Theory and interpretability of Quantum Extreme Learning Machines: a Pauli-transfer matrix approach

Markus Gross[1,*] and Hans-Martin Rieser[1]

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

(Dated: April 27, 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. We apply the Pauli transfer matrix (PTM) formalism to theoretically analyze the influence of encoding, reservoir dynamics, and measurement operations (including temporal multiplexing) on the QELM performance. This formalism reveals the complete set of (nonlinear) features generated by the encoding, and shows how the subsequent quantum channels linearly transform these Pauli features before they are probed by the chosen measurement operators. Optimizing such a QELM can therefore be cast as a decoding problem in which one shapes the channel-induced transformations such that task-relevant features become available to the regressor, effectively reversing the information scrambling of a unitary. Operator spreading under unitary evolution determines decodability of Pauli features, which underlies the nonlinear processing capacity of the reservoir. When paired with certain observables, structured Hamiltonians can reduce model expressivity, as reflected in a low readout rank. We trace this effect to Hamiltonian symmetries and derive asymptotic rank estimates for symmetry-resolved observable families. The PTM formalism yields a nonlinear vector (auto-)regression model as an interpretable classical representation of a QELM. As a specific application, we focus on forecasting nonlinear dynamical systems and show that a QELM trained on such trajectories learns a surrogate-approximation to the underlying flow map.

I. INTRODUCTION

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 post-processing of quantum experiments [13–16]. A major advantage of QRCs is their suitability for noisy intermediate-scale quantum (NISQ) 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 relevance of nonlinear input transformations has been analyzed for QRCs in [23–25], and universality conditions have been derived in [6, 26–30]. Notably, QRCs are closely related to quantum kernel methods [31]—a fact that can be utilized to find optimal measurement operators [32] and harness insights from the classical domain [33–35]. For general overviews on QRCs, we refer to [36, 37].

A specific QRC variant without any coherently evolving internal state is the QELM [2, 36–38]. Its classical counterpart, the 'extreme learning machine', is essentially a feedforward network with fixed internal architecture [39], similarly to a random-feature model [40]. QELMs are capable of

* markus.gross@dlr.de

various ML tasks, such as image classification and time series forecasting. A QELM differs from a conventional quantum neural network with fixed internal parameters in two important ways: first, instead of gate-based architectures, it typically 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 to maximize information processing and being able to properly decode relevant features from the measurements [14, 41–50]. Identifying possible quantum advantages, e.g., related to the exponentially large Hilbert space into which data is encoded, is an active research topic [3, 37, 44, 48, 51, 52]. However, the exponential concentration effect emerging for large qubit numbers is a phenomenon conceptually related to barren plateaus and can strongly reduce QELM performance [53].

The present study aims at an interpretable framework that can explain why and which failures in QELMs occur and thereby support their systematic design. To this end, we will discuss temporal multiplexing [1] as well as information processing capacity [54, 55] in the light of the PTM formalism. 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 developed formalism applies to n-qubit QELMs in which the encoding is time-independent and separated from reservoir evolution. Models with dynamical input injection [14, 57] or generic continuous-variable (infinite-dimensional) encodings [24, 58, 59] are therefore excluded. Our findings also hold implications for other QML models that fall into this category.

II. MODEL

We briefly present the formalism of QELMs and the specific implementation used in this work [1, 36, 37]. Let $\mathrm{Herm}(\mathcal{H})$ 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

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

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

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

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 $\mathcal{L}_{\mathrm{hinge}}(\hat{y}, y) = \max(0, 1 - y\hat{y})$, with $y \in \{-1, +1\}$ for

classification. In the case of the least-squares loss, the optimal weights are given by

$$\hat{\mathbf{w}} = (\mathbf{\Psi}\mathbf{\Psi}^\top + \lambda \mathbb{I}_m)^{-1}\mathbf{\Psi}\mathbf{y}, \qquad \Psi_{ki} = \mathrm{tr}(M_k \rho(\mathbf{x}_i)), \tag{2.3}$$

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

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

$$|\psi(\mathbf{x})\rangle = U_R(t)S(\mathbf{x})|\Phi_0\rangle. \tag{2.4}$$

Here, $S(\mathbf{x})$ 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 $U_R(t) = \exp(-\mathrm{i}tH_R)$ 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)

$$H_{\mathrm{TFIM}}^{\text{zz-x}} = -J\sum_{j=1}^{n-1} Z_jZ_{j+1} - h\sum_{j=1}^{n} X_j, \tag{2.5a}$$

$$H_{\mathrm{TFIM}}^{\text{xx-z}} = -J\sum_{j=1}^{n-1} X_jX_{j+1} - h\sum_{j=1}^{n} Z_j, \tag{2.5b}$$

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_jZ_{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 $U_{\mathrm{H}} = \mathrm{H}^{\otimes n}$:

$$U_{\mathrm{H}}X_jU_{\mathrm{H}}^\dagger = Z_j, \quad U_{\mathrm{H}}Y_jU_{\mathrm{H}}^\dagger = -Y_j, \quad U_{\mathrm{H}}Z_jU_{\mathrm{H}}^\dagger = X_j, \quad U_{\mathrm{H}}H_{\mathrm{TFIM}}^{\text{zz-x}}U_{\mathrm{H}}^\dagger = H_{\mathrm{TFIM}}^{\text{xx-z}}. \tag{2.6}$$

B. Encoding schemes

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

$$\rho_{\mathrm{enc}}(\mathbf{x}) = S(\mathbf{x})\rho_0 S(\mathbf{x})^\dagger = \bigotimes_{j=1}^{n} |\psi_{\mathrm{enc}}(x_j)\rangle\langle\psi_{\mathrm{enc}}(x_j)|, \qquad S(\mathbf{x}) = \bigotimes_{j=1}^{n} S_j(x_j), \tag{2.7}$$

where the unitary S_j act on the j-th qubit. In our numerical experiments we employ amplitude and rotational encoding schemes [60]. The former comes in two variants, where one either encodes data on half a great circle of the Bloch sphere:

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

or on the full great circle:

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

Rotational encoding is defined by the unitary

$$S_j(x_j) = \exp(-i\pi x_j \sigma^{p_j}) = R_{p_j}(2\pi x_j), \qquad x \in [0,1], \tag{2.10}$$

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, one typically uses R_Y gates, which yield the single-qubit state

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

III. THEORY

QML models, including QELMs, have often been analyzed from the Fourier perspective, which is natural for Hamiltonian-based encoding [51–53]. In the following, we instead employ the Pauli transfer matrix (PTM) formalism, which allows a more direct identification of the relevant features for models with initial-state encoding [36, 46, 61–63] and has been previously used for the analysis of QRCs [28, 36, 49].

A. Pauli-transfer formulation of a QELM

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

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

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

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

giving rise to a vector $\boldsymbol{\phi}(\mathbf{x}) \in \mathbb{R}^{d^2}$ 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 Section III B). Any quantum channel $\mathcal{E}$ acts linearly on this vector, $\boldsymbol{\phi}' = T_{\mathcal{E}}\, \boldsymbol{\phi}$, via the Pauli transfer matrix:

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

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 $F_k(\mathbf{x}) = \mathrm{tr}\big[P_k\,\mathcal{E}(\rho_{\mathrm{enc}}(\mathbf{x}))\big]$. Expanding ρ_{enc} as in Eq. (3.2) and using Eq. (3.3) gives the model output function

$$f(\mathbf{x}) = \mathbf{w}^\top \mathbf{F}(\mathbf{x}) = \mathbf{w}^\top T_{\mathcal{E}}^{(m)}\,\boldsymbol{\phi}(\mathbf{x}) = \sum_{k\in\mathcal{S}} w_k \sum_{j=0}^{d^2-1} (ST_{\mathcal{E}})_{kj}\phi_j(\mathbf{x}), \tag{3.4}$$

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 $\boldsymbol{\phi}(\mathbf{x})$ generated solely by the encoding and mixed by the reservoir channel $T_{\mathcal{E}}$.

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

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

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

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

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

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

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

$$V_{kj}(t) = \frac{1}{d}\mathrm{tr}\big[P_k\,U(t)P_jU(t)^\dagger\big], \tag{3.8}$$

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

B. Feature encoding

The structure of the Pauli feature vector $\boldsymbol{\phi}(\mathbf{x})$ 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

$$\rho_{\mathrm{enc}}(\mathbf{x}) = S(\mathbf{x})\,|0\rangle\langle 0|^{\otimes n}\,S(\mathbf{x})^\dagger = \bigotimes_{j=1}^{n} \rho_j(x_j), \qquad \rho_j(x_j) = S_j(x_j)|0\rangle\langle 0|S_j(x_j)^\dagger. \tag{3.9}$$

It is convenient to introduce the single-qubit Pauli features

$$\boldsymbol{\phi}^{(j)}(x_j) = \big(1, \phi_x^{(j)}(x_j), \phi_y^{(j)}(x_j), \phi_z^{(j)}(x_j)\big)^\top, \qquad \phi_a^{(j)}(x_j) = \mathrm{tr}\big[\sigma^a\,\rho_j(x_j)\big], \tag{3.10}$$

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

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). 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} (1st column) in an observable P produces the corresponding single-qubit feature $\phi_{a_k}^{(k)}(u_k) := \text{tr}\big(\sigma_{a_k}\, \rho_k(u_k)\big)$ (2nd–4th columns). For instance, the X_k row gives $\phi_x^{(k)}(u_k)$ and the Z_k row gives $\phi_z^{(k)}(u_k)$.

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 $\boldsymbol{\phi}^{(j)}$ define the Bloch vector, which has unit magnitude for pure states (see Appendix D for further discussion). Since both the state and the Pauli basis factorize across qubits, the n-qubit feature vector is

$$\boldsymbol{\phi}(\mathbf{x}) = \bigotimes_{j=1}^{n} \boldsymbol{\phi}^{(j)}(x_j), \tag{3.11}$$

and its components are given by

$$\phi_{a_1 \ldots a_n}(\mathbf{x}) = \text{tr}\Big[\Big(\bigotimes_{j=1}^{n} \sigma^{a_j}\Big)\, \rho_{\text{enc}}(\mathbf{x})\Big] = \prod_{j=1}^{n} \phi_{a_j}^{(j)}(x_j), \qquad a_j \in \{0, x, y, z\}. \tag{3.12}$$

The subscript $a_1 \ldots a_n$ 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. (3.10): 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.

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 Section D 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, 45, 47, 49]:

1. Decide which types of features $\{\phi_r\}$ are best suited to learn the dataset. For example, non-Markovian time series (including harmonic signals) require delay coordinates [64, 65], whereas Markovian dynamical systems can already be modeled by instantaneous monomial features.

While commonly used encodings support universal function approximation [6, 26–30], they can differ in their approximation errors, generalization behavior [51, 66**?**], or robustness during autoregressive rollout.

2. Select suitable measurement observables $\{P_k\}$. Ideally, the Heisenberg-evolved observables $P_k^{(\mathrm{H})}(t)$ 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 [32]. Another possibility is temporal multiplexing [1, 67–69], discussed further in the PTM framework below (see Section IV B).

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 decodability means that, for each such ϕ_r, there exists a weight vector $\mathbf{w}_r$ such that $\mathbf{w}_r T_{\mathcal{E}} \boldsymbol{\Phi} \propto \phi_r$ (see Section IV C). 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 Section IV D). We will show that temporal multiplexing directly enhances decodability by increasing the operator support.

In temporal multiplexing [1, 70], 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 and evaluates them at several times $t_1, \ldots, t_L$ [cf. Eq. (3.4)]. One then stacks the corresponding rows of $V(t)$ into a $mL \times d^2$ observability matrix

$$R_L = \begin{pmatrix} SV(t_1) \\ SV(t_2) \\ \vdots \\ SV(t_L) \end{pmatrix}. \tag{4.1}$$

(We will usually denote this simply by R.) The case without temporal multiplexing corresponds to $L = 1$. The full unitary PTM $R = V(t)$ is then recovered for $S = \mathbb{I}_{d^2}$.

A. Classical representation

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,

$$f(\mathbf{u}) = \mathbf{w}^\top \mathbf{F}(\mathbf{u}), \qquad \mathbf{F}(\mathbf{u}) \equiv R\boldsymbol{\Phi}(\mathbf{u}), \tag{4.2}$$

where $R \in \mathbb{R}^{B \times q}$ is the effective PTM (4.1) (including any row-selection and temporal multiplexing) and $F_k = R_{kj}\phi_j$ are the Pauli features available at the readout.

If R has full column rank, it is left-invertible [71] and thus the regressor can always find weights $\mathbf{w} \propto (R^+)_r$ that isolate 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:

$$\hat{\mathbf{w}} = (GG^\top + \lambda \mathbb{I}_B)^{-1} G\mathbf{y} = R(CR^\top R + \lambda \mathbb{I}_q)^{-1} \Phi \mathbf{y}, \tag{4.3}$$

where we introduced the feature design and Gram matrices

$$\Phi \equiv [\boldsymbol{\phi}(\mathbf{u}_1)\cdots\boldsymbol{\phi}(\mathbf{u}_P)] \in \mathbb{R}^{q\times P}, \quad C = \Phi\Phi^\top \in \mathbb{R}^{q\times q}, \tag{4.4}$$

and the corresponding readout design and Gram matrices

$$G \equiv [\mathbf{F}(\mathbf{u}_1)\cdots\mathbf{F}(\mathbf{u}_P)] = R\Phi \in \mathbb{R}^{B\times P}, \quad GG^\top = RCR^\top \in \mathbb{R}^{B\times B}. \tag{4.5}$$

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 (4.3) (which follows from the Woodbury 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:

$$\hat{\mathbf{w}} = V(C + \lambda\mathbb{I}_q)^{-1}\Phi\mathbf{y}. \tag{4.6}$$

Moreover, since V is orthogonal, it drops out from the predictor (4.2), 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 $\boldsymbol{\phi}$. In the context of time series processing (especially when using some inputs u_j to encode delay states) this is reduces to a "next-generation" reservoir computer (NG-RC) [72] or, equivalently, a nonlinear vector autoregressive process (NVAR). When R merely has full column rank, it can be interpreted to induce a regressor operating on the mixed feature set $\boldsymbol{\phi}' = R\boldsymbol{\phi}$.

A coarse measure for expressivity is the rank of the readout Gram matrix in Eq. (4.3), which can be bounded as

$$\operatorname{rank}(G) \le \min(\operatorname{rank}(R), \operatorname{rank}(\Phi)) \le \min(B, P, q). \tag{4.7}$$

For sufficiently diverse data, Φ typically achieves a rank near its maximum $\min(q, P)$, but can suffer from ill-conditioning since the features ϕ_k are in general not orthogonal. Accordingly, for large sample number P, the rank of the readout approximately follows the rank of the effective PTM:

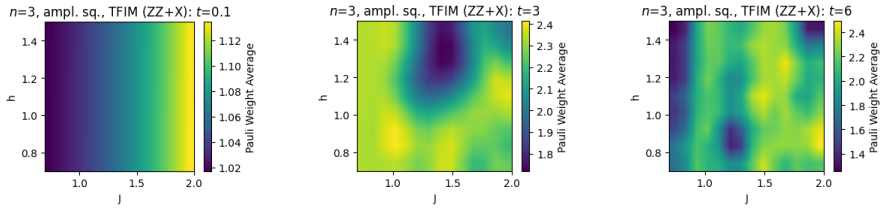
$$\operatorname{rank}(G) \simeq \operatorname{rank}(R). \tag{4.8}$$

For random Hamiltonians, $\operatorname{rank}(R)$ typically grows proportionally to the measurement budget B and then saturates near q (see Fig. 4). By contrast, structured Hamiltonians like the TFIM are not fully scrambling, resulting in $\operatorname{rank}(R) \ll q$. The achieved ranks indicate the quality of the measurement strategy (see Section IV C) and thus of the resulting training accuracy (see Fig. 11).

B. 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, 68, 73], which we summarize here in our setting. Specifically, the unitary Heisenberg-evolution [Eq. (3.7)] of a given Pauli observable P_k satisfies,

$$P_k^{(\mathrm{H})}(t) \equiv U(t)^\dagger P_k U(t) = e^{iHt}P_k e^{-iHt} = e^{t\mathcal{L}}(P_k) = \sum_{\ell=0}^{\infty}\frac{t^\ell}{\ell!}\,\mathcal{L}^\ell(P_k), \tag{4.9}$$



(a) (b) (c)

Figure 1. Time evolution of the Pauli weight average $\bar{\nu}_k$ [Eq. (4.13)] of a single-site Pauli σ_k in the h–J coupling parameter space of the TFIM (2.5a) for $n = 3$ qubits.

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

$$\mathcal{K}_\ell(P_k) = \mathrm{span}\big\{P_k,\, \mathcal{L}(P_k),\, \ldots,\, \mathcal{L}^{\ell-1}(P_k)\big\}. \tag{4.10}$$

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

$$(Z_{j_1} \cdots Z_{j_k})^{\mathrm{H}}(t)\big|_{\text{zz-x}} = U_{\mathrm{H}}^\dagger (X_{j_1} \cdots X_{j_k})^{\mathrm{H}}(t)\big|_{\text{xx-z}} U_{\mathrm{H}}, \tag{4.11}$$

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

In the PTM approach, Krylov growth can be represented in terms of *operator spreading* in the Pauli basis [74]. Accordingly, expanding the time-evolved observable $P_k^{(\mathrm{H})}(t)$ gives

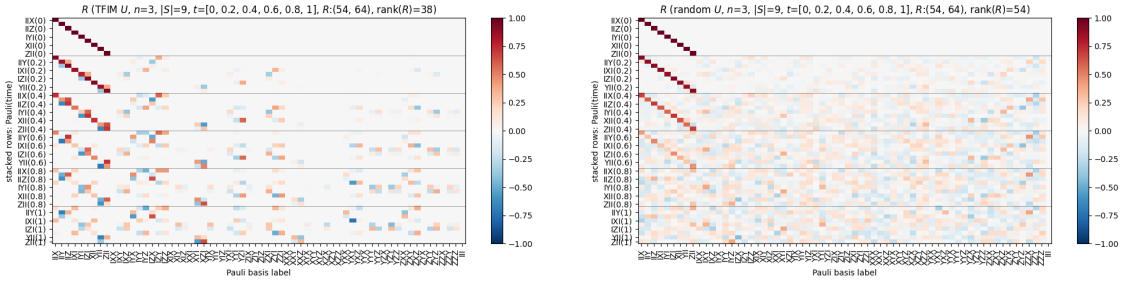
$$P_k^{(\mathrm{H})}(t) = \sum_{j=0}^{d^2-1} V_{kj}(t) P_j, \tag{4.12}$$

where we used cyclicity of the trace to introduce $V(t)$ from Eq. (3.8). Thus, for non-Clifford H, 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 $\nu(P_j) \in \{0, \ldots, n\}$, i.e., the number of non-identity factors in P_j, and its coefficient-weighted average

$$\bar{\nu}_k(t) = \sum_{j=0}^{d^2-1} \nu(P_j)\, V_{kj}(t)^2. \tag{4.13}$$

For local (including nonintegrable/chaotic ones) Hamiltonians, the Lieb–Robinson bound [50, 75–78] heuristically implies that an initially local operator can only develop support within distance $\lesssim v_{\mathrm{LR}} t$ until finite-size saturation:

$$\bar{\nu}_k(t) \sim \begin{cases} 1 + \mathrm{const} \cdot v_{\mathrm{LR}} t & \text{(intermediate times)} \\ \mathcal{O}(n) & \text{(late times)} \end{cases} \tag{4.14}$$



(a) (b)

Figure 2. Visualization of the operator spreading (4.12) via the effective PTM R [Eq. (4.1)] constructed in the temporal multiplexing scheme for the TFIM (a) and random unitary (b). The selected observable set $\mathcal{S}$ are all weight-1 Pauli operators of a $n = 3$ 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. (4.12)]. In the TFIM case, the supporting basis exhibits a sparse pattern, whereas it is dense for a random unitary.

with a typical late-time value $\bar{\nu}_k \approx \frac{3}{4}n$, corresponding to a uniform distribution of coefficients V_{kj}^2 over Pauli strings. From the definition (3.7) and the expansion (4.12), 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. Consequently, for Haar-random unitaries, one expects a typical late-time magnitude

$$|V_{kj}(t \to \infty)| = 2^{-n}. \tag{4.15}$$

This is the origin of the exponential concentration effect, which makes accessing information about the input infeasible for large qubit numbers $n \to \infty$ [53]. Integrable models (such as the TFIM at its integrable points) 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 [79–83]. The evolution of the Pauli weight average is illustrated in Fig. 1. As shown in Section E 1, the interplay of Hamiltonian structure and observables can have significant impact on the Heisenberg evolution and thus of the resulting Krylov spans.

C. Pauli feature decoding

Following Eq. (4.2), we denote the set of Pauli features available at the readout by $F_k = R_{kj}\phi_j$ ($k = 1, \ldots, B$), which emerge by linear transformations of the input features ϕ_j ($j = 1, \ldots, q \leq d^2$) via the effective PTM $R \in \mathbb{R}^{B\times q}$ [Eq. (4.1)], possibly including temporal multiplexing. We now ask for conditions that enable the regressor to isolate a given input feature $\phi_r(\mathbf{x})$ by finding weights $\mathbf{w}$ such that (approximately)

$$\mathbf{w}\cdot\mathbf{F} \propto \phi_r \quad \Leftrightarrow \quad \mathbf{w}^\top R = \omega \mathbf{e}_r^\top, \tag{4.16}$$

where ω is a free parameter and $\mathbf{e}_r$ is the r-th basis vector of $\mathbb{R}^q$.

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

$$\mathbf{w} = \omega V(t)\mathbf{e}_r. \qquad \text{(unitary channel)} \tag{4.17}$$

If we select subsets of observables, then Eq. (4.16) can in general *not* be satisfied for arbitrary r since features get mixed under a quantum channel. The exact condition derived in Appendix B provides a Pauli feature decodability score:

$$\gamma_r^2 \equiv (R^+R)_{rr}, \qquad \sum_{r=1}^{q} \gamma_r^2 = \mathrm{rank}(R) \le q, \tag{4.18}$$

and associated decoding weights

$$\mathbf{w} = \omega \left((R^+)_{r,:}\right)^\top, \tag{4.19}$$

in terms of the Moore–Penrose pseudoinverse R^+. These coincide with the minimum-norm least-squares solution. 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 $\sim \sqrt{1-\gamma_r^2}\|\phi\|_2$. We take $\gamma_r^2 \gtrsim 0.5$ as a useful decodability threshold. Note that γ_r^2 is a geometric notion and ignores statistical correlations or noise floor (see Appendix C for further discussion).

Both score and weights preserve tensor product structure of the general PTM R: if $R = \otimes_{i=1}^n R_i$ then $\gamma_r^2 = \prod_{i=1}^n \gamma_{r,i}^2$ and $\mathbf{w} = \otimes_{i=1}^n \mathbf{w}_i$. For $R = SV(t)$, 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 $\mathcal{S}$.[84] Specifically, for Haar-random unitaries, late-time decodability decays exponentially with qubit number n [see Eq. (4.15)],

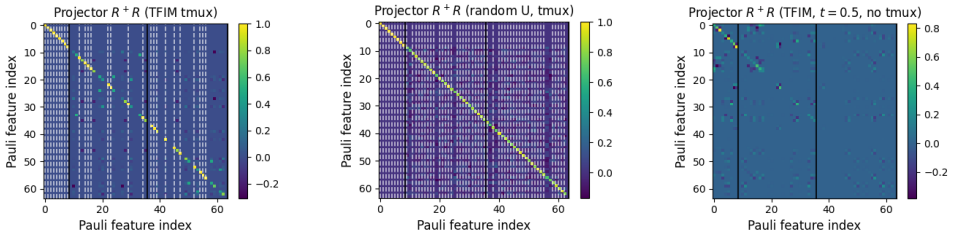
$$\gamma_r^2\big|_{t\to\infty} \sim |\mathcal{S}|4^{-n}. \tag{4.20}$$

Noisy channels can have invertible PTMs [see Section D 2] and then do not strictly reduce the accessible feature subspace. If they are contracting, they effectively suppress high-weight Pauli features [see Eq. (D12)], 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 matrices R^+R are shown in Fig. 3(a,b). Since we include here $t = 0$ in the multiplexing scheme, the first $|\mathcal{S}| = 9$ rows of R have 1 on the diagonal (no spreading), leading to perfect decodability of the weight-1 Pauli sector. Without temporal multiplexing and nonzero initial time, perfect feature reconstruction is not possible, see Fig. 3(c).

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 [85]. 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 Fig. 4(c) and (g), Pauli-Z's evolving under the Hamiltonian $H_{\mathrm{TFIM}}^{\text{xx-z}}$ render only weak scores compared to the Hamiltonian $H_{\mathrm{TFIM}}^{\text{zz-x}}$ [86].

A more detailed analysis using the Jordan-Wigner/Majorana formalism (see Appendix E) shows that, under $H_{\mathrm{TFIM}}^{\text{xx-z}}$, the Pauli strings generated from an initial Z are mainly consisting of Z-strings with X/Y endpoints, of which there are $\sim \mathcal{O}(n^2)$. By contrast, spreading under $H_{\mathrm{TFIM}}^{\text{zz-x}}$ almost fully saturates the Pauli basis, producing $\sim \mathcal{O}(4^n)$ strings (up to sub-exponential factors). The



(a) (b) (c)

Figure 3. Projector R^+R of the effective PTM [Eq. (4.1)] for different unitary types and a complete weight-1 Pauli observable set. Panels (a,b) directly correspond to the R shown in Fig. 2, which are constructed by temporal multiplexing. Panel (c) shows the projector based on $R = V(t = 0.5)$ (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.

Hamiltonian	Observables	Asymptotic saturated rank r_∞
$H_{\text{TFIM}}^{\text{xx-z}}$	Z_j	$\mathcal{O}(n^2)$
$H_{\text{TFIM}}^{\text{xx-z}}$	$Z_j,\ Z_iZ_j$	$\mathcal{O}(n^4)$
$H_{\text{TFIM}}^{\text{xx-z}}$	X_j, Y_j, Z_j	$\mathcal{O}(4^n)$ (up to poly. factors)
$H_{\text{TFIM}}^{\text{zz-x}}$	Z_j	$\mathcal{O}(4^n)$ (up to poly. factors)
$H_{\text{TFIM}}^{\text{zz-x}}$	$Z_j,\ Z_iZ_j$	$\mathcal{O}(4^n)$ (near full)
$H_{\text{TFIM}}^{\text{zz-x}}$	X_j, Y_j, Z_j	$\mathcal{O}(4^n)$ (up to poly. factors)

Table II. Asymptotic saturated rank $r_\infty := \lim_{L\to\infty} \text{rank}(R_L)$ (see Section E 2) of the temporally multiplexed PTM (4.1) 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 $H_{\text{TFIM}}^{\text{xx-z}}$ leads to exponentially lower expressivity, reflected by polynomial-in-n ranks $\ll 4^n$.

implications for the asymptotic rank $r_\infty := \lim_{L\to\infty} \text{rank}(R_L)$ are summarized in Table II and numerically confirmed in Fig. 5(b). Additionally, Fig. 5(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 F.

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 sectors for large L. In general, we find that temporal multiplexing with Pauli Z and ZZ observables and random unitaries requires $L \sim d^2/|\mathcal{S}|$ 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 much fewer measurements (see Figs. 7 and 8 below).

Summary. The decodability framework presented here extents the Krylov observability theory of [47, 68] by resolving geometry on the level of individual Pauli features. We remark that the Krylov grade M discussed in [47] corresponds to the asymptotic rank of a single observable, $r_\infty(\mathcal{S} = \{O\}) = M$ [Eq. (E11)], while the effective dimension score O_K can be regarded as a scalar proxy for our multi-observable $\text{rank}(R_L)$. The geometric decodability score γ_r^2 of a feature r [Eq. (4.18)]

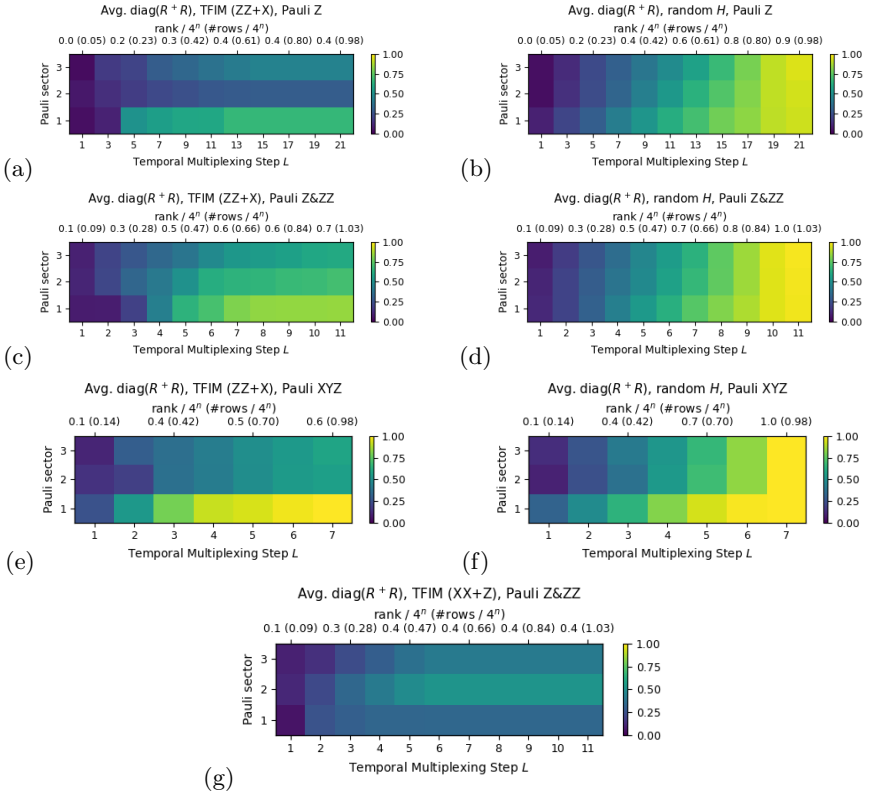


Figure 4. Feature decodability score $\bar{\gamma}_j^2$ obtained from the effective PTM R [Eq. (4.18)], averaged over each Pauli weight sector, of a $n = 3$ qubit model as a function of temporal multiplexing iteration length L [Eq. (4.1)]. The unitary evolution time is given by $t_L = L$ (i.e., initial 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,e), the TFIM (2.5b) in (g), and random Hamiltonians in (b,d,f). The observable set $\mathcal{S}$ is either only Z-Paulis (a,b), Z- and ZZ-Paulis (c,d,g), or all weight-1 Paulis (e,f).

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, 69]. While structured Hamiltonians can be advantageous in the large-n limit to avoid exponential concentration issues [53]. However, they can have non-trivial interplay with observables and thereby negatively impact expressivity of the model under temporal multiplexing, as reflected in sub-exponential scaling of the PTM rank with qubit number n (see Table II and Fig. 2).

D. Statistical decodability of input data

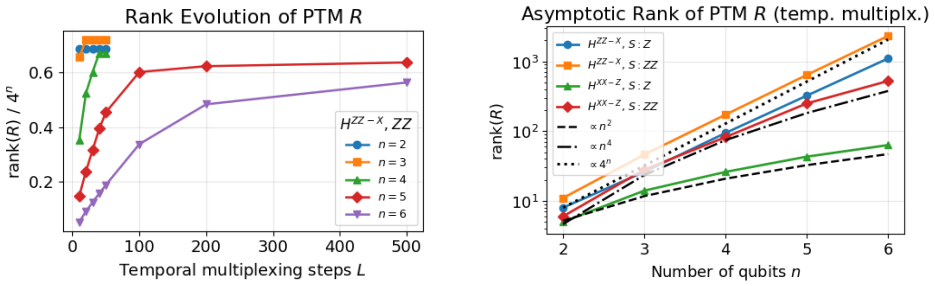
We focused above on the Pauli features ϕ_j, which are generated by the encoding and transform linearly under the action of the QELM. However, they are themselves nonlinear functions of the classical input $\mathbf{u} \in \mathbb{R}^D$. For example, with amplitude encoding (2.8), a single Pauli-Z generates a linear readout $\langle Z \rangle \propto u$, 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 initially as memory capacity [87] and later generalized in terms of information processing capacity [54, 55, 67, 88]. We discuss this, as well as an alternative approach, next.

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 nonlinear degree within this statistical approach. Following [54, 67], we therefore use orthogonal polynomials $p_j(u_\alpha)$ (degree j) to construct orthogonal degree-k targets as $y_{\mathbf{e}}(\mathbf{u}) = \prod_{\alpha=1}^{D} p_{e_\alpha}(u_\alpha)$, with $\sum_{\alpha=1}^{D} e_\alpha = k$. Specifically, for amplitude encoding, we use (shifted) Legendre polynomials, which are orthonormal w.r.t. the uniform distribution $[-1,1]$ or $[0,1]$; for angle encoding, we use Hermite polynomials, which are orthonormal w.r.t. the Gaussian distribution $\mathcal{N}(0,1)$ [89].

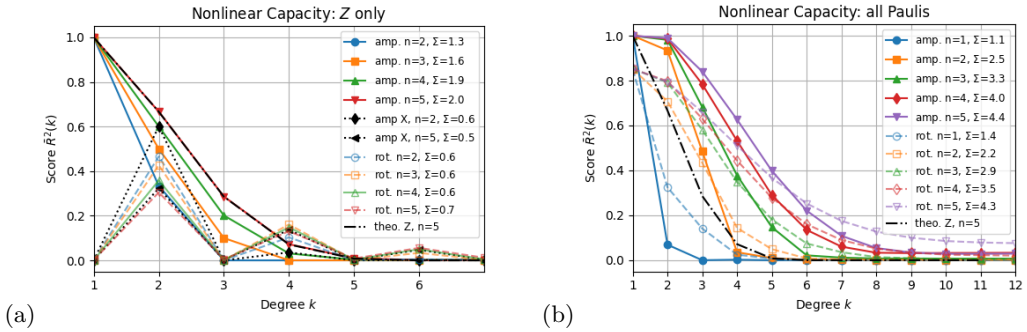
We take the coefficient of determination [54, 55]

$$\mathcal{R}^2(k) = \frac{1}{N_k} \sum_{|\mathbf{e}|=k} \mathcal{R}^2[y_{\mathbf{e}}], \qquad \mathcal{R}^2[y_{\mathbf{e}}] = 1 - \frac{\mathbb{E}_{\mathbf{u}}(y_{\mathbf{e}}(\mathbf{u}) - \hat{f}(\mathbf{u}))^2}{\mathrm{Var}_{\mathbf{u}} y_{\mathbf{e}}(\mathbf{u})}, \tag{4.21}$$



(a) (b)

Figure 5. Rank of the effective PTM R [Eq. (4.1)] obtain 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 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, $\lim_{L\to\infty} \mathrm{rank}(R)$, obtained numerically (solid curves), compared to the theoretical scaling predictions from Table II (broken lines).



(a) (b)

Figure 6. Ideal nonlinear capacity score $\mathcal{R}^2(k)$ [Eq. (4.21)] for amplitude and rotational encodings [Eqs. (2.8) and (2.10)], indicating the ability of an ideal QELM $f(\mathbf{u}) = \mathbf{w}^\top \boldsymbol{\phi}(\mathbf{u})$ to generate degree-k monomials of the input u_α ($\alpha = 1, \ldots, D$ and $n = D$). 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 dash-dotted curve represents the theoretical score for $\mathcal{R}^2_Z(k)$ with amplitude encoding and $n = 5$ input qubits [Eq. (4.22)], which agrees exactly with the numerical results. Σ denotes the integrated capacity over all k.

averaged over all $N_k = \binom{D+k-1}{k}$ degree-k monomials, to indicate whether the trained predictor $\hat{f}(\mathbf{u})$ can generate nonlinearities of degree k in the $(D = n)$-dimensional input [90]. $\mathcal{R}^2$ generalizes the geometric score (4.18) by taking statistical correlations into account (see Appendix C). In the special case where the targets are whitened orthonormal features, one has $\mathcal{R}^2[y_r] = \gamma_r^2$. In the case of amplitude encoding (2.8) with Pauli-Z readout, the number of representable degree-k targets is $\binom{D}{k}$, giving an averaged score

$$\mathcal{R}^2_Z(k) = \begin{cases} \binom{D}{k} / \binom{D+k-1}{k} & \text{for } k \leq D, \\ 0 & \text{for } k > D \end{cases} \tag{4.22}$$

which exponentially decreases with k.

In Fig. 6, we first display the nonlinear capacity that can be *ideally* achieved from the encoded state without effects of feature mixing, i.e., we take $\mathbf{F} = \boldsymbol{\phi}$ in (4.2). As shown in panel (a), a Pauli-X readout (black symbols with dotted curves) generates even degree monomials, behaving similarly to angle encoding with Pauli-Z observables (connected open symbols). In Fig. 6(b) we show the nonlinear capacity score $\mathcal{R}^2(k)$ 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 dash-dotted curve corresponding to $\mathcal{R}^2_Z(k)$ for $D = n = 5$ [Eq. (4.22)]. Moreover, angle and amplitude encodings show a similar behavior. We attribute the decay of $\mathcal{R}^2(k)$ with k mainly to 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 [91].

Figure 7 illustrates the nonlinear capacity score $\mathcal{R}^2(k)$ obtained for a QELM under temporal multiplexing of Pauli-Z/ZZ observables with the TFIM or random Hamiltonians. 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)

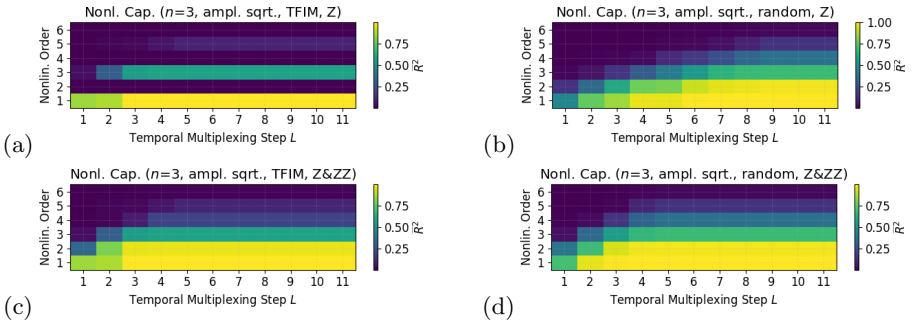


Figure 7. Nonlinear capacity score $\mathcal{R}^2(k)$ [Eq. (4.21)] as a function of the nonlinear order k and the temporal multiplexing iteration length L [Eq. (4.1)]. The unitary evolution time is given by $t_L = L$. We consider a $n = 3$ qubit model in the same setup as in Fig. 4. The unitary is based on the TFIM (2.5a) (a,c) or a random Hamiltonian (b,d), while the observable set $\mathcal{S}$ consists of either only Z-Paulis (a,b) or Z- and ZZ-Paulis (c,d).

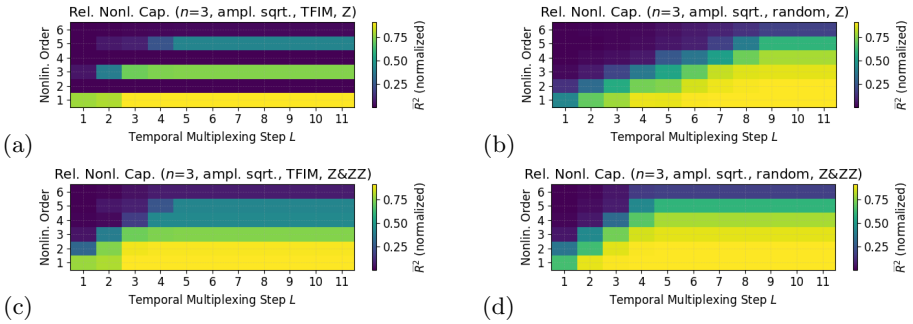


Figure 8. Relative nonlinear capacity score $\mathcal{R}^2(k)/(\mathcal{R}^2_{\text{all}}(k) + \delta)$ [Eq. (4.21)] as a function of the nonlinear order k and the temporal multiplexing iteration length L [Eq. (4.1)]. $\mathcal{R}^2_{\text{all}}$ is the nonlinear capacity score obtained for a complete Pauli observable set, and we choose $\delta = 0.1$ as a cut-off since scores $\mathcal{R}^2_{\text{all}} \lesssim 0.1$ are not considered meaningful. We use the same setup as in Fig. 7. The unitary is based on the TFIM (2.5a) (a,c) or a random Hamiltonian (b,d), while the observable set $\mathcal{S}$ consists of either only Z-Paulis (a,b) or Z- and ZZ-Paulis (c,d).

the encodings used here can generate only a small subset of all possible degree-k monomials (see Eq. (4.22) and Fig. 6). A clearer measure of the effectiveness of temporal multiplexing to generate nonlinearities is thus the score $\mathcal{R}^2(k)$ normalized by the score $\mathcal{R}^2_{\text{all}}(k)$ obtained when measuring over the complete Pauli observable set (see Fig. 8). (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.

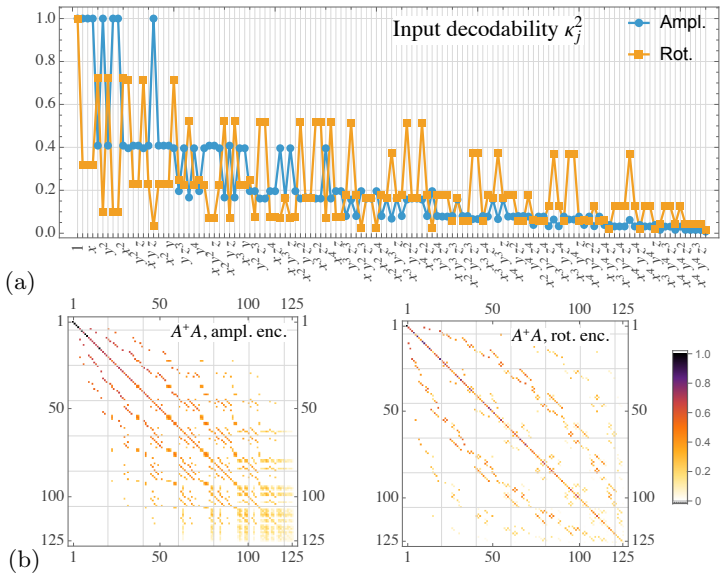


Figure 9. Ability to construct monomials of the input $\{x, y, z\}$ for amplitude encoding (2.8) and rotational encoding (2.10) with $n = D = 3$. (a) Decodability score κ_j^2 according to Eq. (4.23). (b) Projector matrix A^+A, where nonzero off-diagonals correspond to $\kappa_j^2 < 1$ and indicate decoding is possible only within a certain accuracy. The underlying Taylor expansion is performed up to order $r = 4$ in each $u_j = \{x, y, z\}$ 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).

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 $\mathcal{B}(\mathbf{u}) = \{1, u_1, \ldots, u_1u_2, \ldots, \prod_\alpha u_\alpha^r\}$ be the vector of size b of monomials of $\{u_\alpha\}_{1\le\alpha\le D}$ up to degree r. The order-r Taylor expansion of the Pauli features $\{\phi_j(\mathbf{u})\}_{1\le j\le q}$ within a small region around some operating point $\mathbf{u}_0$ (e.g., data mean) can then be written as $\boldsymbol{\phi}(\mathbf{u}) \simeq A(\mathbf{u}_0)\mathcal{B}(\mathbf{u} - \mathbf{u}_0)$ with a coefficient matrix $A(\mathbf{u}_0) \in \mathbb{R}^{q\times b}$. As before, we can define a geometric decodability score

$$\kappa_j^2 \equiv (A^+A)_{jj}, \qquad \sum_{j=1}^{b} \kappa_j^2 = \mathrm{rank}(A) \le \min(q, b), \tag{4.23}$$

which measures how accurately a certain monomial in $\mathcal{B}$ can be constructed from linear combinations of ϕ_j. This score depends on the operating point $\mathbf{u}_0$, but is independent of quantum properties of the reservoir, and ranges between 0 and 1 (no/exact construction possible). Intermediate 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. 9. 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. 9(b)], and contributes to the robust performance often observed for rotational encoding.

E. Summary and Discussion

It is useful to briefly discuss the *injectivity* property of the feature map [49], which ensures that distinct inputs map to distinct readouts. A standard condition for local injectivity of $\mathbf{F}(\mathbf{u}) = R\boldsymbol{\phi}(\mathbf{u})$ [Eq. (4.2)] is a full-rank Jacobian:

$$\operatorname{rank} RJ_{\boldsymbol{\phi}}(\mathbf{u}) = D, \qquad J_{\boldsymbol{\phi}}(\mathbf{u}) = \frac{\partial \boldsymbol{\phi}(\mathbf{u})}{\partial \mathbf{u}} \in \mathbb{R}^{q\times D}. \tag{4.24}$$

For the considered product state encodings it is straightforward to check that $\boldsymbol{\phi}(\mathbf{u})$ is injective on the respective domains in $\mathbb{R}^D$ (possibly after removing one endpoint). Hence, local injectivity of $\mathbf{F}(\mathbf{u})$ is determined by whether R is injective on the tangent space range $J_{\boldsymbol{\phi}}(\mathbf{u})$, i.e. whether $\ker R \cap \operatorname{im} J_{\boldsymbol{\phi}}(\mathbf{u}) = \{0\}$. In a generic setting, this condition is satisfied (at fixed $\mathbf{u}$) if $\operatorname{rank} R \geq D$, which is the case for adequate measurement schemes and sufficiently scrambling dynamics.

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. 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, 44, 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., [41, 92, 93]). An alternative to increasing randomness is to devise task-aligned reservoir structures and measurement schemes [2, 32, 43, 50, 94]. Knowing the connection between the readout and the encoding-generated features via the PTM framework can thus inform feature engineering strategies to enhance expressivity.

V. INTERPRETABLE LEARNING OF DYNAMICAL SYSTEMS

A. Time series modeling with QELMs

The classical representation (4.2) allows us to identify the surrogate model a QELM learns during the training process. Specifically, for the product state encodings considered in Section II B, the function library $\boldsymbol{\phi} : \mathbb{R}^D \to \mathbb{R}^q$ 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

$$\text{dynamical system:} \qquad \dot{\mathbf{u}} = \mathbf{g}(\mathbf{u}), \qquad \mathbf{u}_{t+1} = \boldsymbol{\psi}_{\mathbf{g}}(\mathbf{u}_t), \tag{5.1}$$

where $\boldsymbol{\psi}_{\mathbf{g}}(\mathbf{u}) = \mathbf{u} + \Delta t \mathbf{g}(\mathbf{u}) + \mathcal{O}(\Delta t^2)$ 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:

$$\boldsymbol{\psi}_{\mathbf{g}}(\mathbf{u}) = \exp(\Delta t\, \mathcal{L}_{\mathbf{g}})\, \mathbf{u} = \sum_{k=0}^{\infty} \frac{\Delta t^k}{k!} (\mathcal{L}_{\mathbf{g}}^k \mathbf{u}), \qquad (\mathcal{L}_{\mathbf{g}} f)(\mathbf{x}) = \mathbf{g}(\mathbf{x}) \cdot \nabla f(\mathbf{x}). \tag{5.2}$$

The non-integrable 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, 95].

If the dynamics is Markovian, it can be modeled solely with a NG-RC or QELM based on an instantaneous feature map (see [32] for examples). Accordingly, we formulate the training problem of the QELM (4.2) as

$$\text{QELM:} \qquad \mathbf{u}_{t+1} = \mathbf{f}(\mathbf{u}_t) = W^\top R \boldsymbol{\phi}(\mathbf{u}_t) \tag{5.3}$$

and seek to learn an approximation to the flow map $\boldsymbol{\psi}_{\mathbf{g}}$ 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 $W = [\mathbf{w}_1, \ldots, \mathbf{w}_D] \in \mathbb{R}^{B \times D}$.

We use the Lorenz-63 and Halvorsen models as benchmarks ($D = 3$ dimensions) [96]:

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

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:

$$\begin{aligned} \boldsymbol{\Phi}_{\text{amp}}^{n=3}(\mathbf{u}) = \Big\{ & 1, 2\sqrt{(1-x)x}, 2x-1, 2\sqrt{(1-y)y}, 2(2x-1)\sqrt{-((y-1)y)}, 4\sqrt{(x-1)x(y-1)y}, 2y-1, \\ & 2\sqrt{-((x-1)x)}(2y-1), (2x-1)(2y-1), 2\sqrt{(1-z)z}, 2(2x-1)\sqrt{-((z-1)z)}, \\ & 2(2y-1)\sqrt{-((z-1)z)}, 2(2x-1)(2y-1)\sqrt{-((z-1)z)}, 4\sqrt{(x-1)x(z-1)z}, \\ & 4(2y-1)\sqrt{(x-1)x(z-1)z}, 4\sqrt{(y-1)y(z-1)z}, 4(2x-1)\sqrt{(y-1)y(z-1)z}, \\ & 8\sqrt{-((x-1)x(y-1)y(z-1)z)}, 2z-1, 2\sqrt{-((x-1)x)}(2z-1), (2x-1)(2z-1), \\ & 2\sqrt{-((y-1)y)}(2z-1), 2(2x-1)\sqrt{-((y-1)y)}(2z-1), 4(2z-1)\sqrt{(x-1)x(y-1)y}, \\ & (2y-1)(2z-1), 2\sqrt{-((x-1)x)}(2y-1)(2z-1), (2x-1)(2y-1)(2z-1) \Big\} \end{aligned} \tag{5.5}$$

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 $\mathbf{u} \to \mathbf{u}' = \boldsymbol{\alpha} \odot (\mathbf{u} - \mathbf{m})$ with $\boldsymbol{\alpha}, \mathbf{m} \in \mathbb{R}^3$ to the state (we typically use the same α for all coordinates). This leads to the rescaled Lorenz-63 model:

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

and the rescaled Halvorsen model:

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

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. (4.6)]. We generate trajectories using a 5th-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. (4.2) the learned weights $\hat{\mathbf{w}}$.

The RMS training error

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

measures one-step accuracy of the trained predictor $\hat{\mathbf{f}}$. As a measure of multi-step 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

$$T_{\text{fch}}/\Delta t = \inf \{ t : \|\mathbf{u}_t - \hat{\mathbf{u}}_t\| > \sigma \}, \tag{5.9}$$

where $\bar{\mathbf{u}} = \frac{1}{P}\sum_{t=1}^{P} \mathbf{u}_t$ and $\sigma^2 = \frac{1}{P}\sum_{t=1}^{P} \|\bar{\mathbf{u}} - \mathbf{u}_t\|^2$ 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 $T_L = 1/\lambda_{\max}$.

Results of training a QELM in various setups are summarized in Table III. As exemplified for the rescaled *Lorenz-63* model in Fig. 10, we find exact agreement between the autonomously generated prediction of the trained QELM and the iterated classical 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) [97].

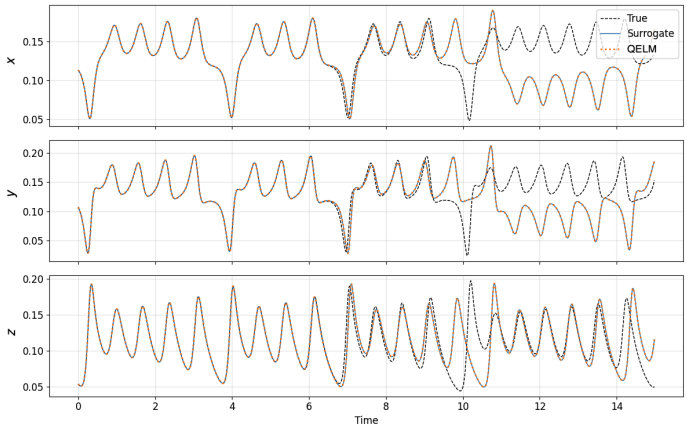


Figure 10. QELM vs. its classical representation: The forecast obtained from a QELM [Eq. (2.1)] trained on the rescaled Lorenz-63 model is matched exactly by the forecast of its iterated classical (surrogate) representation [Eq. (4.2)]. Both eventually diverge from the true trajectory due to the Lyapunov 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 $\alpha_\kappa = 0.004$ and $\mathbf{m} = [-30, -30, -5]$.

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 III). Alternatively, a similar performance can also be achieved when using the amplitude

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.	...+ 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.	...+ 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

Table III. 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. The column 'type' indicates whether we use the standard [Eq. (5.4)] or rescaled dynamical system. In the latter case, α gives the rescaling factor in Eqs. (5.6) and (5.7) (same for all coordinates).

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 [69, 72, 98]. 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. (4.5)] can serve as an indicator of the training accuracy, as illustrated in Fig. 11. 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], achieving $\mathrm{rank}(G) \simeq q$ makes essentially all Pauli features available to the readout and can thus improve performance. Indeed, the QELM in Fig. 11(b) achieves training error and forecast horizon essentially identical to the ones reported in Table III obtained with a complete Pauli observable set. By contrast, Fig. 11(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 $\mathrm{rank}(G) \simeq \mathrm{rank}(R) \ll q$ [see Eq. (4.8) and Fig. 4]. This effect is most pronounced when using the specific TFIM variant $H^{\text{xx-z}}_{\text{TFIM}}$ [Eq. (2.5b)] with Z-type observables. Numerical experiments confirm the close relation (4.8) between the readout and PTM rank.

C. Interpretability

When training a NG-RC with an instantaneous feature library on a Markovian dynamical system, the model learns an approximation of the flow map [72, 99–102]. Libraries $\boldsymbol{\Phi}$ of the form (5.5), however, do not render directly interpretable models due to the structural mismatch with the ground truth model [103]. In the present teacher-forcing setup, a way to render the model interpretable is

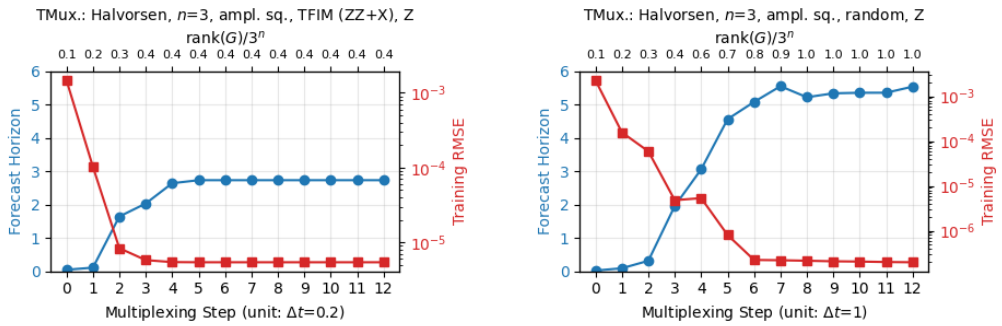


Figure 11. Prediction performance (forecast horizon and RMS training error) under temporal multiplexing of a QELM trained on the Halvorsen model (5.7). We use amplitude encoding (2.9) with $n = 3$ qubits, single-qubit Pauli-Z observables, and (a) the TFIM Hamiltonian $H^{\text{zz-x}}_{\text{TFIM}}$ [Eq. (2.5a)] or (b) a random Hamiltonian. The top horizontal axis shows the numerically obtained rank of the readout matrix G [Eq. (4.5)] relative to the maximum achievable rank $q = 3^n$ [Eq. (4.7)]. Forecast horizon is given in units of the Lyapunov time T_L; no additional features are added to the readout. For the TFIM variant $H^{\text{xx-z}}_{\text{TFIM}}$ [Eq. (2.5b)], we obtain a maximum rank of $\mathrm{rank}(G)/3^n \simeq 0.22$, RMSE $\epsilon_{\text{train}} \simeq 7 \times 10^{-4}$, and forecast horizon $T_{\text{fch}} \simeq 0$.

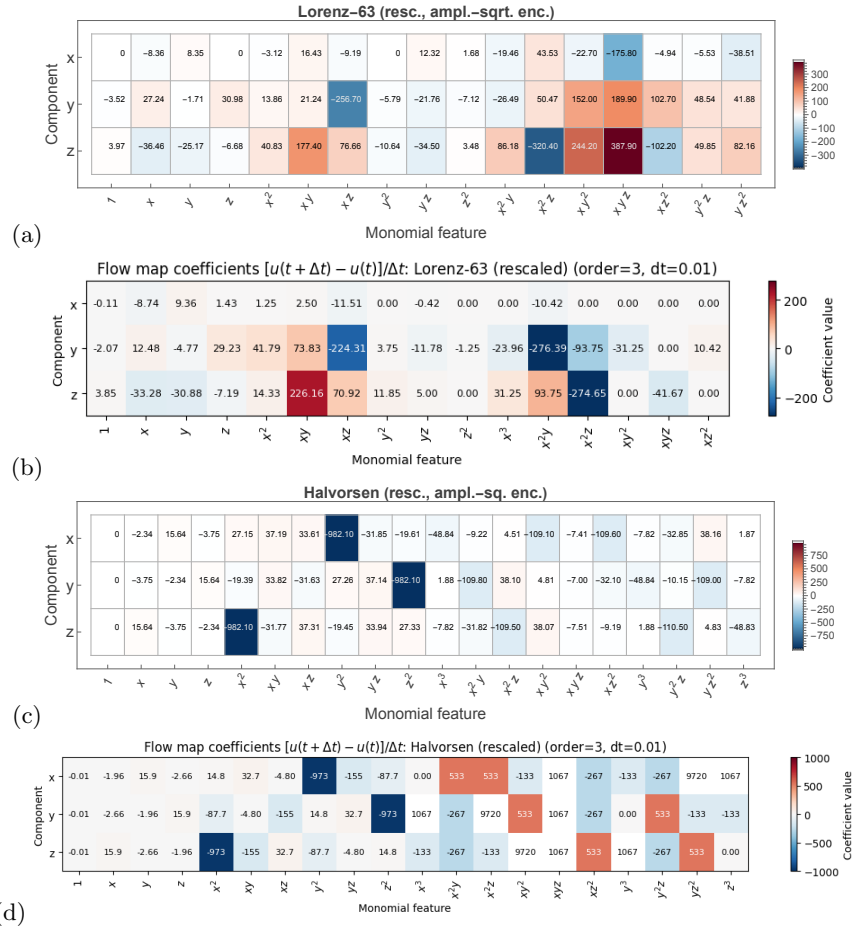


Figure 12. Illustration of the flow map learned by a QELM, obtained from the Taylor-approximation approach Eq. (5.11). The QELM is trained on (a) the Lorenz-63 model (5.6) and (c) the Halvorsen model (5.7), using the *amplitude* encoding variants (2.8) and (2.9), respectively. The rescaling parameters are here $\mathbf{m} = [-30, -30, 5], \alpha_\kappa = 0.004$ (Lorenz-63) and $\mathbf{m} = [-10, -10, -10], \alpha_\kappa = 0.001$ (Halvorsen). The corresponding true flow maps $\boldsymbol{\psi}_\mathbf{g}$ [Eq. (5.1)], expanded to third order in the state variables, are shown in (b,d) for comparison (terms identically zero are omitted).

thus to express the feature map $\boldsymbol{\phi}$ in terms of a N-dimensional library $\mathcal{B}$ of functions that make up the flow map $\boldsymbol{\psi}_\mathbf{g}$ [Eq. (5.1)],

$$\boldsymbol{\phi}(\mathbf{u}) \simeq K\mathcal{B}(\mathbf{u}), \tag{5.10}$$

and insert this into the learned model (4.2),

$$\hat{\mathbf{f}}(\mathbf{u}) = \hat{W}^\top RK\mathcal{B}(\mathbf{u}) \simeq \boldsymbol{\psi}_{\text{approx}}(\mathbf{u}). \tag{5.11}$$

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

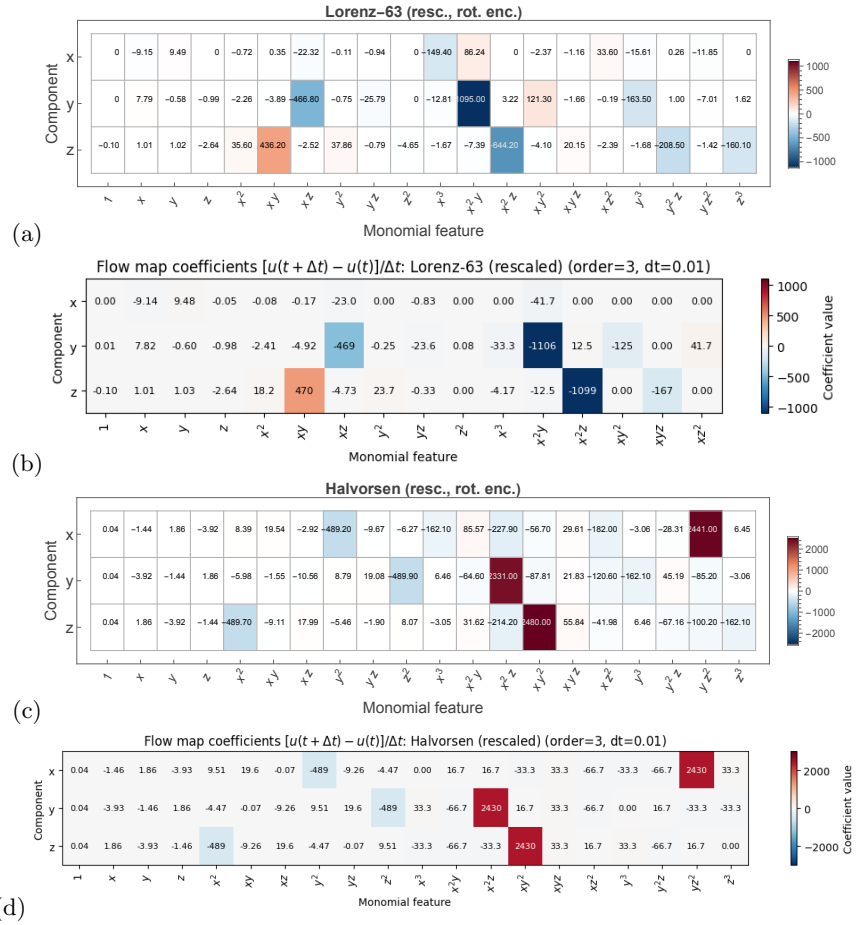


Figure 13. Illustration of the flow map learned by a QELM, obtained from the Taylor-approximation approach Eq. (5.11). The QELM uses *rotational* encoding [Eq. (2.10)] and is trained on (a) the Lorenz-63 model (5.6) and (c) the Halvorsen model (5.7). The rescaling parameters are here $\mathbf{m} = [1, 1, 20], \alpha_\kappa = 0.002$ (Lorenz-63) and $\mathbf{m} = [-3, -3, -3], \alpha_\kappa = 0.002$ (Halvorsen). The corresponding true flow maps $\boldsymbol{\psi}_\mathbf{g}$ [Eq. (5.1)], expanded to third order in the state variables, are shown in (b,d) for comparison. The coefficients differ from Fig. 12 due to the different expansion points.

While there are several methods to determine K (such as collocation), we perform here a Taylor expansion of $\boldsymbol{\phi}(\mathbf{u})$ 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 $\boldsymbol{\phi}$ [see Section IV D 2]. However, due to the tensor product structure of $\boldsymbol{\phi}$ [see Eq. (3.11)] and the fact, that the derivatives of the single-qubit feature vector lie in a 2-dimensional space (vanishing Pauli-Y), K typically has here not full column rank when $N < q$. A consequence of this is that only a rank(K)-dimensional subspace of the library $\mathcal{B}$ is accessible through the encoding features $\boldsymbol{\phi}$ and that hence some monomials in $\mathcal{B}$ are not independently identifiable. Nevertheless, deviations are small for the settings considered here [104].

Figure 12 shows the learned flow maps, 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 coefficients, their influence on the overall training error is still small, since they are multiplied by increasing powers of Δt.

As illustrated in Fig. 13, also in the case of rotational encoding [Eq. (2.10)] the QELM forms linear combinations of the feature library $\boldsymbol{\phi}(\mathbf{u})$ so that its Taylor expansion approximately matches the true flow map $\boldsymbol{\psi}_{\mathbf{g}}$ [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

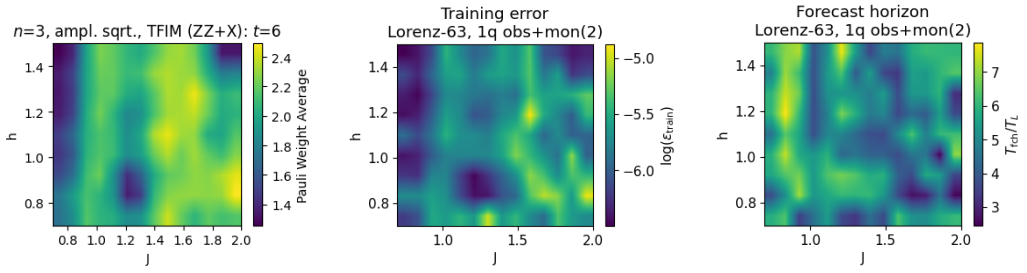
We have employed the Pauli transfer matrix (PTM) formalism to describe a memoryless quantum reservoir computer (quantum extreme learning machine, QELM) with initial-state encoding and data-independent reservoir dynamics. Writing states and channels in the Pauli basis (or any other orthogonal operator basis) turns the QELM into a sequence of real linear maps $T_{\mathcal{E}_j}$ acting on the vector of Pauli features $\boldsymbol{\phi}(\mathbf{u})$ representing the encoded state $\rho_{\mathrm{enc}}(\mathbf{u})$. These features carry the full dependence of the QELM on the encoded input $\mathbf{u}$ and determine the nonlinearities that the model can possibly express. Unitaries appear as orthogonal rotations on the traceless sector of $\boldsymbol{\phi}(\mathbf{u})$, while general CPTP maps act affinely. Training a QELM reduces to classical linear regression over the components $\phi_k(\mathbf{u})$ and can be framed as a decoding problem of features mixed by quantum channels.

While the PTM formalism has been previously been applied to QRCs [28, 36, 46, 49], we focused here on its implications for time-series prediction and system identification with QELMs. First, this approach allows one to cast optimization of a QELM as an inversion problem: knowing the target features can point to an ideal encoding scheme as well as to constraints on the decodability paths. This goes beyond universal approximation theorems [6, 26–30] and allows one to gain more control over the training performance.

Second, we have determined a classical and interpretable representation of the QELM in terms of a nonlinear vector (auto-)regression model (or "next-generation" RC [72]). The question of classical representations of QML models has been explored in numerous works [50, 105–107]. These insights can help to construct reduced-order models in cases where the exact data-generating process is unknown and address the problem of black-box system identification [95, 108].

Under 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 parition the reachable Krylov space (see Appendix F), 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 E).

Numerous previous studies noted that simply enhancing the quantum nature of a QML model often does not improve its performance on practical datasets [109–115]. The present study adds further perspectives on this finding: first, a reservoir unitary mixes features and reversing this potentially detrimental effect requires complex measurement schemes (i.e., high-order Pauli observables, optimized pre-measurement unitaries, or temporal multiplexing schemes). This insight is



(a) (b) (c)

Figure 14. Pauli weight average (a) of Heisenberg-evolved measurement observables can predict the training error (b) and forecast horizon (c) of a QELM. We use a TFIM Hamiltonian (varying coupling parameters J, h, fixed $t = 6$) with $n = 3$ qubits and amplitude encoding (2.8) (no temporal multiplexing). The Pauli weight average is data-independent, while the QELM in (b,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.

supported directly by comparing the Pauli weight average of single-qubit observables to its training accuracy: as illustrated in Fig. 14, for a minimal QELM without temporal multiplexing, large Pauli weight average strongly correlates with high training error and short forecast horizon on a time-series task. Consistent with [47, 68], what matters for task performance is therefore not just how a quantum system evolves, but what part of the Heisenberg/Krylov operator space is actually accessible under the chosen observables and temporal multiplexing. In conventional QML models, instead, feature mixing must be controlled by training of variational gates.

Moreover, learning of the flow map of a dynamical system benefits from a structurally aligned feature library [99, 102]. If the encoding scheme does not provide this, the resulting surrogate model can exhibit significant errors (cf. Fig. 12). In the case of chaotic systems, such errors are magnified by the Lyapunov instability and yield short forecast horizons T_{fch} [116]. Beside optimizing the measurement process [32, 45] and carefully structuring the Hamiltonian [43, 50, 94], the present work thus points to the encoding as a central turning knob for performance of QELMs and, more generally, QML models (see, e.g., [117, 118] for further insights along this line). How this surrogate-model argument gets modified when training on datasets with unknown ground truths is an interesting avenue for future work [119].

The 4^n-dimensional space of Pauli features [120] can be compared to the exponential frequency content of QML models that employ gate-based rotational encoding [51, 52, 66]. 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.

Symbol	Meaning
$\mathcal{H}$	Hilbert space
$\mathrm{Herm}(\mathcal{H})$	Space of Hermitian operators on $\mathcal{H}$
n	Number of qubits
d	Hilbert space dimension ($d = 2^n$)
D	Input data dimension
P	Number of training samples
m	Number of measurement operators
L	Number of temporal multiplexing steps
B	Total number of measurements (e.g., $B = mL$)
q	Number of nonzero Pauli features ($q \leq d^2$)
$\mathbf{x}, \mathbf{u}$	Input data ($\mathbf{x}, \mathbf{u} \in \mathbb{R}^D$)
y	Target value ($y \in \mathbb{R}$)
$\rho(\mathbf{x})$	Density matrix of the quantum state (feature map)
M_k	Measurement operators
P_k	n-qubit Pauli operators ($k = 0, \ldots, d^2 - 1$)
σ_j	single-site Pauli matrices ($\mathbb{I}$, X, Y, Z)
U	Unitary evolution operator
$T_{\mathcal{E}}$	Pauli transfer matrix (channel $\mathcal{E}$, $T_{\mathcal{E}} \in \mathbb{R}^{d^2 \times d^2}$)
V	Unitary PTM ($V \in \mathrm{SO}(d^2)$)
R	Effective PTM (temporal multiplexing) ($R \in \mathbb{R}^{B \times d^2}$)
$\mathcal{S}$	Selected observable set ($\lvert\mathcal{S}\rvert = m$)
S	Selector matrix corresponding to $\mathcal{S}$
$\boldsymbol{\phi}(\mathbf{x})$	Pauli feature vector ($\boldsymbol{\phi}(\mathbf{x}) \in \mathbb{R}^{d^2}$)
$\mathbf{F}(\mathbf{x})$	Readout feature vector ($\mathbf{F}(\mathbf{x}) = R\boldsymbol{\phi}(\mathbf{x})$)
$\mathbf{w}$	Readout weights ($\mathbf{w} \in \mathbb{R}^B$)
λ	Regularization parameter
$f(\mathbf{x})$	QELM output function ($f(\mathbf{x}) = \mathbf{w}^\top \mathbf{F}(\mathbf{x})$)
ν	Pauli weight
γ	Noise parameter

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

ACKNOWLEDGMENTS

This project was made possible by the DLR Quantum Computing Initiative and the Federal Ministry for Research, Technology and Space; https://qci.dlr.de/NeMoQC

Appendix A: Kernel representation

Training over a complete basis of measurement operators $\{M_k\}_{k=1}^{d^2}$ results in a quantum kernel form of the QELM [32]:

$$f(\mathbf{x}) = \sum_{j=1}^{P} \alpha_j^* K(\mathbf{x}, \mathbf{x}_j) = \mathrm{tr}(M^* \rho(\mathbf{x})), \qquad K(\mathbf{x}, \mathbf{x}') = \mathrm{tr}\big(\rho(\mathbf{x})\rho(\mathbf{x}')\big) \tag{A1}$$

For the least-squares loss, the optimal parameters have the explicit representation $\boldsymbol{\alpha}^* = (\mathbf{K} + \lambda \mathbb{I})^{-1}\mathbf{y}$, with the Gram matrix $K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$ and the sample vector $\mathbf{y} \in \mathbb{R}^P$. If the reservoir consists only of a data-independent unitary, it drops out from Eq. (A1),

$$K(\mathbf{x}, \mathbf{x}') = \mathrm{tr}\big(\rho_{\mathrm{enc}}(\mathbf{x})\rho_{\mathrm{enc}}(\mathbf{x}')\big), \qquad \text{(unitary reservoir)} \tag{A2}$$

showing that such a QELM only depends on the encoding state $\rho_{\mathrm{enc}}(\mathbf{x})$. An optimal measurement operator can be identified from Eq. (A1) as

$$M^* = \sum_{j=1}^{P} \alpha_j^* \rho(\mathbf{x}_j). \tag{A3}$$

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

Appendix B: Exact feature isolation criterion: null-space condition

Equation (4.16) is equivalent to requiring $\mathbf{w}^\top R = \omega e_r^\top$ with some scalar weight ω, i.e., $e_r^\top \in \mathbb{R}^q$ belongs to the row-space of the effective PTM $R \in \mathbb{R}^{m \times q}$. This, in turn, is equivalent to

$$\mathbf{e}_r \perp \mathrm{null}(R), \qquad (r\text{-th feature isolatable}) \tag{B1}$$

i.e., the r-th component of every (basis) vector of null(R) must be zero. If Eq. (B1) is met, the solution for $\mathbf{w}$ is provided by the pseudoinverse R^+ via

$$\mathbf{w}^\top = \omega \mathbf{e}_r^\top R^+ \tag{B2}$$

up to a term $\propto h(\mathbb{I}_m - RR^+)$ with an arbitrary vector $h \in \mathbb{R}^{1\times m}$ accounting for the non-uniqueness of the solution if R does not have full row rank. Without this term, Eq. (B2) represents the minimum-norm solution.

Consider the case where $R \in \mathbb{R}^{m\times q}$ is dense and *quasi-random*, 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 null(R) is essentially zero, such that no features are isolatable. Once $m \geq q$, R has full column rank, hence null$(R) = 0$ and the null-space condition in Eq. (B1) is trivially fulfilled, implying that all q features become instantly isolatable 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, row(R) simply consists of a subset of signed standard basis vectors, hence rank(R) features are isolatable even if R is

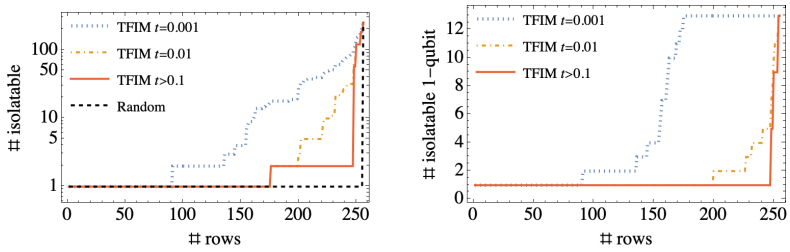


Figure 15. Null-space criterion (B1): Number of isolatable features as a function of the number of observables (=rows of the PTM) m, for (a) the TFIM at various evolution times t and (b) a random unitary, for $n = 4$ qubits ($d^2 = 256$). Panel (a) shows the total number of isolatable features, while panel (b) shows the number of isolatable 1-qubit features. In contrast to statistical decodability (cf. Appendix C), isolatability is an exact geometric criterion independent of the data.

wide. For unitaries based on *structured Hamiltonians* or at small evolution times, one may expect Eq. (B1) 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. 15, where the behavior of the feature isolation criterion (B1) is illustrated for various unitaries as a function of the number of observables m (rows of R).

Appendix C: Decodability scores and coefficient of determination

The decodability score γ_r^2 [Eq. (4.18)] tells to which extent a feature can be geometrically reconstructed from the readout. It disregards statistical correlations and effects of noise, i.e., features with small singular values in the effective PTM R could be considered decodable, although it would lead to huge weight norms. A straightforward way to limit noise amplification is to replace the pseudoinverse by its Tikhonov regularized form

$$\gamma_{r,\lambda}^2 = (R_\lambda^+ R)_{rr} = \left[(R^\top R + \lambda \mathbb{I})^{-1} R^\top R\right]_{rr} = \sum_{i=1}^{s} \frac{\sigma_i^2}{\sigma_i^2 + \lambda} v_{ri}^2, \qquad \gamma_{r,\lambda\to 0}^2 \to \gamma_r^2 \tag{C1}$$

where $\sigma_{i=1,\ldots,s}$ are the singular values of R and $\mathbf{v}_i$ the right-singular vectors. Directions with $\sigma_i^2 \ll \lambda$ contribute little to $\gamma_{r,\lambda}^2$ and are thus considered undecodable.

To account for data correlations within a refined $\mathcal{R}^2$-score, consider the regression solution in Section IV A, assume all quantities are centered, and define the target alignment vector

$$\mathbf{b}_k := \mathbb{E}[\boldsymbol{\phi}(\mathbf{u}) y_k(\mathbf{u})] \in \mathbb{R}^q. \tag{C2}$$

Using Eq. (4.3), the mean-squared error satisfies

$$\mathbb{E}_{\mathbf{u}}\left[(y_k - f)^2\right] = \mathrm{Var}_{\mathbf{u}}(y_k) - (R\mathbf{b}_k)^\top \hat{\mathbf{w}}. \tag{C3}$$

Plugging this into the definition in Eq. (4.21) gives the expression for the coefficient of determination (in the unregularized case $\lambda = 0$)

$$\mathcal{R}^2[y_k] = \frac{\mathbf{b}_k^\top R^\top (RCR^\top)^+ R\mathbf{b}_k}{\mathrm{Var}(y_k)}, \tag{C4}$$

where $C = \mathbb{E}[\boldsymbol{\phi}\boldsymbol{\phi}^\top]$ is the feature covariance matrix. All quantities herein are to be averaged over the input distribution of $\mathbf{u}$.

In the special where the targets are the Pauli features, $y_r(\mathbf{u}) = \phi_r(\mathbf{u})$, and under the additional assumption of orthonormality (whitened features),

$$\mathbb{E}[y_r y_{r'}] = \delta_{rr'} \implies C = \mathbb{I}_q, \quad \mathbf{b}_r = \mathbf{e}_r, \quad \mathrm{Var}(y_r) = 1, \tag{C5}$$

the coefficient of determination $\mathcal{R}^2[y_r]$ in Eq. (C4) reduces to the decodability score

$$\mathcal{R}^2[y_r] = \mathbf{b}_r^\top R^\top (RR^\top)^+ R\mathbf{b}_r = \mathbf{e}_r^\top (R^+R)\mathbf{e}_r = \gamma_r^2. \tag{C6}$$

In the generic non-orthogonal case, this simple relation breaks down and $\mathcal{R}^2$ depends also on the covariance C and target-feature alignment. Then, $\mathcal{R}^2$ measures the extent to which the target lies in the decodable subspace. As a consequence, it is then possible to have $\gamma_r^2 = 0$ (feature direction not present in row-space of R), yet still $\mathcal{R}^2[y_r] > 0$ if ϕ_r is statistically correlated with other ϕ_j that are present. Hence, $\mathcal{R}^2$ in Eq. (C4) generally measures predictability from correlations and not just isolatability of a feature.

Appendix D: Pauli-Transfer Matrix Formalism: Further Details

We provide here some further details on the Pauli-Transfer Matrix (PTM) formalism for QELMs [28, 36, 49], including some geometrical intuitions [121] and discussion of noisy channels.

When ordering the basis with identity first, we have $\phi = (\phi_0, \mathbf{s})$, where $\phi_0 = \mathrm{tr}(\rho) = 1$ 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 ($\mathrm{tr}(\rho^2) = 1$) 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

$$T_{U,\alpha\beta} = \frac{1}{2^n}\,\mathrm{tr}\big(P_\alpha\, U P_\beta U^\dagger\big), \tag{D1}$$

Trace preservation and unitality imply the form

$$T_U = \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \Omega_U \end{pmatrix}, \tag{D2}$$

i.e., the traceless coefficients undergo a real orthogonal rotation with $\Omega_U \in \mathrm{SO}(d^2-1)$. 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 $h/J \to \infty$ [Eq. (2.5)]. This implies a Kronecker-product structure of the unitary PTM (D1) (and similarly for any other product quantum channel),

$$T_U(t) = V^{(1)}(t) \otimes \cdots \otimes V^{(n)}(t), \tag{D3}$$

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

$$\boldsymbol{\phi}'(\mathbf{x}) = T_U(t)\boldsymbol{\phi}(\mathbf{x}) = \big(V^{(1)}(t)\boldsymbol{\phi}^{(1)}(x_1)\big) \otimes \cdots \otimes \big(V^{(n)}(t)\boldsymbol{\phi}^{(n)}(x_n)\big). \tag{D4}$$

Thus, each Pauli-string feature $\phi_{a_1 \ldots a_n}$ 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 $U_k(t) = e^{-ith_k}$ 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 $\tilde{\boldsymbol{\phi}}^{(k)} \equiv (\phi_x^{(k)}, \phi_y^{(k)}, \phi_z^{(k)})^\top$, any single-qubit unitary induces a SO(3) rotation

$$\tilde{\boldsymbol{\phi}}^{'(k)}(u_k) = R^{(k)}(t)\, \tilde{\boldsymbol{\phi}}^{(k)}(u_k), \qquad \phi_0^{'(k)} = \phi_0^{(k)} = 1, \tag{D5}$$

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

$$U_k(t)^\dagger X U_k(t) = \cos(2t)\, X - \sin(2t)\, Y, \qquad U_k(t)^\dagger Y U_k(t) = \cos(2t)\, Y + \sin(2t)\, X, \tag{D6}$$

so that $\phi_y^{'(k)}(u_k) = \sin(2t)\, \phi_x^{(k)}(u_k)$ and, correspondingly, $\phi'_{\ldots y_k \ldots}(\mathbf{x}) = \sin(2t)\, \phi_{\ldots x_k \ldots}(\mathbf{x})$, 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

Consider a channel $\mathcal{E}$ that is completely positive and trace-preserving (CPTP) and has associated Kraus operators $\{K_\ell\}$. Then

$$T_{\mathcal{E},\alpha\beta} = \frac{1}{2^n} \sum_\ell \mathrm{tr}\big(P_\alpha K_\ell P_\beta K_\ell^\dagger\big), \tag{D7}$$

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

$$T_{\mathcal{E}} = \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{t} & \Omega \end{pmatrix}, \qquad \mathbf{s}' = \mathbf{t} + \Omega \mathbf{s}, \tag{D8}$$

where $\mathbf{t} = \mathbf{0}$ for unital channels. For a single qubit, Ω is a contracting map (singular values ≤ 1). Such a map can improve numerical robustness by damping high-weight Pauli components that are often noise sensitive. Non-unitality 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 non-unitary *noise channels* [122–124]:

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

$$T_{\mathrm{dep}}(\lambda) = \mathrm{diag}(1, \lambda, \lambda, \lambda). \tag{D9}$$

  For a global isotropic depolarizing channel on dimension $d = 2^n$, $T_{\mathrm{dep}}^{(n)}(\lambda) = \mathrm{diag}\big(1, \lambda I_{d^2-1}\big)$.

- Z-dephasing with probability p:

$$T_{\mathrm{dephZ}}(p) = \mathrm{diag}(1, 1-2p, 1-2p, 1). \tag{D10}$$

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

$$T_{\mathrm{AD}}(\gamma) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-\gamma} & 0 & 0 \\ 0 & 0 & \sqrt{1-\gamma} & 0 \\ z_*\gamma & 0 & 0 & 1-\gamma \end{pmatrix}. \tag{D11}$$

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, 125]. Since the PTMs (D9) and (D10) are Pauli-diagonal, they do not lead to Pauli spreading and hence do not enlarge the accessible feature subspace. Amplitude damping [Section D 2] 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 $\mathcal{D}^{(n)} = \otimes_{i=1}^{n} \mathcal{D}_\lambda$ [see Eq. (D9)], any Pauli string P of weight $\nu(P)$ is an eigenoperator with

$$\mathcal{D}^{(n)}(P) = \lambda^{\nu(P)} P. \tag{D12}$$

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 E: TFIM symmetries and their impact on temporal multiplexing

We now specialize the general discussion of Section IV B 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/Majorana formalism for the Ising chain [126, 127]. Once the Lieb–Robinson lightcone 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

For the TFIM in the form Eq. (2.5b), we introduce Jordan–Wigner Majorana quasiparticle operators

$$\gamma_{2j-1} = \Big(\prod_{m<j} Z_m\Big) X_j, \qquad \gamma_{2j} = \Big(\prod_{m<j} Z_m\Big) Y_j, \qquad j = 1, \ldots, n, \tag{E1}$$

resulting in

$$H_{\mathrm{TFIM}}^{\text{xx-z}} = iJ\sum_{j=1}^{n-1}\gamma_{2j}\gamma_{2j+1} + ih\sum_{j=1}^{n}\gamma_{2j-1}\gamma_{2j}. \tag{E2}$$

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

$$\gamma_a^{(\mathrm{H})}(t) = \sum_{b=1}^{2n} R_{ab}(t)\,\gamma_b, \qquad R(t)\in \mathrm{SO}(2n), \tag{E3}$$

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

$$\mathcal{V}_r = \mathrm{span}\{\gamma_{a_1}\cdots\gamma_{a_r} : 1\le a_1<\cdots<a_r\le 2n\} \tag{E4}$$

is invariant and has dimension

$$\dim \mathcal{V}_r = \binom{2n}{r}. \tag{E5}$$

The observables relevant for the QELM then fall into simple sectors:

$$Z_j = -i\gamma_{2j-1}\gamma_{2j}\in\mathcal{V}_2, \qquad Z_iZ_j\in\mathcal{V}_4 \quad (i\neq j), \tag{E6}$$

while

$$X_j = \Big[\prod_{m<j}(-i\gamma_{2m-1}\gamma_{2m})\Big]\gamma_{2j-1}, \qquad Y_j = \Big[\prod_{m<j}(-i\gamma_{2m-1}\gamma_{2m})\Big]\gamma_{2j}, \tag{E7}$$

so that $X_j, Y_j\in\mathcal{V}_{2j-1}$. Thus Z-type observables under $H_{\mathrm{TFIM}}^{\text{xx-z}}$ stay in low-degree Gaussian sectors, whereas X_j and Y_j occupy odd sectors whose degree grows with the site index. In particular, for a bulk site j, a single X_j or Y_j already 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 $H_{\mathrm{TFIM}}^{\text{xx-z}}$ is restricted to

$$Z_j^{(\mathrm{H})}(t) = \sum_{p=1}^{n} a_p(t)\,Z_p + \sum_{1\le p<q\le n}\ \sum_{\alpha,\beta\in\{X,Y\}} b_{pq}^{\alpha\beta}(t)\,\alpha_p\,Z_{p+1}\cdots Z_{q-1}\,\beta_q, \tag{E8}$$

i.e., to strings with two X/Y endpoints and a Z-string in between. Hence, after full spatial spreading the number of Pauli strings remains $\sim\mathcal{O}(n^2)$. Likewise, $Z_iZ_j^{(\mathrm{H})}(t)$ remains in the quartic sector and becomes a sum of products of two such interval strings. These strings have at most four X/Y endpoints 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):

$$O^{(\mathrm{H})}(t)\big|_{\text{zz-x}} = U_{\mathrm{H}}^{\dagger}\Big[(U_{\mathrm{H}}OU_{\mathrm{H}}^{\dagger})^{(\mathrm{H})}(t)\big|_{\text{xx-z}}\Big]U_{\mathrm{H}}. \tag{E9}$$

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

$$Z_j \;\longleftrightarrow\; X_j\in\mathcal{V}_{2j-1}, \qquad Z_iZ_j \;\longleftrightarrow\; X_iX_j\in\mathcal{V}_{2|i-j|} \quad (i\neq j). \tag{E10}$$

Hence $Z_j^{(\mathrm{H})}(t)$ under $H_{\mathrm{TFIM}}^{\text{zz-x}}$ can occupy exponentially large sectors for bulk j, and $Z_iZ_j^{(\mathrm{H})}(t)$ does the same whenever $|i-j| \sim \mathcal{O}(n)$. Z-type observables are low-complexity under $H_{\mathrm{TFIM}}^{\text{xx-z}}$, but high-complexity under $H_{\mathrm{TFIM}}^{\text{zz-x}}$. 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.

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. (4.1) evaluated for L multiplexing steps and $\mathcal{L}(\cdot) = i[H, \cdot]$. For generic couplings $J, h \neq 0$ and generic sampling times $\{t_\ell\}$, the large-L rank is the dimension of the Heisenberg orbit, equivalently of the Krylov span [47]:

$$r_\infty(H, \mathcal{S}) := \lim_{L\to\infty} \operatorname{rank}(R_L) = \dim \operatorname{span}\{e^{t\mathcal{L}}(P) : P \in \mathcal{S},\, t \in \mathbb{R}\} = \dim \operatorname{span}\{\mathcal{L}^m(P) : P \in \mathcal{S},\, m \geq 0\}. \tag{E11}$$

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

$$\operatorname{rank}(R_L) \leq \min(q,\ r_\infty(H, \mathcal{S})), \tag{E12}$$

with q the effective number of nonzero encoded features from Section IV A.

We consider the three observable families

$$\mathcal{S}_Z = \{Z_j\}_{j=1}^n, \qquad \mathcal{S}_{Z,ZZ} = \{Z_j,\ Z_iZ_j\}_{i,j=1}^n, \qquad \mathcal{S}_{XYZ} = \{X_j, Y_j, Z_j\}_{j=1}^n. \tag{E13}$$

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

Using the sector decomposition from Section E 1, the rank estimates for the TFIM $H_{\mathrm{TFIM}}^{\text{xx-z}}$ can be obtained as:

$$r_\infty(H_{\mathrm{TFIM}}^{\text{xx-z}}, \mathcal{S}_Z) \leq \binom{2n}{2} = n(2n-1) \sim \mathcal{O}(n^2), \tag{E14}$$

$$r_\infty(H_{\mathrm{TFIM}}^{\text{xx-z}}, \mathcal{S}_{Z,ZZ}) \leq 1 + \binom{2n}{2} + \binom{2n}{4} \sim \mathcal{O}(n^4), \tag{E15}$$

$$r_\infty(H_{\mathrm{TFIM}}^{\text{xx-z}}, \mathcal{S}_{XYZ}) \leq \sum_{\substack{r=1 \\ r \text{ odd}}}^{2n-1} \binom{2n}{r} + \binom{2n}{2} = 2^{2n-1} + \binom{2n}{2} \sim \mathcal{O}(4^n). \tag{E16}$$

Indeed, a single bulk X_j or Y_j already contributes a sector of size $\binom{2n}{n} \sim 4^n/\sqrt{\pi n}$, so $r_\infty(H_{\mathrm{TFIM}}^{\text{xx-z}}, \mathcal{S}_{XYZ})$ is $\propto 4^n$ up to polynomial factors.

For $H_{\mathrm{TFIM}}^{\text{zz-x}}$, the Hadamard equivalence (E9) maps the observable families to the corresponding X-type families of $H_{\mathrm{TFIM}}^{\text{xx-z}}$. Hence

$$\binom{2n}{n} \lesssim r_\infty(H_{\mathrm{TFIM}}^{\text{zz-x}}, \mathcal{S}_Z) \approx r_\infty(H_{\mathrm{TFIM}}^{\text{zz-x}}, \mathcal{S}_{Z,ZZ}) \sim \mathcal{O}(4^n), \tag{E17}$$

and both are exponential in n. In fact, $\mathcal{S}_Z$ contributes odd Majorana degrees, whereas $\mathcal{S}_{ZZ}$ contributes even degrees $2, 4, \ldots, 2n-2$, so only the top sector $r = 2n$ is generically absent. Finally,

since U_{H} swaps $X \leftrightarrow Z$ and sends $Y \to -Y$, the single-site Pauli family is invariant up to relabeling, and therefore $r_\infty(H^{\text{zz-x}}_{\mathrm{TFIM}}, \mathcal{S}_{XYZ})$ has the same exponential scaling as $r_\infty(H^{\text{xx-z}}_{\mathrm{TFIM}}, \mathcal{S}_{XYZ})$.

In summary, temporal multiplexing of Z-type observables under $H^{\text{xx-z}}_{\mathrm{TFIM}}$ explores only low-degree Gaussian sectors and the effective PTM rank grows polynomially. By contrast, under $H^{\text{zz-x}}_{\mathrm{TFIM}}$, the same Z-type observables map to X-type order operators and access exponentially large sectors, so the rank becomes exponential in n and, for $\mathcal{S}_{Z,ZZ}$, nearly fills the full 4^n-dimensional Pauli operator space. Fine-tuned couplings or commensurate sampling times can reduce these generic ranks.

Appendix F: Influence of symmetries

In Appendix E we derived the asymptotic rank of the temporally multiplexed PTM for the TFIM by exploiting a special operator-space sector decomposition. We will frame here these results in the language of the symmetries of the Liouvillian. Although we phrase the discussion for unitary evolution generated by a Hamiltonian, the same arguments apply to any linear Heisenberg generator whose symmetry projectors commute with the generator.

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 $\langle A, B\rangle = \frac{1}{d}\mathrm{tr}(A^\dagger B)$. A Hamiltonian H induces the Heisenberg evolution $O^{(H)}(t) = e^{t\mathcal{L}}O$ with Liouvillian $\mathcal{L}(O) = i[H, O]$. Given a selected observable family $\mathcal{S}$, the "seed" space of measurement operators is

$$\mathcal{M} = \mathrm{span}\{P_k : k \in \mathcal{S}\} \subseteq \mathcal{O}. \tag{F1}$$

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 Section IV A. We write $\Pi_{\mathcal{F}}$ for the orthogonal projector onto $\mathcal{F}$. For temporal multiplexing at times $t_1, \ldots, t_L$, define

$$\mathcal{W}_L(H, \mathcal{M}; \mathcal{F}) = \mathrm{span}\left\{\Pi_{\mathcal{F}} e^{t_\ell \mathcal{L}}(M) : M \in \mathcal{M},\ \ell = 1, \ldots, L\right\}. \tag{F2}$$

In any orthonormal operator basis whose first q basis elements span $\mathcal{F}$, the effective PTM R_L in Eq. (4.1) is the coordinate matrix of these vectors, hence

$$\mathrm{rank}(R_L) = \dim \mathcal{W}_L(H, \mathcal{M}; \mathcal{F}). \tag{F3}$$

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

$$r_\infty(H, \mathcal{M}) := \lim_{L\to\infty} \mathrm{rank}(R_L) = \dim \mathrm{span}\{e^{t\mathcal{L}}(M) : M \in \mathcal{M},\ t \in \mathbb{R}\}. \tag{F4}$$

For generic sampling times, this coincides with the Krylov dimension [Eq. (E11)],

$$r_\infty(H, \mathcal{M}) = \dim \mathrm{span}\{\mathcal{L}^m(M) : M \in \mathcal{M},\ m \geq 0\}, \tag{F5}$$

2. Invariant sectors induced by symmetries

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

$$[\Gamma, \mathcal{L}] = 0. \tag{F6}$$

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

$$\mathcal{O} = \bigoplus_\alpha \mathcal{O}_\alpha, \qquad \Pi_\alpha \mathcal{O} = \mathcal{O}_\alpha, \qquad \Pi_\alpha \Pi_\beta = \delta_{\alpha\beta} \Pi_\alpha, \qquad \sum_\alpha \Pi_\alpha = \mathbb{I}_{\mathcal{O}}. \tag{F7}$$

Since Equation (F6) implies

$$[\Pi_\alpha, \mathcal{L}] = 0 \qquad \text{for all } \alpha, \tag{F8}$$

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

$$\mathcal{L}(\mathcal{O}_\alpha) \subseteq \mathcal{O}_\alpha, \qquad e^{t\mathcal{L}} \mathcal{O}_\alpha \subseteq \mathcal{O}_\alpha. \tag{F9}$$

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

$$\mathcal{M} = \bigoplus_\alpha \mathcal{M}_\alpha, \qquad \mathcal{F} = \bigoplus_\alpha \mathcal{F}_\alpha, \tag{F10}$$

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,

$$\mathcal{W}_L(H, \mathcal{M}; \mathcal{F}) = \bigoplus_\alpha \mathcal{W}_{L,\alpha}(H, \mathcal{M}_\alpha; \mathcal{F}_\alpha), \tag{F11}$$

where

$$\mathcal{W}_{L,\alpha}(H, \mathcal{M}_\alpha; \mathcal{F}_\alpha) = \operatorname{span}\left\{ \Pi_{\mathcal{F}_\alpha} e^{t_\ell \mathcal{L}}(M) : M \in \mathcal{M}_\alpha,\ \ell = 1, \ldots, L \right\}. \tag{F12}$$

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

$$\operatorname{rank}(R_L) = \sum_\alpha \operatorname{rank}(R_{L,\alpha}), \tag{F13}$$

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

$$r_{\infty,\alpha}(H, \mathcal{M}_\alpha) = \dim \operatorname{span}\{ e^{t\mathcal{L}}(M) : M \in \mathcal{M}_\alpha,\ t \in \mathbb{R} \}, \tag{F14}$$

one has the sectorwise bound

$$\operatorname{rank}(R_{L,\alpha}) \le \min\big(L \dim \mathcal{M}_\alpha,\ \dim \mathcal{F}_\alpha,\ r_{\infty,\alpha}(H, \mathcal{M}_\alpha) \big). \tag{F15}$$

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

$$r_\infty(H, \mathcal{M}; \mathcal{F}) \le \sum_{\alpha : \mathcal{M}_\alpha \neq 0} \min\big(\dim \mathcal{F}_\alpha,\ \dim \mathcal{O}_\alpha \big). \tag{F16}$$

3. Hamiltonian criteria for underexploration

The symmetry decomposition in Eq. (F16) 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

$$\mathcal{L}_\alpha = \sum_{\omega\in\Omega_\alpha} i\omega\,\Pi_{\alpha,\omega}, \qquad \mathcal{O}_{\alpha,\omega} := \Pi_{\alpha,\omega}\mathcal{O}_\alpha, \tag{F17}$$

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 $M_1, \ldots, M_{\dim \mathcal{M}_\alpha}$ 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

$$\Pi_{\alpha,\omega} M_k = \sum_{s=1}^{\dim \mathcal{O}_{\alpha,\omega}} (C_{\alpha,\omega})_{s,k} F_{\alpha,\omega,s}, \tag{F18}$$

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

$$r_{\infty,\alpha}(H, \mathcal{M}_\alpha) = \sum_{\omega\in\Omega_\alpha} \operatorname{rank} C_{\alpha,\omega}. \tag{F19}$$

This result shows that underexploration occurs because several operator directions share the same frequency (l.h.s. of Eq. (F18)) while the seed subspace supplies too few independent directions inside that degenerate eigenspace (r.h.s. of Eq. (F18)).

A direct consequence of Eq. (F19) is the general bound

$$r_{\infty,\alpha}(H, \mathcal{M}_\alpha) \le \sum_{\omega\in\Omega_\alpha} \min\big(\dim \mathcal{O}_{\alpha,\omega}, \dim \mathcal{M}_\alpha\big). \tag{F20}$$

For a generic $(\dim \mathcal{M}_\alpha)$-dimensional choice of seed subspace inside $\mathcal{O}_\alpha$, one expects equality in Eq. (F20). Hence the Hamiltonian alone determines the generic underexploration deficit

$$\dim \mathcal{O}_\alpha - \sum_{\omega\in\Omega_\alpha} \min\big(\dim \mathcal{O}_{\alpha,\omega}, \dim \mathcal{M}_\alpha\big) = \sum_{\omega\in\Omega_\alpha} \max\big(\dim \mathcal{O}_{\alpha,\omega} - \dim \mathcal{M}_\alpha, 0\big). \tag{F21}$$

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:

$$\dim \mathcal{O}_{\alpha,\omega} \le \dim \mathcal{M}_\alpha \qquad \text{for all } \omega \in \Omega_\alpha. \tag{F22}$$

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 form a single frequency sector with $\omega = 0$. 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 [47]. We set thus $\mathcal{F} = \mathcal{O}$ and choose a one-dimensional seed space $\mathcal{M} = \mathrm{span}\{O\}$ for a single observable O. Then the Krylov grade M is the dimension of the minimal Liouvillian Krylov space generated by O [47],

$$M := \dim \mathrm{span}\{O, \mathcal{L}(O), \mathcal{L}^2(O), \ldots\} = r_\infty(H, \mathcal{M}), \tag{F23}$$

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

$$\mathcal{L} = \sum_{\omega \in \Omega} i\omega\, \Pi_\omega, \qquad \Omega = \{\omega_P\}_{P=0}^{N_\omega - 1}, \tag{F24}$$

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

$$\sigma_P := \Pi_{\omega_P} O, \tag{F25}$$

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

$$N_1 := \big|\{P \in \{0, \ldots, N_\omega - 1\} : \sigma_P = 0\}\big| \tag{F26}$$

be the number of vanishing frequency blocks. Since $\dim \mathcal{M} = 1$, each coefficient matrix C_{ω_P} in Eq. (F18) is a single column, and therefore

$$\mathrm{rank}\, C_{\omega_P} = \begin{cases} 1, & \sigma_P = \Pi_{\omega_P} O \neq 0, \\ 0, & \sigma_P = \Pi_{\omega_P} O = 0. \end{cases} \tag{F27}$$

Hence Eqs. (F19) and (F23) give

$$M = r_\infty(H, \mathcal{M}) = \sum_{P=0}^{N_\omega - 1} \mathrm{rank}\, C_{\omega_P} = \big|\{P : \sigma_P \neq 0\}\big| = N_\omega - N_1, \tag{F28}$$

recovering the corresponding theorem in [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 symmetries with $UHU^\dagger = H$. Then the superoperator

$$\Gamma_U(O) = UOU^\dagger \tag{F29}$$

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 superoperator is

$$\Pi_\pm = \frac{1}{2}(\mathbb{I}_\mathcal{O} \pm \Gamma_U). \tag{F30}$$

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

$$\Pi_{qq'}(O) = P_q O P_{q'} \tag{F31}$$

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

$$\mathcal{O} = \bigoplus_{q,q'} P_q \mathcal{O} P_{q'}. \tag{F32}$$

If the measured observables commute with Q, only the diagonal sectors $\mathcal{O}_{qq}$ can be seeded, and therefore

$$r_\infty(H, \mathcal{M}) \leq \sum_q d_q^2. \tag{F33}$$

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 E 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_{TFIM}^{xx-z} in Eq. (E2), the operator space decomposes as

$$\mathcal{O} = \bigoplus_{r=0}^{2n} \mathcal{V}_r, \qquad \dim \mathcal{V}_r = \binom{2n}{r}, \tag{F34}$$

which is the sector structure used in Section E 1. Since the usual observable families are sector-compatible,

$$Z_j \in \mathcal{V}_2, \qquad Z_i Z_j \in \mathcal{V}_4, \qquad X_j, Y_j \in \mathcal{V}_{2j-1}, \tag{F35}$$

the general bound (F16) reproduces the polynomial versus exponential behavior found in Section E 2. In particular, for H_{TFIM}^{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_{TFIM}^{zz-x}, the same Z-type observables are mapped to the high-degree order operators discussed around Eqs. (E9) and (E10).

6. Consequences for PTM analysis and feature decodability

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

$$V(t) = \bigoplus_\alpha V_\alpha(t), \tag{F36}$$

Accordingly, the temporally multiplexed effective PTM (4.1) can be brought into a block form

$$R_L \sim \bigoplus_\alpha R_{L,\alpha}, \tag{F37}$$

and the projector defined in Eq. (4.18) decomposes as

$$R_L^+ R_L \sim \bigoplus_\alpha (R_{L,\alpha}^+ R_{L,\alpha}). \tag{F38}$$

Thus, if a feature lies in a with $\mathcal{M}_\alpha = 0$, then it is symmetry-forbidden and cannot be 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. Applied **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. Applied **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, F. Huber, P. N. Jepsen, N. Jia, T. Karolyshyn, M. Kwon, J. Long, J. Lopatin, A. Lukin, T. Macrì, O. Marković, L. A. Martínez-Martínez, X. Meng, E. Ostroumov, D. Paquette, J. Robinson, P. S. Rodriguez, A. Singh, N. Sinha, H. Thoreen, N. Wan, D. Waxman-Lenz, T. Wong, K.-H. Wu, P. L. S. Lopes, Y. Boger, N. Gemelke, T. Kitagawa, A. Keesling, X. Gao, A. Bylinskii, S. F. Yelin, F. Liu, and S.-T. Wang, Large-scale quantum reservoir learning with an analog quantum computer (2024), arXiv:2407.02553 [quant-ph].

[4] A. Sakurai, A. Hayashi, W. J. Munro, and K. Nemoto, Simple Hamiltonian dynamics is a powerful quantum processing resource (2025), arXiv:2405.14245 [quant-ph].

[5] A. D. Lorenzis, M. P. Casado, M. P. Estarellas, N. L. Gullo, T. Lux, F. Plastina, A. Riera, and J. Settino, Harnessing Quantum Extreme Learning Machines for image classification, Phys. Rev. Applied **23**, 044024 (2025), arXiv:2409.00998 [quant-ph].

[6] J. Chen, H. I. Nurdin, and N. Yamamoto, Temporal Information Processing on Noisy Quantum Computers, Phys. Rev. Applied **14**, 024065 (2020).

[7] Y. Suzuki, Q. Gao, K. C. Pradel, K. Yasuoka, and N. Yamamoto, Natural quantum reservoir computing for temporal information processing, Scientific Reports **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 Information **9**, 16 (2023), arXiv:2205.06809 [quant-ph].

[9] O. Ahmed, F. Tennie, and L. Magri, Prediction of chaotic dynamics and extreme events: A recurrence-free quantum reservoir computing approach (2024), arXiv:2405.03390 [quant-ph].

[10] O. Ahmed, F. Tennie, and L. Magri, Optimal training of finitely-sampled quantum reservoir computers for forecasting of chaotic dynamics (2024), arXiv:2409.01394 [quant-ph].

[11] P. Pfeffer, F. Heyder, and J. Schumacher, Hybrid quantum-classical reservoir computing of thermal convection flow, Phys. Rev. Research **4**, 033176 (2022), arXiv:2204.13951 [quant-ph].

[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. Research **5**, 043242 (2023), arXiv:2307.03053 [physics].

[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 Information **11**, 20 (2025).

[15] D. Zia, L. Innocenti, G. Minati, S. Lorenzo, A. Suprano, R. D. Bartolo, N. Spagnolo, T. Giordani, V. Cimini, G. M. Palma, A. Ferraro, F. Sciarrino, and M. Paternostro, Quantum reservoir computing for photonic entanglement witnessing, Science Advances **11**, 10.1126/sciadv.ady7987 (2025), arXiv:2502.18361 [quant-ph].
[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, Nature Communications **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, Nature Reviews Physics **7**, 174 (2025), arXiv:2405.00781 [quant-ph].
[19] L. Domingo, G. Carlo, and F. Borondo, Taking advantage of noise in quantum reservoir computing (2023), arXiv:2301.06814 [quant-ph].
[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, Scientific Reports **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 (2024), arXiv:2409.07886 [quant-ph].
[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, Journal of Physics: 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, Communications Physics **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, Neuromorphic Computing and Engineering **2**, 014008 (2022).
[26] J. Chen and H. I. Nurdin, Learning Nonlinear Input-Output Maps with Dissipative Quantum Systems, Quantum Information Processing **18**, 198 (2019), arXiv:1901.01653 [quant-ph].
[27] 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), arXiv:2009.00298 [quant-ph].
[28] R. Martínez-Peña and J.-P. Ortega, Quantum reservoir computing in finite dimensions, Phys. Rev. E **107**, 035306 (2023).
[29] F. Monzani and E. Prati, Universality conditions of unified classical and quantum reservoir computing (2024), arXiv:2401.15067 [quant-ph].
[30] L. Gonon and A. Jacquier, Universal Approximation Theorem and Error Bounds for Quantum Neural Networks and Quantum Reservoirs, IEEE Transactions on Neural Networks and Learning Systems **36**, 11355 (2025).
[31] M. Schuld, Supervised quantum machine learning models are kernel methods, Journal of Machine Learning Research 10.1145/3451902.3451903 (2021).
[32] M. Gross and H.-M. Rieser, Kernel-based optimization of measurement operators for quantum reservoir computers (2026), arXiv:2602.14677 [quant-ph].
[33] A. Canatar, E. Peters, C. Pehlevan, and S. Wild, Bandwidth enables generalization in quantum kernel models, Trans. Mach. Learn. Res. **2023** (2022).
[34] 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 (Ciaco - i6doc.com, Bruges, Belgium, 2025) pp. 543–548.
[35] 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), arXiv:2501.10077

[quant-ph].
[36] K. Fujii and K. Nakajima, Quantum reservoir computing: A reservoir approach toward quantum machine learning on near-term quantum devices (2020), arXiv:2011.04890 [quant-ph].
[37] P. Mujal, R. Martínez-Peña, J. Nokkala, J. García-Beni, G. L. Giorgi, M. C. Soriano, and R. Zambrini, Opportunities in Quantum Reservoir Computing and Extreme Learning Machines, Advanced Quantum Technologies **4**, 2100027 (2021).
[38] 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 (2019), arXiv:1806.08321 [quant-ph].
[39] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, Extreme learning machine: Theory and applications, Neurocomputing Neural Networks, **70**, 489 (2006).
[40] A. Rahimi and B. Recht, Random Features for Large-Scale Kernel Machines, in *Advances in Neural Information Processing Systems*, Vol. 20 (Curran Associates, Inc., 2007).
[41] 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).
[42] L. Domingo, G. Carlo, and F. Borondo, Optimal quantum reservoir computing for the noisy intermediate-scale quantum era, Phys. Rev. E **106**, L043301 (2022).
[43] 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), arXiv:2211.07841 [quant-ph].
[44] N. Götting, F. Lohof, and C. Gies, Exploring quantumness in quantum reservoir computing, Phys. Rev. A **108**, 052427 (2023).
[45] L. Innocenti, S. Lorenzo, I. Palmisano, A. Ferraro, M. Paternostro, and G. M. Palma, Potential and limitations of quantum extreme learning machines, Communications Physics **6**, 118 (2023).
[46] S. Kobayashi, Q. H. Tran, and K. Nakajima, Coherence influx is indispensable for quantum reservoir computing (2024), arXiv:2409.12693 [quant-ph].
[47] S. Čindrak, L. Jaurigue, and K. Lüdge, Engineering Quantum Reservoirs through Krylov Complexity, Expressivity and Observability (2025), arXiv:2409.12079 [quant-ph].
[48] N.-E. Schütte, N. Götting, H. Müntinga, M. List, D. Brunner, and C. Gies, Expressivity Limits of Quantum Reservoir Computing (2025), arXiv:2501.15528 [quant-ph].
[49] R. Martínez-Peña and J.-P. Ortega, Input-dependence in quantum reservoir computing, Phys. Rev. E **111**, 065306 (2025).
[50] A. D. Lorenzis, M. P. Casado, N. L. Gullo, T. Lux, F. Plastina, and A. Riera, Entanglement and Classical Simulability in Quantum Extreme Learning Machines (2025), arXiv:2509.06873 [quant-ph].
[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 (2023), arXiv:2206.12105.
[53] W. Xiong, G. Facelli, M. Sahebi, O. Agnel, T. Chotibut, S. Thanasilp, and Z. Holmes, On fundamental aspects of quantum extreme learning machines (2023), arXiv:2312.15124 [quant-ph, stat].
[54] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, Information Processing Capacity of Dynamical Systems, Scientific Reports **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, Cognitive Computation **15**, 1440 (2023).
[56] Z. Shumaylov, P. Zaika, P. Scholl, G. Kutyniok, L. Horesh, and C.-B. Schönlieb, When is a System Discoverable from Data? Discovery Requires Chaos (2025), arXiv:2511.08860 [math].
[57] G. McCaul, J. S. Totero Gongora, W. Otieno, S. Savelev, A. Zagoskin, and A. G. Balanov, Minimal quantum reservoirs with Hamiltonian encoding, Chaos: An Interdisciplinary Journal of Nonlinear Science **35**, 093135 (2025).
[58] C. Weedbrook, S. Pirandola, R. Garcia-Patron, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, Gaussian Quantum Information, Reviews of Modern Physics **84**, 621 (2012), arXiv:1110.3234 [quant-ph].

[59] L. C. G. Govia, G. J. Ribeill, G. E. Rowlands, H. K. Krovi, and T. A. Ohki, Quantum reservoir computing with a single nonlinear oscillator, Phys. Rev. Research **3**, 013077 (2021), arXiv:2004.14965 [quant-ph].
[60] M. Schuld and F. Petruccione, *Machine Learning with Quantum Computers* (Springer, Cham, 2021).
[61] M. C. Caro, Learning Quantum Processes and Hamiltonians via the Pauli Transfer Matrix, ACM Transactions on Quantum Computing **5**, 1 (2024), arXiv:2212.04471 [quant-ph].
[62] L. Hantzko, L. Binkowski, and S. Gupta, Fast generation of Pauli transfer matrices utilizing tensor product structure, Physica Scripta **100**, 075125 (2025), arXiv:2411.00526 [quant-ph].
[63] A. Angrisani, A. Schmidhuber, M. S. Rudolph, M. Cerezo, Z. Holmes, and H.-Y. Huang, Classically estimating observables of noiseless quantum circuits (2025), arXiv:2409.01706 [quant-ph].
[64] H. So, K. W. Chan, Y. Chan, and K. Ho, Linear prediction approach for efficient frequency estimation of multiple real sinusoids: Algorithms and analyses, IEEE Transactions on Signal Processing **53**, 2290 (2005).
[65] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction* (Wiley, Chichester, U.K, 2008).
[66] E. Peters and M. Schuld, Generalization despite overfitting in quantum machine learning models, Quantum **6**, 777 (2022).
[67] 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. Research **6**, 013051 (2024).
[68] S. Čindrak, K. Lüdge, and L. Jaurigue, From Krylov Complexity to Observability: Capturing Phase Space Dimension with Applications in Quantum Reservoir Computing (2025), arXiv:2502.12157 [quant-ph].
[69] J. Steinegger and C. Räth, Predicting three-dimensional chaotic systems with four qubit quantum systems, Scientific Reports **15**, 6201 (2025), arXiv:2501.15191 [quant-ph].
[70] 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, Nature Communications **2**, 468 (2011).
[71] This allows for the existence of a pseudoinverse such that $R^{+}R = \mathbb{I}_q$.
[72] D. J. Gauthier, E. Bollt, A. Griffith, and W. A. S. Barbosa, Next generation reservoir computing, Nature Communications **12**, 5564 (2021).
[73] D. E. Parker, X. Cao, A. Avdoshkin, T. Scaffidi, and E. Altman, A Universal Operator Growth Hypothesis, Phys. Rev. X **9**, 041017 (2019), arXiv:1812.08657 [cond-mat].
[74] A. Nahum, S. Vijay, and J. Haah, Operator Spreading in Random Unitary Circuits, Phys. Rev. X **8**, 021014 (2018).
[75] E. H. Lieb and D. W. Robinson, The finite group velocity of quantum spin systems, Communications in Mathematical Physics **28**, 251 (1972).
[76] B. Nachtergaele, Y. Ogata, and R. Sims, Propagation of Correlations in Quantum Lattice Systems, J. Stat. Phys. **124**, 1 (2006).
[77] C.-F. (Anthony) Chen, A. Lucas, and C. Yin, Speed limits and locality in many-body quantum dynamics, Reports on Progress in Physics **86**, 116001 (2023).
[78] B. J. Mahoney and C. S. Lent, Lieb-Robinson correlation function for the quantum transverse-field Ising model, Phys. Rev. Research **6**, 023286 (2024).
[79] 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 (2018), arXiv:1809.02126 [cond-mat].
[80] A. Avdoshkin and A. Dymarsky, Euclidean operator growth and quantum chaos, Phys. Rev. Research **2**, 043234 (2020).
[81] J. D. Noh, Operator growth in the transverse-field Ising spin chain with integrability-breaking longitudinal field, Phys. Rev. E **104**, 034112 (2021).
[82] R. Kaneko and I. Danshita, Dynamics of correlation spreading in low-dimensional transverse-field Ising models, Phys. Rev. A **108**, 023301 (2023).

[83] 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).

[84] 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 $\mathbf{e}_j \in \mathrm{row}(R)$. Since R^+R is the orthogonal projector onto $\mathrm{row}(R)$, $(R^+R)\mathbf{e}_j = \mathbf{e}_j$ and thus $\gamma_j^2 = (R^+R)_{jj} = 1$.

[85] 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.

[86] By global Hadamard equivalence, the statement flips for Pauli-X as initial operators.

[87] H. Jaeger, Short term memory in echo state networks (2001).

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

[89] Due to periodicity of rotational encoding [Eq. (2.10)], the achievable capacity can be slightly reduced when sampling $u \sim \mathcal{N}$.

[90] In Eq. (4.21), $\mathrm{Var}_{\mathbf{u}} y(\mathbf{u})$ denotes the variance of $y(\mathbf{u})$ over the distribution of $\mathbf{u}$.

[91] 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.

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

[93] M. N. Ivaki, A. Lazarides, and T. Ala-Nissila, Quantum reservoir computing on random regular graphs, Phys. Rev. A **112**, 012622 (2025).

[94] A. Palacios, R. Martínez-Peña, M. C. Soriano, G. L. Giorgi, and R. Zambrini, Role of coherence in many-body Quantum Reservoir Computing, Communications Physics **7**, 369 (2024).

[95] L. Ljung, *System Identification: Theory for the User* (Pearson, Upper Saddle River, NJ, 1999).

[96] J. C. Sprott, *Elegant Chaos: Algebraically Simple Chaotic Flows* (World Scientific, 2010).

[97] 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 $\Delta t \lesssim 0.001$.

[98] C. Yang, J. Qiao, H. Han, and L. Wang, Design of polynomial echo state networks for time series prediction, Neurocomputing **290**, 148 (2018).

[99] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, Proceedings of the National Academy of Sciences **113**, 3932 (2016).

[100] 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 (2022), arXiv:2201.05193 [cs].

[101] Y. Zhang, E. R. Santos, and S. P. Cornelius, How more data can hurt: Instability and regularization in next-generation reservoir computing (2025), arXiv:2407.08641 [cs].

[102] M. Gross, Flow Map Learning in Nonlinear Vector Autoregressive Models (2026).

[103] Moreover, non-orthogonal function libraries are typically highly redundant, reflected in large condition numbers of the corresponding design matrices [Eqs. (4.4) and (4.5)].

[104] 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 $\mathrm{rank}(K) = 16$.

[105] R. Sweke, D. P. Recio, S. Jerbi, E. Gil-Fuster, and E. Fuller, Potential and limitations of random fourier features for dequantizing quantum machine learning, Quantum **7**, 872 (2023).

[106] F. J. Schreiber, J. Eisert, and J. J. Meyer, Classical Surrogates for Quantum Learning Models, Phys. Rev. Lett. **131**, 100803 (2023).

[107] 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 (2025), arXiv:2512.15661 [quant-ph].

[108] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control* (Cambridge University Press, Cambridge, 2019).

[109] J. Bowles, S. Ahmed, and M. Schuld, Better than classical? The subtle art of benchmarking quantum machine learning models, arXiv.org 10.48550/arXiv.2403.07059 (2024).

[110] J. Schnabel and M. Roth, Quantum kernel methods under scrutiny: A benchmarking study, Quantum Machine Intelligence **7**, 58 (2025).
[111] 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 Machine Intelligence **7**, 110 (2025).
[112] T. Rohe, D. Schuman, J. Nüßlein, L. Sünkel, J. Stein, and C. Linnhoff-Popien, The Questionable Influence of Entanglement in Quantum Optimisation Algorithms (2025), arXiv:2407.17204 [quant-ph].
[113] H. Gundlach, H. Kukina, J. Lynch, and N. Thompson, Quantum Deep Learning Still Needs a Quantum Leap (2025), arXiv:2511.01253 [quant-ph].
[114] T. Fellner, D. Kreplin, S. Tovey, and C. Holm, Quantum vs. classical: A comprehensive benchmark study for predicting time series with variational quantum machine learning, Machine Learning: Science and Technology **7**, 010501 (2026), arXiv:2504.12416 [quant-ph].
[115] D. Freinberger and P. Moser, The Role of Quantum in Hybrid Quantum-Classical Neural Networks: A Realistic Assessment (2026), arXiv:2601.04732 [quant-ph].
[116] This is in contrast to the one-step training error, which can be decreased even by linear delay features.
[117] D. Giannakis, A. Ourmazd, P. Pfeffer, J. Schumacher, and J. Slawinska, Embedding classical dynamics in a quantum computer (2021), arXiv:2012.06097 [quant-ph].
[118] S. Han, L. Jia, and L. Guo, Multiple Embeddings for Quantum Machine Learning, arXiv.org 10.48550/arXiv.2503.22758 (2025).
[119] E. Gil-Fuster, J. Eisert, and C. Bravo-Prieto, Understanding quantum machine learning also requires rethinking generalization, Nature Communications **15**, 2277 (2024).
[120] Due to symmetries, the actual effective dimension is lower, and typically $\sim 3^n$.
[121] M. Beth Ruskai, S. Szarek, and E. Werner, An analysis of completely-positive trace-preserving maps on M2, Linear Algebra and its Applications **347**, 159 (2002).
[122] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge ; New York, 2010).
[123] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, Cambridge, UK ; New York, 2017).
[124] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, 2018).
[125] T. Kubota, Y. Suzuki, S. Kobayashi, Q. H. Tran, N. Yamamoto, and K. Nakajima, Quantum Noise-Induced Reservoir Computing (2022), arXiv:2207.07924 [quant-ph].
[126] T. Prosen and I. Pižorn, Operator space entanglement entropy in a transverse Ising chain, Phys. Rev. A **76**, 032316 (2007).
[127] G. B. Mbeng, A. Russomanno, and G. E. Santoro, The quantum Ising chain for beginners, SciPost Physics Lecture Notes , 082 (2024).