

Unraveling quantum environments: Transformer-assisted learning in Lindblad dynamics

Chi-Sheng Chen¹ and En-Jui Kuo^{1,2}

¹*Department of Electrophysics, National Yang Ming Chiao Tung University, Hsinchu, Taiwan, Republic of China*

²*Center for Theoretical and Computational Physics, National Yang Ming Chiao Tung University, Hsinchu 300093, Taiwan*



(Received 13 May 2025; revised 4 August 2025; accepted 16 September 2025; published 28 October 2025)

Understanding dissipation in open quantum systems is crucial for the development of robust quantum technologies. In this work, we introduce a transformer-based machine learning framework to infer time-dependent dissipation rates in systems governed by the Lindblad master equation. Our method requires only time series of observable quantities, such as $\langle\sigma_x(t)\rangle$, $\langle\sigma_y(t)\rangle$, and $\langle\sigma_z(t)\rangle$, and does not require prior knowledge of the initial quantum state or the system Hamiltonian during inference. We demonstrate the effectiveness of our approach across a hierarchy of open quantum models, including single-qubit systems with time-independent and time-dependent jump rates, two-qubit interacting systems (e.g., Heisenberg and transverse Ising models), and the Jaynes-Cummings model with cavity loss and time-dependent decay. Our method accurately reconstructs both fixed and time-varying dissipation profiles from observable trajectories. We further show that under reasonable assumptions, the jump rates in these models are uniquely determined by a small set of measurable observables. The results suggest that transformer-based learning provides a scalable and data-driven route for identifying environmental effects in open quantum systems.

DOI: [10.1103/gsxk-45mk](https://doi.org/10.1103/gsxk-45mk)

I. INTRODUCTION

Open quantum systems play a fundamental role in quantum information, condensed matter physics, and quantum thermodynamics [1,2]. Unlike closed systems, which evolve unitarily under the Schrödinger equation, open quantum systems interact with external environments, resulting in nonunitary dynamics. Environmental coupling introduces dissipation and decoherence, making the analysis of such systems essential for understanding real-world quantum technologies [3].

In the weak-coupling regime, the dynamics of an open system's density matrix ρ are often modeled by the Lindblad master equation [1],

$$\dot{\rho} = -i[H, \rho] + \sum_i \gamma_i \left(L_i \rho L_i^\dagger - \frac{1}{2} \{L_i^\dagger L_i, \rho\} \right), \quad (1)$$

where H governs the unitary evolution, L_i are jump operators that encode system-environment interactions, and $\gamma_i \geq 0$ are decay rates that quantify the coupling strength.

In many experimental settings, both the jump operators and the dissipation rates γ_i can be complex or time dependent [4–7]. This uncertainty motivates a sensing-based approach: embed a known quantum system (e.g., a qubit or photonic state) into an unknown environment and infer the dissipation parameters from observable dynamics. This concept underlies quantum sensing protocols [8], which have been demonstrated on platforms including superconducting flux qubits [9], photonic systems [10], and trapped ions [11].

In this work, we explore whether such observable time-series data can be used to infer environmental properties, and propose using modern sequence modeling architectures—specifically the transformer [12–15]—to learn time-dependent dissipation rates in open quantum systems.

We begin with the single-qubit case, where the jump operators are fixed but the dissipation rate $\gamma(t)$ is unknown. The model input consists of local observable trajectories: $\langle\sigma_x(t)\rangle$, $\langle\sigma_y(t)\rangle$, and $\langle\sigma_z(t)\rangle$. We demonstrate that a transformer-based regression model can accurately reconstruct the time-dependent $\gamma(t)$ from these observables.

We then extend our analysis to two-qubit systems with interacting Hamiltonians. In particular, we consider the anisotropic Heisenberg model (realized in quantum dot arrays [16–20]) and the transverse-field Ising model (implemented in ion-trap and Rydberg platforms). In both cases, the transformer model successfully predicts the dissipation rates associated with each qubit, showing strong generalization across coupling regimes.

To further test the method in hybrid quantum systems, we apply it to the open Jaynes-Cummings model [21], which describes light-matter interaction between a qubit and a photon mode. Here, the model predicts six time-dependent coefficients parametrizing the dissipation rates $\kappa(t)$ and $\gamma(t)$, based on the observable sequences $\langle\sigma_s(t)\rangle$, $s \in \{x, y, z\}$, $\langle a^\dagger a(t) \rangle$, and $\langle X \otimes \sigma_z(t) \rangle$.

In all cases, the transformer model achieves high predictive accuracy. We summarize full architectural and training details in Appendixes B–D. Moreover, we provide theoretical results justifying the identifiability of dissipation rates from observable trajectories under mild assumptions. Analytical derivations for both the single-qubit Lindblad case and the Jaynes-Cummings model are presented in Appendix E.

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

Our results suggest that data-driven sequence models can serve as effective and scalable tools for extracting physical parameters in open quantum systems, offering a robust alternative to analytic inversion or system identification methods.

We note that our current model is restricted to time-local Lindblad dynamics with non-negative rates. In particular, we assume that the Lindblad structure and jump operators are known *a priori* (e.g., amplitude damping via σ_- , or cavity loss via a). This is consistent with many realistic scenarios in quantum hardware, where dominant decoherence channels are experimentally well characterized. Under this assumption, our method learns the effective dissipation rates $\gamma_i(t)$ within a fixed operator representation and does not attempt to reconstruct microscopic bath parameters such as spectral density or temperature. When the true dynamics exhibits non-Markovian features, such as temporarily negative $\gamma(t)$, the inferred rates may not correspond to a physically consistent generator. Extending the present framework to capture such memory effects—e.g., via learning nonlocal memory kernels or non-Lindbladian propagators—is an interesting direction for future work.

Related work. Recent efforts have applied machine learning to characterize quantum dynamics, including noise classification [22,23], effective Lindbladian recovery from unitary simulations [24], and dynamical inference using tensor networks [25] or dynamical maps [26]. While powerful, these methods often require full density matrices, high-fidelity state tomography, or are restricted to classification tasks. Closer in spirit, Bairey *et al.* [27] reconstruct Lindbladians from steady states, though such approaches may fail in systems with trivial attractors (e.g., vacuum under bosonic loss). Our method instead focuses on regressing time-varying dissipation rates from limited observables, under known operator structure, and requires no access to initial states or full tomography. This positions our framework as a scalable, data-driven tool for dissipation identification in realistic NISQ platforms.

Machine learning method basics

In recent years, machine learning has emerged as a powerful tool across various domains of physics [28]. Applications include phase classification in condensed matter systems [29–32], modeling nonequilibrium dynamics [33,34], exploring the string theory landscape and holographic dualities [35,36], simulating quantum many-body systems [37], quantum state tomography [38,39], enhancing quantum hardware performance [40], and quantum machine learning [41,42]. In molecular physics and chemistry, machine learning has also been applied to accelerate molecular dynamics simulations [43,44].

When specialized in time-series prediction or natural language processing (NLP), a wide array of architectures have been developed to capture temporal correlations, starting with recurrent neural networks (RNNs) [45], gated recurrent units (GRUs) [46], and several gated variants such as long short-term memory (LSTM) networks [47]. More recently, the transformer architecture [12] has revolutionized the field by replacing recurrence with self-attention mechanisms, enabling scalable modeling of long-range dependencies. Its success in

NLP has inspired applications across fields, including physics [48,49].

II. MODEL AND QUANTITIES OF INTEREST

We now describe the learning task and modeling setup. As discussed in Eq. (1), the dynamics of open quantum systems are governed by the Lindblad master equation [50,51], where H is the system Hamiltonian, L_i are jump operators, and $\gamma_i(t) \geq 0$ are time-dependent dissipation rates. This form ensures the dynamics remains completely positive and trace preserving (CPTP) [52]. Conversely, any divisible CPTP map admits a representation of this structure.

We assume that both the Hamiltonian H and jump operators $\{L_i\}$ are known during data generation, as they are required to simulate time-series observables for training. However, the inference model itself does not receive H or the coupling constants as input. Our goal is to learn the time-dependent dissipation rates $\{\gamma_i(t)\}$ from observable trajectories alone.

For example, consider a single-qubit system with Hamiltonian $H = \sigma_z$ and a single jump operator σ_- with time-dependent dissipation $\gamma(t)$,

$$\dot{\rho} = -i[H, \rho] + \gamma(t)(\sigma_- \rho \sigma_+ - \frac{1}{2}\{\sigma_+ \sigma_-, \rho\}). \quad (2)$$

If the system is initialized in multiple states and we record observables such as $\langle \sigma_x(t) \rangle$, $\langle \sigma_y(t) \rangle$, and $\langle \sigma_z(t) \rangle$ over time, then, in principle, $\gamma(t)$ can be reconstructed, as the Pauli observables span the full space of Hermitian operators on a qubit. However, the exact inversion formulas are analytically involved; see Appendix E for explicit derivations. This reflects a realistic scenario in experiments, where the form of the Hamiltonian is engineered and known, but the dissipation profile remains uncertain.

To proceed, we cast the reconstruction of $\gamma(t)$ as a regression problem. The input consists of observable trajectories generated from random initial states and the output is a finite-dimensional parametrization of $\gamma(t)$. To represent the time-dependent dissipation profile, we adopt the Bernstein polynomial basis of fixed degree n . This basis offers a simple yet expressive representation for non-negative functions on $[0,1]$. Specifically, the Bernstein basis of degree n consists of

$$b_{j,n}(t) = \binom{n}{j} t^j (1-t)^{n-j}, \quad j = 0, 1, \dots, n, \quad (3)$$

and any non-negative linear combination $\gamma(t) = \sum_{j=0}^n a_j b_{j,n}(t)$ with $a_j \geq 0$ guarantees $\gamma(t) \geq 0$ on $[0,1]$. This avoids the sign indefiniteness inherent in monomial bases (e.g., $\{1, t, t^2\}$) and enables random sampling or learning over the coefficient space. A detailed discussion of the Bernstein basis and its approximation properties is provided in Appendix A.

The learning task is thus reduced to predicting the coefficient vector $\{a_0, \dots, a_n\}$ from time-series input features. In our implementation, we use a transformer model trained on multiple trajectories to output these coefficients. The model is trained using a supervised loss over many initial states and dissipation profiles.

An overview of the entire pipeline is depicted in Fig. 1. Observable time series such as $\langle \sigma_x(t) \rangle$, $\langle \sigma_y(t) \rangle$, and $\langle \sigma_z(t) \rangle$

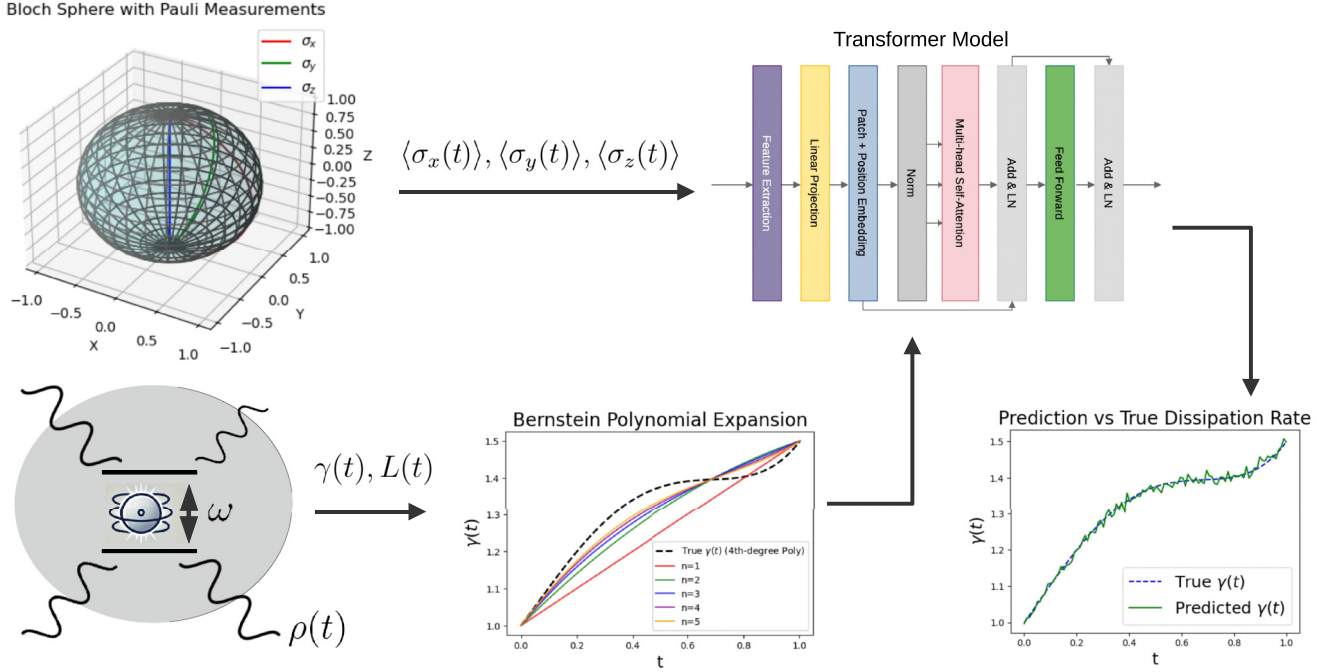


FIG. 1. Schematic illustration of the machine learning pipeline. Time-series data of local observables $\langle \sigma_{x,y,z}(t) \rangle$ are used as input features to a transformer model trained to regress the time-dependent dissipation rate $\gamma(t)$. The dissipation profile is parameterized using a Bernstein polynomial expansion. This approach is applied to single- and two-qubit systems, as well as light-matter interaction models.

are encoded as fixed-length feature vectors. The transformer learns to regress the dissipation profile by predicting the corresponding Bernstein coefficients. This framework will be applied in the next sections to various systems, including single-qubit models, two-qubit interactions, and the Jaynes-Cummings model.

III. EXAMPLES

In this section, we apply our framework to several concrete systems and summarize the setup of each in Table I.

A. Single-qubit dissipative systems

We begin with four representative examples involving single-qubit open dynamics. The first case considers a simple Hamiltonian $H = \sigma_z$ with a constant but unknown dissipation rate γ_- associated with the jump operator σ_- . The goal is to

predict the scalar value of γ_- from the observable trajectory $\langle \sigma_z(t) \rangle$.

The second case extends this setup to time-dependent dissipation: $\gamma_-(t)$ is parameterized by a second-order Bernstein polynomial. In both the first and second examples, the system is initialized in the ground state $|0\rangle$.

In the third and fourth examples, the initial state is drawn randomly from the Haar measure over pure qubit states. This is achieved by sampling the components α_j from independent complex Gaussian distributions $\mathcal{CN}(0, 1)$ and normalizing the resulting vector. These random initializations provide a more general testing ground for the model's robustness across input variability.

In all four cases, we generate 1000 simulation trajectories using the `qutip` package [53], each with randomly sampled dissipation rates. The input features consist of observable time series, either $\langle \sigma_z(t) \rangle$ or $\langle \sigma_{x,y,z}(t) \rangle$, depending on the initial

TABLE I. Details of examples shown in the single-qubit space $\mathbb{C}^2$. For the first and third examples, we pick γ_+ , γ_- uniformly at random in $(0, 2)$. For the third and fourth examples, we pick $\gamma_{\pm}(t) = \sum_{j=0}^2 \gamma_{\pm,j} b_{j,2}(t)$, where $b_{j,2}(t)$ is the j th Bernstein polynomial of degree 2, with all $\gamma_{\pm,j} \in (0.1, 2)$. We denote $\mathcal{H}_p$ to be the Hilbert space of a single photon.

Hilbert Space	Hamiltonian	Initial	Jump operator L , rate γ	Input
$\mathbb{C}^2$	σ_z	$ 0\rangle$	(σ_-, γ_-)	$\langle \sigma_z(t) \rangle$
$\mathbb{C}^2$	σ_z	Haar random	$(\sigma_-, \gamma_-), (\sigma_+, \gamma_+)$	$\langle \sigma_i(t) \rangle, i \in \{x, y, z\}$
$\mathbb{C}^2$	σ_z	$ 0\rangle$	$[\sigma_-, \gamma_-(t)]$	$\langle \sigma_z(t) \rangle$
$\mathbb{C}^2$	σ_z	Haar random	$[\sigma_-, \gamma_-(t)], (\sigma_+, \gamma_+(t))$	$\langle \sigma_i(t) \rangle, i \in \{x, y, z\}$
$\mathbb{C}^2 \otimes \mathbb{C}^2$	$J(\sigma_x^{(1)}\sigma_x^{(2)} + \sigma_y^{(1)}\sigma_y^{(2)} + \sigma_z^{(1)}\sigma_z^{(2)})$	Haar random	$(\sigma_{-,j}, \gamma_{-,j}), (\sigma_{+,j}, \gamma_{+,j})$	$\langle \sigma_i^j(t) \rangle, i \in \{x, y, z\}, j \in \{1, 2\}$
$\mathbb{C}^2 \otimes \mathbb{C}^2$	$J\sigma_z^{(1)}\sigma_z^{(2)} + h(\sigma_x^{(1)} + \sigma_x^{(2)})$	Haar random	$(\sigma_{-,j}, \gamma_{-,j}), (\sigma_{+,j}, \gamma_{+,j})$	$\langle \sigma_i^j(t) \rangle, i \in \{x, y, z\}, j \in \{1, 2\}$
$\mathbb{C}^2 \otimes \mathcal{H}_p$	$\omega_c a^\dagger a + \frac{\omega_q}{2} \sigma_z + g(\sigma_+ a + \sigma_- a^\dagger)$	Haar random $\otimes n\rangle$	$[\sigma_-, \gamma_-(t)], [a, \kappa(t)]$	$\langle \sigma_i(t) \rangle, i \in \{x, y, z\}, \langle n(t) \rangle, \langle \sigma_z \otimes X_{\text{photon}}(t) \rangle$

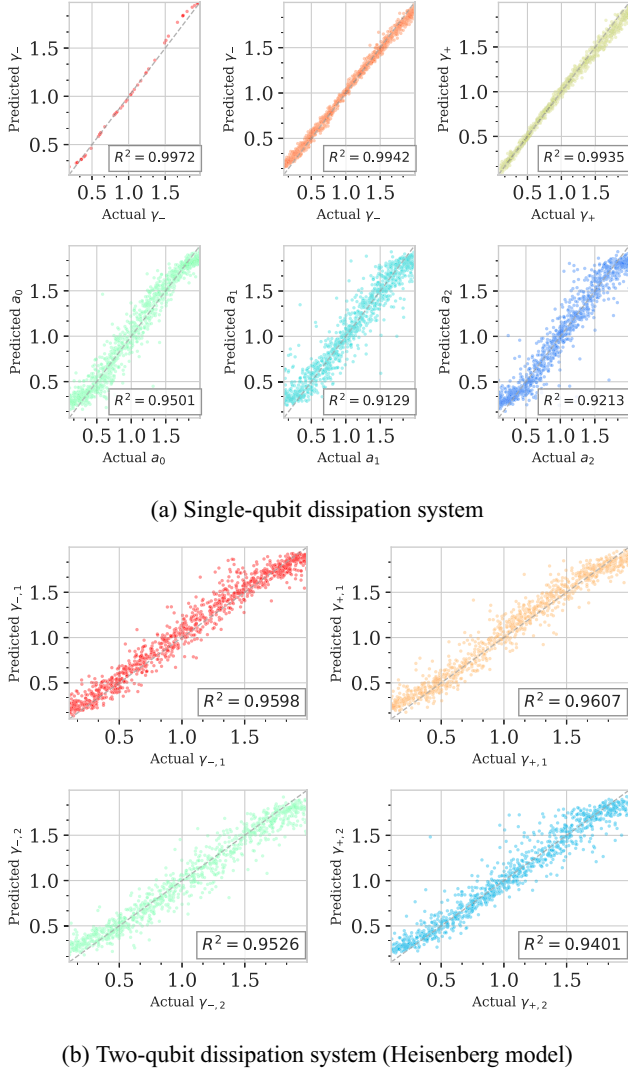


FIG. 2. Predictions of dissipation rates for single- and two-qubit systems. (a) Results from the first three single-qubit examples in Table I. The top row illustrates prediction of time-dependent dissipation rates. The first panel corresponds to Example 1, with a single channel $\gamma_-(t)$ and jump operator σ_- . The second and third panels correspond to Example 2, which includes both $\gamma_+(t)$ and $\gamma_-(t)$ associated with jump operators σ_+ and σ_- . The bottom row shows the predicted Bernstein coefficients a_0 , a_1 , and a_2 for Example 3, where $\gamma_-(t) = \sum_{k=0}^2 a_k b_{k,2}(t)$ is expanded in the Bernstein basis. (b) The predicted dissipation rates for the Heisenberg two-qubit model, involving four constant rates $\gamma_{\pm,j}$ acting on each qubit $j = 1, 2$.

state. The model is trained to regress either the constant γ or the Bernstein coefficients if $\gamma(t)$ is time dependent. The dataset is split into 80% training and 20% testing.

Model architecture, feature extraction techniques, and additional training details are provided in Appendix B. Results for the first three examples are shown in Fig. 2(a). The top panel corresponds to the first example (constant γ_-); the second and third panels illustrate the performance on time-dependent $\gamma_-(t)$ with deterministic and random initial states, respectively. The R^2 scores in all three cases exceed 0.99, indicating excellent predictive performance.

The fourth example introduces two time-dependent dissipation channels $\gamma_+(t)$ and $\gamma_-(t)$, corresponding to jump operators σ_+ and σ_- . Each is represented using a second-order Bernstein basis. This results in a six-dimensional output space. Due to space constraints, results for this case are included in Appendix B.

B. Two-qubit dissipative systems

After validating our approach on single-qubit systems, we extend the framework to two-qubit dissipative systems with constant pumping and loss channels. We consider two widely studied Hamiltonians: The Heisenberg exchange Hamiltonian, $H = J(\sigma_x^{(1)}\sigma_x^{(2)} + \sigma_y^{(1)}\sigma_y^{(2)} + \sigma_z^{(1)}\sigma_z^{(2)})$, relevant to quantum dot architectures [16–20], and the transverse-field Ising model, $H = J\sigma_z^{(1)}\sigma_z^{(2)} + h(\sigma_x^{(1)} + \sigma_x^{(2)})$, commonly realized in superconducting circuits, trapped ions, and cold-atom platforms [54–56].

Both models serve as minimal testbeds for studying dissipation, entanglement, and open-system dynamics under tunable interactions. For the Heisenberg model, we sample $J \in (0, 2)$ uniformly. For the transverse-field Ising model, we sample $(J, h) \in (0.1, 2)^2$ and assign four dissipation rates $\gamma_{\pm,j} \in (0.1, 2)$ for $j = 1, 2$, corresponding to jump operators $\sigma_{\pm,j}$ acting on each qubit.

The model input consists of local observables $\langle \sigma_{x,y,z}^{(1)}(t) \rangle$ and $\langle \sigma_{x,y,z}^{(2)}(t) \rangle$ and the regression target is the set of unknown rates $\{\gamma_{\pm,1}, \gamma_{\pm,2}\}$. Results for the Heisenberg model are shown in Fig. 2(b), where the transformer accurately predicts all four rates with $R^2 \geq 0.95$.

Each trajectory is generated from a product of Haar-random pure states and evolved under the Lindblad master equation. We simulate 3000 examples for each model. The dataset is split into 80% training and 20% testing. Feature extraction techniques are applied to improve performance, and full implementation details are provided in Appendix C.

These two-qubit models are summarized in columns 5 and 6 of Table I. The dataset offers a testbed for learning dissipative dynamics with tunable interactions and multi-qubit decoherence. Notably, our model achieves this inference without access to the coupling constants or the underlying Hamiltonian parameters, relying solely on observable trajectories. Theoretical justification for identifiability of the dissipation rates in this setting is given in Appendix E.

While the two-qubit benchmarks already demonstrate the ability of our framework to infer multiple dissipation channels in interacting systems, such small-scale examples leave open the question of scalability. To further assess performance in larger settings, we now turn to Heisenberg spin chains with up to nine qubits, where the number of degrees of freedom and dissipation parameters grow systematically with n . These benchmarks allow us to test whether the method remains effective as the system size increases.

C. Scaling tests on multiqubit Heisenberg chains

To further examine scalability beyond the one- and two-qubit cases, we benchmark our framework on Heisenberg spin chains of up to nine qubits. We consider two scenarios: a *sparse test*, involving a single dissipation channel, and a

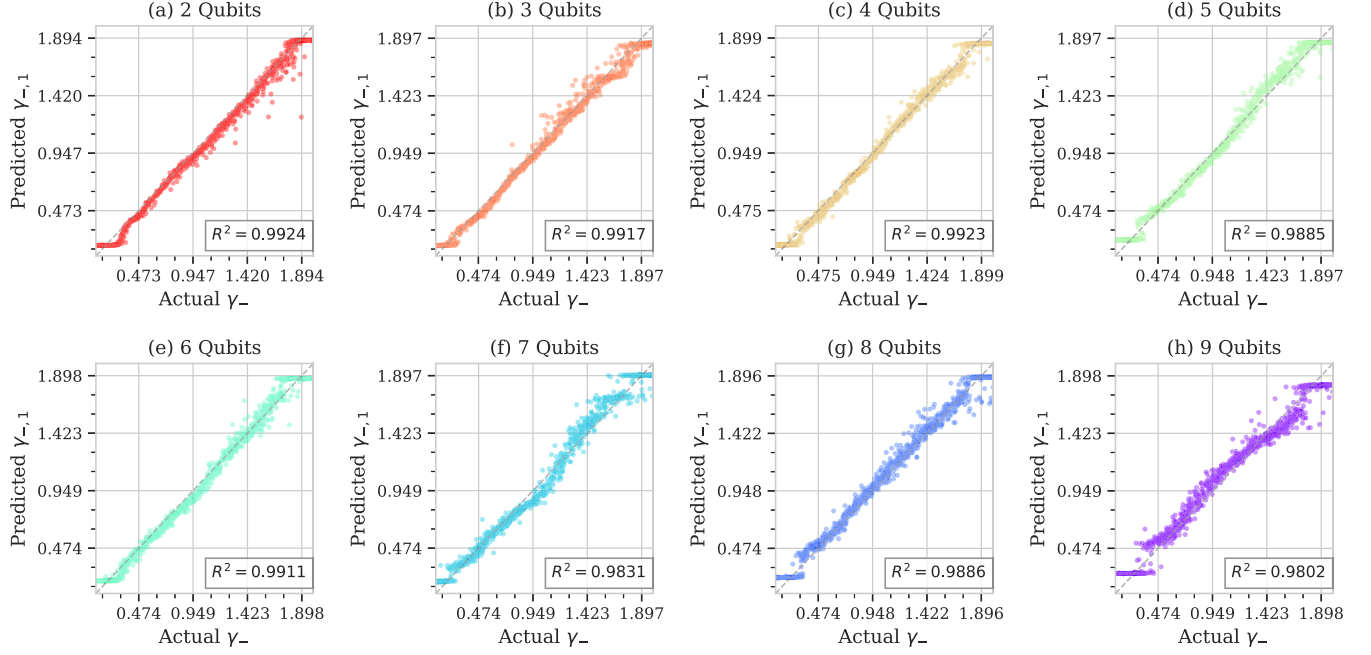


FIG. 3. Sparse scaling results. Prediction performance for Heisenberg chains of size $n = 2-9$ with a single dissipation channel. The transformer maintains $R^2 > 0.97$ across all sizes, demonstrating robust scalability under simple dissipation.

dense test, where each qubit is subject to independent loss and excitation processes.

Sparse test. In this setting, an n -qubit Heisenberg chain with nearest-neighbor coupling,

$$H = J \sum_{j=1}^{n-1} (\sigma_x^{(j)} \sigma_x^{(j+1)} + \sigma_y^{(j)} \sigma_y^{(j+1)} + \sigma_z^{(j)} \sigma_z^{(j+1)}), \quad J = 1,$$

is initialized in Haar-random product states. A single dissipation channel acts on the first qubit with jump operator $L = \sigma_-^{(1)}$ and random rate $\gamma \in (0.1, 2.0)$. For each $n = 2, \dots, 9$, we generate 1000 trajectories, each consisting of local magnetizations $\langle \sigma_z^{(j)}(t) \rangle$ sampled at 100 time steps over $t \in [0, 5]$. The regression task is to infer γ directly from this $(n \times 100)$ matrix of observables. Across all sizes, the transformer achieves $R^2 > 0.97$, showing that the difficulty of the learning task grows only mildly with n . Figure 3 illustrates the predicted versus true values of γ , confirming robustness even at $n = 9$.

Dense test. We next consider a more challenging regime in which each qubit j carries two independent Lindblad channels $\{\sigma_+^{(j)}, \sigma_-^{(j)}\}$ with random rates $\gamma_{j,\pm} \in (0.1, 2.0)$. The system evolves under the same Heisenberg Hamiltonian, with trajectories recorded for local observables $\langle \sigma_{x,y,z}^{(j)}(t) \rangle$ over 100 time steps. Each training example thus consists of $3n$ time series, and the regression target is the full set of $2n$ rates. We generate 1000 trajectories for each system size $n = 2, \dots, 7$. As shown in Fig. 4, the average R^2 score per rate degrades only linearly with n , remaining above 0.85 even for $n = 7$. This indicates that meaningful inference remains possible even in densely parameterized, strongly dissipative regimes.

These benchmarks demonstrate that the proposed framework scales favorably with system size, even though the Hilbert space dimension grows exponentially. The accuracy of

inference is governed primarily by the number of dissipation parameters (linear in n under local noise assumptions), not by the full state space dimension. We emphasize that the same transformer architecture and hyperparameters were used throughout, without task-specific tuning. Full details of the network design, feature extraction, and training protocol are provided in Appendix F.

D. Photon-spin coupling systems

We now apply our framework to hybrid light-matter systems, focusing on the Jaynes-Cummings model [21]. The

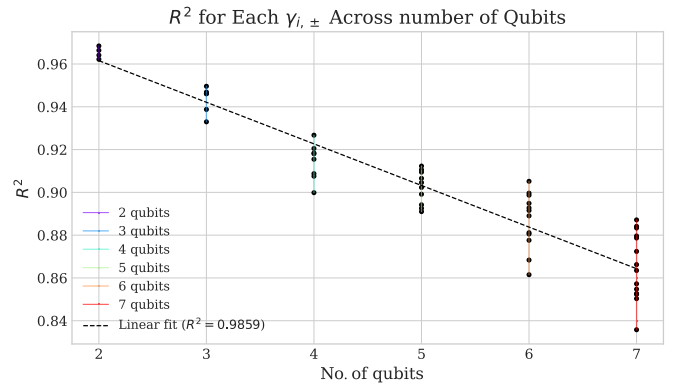


FIG. 4. Dense scaling results. Prediction accuracy for Heisenberg chains of size $n = 2-7$ with two dissipation channels per qubit. Colored curves show R^2 scores for individual rates, while the dashed line indicates the average over all $2n$ parameters. Performance decreases only mildly with n , confirming robustness in high-dimensional dissipation models.

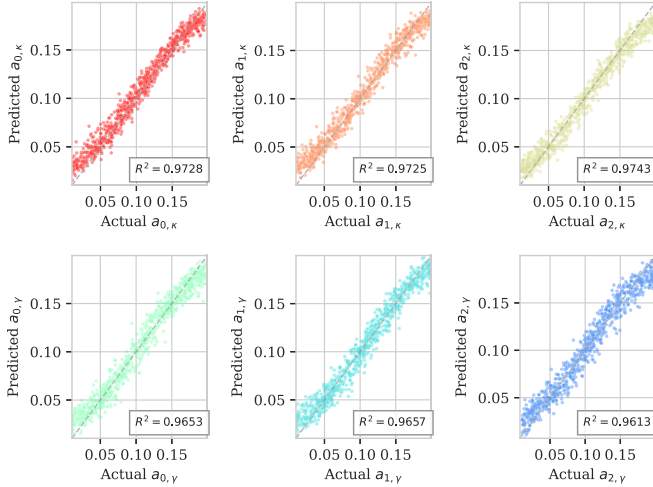


FIG. 5. Prediction of time-dependent dissipation rates in the open Jaynes-Cummings model. The decay rates $\gamma(t)$ and $\kappa(t)$ are each parameterized by a second-order Bernstein polynomial with three unknown coefficients. The model accurately reconstructs all six coefficients from time series of observables $\langle\sigma_z(t)\rangle$, $\langle a^\dagger a(t)\rangle$, and $\langle X \otimes \sigma_z(t)\rangle$. See Table I, column 7, for full input-output specifications.

system is governed by the Hamiltonian

$$H = \omega_c a^\dagger a + \frac{\omega_q}{2} \sigma_z + g(\sigma_+ a + \sigma_- a^\dagger), \quad (4)$$

where ω_c and ω_q denote the cavity and qubit frequencies, respectively, and g is the light-matter coupling strength.

To reflect realistic parameter regimes, we sample (ω_c, ω_q) uniformly from $(0.8, 1.2)$ and g from $(0.01, 0.1)$. The system is initialized in a product state $|n\rangle \otimes |g\rangle$, where $1 \leq n \leq 10$ is drawn randomly. The open-system dynamics is governed by a time-dependent Lindblad master equation,

$$\dot{\rho} = -i[H, \rho] + \kappa(t) \mathcal{D}[a]\rho + \gamma(t) \mathcal{D}[\sigma_-]\rho, \quad (5)$$

where $\kappa(t)$ and $\gamma(t)$ represent photon and qubit decay rates, respectively. Both are parameterized as second-order Bernstein polynomials,

$$\kappa(t) = \sum_{j=0}^2 a_{j,\kappa} b_{j,2}(t), \quad \gamma(t) = \sum_{j=0}^2 a_{j,\gamma} b_{j,2}(t),$$

with coefficients $a_{j,\kappa}, a_{j,\gamma}$ sampled uniformly from $(0.01, 0.2)$.

The model input consists of time-series observables $\langle\sigma_s(t)\rangle$ $s \in \{x, y, z\}$, $\langle a^\dagger a(t)\rangle$, and $\langle X \otimes \sigma_z(t)\rangle$, recorded at 100 points over $t \in [0, 10]$. The physical parameters (ω_c, ω_q, g) are excluded from the input. The regression target is the six Bernstein coefficients $\{a_{j,\gamma}, a_{j,\kappa}\}_{j=0}^2$ encoding the dissipation rates.

We generate 5000 training examples and train the model in a supervised setting with the dataset split into 80% training and 20% testing. Results are shown in Fig. 5, where all predicted coefficients achieve R^2 scores above 0.9. Full model specifications and training procedures are described in Appendix D. The identifiability of $\gamma(t)$ and $\kappa(t)$ from these

observables is formally justified in Appendix E. The setup for this case is summarized in column 7 of Table I.

IV. DISCUSSION

We have introduced a transformer-based framework for learning time-dependent dissipation rates in open quantum systems from observable trajectories such as $\langle\sigma_x(t)\rangle$, $\langle\sigma_y(t)\rangle$, and $\langle\sigma_z(t)\rangle$. The model achieves accurate reconstruction across diverse settings—including two-qubit and Jaynes-Cummings models—without requiring knowledge of the initial quantum state, thus enhancing robustness to state preparation noise [57]. While Bernstein polynomials were introduced to encode non-negativity, their primary role is to generate smooth, physically consistent training data that support efficient learning.

Our study focuses on a practically motivated and experimentally realistic regime: time-local Lindblad dynamics under weak coupling, with known Hamiltonians and jump operators. This setting reflects common assumptions in NISQ-era platforms, where the system Hamiltonian is programmable and can be independently calibrated, while time-varying dissipation remains harder to characterize and fluctuates across hardware implementations. Our method addresses this gap by inferring dissipation profiles directly from local observable data.

Although the Hilbert space dimension grows exponentially, we show that a linearly scaling set of local observables—namely, single-qubit $\langle\sigma_{x,y,z}^{(j)}(t)\rangle$ —is sufficient to reconstruct dense dissipation profiles with high fidelity. In contrast, even full knowledge of all 4^n Pauli expectation values would still require solving exponentially large inverse problems (see Appendix E). Learning from partial yet physically meaningful data is therefore both necessary and advantageous.

We emphasize that our framework is not designed for arbitrarily complex or highly nonlocal dissipative dynamics. While the Lindblad operator structure could, in principle, involve nonlocal couplings, such settings rarely arise in current experiments. Most practical systems are governed by local, Markovian noise processes with known coupling structure but unknown time dependence. Our results suggest that in this regime, transformer architectures can offer a scalable and interpretable route for dissipation inference.

Extending this approach to more general settings—such as multisite dissipation, strongly correlated environments, or non-Markovianity—will likely require new architectures, structured priors, or hybrid data-theory approaches. Identifying the minimal observable sets, quantifying identifiability thresholds, and scaling to entangled many-body dynamics remain compelling directions for future research.

ACKNOWLEDGMENTS

E.J.K. acknowledges financial support from the National Science and Technology Council (NSTC) of Taiwan under Grant No. NSTC 114-2112-M-A49-036-MY3. C.S.C. thanks Neuro Industry, Inc., for its support. We also thank the Center for Theoretical and Computational Physics (CTCP), National Yang Ming Chiao Tung University (NYCU), Taiwan, for full support.

DATA AVAILABILITY

The data that support the findings of this article are openly available [58]; embargo periods may apply.

APPENDIX A: BERNSTEIN POLYNOMIAL APPROXIMATION

Let $\{b_{v,n}(t)\}_{v=0}^n$ denote the Bernstein basis polynomials of degree n , defined on $[0, 1]$ by

$$b_{v,n}(t) = \binom{n}{v} t^v (1-t)^{n-v}, \quad v = 0, 1, \dots, n. \quad (\text{A1})$$

These polynomials form a non-negative partition of unity and span the space Π_n of real polynomials of degree at most n . Given a continuous function $f : [0, 1] \rightarrow \mathbb{R}$, the Bernstein approximation of f is

$$B_n[f](t) = \sum_{v=0}^n f\left(\frac{v}{n}\right) b_{v,n}(t). \quad (\text{A2})$$

By the classical Weierstrass approximation theorem, $B_n[f] \rightarrow f$ uniformly as $n \rightarrow \infty$ for any continuous f [59].

Compared to Taylor expansions, Bernstein polynomials have two advantages: (i) they preserve non-negativity if all coefficients are non-negative, and (ii) they are globally defined over the entire interval. For example, the quadratic polynomial $f(t) = t^2 - 2t + 2$ is strictly positive on $[0, 1]$, but has negative Taylor coefficients. In contrast, any non-negative quadratic function on $[0, 1]$ can be expressed as a positive linear combination of the Bernstein basis $\{(1-t)^2, t(1-t), t^2\}$.

This positivity makes the Bernstein basis particularly useful for generating random non-negative decay functions $\gamma(t)$ in open quantum systems. In our work, we represent unknown time-dependent dissipation rates using degree-2 Bernstein polynomials,

$$\gamma(t) = \sum_{v=0}^2 a_v b_{v,2}(t), \quad a_v \geq 0. \quad (\text{A3})$$

We show in Appendixes B–D that these coefficients can be effectively learned from time-series observables via transformer regression. The robustness of this parametrization is demonstrated across all examples.

APPENDIX B: SINGLE-QUBIT MODEL DETAILS AND ADDITIONAL EXAMPLE WITH TWO TIME-DEPENDENT $\gamma_{\pm}(t)$

1. Training architecture and feature design

Our core model is based on the transformer encoder architecture [12], which learns temporal dependencies via self-attention. Given an input matrix $X \in \mathbb{R}^{n \times d}$, where n is the sequence length and d is the feature dimension, the self-attention mechanism computes,

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V, \quad (\text{B1})$$

where the query, key, and value matrices are defined as

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V,$$

with $W^Q, W^K, W^V \in \mathbb{R}^{d \times d_k}$ being learned projection matrices. The dot product $QK^\top$ measures the similarity between time steps, and the softmax weights modulate which values V to attend to at each position.

For richer representation learning, we use multihead attention,

$$\text{Multi Head}(X) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O, \quad (\text{B2})$$

$$\text{head}_i = \text{Attention}(XW_i^Q, XW_i^K, XW_i^V), \quad (\text{B3})$$

where $W^O \in \mathbb{R}^{hd_k \times d}$ is an output projection and h is the number of attention heads (we use $h = 8$). This allows the model to jointly learn from multiple subspaces of the input.

Each encoder layer consists of multihead self-attention, followed by a feedforward block, layer normalization, and residual connections. The final representation is passed to a multilayer perceptron (MLP) to predict the target coefficients.

Our regression pipeline is based on a transformer encoder model designed to predict dissipation profiles from Pauli observable trajectories. Each single-qubit dataset consists of three input time series $\sigma_x(t)$, $\sigma_y(t)$, and $\sigma_z(t)$, evaluated at 100 uniform time points over $[0, 10]$.

We extract 18 handcrafted features per observable, including the following:

- (a) Basic statistics: mean, std, max, min, median, 25th/75th percentiles
- (b) Temporal variation: range, interquartile range (IQR), slope, mean/std of Δx and $\Delta^2 x$
- (c) Distributional shape: skewness and kurtosis
- (d) Spectral: max and average fast Fourier transform (FFT) magnitude (excluding DC component).

These yield a 54-dimensional feature vector per sample. All features and targets are standardized using z-score normalization. The model architecture includes the following:

- (a) Linear projection $\rightarrow$ LayerNorm $\rightarrow$ ReLU $\rightarrow$ Dropout
- (b) 4-layer transformer encoder with 8 attention heads, feedforward dimension 512
- (c) MLP head with layer sizes 256–128–64 and linear output layer.

We train using AdamW optimizer ($\text{lr}=10^{-3}$, weight decay=0.01), with mean squared error (MSE) loss, ReduceLROnPlateau scheduler (patience 20), and early stopping (patience 30). Each model is trained up to 300 epochs with batch size 32. Final predictions are inverse transformed, and evaluated using R^2 and MSE on the test set.

2. Example-specific configurations (examples 1–4 in the main text)

Example 1: Constant γ . A minimal transformer (2 encoder layers, 4 heads) predicts a single scalar γ from the time-averaged $\sigma_{x,y,z}$. This baseline achieves $R^2 \geq 0.99$.

Example 2: Constant $\gamma_{\pm}$. Predicts two values γ_+ , γ_- using 30 input features (10 per Pauli observable). A 3-layer transformer encoder with 8 heads is used. The model learns both rates jointly with high accuracy.

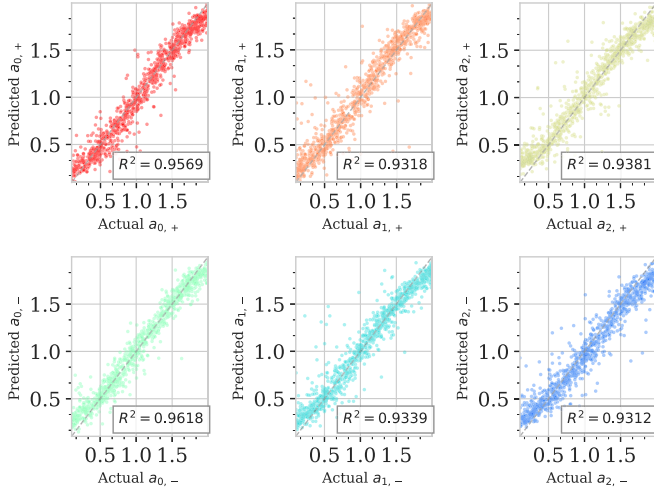


FIG. 6. Prediction results for Example 4 (main text) involving a single-qubit system with two quadratic time-dependent loss channels. The dissipation rates are represented by second-order Bernstein polynomials $\gamma_{\pm}(t) = \sum_{j=0}^2 a_{j,\pm} b_{j,2}(t)$, and six coefficients are predicted from local observables.

Examples 3 and 4: Time-dependent $\gamma(t)$. We represent each dissipation rate using the degree-2 Bernstein polynomial,

$$\gamma(t) = \sum_{j=0}^2 a_j b_{j,2}(t), \quad b_{j,2}(t) = \binom{2}{j} t^j (1-t)^{2-j}.$$

The model predicts the coefficient vector (a_0, a_1, a_2) per channel. In Example 3, only one time-dependent $\gamma(t)$ is present; in Example 4, both $\gamma_+(t)$ and $\gamma_-(t)$ are time dependent, requiring six total coefficients. Both models use the same 4-layer transformer encoder as described above.

3. Additional results for example 4 (two time-dependent dissipation channels)

We provide a detailed prediction plot for Example 4 (see Table I in the main text) in Fig. 6. This task involves learning six Bernstein coefficients from $\sigma_{x,y,z}$ time series. The transformer accurately recovers all coefficients, demonstrating a strong generalization in the time-dependent multichannel setting.

APPENDIX C: TWO-QUBIT MODELS WITH CONSTANT DISSIPATION RATES

This appendix provides the training details for both two-qubit models: the Heisenberg model presented in the main text and the transverse-field Ising model included here as additional results (see Fig. 7). While their Hamiltonians differ, the dissipation structure and model training pipeline remain identical.

Heisenberg model (main text). The Hamiltonian is given by

$$H = J(\sigma_x^{(1)}\sigma_x^{(2)} + \sigma_y^{(1)}\sigma_y^{(2)} + \sigma_z^{(1)}\sigma_z^{(2)}), \quad (C1)$$

where $J \sim \text{Uniform}(0, 2)$ is the exchange coupling strength. Each qubit is coupled to a local environment with constant dissipation and gain rates $\gamma_{\pm,j}$ for $j = 1, 2$.

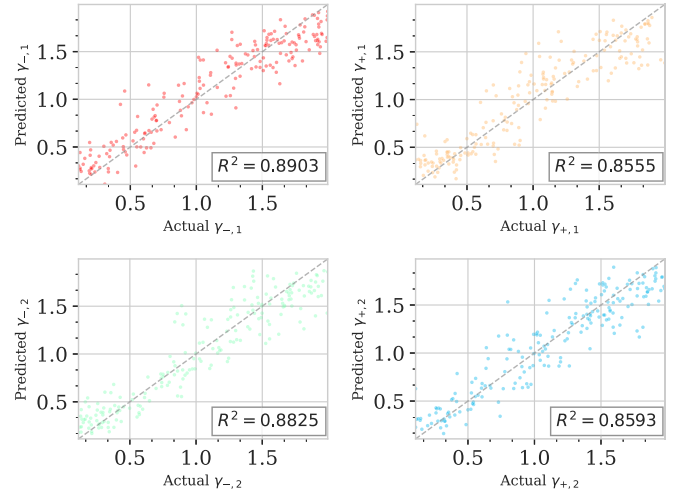


FIG. 7. Prediction results for the transverse-field Ising model with four constant dissipation rates $\gamma_{\pm,j}$ ($j = 1, 2$). Each dissipation rate is inferred from local observables $\langle \sigma_{x,y,z}^{(j)}(t) \rangle$ using a transformer-based regression model. This confirms the method's generalizability beyond the Heisenberg model shown in the main text.

Transverse-field Ising model (supplemental result). The Hamiltonian reads

$$H = J\sigma_z^{(1)}\sigma_z^{(2)} + h(\sigma_x^{(1)} + \sigma_x^{(2)}), \quad (C2)$$

with $(J, h) \sim \text{Uniform}(0.1, 2)^2$. The jump operators and dissipation structure are the same as in the Heisenberg case.

Training method (shared by both models)

For both models, we simulate 3000 trajectories. Each initial state is prepared as a tensor product of two Haar-random pure states. The systems evolve under the Lindblad equation with fixed dissipation rates $\gamma_{\pm,j}$. Time series of $\langle \sigma_{x,y,z}^{(j)}(t) \rangle$ for $j = 1, 2$ are recorded at 100 time steps over $t \in [0, 1]$.

From each observable component, 10 statistical features are extracted—mean, standard deviation, extrema, percentiles, slope, and interquartile range—resulting in 60-dimensional input vectors. These vectors are standardized and projected into a 128-dimensional embedding space. We use a 4-layer transformer encoder (8 heads, feedforward size 512), followed by an MLP decoder that predicts the four dissipation rates.

Training employs the AdamW optimizer ($\text{lr} = 10^{-3}$, weight decay 0.01), with ReduceLROnPlateau scheduling and early stopping (patience 30). Batch size is 32 and the model is trained for up to 300 epochs. The final models achieve $R^2 \geq 0.95$ on all outputs for both systems.

APPENDIX D: TRAINING DETAILS FOR THE PHOTON-SPIN COUPLING SYSTEM

This Appendix describes the training procedure for the transformer model used in the Jaynes-Cummings light-matter interaction setting. The goal is to predict six coefficients $(a_{0,\kappa}, a_{1,\kappa}, a_{2,\kappa}, a_{0,\gamma}, a_{1,\gamma}, a_{2,\gamma})$ that parametrize the time-dependent decay functions $\kappa(t)$ and $\gamma(t)$ as quadratic Bernstein polynomials.

Each data sample contains five observables recorded over 100 time steps: $\langle\sigma_x(t)\rangle$, $\langle\sigma_y(t)\rangle$, $\langle\sigma_z(t)\rangle$, $\langle a^\dagger a(t)\rangle$, and $\langle X_{\text{photon}} \otimes \sigma_z(t)\rangle$. From each signal, 18 time-series features are extracted—including statistical descriptors (mean, std, min, max, median, percentiles), shape (skewness, kurtosis), variation (range, IQR), temporal differences (mean/std of Δx , $\Delta^2 x$), slope, and spectral features (mean/max non-DC discrete Fourier transform amplitude). This results in a 90-dimensional input vector per sample.

All inputs and targets are standardized using z -score normalization. The 90-dimensional feature vector is projected into a 128-dimensional latent space via a feedforward block (Linear $\rightarrow$ LayerNorm $\rightarrow$ ReLU $\rightarrow$ Dropout). This is followed by a 4-layer transformer encoder (8 heads per layer, feedforward dimension 512). The encoded representation is passed through an MLP with layers of size 256, 128, and 64, followed by a final output layer producing six regression values.

Training uses the AdamW optimizer (initial learning rate 10^{-3} , weight decay 0.01) and mean squared error (MSE) loss. A ReduceLROnPlateau scheduler lowers the learning rate by a factor of 0.5 when validation loss stalls for 20 epochs. Early stopping is triggered after 30 epochs without improvement. We train with a batch size of 64 for up to 300 epochs. Gradient clipping (max norm 1.0) is applied for numerical stability.

Model performance is evaluated using the R^2 score and MSE on both training and held-out test sets. Predictions are inverse transformed to physical units and scatterplots are generated for each output coefficient. All models, logs, and output results are stored in time-stamped directories for full reproducibility.

APPENDIX E: LEARNING IDENTIFIABILITY IN LINDBLADIAN DYNAMICS

We begin with a basic identifiability result showing that the time-dependent decay rate $\gamma(t)$ can be uniquely recovered from local observables in a single-qubit open quantum system governed by the Lindblad equation.

Theorem E1. Consider a single qubit in Hilbert space $\mathbb{C}^2$ with a time-independent Hamiltonian $H = \sigma_z$ and a single jump operator $L = \sigma_-$. Then the dissipation rate $\gamma(t)$ is uniquely determined by the observable $\langle\sigma_z(t)\rangle$ alone, without knowledge of the initial state.

Proof. The Lindblad equation is given by

$$\dot{\rho} = -i[\sigma_z, \rho] + \gamma(t)(\sigma_- \rho \sigma_+ - \frac{1}{2}\{\sigma_+ \sigma_-, \rho\}).$$

We parametrize the density matrix as

$$\rho(t) = \begin{bmatrix} p_0(t) & \rho_{01}(t) \\ \rho_{01}^*(t) & p_1(t) \end{bmatrix}, \quad p_0(t) + p_1(t) = 1.$$

Compute each term in the equation:

(i) The unitary part:

$$-i[\sigma_z, \rho] = \begin{bmatrix} 0 & -2i\rho_{01} \\ 2i\rho_{01}^* & 0 \end{bmatrix}.$$

(ii) The dissipative part:

$$\sigma_- \rho \sigma_+ = \begin{bmatrix} 0 & 0 \\ p_0 & 0 \end{bmatrix}, \quad \frac{1}{2}\{\sigma_+ \sigma_-, \rho\} = \begin{bmatrix} p_0 & \rho_{01} \\ \rho_{01}^* & p_1 \end{bmatrix},$$

so the dissipator becomes

$$\mathcal{D}[\sigma_-](\rho) = \begin{bmatrix} -p_0 & -\rho_{01} \\ p_0 - \rho_{01}^* & -p_1 \end{bmatrix}.$$

Combining both terms, we obtain the differential equations:

$$\dot{p}_0 = -\gamma(t)p_0, \quad (\text{E1})$$

$$\dot{p}_1 = \gamma(t)p_0, \quad (\text{E2})$$

$$\dot{\rho}_{01} = -2i\rho_{01} - \gamma(t)\rho_{01}. \quad (\text{E3})$$

Now, $\langle\sigma_z\rangle = p_0 - p_1 = 2p_0 - 1$, and hence

$$p_0(t) = \frac{1 + \langle\sigma_z(t)\rangle}{2}.$$

Solving $\dot{p}_0 = -\gamma(t)p_0$ gives

$$p_0(t) = p_0(0)e^{-\int_0^t \gamma(s) ds},$$

Therefore,

$$\langle\sigma_z(t)\rangle = -1 + 2p_0(0)e^{-\int_0^t \gamma(s) ds}.$$

By differentiating, we get

$$\frac{d}{dt}\langle\sigma_z(t)\rangle = -2p_0(0)\gamma(t)e^{-\int_0^t \gamma(s) ds}.$$

By rewriting, we get

$$\gamma(t) = -\frac{1}{1 + \langle\sigma_z(t)\rangle} \frac{d}{dt}\langle\sigma_z(t)\rangle. \quad (\text{E4})$$

This expression shows that $\gamma(t)$ can be reconstructed from $\langle\sigma_z(t)\rangle$ alone, regardless of the initial state. This validates the ability of our model to infer $\gamma(t)$ directly from observable trajectories. ■

Theorem E2. Consider a single qubit in $\mathbb{C}^2$ with time-independent Hamiltonian $H = \sigma_z$, and jump operators $L_1 = \sigma_-$ and $L_2 = \sigma_+$ associated with time-dependent dissipation rates $\gamma_-(t)$ and $\gamma_+(t)$, respectively. Then, the functions $\gamma_+(t)$ and $\gamma_-(t)$ can be uniquely determined from the observables $\langle\sigma_x(t)\rangle$, $\langle\sigma_y(t)\rangle$, and $\langle\sigma_z(t)\rangle$.

Proof. Following the single-jump case, we now consider a more general dissipative setting with two jump channels. The master equation becomes

$$\begin{aligned} \dot{\rho} = & -i[\sigma_z, \rho] + \gamma_-(t)(\sigma_- \rho \sigma_+ - \frac{1}{2}\{\sigma_+ \sigma_-, \rho\}) \\ & + \gamma_+(t)(\sigma_+ \rho \sigma_- - \frac{1}{2}\{\sigma_- \sigma_+, \rho\}). \end{aligned} \quad (\text{E5})$$

Let $\rho(t)$ be written in the standard basis:

$$\rho(t) = \begin{bmatrix} p_0(t) & \rho_{01}(t) \\ \rho_{01}^*(t) & p_1(t) \end{bmatrix}, \quad p_0(t) + p_1(t) = 1.$$

Computing each Lindblad term yields

$$\begin{aligned} \sigma_- \rho \sigma_+ &= \begin{bmatrix} 0 & 0 \\ p_0 & 0 \end{bmatrix}, & \sigma_+ \rho \sigma_- &= \begin{bmatrix} 0 & p_1 \\ 0 & 0 \end{bmatrix}, & (\text{E6}) \\ \{\sigma_+ \sigma_-, \rho\}/2 &= \begin{bmatrix} p_0 & \rho_{01} \\ \rho_{01}^* & p_1 \end{bmatrix}, & \{\sigma_- \sigma_+, \rho\}/2 &= \begin{bmatrix} p_0 & \rho_{01} \\ \rho_{01}^* & p_1 \end{bmatrix}. & (\text{E7}) \end{aligned}$$

Therefore, the two dissipative contributions are

$$\gamma_-(t)(\cdots) = \gamma_-(t) \begin{bmatrix} -p_0 & -\rho_{01} \\ p_0 - \rho_{01}^* & -p_1 \end{bmatrix}, \quad (\text{E8})$$

$$\gamma_+(t)(\cdots) = \gamma_+(t) \begin{bmatrix} -p_0 & p_1 - \rho_{01} \\ -\rho_{01}^* & -p_1 \end{bmatrix}. \quad (\text{E9})$$

Combining all terms, the equations of motion for each element of ρ are

$$\dot{p}_0 = -\gamma_-(t) p_0 + \gamma_+(t) p_1, \quad (\text{E10})$$

$$\dot{p}_1 = -\dot{p}_0, \quad (\text{E11})$$

$$\dot{\rho}_{01} = -2i\rho_{01} - \frac{1}{2}(\gamma_-(t) + \gamma_+(t))\rho_{01}. \quad (\text{E12})$$

The Pauli expectation values are

$$\langle \sigma_x \rangle = 2 \text{Re}(\rho_{01}), \quad \langle \sigma_y \rangle = -2 \text{Im}(\rho_{01}), \quad (\text{E13})$$

$$\gamma_-(t) = \frac{4\langle \sigma_x \rangle \langle \sigma_z \rangle - i\langle \sigma_x \rangle \langle \dot{\sigma}_z \rangle - 4i\langle \sigma_y \rangle \langle \sigma_z \rangle - \langle \sigma_y \rangle \langle \dot{\sigma}_z \rangle - 2i\langle \sigma_z \rangle \langle \dot{\sigma}_x \rangle - 2\langle \sigma_z \rangle \langle \dot{\sigma}_y \rangle - 6i\langle \dot{\sigma}_x \rangle - 6\langle \dot{\sigma}_y \rangle}{2i\langle \sigma_x \rangle \langle \sigma_z \rangle + 4i\langle \sigma_x \rangle + 2\langle \sigma_y \rangle \langle \sigma_z \rangle + 4\langle \sigma_y \rangle}. \quad (\text{E17})$$

Once $\gamma_-(t)$ is obtained, $\gamma_+(t)$ can be recovered by substitution into the differential equations above. These expressions are exact and can be validated numerically against simulation data generated using qutip. ■

We now generalize our theoretical analysis to multiqubit open quantum systems. While exact analytical formulas such as those in the single-qubit case become intractable, we can still express the system evolution in a linear operator basis and outline the principle of identifiability. This justifies the feasibility of using machine learning to approximate the inversion from observable trajectories to underlying dissipation rates.

1. Pauli basis representation

Let ρ be the n -qubit density matrix, which can be expanded in the Pauli basis as

$$\rho(t) = \sum_{\mu \in G_n} a_\mu(t) \mu,$$

where G_n denotes the n -qubit Pauli group and $\{a_\mu(t)\}$ are real coefficients due to the Hermiticity of ρ . The set G_n spans a basis of dimension 4^n and the orthogonality relation $\text{Tr}(\mu\nu) = \delta_{\mu,\nu}$ holds for $\mu, \nu \in G_n$.

Assume the Hamiltonian and Lindblad operators admit the following Pauli expansions:

$$H = \sum_{\mu \in G_n} h_\mu \mu, \quad L_j = \sum_{\mu \in G_n} L_j^\mu \mu.$$

We define the structure constants of the Pauli algebra via

$$[\mu, \nu] = \sum_{\lambda \in G_n} C_{\mu,\nu}^\lambda \lambda.$$

$$\langle \sigma_z \rangle = p_0 - p_1 = 2p_0 - 1. \quad (\text{E14})$$

Define the complex observable:

$$\frac{\langle \sigma_x \rangle - i\langle \sigma_y \rangle}{2} = \rho_{01}.$$

Taking time derivatives, we obtain two equations involving $\gamma_\pm(t)$:

$$\frac{1}{2} \frac{d}{dt} \langle \sigma_z(t) \rangle = \gamma_+(t) - \gamma_-(t) \cdot \frac{1 + \langle \sigma_z(t) \rangle}{2}, \quad (\text{E15})$$

$$\frac{d}{dt} \left(\frac{\langle \sigma_x \rangle - i\langle \sigma_y \rangle}{2} \right) = \left\{ -2i - \frac{1}{2}[\gamma_+(t) + \gamma_-(t)] \right\} \cdot \rho_{01}. \quad (\text{E16})$$

Solving these equations algebraically yields explicit expressions for $\gamma_+(t)$ and $\gamma_-(t)$ in terms of the observables and their time derivatives. We denote time derivatives as $\langle \dot{O} \rangle := \frac{d}{dt} \langle O(t) \rangle$ for any observable O . One can solve for $\gamma_-(t)$ as

2. Time evolution of coefficients

Substituting the Pauli expansions into the Lindblad master equation, the derivative of the density matrix is

$$\begin{aligned} \sum_{\xi} \dot{a}_\xi \xi &= -i \sum_{\mu, \nu, \xi} h_\mu a_\nu C_{\mu,\nu}^\xi \xi + \sum_j \gamma_j \sum_{\mu, \nu, \lambda, \omega, \xi} \\ &\times \left\{ L_j^\mu a_\nu (L_j^\lambda)^* - \frac{1}{2} [(L_j^\mu)^* a_\lambda L_j^\nu + (L_j^\nu)^* a_\mu L_j^\lambda] \right\} \\ &\times C_{\mu,\nu}^\omega C_{\omega,\lambda}^\xi \xi. \end{aligned} \quad (\text{E18})$$

Here, $\dot{a}_\xi$ represents the time derivative $\frac{da_\xi(t)}{dt}$. By orthogonality, this yields a linear system for the coefficients,

$$\begin{aligned} \dot{a}_\xi &= -i \sum_{\mu, \nu} h_\mu C_{\mu,\nu}^\xi a_\nu + \sum_j \gamma_j \sum_{\mu, \nu, \lambda, \omega} \\ &\times \left\{ L_j^\mu a_\nu (L_j^\lambda)^* - \frac{1}{2} [(L_j^\mu)^* a_\lambda L_j^\nu + (L_j^\nu)^* a_\mu L_j^\lambda] \right\} \\ &\times C_{\mu,\nu}^\omega C_{\omega,\lambda}^\xi. \end{aligned} \quad (\text{E19})$$

This equation expresses $\dot{a}_\xi$ as a linear combination of γ_j with coefficients that depend polynomially on $\{a_\mu(t)\}_{\mu \in G_n}$. Therefore, if the set $\{L_j\}$ is linearly independent as operators on $\mathbb{C}^{2^n}$, then for any ξ , the corresponding differential equations can be inverted to recover $\gamma_j(t)$, in principle.

3. Discussion and practical implications

Although this linear system is exact, the inversion is prohibitively complex even for moderate n , involving $O(4^n)$ terms and tensor contractions over multiple indices. Moreover, estimating $\{a_\mu(t)\}$ from experiment requires full quantum state tomography. The sample complexity for estimating a rank- r n -qubit state to Frobenius error ϵ is $O(2^n/\epsilon)$ [60].

Several efficient numerical methods exist, including Riemannian gradient descent over the quantum state manifold [61].

This analysis supports our learning framework: instead of symbolically inverting this high-dimensional linear system, we use time-series observables as proxy input and train a transformer to approximate the underlying mapping to $\gamma_j(t)$. The supervised data are generated by simulating the Lindblad evolution using tools such as qutip [53].

In principle, a universal estimator can learn the function class $\{\gamma_j(t)\}$ through such indirect supervision, even when exact formulas are intractable.

4. Photonic system learning ability

Theorem E3. Consider the open Jaynes-Cummings model with master equation

$$\dot{\rho}(t) = -i[H, \rho(t)] + \kappa(t)\mathcal{D}[a]\rho(t) + \gamma(t)\mathcal{D}[\sigma_-]\rho(t), \quad (\text{E20})$$

where $H = \omega_c a^\dagger a + \frac{\omega_q}{2}\sigma_z + g(a^\dagger\sigma_- + a\sigma_+)$ is the Jaynes-Cummings Hamiltonian and $\mathcal{D}[L]\rho = L\rho L^\dagger - \frac{1}{2}\{L^\dagger L, \rho\}$ is the Lindblad dissipator. Then, given a fixed initial state $\rho(0)$, the dissipation rates $\gamma(t)$ and $\kappa(t)$ are uniquely determined over $[0, T]$ by the time series of five observables:

$$\{\langle\sigma_x(t)\rangle, \langle\sigma_y(t)\rangle, \langle\sigma_z(t)\rangle, \langle a^\dagger a(t)\rangle, \langle X \otimes \sigma_z(t)\rangle\}.$$

Proof. Under the Heisenberg evolution, the observables satisfy the following coupled ordinary differential equations:

$$\frac{d}{dt}\langle\sigma_x\rangle = \omega_q\langle\sigma_y\rangle + ig\langle(a^\dagger - a)\sigma_z\rangle - \frac{\gamma(t)}{2}\langle\sigma_x\rangle, \quad (\text{E21})$$

$$\frac{d}{dt}\langle\sigma_y\rangle = -\omega_q\langle\sigma_x\rangle - g\langle(a + a^\dagger)\sigma_z\rangle - \frac{\gamma(t)}{2}\langle\sigma_y\rangle, \quad (\text{E22})$$

$$\frac{d}{dt}\langle\sigma_z\rangle = -2ig\langle(a\sigma_+ - a^\dagger\sigma_-)\rangle - \gamma(t)(1 + \langle\sigma_z\rangle), \quad (\text{E23})$$

$$\frac{d}{dt}\langle a^\dagger a\rangle = -ig\langle(a^\dagger\sigma_- - a\sigma_+)\rangle - \kappa(t)\langle a^\dagger a\rangle. \quad (\text{E24})$$

The terms involving $\langle a\sigma_+ \rangle$ and $\langle a^\dagger\sigma_- \rangle$ are internal to the dynamics and fully determined by the state $\rho(t)$. Therefore, $\gamma(t)$ and $\kappa(t)$ can be algebraically inferred from observable trajectories, provided the system is fully observed.

In practice, we found that using only $\langle\sigma_{x,y,z}(t)\rangle$ and $\langle a^\dagger a(t)\rangle$ leads to degraded performance in learning. To address this, we include the Hermitian cross observable $\langle X_{\text{photon}} \otimes \sigma_z \rangle$, which empirically improves identifiability and allows the transformer to accurately recover the dissipation profiles. This confirms that under appropriate observability conditions, the decay rates are theoretically and practically identifiable. ■

APPENDIX F: IMPLEMENTATION DETAILS FOR SCALING TESTS

Here we summarize the architecture and training setup used in the scaling benchmarks presented in Sec. III.

1. Architectures of sparse test

While the two-qubit benchmarks already demonstrate the ability of our framework to infer multiple dissipation channels in interacting systems, such small-scale examples leave open

the question of scalability. To further assess performance in larger settings, we now turn to Heisenberg spin chains with up to nine qubits, where the number of degrees of freedom and dissipation parameters grow systematically with n . These benchmarks allow us to test whether the method remains effective as the system size increases.

2. Scaling tests on multiqubit Heisenberg chains

For an N -qubit chain ($N = 2, \dots, 9$) evolving under homogeneous Heisenberg dynamics, we design a size-agnostic transformer [12] $\mathcal{H}_N$ that infers the uniform exchange coupling J from local spin-projection traces alone. From every time series $\langle\sigma_z^{(k)}(t)\rangle$, we extract the same 18-dimensional descriptor used in the Lindblad study, basic statistics, linear trend, first- and second-order differentials, shape factors, and spectral extrema, yielding an input vector $x \in \mathbb{R}^{18N}$ that grows linearly from 36 ($N = 2$) to 162 ($N = 9$).

A linear + LayerNorm + ReLU front end embeds x into a $d_{\text{model}} = 128$ latent space processed by four transformer-encoder blocks with $h = 8$ heads and 512-unit feedforward layers (drop-out 0.2). The context token is funnelled through a three-stage LayerNorm MLP ($256 \rightarrow 128 \rightarrow 64$) that outputs a single scalar, providing J in one forward pass. All hyperparameters are kept fixed across N , giving $\sim 10^5$ weights and enabling parameter-efficient scaling from two to nine qubits. Training on synthetic data generated from exact Heisenberg time evolutions employs AdamW, a cosine-annealed learning-rate schedule, early stopping, and an ℓ_2 loss over 400 epochs, achieving subpercent test MSE throughout. These results show that global self-attention acting on carefully engineered one-dimensional Pauli signatures suffices to reconstruct collective exchange interactions, underscoring the suitability of attention mechanisms for efficient Hamiltonian learning.

3. Architectures of dense test

We construct a scalable family of transformer [12] regressors $\mathcal{T}_N$ ($N = 2, \dots, 7$) that predicts the $2N$ dissipative rates $\{\gamma_{\pm k}\}_{k=1}^N$ of an N -qubit dense-Lindblad master equation directly from single-qubit Pauli expectation traces. From every time series $\langle\sigma_\alpha^{(k)}(t)\rangle$ ($\alpha = x, y, z$), we extract an 18-dimensional descriptor comprising basic statistics (mean, standard deviation, extrema, quartiles, interquartile range), linear trend, first- and second-order finite differences, shape factors (skewness, kurtosis), and spectral fingerprints (mean and peak FFT amplitude). Concatenating the descriptors of the $3N$ traces yields a fixed-length input $x \in \mathbb{R}^{18 \cdot 3N}$ (108–378 features for $N = 2$ –7).

A linear-LayerNorm-ReLU projection embeds x into a 128-dimensional latent space that is processed by $L = 4$ stacked transformer encoder blocks with $h = 8$ attention heads and 512-unit positionwise feedforward layers (drop-out 0.2). The context vector is mapped by a three-stage LayerNorm MLP to the target dimension $2N$, yielding the Lindblad-rate estimates in a single forward pass. All hyperparameters are held fixed across N , resulting in $\sim 10^5$ trainable weights and enabling parameter-efficient scalability.

Training on synthetic trajectories generated from exact dense-Lindblad dynamics employs AdamW, cosine-decayed learning rate, early stopping, and mean squared error loss over

400 epochs. The models achieve subpercent test MSE across all N , demonstrating that carefully engineered local Pauli signatures, when coupled with global self-attention, suffice to

reconstruct the full dissipative kernel—underscoring the potential of attention mechanisms for efficient quantum-process tomography.

-
- [1] H.-P. Breuer, E.-M. Laine, J. Piilo, and B. Vacchini, Colloquium: Non-Markovian dynamics in open quantum systems, *Rev. Mod. Phys.* **88**, 021002 (2016).
 - [2] Á. Rivas and S. F. Huelga, *Open Quantum Systems*, Vol. 10 (Springer, New York, 2012).
 - [3] M. S. Sarandy and D. A. Lidar, Adiabatic quantum computation in open systems, *Phys. Rev. Lett.* **95**, 250503 (2005).
 - [4] L. Banchi, E. Grant, A. Rocchetto, and S. Severini, Modelling non-Markovian quantum processes with recurrent neural networks, *New J. Phys.* **20**, 123030 (2018).
 - [5] K. Fujimoto, R. Hamazaki, and Y. Kawaguchi, Impact of dissipation on universal fluctuation dynamics in open quantum systems, *Phys. Rev. Lett.* **129**, 110403 (2022).
 - [6] S. Yin, P. Mai, and F. Zhong, Nonequilibrium quantum criticality in open systems: The dissipation rate as an additional indispensable scaling variable, *Phys. Rev. B* **89**, 094108 (2014).
 - [7] G. Barontini, R. Labouvie, F. Stubenrauch, A. Vogler, V. Guarrera, and H. Ott, Controlling the dynamics of an open many-body quantum system with localized dissipation, *Phys. Rev. Lett.* **110**, 035302 (2013).
 - [8] C. L. Degen, F. Reinhard, and P. Cappellaro, Quantum sensing, *Rev. Mod. Phys.* **89**, 035002 (2017).
 - [9] J. Bylander, S. Gustavsson, F. Yan, F. Yoshihara, K. Harrabi, G. Fitch, D. G. Cory, Y. Nakamura, J.-S. Tsai, and W. D. Oliver, Noise spectroscopy through dynamical decoupling with a superconducting flux qubit, *Nat. Phys.* **7**, 565 (2011).
 - [10] J. Wang, L. Davidovich, and G. S. Agarwal, Quantum sensing of open systems: Estimation of damping constants and temperature, *Phys. Rev. Res.* **2**, 033389 (2020).
 - [11] F. Wolf and P. O. Schmidt, Quantum sensing of oscillating electric fields with trapped ions, *Measur. Sensors* **18**, 100271 (2021).
 - [12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, Attention is all you need, in *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, USA, 2017), pp. 6000–6010.
 - [13] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y. Wang, and X. Yan, Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting, in *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, USA, 2019), pp. 5243–5253.
 - [14] G. Zerveas, S. Jayaraman, D. Patel, A. Bhamidipaty, and C. Eickhoff, A transformer-based framework for multivariate time series representation learning, in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD '21)* (unpublished).
 - [15] B. Lim and S. Zohren, Nicolas Loeff, and Tomas Pfister, Temporal fusion transformers” for interpretable multi-horizon time series forecasting, *Intl. J. Forecast.* **37**, 1748 (2021).
 - [16] H. Qiao, Y. P. Kandel, K. Deng, S. Fallahi, G. C. Gardner, M. J. Manfra, E. Barnes, and J. M. Nichol, Coherent multispin exchange coupling in a quantum-dot spin chain, *Phys. Rev. X* **10**, 031006 (2020).
 - [17] Z. Zhou, Y. Li, Z. Wu, X. Ma, S. Fan, and S. Huang, The exchange interaction between neighboring quantum dots: Physics and applications in quantum information processing, *J. Semicond.* **45**, 101701 (2024).
 - [18] X. Hu and S. Das Sarma, Overview of spin-based quantum dot quantum computation, *Phys. Status Solidi B* **238**, 360 (2003).
 - [19] K. W. Chan, H. Sahasrabudhe, W. Huang, Y. Wang, H. C. Yang, M. Veldhorst, J. C. C. Hwang, F. A. Mohiyaddin, F. E. Hudson, K. M. Itoh, and A. Morello, Exchange coupling in a linear chain of three quantum-dot spin qubits in silicon, *Nano Lett.* **21**, 1517 (2021).
 - [20] Y. P. Kandel, H. Qiao, S. Fallahi, G. C. Gardner, M. J. Manfra, and J. M. Nichol, Coherent spin-state transfer via Heisenberg exchange, *Nature (London)* **573**, 553 (2019).
 - [21] B. W. Shore and P. L. Knight, The Jaynes-Cummings model, *J. Mod. Opt.* **40**, 1195 (1993).
 - [22] A. R. Morgillo, M. F. Sacchi, and C. Macchiavello, Detecting markovianity of quantum processes via recurrent neural networks, *Quantum Mach. Intell.* **7**, 47 (2025).
 - [23] S. Martina, S. Gherardini, and F. Caruso, Machine learning classification of non-markovian noise disturbing quantum dynamics, *Physica Scr.* **98**, 035104 (2023).
 - [24] P. P. Mazza, D. Zietlow, F. Carollo, S. Andergassen, G. Martius, and I. Lesanovsky, Machine learning time-local generators of open quantum dynamics, *Phys. Rev. Res.* **3**, 023084 (2021).
 - [25] C. Guo, K. Modi, and D. Poletti, Tensor-network-based machine learning of non-Markovian quantum processes, *Phys. Rev. A* **102**, 062414 (2020).
 - [26] I. A. Luchnikov, E. O. Kiktenko, M. A. Gavreev, H. Ouerdane, S. N. Filippov, and A. K. Fedorov, Probing non-Markovian quantum dynamics with data-driven analysis: Beyond “black-box” machine-learning models, *Phys. Rev. Res.* **4**, 043002 (2022).
 - [27] E. Bairey, C. Guo, D. Poletti, N. H. Lindner, and I. Arad, Learning the dynamics of open quantum systems from their steady states, *New J. Phys.* **22**, 032001 (2020).
 - [28] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, *Rev. Mod. Phys.* **91**, 045002 (2019).
 - [29] S. J. Wetzel, Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders, *Phys. Rev. E* **96**, 022140 (2017).
 - [30] E. P. L. Van Nieuwenburg, Y.-H. Liu, and S. D. Huber, Learning phase transitions by confusion, *Nat. Phys.* **13**, 435 (2017).
 - [31] E.-J. Kuo, A. Seif, R. Lundgren, S. Whitsitt, and M. Hafezi, Decoding conformal field theories: From supervised to unsupervised learning, *Phys. Rev. Res.* **4**, 043031 (2022).
 - [32] E.-J. Kuo and H. Dehghani, Unsupervised learning of interacting topological and symmetry-breaking phase transitions, *Phys. Rev. B* **105**, 235136 (2022).

- [33] E. van Nieuwenburg, E. Bairey, and G. Refael, Learning phase transitions from dynamics, *Phys. Rev. B* **98**, 060301(R) (2018).
- [34] F. Schindler, N. Regnault, and T. Neupert, Probing many-body localization with neural networks, *Phys. Rev. B* **95**, 245134 (2017).
- [35] J. Carifio, J. Halverson, D. Krioukov, and B. D. Nelson, Machine learning in the string landscape, *J. High Energy Phys.* **09** (2017) 157.
- [36] K. Hashimoto, S. Sugishita, A. Tanaka, and A. Tomiya, Deep learning and the AdS/CFT correspondence, *Phys. Rev. D* **98**, 046019 (2018).
- [37] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [38] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Neural-network quantum state tomography, *Nat. Phys.* **14**, 447 (2018).
- [39] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, Reconstructing quantum states with generative models, *Nat. Mach. Intell.* **1**, 155 (2019).
- [40] G. Torlai, B. Timar, E. P. L. Van Nieuwenburg, H. Levine, A. Omran, A. Keesling, H. Bernien, M. Greiner, V. Vuletić, M. D. Lukin *et al.*, Integrating neural networks with a quantum simulator for state reconstruction, *Phys. Rev. Lett.* **123**, 230504 (2019).
- [41] C.-Y. Liu, E.-J. Kuo, C.-H. A. Lin, S. Chen, J. G. Young, Y.-J. Chang, and M.-H. Hsieh, Training classical neural networks by quantum machine learning, in *IEEE International Conference on Quantum Computing and Engineering (QCE)*, Vol. 2 (IEEE, 2024), pp. 34–38.
- [42] C.-Y. Liu, E.-J. Kuo, C.-H. A. Lin, J. G. Young, Y.-J. Chang, M.-H. Hsieh, and H.-S. Goan, Quantum-train: Rethinking hybrid quantum-classical machine learning in the model compression perspective, *Quantum Mach. Intell.* **7**, 80 (2025).
- [43] S.-T. Tsai, E.-J. Kuo, and P. Tiwary, Learning molecular dynamics with simple language model built upon long short-term memory neural network, *Nat. Commun.* **11**, 5115 (2020).
- [44] S.-T. Tsai, E. Fields, Y. Xu, E.-J. Kuo, and P. Tiwary, Path sampling of recurrent neural networks by incorporating known physics, *Nat. Commun.* **13**, 7231 (2022).
- [45] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning representations by back-propagating errors, *Nature (London)* **323**, 533 (1986).
- [46] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, [arXiv:1406.1078](https://arxiv.org/abs/1406.1078).
- [47] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Comput.* **9**, 1735 (1997).
- [48] J. Kaplan, S. McCandlish, T. Henighan *et al.*, Scaling laws for neural language models, [arXiv:2001.08361](https://arxiv.org/abs/2001.08361).
- [49] Z. Li, H. Zheng, N. Kovachki, D. Jin, H. Chen, B. Liu, K. Azizzadenesheli, and A. Anandkumar, Physics-informed neural operator for learning partial differential equations, *ACM/IMS J. Data Science* **1**, 1 (2024).
- [50] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2002).
- [51] P. Pearle, Simple derivation of the Lindblad equation, *Eur. J. Phys.* **33**, 805 (2012).
- [52] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2002).
- [53] J. R. Johansson, P. D. Nation, and F. Nori, Qutip: An open-source python framework for the dynamics of open quantum systems, *Comput. Phys. Commun.* **183**, 1760 (2012).
- [54] S. Sachdev, *Quantum Phase Transitions* (Cambridge University Press, Cambridge, 2011).
- [55] R. Islam, E. E. Edwards, K. Kim, S. Korenblit, C. Noh, H. Carmichael, and C. Monroe, Onset of a quantum phase transition with a trapped ion quantum simulator, *Nat. Commun.* **2**, 377 (2011).
- [56] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, and C. Monroe, Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator, *Nature (London)* **551**, 601 (2017).
- [57] I. Georgescu, The Divincenzo criteria 20 years on, *Nat. Rev. Phys.* **2**, 666 (2020).
- [58] E.-J. Kuo and C.-S. Chen, Data for “Unraveling quantum environments: Transformer-assisted learning in Lindblad dynamics” (v1.0) [Data set], Zenodo (2025), doi:[10.5281/zenodo.17139233](https://doi.org/10.5281/zenodo.17139233).
- [59] M. H. Stone, The generalized Weierstrass approximation theorem, *Math. Mag.* **21**, 237 (1948).
- [60] R. O’Donnell and J. Wright, Efficient quantum tomography, in *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2016), pp. 899–912.
- [61] M.-C. Hsu, E.-J. Kuo, W.-H. Yu, J.-F. Cai, and M.-H. Hsieh, Quantum state tomography via nonconvex Riemannian gradient descent, *Phys. Rev. Lett.* **132**, 240804 (2024).