho_{\text{enc}}(\mathbf{x}) U_R^\dagger$. For a complete measurement operator basis, unitary invariance of the Hilbert-Schmidt inner product gives

$$K(\mathbf{x}, \mathbf{x}') = \text{tr}(\rho_U(\mathbf{x}) \rho_U(\mathbf{x}')) = \text{tr}(\rho_{\text{enc}}(\mathbf{x}) \rho_{\text{enc}}(\mathbf{x}')), \quad (\text{A3})$$

which is the usual fidelity kernel. Thus the Gram matrix, dual coefficients, and predictions are independent of U_R , whereas the optimal observable transforms covariantly as

$M_U^* = U_R M_{\text{enc}}^* U_R^\dagger$, with M_{enc}^* obtained from Eq. (A2) using ρ_{enc} .

Consider now a restricted Pauli observable set $\mathcal{S}$. In this case, the kernel induced by the primal model is generally not the full fidelity kernel. To see this, define the Hilbert-Schmidt projector onto the measured operator subspace by

$$\Pi_{\mathcal{S}}(\rho) = \frac{1}{d} \sum_{k \in \mathcal{S}} \text{tr}(P_k \rho) P_k. \quad (\text{A4})$$

Using the normalized Pauli basis $P_k/\sqrt{d}$, the corresponding kernel is

$$\begin{aligned} K_{\mathcal{S}, U}(\mathbf{x}, \mathbf{x}') &= \frac{1}{d} \sum_{k \in \mathcal{S}} \text{tr}(P_k \rho_U(\mathbf{x})) \text{tr}(P_k \rho_U(\mathbf{x}')) \\ &= \text{tr}[\Pi_{\mathcal{S}}(\rho_U(\mathbf{x})) \Pi_{\mathcal{S}}(\rho_U(\mathbf{x}'))]. \end{aligned} \quad (\text{A5})$$

The reservoir unitary drops out of this restricted kernel if conjugation preserves the measured subspace,

$$U_R^\dagger \mathcal{V}_{\mathcal{S}} U_R = \mathcal{V}_{\mathcal{S}}, \quad \mathcal{V}_{\mathcal{S}} = \text{span}\{P_l : l \in \mathcal{S}\}, \quad (\text{A6})$$

which follows from cyclicity of the trace, $\text{tr}(P_k U_R \rho_{\text{enc}} U_R^\dagger) = \text{tr}(Q \rho_{\text{enc}})$ with some $Q \in \mathcal{V}_S$. In this case, the restricted feature vector changes only by an orthogonal transformation, so ridge regression results in the same Gram matrix and dual solution. This condition holds, for instance, for certain Clifford unitaries that map $\mathcal{S}$ onto itself up to signs (such as CNOT circuits when $\mathcal{S}$ comprises all Z strings). However, a generic Pauli subset $\mathcal{S}$ does not satisfy Eq. (A6) and the corresponding restricted primal QELM is then represented by the kernel (A5), which still depends on the reservoir unitary U_R .

APPENDIX B: DECODING REFORMULATION OF LINEAR REGRESSION

Let $\Phi \in \mathbb{R}^q$ be a vector and $R \in \mathbb{R}^{B \times q}$ be a fixed mixing matrix, and $\mathbf{e}_r \in \mathbb{R}^q$ the r th standard basis vector. We ask for the reconstruction of a single component $\phi_r = \mathbf{e}_r^\top \Phi$ by finding weights $\mathbf{w}_r \in \mathbb{R}^B$ such that

$$\phi_r \simeq \mathbf{w}_r^\top R \Phi \quad (\text{B1})$$

to within a certain error tolerance. We discuss below situations where equality holds and where reconstruction can only be approximate. We remark that related ideas are pursued in compressed sensing [110].

In general, this problem is solved by ordinary least squares (OLS) regression. Collecting P centered training samples as $\Phi = [\Phi_1, \dots, \Phi_P] \in \mathbb{R}^{q \times P}$, the corresponding target vector is $\mathbf{e}_r^\top \Phi$. The OLS problem

$$\hat{\mathbf{w}}_r = \underset{\mathbf{w} \in \mathbb{R}^B}{\text{argmin}} \frac{1}{P} \|\mathbf{e}_r^\top \Phi - \mathbf{w}^\top R \Phi\|_2^2 \quad (\text{B2})$$

then has the well-known minimum-Euclidean-norm solution

$$\hat{\mathbf{w}}_r = (R \Sigma R^\top)^+ R \Sigma \mathbf{e}_r, \quad \Sigma = \frac{1}{P} \Phi \Phi^\top, \quad (\text{B3})$$

with Σ the empirical covariance matrix of the training data. If the data are not centered, then an intercept can be included, which takes the form $b_r = \bar{\phi}_r - \mathbf{w}_r^\top R \bar{\Phi}$. We now discuss how the properties of R and the data covariance Σ determine the exactness of the reconstruction of ϕ_r .

1. Geometric decodability

Exact geometric reconstruction of the single component ϕ_r is possible if $\mathbf{e}_r \in \text{im}(R^\top)$ or, equivalently, if there is a weight vector $R^\top \mathbf{w}_r = \mathbf{e}_r$. This, in turn, is equivalent to

$$\mathbf{e}_r^\top \mathbf{v} = 0 \quad \text{for every } \mathbf{v} \in \text{null}(R), \quad (\text{B4})$$

i.e., every vector of the nullspace of R must have zero r th component. If this condition is met, then the minimum-norm solution for the weights is provided by

$$\mathbf{w}_r^* = (R^+)^{\top} \mathbf{e}_r. \quad (\text{B5})$$

However, the solution need not be unique, since any term in $\text{null}(R^\top)$ can be added, so that $\mathbf{w}_r = (R^+)^{\top} \mathbf{e}_r + (\mathbb{I}_B - R R^+) \mathbf{u}$, with $\mathbf{u} \in \mathbb{R}^B$. The pseudoinverse singles out the solution with minimum Euclidean norm. Geometric decodability is thus a property of the mixing matrix R alone.

In certain special cases, the test (B4) can be circumvented:

a. Square and invertible mixing matrix

The first case is an invertible R with $B = q$, and positive definite Σ . Then $(R \Sigma R^\top)^{-1} R \Sigma = R^{-\top}$ and the minimum-norm solution is

$$\mathbf{w}_r^* = R^{-\top} \mathbf{e}_r, \quad (\text{B6})$$

so that ϕ_r can be exactly reconstructed, $\hat{\phi}_r = \mathbf{e}_r^\top R^{-1} R \Phi = \phi_r$.

b. Tall mixing matrix with full column rank

In the case $B \geq q$ and $\text{rank}(R) = q$, the mapping $\Phi \mapsto R \Phi$ is injective, and every component can be reconstructed exactly. The minimum-norm weights are

$$\mathbf{w}_r^* = (R^+)^{\top} \mathbf{e}_r = R(R^\top R)^{-1} \mathbf{e}_r, \quad (\text{B7})$$

which fulfill $\mathbf{w}_r^{*\top} R \Phi = \phi_r$.

2. Statistical decodability

When R is rank deficient ($\text{rank}(R) < q$) and Eq. (B4) is not satisfied, a feature ϕ_r may still be decodable using statistical correlations. To this end, recall that, from Eq. (B2), the minimum mean-squared error is $\mathcal{E}_r^{\min} = \min_{\mathbf{w}} (\mathbf{e}_r - R^\top \mathbf{w})^\top \Sigma (\mathbf{e}_r - R^\top \mathbf{w})$. Since Σ is positive semidefinite, this error is zero only if the residual vector $\mathbf{e}_r - R^\top \mathbf{w}$ lies in $\text{null}(\Sigma)$ for some $\mathbf{w}$ or, equivalently, if $\mathbf{e}_r \in \text{im}(R^\top) + \text{null}(\Sigma)$. This, in turn, is equivalent to

$$\mathbf{e}_r^\top \mathbf{v} = 0 \quad \text{for every } \mathbf{v} \in \text{null}(R) \cap \text{im}(\Sigma), \quad (\text{B8})$$

which says that two statistically admissible inputs producing the same mixed vector may differ in unobserved directions, but such differences must not alter the component ϕ_r . Exact recovery through correlations requires an exact, lower-dimensional relation in the data, while strong but imperfect correlations reduce the error but do not make it vanish. We analyze now a number of specific cases.

a. Isotropic data

For isotropic data, $\Sigma = \sigma_\phi^2 \mathbb{I}_q$, the minimum-norm solution (B3) reduces to

$$\mathbf{w}_r^* = (R R^\top)^+ R \mathbf{e}_r = (R^+)^{\top} \mathbf{e}_r. \quad (\text{B9})$$

The resulting predictor is $\hat{\phi}_r = \mathbf{e}_r^\top R^+ R \Phi$, wherein the component of Φ in $\text{null}(R)$ is set to zero.

b. Positive definite covariance matrix

If Σ is positive definite, then $\text{null}(\Sigma) = \{\mathbf{0}\}$, and the statistical-support criterion reduces to Eq. (B4), i.e., $\mathbf{e}_r \in \text{im}(R^\top)$, with corresponding minimum-norm weights Eq. (B5).

c. Correlated data

If $\mathbf{e}_r \notin \text{im}(R^\top)$, then part of ϕ_r depends on directions that are not directly observable through R , but may still be linearly predictable from second-order correlations. In order to quantify the effect, consider the compact singular value decomposition [$s = \text{rank}(R)$],

$$R = U_s D_s V_s^\top, \quad D_s = \text{diag}(d_1, \dots, d_s), \quad (\text{B10})$$

and complete V_s to an orthogonal basis, $V = [V_s \ V_0]$, where the columns of V_0 span $\text{null}(R)$. This allows us to decompose the input into observable and unobservable coordinates,

$$\mathbf{a} = V_s^\top \Phi, \quad \mathbf{b} = V_0^\top \Phi, \quad (\text{B11})$$

so that

$$\Phi = V_s \mathbf{a} + V_0 \mathbf{b}, \quad R\Phi = U_s D_s \mathbf{a}. \quad (\text{B12})$$

Thus, the mixed vector $R\Phi$ determines only $\mathbf{a}$ exactly. In this basis, the target component can be written as

$$\phi_r = \mathbf{v}_a^\top \mathbf{a} + \mathbf{v}_b^\top \mathbf{b}, \quad \mathbf{v}_a \equiv V_s^\top \mathbf{e}_r, \quad \mathbf{v}_b \equiv V_0^\top \mathbf{e}_r \quad (\text{B13})$$

while the covariance matrix takes the form

$$V^\top \Sigma V = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}. \quad (\text{B14})$$

Using the SVD (B10) and assuming that Σ_{aa} is invertible, the first term in the weight solution (B3) results as

$$\begin{aligned} (R\Sigma R^\top)^+ &= (U_s D_s V_s^\top \Sigma V_s D_s U_s^\top)^+ = (U_s D_s \Sigma_{aa} D_s U_s^\top)^+ \\ &= U_s D_s^{-1} \Sigma_{aa}^{-1} D_s^{-1} U_s^\top. \end{aligned} \quad (\text{B15})$$

Since $\mathbf{e}_r = V_s \mathbf{v}_a + V_0 \mathbf{v}_b$, the second term in (B3) reads

$$R\Sigma \mathbf{e}_r = U_s D_s V_s^\top \Sigma (V_s \mathbf{v}_a + V_0 \mathbf{v}_b) = U_s D_s (\Sigma_{aa} \mathbf{v}_a + \Sigma_{ab} \mathbf{v}_b). \quad (\text{B16})$$

Altogether, one obtains

$$\mathbf{w}_r^* = U_s D_s^{-1} (\mathbf{v}_a + \Sigma_{aa}^{-1} \Sigma_{ab} \mathbf{v}_b), \quad (\text{B17})$$

which results in the predictor

$$\hat{\phi}_r = \mathbf{v}_a^\top \mathbf{a} + \mathbf{v}_b^\top \Sigma_{ba} \Sigma_{aa}^{-1} \mathbf{a}. \quad (\text{B18})$$

The first term directly reconstructs the observable contribution to ϕ_r , while the second term predicts its unobservable contribution using correlations between $\mathbf{a}$ and $\mathbf{b}$. The residual error variance is given by

$$\begin{aligned} \mathcal{E}_r &= \mathbb{E}[(\phi_r - \hat{\phi}_r)^2] = \mathbb{E}[(\mathbf{v}_b^\top (\mathbf{b} - \Sigma_{ba} \Sigma_{aa}^{-1} \mathbf{a}))^2] \\ &= \mathbf{v}_b^\top (\Sigma_{bb} - \Sigma_{ba} \Sigma_{aa}^{-1} \Sigma_{ab}) \mathbf{v}_b, \end{aligned} \quad (\text{B19})$$

where the matrix in parentheses is the residual covariance of the unobserved coordinates given the observed ones. If $\mathbf{v}_b = \mathbf{0}$, then the target is directly observable and no statistical prediction is needed. If $\mathbf{v}_b \neq \mathbf{0}$ but $\Sigma_{ab} \mathbf{v}_b = \mathbf{0}$, then correlations provide no improvement. Otherwise, correlations reduce the reconstruction error by predicting the target-relevant part of the unobserved coordinates.

d. Simple example: How correlations help

We give a simple example that shows how correlations help reconstruct targets not directly decodable from R . Consider $q = 2, B = 1$, and

$$R = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad (\text{B20})$$

so that the observed variable is $R\Phi = \phi_1$. Suppose the goal is to reconstruct ϕ_2 , and let the data covariance be

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}. \quad (\text{B21})$$

The optimal weight and predictor follow as

$$w_2^* = \rho \frac{\sigma_2}{\sigma_1}, \quad \hat{\phi}_2 = \rho \frac{\sigma_2}{\sigma_1} \phi_1. \quad (\text{B22})$$

Thus, without correlation ($\rho = 0$), ϕ_1 provides no linear information about ϕ_2 . As $|\rho|$ approaches one, the reconstruction error $\mathcal{E}_2 = \mathbb{E}[(\phi_2 - \hat{\phi}_2)^2] = \sigma_2^2(1 - \rho^2)$ decreases. For $|\rho| = 1$, the two components satisfy an exact linear relation and ϕ_2 can be reconstructed perfectly, even though R is rank deficient.

APPENDIX C: GENERAL FORMALISM OF DECODABILITY SCORES

We now develop a formalism of decodability scores based on the least-squares approach of Appendix B. Appendices C 1 and C 2 closely follow the information processing capacity formulation in Ref. [54]. Going beyond this, the subsequent sections show how the decodability scores factorize into quantum encoding and channel parts, and how to distinguish statistical from geometric contributions. We retain the centered finite-sample setting of Appendix B and the notation of Sec. III C, including the following definitions:

$$\begin{aligned} \Phi &= [\Phi_1, \dots, \Phi_P] \in \mathbb{R}^{q \times P}, \quad R \in \mathbb{R}^{B \times q}, \\ G &= R\Phi \in \mathbb{R}^{B \times P}, \quad \Sigma = \frac{1}{P} \Phi \Phi^\top, \quad \mathbf{y} = (y(\mathbf{u}_1), \dots, y(\mathbf{u}_P))^\top, \\ \mathbf{b}_y &= \frac{1}{P} \Phi \mathbf{y}, \quad s_y^2 = \frac{1}{P} \mathbf{y}^\top \mathbf{y} > 0, \end{aligned} \quad (\text{C1})$$

where y is a centered scalar target. An intercept can be included before centering, as described in Appendix B. Unless stated otherwise, the capacities below refer to the unregularized least-squares method.

1. General decodability score

The minimum-norm OLS weights and fitted target are

$$\begin{aligned} \hat{\mathbf{w}}_y &= (R\Sigma R^\top)^+ R \mathbf{b}_y, \quad \hat{\mathbf{y}}_R = G^\top \hat{\mathbf{w}}_y = \Pi_G \mathbf{y}, \\ \Pi_G &\equiv G^\top (G G^\top)^+ G, \end{aligned} \quad (\text{C2})$$

where Π_G projects targets onto the subspace of functions representable by the measured QELM features $G = R\Phi$ [i.e., the row space $\text{row}(G) \subseteq \mathbb{R}^P$]. We define the coefficient of determination $\mathcal{R}^2$ as the *general decodability score*

$$\begin{aligned} \mathcal{D}_R[y] &\equiv 1 - \frac{\min_{\mathbf{w}} P^{-1} \|\mathbf{y} - G^\top \mathbf{w}\|_2^2}{s_y^2} = \frac{\mathbf{y}^\top \Pi_G \mathbf{y}}{\mathbf{y}^\top \mathbf{y}} \\ &= \frac{\mathbf{b}_y^\top R^\top (R\Sigma R^\top)^+ R \mathbf{b}_y}{s_y^2}, \end{aligned} \quad (\text{C3})$$

which corresponds to the information processing capacity introduced in Ref. [54]. The second equality follows because the OLS residual is orthogonal to the fitted vector. The last equality follows from $G G^\top = P R \Sigma R^\top$ and $G \mathbf{y} = P R \mathbf{b}_y$. Consequently, we have $0 \leq \mathcal{D}_R[y] \leq 1$ and

$$\mathcal{D}_R[y] = 1 \iff \mathbf{y} \in \text{row}(G). \quad (\text{C4})$$

2. Feature-direction, statistical, and geometric scores

For a feature-space direction $\mathbf{a} \in \mathbb{R}^q$, consider the target $y_{\mathbf{a}}(\mathbf{u}_p) = \mathbf{a}^\top \boldsymbol{\Phi}_p$, with variance $s_{\mathbf{a}}^2 = \mathbf{a}^\top \Sigma \mathbf{a}$. If $\mathbf{a} \notin \text{null}(\Sigma)$, then Eq. (C3) gives

$$\mathcal{D}_{R,\Sigma}(\mathbf{a}) \equiv \mathcal{D}_R[y_{\mathbf{a}}] = \frac{\mathbf{a}^\top \Sigma R^\top (R \Sigma R^\top)^+ R \Sigma \mathbf{a}}{\mathbf{a}^\top \Sigma \mathbf{a}}. \quad (\text{C5})$$

It is convenient to introduce the whitened quantities

$$\tilde{R} = R \Sigma^{1/2}, \quad \tilde{\mathbf{a}} = \Sigma^{1/2} \mathbf{a}, \quad \Pi_{\tilde{R}} = \tilde{R}^+ \tilde{R}, \quad (\text{C6})$$

in terms of which Eq. (C5) takes the form

$$\mathcal{D}_{R,\Sigma}(\mathbf{a}) = \frac{\tilde{\mathbf{a}}^\top \Pi_{\tilde{R}} \tilde{\mathbf{a}}}{\tilde{\mathbf{a}}^\top \tilde{\mathbf{a}}} = 1 - \frac{\|(\mathbb{I} - \Pi_{\tilde{R}}) \tilde{\mathbf{a}}\|_2^2}{\|\tilde{\mathbf{a}}\|_2^2}, \quad (\text{C7})$$

where $\Sigma^{1/2}$ is the symmetric square root. Note that the formulas remain valid for singular Σ , with whitening understood on $\text{im}(\Sigma)$.

Since $\Pi_{\tilde{R}}$ is an orthogonal projector, exact and vanishing decodability are characterized by

$$\begin{aligned} \mathcal{D}_{R,\Sigma}(\mathbf{a}) = 1 &\iff \mathbf{a} \in \text{row}(R) + \text{null}(\Sigma), \\ \mathcal{D}_{R,\Sigma}(\mathbf{a}) = 0 &\iff R \Sigma \mathbf{a} = 0. \end{aligned} \quad (\text{C8})$$

For $\mathbf{a} = \mathbf{e}_r$, the first relation reproduces the statistical recovery criterion of Eq. (B8). The corresponding single-feature statistical score is

$$\mathcal{D}_{R,\Sigma}(\mathbf{e}_r) = \frac{\mathbf{e}_r^\top \Sigma R^\top (R \Sigma R^\top)^+ R \Sigma \mathbf{e}_r}{\Sigma_{rr}} = \mathcal{R}_R^2[\phi_r]. \quad (\text{C9})$$

This quantity can be nonzero even when the direction $\mathbf{e}_r$ is absent from $\text{row}(R)$, because correlations encoded in Σ can predict the unresolved component.

For isotropic data $\Sigma = \sigma_\phi^2 \mathbb{I}$, we obtain the geometric decodability score:

$$\mathcal{D}_{R,\sigma_\phi^2 \mathbb{I}}(\mathbf{e}_r) = \mathbf{e}_r^\top R^+ R \mathbf{e}_r \equiv \gamma_r^2, \quad (\text{C10})$$

consistent with the OLS analysis in Appendix B 1. Moreover, $1 - \gamma_r^2 = \|(\mathbb{I} - R^+ R) \mathbf{e}_r\|_2^2$ is the squared coefficient-space projection error, and

$$\sum_{r=1}^q \gamma_r^2 = \text{tr}(R^+ R) = \text{rank}(R). \quad (\text{C11})$$

If Σ is nonsingular (or after restriction to its support), then the single-coordinate scores in whitened feature coordinates are

$$\tilde{\gamma}_r^2 \equiv (\Pi_{\tilde{R}})_{rr} = \mathcal{D}_{\tilde{R},\mathbb{I}}(\tilde{\mathbf{e}}_r), \quad \sum_r \tilde{\gamma}_r^2 = \text{rank}(R \Sigma^{1/2}), \quad (\text{C12})$$

which generalizes Eq. (C10).

The tensor-product statement used in the main text follows directly from $(\bigotimes_i R_i)^+ = \bigotimes_i R_i^+$. If $R = \bigotimes_{i=1}^n R_i$ and $\mathbf{e}_r = \bigotimes_{i=1}^n \mathbf{e}_{r_i}$, then

$$\begin{aligned} R^+ R &= \bigotimes_{i=1}^n (R_i^+ R_i), \quad \gamma_r^2 = \prod_{i=1}^n \gamma_{r_i,i}^2, \\ \mathbf{w}_r^* &= (R^+)^T \mathbf{e}_r = \bigotimes_{i=1}^n (R_i^+)^T \mathbf{e}_{r_i}. \end{aligned} \quad (\text{C13})$$

More generally, if also $\Sigma = \bigotimes_i \Sigma_i$ and $\mathbf{a} = \bigotimes_i \mathbf{a}_i$, then the statistical score factorizes as

$$\mathcal{D}_{\bigotimes_i R_i, \bigotimes_i \Sigma_i} \left(\bigotimes_i \mathbf{a}_i \right) = \prod_i \mathcal{D}_{R_i, \Sigma_i}(\mathbf{a}_i). \quad (\text{C14})$$

3. Representation-transmission factorization

Consider the orthogonal projector Π_Φ from the target space onto the subspace representable by the feature library Φ (unrestricted by measurement or reservoir mixing):

$$\Pi_\Phi = \Phi^\top (\Phi \Phi^\top)^+ \Phi, \quad \hat{\mathbf{y}}_\phi = \Pi_\Phi \mathbf{y}, \quad \mathbf{a}_y = \Sigma^+ \mathbf{b}_y, \quad (\text{C15})$$

so that $\hat{\mathbf{y}}_\phi = \Phi^\top \mathbf{a}_y$ and $\Sigma \mathbf{a}_y = \mathbf{b}_y$. Here $\mathbf{a}_y$ is the minimum-norm feature-space representative of the target projection. The ideal feature-representation score, corresponding to a QELM with $R = \mathbb{I}$ (see Sec. IV C 1), is

$$\mathcal{R}_\mathbb{I}^2[y] \equiv \mathcal{D}_\mathbb{I}[y] = \frac{\|\hat{\mathbf{y}}_\phi\|_2^2}{\|\mathbf{y}\|_2^2} = \frac{\mathbf{b}_y^\top \Sigma^+ \mathbf{b}_y}{s_y^2} = \frac{\mathbf{a}_y^\top \Sigma \mathbf{a}_y}{s_y^2}. \quad (\text{C16})$$

This quantity measures the fraction of the target variance that can be represented by the encoding-generated feature library, independently of subsequent reservoir or measurement action.

For $\hat{\mathbf{y}}_\phi \neq \mathbf{0}$, we define the *channel-transmission* score

$$\begin{aligned} \tau_R[y] &\equiv \frac{\|\Pi_G \hat{\mathbf{y}}_\phi\|_2^2}{\|\hat{\mathbf{y}}_\phi\|_2^2} = \mathcal{D}_{R,\Sigma}(\mathbf{a}_y) \\ &= \frac{\mathbf{b}_y^\top R^\top (R \Sigma R^\top)^+ R \mathbf{b}_y}{\mathbf{b}_y^\top \Sigma^+ \mathbf{b}_y} = \frac{\tilde{\mathbf{b}}_y^\top \Pi_{\tilde{R}} \tilde{\mathbf{b}}_y}{\tilde{\mathbf{b}}_y^\top \tilde{\mathbf{b}}_y}, \end{aligned} \quad (\text{C17})$$

where $\tilde{\mathbf{b}}_y \equiv (\Sigma^+)^{1/2} \mathbf{b}_y$, $\tilde{R} = R \Sigma^{1/2}$, and $\Pi_{\tilde{R}} = \tilde{R}^+ \tilde{R}$. If $\hat{\mathbf{y}}_\phi = \mathbf{0}$, then $\mathcal{R}_R^2[y] = 0$ for every R , and the channel factor is immaterial. The general score factorizes as

$$\mathcal{D}_R[y] = \mathcal{R}_R^2[y] = \mathcal{R}_\mathbb{I}^2[y] \tau_R[y] = \mathcal{D}_\mathbb{I}[y] \mathcal{D}_{R,\Sigma}(\mathbf{a}_y). \quad (\text{C18})$$

This follows from $\|\Pi_G \mathbf{y}\|_2^2 = \|\Pi_G \Pi_\Phi \mathbf{y}\|_2^2 = \tau_R[y] \|\Pi_\Phi \mathbf{y}\|_2^2$, since $\text{row}(G) \subseteq \text{row}(\Phi)$ and hence the fitted component can first be projected onto the feature library and then onto the readout subspace. Equivalently, the result is obtained by substituting $\Sigma \mathbf{a}_y = \mathbf{b}_y$ into Eq. (C5).

4. Geometric and statistical contributions to the channel-transmission

We now split $\tau_R[y]$ [Eq. (C17)] into a geometric and a statistical contribution. Let

$$R = U_s D_s V_s^\top \quad (\text{C19})$$

be the compact singular value decomposition of R , with $s = \text{rank}(R)$. The columns of $V_s \in \mathbb{R}^{q \times s}$ span $\text{row}(R)$, while the columns of V_0 span $\text{null}(R)$. Writing $V = [V_s \ V_0]$, we decompose the feature vector and the ideal target direction as

$$\mathbf{z} = V_s^\top \boldsymbol{\Phi}, \quad \boldsymbol{\eta} = V_0^\top \boldsymbol{\Phi}, \quad \mathbf{a}_\parallel = V_s^\top \mathbf{a}_y, \quad \mathbf{a}_\perp = V_0^\top \mathbf{a}_y. \quad (\text{C20})$$

Since $R \boldsymbol{\Phi} = U_s D_s \mathbf{z}$, the coordinates $\mathbf{z}$ are geometrically accessible from the readout. By contrast, the coordinates $\boldsymbol{\eta}$ lie in $\text{null}(R)$ and are not geometrically accessible.

The ideal feature representative decomposes as

$$\hat{\mathbf{y}}_\phi(\mathbf{u}) = \mathbf{a}_\parallel^\top \mathbf{z}(\mathbf{u}) + \underbrace{\mathbf{a}_\perp^\top \boldsymbol{\eta}(\mathbf{u})}_{h_y(\mathbf{u})}. \quad (\text{C21})$$

The first term is directly transmitted by R . The second term,

$$h_y(\mathbf{u}) = \mathbf{a}_\perp^\top V_0^\top \boldsymbol{\Phi}(\mathbf{u}), \quad (\text{C22})$$

is geometrically hidden from the readout. It may nevertheless be statistically predictable from the accessible coordinates $\mathbf{z}$. To quantify this effect, let

$$V^\top \Sigma V = \begin{pmatrix} \Sigma_{\parallel\parallel} & \Sigma_{\parallel\perp} \\ \Sigma_{\perp\parallel} & \Sigma_{\perp\perp} \end{pmatrix}. \quad (\text{C23})$$

The variance of the hidden component is then

$$\text{var}(h_y) = \mathbf{a}_\perp^\top \Sigma_{\perp\perp} \mathbf{a}_\perp, \quad (\text{C24})$$

while the part of h_y linearly predictable from the accessible coordinates $\mathbf{z}$ has variance

$$\text{var}(\Pi_z h_y) = \mathbf{a}_\perp^\top \Sigma_{\perp\parallel} \Sigma_{\parallel\parallel}^+ \Sigma_{\parallel\perp} \mathbf{a}_\perp, \quad (\text{C25})$$

where Π_z denotes the projection onto the span of the rowspace coordinates $\mathbf{z}$. This motivates the *statistical rescue* score

$$\chi_R[y] \equiv \begin{cases} \frac{\mathbf{a}_\perp^\top \Sigma_{\perp\parallel} \Sigma_{\parallel\parallel}^+ \Sigma_{\parallel\perp} \mathbf{a}_\perp}{\mathbf{a}_\perp^\top \Sigma_{\perp\perp} \mathbf{a}_\perp}, & \mathbf{a}_\perp^\top \Sigma_{\perp\perp} \mathbf{a}_\perp > 0, \\ 0, & \mathbf{a}_\perp^\top \Sigma_{\perp\perp} \mathbf{a}_\perp = 0, \end{cases} \quad (\text{C26})$$

which fulfills $0 \leq \chi_R[y] \leq 1$. A value $\chi_R[y] \simeq 0$ means that the geometrically hidden part of the target-relevant feature representative is also not statistically predictable from the measured rowspace. By contrast, $\chi_R[y] \simeq 1$ indicates that the hidden component is almost fully recoverable from correlations in the data.

However, $\chi_R[y] \simeq 1$ does not exclude that the decodability character of a task is actually geometric, which is the case if almost no task-relevant feature lies in $\text{null}(R)$. To quantify how much of a feature lies in geometrically hidden directions, we define the *nullspace load*

$$\nu_R[y] \equiv \frac{\mathbf{a}_\perp^\top \Sigma_{\perp\perp} \mathbf{a}_\perp}{\mathbf{a}_y^\top \Sigma \mathbf{a}_y}. \quad (\text{C27})$$

In general, since $\Sigma_{\parallel\perp}$ need not vanish, $\nu_R[y]$ can exceed unity and should be understood as a nullspace-load diagnostic. By contrast, the product $\nu_R[y](1 - \chi_R[y]) \in [0, 1]$ entering Eq. (C30) is a proper fraction. We now collect the coordinates defined in Eq. (C20) over the training set and obtain coordinate matrices $Z = V_s^\top \Phi$ and $N = V_0^\top \Phi$, so that

$$\hat{\mathbf{y}}_\phi = Z^\top \mathbf{a}_\parallel + N^\top \mathbf{a}_\perp. \quad (\text{C28})$$

Since $G = R\Phi = U_s D_s Z$, we have $\Pi_G = \Pi_Z$, where $\Pi_Z = Z^\top (ZZ^\top)^+ Z$. The visible component $Z^\top \mathbf{a}_\parallel$ already lies in $\text{row}(Z)$, and therefore the only channel error comes from the part of the hidden component not predictable from Z :

$$\frac{1}{P} \|\hat{\mathbf{y}}_\phi - \Pi_G \hat{\mathbf{y}}_\phi\|_2^2 = \mathbf{a}_\perp^\top (\Sigma_{\perp\perp} - \Sigma_{\perp\parallel} \Sigma_{\parallel\parallel}^+ \Sigma_{\parallel\perp}) \mathbf{a}_\perp. \quad (\text{C29})$$

Dividing by $\|\hat{\mathbf{y}}_\phi\|_2^2/P = \mathbf{a}_y^\top \Sigma \mathbf{a}_y$ yields the refined transmission score

$$\tau_R[y] = 1 - \nu_R[y](1 - \chi_R[y]). \quad (\text{C30})$$

For a fixed target, full channel transmission thus occurs if the geometrically hidden component is either absent or exactly predictable from the accessible rowspace:

$$\begin{aligned} \tau_R[y] = 1 &\iff (\mathbb{I} - \Pi_Z) N^\top \mathbf{a}_\perp = \mathbf{0} \\ &\iff \nu_R[y](1 - \chi_R[y]) = 0. \end{aligned} \quad (\text{C31})$$

Full transmission for every target is equivalent to

$$\text{null}(R) \cap \text{im}(\Sigma) = \{\mathbf{0}\} \iff \text{rank}(R\Sigma^{1/2}) = \text{rank}(\Sigma). \quad (\text{C32})$$

Thus full column rank of R is sufficient but not necessary, since it is enough that R be injective on the populated feature subspace.

Upon inserting Eq. (C30) into Eq. (C18), the decodability score takes the form

$$\mathcal{R}_R^2[y] = \mathcal{R}_\parallel^2[y][1 - \nu_R[y](1 - \chi_R[y])]. \quad (\text{C33})$$

Equivalently, the normalized target variance can be decomposed as

$$\begin{aligned} 1 &= \underbrace{1 - \mathcal{R}_\parallel^2[y]}_{\text{not represented by encoding}} \\ &+ \underbrace{\mathcal{R}_\parallel^2[y] \nu_R[y](1 - \chi_R[y])}_{\text{geometrically hidden and not statistically rescued}} \\ &+ \underbrace{\mathcal{R}_R^2[y]}_{\text{decoded}}. \end{aligned} \quad (\text{C34})$$

The channel-transmission deficit $1 - \tau_R[y]$ is the amount of target-relevant feature weight lying outside $\text{row}(R)$ that is *not* statistically rescued by correlations. The quantity

$$\begin{aligned} &\mathcal{R}_\parallel^2[y] \nu_R[y] \chi_R[y] \\ &(\text{geometrically hidden but statistically rescued}) \end{aligned} \quad (\text{C35})$$

measures the reduction in normalized mean-squared error relative to predicting with only the directly visible component.

For a nonsingular Σ , or in an orthonormal basis of its support, write $\mathbf{b}_y/\|\mathbf{b}_y\|_2 = \sum_{r \in \mathcal{J}} c_r \mathbf{e}_r$, with $\sum_r c_r^2 = 1$. The whitened single-direction scores of Eq. (C12) bound the transmission loss according to

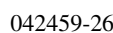
$$1 - \tau_R[y] \leq \left(\sum_{r \in \mathcal{J}} |c_r| \sqrt{1 - \tilde{\gamma}_r^2} \right)^2 \leq \sum_{r \in \mathcal{J}} (1 - \tilde{\gamma}_r^2), \quad (\text{C36})$$

which follows from the triangle inequality applied to $(\mathbb{I} - \Pi_{\tilde{R}}) \sum_r c_r \mathbf{e}_r$ and Cauchy-Schwarz.

5. Specializations to nonlinear-capacity and monomial scores

For the degree- k target family $\{y_e : |\mathbf{e}| = k\}$ used in Eq. (4.13), the nonlinear processing capacity is simply the average of the general score,

$$\mathcal{R}_R^2(k) = \frac{1}{N_k} \sum_{|\mathbf{e}|=k} \mathcal{D}_R[y_e]. \quad (\text{C37})$$



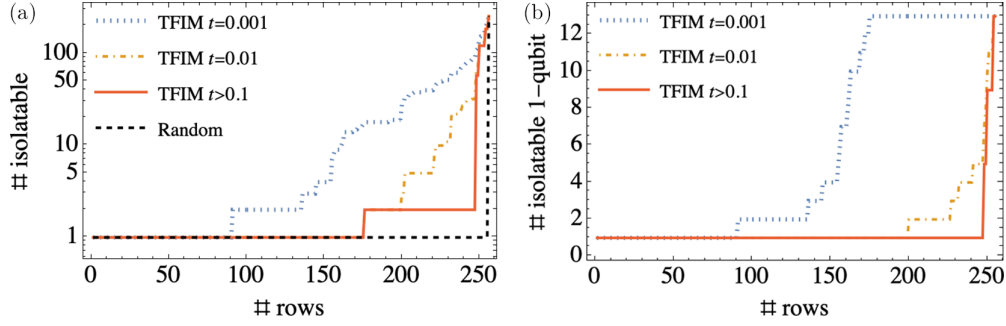


FIG. 18. Geometric decodability according to the null-space criterion (B4): Number of exactly reconstructible features as a function of the number of observables (=rows of the PTM) m , for the TFIM with $n = 4$ qubits ($d^2 = 256$) at various evolution times t . Panel (a) shows the total number of reconstructible features, including a random-unitary reference, while panel (b) shows the number of reconstructible one-qubit features. In contrast to statistical decodability, geometric decodability is an exact criterion independent of the data (see Appendix B).

is directly reflected by the scores. In fact, for the amplitude encoding (2.9), the temporally multiplexed Z-readout satisfies $\mathbf{F}(-\mathbf{u}) = \mathbf{F}(\mathbf{u})$, because $[P_Z, H_{\text{TFIM}}^{\text{xx-z}}] = 0$ and $\rho_{\text{enc}}(-\mathbf{u}) = P_Z \rho_{\text{enc}}(\mathbf{u}) P_Z$ with $P_Z = Z_1 \cdots Z_n$. Hence, under a symmetric input distribution on $[-a, a]$ ($a < 1$), every odd-degree monomial is orthogonal to the readout space and thus has vanishing capacity, although they are available under full Pauli readout ($R = \mathbb{I}$). This is not necessarily the case for other encodings, such as the variant (2.8), for which higher-order monomials have a high statistical decodability.

Note that, while $0 \leq \chi_R[y] \leq 1$, the nullspace load $\nu_R[y]$ can exceed 1. Hence, the quantity (C35) can in principle exceed $\mathcal{R}_R^2$: it is the error reduction when using correlations relative to using only directly visible components for predictions. We therefore show in the plots (also in Fig. 11) a clipped and scaled version $\min(\mathcal{R}_R^2 \nu_R \chi_R, \mathcal{R}_R^2) / \mathcal{R}_R^2$.

APPENDIX D: GEOMETRIC DECODABILITY FOR COMMON PTMS

We analyze here the geometric decodability criterion (B4) for common PTMs. Consider the case where $R \in \mathbb{R}^{m \times q}$ is dense and *quasirandom*, arising, e.g., for a Haar-random unitary U . For $m < q$, the null space is necessarily nonzero and in random orientation. Accordingly, the probability that the same component vanishes for all vectors of $\text{null}(R)$ is essentially zero, such that no features are geometrically decodable. Once $m \geq q$, R has full column rank, hence $\text{null}(R) = \{0\}$ and the null-space condition in Eq. (B4) is trivially fulfilled, implying that all q features become instantly reconstructible for a random R . In this case one may define $W = R^+ = (R^\top R)^{-1} R^\top$ and assign $\mathbf{w}^\top = \mathbf{e}_r^\top W$.

For a *Clifford* unitary, which maps between elements of the Pauli group, $\text{row}(R)$ is spanned by a subset of standard basis vectors, hence $\text{rank}(R)$ features are exactly reconstructible even if R is wide. For unitaries based on *structured Hamiltonians* or at small evolution times, one may expect Eq. (B4) to be fulfilled for a certain set of features, provided m is sufficiently large (but can still be $m \ll q$). This is confirmed numerically in Fig. 18, where the behavior of the feature isolation criterion (B4) is illustrated for various unitaries as a function of the number of observables m (rows of R).

APPENDIX E: PAULI TRANSFER MATRIX FORMALISM: FURTHER DETAILS

We provide here some further useful information on the PTM formalism for QELMs. When ordering the basis with identity first, we have $\Phi = (\phi_0, \mathbf{s})$, where $\phi_0 = \text{tr}(\rho) = 1$ and $\mathbf{s} \in \mathbb{R}^{d^2-1}$ is the *traceless (Bloch)* part. A maximally mixed state has $\mathbf{s} = 0$, i.e., is in the center of the generalized Bloch sphere. By contrast, a pure state [$\text{tr}(\rho^2) = 1$] has $|\mathbf{s}|^2 = d - 1$ and thus lies on the hypersphere of radius $\sqrt{d - 1}$ (although not all points on it are valid quantum states).

1. Unitary channels

For a unitary channel $\mathcal{U}(\rho) = U \rho U^\dagger$, one has

$$T_{U, \alpha\beta} = \frac{1}{2^n} \text{tr}(P_\alpha U P_\beta U^\dagger). \quad (\text{E1})$$

Trace preservation and unitality imply the form

$$T_U = \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \Omega_U \end{pmatrix}, \quad (\text{E2})$$

i.e., the traceless coefficients undergo a real orthogonal rotation with $\Omega_U \in \text{SO}(d^2 - 1)$. For Clifford unitaries, T_U is a signed permutation of Pauli strings.

Consider a *separable* reservoir Hamiltonian $H_R = \sum_{j=1}^n h_R^{(j)}$, such that the reservoir unitary factorizes as $U(t) = \bigotimes_{j=1}^n U_j(t)$. Such a structure emerges, e.g., for a TFIM with $h/J \rightarrow \infty$ [Eq. (2.5)]. This implies a Kronecker-product structure of the unitary PTM (E1) (and similarly for any other product quantum channel),

$$T_U(t) = V^{(1)}(t) \otimes \cdots \otimes V^{(n)}(t), \quad (\text{E3})$$

where each $V^{(j)}(t) \in \text{SO}(4)$ acts on the single-qubit feature vector $\Phi^{(j)}$. Consequently, the evolved features remain factorized:

$$\begin{aligned} \Phi'(\mathbf{x}) &= T_U(t) \Phi(\mathbf{x}) \\ &= (V^{(1)}(t) \Phi^{(1)}(x_1)) \otimes \cdots \otimes (V^{(n)}(t) \Phi^{(n)}(x_n)). \end{aligned} \quad (\text{E4})$$

Thus, each Pauli-string feature $\phi_{a_1 \dots a_n}$ evolves only by local mixing among the single-qubit labels $a_j \in \{0, x, y, z\}$, without coupling different qubits and without operator spreading.

Each local unitary $U_k(t) = e^{-i t h_k}$ rotates the local Pauli components, so Y strings can become nonzero only through

this local mixing. More explicitly, writing the local Pauli feature vector as $\tilde{\Phi}^{(k)} \equiv (\phi_x^{(k)}, \phi_y^{(k)}, \phi_z^{(k)})^\top$, any single-qubit unitary induces an $\text{SO}(3)$ rotation

$$\tilde{\Phi}'^{(k)}(u_k) = \Omega^{(k)}(t) \tilde{\Phi}^{(k)}(u_k), \quad \phi_0'^{(k)} = \phi_0^{(k)} = 1, \quad (\text{E5})$$

i.e., $\phi_a'^{(k)} = \sum_{b \in \{x,y,z\}} \Omega_{ab}^{(k)}(t) \phi_b^{(k)}$, with $\Omega^{(k)}$ the adjoint representation of $U_k(t)$ on the Pauli vector. For example, for a local Z -field $h_k = Z$,

$$U_k(t)^\dagger X U_k(t) = \cos(2t) X - \sin(2t) Y, \\ U_k(t)^\dagger Y U_k(t) = \cos(2t) Y + \sin(2t) X, \quad (\text{E6})$$

so that $\phi_y'^{(k)}(u_k) = \sin(2t) \phi_x^{(k)}(u_k)$ and, correspondingly, $\phi'_{\dots y k \dots}(\mathbf{x}) = \sin(2t) \phi_{\dots x k \dots}(\mathbf{x})$, while Z features remain invariant. This shows how unitaries act to make specific features accessible to the readout when the latter uses a restricted set of observables.

2. General quantum channels: Affine action on the Bloch part

We discuss now briefly the representation of general noisy channels in the PTM formalism, including some geometrical intuitions [62,125]. Consider a CPTP channel $\mathcal{E}$ with associated Kraus operators $\{K_\ell\}$. Then

$$T_{\mathcal{E},\alpha\beta} = \frac{1}{2^n} \sum_\ell \text{tr}(P_\alpha K_\ell P_\beta K_\ell^\dagger), \quad (\text{E7})$$

which corresponds in block form (compatible with $\Phi = (1, \mathbf{s})$) to

$$T_{\mathcal{E}} = \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{t} & \Omega \end{pmatrix}, \quad \mathbf{s}' = \mathbf{t} + \Omega \mathbf{s}, \quad (\text{E8})$$

where $\mathbf{t} = \mathbf{0}$ for unital channels. For a single qubit, Ω is a contracting map (singular values ≤ 1). Nonunitarity produces a translation $\mathbf{t}$ of the Bloch part, e.g., a maximally mixed state would be shifted away from the center of the Bloch sphere.

Consider the following single-qubit PTMs for nonunitary noise channels [126–128]:

(1) Isotropic depolarization $\mathcal{D}_\lambda(\rho) = \lambda\rho + (1-\lambda)\mathbb{I}/2$:

$$T_{\text{dep}}(\lambda) = \text{diag}(1, \lambda, \lambda, \lambda). \quad (\text{E9})$$

For a global isotropic depolarizing channel on dimension $d = 2^n$, $T_{\text{dep}}^{(n)}(\lambda) = \text{diag}(1, \lambda \mathbb{I}_{d^2-1})$.

(2) Z -dephasing with probability p :

$$T_{\text{dephZ}}(p) = \text{diag}(1, 1-2p, 1-2p, 1). \quad (\text{E10})$$

(3) Amplitude damping with parameter $\gamma \in [0, 1]$ and polarization $z_* \in [-1, 1]$ (with $z_* = 1$ corresponding to the zero-temperature fixed point $|0\rangle\langle 0|$):

$$T_{\text{AD}}(\gamma) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-\gamma} & 0 & 0 \\ 0 & 0 & \sqrt{1-\gamma} & 0 \\ z_*\gamma & 0 & 0 & 1-\gamma \end{pmatrix}. \quad (\text{E11})$$

Modeling n -qubit noise as tensor products of single-qubit channels $\mathcal{E}^{(n)} = \bigotimes_{i=1}^n \mathcal{E}^{(i)}$ is often used as a baseline in QRC models [19,22,49,129]. Since the PTMs (E9) and (E10) are Pauli diagonal, they do not lead to Pauli spreading and hence do not enlarge the accessible feature subspace. Amplitude

damping [Eq. (E11)] produces Pauli- I/Z mixing, which causes operator weight decay.

A particularly transparent characterization emerges if the noise channel is approximately diagonal. For instance, for the n -qubit depolarizing channel $\mathcal{D}^{(n)} = \bigotimes_{i=1}^n \mathcal{D}_\lambda$ [see Eq. (E9)], any Pauli string P of weight $v(P)$ is an eigenoperator with

$$\mathcal{D}^{(n)}(P) = \lambda^{v(P)} P. \quad (\text{E12})$$

Accordingly, high-weight Pauli strings are exponentially suppressed by near-diagonal noise channels. In conjunction with unitary-induced operator spreading, this can increase information loss under finite sampling. On the other hand, amplitude damping noise is often observed to improve QRC performance as the associated superoperator spreads the state's probability weight into more Pauli directions, thereby enriching the feature map [19]. This has been related to the “coherence influx” present for amplitude damping but not depolarization [46].

APPENDIX F: TFIM SYMMETRIES AND THEIR IMPACT ON TEMPORAL MULTIPLEXING

We now specialize the general discussion of Sec. IV A to the TFIM reservoirs in Eq. (2.5) and to the observable families used in the numerics. An analytical discussion is possible based on the Jordan-Wigner transformation and the associated Majorana representation for the Ising chain [130,131]. Once the Lieb-Robinson light cone has covered the chain, the support of an initially local operator has saturated. The accessible operator subspace controls the achievable rank of the temporally multiplexed PTM.

1. Heisenberg sectors of the relevant observables

Introducing the Majorana operators obtained from the Jordan-Wigner transformation

$$\gamma_{2j-1} = \left(\prod_{m<j} Z_m \right) X_j, \quad \gamma_{2j} = \left(\prod_{m<j} Z_m \right) Y_j, \quad j = 1, \dots, n, \quad (\text{F1})$$

the TFIM in Eq. (2.5b) can be written as

$$H_{\text{TFIM}}^{\text{xx-z}} = iJ \sum_{j=1}^{n-1} \gamma_{2j} \gamma_{2j+1} + i h \sum_{j=1}^n \gamma_{2j-1} \gamma_{2j}. \quad (\text{F2})$$

Since the Hamiltonian is quadratic in the Majoranas, Heisenberg evolution acts linearly on them,

$$\gamma_a^{(H)}(t) = \sum_{b=1}^{2n} \mathcal{O}_{ab}(t) \gamma_b, \quad \mathcal{O}(t) \in \text{SO}(2n), \quad (\text{F3})$$

and therefore preserves the *Majorana degree* r , i.e., the number of Majorana factors in a monomial. The degree- r operator sector

$$\mathcal{V}_r = \text{span} \{ \gamma_{a_1} \cdots \gamma_{a_r} : 1 \leq a_1 < \cdots < a_r \leq 2n \} \quad (\text{F4})$$

with dimension

$$\dim \mathcal{V}_r = \binom{2n}{r}, \quad (\text{F5})$$

is thus invariant.

The observables relevant for the QELM fall into simple sectors:

$$Z_j = -i\gamma_{2j-1}\gamma_{2j} \in \mathcal{V}_2, \quad Z_i Z_j \in \mathcal{V}_4 \quad (i \neq j), \quad (\text{F6})$$

while

$$\begin{aligned} X_j &= \left[\prod_{m < j} (-i\gamma_{2m-1}\gamma_{2m}) \right] \gamma_{2j-1} \in \mathcal{V}_{2j-1}, \\ Y_j &= \left[\prod_{m < j} (-i\gamma_{2m-1}\gamma_{2m}) \right] \gamma_{2j} \in \mathcal{V}_{2j-1}. \end{aligned} \quad (\text{F7})$$

Thus Z -type observables under $H_{\text{TFIM}}^{\text{xx-z}}$ stay in low-degree Gaussian sectors, whereas X_j and Y_j occupy odd sectors whose degree grows with the site index. In particular, a bulk X_j or Y_j lives in a sector of dimension $\binom{2n}{2j-1} \sim \mathcal{O}(4^n)$.

These sectors have a highly structured Pauli-string content. From the bilinear form of Z_j , its Heisenberg evolution under $H_{\text{TFIM}}^{\text{xx-z}}$ is restricted to

$$\begin{aligned} Z_j^{(\text{H})}(t) &= \sum_{p=1}^n a_p(t) Z_p \\ &+ \sum_{1 \leq p < q \leq n} \sum_{\alpha, \beta \in \{X, Y\}} b_{pq}^{\alpha\beta}(t) \alpha_p Z_{p+1} \cdots Z_{q-1} \beta_q, \end{aligned} \quad (\text{F8})$$

i.e., to strings with two X/Y end points and a Z string in between. Hence, after full spatial spreading the number of Pauli strings remains $\sim \mathcal{O}(n^2)$. Likewise, $(Z_i Z_j)^{(\text{H})}(t)$ remains in the quartic sector and becomes a sum of products of two such interval strings. These strings have at most four X/Y end points connected by Z segments, and their total number is $\sim \mathcal{O}(n^4)$.

The second TFIM in Eq. (2.5a) follows from the Hadamard equivalence (2.6):

$$O^{(\text{H})}(t)|_{\text{zz-x}} = U_H^\dagger [(U_H O U_H^\dagger)^{(\text{H})}(t)]|_{\text{xx-z}} U_H. \quad (\text{F9})$$

Therefore Z -type observables under $H_{\text{TFIM}}^{\text{zz-x}}$ inherit the same operator-space complexity as X -type observables under $H_{\text{TFIM}}^{\text{xx-z}}$:

$$Z_j \longleftrightarrow X_j \in \mathcal{V}_{2j-1}, \quad Z_i Z_j \longleftrightarrow X_i X_j \in \mathcal{V}_{2|i-j|} \quad (i \neq j). \quad (\text{F10})$$

Hence $Z_j^{(\text{H})}(t)$ under $H_{\text{TFIM}}^{\text{zz-x}}$ can occupy exponentially large sectors for bulk j , and $(Z_i Z_j)^{(\text{H})}(t)$ does the same for separations $|i - j| \approx n/2$. Z -type observables are low-complexity under $H_{\text{TFIM}}^{\text{xx-z}}$, but high-complexity under $H_{\text{TFIM}}^{\text{zz-x}}$. By contrast, the full single-site Pauli family $\{X_j, Y_j, Z_j\}_{j=1}^n$ already contains bulk odd-degree operators in either TFIM and therefore has exponential Heisenberg complexity for both variants.

We remark that the Majorana degree r of a Pauli observable is defined only after choosing a Jordan-Wigner (Majorana) representation. Here, we use the convention in which the corresponding TFIM Hamiltonian is quadratic in the Majorana operators, so that Heisenberg evolution acts linearly on the Majoranas and preserves the degree- r sectors.

2. Asymptotic rank of the temporally multiplexed PTM

We now connect the above Heisenberg-sector structure to the effective PTM in temporal multiplexing. Denote by R_L the effective PTM in Eq. (3.9) evaluated for L multiplexing steps and $\mathcal{L}(\cdot) = i[H, \cdot]$. For generic sampling times $\{t_\ell\}$, the large- L rank is the dimension of the Heisenberg orbit, equivalently of the Krylov span [47]:

$$\begin{aligned} r_\infty(H, \mathcal{S}) &:= \lim_{L \rightarrow \infty} \text{rank}(R_L) \\ &= \dim \text{span}\{e^{t\mathcal{L}}(P) : P \in \mathcal{S}, t \in \mathbb{R}\} \\ &= \dim \text{span}\{\mathcal{L}^m(P) : P \in \mathcal{S}, m \geq 0\}. \end{aligned} \quad (\text{F11})$$

This is the operator-space saturation rank before feature clipping. The actual effective PTM rank is therefore bounded by

$$\text{rank}(R_L) \leq \min(q, r_\infty(H, \mathcal{S})), \quad (\text{F12})$$

with q the effective number of nonzero encoded features from Sec. III C.

We consider the three observable families

$$\begin{aligned} \mathcal{S}_Z &= \{Z_j\}_{j=1}^n, \quad \mathcal{S}_{Z,ZZ} = \{Z_j, Z_i Z_j\}_{i,j=1}^n, \\ \mathcal{S}_{XYZ} &= \{X_j, Y_j, Z_j\}_{j=1}^n. \end{aligned} \quad (\text{F13})$$

For $\mathcal{S}_{Z,ZZ}$, ordered pairs do not increase the operator-space dimension since $Z_i Z_j = Z_j Z_i$, while the diagonal $i = j$ contributes only the trivial identity operator.

Using the sector decomposition from Appendix F 1, one obtains the following sector-dimension upper bounds for $H_{\text{TFIM}}^{\text{xx-z}}$:

$$r_\infty(H_{\text{TFIM}}^{\text{xx-z}}, \mathcal{S}_Z) \leq \binom{2n}{2} = n(2n-1) \sim \mathcal{O}(n^2), \quad (\text{F14})$$

$$r_\infty(H_{\text{TFIM}}^{\text{xx-z}}, \mathcal{S}_{Z,ZZ}) \leq 1 + \binom{2n}{2} + \binom{2n}{4} \sim \mathcal{O}(n^4). \quad (\text{F15})$$

The term 1 in the second line is the static identity contribution arising from $Z_i Z_i = \mathbb{I}$ and should be omitted if only $i < j$ is included in $\mathcal{S}_{Z,ZZ}$.

The sector dimensions are ceilings on the temporal Krylov rank and need not be saturated. In particular, for a fixed quadratic Hamiltonian the rank is further restricted by degeneracies of the Liouvillian frequencies. After Bogoliubov diagonalization, the TFIM can be written as $H = E_0 + \sum_{a=1}^n \epsilon_a (f_a^\dagger f_a - \frac{1}{2})$ and every Liouvillian frequency therefore has the form

$$\omega_{\mathbf{q}} = \sum_{a=1}^n q_a \epsilon_a, \quad q_a \in \{-1, 0, 1\}. \quad (\text{F16})$$

Consequently, the quadratic TFIM has at most $|\Omega| = 3^n$ distinct Liouvillian frequencies. With $m_{\mathcal{S}} := \dim \text{span } \mathcal{S}$, the exact frequency-resolved rank formula (G19) implies

$$r_\infty(H, \mathcal{S}) \leq m_{\mathcal{S}} 3^n. \quad (\text{F17})$$

An improved bound becomes available when the frequencies are nonresonant and the projections of the seeded operators onto the Liouvillian-frequency subspaces are nonzero. Here nonresonance means that distinct vectors $\mathbf{q} \in \{-1, 0, 1\}^n$ give distinct sums $\omega_{\mathbf{q}} = \sum_a q_a \epsilon_a$. In a Majorana-degree- r sector, the number of compatible distinct frequencies

is

$$N_r = \sum_{\substack{0 \leq s \leq \min(r, 2n-r) \\ s \equiv r \pmod{2}}} \binom{n}{s} 2^s, \quad (\text{F18})$$

where s is the number of modes with $q_a = \pm 1$. Under the Hadamard equivalence, the observables in $\mathcal{S}_Z$ for $H_{\text{TFIM}}^{\text{ZZ-X}}$ map to $X_j \in \mathcal{V}_{2j-1}$ for $H_{\text{TFIM}}^{\text{XX-Z}}$. Therefore,

$$\begin{aligned} r_\infty(H_{\text{TFIM}}^{\text{ZZ-X}}, \mathcal{S}_Z) &= \sum_{j=1}^n N_{2j-1} = \sum_{\substack{s=1 \\ s \text{ odd}}}^n \binom{n}{s} 2^s (n-s+1) \\ &= \frac{(n+3)3^{n-1} + (n-1)(-1)^n}{2} = \Theta(n3^n), \end{aligned} \quad (\text{F19})$$

where the notation $\Theta(\cdot)$ indicates a tight bound. For $H_{\text{TFIM}}^{\text{XX-Z}}$ one has $\mathcal{L}(X_j) = 2hY_j$ ($h \neq 0$), so including Y_j does not enlarge the Krylov space already generated by X_j . The Z_j observables lie in the separate degree-two sector and it follows that

$$\begin{aligned} r_\infty(H_{\text{TFIM}}^{\text{XX-Z}}, \mathcal{S}_{XYZ}) &= r_\infty(H_{\text{TFIM}}^{\text{ZZ-X}}, \mathcal{S}_Z) + r_\infty(H_{\text{TFIM}}^{\text{XX-Z}}, \mathcal{S}_Z) \\ &= \Theta(n3^n). \end{aligned} \quad (\text{F20})$$

Since the full single-site Pauli family is invariant under the global Hadamard transformation up to signs and relabeling,

$$r_\infty(H_{\text{TFIM}}^{\text{ZZ-X}}, \mathcal{S}_{XYZ}) = r_\infty(H_{\text{TFIM}}^{\text{XX-Z}}, \mathcal{S}_{XYZ}) = \Theta(n3^n). \quad (\text{F21})$$

Finally, the frequency bound gives

$$\begin{aligned} r_\infty(H_{\text{TFIM}}^{\text{ZZ-X}}, \mathcal{S}_Z) &\leq r_\infty(H_{\text{TFIM}}^{\text{ZZ-X}}, \mathcal{S}_{ZZZ}) \\ &\leq \left(1 + n + \binom{n}{2}\right) 3^n \leq \mathcal{O}(n^2 3^n). \end{aligned} \quad (\text{F22})$$

In summary, temporal multiplexing of Z-type observables under $H_{\text{TFIM}}^{\text{XX-Z}}$ explores only low-degree Majorana sectors and gives a polynomial rank. Under $H_{\text{TFIM}}^{\text{ZZ-X}}$, these observables touch high-degree sectors and the rank is exponential, since the growth is governed by the 3^n Liouvillian frequencies. The Majorana-sector dimension characterizes the possible Pauli support of a fully evolved operator. The temporal Krylov rank depends on the multiplicities of the Liouvillian frequencies and can be reduced by resonances or commensurate sampling times.

APPENDIX G: INFLUENCE OF GENERAL LIOUVILLIAN SYMMETRIES

In Appendix F we derived the asymptotic rank of the temporally multiplexed PTM for the TFIM by exploiting a special operator-space sector decomposition. We will generalize here these results in terms of symmetries of the Liouvillian $\mathcal{L}$. Although we phrase the discussion for unitary evolution generated by a Hamiltonian, the same sector-decomposition arguments apply to any linear Heisenberg generator $\mathcal{L}$.

1. Operator-space formulation of the effective PTM

Let $\mathcal{H}$ be a d -dimensional Hilbert space and let $\mathcal{O}$ be the associated operator space equipped with the Hilbert-Schmidt inner product $\langle A, B \rangle = \frac{1}{d} \text{tr}(A^\dagger B)$. A Hamiltonian H induces

the Heisenberg evolution $\mathcal{O}^{(H)}(t) = e^{t\mathcal{L}}\mathcal{O}$ with Liouvillian $\mathcal{L}(\mathcal{O}) = i[H, \mathcal{O}]$. Given a selected observable family $\mathcal{S}$, the “seed” space of measurement operators is

$$\mathcal{M} = \text{span}\{P_k : k \in \mathcal{S}\} \subseteq \mathcal{O}. \quad (\text{G1})$$

Likewise, let $\mathcal{F} \subseteq \mathcal{O}$ denote the feature subspace spanned by those Pauli operators whose input features are not identically zero under the chosen encoding, so that $\dim \mathcal{F} = q$ in the notation of Sec. III C. We write $\Pi_{\mathcal{F}}$ for the orthogonal projector onto $\mathcal{F}$. For temporal multiplexing at times $t_1, \dots, t_L$, define

$$\begin{aligned} \mathcal{W}_L(H, \mathcal{M}; \mathcal{F}) &= \text{span}\{\Pi_{\mathcal{F}} e^{t_\ell \mathcal{L}}(M) : \\ &M \in \mathcal{M}, \ell = 1, \dots, L\}. \end{aligned} \quad (\text{G2})$$

In any orthonormal operator basis whose first q basis elements span $\mathcal{F}$, the effective PTM R_L in Eq. (3.9) is the coordinate matrix of these vectors, hence

$$\text{rank}(R_L) = \dim \mathcal{W}_L(H, \mathcal{M}; \mathcal{F}). \quad (\text{G3})$$

In the absence of feature clipping, one may set $\mathcal{F} = \mathcal{O}$ and obtain the operator-space saturation rank

$$\begin{aligned} r_\infty(H, \mathcal{M}) &:= \lim_{L \rightarrow \infty} \text{rank}(R_L) \\ &= \dim \text{span}\{e^{t\mathcal{L}}(M) : M \in \mathcal{M}, t \in \mathbb{R}\}. \end{aligned} \quad (\text{G4})$$

For generic sampling times, this coincides with the Krylov dimension [Eq. (F11)],

$$r_\infty(H, \mathcal{M}) = \dim \text{span}\{\mathcal{L}^m(M) : M \in \mathcal{M}, m \geq 0\}. \quad (\text{G5})$$

2. Invariant sectors induced by symmetries

Assume a symmetry of the Liouvillian described by the superoperator $\Gamma : \mathcal{O} \rightarrow \mathcal{O}$ satisfying

$$[\Gamma, \mathcal{L}] = 0. \quad (\text{G6})$$

In the physically relevant cases of a unitary or Hermitian Γ , the spectral theorem yields $\Gamma = \sum_\alpha \lambda_\alpha \Pi_\alpha$, where the Π_α are mutually orthogonal spectral projectors. These render the orthogonal decomposition

$$\begin{aligned} \mathcal{O} &= \bigoplus_\alpha \mathcal{O}_\alpha, \quad \Pi_\alpha \mathcal{O} = \mathcal{O}_\alpha, \quad \Pi_\alpha \Pi_\beta = \delta_{\alpha\beta} \Pi_\alpha, \\ \sum_\alpha \Pi_\alpha &= \mathbb{I}_{\mathcal{O}}. \end{aligned} \quad (\text{G7})$$

Since Eq. (G6) implies

$$[\Pi_\alpha, \mathcal{L}] = 0 \quad \text{for all } \alpha, \quad (\text{G8})$$

every sector is invariant under Heisenberg evolution, i.e.,

$$\mathcal{L}(\mathcal{O}_\alpha) \subseteq \mathcal{O}_\alpha, \quad e^{t\mathcal{L}}\mathcal{O}_\alpha \subseteq \mathcal{O}_\alpha. \quad (\text{G9})$$

When both $\mathcal{M}$ and $\mathcal{F}$ are sector-compatible, i.e., the measured observables and the retained feature basis can each be chosen sector by sector, then

$$\mathcal{M} = \bigoplus_\alpha \mathcal{M}_\alpha, \quad \mathcal{F} = \bigoplus_\alpha \mathcal{F}_\alpha, \quad (\text{G10})$$

with $\mathcal{M}_\alpha = \Pi_\alpha \mathcal{M} \subseteq \mathcal{O}_\alpha$ and $\mathcal{F}_\alpha = \Pi_\alpha \mathcal{F} \subseteq \mathcal{O}_\alpha$. In that case the accessible space decomposes orthogonally,

$$\mathcal{W}_L(H, \mathcal{M}; \mathcal{F}) = \bigoplus_\alpha \mathcal{W}_{L,\alpha}(H, \mathcal{M}_\alpha; \mathcal{F}_\alpha), \quad (\text{G11})$$

where

$$\mathcal{W}_{L,\alpha}(H, \mathcal{M}_\alpha; \mathcal{F}_\alpha) = \text{span}\{\Pi_{\mathcal{F}_\alpha} e^{t_\ell \mathcal{L}}(M) : M \in \mathcal{M}_\alpha, \ell = 1, \dots, L\}. \quad (\text{G12})$$

Hence symmetries limit the accessible operator space by restricting the dynamics to the sectors touched by the measurement subspace. Consequently,

$$\text{rank}(R_L) = \sum_\alpha \text{rank}(R_{L,\alpha}), \quad (\text{G13})$$

with one effective PTM block $R_{L,\alpha}$ per active sector. Writing

$$r_{\infty,\alpha}(H, \mathcal{M}_\alpha) = \dim \text{span}\{e^{t\mathcal{L}}(M) : M \in \mathcal{M}_\alpha, t \in \mathbb{R}\}, \quad (\text{G14})$$

one has the sectorwise bound

$$\text{rank}(R_{L,\alpha}) \leq \min(L \dim \mathcal{M}_\alpha, \dim \mathcal{F}_\alpha, r_{\infty,\alpha}(H, \mathcal{M}_\alpha)). \quad (\text{G15})$$

Since $r_{\infty,\alpha}(H, \mathcal{M}_\alpha) \leq \dim \mathcal{O}_\alpha$, the symmetry-resolved PTM rank bound follows as

$$\begin{aligned} r_\infty(H, \mathcal{M}; \mathcal{F}) &:= \lim_{L \rightarrow \infty} \dim \mathcal{W}_L(H, \mathcal{M}; \mathcal{F}) \\ &\leq \sum_{\alpha: \mathcal{M}_\alpha \neq \emptyset} \min(\dim \mathcal{F}_\alpha, r_{\infty,\alpha}(H, \mathcal{M}_\alpha)) \\ &\leq \sum_{\alpha: \mathcal{M}_\alpha \neq \emptyset} \dim \mathcal{F}_\alpha. \end{aligned} \quad (\text{G16})$$

If the measured observables are not sector-compatible, then the above reasoning yields an upper bound after decomposing them into sector components, but the exact sum rule (G13) can fail because a single observable may contain linear combinations of several sectors.

3. Hamiltonian criteria for underexploration

The symmetry decomposition in Eq. (G16) determines only the maximal accessible dimension. Whether a populated sector is fully explored depends on the restricted generator $\mathcal{L}|_{\mathcal{O}_\alpha}$ and on the choice of seed subspace $\mathcal{M}_\alpha$. Let $\mathcal{O}_\alpha$ be a symmetry-compatible invariant sector and $\mathcal{L}_\alpha := \mathcal{L}|_{\mathcal{O}_\alpha}$. As above, we consider the spectral decomposition

$$\mathcal{L}_\alpha = \sum_{\omega \in \Omega_\alpha} i\omega \Pi_{\alpha,\omega}, \quad \mathcal{O}_{\alpha,\omega} := \Pi_{\alpha,\omega} \mathcal{O}_\alpha, \quad (\text{G17})$$

where $\Omega_\alpha \subset \mathbb{R}$ is the set of Bohr frequencies ω appearing in sector α , with multiplicity $\dim \mathcal{O}_{\alpha,\omega}$. The Bohr frequencies constitute the set $\{E_n - E_m : E_n, E_m \in \sigma(H)\}$ in terms of the eigenvalues E_n of the Hamiltonian. We choose now a basis $M_1, \dots, M_{\dim \mathcal{M}_\alpha}$ of $\mathcal{M}_\alpha$ and an orthonormal basis $\{F_{\alpha,\omega,s}\}_{s=1}^{\dim \mathcal{O}_{\alpha,\omega}}$ of each eigenspace $\mathcal{O}_{\alpha,\omega}$. Accordingly, each

observable M_k expands as

$$\Pi_{\alpha,\omega} M_k = \sum_{s=1}^{\dim \mathcal{O}_{\alpha,\omega}} (C_{\alpha,\omega})_{s,k} F_{\alpha,\omega,s}, \quad (\text{G18})$$

with the coefficient matrix $C_{\alpha,\omega} \in \mathbb{C}^{\dim \mathcal{O}_{\alpha,\omega} \times \dim \mathcal{M}_\alpha}$. Then the asymptotic rank contributed by sector α is exactly

$$r_{\infty,\alpha}(H, \mathcal{M}_\alpha) = \sum_{\omega \in \Omega_\alpha} \text{rank } C_{\alpha,\omega}. \quad (\text{G19})$$

This result shows that underexploration occurs because several operator directions share the same frequency [large $\dim \mathcal{O}_{\alpha,\omega}$ in Eq. (G18)], while the seed subspace supplies too few independent directions inside that degenerate eigenspace (small $\text{rank } C_{\alpha,\omega}$).

A direct consequence of Eq. (G19) is the general bound

$$r_{\infty,\alpha}(H, \mathcal{M}_\alpha) \leq \sum_{\omega \in \Omega_\alpha} \min(\dim \mathcal{O}_{\alpha,\omega}, \dim \mathcal{M}_\alpha). \quad (\text{G20})$$

For a generic ($\dim \mathcal{M}_\alpha$)-dimensional choice of seed subspace inside $\mathcal{O}_\alpha$, one expects equality in Eq. (G20). Hence the Hamiltonian alone determines the generic underexploration deficit

$$\begin{aligned} \dim \mathcal{O}_\alpha - \sum_{\omega \in \Omega_\alpha} \min(\dim \mathcal{O}_{\alpha,\omega}, \dim \mathcal{M}_\alpha) \\ = \sum_{\omega \in \Omega_\alpha} \max(\dim \mathcal{O}_{\alpha,\omega} - \dim \mathcal{M}_\alpha, 0). \end{aligned} \quad (\text{G21})$$

In particular, full sector saturation is generically possible only if the multiplicity of each Bohr frequency does not exceed the dimension of the seed subspace:

$$\dim \mathcal{O}_{\alpha,\omega} \leq \dim \mathcal{M}_\alpha \quad \text{for all } \omega \in \Omega_\alpha. \quad (\text{G22})$$

The above approach avoids the construction of Krylov spaces and instead relies on the spectral properties of the Hamiltonian or Liouvillian.

It is useful to note that operators diagonal in the energy eigenbasis belong to the $\omega = 0$ eigenspace. Hence, temporal multiplexing cannot increase their span and the asymptotic rank equals the dimension of the initially seeded diagonal subspace.

4. Single-observable limit

We now apply the above formalism to the Krylov grade of a single observable, recovering a key result in Ref. [47]. We set thus $\mathcal{F} = \mathcal{O}$ and choose a one-dimensional seed space $\mathcal{M} = \text{span}\{O\}$ for a single observable O . Then the Krylov grade M is the dimension of the minimal Liouvillian Krylov space generated by O [47],

$$M := \dim \text{span}\{O, \mathcal{L}(O), \mathcal{L}^2(O), \dots\} = r_\infty(H, \mathcal{M}). \quad (\text{G23})$$

Suppressing the symmetry label α for simplicity, write the Liouvillian spectral decomposition as

$$\mathcal{L} = \sum_{\omega \in \Omega} i\omega \Pi_\omega, \quad \Omega = \{\omega_P\}_{P=0}^{N_\omega-1}, \quad (\text{G24})$$

where $\omega_P = E_m - E_n$ with $E_m, E_n \in \sigma(H)$ are the pairwise distinct Bohr frequencies of H , so that $N_\omega := |\Omega|$ is the num-

ber of distinct transition frequencies. For the single seed O , define

$$\sigma_P := \Pi_{\omega_P} O, \quad (\text{G25})$$

so that $O(t) = \sum_{P=0}^{N_\omega-1} e^{i\omega_P t} \sigma_P$. Let

$$N_1 := |\{P \in \{0, \dots, N_\omega - 1\} : \sigma_P = 0\}| \quad (\text{G26})$$

be the number of vanishing frequency blocks. Since $\dim \mathcal{M} = 1$, each coefficient matrix C_{ω_P} in Eq. (G18) is a single column, and therefore

$$\text{rank } C_{\omega_P} = \begin{cases} 1, & \sigma_P = \Pi_{\omega_P} O \neq 0, \\ 0, & \sigma_P = \Pi_{\omega_P} O = 0. \end{cases} \quad (\text{G27})$$

Hence Eqs. (G19) and (G23) give

$$\begin{aligned} M = r_\infty(H, \mathcal{M}) &= \sum_{P=0}^{N_\omega-1} \text{rank } C_{\omega_P} \\ &= |\{P : \sigma_P \neq 0\}| = N_\omega - N_1, \end{aligned} \quad (\text{G28})$$

recovering the corresponding theorem in Ref. [47] as the single observable specialization of the present formalism.

5. Examples of symmetry-induced rank bounds

a. Hilbert-space symmetries

Consider an ordinary Hilbert-space symmetry with $UHU^\dagger = H$. Then the superoperator

$$\Gamma_U(O) = UOU^\dagger \quad (\text{G29})$$

is unitary on $\mathcal{O}$ and satisfies $[\Gamma_U, \mathcal{L}] = 0$. Similarly, for a $\mathbb{Z}_2$ symmetry with $U^2 = \mathbb{I}$, the corresponding spectral projectors are

$$\Pi_\pm = \frac{1}{2}(\mathbb{I}_\mathcal{O} \pm \Gamma_U). \quad (\text{G30})$$

If $Q = \sum_q qP_q$ is a conserved charge with $[H, Q] = 0$, then

$$\Pi_{qq'}(O) = P_q O P_{q'} \quad (\text{G31})$$

defines orthogonal projectors commuting with $\mathcal{L}$, which yields the block decomposition

$$\mathcal{O} = \bigoplus_{q,q'} P_q \mathcal{O} P_{q'}. \quad (\text{G32})$$

If the measured observables commute with Q , then only the diagonal sectors $\mathcal{O}_{qq}$ can be seeded, and therefore ($d_q = \text{rank } P_q$)

$$r_\infty(H, \mathcal{M}) \leq \sum_q d_q^2. \quad (\text{G33})$$

Thus a symmetry reduces the PTM rank because temporal multiplexing can only enlarge the Krylov span inside the sectors allowed by the symmetry and by the measured observables.

b. Majorana symmetries of the TFIM

The TFIM analysis in Appendix F is a direct instance of the above framework, but with a Liouville-space symmetry that is stronger than a conventional conserved charge. For the quadratic Majorana form of $H_{\text{TFIM}}^{\text{xx-z}}$ in Eq. (F2), the operator space decomposes as

$$\mathcal{O} = \bigoplus_{r=0}^{2n} \mathcal{V}_r, \quad \dim \mathcal{V}_r = \binom{2n}{r}, \quad (\text{G34})$$

which is the sector structure used in Appendix F1. Since the usual observable families are sector-compatible,

$$Z_j \in \mathcal{V}_2, \quad Z_i Z_{j \neq i} \in \mathcal{V}_4, \quad X_j, Y_j \in \mathcal{V}_{2j-1}, \quad (\text{G35})$$

the general bound (G16) reproduces the polynomial versus exponential behavior found in Appendix F2. In particular, for $H_{\text{TFIM}}^{\text{xx-z}}$, Z-type observables seed only the low-degree sectors $\mathcal{V}_2$ and $\mathcal{V}_4$, which yields polynomial asymptotic ranks, while inclusion of X_j or Y_j activates odd sectors whose dimensions are already exponential for bulk sites. By contrast, under the Hadamard-related Hamiltonian $H_{\text{TFIM}}^{\text{zz-x}}$, the same Z-type observables are mapped to the high-degree operators discussed around Eqs. (F9) and (F10).

6. Consequences for PTM analysis and feature decodability

In a symmetry-adapted orthonormal operator basis, the unitary PTM becomes block diagonal,

$$V(t) = \bigoplus_{\alpha} V_{\alpha}(t). \quad (\text{G36})$$

Under the sector-compatibility assumptions in Eq. (G10), the temporally multiplexed effective PTM (3.9) can be brought into a block form

$$R_L \sim \bigoplus_{\alpha} R_{L,\alpha}, \quad (\text{G37})$$

and the projector defined in Eq. (4.10) decomposes as

$$R_L^+ R_L \sim \bigoplus_{\alpha} (R_{L,\alpha}^+ R_{L,\alpha}). \quad (\text{G38})$$

Thus, if a feature lies in a sector with $\mathcal{M}_{\alpha} = 0$, then it is symmetry-forbidden and cannot be (geometrically) decoded by the chosen observable family, i.e., $\gamma_r^2 = (R_L^+ R_L)_{rr} = 0$. Note that, in the raw Pauli basis, symmetry-induced restrictions need not appear as perfectly sharp blocks, because the symmetry-adapted basis is generally different from the Pauli basis.

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

[2] A. Sakurai, M. P. Estarellas, W. J. Munro, and K. Nemoto, Quantum extreme reservoir computation utilizing scale-free networks, *Phys. Rev. Appl.* **17**, 064044 (2022).

- [3] M. Kornjača, H.-Y. Hu, C. Zhao, J. Wurtz, P. Weinberg, M. Hamdan, A. Zhdanov, S. H. Cantu, H. Zhou, R. A. Bravo, K. Bagnall, J. I. Basham, J. Campo, A. Choukri, R. DeAngelo, P. Frederick, D. Haines, J. Hammett, N. Hsu, M.-G. Hu, *et al.*, Large-scale quantum reservoir learning with an analog quantum computer, [arXiv:2407.02553](#).
- [4] A. Sakurai, A. Hayashi, W. J. Munro, and K. Nemoto, Simple Hamiltonian dynamics as a powerful resource for image classification, *Phys. Rev. A* **111**, 052432 (2025).
- [5] A. De Lorenzis, M. Casado, M. Estarellas, N. Lo Gullo, T. Lux, F. Plastina, A. Riera, and J. Settimo, Harnessing quantum extreme learning machines for image classification, *Phys. Rev. Appl.* **23**, 044024 (2025).
- [6] J. Chen, H. I. Nurdin, and N. Yamamoto, Temporal information processing on noisy quantum computers, *Phys. Rev. Appl.* **14**, 024065 (2020).
- [7] 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).
- [8] 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).
- [9] 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).
- [10] O. Ahmed, F. Tennie, and L. Magri, Optimal training of finitely sampled quantum reservoir computers for forecasting of chaotic dynamics, *Quantum Mach. Intell.* **7**, 31 (2025).
- [11] P. Pfeffer, F. Heyder, and J. Schumacher, Hybrid quantum-classical reservoir computing of thermal convection flow, *Phys. Rev. Res.* **4**, 033176 (2022).
- [12] P. Pfeffer, F. Heyder, and J. Schumacher, Reduced-order modeling of two-dimensional turbulent Rayleigh-Bénard flow by hybrid quantum-classical reservoir computing, *Phys. Rev. Res.* **5**, 043242 (2023).
- [13] 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).
- [14] 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).
- [15] D. Zia, L. Innocenti, G. Minati, S. Lorenzo, A. Suprano, R. Di Bartolo, N. Spagnolo, T. Giordani, V. Cimini, G. M. Palma, A. Ferraro, F. Sciarrino, and M. Paternostro, Quantum reservoir computing for photonic entanglement witnessing, *Sci. Adv.* **11**, eady7987 (2025).
- [16] H. Assil, A. El Allati, and G. L. Giorgi, Entanglement estimation of Werner states with a quantum extreme learning machine, *Phys. Rev. A* **111**, 022412 (2025).
- [17] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, *Nat. Commun.* **9**, 4812 (2018).
- [18] M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes, and M. Cerezo, Barren plateaus in variational quantum computing, *Nat. Rev. Phys.* **7**, 174 (2025).
- [19] L. Domingo, G. Carlo, and F. Borondo, Taking advantage of noise in quantum reservoir computing, *Sci. Rep.* **13**, 8790 (2023).
- [20] D. Fry, A. Deshmukh, S. Y.-C. Chen, V. Rastunkov, and V. Markov, Optimizing quantum noise-induced reservoir computing for nonlinear and chaotic time series prediction, *Sci. Rep.* **13**, 19326 (2023).
- [21] 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).
- [22] F. Monzani, E. Ricci, L. Nigro, and E. Prati, Non-unital noise in a superconducting quantum computer as a computational resource for reservoir computing, [arXiv:2409.07886](#).
- [23] 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.: Complexity* **2**, 045008 (2021).
- [24] J. Nokkala, R. Martínez-Peña, G. L. Giorgi, V. Parigi, M. C. Soriano, and R. Zambrini, Gaussian states of continuous-variable quantum systems provide universal and versatile reservoir computing, *Commun. Phys.* **4**, 53 (2021).
- [25] L. C. G. Govia, G. J. Ribeill, G. E. Rowlands, and T. A. Ohki, Nonlinear input transformations are ubiquitous in quantum reservoir computing, *Neuromorph. Comput. Eng.* **2**, 014008 (2022).
- [26] 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).
- [27] J. Chen and H. I. Nurdin, Learning nonlinear input-output maps with dissipative quantum systems, *Quantum Info. Proc.* **18**, 198 (2019).
- [28] T. Goto, Q. H. Tran, and K. Nakajima, Universal approximation property of quantum machine learning models in quantum-enhanced feature spaces, *Phys. Rev. Lett.* **127**, 090506 (2021).
- [29] R. Martínez-Peña and J.-P. Ortega, Quantum reservoir computing in finite dimensions, *Phys. Rev. E* **107**, 035306 (2023).
- [30] F. Monzani and E. Prati, Universality conditions of unified classical and quantum reservoir computing, *Neurocomputing* **643**, 130391 (2025).
- [31] L. Gonon and A. Jacquier, Universal approximation theorem and error bounds for quantum neural networks and quantum reservoirs, *IEEE Trans. Neural Netw. Learn. Syst.* **36**, 11355 (2025).
- [32] M. Schuld, Supervised quantum machine learning models are kernel methods, [arXiv:2101.11020](#).
- [33] M. Gross and H.-M. Rieser, Kernel-based optimization of measurement operators for quantum reservoir computers, [arXiv:2602.14677](#).
- [34] A. Canatar, E. Peters, C. Pehlevan, S. M. Wild, and R. Shaydulin, Bandwidth enables generalization in quantum kernel models, *Transactions on Machine Learning Research* (2023).
- [35] M. Gross, M. Lange, B. Bujnowski, and H.-M. Rieser, Expressivity vs. generalization in quantum kernel methods, in *33rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2025*, edited

- by M. Verleysen (Ciacio - i6doc.com, Bruges, Belgium, 2025), pp. 543–548.
- [36] M. Kempkes, A. Ijaz, E. Gil-Fuster, C. Bravo-Prieto, J. Spiegelberg, E. van Nieuwenburg, and V. Dunjko, Double descent in quantum kernel methods, *PRX Quantum* **7**, 010312 (2026).
 - [37] K. Fujii and K. Nakajima, Quantum reservoir computing: A reservoir approach toward quantum machine learning on near-term quantum devices, in *Reservoir Computing: Theory, Physical Implementations, and Applications*, Natural Computing Series, edited by K. Nakajima and I. Fischer (Springer, Singapore, 2021), pp. 423–450.
 - [38] P. Mujal, R. Martínez-Peña, J. Nokkala, J. García-Bení, G. L. Giorgi, M. C. Soriano, and R. Zambrini, Opportunities in quantum reservoir computing and extreme learning machines, *Adv. Quantum Technol.* **4**, 2100027 (2021).
 - [39] C. M. Wilson, J. S. Otterbach, N. Tezak, R. S. Smith, A. M. Polloreno, P. J. Karalekas, S. Heidel, M. S. Alam, G. E. Crooks, and M. P. da Silva, Quantum kitchen sinks: An algorithm for machine learning on near-term quantum computers, [arXiv:1806.08321](#).
 - [40] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, Extreme learning machine: Theory and applications, *Neurocomputing* **70**, 489 (2006).
 - [41] A. Rahimi and B. Recht, Random features for large-scale kernel machines, in *Advances in Neural Information Processing Systems* (Curran Associates, Red Hook, NY, 2007), Vol. 20, pp. 1177–1184.
 - [42] 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).
 - [43] L. Domingo, G. Carlo, and F. Borondo, Optimal quantum reservoir computing for the noisy intermediate-scale quantum era, *Phys. Rev. E* **106**, L043301 (2022).
 - [44] A. Hayashi, A. Sakurai, S. Nishio, W. J. Munro, and K. Nemoto, Impact of the form of weighted networks on the quantum extreme reservoir computation, *Phys. Rev. A* **108**, 042609 (2023).
 - [45] N. Götting, F. Lohof, and C. Gies, Exploring quantumness in quantum reservoir computing, *Phys. Rev. A* **108**, 052427 (2023).
 - [46] S. Kobayashi, Q. H. Tran, and K. Nakajima, Coherence influx is indispensable for quantum reservoir computing, [arXiv:2409.12693](#).
 - [47] S. Čindrak, L. Jaurigue, and K. Lüdge, Engineering quantum reservoirs through Krylov complexity, expressivity, and observability, *Phys. Rev. Res.* **7**, 043190 (2025).
 - [48] N.-E. Schütte, N. Götting, H. Müntinga, M. List, D. Brunner, and C. Gies, Expressivity limits of quantum reservoir computing, [arXiv:2501.15528](#).
 - [49] R. Martínez-Peña and J.-P. Ortega, Input-dependence in quantum reservoir computing, *Phys. Rev. E* **111**, 065306 (2025).
 - [50] A. De Lorenzis, M. P. Casado, N. Lo Gullo, T. Lux, F. Plastina, and A. Riera, Entanglement and classical simulability in quantum extreme learning machines, [arXiv:2509.06873](#).
 - [51] 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).
 - [52] S. Shin, Y. S. Teo, and H. Jeong, Exponential data encoding for quantum supervised learning, *Phys. Rev. A* **107**, 012422 (2023).
 - [53] 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).
 - [54] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, Information processing capacity of dynamical systems, *Sci. Rep.* **2**, 514 (2012).
 - [55] R. Martínez-Peña, J. Nokkala, G. L. Giorgi, R. Zambrini, and M. C. Soriano, Information processing capacity of spin-based quantum reservoir computing systems, *Cognit. Comput.* **15**, 1440 (2023).
 - [56] Z. Shumaylov, P. Zaika, P. Scholl, G. Kutyniok, L. Horeh, and C.-B. Schönlieb, When is a system discoverable from data? Discovery requires chaos, [arXiv:2511.08860](#).
 - [57] G. McCaul, J. S. Toterogongora, W. Otieno, S. Savel'ev, A. Zagorskin, and A. G. Balanov, Minimal quantum reservoirs with Hamiltonian encoding, *Chaos* **35**, 093135 (2025).
 - [58] M. Schuld and F. Petruccione, *Machine Learning with Quantum Computers*, 2nd ed., Quantum Science and Technology (Springer, Cham, 2021).
 - [59] M. C. Caro, Learning quantum processes and Hamiltonians via the Pauli transfer matrix, *ACM Trans. Quantum Comput.* **5**, 1 (2024).
 - [60] L. Hantzko, L. Binkowski, and S. Gupta, Fast generation of Pauli transfer matrices utilizing tensor product structure, *Phys. Scr.* **100**, 075125 (2025).
 - [61] A. Angrisani, A. Schmidhuber, M. S. Rudolph, M. Cerezo, Z. Holmes, and H.-Y. Huang, Classically estimating observables of noiseless quantum circuits, *Phys. Rev. Lett.* **135**, 170602 (2025).
 - [62] A. Angrisani, A. A. Mele, M. S. Rudolph, M. Cerezo, and Z. Holmes, Simulating quantum circuits with arbitrary local noise using Pauli propagation, *PRX Quantum* **7**, 020313 (2026).
 - [63] M. S. Rudolph, T. Jones, Y. Teng, A. Angrisani, and Z. Holmes, Pauli propagation: A computational framework for simulating quantum systems, *PRX Quantum* **7**, 032001 (2026).
 - [64] S. Čindrak, B. Donvil, K. Lüdge, and L. Jaurigue, Enhancing the performance of quantum reservoir computing and solving the time-complexity problem by artificial memory restriction, *Phys. Rev. Res.* **6**, 013051 (2024).
 - [65] J. Steinegger and C. Răth, Predicting three-dimensional chaotic systems with four qubit quantum systems, *Sci. Rep.* **15**, 6201 (2025).
 - [66] This allows for the existence of a pseudoinverse such that $R^+R = \mathbb{I}_q$.
 - [67] D. J. Gauthier, E. Bollt, A. Griffith, and W. A. S. Barbosa, Next generation reservoir computing, *Nat. Commun.* **12**, 5564 (2021).
 - [68] H. C. So, K. W. Chan, Y. T. Chan, and K. C. Ho, Linear prediction approach for efficient frequency estimation of multiple real sinusoids: Algorithms and analyses, *IEEE Trans. Signal Process.* **53**, 2290 (2005).
 - [69] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, 4th ed. (John Wiley & Sons, Chichester, UK, 2008).

- [70] M. Gross, Flow map learning in nonlinear vector autoregressive models: Influence of the feature-library structure on the training error, [arXiv:2605.31438](#).
- [71] M. C. Caro, E. Gil-Fuster, J. J. Meyer, J. Eisert, and R. Sweke, Encoding-dependent generalization bounds for parametrized quantum circuits, *Quantum* **5**, 582 (2021).
- [72] E. Peters and M. Schuld, Generalization despite overfitting in quantum machine learning models, *Quantum* **7**, 1210 (2023).
- [73] 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).
- [74] S. Čindrak, K. Lüdge, and L. Jaurigue, From Krylov complexity to observability: Capturing phase space dimension with applications in quantum reservoir computing, *Phys. Rev. Res.* **7**, L042039 (2025).
- [75] D. E. Parker, X. Cao, A. Avdoshkin, T. Scaffidi, and E. Altman, A universal operator growth hypothesis, *Phys. Rev. X* **9**, 041017 (2019).
- [76] A. Nahum, S. Vijay, and J. Haah, Operator spreading in random unitary circuits, *Phys. Rev. X* **8**, 021014 (2018).
- [77] E. H. Lieb and D. W. Robinson, The finite group velocity of quantum spin systems, *Commun. Math. Phys.* **28**, 251 (1972).
- [78] B. Nachtergaele, Y. Ogata, and R. Sims, Propagation of correlations in quantum lattice systems, *J. Stat. Phys.* **124**, 1 (2006).
- [79] C.-F. (Anthony) Chen, A. Lucas, and C. Yin, Speed limits and locality in many-body quantum dynamics, *Rep. Prog. Phys.* **86**, 116001 (2023).
- [80] B. J. Mahoney and C. S. Lent, Lieb-Robinson correlation function for the quantum transverse-field Ising model, *Phys. Rev. Res.* **6**, 023286 (2024).
- [81] A. Sannia, G. L. Giorgi, and R. Zambrini, Exponential concentration and symmetries in quantum reservoir computing, [arXiv:2505.10062](#).
- [82] S. Gopalakrishnan, D. A. Huse, V. Khemani, and R. Vasseur, Hydrodynamics of operator spreading and quasiparticle diffusion in interacting integrable systems, *Phys. Rev. B* **98**, 220303(R) (2018).
- [83] A. Avdoshkin and A. Dymarsky, Euclidean operator growth and quantum chaos, *Phys. Rev. Res.* **2**, 043234 (2020).
- [84] J. D. Noh, Operator growth in the transverse-field Ising spin chain with integrability-breaking longitudinal field, *Phys. Rev. E* **104**, 034112 (2021).
- [85] R. Kaneko and I. Danshita, Dynamics of correlation spreading in low-dimensional transverse-field Ising models, *Phys. Rev. A* **108**, 023301 (2023).
- [86] A. Sahu, N. D. Varikuti, B. K. Das, and V. Madhok, Quantifying operator spreading and chaos in Krylov subspaces with quantum state reconstruction, *Phys. Rev. B* **108**, 224306 (2023).
- [87] If R contains a $t = 0$ block, then since $V(0) = \mathbb{I}_{d^2}$ the corresponding rows of R include $\mathbf{e}_j^\top$ for all $j \in \mathcal{S}$ (up to basis ordering), hence $\mathbf{e}_j \in \text{row}(R)$. Since $R^\top R$ is the orthogonal projector onto $\text{row}(R)$, $(R^\top R)\mathbf{e}_j = \mathbf{e}_j$ and thus $\gamma_j^2 = (R^\top R)_{jj} = 1$.
- [88] Including $t \approx 0$ in the temporal multiplexing scheme improves decodability at early iterations but does not affect the final decodability for sufficiently large L .
- [89] By global Hadamard equivalence, the statement flips for Pauli- X as initial operators.
- [90] H. Jaeger, Short Term Memory in Echo State Networks, GMD Report 152 (GMD Forschungszentrum Informationstechnik, Sankt Augustin, 2001).
- [91] T. Kubota, H. Takahashi, and K. Nakajima, Unifying framework for information processing in stochastically driven dynamical systems, *Phys. Rev. Res.* **3**, 043135 (2021).
- [92] In Eq. (4.13), $\text{Var}_{\mathbf{u}} y(\mathbf{u})$ denotes the variance of $y(\mathbf{u})$ over the distribution of $\mathbf{u}$.
- [93] This argument is not fully rigorous, since features like $\phi_x^{(i)}(u_i)$ have infinite Taylor expansion in u_i and can thus correlate locally with raw monomials u_i^m .
- [94] W. Xia, J. Zou, X. Qiu, and X. Li, The reservoir learning power across quantum many-body localization transition, *Front. Phys.* **17**, 33506 (2022).
- [95] M. N. Ivaki, A. Lazarides, and T. Ala-Nissila, Quantum reservoir computing on random regular graphs, *Phys. Rev. A* **112**, 012622 (2025).
- [96] A. Palacios, R. Martínez-Peña, M. C. Soriano, G. L. Giorgi, and R. Zambrini, Role of coherence in many-body quantum reservoir computing, *Commun. Phys.* **7**, 369 (2024).
- [97] L. Ljung, *System Identification: Theory for the User*, 2nd ed. (Prentice Hall PTR, Upper Saddle River, NJ, 1999).
- [98] J. C. Sprott, *Elegant Chaos: Algebraically Simple Chaotic Flows* (World Scientific, Singapore, 2010).
- [99] We find that training on increments via $\mathbf{u}_{t+1} - \mathbf{u}_t = \mathbf{f}(\mathbf{u}_t)$ does not significantly improve the forecast horizon, unless one uses small integrator time steps $\Delta t \lesssim 0.001$.
- [100] C. Yang, J. Qiao, H. Han, and L. Wang, Design of polynomial echo state networks for time series prediction, *Neurocomputing* **290**, 148 (2018).
- [101] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, *Proc. Natl. Acad. Sci. USA* **113**, 3932 (2016).
- [102] T.-C. Chen, S. G. Penny, T. A. Smith, and J. A. Platt, ‘Next generation’ reservoir computing: An empirical data-driven expression of dynamical equations in time-stepping form, [arXiv:2201.05193](#).
- [103] Y. Zhang, E. Roque dos Santos, H. Zhang, and S. P. Cornelius, How more data can hurt: Instability and regularization in next-generation reservoir computing, *Chaos* **35**, 073102 (2025).
- [104] Moreover, nonorthogonal function libraries are typically highly redundant, reflected in large condition numbers of the corresponding design matrices [Eqs. (3.16) and (3.17)].
- [105] Deviations are small for $N < q$: in practice, for $n = D = 3$ (thus $q = 27$) and monomials of order $r = 3$ (thus $N = 19$ excluding the constant term), we get $\text{rank}(K) = 16$.
- [106] R. Sweke, E. Recio-Armengol, S. Jerbi, E. Gil-Fuster, B. Fuller, J. Eisert, and J. J. Meyer, Potential and limitations of random Fourier features for dequantizing quantum machine learning, *Quantum* **9**, 1640 (2025).
- [107] F. J. Schreiber, J. Eisert, and J. J. Meyer, Classical surrogates for quantum learning models, *Phys. Rev. Lett.* **131**, 100803 (2023).
- [108] C. Mingard, J. Pointing, C. London, Y. Nam, and A. A. Louis, Exploiting the equivalence between quantum neural networks and perceptrons, [arXiv:2407.04371](#).
- [109] S. Masot-Llima, E. Gil-Fuster, C. Bravo-Prieto, J. Eisert, and T. Guaita, Prospects for quantum advantage in machine learning from the representability of functions, [arXiv:2512.15661](#).

- [110] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control* (Cambridge University Press, Cambridge, UK, 2019).
- [111] E. Gil-Fuster, J. Eisert, and C. Bravo-Prieto, Understanding quantum machine learning also requires rethinking generalization, *Nat. Commun.* **15**, 2277 (2024).
- [112] S. Han, L. Jia, and L. Guo, Multiple embeddings for quantum machine learning, [arXiv:2503.22758](https://arxiv.org/abs/2503.22758).
- [113] J. Bowles, S. Ahmed, and M. Schuld, Better than classical? The subtle art of benchmarking quantum machine learning models, [arXiv:2403.07059](https://arxiv.org/abs/2403.07059).
- [114] J. Schnabel and M. Roth, Quantum kernel methods under scrutiny: A benchmarking study, *Quantum Mach. Intell.* **7**, 58 (2025).
- [115] D. Basilewitsch, J. F. Bravo, C. Tutschku, and F. Struckmeier, Quantum neural networks in practice: A comparative study with classical models from standard data sets to industrial images, *Quantum Mach. Intell.* **7**, 110 (2025).
- [116] T. Rohe, D. Schuman, J. Nüßlein, L. Sünkel, J. Stein, and C. Linnhoff-Popien, The questionable influence of entanglement in quantum optimisation algorithms, [arXiv:2407.17204](https://arxiv.org/abs/2407.17204).
- [117] H. Gundlach, H. Kukina, J. Lynch, and N. Thompson, Quantum deep learning still needs a quantum leap, in *Proceedings of the 43rd ICML*, Vol. 306 (PMLR, 2026), pp. 170357–170374.
- [118] T. Fellner, D. A. Kreplin, S. Tovey, and C. Holm, Quantum vs. classical: A comprehensive benchmark study for predicting time series with variational quantum machine learning, *Mach. Learn.: Sci. Technol.* **7**, 010501 (2026).
- [119] D. Freinberger and P. Moser, The role of quantum in hybrid quantum-classical neural networks: A realistic assessment, [arXiv:2601.04732](https://arxiv.org/abs/2601.04732).
- [120] F. J. Gil Vidal and D. O. Theis, Input redundancy for parameterized quantum circuits, *Front. Phys.* **8**, 297 (2020).
- [121] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, Data re-uploading for a universal quantum classifier, *Quantum* **4**, 226 (2020).
- [122] J. Kübler, S. Buchholz, and B. Schölkopf, The inductive bias of quantum kernels, in *Advances in Neural Information Processing Systems*, edited by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, and J. W. Vaughan (Curran Associates, Red Hook, NY, 2021), Vol. 34, pp. 12661–12673.
- [123] For the real product encodings considered here, any Pauli string containing a Y factor has zero pre-reservoir expectation, leaving at most $\sim 3^n$ features.
- [124] <https://qci.dlr.de/NeMoQC>.
- [125] M. B. Ruskai, S. Szarek, and E. Werner, An analysis of completely-positive trace-preserving maps on M_2 , *Lin. Algebr. Appl.* **347**, 159 (2002).
- [126] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th ed. (Cambridge University Press, Cambridge, UK, 2010).
- [127] M. M. Wilde, *Quantum Information Theory*, 2nd ed. (Cambridge University Press, Cambridge, UK, 2017).
- [128] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, UK, 2018).
- [129] T. Kubota, Y. Suzuki, S. Kobayashi, Q. H. Tran, N. Yamamoto, and K. Nakajima, Temporal information processing induced by quantum noise, *Phys. Rev. Res.* **5**, 023057 (2023).
- [130] T. Prosen and I. Pižorn, Operator space entanglement entropy in a transverse Ising chain, *Phys. Rev. A* **76**, 032316 (2007).
- [131] G. B. Mbeng, A. Russomanno, and G. E. Santoro, The quantum Ising chain for beginners, *SciPost Phys. Lect. Notes* **82** (2024).