







Probing non-Markovian quantum dynamics with data-driven analysis: Beyond “black-box” machine-learning models

I. A. Luchnikov ^{1,2,3,*}, E. O. Kiktenko ^{1,4,5}, M. A. Gavreev ^{1,3,4}, H. Ouerdane ²,
S. N. Filippov ^{4,5,6} and A. K. Fedorov ^{1,3,4,†}

¹Russian Quantum Center, Skolkovo, Moscow 143025, Russia

²Skolkovo Institute of Science and Technology, Moscow 121205, Russia

³National University of Science and Technology “MISIS”, Moscow 119049, Russia

⁴Moscow Institute of Physics and Technology, Moscow Region 141700, Russia

⁵Department of Mathematical Methods for Quantum Technologies,

Sklov Mathematical Institute of Russian Academy of Sciences, Moscow 119991, Russia

⁶Valiev Institute of Physics and Technology of Russian Academy of Sciences, Moscow 117218, Russia



(Received 11 June 2021; revised 16 May 2022; accepted 19 July 2022; published 3 October 2022)

A precise understanding of the influence of a quantum system’s environment on its dynamics, which is at the heart of the theory of open quantum systems, is crucial for further progress in the development of controllable large-scale quantum systems. However, existing approaches to account for complex system-environment interaction in the presence of memory effects are either based on heuristic and oversimplified principles or give rise to computational difficulties. In practice, one can leverage on available experimental data and replace first-principles simulations with a data-driven analysis that is often much simpler. Inspired by recent advances in data analysis and machine learning, we propose a data-driven approach to the analysis of the non-Markovian dynamics of open quantum systems. Our method allows, on the one hand, capturing the most important characteristics of open quantum systems such as the effective dimension of the environment and the spectrum of the joint system-environment quantum dynamics, and, on the other hand, reconstructing a predictive model of non-Markovian quantum dynamics, and denoising the measured quantum trajectories. We demonstrate the performance of the proposed approach with various models of open quantum systems, including a qubit coupled with a finite environment, a spin-boson model, and the damped Jaynes-Cummings model.

DOI: [10.1103/PhysRevResearch.4.043002](https://doi.org/10.1103/PhysRevResearch.4.043002)

I. INTRODUCTION

Understanding and predicting dynamics of quantum many-body systems is a formidable challenge, and an essential step toward the development of quantum computing devices [1]. Recent experiments [2–9] have demonstrated the realization of quantum computing and simulation protocols with quantum many-body systems of an intermediate scale (dozens of particles). Nevertheless, even with the achieved outstanding level of control in these experiments it remains challenging to isolate a quantum system from the environment [9], which is the main source of decoherence causing errors. This fact significantly limits applications of such systems for quantum information processing and prevents the full experimental verification of various effects related to the nonequilibrium dynamics [10–12].

The theory of open quantum systems [13–16] aims to provide a complete description of such nonequilibrium dynamical effects at the quantum level. Within such a framework, one can arbitrarily partition a quantum many-body system and its environment and describe each part as an individual open system. This powerful idea allows studying a nonequilibrium many-body system by considering smaller parts of it; for example, an individual qubit in a large ensemble of qubits interacting with each other and their environment. At first sight, dealing with one qubit instead of an ensemble of entangled qubits seems to be a much simpler problem, because it does not suffer from the computational difficulties raised by the exponentially large Hilbert space in the number of subsystems. Indeed, the dynamics of the subsystem’s density matrix (e.g., one qubit density matrix) can be described exactly by the Nakajima-Zwanzig equation [17,18], which contains a convolution (memory) kernel. Time-convolution is necessary to take into account the non-Markovianity of the general open quantum dynamics. Non-Markovianity, i.e., effects of memory, are among the most challenging problems to solve for the proper description of the general open quantum system [14]. The problem is that the exact derivation of the Nakajima-Zwanzig equation is of the same complexity as the calculation of the coupled system and environment quantum dynamics [13].

*luchnikovilya@gmail.com

†akf@rqc.ru

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.

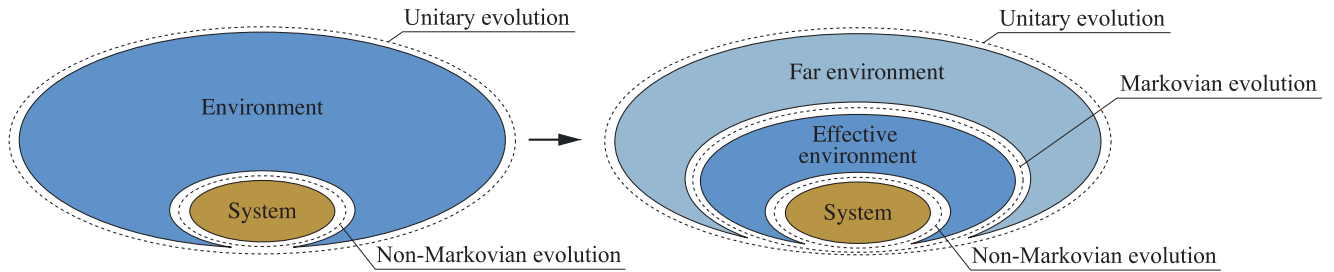


FIG. 1. Physical intuition underpinning the Markovian embedding for the non-Markovian dynamics of the finite-dimensional system interacting with a high-dimensional (possibly infinite-dimensional) environment. One can split the system’s environment into two parts: The finite-dimensional effective environment, which is responsible for memory effects in the system’s dynamics, and the remaining far environment responsible for dissipation. It allows one to consider a Markovian dissipative dynamics of the finite-dimensional extended system consisting of the original system and its effective environment.

This problem can be approached by considering different approximations such as the Born-Markov approximation [16,19] or simplified models of open quantum dynamics that are exactly solvable [20–23]. However, the class of problems that can be studied analytically is limited to a handful of exceptional situations. An alternative approach is to use numerical techniques such as the non-Markovian quantum state diffusion [24–27], the hierarchical equations of motion [28,29], the time-evolving matrix product operators [30], the method based on optimized auxiliary oscillators [31], the dressed quantum trajectories method [32], the Dirac-Frenkel time-dependent variational approach with the Davydov ansatz [33], or the time-evolving density with orthogonal polynomials algorithm [34,35], just to name a few. At the same time, open quantum systems studied under real experimental conditions are usually too complex and cannot always be described with existing numerical methods developed for illustrative models. For instance, most of the numerical methods, including those listed above, are devoted to simulating systems with an environment made of noninteracting quantum oscillators [14] or noninteracting fermions [36,37], so they are unable to simulate systems with more complex environments, such as the spin environment [38–42]. Moreover, since in a real experimental setting the joint Hamiltonian of a system and its environment is often unknown, exhaustive spectroscopy experiments are required to recover the Hamiltonian parameters and simulate the actual system’s dynamics at the microscopic level [43,44].

Since building a precise analytical or numerical description of open quantum dynamics from a microscopic model is a notoriously difficult problem in the general case, one can try to build a precise data-driven model of open quantum dynamics using experimentally measured data. In addition, data-driven approaches do not require reconstruction of the joint system and environment microscopic model. A number of data-driven techniques for the analysis of the dynamics of open systems have been recently proposed. For example, Ref. [45] introduces a method for a data-driven reconstruction of the discrete-time Nakajima-Zwanzig equation, namely the transfer-tensor method (TTM) [46–54], that can be used for dynamics prediction. In Ref. [55] a recurrent neural network based method [56] for data-driven identification of open quantum dynamics has been developed. Although these methods provide efficient predictive models, they serve as “black

boxes,” i.e., they do not allow one to unravel the physical picture of the underlying processes [57–59].

One of the most natural ways to describe open quantum dynamics is embedding the non-Markovian system dynamics into a Markovian dynamics of the system and the effective environment of a finite dimension [60]. The physical intuition behind such embedding is summarized in Fig. 1. One can think of the environment that induces the system’s non-Markovian dynamics as a two-component system consisting of effective and far environments. The effective environment is responsible for memory effects: some information about the system is recorded into the effective environment and then affects the system at later time. The dimension of the effective environment determines the complexity of the non-Markovian system’s dynamics. In contrast to the effective environment, the interaction with the far environment does not entail a backflow of information, and leads to a purely dissipative dynamics. In this way, the system and the effective environment instantiate a Markovian embedding, and their collective dynamics is both dissipative and Markovian. We emphasize that Markovian embedding is not just a “black-box” model of open quantum dynamics since it provides insights into the properties of both the system and its environment. Moreover, Markovian embedding can be reconstructed with only information about the non-Markovian dynamics of a system, which is accessible in real experimental conditions [61,62].

In this work, we harness data-driven Markovian embedding reconstruction, and present a complete framework for general open quantum dynamics analysis. Our framework can be seen as a combination of the alternative “data-friendly” view on non-Markovian quantum dynamics, that we develop in first sections, and linear machine-learning methods [63–66]. Input data include trajectories of a system at discrete time steps, a guess of the memory depth of the non-Markovian process, and the level of noise generated during experimental reconstruction of trajectories. As output, our framework returns the predictive model of non-Markovian quantum dynamics, the dimension of the effective environment [62,67,68], the eigenfrequencies of the joint system and environment quantum dynamics, and the denoised trajectories of the system as a valuable byproduct. Note that our framework has the distinct advantage to rely on the linear methods of machine learning, which are known to be scalable, data efficient and yield the

exact solution. We illustrate the performance of our scheme with several paradigmatic examples of realistic models such as a qubit coupled with a finite-dimensional environment, a spin-boson model, and the damped Jaynes-Cummings model.

II. TIME-DELAY EMBEDDING OF NON-MARKOVIAN QUANTUM DYNAMICS

A. Quantum trajectories' Markovian dynamics

In this section we discuss an alternative view on non-Markovian open quantum dynamics that is the basis of the proposed data processing scheme. The alternative lies in turning from the density matrix based description of non-Markovian quantum dynamics to quantum trajectories based description, where a quantum trajectory is seen as a concatenation of several subsequent in time density matrices. This description is less physics-motivated and more suitable for the data-driven analysis we propose. Consider a d -dimensional quantum system undergoing non-Markovian dynamics [69] due to the interaction with a d_E -dimensional environment. Throughout the paper, we assume that the Hamiltonian driving dynamics of the system and environment as a whole remains time-independent. In the development that follows, d and d_E are finite, though the environment may consist of a large number of subsystems (particles), so one can have d_E exponentially large in the number of particles. Assume one has access to the system states (density matrices) $\varrho(t) \in \mathbb{C}^{d \times d}$ at consecutive and equidistant time instants $t = 0, 1, 2, \dots$, separated by the constant experimental time resolution τ , and one is able to reconstruct $\varrho(t)$ using quantum state tomography. Throughout the paper we use τ as time units, i.e., the physical time is expressed as τt . Performing the tomographic experiment multiple times, one reconstructs K -element sequences of the system's states:

$$T_K(t) := (\varrho(t), \varrho(t + 1), \dots, \varrho(t + K)) \in \mathbb{C}^{K \times d \times d}, \quad (1)$$

which we refer to as system *trajectories*. Assuming that the observed non-Markovian dynamics has a finite memory depth, for a sufficiently large value of K , there exists a linear map M that connects two subsequent trajectories through a Markovian master equation:

$$T_K(t + 1) = M[T_K(t)]. \quad (2)$$

The rigorous criteria for K to be sufficient (existence of M) is given in Appendix A. One can think of K as the number of previous in-time density matrices of the system, which is enough to determine the next in-time state of the system. Note also that while one can express the physical memory depth as τK , which is a more fundamental quantity, K itself is more suitable for data-driven analysis. In Appendix A we also show that for a finite d_E one always has a finite sufficient K that does not exceed $d^2 d_E^2$. Note, however, that in the worst case, the minimal sufficient K scales with the number of environment's subsystems (particles) exponentially and Eq. (2) quickly becomes intractable for a numerical analysis. Our conjecture is that this worst case corresponds to a dynamical chaos regime, as we discussed in more detail in Appendix A, for which long-time dynamics prediction is impossible. Although, one can use the data-driven analysis (discussed in the next section)

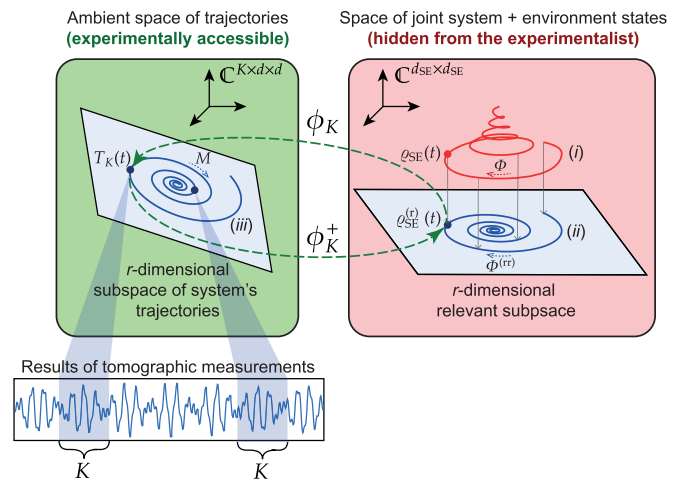


FIG. 2. Connection between the joint dynamics of the system and its environment with the observed dynamics of trajectories. Curve (i) sketches the joint dynamics of the system and its environment $\varrho_{SE}(t)$ driven by the quantum channel Φ . Curve (ii) sketches the joint dynamics of the system and the effective environment $\varrho_{SE}^{(r)}(t)$ driven by the map $\Phi^{(r)}$. Curve (iii) sketches the observed trajectories dynamics $T_K(t)$ driven by the map M . There is the one-to-one correspondence between curves (ii) and (iii) that is described by the linear map ϕ_K . Both dynamics (ii) and (iii) are Markovian, their Master equations are connected via ϕ_K and they can be seen as “shadows” of dynamics (i).

to identify chaos in the system's dynamics, our scheme is suitable only for finite-memory (nonchaotic) non-Markovian dynamics. In what follows, K is always assumed to be sufficient. We also refer to a sufficient K as to a *memory length*.

The general idea of considering the Markovian dynamics of $T_K(t)$ instead of the non-Markovian dynamics of $\varrho(t)$ is also known as a *time-delay embedding* and it is routinely employed across wide variety of contexts, including Koopman operators [70–73], closure modeling [74], time-series modeling [75,76], and reinforcement learning [77]. One can also think of the time-delay embedding as a particular way to construct a *Markovian embedding* [60,61,67,78–81] of non-Markovian quantum dynamics. The framework being developed here also seems very close in spirit to the TTM. However, there is a substantial difference. The TTM operates with a map from the space of trajectories to the space of density matrices, while here we operate with a map from the space of trajectories to the same space. This framework turns out to be more fruitful since it allows us to study the environment properties in addition to the system's dynamics prediction. In particular we show that the observed trajectories dynamics is the same, up to linear rescaling, as joint dynamics of the system and so called effective part of the environment. This finding is illustrated in Fig. 2. The given figure and the concepts involved are discussed in detail below. Since the main distinguishing feature of the presented method is its ability to explore the physics of the environment, the central question we address in this section is: what knowledge about the system and its environment causing non-Markovianity one can extract from the analysis of the Markovian dynamics of system's trajectories?

B. Unraveling the joint system and environment dynamics via the trajectories' dynamics

To start addressing this question, let us first note that states $\varrho(t)$ can be expressed via the corresponding joint system and environment states $\varrho_{SE}(t) \in \mathbb{C}^{d_{SE} \times d_{SE}}$ as follows:

$$\varrho(t) = \text{Tr}_E[\varrho_{SE}(t)], \quad (3)$$

where $d_{SE} := dd_E$ and Tr_E denotes a partial trace over the environment. Let us also introduce a quantum channel Φ describing the discrete-in-time evolution of the joint states:

$$\varrho_{SE}(t+1) = \Phi[\varrho_{SE}(t)]. \quad (4)$$

Since Tr_E is a linear map and the dynamics of $\varrho_{SE}(t)$ is linear as well, one can introduce a linear map ϕ_K returning $T_K(t)$ from $\varrho_{SE}(t)$:

$$T_K(t) = \phi_K[\varrho_{SE}(t)]. \quad (5)$$

The precise definition of ϕ_K is given in Appendix A; for the purposes of this section we only need the linearity property of this map.

The analysis of ϕ_K provides insights into the connection between the system trajectories' dynamics and the actual joint system and environment dynamics. It turns out that the linear map ϕ_K defines which ‘‘part’’ of the joint system and environment dynamics affects the observed trajectories. Let us call the *irrelevant (relevant) subspace* the kernel (the support) of ϕ_K . The corresponding projectors take the form

$$\pi_K^{(i)} = \text{Id} - \phi_K^+ \phi_K, \quad \pi_K^{(r)} = \phi_K^+ \phi_K, \quad (6)$$

where Id denotes the identity map, ϕ_K^+ is the Moore-Penrose inverse of ϕ_K , and superscripts (i) and (r) stand for irrelevant and relevant respectively. These projectors allow us to introduce the ‘irrelevant’ and ‘relevant’ parts of the joint state that read

$$\varrho_{SE}^{(i)}(t) = \pi_K^{(i)}[\varrho_{SE}(t)], \quad \varrho_{SE}^{(r)}(t) = \pi_K^{(r)}[\varrho_{SE}(t)]. \quad (7)$$

Since $\varrho_{SE}^{(i)}(t)$ belongs to the kernel of ϕ_K , only $\varrho_{SE}^{(r)}(t)$ defines $T_K(t)$. Therefore, $\varrho_{SE}^{(i)}(t)$ and $\varrho_{SE}^{(r)}(t)$ have a clear operational meaning: The dynamics of $\varrho_{SE}^{(i)}(t)$ is hidden, while $\varrho_{SE}^{(r)}(t)$ can be considered as a joint state of the system and the ‘‘effective’’ environment contributing to the observed non-Markovian dynamics of the system. Note, however, that $\varrho_{SE}^{(i)}(t)$ and $\varrho_{SE}^{(r)}(t)$ no longer have a unit trace. Moreover, there is no guarantee that they are positive semidefinite or even Hermitian. One can think of the projection on the relevant subspace as an analog of the partial trace over those degrees of freedom of the environment that do not contribute to the memory effects. We note that the similar projections on the relevant and irrelevant subspaces, yet with different meanings, are employed in the Nakajima-Zwanzig projection method [13].

An important characteristic of the linear map ϕ_K is the rank r of its matrix. It provides the dimension of the relevant subspace. In the next subsection, we show that the dynamics in the relevant subspace is independent of the dynamics in the irrelevant subspace. So, to fully describe the non-Markovian dynamics of the system, it is enough to consider only the relevant subspace that can be seen now as the space of the joint system and effective environment density matrices. By

the construction, d_E^{eff} provides the minimal dimensionality of the environment that can describe the non-Markovian behavior of the system. This observation allows us to introduce the dimension of the effective environment through the following relation $r = d^2[d_E^{\text{eff}}]^2$, which simply states that the joint system and effective environment density matrix has size $dd_E^{\text{eff}} \times dd_E^{\text{eff}}$. Note, however, that there is no guarantee that $d_E^{\text{eff}} = \sqrt{r/d^2}$ results in an integer number, since r can be an arbitrary integer. So, to make d_E^{eff} an integer, one can either round it up or down. Rounding down contradicts the fact that $\sqrt{r/d^2}$ is an underestimation of d_E^{eff} since some of the degrees of freedom of the environment may not affect the system dynamics. Therefore, one needs to round $\sqrt{r/d^2}$ up, or in other words one needs to complement the relevant subspace with additional degrees of freedom that, however, do not affect the dynamics of the system. Clearly, this can always be done and not in a unique way. We use this trick only here and for the sole purpose to justify the following final estimation of the effective environment dimension that reads:

$$d_E^{\text{eff}} := \lceil \sqrt{r/d^2} \rceil, \quad (8)$$

where $\lceil \cdot \rceil$ stands for the rounding up operation. Note that the value of r and d_E^{eff} are independent of K because K is sufficient (see Appendix A). The concept of the effective environment's dimension has been actively studied in recent literature [61,67,68,82,83] and has been shown to be an important characteristic of the environment. One can think of d_E^{eff} as a quantitative indicator of non-Markovianity, the dimensionality of the memory, or the complexity of non-Markovian dynamics. We remark that an exponentially large d_E^{eff} corresponds to utterly complex dynamics/dynamical chaos.

C. Equivalence between the trajectories' dynamics and the dynamics of the system and its effective environment

Next, consider the r -dimensional subspace $\mathcal{C}_K \equiv \text{Im}(\phi_K) \subseteq \mathbb{C}^{K \times d \times d}$, where Im stands for the image of ϕ_K . All experimentally accessible trajectories $T_K(t)$ lie in this subspace by definition; therefore, we refer to it as *trajectories subspace*. Note that ϕ_K is the linear bijection between the relevant subspace and the trajectories subspace. The Moore-Penrose inverse ϕ^+ inverts this bijection. The bijection between this subspaces is complemented by the following identity between Φ , M , and ϕ_K :

$$M\phi_K = \phi_K\Phi. \quad (9)$$

One can think about Eq. (9) as about the following commutative diagram:

$$\begin{array}{ccc} \varrho_{SE}(t) & \xrightarrow{\Phi} & \varrho_{SE}(t+1) \\ \downarrow \phi_K & & \downarrow \phi_K \\ T_K(t) & \xrightarrow{M} & T_K(t+1) \end{array} .$$

It says that one can first make a time step forward in the space of the joint system and environment states and then go to the corresponding trajectory via ϕ_K or first go to a trajectory via ϕ_K from a given joint state and then make a time step forward by applying M in the trajectories subspace. Note that the identity Eq. (9) follows from the fact that M describes

the temporal evolution of any trajectory. The existence of the bijection between the relevant subspace and \mathcal{C}_K and Eq. (9) mean that the trajectories' dynamics and the joint system and effective environment dynamics are identical in a sense that is explained below. Let us decompose Φ into four terms

$$\Phi = \Phi^{(ii)} + \Phi^{(ir)} + \Phi^{(ri)} + \Phi^{(rr)}, \quad (10)$$

where $\Phi^{(xy)} := \pi_K^{(x)} \Phi \pi_K^{(y)}$, $x, y \in \{i, r\}$. One can see that each term has a clear operational meaning: $\Phi^{(rr)}$ and $\Phi^{(ii)}$ govern the dynamics in the relevant and irrelevant subspace respectively, while $\Phi^{(ir)}$ and $\Phi^{(ri)}$ correspond to ‘‘cross-flows’’ between the subspaces. Note, however, that $\Phi^{(xy)}$ are neither trace-preserving nor completely positive maps in general. Let us substitute the decomposition Eq. (10) in Eq. (9):

$$M\phi_K = \phi_K \Phi^{(ii)} + \phi_K \Phi^{(ir)} + \phi_K \Phi^{(ri)} + \phi_K \Phi^{(rr)}. \quad (11)$$

One can compare supports of all terms of the identity above and identifies two linearly independent identities that read

$$M\phi_K = \phi_K \Phi^{(ir)} + \phi_K \Phi^{(rr)}, \quad 0 = \phi_K \Phi^{(ii)} + \phi_K \Phi^{(ri)}. \quad (12)$$

By noticing that $\phi_K \Phi^{(ii)} = 0$ and $\phi_K \Phi^{(ri)} = 0$ due to the fact that ϕ_K maps any vector from the irrelevant subspace to zero one finally gets

$$M\phi_K = \phi_K \Phi^{(rr)}, \quad 0 = \phi_K \Phi^{(ri)}. \quad (13)$$

The identity $0 = \phi_K \Phi^{(ri)}$ implies the absence of ‘‘flow’’ from the irrelevant subspace to the relevant one. In other words $\Phi^{(rr)}$ fully describes the dynamics in the relevant subspace:

$$\varrho_{SE}^{(r)}(t+1) = \Phi^{(rr)}[\varrho_{SE}^{(r)}(t)], \quad (14)$$

since there is no effect of the irrelevant subspace on the dynamics in the relevant subspace. Moreover, one can see that Eq. (14) is a Markovian master equation. Let us multiply the identity $M\phi_K = \phi_K \Phi^{(rr)}$ by ϕ_K^+ on the right side

$$M\phi_K \phi_K^+ = \phi_K \Phi^{(rr)} \phi_K^+. \quad (15)$$

Bearing in mind that the action of M on the orthogonal complement of \mathcal{C}_K does not have a physical meaning and can be set arbitrarily, we set it equal to zero. Hence, $M = M\phi_K \phi_K^+$, and we get

$$M = \phi_K \Phi^{(rr)} \phi_K^+. \quad (16)$$

Now we can claim that the dynamics of trajectories and the joint dynamics of the system and its effective environment are identical up to a linear rescaling. Indeed, both dynamics are Markovian, there is the linear bijection between states, i.e., $T(t) = \phi_K[\varrho_{SE}^{(r)}(t)]$ and the corresponding linear identity between master equations $M = \phi_K \Phi^{(rr)} \phi_K^+$. This also implies that M and $\Phi^{(rr)}$ have completely the same eigenvalues and their left and right eigenvectors are connected via ϕ_K . Moreover, since Φ has block triangular form w.r.t. relevant and irrelevant subspaces ($\Phi^{(ri)} = 0$), the set of eigenvalues of Φ includes all the eigenvalues of $\Phi^{(rr)}$, and so the eigenvalues of M . The overall concept described above is illustrated in Fig. 2.

The fact that the dynamics of the system and environment factorizes into two parts: the ‘‘relevant’’ part, which is Markovian, and the independent ‘‘irrelevant’’ part, may seem artificial at first glance. Indeed, this property should imply some symmetry that is not inherent to an arbitrary system.

In fact, the exact factorization in general, except for some trivial cases of two noninteracting subsystems or highly symmetric cases, results trivially in the ‘‘relevant’’ part that covers the entire state space and a one-dimensional irrelevant part. However, an approximate factorization (with some acceptable error) may be far more productive and lead to a significant dimensionality reduction of the ‘‘relevant’’ part compared to the entire state space. The data-driven approach we introduce later builds an approximate factorization by setting a threshold defined by the noise amplitude.

The main results of this section may be summarized as follows. If a non-Markovian dynamics has a finite memory length K , then: (I) The dynamics of trajectories is driven by the Markovian master Eq. (2); (II) All the experimentally accessible trajectories lie in a subspace of dimension $r = \text{rank}(M)$ (trajectories subspace) and form an r -dimensional Markovian embedding; (III) There exists an r -dimensional relevant subspace of the joint system and environment density matrices' space. The dynamics in the relevant subspace is Markovian and identical to the dynamics of trajectories up to the bijective linear transformation ϕ_K . The dynamics in the orthogonal complement of the relevant subspace does not affect dynamics of trajectories at all; (IV) One can think of the dynamics in the relevant subspace as the dynamics of the joint system and effective environment. One can estimate the dimension of the effective environment d_E^{eff} using Eq. (8). Both r and d_E^{eff} are quantitative indicators of the non-Markovian dynamics complexity; (V) As a consequence of (III), all the eigenvalues of the dynamical maps M and $\Phi^{(rr)}$ are the same, and are part of the set of eigenvalues of Φ .

These claims allow one to gain valuable knowledge about the system and its environment from the master equation describing the system's trajectories and the dimension of the subspace where these trajectories are embedded. The only missing ingredient at this stage, is a scheme to extract these objects from noisy measured trajectories.

III. ALGORITHMS FOR PROCESSING OF QUANTUM TRAJECTORIES

A. Effective environment dimension identification and data denoising

As we discussed in the previous section, the only data assumed to be experimentally accessible are the system's discrete-in-time quantum trajectories. Those quantum trajectories can be reconstructed via quantum state tomography [84,85], i.e., for each trajectory, one runs multiple experiments with completely identical initial conditions and collects measurement outcomes statistics to reconstruct underlying density matrices in different moments of time. Since in real experiments each measurement takes some finite time τ_m , one has a natural restriction $\tau \geq \tau_m$. Note also that the time resolution τ should be chosen small enough to capture all necessary short-time dynamical effects but not too small not to make an experiment lasts too long collecting too many data-points. Note, that subsequent experiment runs must be sufficiently separated in time, to eliminate memory effects from the previous run. Let us assume that for L different initial states of the system the measured data set

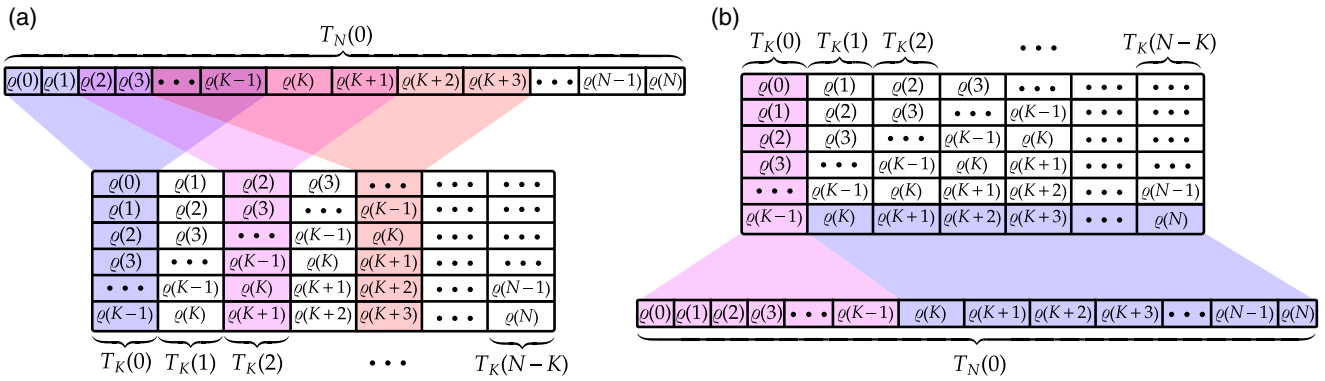


FIG. 3. (a) Graphical interpretation of slicing a trajectory into smaller chunks. (b) Reconstruction of the whole trajectory of length N back from chunks.

consists of quantum trajectories of length $N > K$, where N is the total number of discrete time steps per each trajectory and K is assumed to be sufficient for the Markovian dynamics of trajectories. This set reads

$$\mathcal{S} = \{T_N^{(i)}(0)\}_{i=1}^L, \tag{17}$$

where i enumerates initial states. To prepare data for the analysis we construct their time-delay embedding. We slice each trajectory $T_N^{(i)}(0)$ into $N - K + 1$ smaller K -steps chunks $\{T_K^{(i)}(t)\}_{t=0}^{N-K}$ as it is shown in Fig. 3(a). The resulting set of chunks $\{\{T_K^{(i)}(t)\}_{t=0}^{N-K}\}_{i=1}^L$ are processed further.

First, one can estimate the subspace \mathcal{C}_K as the linear span of all the trajectories chunks, i.e.,

$$\mathcal{C}_K \approx \text{span}\left(\left\{\{T_K^{(i)}(t)\}_{t=0}^{N-K}\right\}_{i=1}^L\right). \tag{18}$$

The question arises: how many trajectories L have to be measured to provide a reliable estimate of \mathcal{C}_K ? For noiseless trajectories it turns out that in most cases one needs only one long enough trajectory. This can be explained as follows. The estimate Eq. (18) for only one trajectory reads

$$\mathcal{C}_K \approx \text{span}\left(\{T_K^{(1)}(t)\}_{t=0}^{N-K}\right) = \text{span}\left(\{M^t T_K^{(1)}(0)\}_{t=0}^{N-K}\right). \tag{19}$$

Note that unless the initial trajectory chunk $T_K^{(1)}(0)$ not in a low-dimensional invariant subspace of M , for $N - K = r$ one has $\text{span}(\{M^t T_K^{(1)}(0)\}_{t=0}^{N-K})$ covering the entire trajectories subspace. Therefore, only one trajectory is enough. This is also supported by the numerical experiments that are demonstrated further. For example, in Fig. 10 one can see that the average error of prediction for quite a few numerical experiments with $L = 1$ and different other parameters, including noise amplitude, is about 0.06, which is not that much. However, for highly symmetric systems it might be the case that $\text{span}(\{M^t T_K^{(1)}(0)\}_{t=0}^{N-K})$ covers only part of the trajectories subspace. In this case, one needs to sample trajectories with

different initial states. It can be done in different ways. For example, before measuring a trajectory one can perturb the system in various ways, e.g., measure it on an arbitrary basis or apply some control signal to it. In our numerical experiments, we prepare a unique initial state as follows. We set the initial joint system and environment state in the stationary thermal state and then replace the system with a new one in some state of our choice. This helps us to explore the trajectories subspace better. To determine the value of L in practice, one can sample a new trajectory with a unique initial state until the estimate Eq. (18) saturates.

All the trajectories bear some noise induced by finiteness of the measurements statistics, imperfections of measurement devices, etc. This results in an additive noise

$$T_K^{(i)}(t) \longrightarrow T_K^{(i)}(t) + \delta T_K^{(i)}(t), \tag{20}$$

where $\delta T_K^{(i)}(t)$ is a noise term. Note that this is not a noise induced by the interaction of the system with the environment that has an entirely different nature. The linear span of noisy trajectories always tends to the ambient space $\mathbb{C}^{d \times d \times K}$ with the growing number of trajectories in a data set and does not correspond to the genuine subspace $\mathcal{C}_K \subseteq \mathbb{C}^{K \times d \times d}$ of trajectories. Thus, one requires a method that identifies the genuine trajectories subspace in the presence of background noise. In other words, one needs to find a low-dimensional hyperplane of an appropriate dimension such that the data points (trajectories) lie as close to this hyperplane as possible. The dimension of the hyperplane is chosen to separate signal from noise. This is a typical machine-learning task usually performed by using a principal component analysis (PCA) with automatic rank determination. We found the automatic rank determination procedure described in Ref. [63] the most suitable for our case. The overall algorithm is summarized here:

Step 1. One concatenates all the trajectories chunks into a matrix

$$H = \underbrace{\begin{bmatrix} T_K^{(1)}(0) & \dots & T_K^{(L)}(0) & T_K^{(1)}(1) & \dots & T_K^{(L)}(1) & \dots & T_K^{(1)}(N-K) & \dots & T_K^{(L)}(N-K) \end{bmatrix}}_{\text{all accessible trajectories}} \tag{21}$$

of size $Kd^2 \times L(N - K + 1)$, which has the form of a Hankel matrix. Here, trajectories chunks can be thought of as vectors of length Kd^2 . In the noiseless case, it is enough to take the linear span of the Hankel matrix columns to determine the subspace \mathcal{C}_K , but we need to process them further to distinguish the signal from background noise.

Step 2. One performs a singular value decomposition (SVD) of the Hankel matrix $H = USV^\dagger$, where U and V are isometric matrices and S is the diagonal matrix with singular values on the diagonal arranged in nonincreasing order.

Step 3. Following the results of Ref. [63] one separates “noisy” singular values from “signal” singular values by comparing them with a threshold

$$s = \sigma \sqrt{2L(N - K + 1)} f \left[\frac{Kd^2}{L(N - K + 1)} \right], \quad (22)$$

where f is defined as follows $f(\beta) = \sqrt{2(\beta + 1) + \frac{8\beta}{(\beta+1) + \sqrt{\beta^2 + 14\beta + 1}}}$, σ is the standard deviation of noise. If a singular value is greater (less) than the threshold, then one considers it as a signal (noise) singular value. Since the quantum state tomography requires an enormous number of measurement outcomes to build a precise estimations of system’s states, the size of measurements statistic is the main bottleneck of a tomographic experiment. Therefore, the observed noise is mainly the result of a statistical error caused by the finiteness of the quantum measurements statistics N_{samples} per time step. This allows one to estimate the standard deviation of noise σ as $N_{\text{samples}}^{-1/2}$. More precisely, the standard deviation σ of the resulting noise reads $\sigma \approx N_{\text{samples}}^{-1/2}$ for the standard tomographic protocols or $\sigma \approx N_{\text{samples}}^{-1}$ for adaptive tomographic protocols with a protocol-specific prefactor [86]. If trajectories result from a Monte Carlo simulation [13], then one can estimate σ from the number of samples as $\sigma \approx N_{\text{samples}}^{-1/2}$. The threshold s differs from the threshold in Ref. [63] by a factor $\sqrt{2}$. This is motivated by the fact that the amplitude of the absolute value of a complex valued noise is $\sqrt{2}$ times larger. The intuition behind such a separation of singular values is simple: one uses the random matrix theory to determine the maximal singular value s_{max} of a random matrix with the standard deviation of a matrix elements σ . Then, $s \approx s_{\text{max}}$ can be seen as a threshold that separates “noisy” singular values from “signal” singular values;

Step 4. One truncates the SVD of the Hankel matrix as follows:

$$\hat{U} = U[:, : \eta], \quad \hat{V} = V[:, : \eta], \quad \hat{S} = S[:, : \eta], \quad (23)$$

where we use MATLAB/NumPy notations to represent slices of matrices (see Appendix C), η is the number of singular values that are greater than s . The corresponding “truncated” Hankel matrix reads $\hat{H} = \hat{U} \hat{S} \hat{V}^\dagger$.

This algorithm provides a data-driven estimation of the genuine trajectories subspace \mathcal{C}_K , its dimension, and the dimension of the effective environment:

$$\mathcal{C}_K \approx \text{span}(\hat{U}), \quad r \approx \eta, \quad d_E^{\text{eff}} \approx \lceil \sqrt{\eta/d^2} \rceil, \quad (24)$$

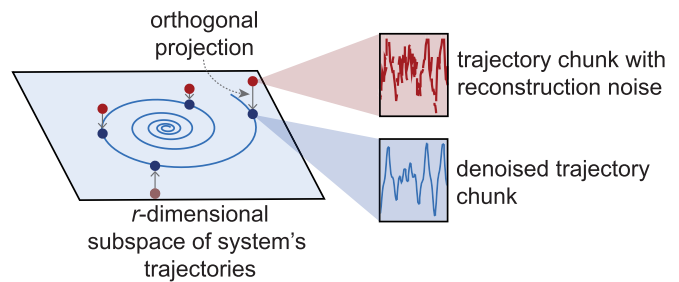


FIG. 4. Illustrative interpretation of the PCA with automatic rank selection algorithm applied to quantum trajectories. The algorithm estimates the low-dimensional subspace of system’s trajectories from measured trajectories and then projects trajectories chunks on this subspace. The orthogonal projection on this subspace eliminates the orthogonal to the subspace component of noise that leads to the noise reduction effect.

where $\text{span}(\cdot)$ denotes the linear span of matrix columns and η serves as a data-driven estimation of the genuine trajectories subspace dimension r . But these are not the only useful outcomes of the algorithm: It also reduces noise in the measured trajectories. The idea is that since noise does not preserve the subspace \mathcal{C}_K , it takes trajectories out of this subspace (see Fig. 4). Meanwhile, the orthogonal projection of trajectories back onto the subspace \mathcal{C}_K reduces the effect of noise because the component of noise orthogonal to the subspace \mathcal{C}_K is thus eliminated. Note, that one can reconstruct the whole trajectory of length N from its chunks as shown in Fig. 3(b). To establish a connection between the proposed denoising method with previously developed techniques for denoising and quantum tomographic reconstruction from limited data, we note that usually their main idea is to employ an assumption about underlying states model imposing certain restrictions on admissible states. In other words, these methods use prior information about an unknown state in addition to the observed measurements data, which mitigates requirements to the amount of data. For example, one can use the low-rank assumption to reconstruct a fairly pure state from limited data [87], or assume limited entanglement to make the reconstruction of a many-body state from measurement outcomes possible [88]. However, our technique utilizes another type of prior information. It uses prior information about the temporal structure of a system’s dynamics, not about its state. It fixes some level of complexity of the dynamics that reduces the required amount of data and leads to noise reduction. Our denoising technique is compatible with conventional methods of noise reduction, and in combination, they can lead to further improvement of data efficiency. We verify the performance of denoising in the numerical experiments discussed further below.

B. Identification of the Markovian master equation governing trajectories dynamics

Denoised trajectories $\hat{T}_K^{(i)}(t)$ which are columns of \hat{H} can be also used to estimate \hat{M} , which could be thought of as a predictive model from the machine-learning point of view. The estimation can be written as the solution of the following

optimization problem:

$$\begin{aligned} & \underset{M}{\text{minimize}} \quad \sum_{i=1}^L \sum_{t=0}^{N-K} \|\hat{T}_K^{(i)}(t+1) - M[\hat{T}_K^{(i)}(t)]\|_2^2, \\ & \text{subject to} \quad \text{rank}(M) = \eta, \end{aligned} \tag{25}$$

where $\|\cdot\|_2$ stands for the 2-norm. The first line corresponds to the minimization of the difference between left and right hand sides of the equation $\hat{T}_K(t+1) = M[\hat{T}_K(t)]$ driving the dynamics of trajectories. The second line restricts consideration only to matrices of rank η . As the dynamics of trajectories is embedded in the subspace of dimension η , there is no need to consider matrices of other ranks. Note that η is not only an estimate of r , but also the optimal choice of one of the hyperparameters of the predictive model. Another hyperparameter of the predictive model is the memory depth K . Search of η corresponds to the selection of the predictive model. The larger the value of the hyperparameter is, the more expressible the corresponding model is (i.e., more complex data can be described), but the more data is needed to reconstruct M (lower data efficiency). This is the bias-variance trade-off typical for the model selection tasks in machine learning [89]. Usually one selects the best model by minimization of the error on a validation set, which is not used for model training. This model selection technique is called cross-validation [90], and requires splitting data on training and validation sets and exhaustive validation set's error minimization via brute force adjustment of hyperparameters. However, for some models, e.g., relevance vector machine [66,91], one selects the best

model automatically without a validation set by using the Bayesian interpretation of the model selection task. The idea of such an automatic model selection follows the Occam's razor principle: one should select the simplest model among all possible that still yet manages to describe the observed data [66]. The correct model selection guarantees the best data efficiency and robustness to noise and corresponds to the best machine-learning practices. One sees that a PCA algorithm with automatic determination of the subspace dimension solves such a model selection task. It chooses the model with the smallest value of the hyperparameter (the simplest model) that still describes the measured data. Therefore, our approach is equipped with the automatic model selection, which improves the data efficiency compared to previously proposed methods such as TTM [45] as discussed in Sec. V. The effect of noise reduction also can be seen as a consequence of the correct model selection. It also makes the model less sensitive to selection of another hyperparameter K . One can safely overestimate K by setting it knowingly large, the automatic model selection prevents the effect of overfitting. This makes the selection procedure for K very simple: one can set K to be comparable with N , e.g., $K = N/2$. The only important restriction is not to make the number of trajectories $(N - K + 1)L$ in a data set too small. The proper selection of η prevents overfitting. We show this numerically in Sec. IV A.

The optimization problem Eq. (25) is exactly solvable. Indeed, this is an example of linear regression among the most robust methods of machine learning. To write down the solution of Eq. (25), let us introduce two matrices

$$\begin{aligned} \hat{X} &= \hat{H}[:, : -L] = \begin{bmatrix} \hat{T}_K^{(1)}(0) & \dots & \hat{T}_K^{(L)}(0) & \hat{T}_K^{(1)}(1) & \dots & \hat{T}_K^{(L)}(1) & \dots & \hat{T}_K^{(1)}(N-K-1) & \dots & \hat{T}_K^{(L)}(N-K-1) \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \end{bmatrix}, \\ \hat{Y} &= \hat{H}[:, L :] = \begin{bmatrix} \hat{T}_K^{(1)}(1) & \dots & \hat{T}_K^{(L)}(1) & \hat{T}_K^{(1)}(2) & \dots & \hat{T}_K^{(L)}(2) & \dots & \hat{T}_K^{(1)}(N-K) & \dots & \hat{T}_K^{(L)}(N-K) \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \end{bmatrix}. \end{aligned} \tag{26}$$

The matrix \hat{X} is the matrix \hat{H} without L last columns and the matrix \hat{Y} is the matrix \hat{H} without L first columns. In other words, columns of \hat{Y} are trajectories subsequent in time to the corresponding columns of \hat{X} . Using this fact, one can rewrite the optimization problem Eq. (25) in the following form:

$$\begin{aligned} & \underset{M}{\text{minimize}} \quad \|\hat{Y} - M\hat{X}\|_F^2, \\ & \text{subject to} \quad \text{rank}(M) = \eta, \end{aligned} \tag{27}$$

where $\|\cdot\|_F$ stands for the Frobenius norm, and M is seen as a matrix rather than as an abstract linear map. The solution of this optimization problem thus reads

$$M = \hat{Y}\hat{X}^+, \tag{28}$$

where \hat{X}^+ is the Moore-Penrose inverse of X . Indeed, M has the rank η by construction and it is the global minimum of the objective function. Note that it is more efficient both for numerical complexity and memory to find the solution Eq. (28) by using the *dynamic mode decomposition*

(DMD) method [64,65], rather than straightforwardly calculate YX^+ . In our numerical experiments we use the DMD to find M .

Finally, we summarize the data processing scheme in Fig. 5, that basically includes two steps, (i) environment dimension identification described in the previous subsection and (ii) predictive model reconstruction described in the given subsection. The proposed data processing scheme allows one to obtain (i) information on the quantum environment such as the dimension of the effective environment, eigenvalues of the quantum channel driving the discrete-in-time joint dynamics of the system and its environment, (ii) the master equation that predicts the system's dynamics, and (iii) denoised data set of trajectories. As inherent features of the scheme we note that all the used algorithms (i.e., PCA with automatic rank selection and the linear-regression-based reconstruction of M) converge with mathematical guarantees to the optimal solution and the scheme is equipped with the automatic model selection that improves data efficiency and robustness to noise.

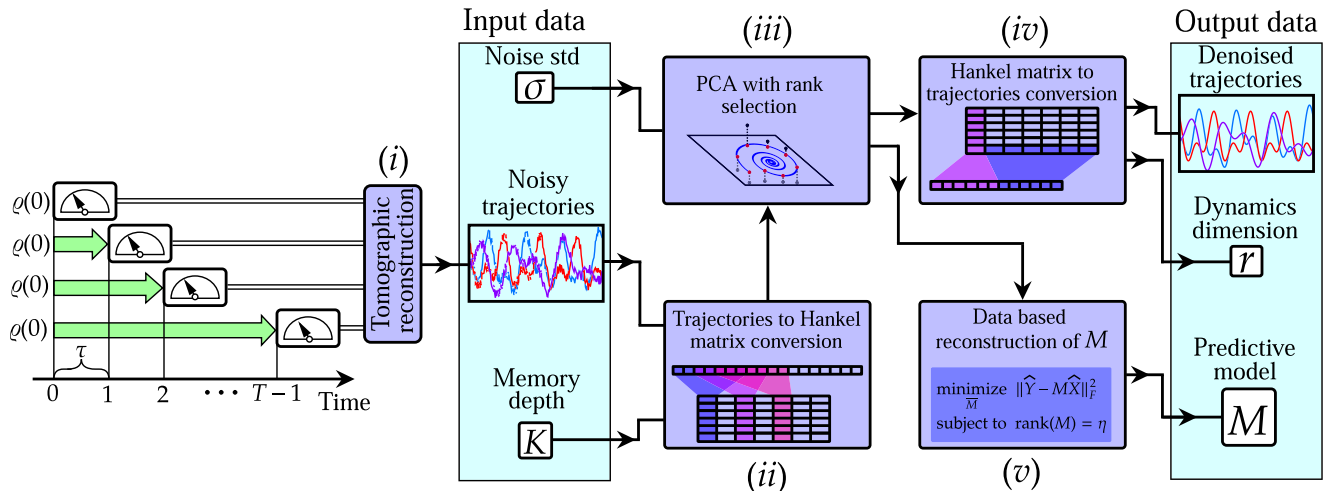


FIG. 5. Summary of the proposed data processing scheme. (i) One starts by collecting a data set of noisy quantum trajectories. Usually, it is obtained via tomographic reconstruction or numerical simulation. As input information the scheme also requires an estimation of the standard deviation of noise and an estimation of the sufficient K . The standard deviation can be estimated from the size of the quantum measurements statistics used per time step. An estimation of the sufficient K does not require precise tuning. It can be set knowingly large. The automatic model selection prevents overfitting for an overestimated K . One of the possible practical recipes of choosing K consists in setting the maximal possible $K = N - 1$ and iterative retraining of the model while decrementing K , until the value of η saturates. (ii) In the next step, one slices trajectories into chunks of size K as shown in Fig. 3(a). (iii) Then, one identifies the subspace of trajectories via a PCA with the automatic dimension selection. It not only reconstructs the space \mathcal{C}_K , but also allows to estimate the effective dimension of the environment d_E^{eff} , and performs the automatic model selection and reduces noise in data. (iv) Denoised chunks of trajectories are combined into a denoised version of initial whole trajectories of length N as shown in Fig. 3(b). (v) Finally, one reconstructs M from the denoised trajectories chunks by solving the optimization problem Eq. (25). The reconstructed M allows predicting the dynamics of a non-Markovian quantum system and its eigenvalues are also eigenvalues of Φ which drives the joint system and environment dynamics.

IV. NUMERICAL RESULTS

A. Probing non-Markovian quantum dynamics with a predefined effective environment

As the first example of non-Markovian quantum dynamics that we analyze with the proposed method, we consider a model with a predefined effective environment, i.e., with a predefined $\Phi^{(\text{tr})}$ driving the dynamics of the system and the effective environment as a whole. We choose $\Phi^{(\text{tr})}$ in such a way, that its dimension can not be reduced further. This model is chosen to validate the claimed ability of the proposed method to reconstruct properties of the effective environment, i.e., the effective dimension of the environment and eigenvalues of the joint system and environment dynamics. It can be easily done since all the properties of $\Phi^{(\text{tr})}$ are predefined and known. In this subsection we also (i) validate the ability of our method to predict a non-Markovian quantum dynamics via the reconstructed M ; (ii) validate the ability of our method to reduce noise in data by checking how the projection of the observed trajectories on the principle subspace reduces the amplitude of noise; (iii) study the sensitivity of the proposed method to a choice of the hyperparameter K ; and finally (iv) we study how the performance of the method depends on the size of data set, i.e., number of trajectories L in a data set.

The discrete-in-time dynamics of the system reads

$$\begin{aligned} \Phi^{(\text{tr})} &= \exp(\tau \mathcal{L}), \quad \varrho_{\text{SE}}^{(r)}(t+1) = \Phi^{(\text{tr})}[\varrho_{\text{SE}}^{(r)}(t)], \\ \varrho(t+1) &= \text{Tr}_E[\varrho_{\text{SE}}^{(r)}(t+1)], \end{aligned} \quad (29)$$

where $\varrho_{\text{SE}}^{(r)} \in \mathbb{C}^{d d_E^{\text{eff}} \times d d_E^{\text{eff}}}$ is the joint system and effective environment density matrix that also plays the role of the state in the relevant subspace, d_E^{eff} is the dimension of the effective environment, ϱ is the state of the system, \mathcal{L} is the Gorini-Kossakowski-Sudarshan-Lindblad (GKSL) generator [13,15,16,92,93] of the Markovian quantum dynamics of the system and the effective environment, τ is the constant discrete time step. One generates the GKSL generator \mathcal{L} randomly as described in Appendix B. The randomness of \mathcal{L} prevents further dimensionality reduction of the effective environment.

We use the model above to generate a number of data sets. We fix $d = 2$ (the system is a qubit) and for d_E^{eff} ranging from 2 to 6 generate a set of random GKSL generators. For each GKSL generator and for total simulation times $N = 150$ and 200 we simulate $L = 4$ quantum trajectories $\{T_N^{(i)}(0)\}_{i=1}^L$ with time step $\tau = 0.2$ using Eq. (29). As an initial joint state for each trajectory we take $\varrho_{\text{SE}}(0) = |\psi\rangle\langle\psi| \otimes \varrho_E(0)$, where $|\psi\rangle$ is sampled uniformly from the Bloch sphere, $\varrho_E(0) = \text{Tr}_S \varrho_{\text{SE}}^{\text{st}}$, and $\varrho_{\text{SE}}^{\text{st}}$ is the stationary state of \mathcal{L} , i.e., $\mathcal{L}[\varrho_{\text{SE}}^{\text{st}}] = 0$. It means that the system and the environment are in a joint thermalized state before the observation process, and at the moment the process starts, we replace the system by a new random pure state. To address dynamics prediction accuracy, for each value of d_E^{eff} , we generate an additional “test” trajectory $T_N^{\text{test}}(0)$, which is not employed in the training procedure, i.e., its initial state is also randomly generated but different with initial states of $\{T_N^{(i)}(0)\}_{i=1}^L$. To simulate noise appearing during data acquisition of trajectories, we

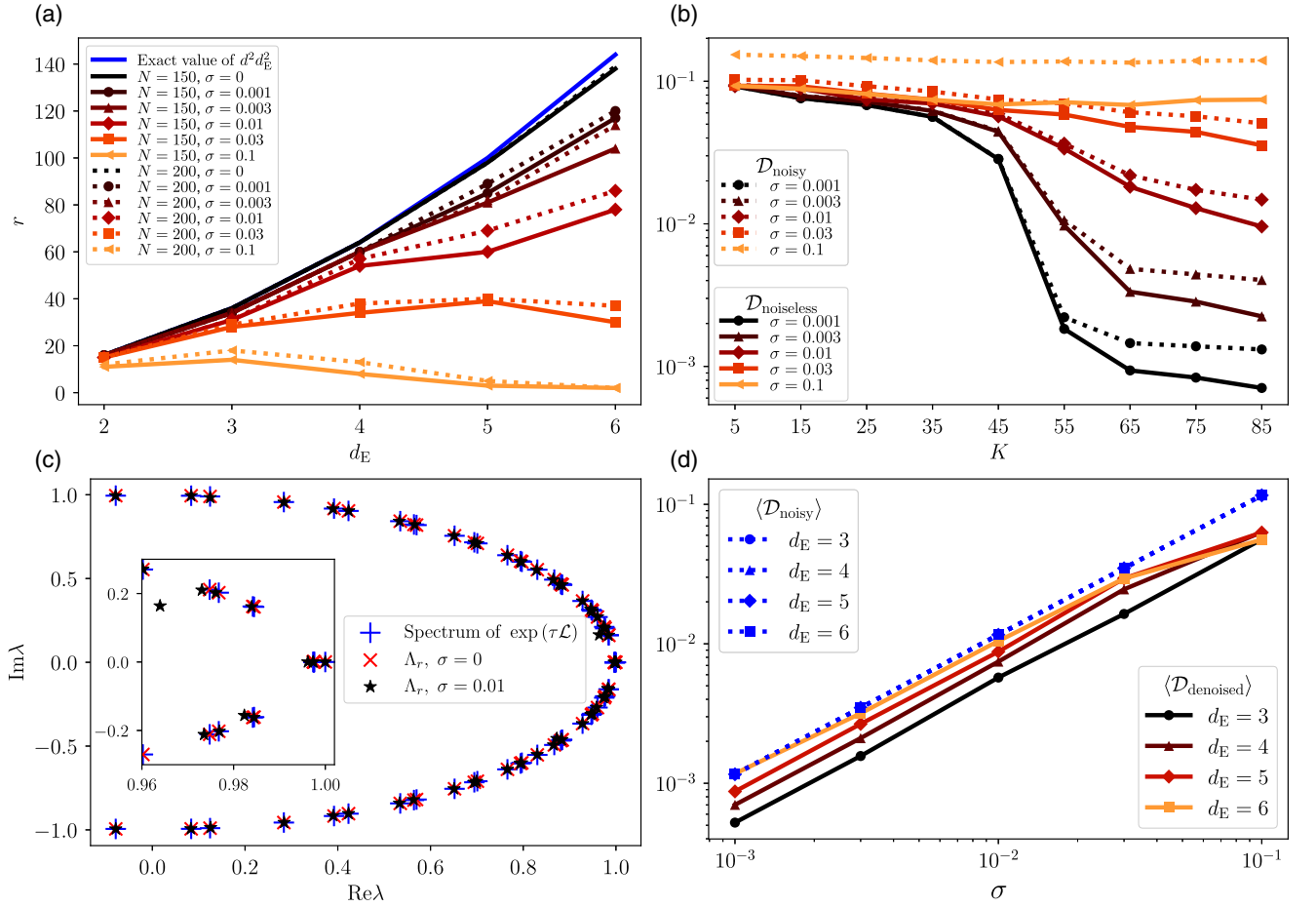


FIG. 6. Results obtained by the application of the proposed method to the non-Markovian dynamics with a predefined effective environment [Eq. (29)]. (a) A comparison of the exact dimension of the trajectories subspace $d^2[d_E^{\text{eff}}]^2$ with the data-driven reconstruction r of it for different data set trajectories lengths N and noise levels σ . (b) The distance between the prediction and the noiseless test trajectory $\mathcal{D}_{\text{noiseless}}$ and the distance between the prediction and the noisy test trajectory $\mathcal{D}_{\text{noisy}}$ as functions of memory depth K for different levels of noise σ . (c) Comparison of the spectrum of the quantum channel $\exp(\tau\mathcal{L})$ driving the dynamics of the joint system and effective environment density matrix with the eigenvalues of M reconstructed from data for $d_E = 4$. (d) Comparison of the distance between noisy and noiseless data sets $\langle \mathcal{D}_{\text{noisy}} \rangle$ and the distance between denoised and noiseless data sets $\langle \mathcal{D}_{\text{denoised}} \rangle$ for several different σ and d_E .

add i.i.d. Gaussian noise with zero mean and variance σ^2 to the real and imaginary parts of each entry of all trajectories (both employed in the training and the test trajectory). In what follows, we consider σ to be known in advance, because in a typical tomographic experiment or in the Monte Carlo simulation one can estimate σ from the statistics size. We also note that the considered values of σ are consistent with typical numbers of samples processed in tomographic experiments $N_{\text{samples}} \sim 10^3 \dots 10^6$ (see, e.g., [86]). To demonstrate the denoising ability of our approach we keep record of both noisy and noiseless versions of each trajectory, i.e., each trajectory of all data sets $\{T_N^{(i)}(0)\}_{i=1}^L$ and all test trajectories $T_N^{\text{test}}(0)$.

We start discussion of the results of data processing from the reconstruction of the trajectories subspace dimension r and the effective dimension of the environment d_E^{eff} . The comparison of the exact r and the reconstructed from data one for the fixed memory depth hyperparameter $K = 75$, several noise amplitudes σ , and two values of a data set trajectories length N , is presented in Fig. 6(a). One can see that the reconstructed r underestimates the exact value of r that is

equal to $d^2 d_E^{\text{eff}}$ though approaches it with decreasing the noise level and increasing the length of trajectories forming a data set. In the case of $\sigma = 0$, reconstructed r precisely fits the exact one. The tiny difference appearing for large r is due to negligible memory effects that do not fit into given memory length K . For finite σ it is impossible to recognize some weak memory effects in front of noise, this leads to underestimated r value. The situation becomes critical when $\sigma = 0.1$, as the algorithm is almost unable to recognize the signal against the background noise. We also compare the exact effective dimensions of the environment d_E^{eff} with its data-driven reconstruction $\lceil \sqrt{r/d^2} \rceil$, where r is reconstructed from the data. The comparison is given in Table I. We see that for small noise amplitudes and sufficiently long trajectories in a data set, the value of d_E^{eff} is reconstructed exactly.

Next, we validate the ability of the proposed method to predict the dynamics of the system. The prediction is built as follows. One reconstructs M that describes Markovian dynamics of trajectories from a noisy data set. Then one uses M to predict dynamics of a test trajectory. One takes a noisy chunk

TABLE I. Comparison of the exact effective dimensions of the environment d_E^{eff} with reconstructed from data ones for different amplitudes of noise σ and different trajectories lengths N . Cases where the data-driven estimation is exact are shown in bold font.

	$\sigma = 10^{-1}$		$\sigma = 10^{-2}$		$\sigma = 10^{-3}$	
	$N = 150$	$N = 200$	$N = 150$	$N = 200$	$N = 150$	$N = 200$
$d_E^{\text{eff}} = 2$	2	2	2	2	2	2
$d_E^{\text{eff}} = 3$	2	2	3	3	3	3
$d_E^{\text{eff}} = 4$	2	2	4	4	4	4
$d_E^{\text{eff}} = 5$	2	2	4	5	5	5
$d_E^{\text{eff}} = 6$	1	1	5	5	5	6

$T_K(0)$ of the test trajectory, also seen as an initial state of Eq. (2) or as a “prehistory” of non-Markovian dynamics, and propagates it forward in time by using the reconstructed from data M and Eq. (2). In Fig. 7(a) we demonstrate the agreement between the predicted behavior of the test trajectory and the test trajectory itself for the following parameters of the data set used for the reconstruction of M : $N = 200$, $d_E = 3$ and several different values of σ . The value of K is set to 75.

One sees that the proposed method predicts non-Markovian dynamics of the system correctly. Even if a data set and a test trajectory are effected by noise with reasonably high standard deviation $\sigma = 0.1$, the approach is capable to predict the system’s dynamics at a small timescale.

Then we compare eigenvalues of M with eigenvalues of the quantum channel $\Phi^{(\text{tr})} = \exp(\tau\mathcal{L})$ driving the joint system and effective environment dynamics. The results for M reconstructed from a data set with $d_E = 4$, $K = 75$, $N = 200$ in the case of noise absence ($\sigma = 0$) and in the case with noise ($\sigma = 0.01$) are presented in Fig. 6(c). One can see that there is a perfect coincidence between exact and reconstructed eigenvalues in the case without noise, which agrees well with the theory (see Sec. II), while in the case with noise the obtained eigenvalues are slightly shifted, and a tiny part of them only is lost. The latter fact can be explained by the indistinguishability of some eigenmodes dynamics from the noise. However, even in the presence of noise, one sees that the method provides valuable information such as the eigenvalues of the joint dynamics.

We also study how the selection of the memory depth K that is a hyperparameter of the model and number of trajectories L (data set size) impact the accuracy of the prediction. For this purpose we introduce the distance between two quantum

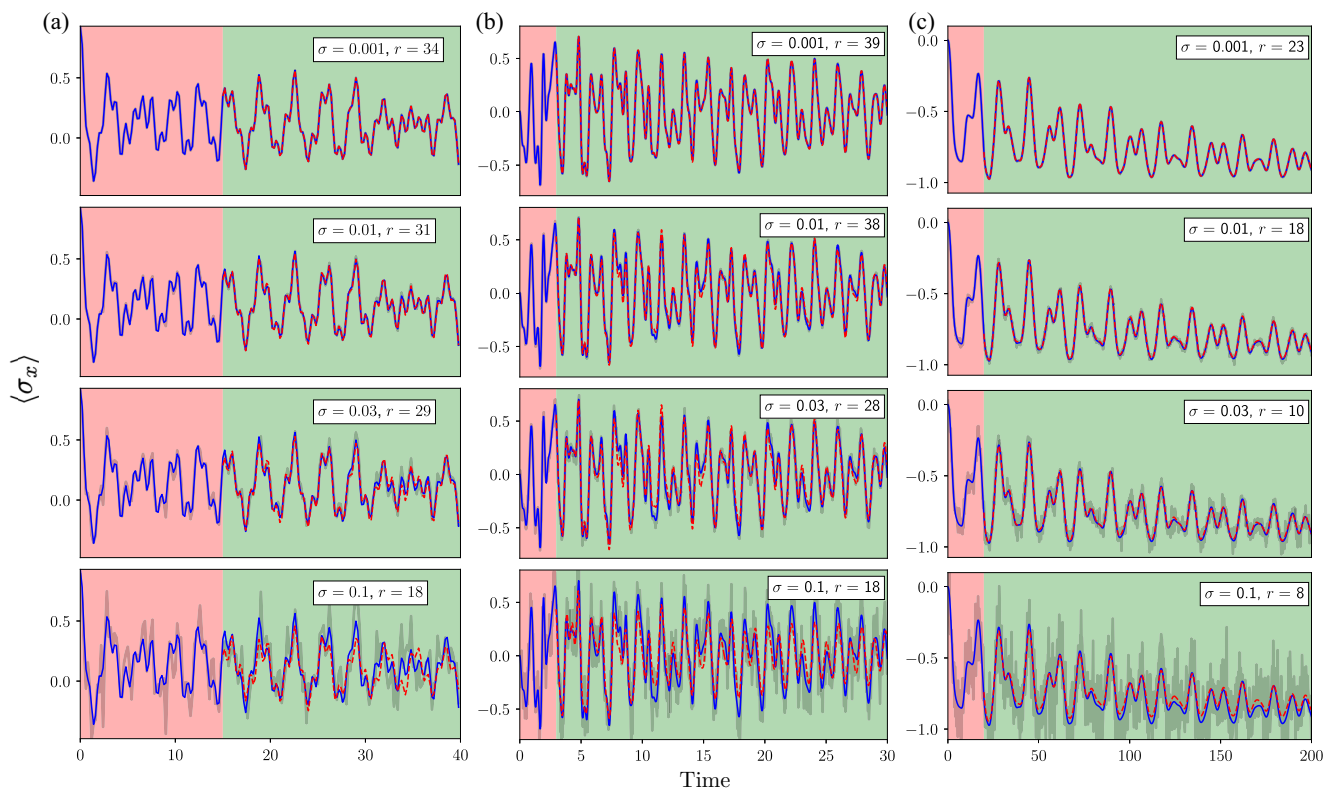


FIG. 7. Comparison of the noiseless dynamics of $\langle \sigma_x \rangle = \text{Tr}(\sigma_x \varrho)$ computed for the test trajectory (blue curve), noisy version of this dynamics (gray thick curve), and the data-based prediction (red dashed curve) within (a) the model with the predefined effective environment [Eq. (29)] of dimension $d_E = 3$, (b) The damped Jaynes-Cummings model [Eq. (31)] with parameters $\gamma = 0.05$, $g = 2.5$, $\alpha = 1.1$, and (c) the spin-boson model [Eq. (33)] with parameters $\gamma = 0.05$, $g = 0.5$, $\Delta = 0.5$. The value of r shows the estimated from data dimension of trajectories subspace. The region where the noisy trajectory chunk $T_K^{\text{test}}(0)$ is assumed to be known and employed for prediction of the further dynamics is highlighted by the red color. The region where one builds the prediction of the further dynamics $T_{N-K}^{\text{test}}(K)$ is highlighted by the green color.

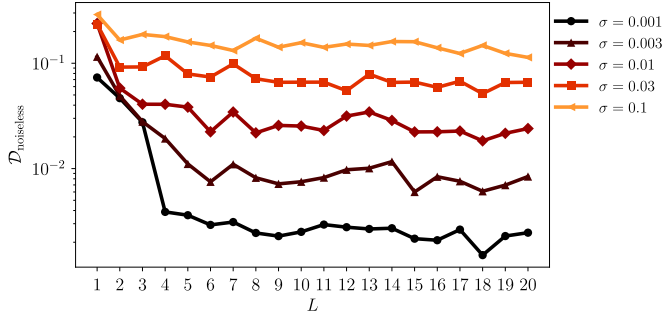


FIG. 8. The distance between prediction and noiseless test trajectory $\mathcal{D}_{\text{noiseless}}$ plotted against L and σ for the model with predefined effective environment [Eq. (29)]. The value of K is set equal to 75.

trajectories $T_R^{(a)}(t)$ and $T_R^{(b)}(t)$ of length R at time t as follows:

$$\mathcal{D}(T_R^{(a)}(t), T_R^{(b)}(t)) = \frac{1}{R} \sum_{k=0}^{R-1} \|\varrho^{(a)}(t+k) - \varrho^{(b)}(t+k)\|_1, \quad (30)$$

where states $\varrho^{(a)}$ and $\varrho^{(b)}$ are taken from the trajectories $T_R^{(a)}(t)$ and $T_R^{(b)}(t)$ correspondingly, and $\|\cdot\|_1$ denotes a trace norm [84]. Let $\mathcal{D}_{\text{noisy}}$ be the distance Eq. (30) between the prediction and the noisy version of a test trajectory $T_{N-K}^{\text{test}}(K)$ and $\mathcal{D}_{\text{noiseless}}$ be the distance Eq. 30 between the prediction and the noiseless version of a test trajectory $T_{N-K}^{\text{test}}(K)$. These distances show how close the predicted trajectory to the noisy test trajectory and to the noiseless test trajectory and thus can be used to validate the accuracy of the prediction. The behavior of the distance as a function of K is presented in Fig. 6(b). As one can see, there is a ‘‘saturation’’ value of K , starting from which the accuracy of prediction is mainly determined by the noise level. This means that the accuracy of the prediction is not sensitive to the choice of K , i.e., there is no effect of overfitting when the accuracy becomes worse with an increase of the model complexity. One can set K knowingly large and this does not lead to the effect of overfitting. This is the consequence of the proper automatic model selection. To validate how the accuracy of prediction depends on choice of L we generated additional data sets for L ranging from 1 to 20, $d_E^{\text{eff}} = 4$, $N = 200$ and for several different σ . For predictions based on these data sets, we plot the dependence of $\mathcal{D}_{\text{noiseless}}$ on L and σ in Fig. 8. One can observe that the accuracy rapidly improves with growing L at the beginning and then saturates. The saturation effect is caused by the fact that the prediction is based on an initial noisy prehistory of constant size (independent on L) and this noise can not be eliminated via an increase of data set size.

Finally, we validate the ability of the method to reduce noise in data. For this purpose we introduce the distance between data sets as the averaged over all trajectories distance introduced in Eq. 30 and denote this distance as $\langle \mathcal{D} \rangle$. One needs to consider the distance between denoised and noiseless data sets $\langle \mathcal{D}_{\text{denoised}} \rangle$ and the distance between noisy and noiseless data sets $\langle \mathcal{D}_{\text{noisy}} \rangle$. If $\langle \mathcal{D}_{\text{denoised}} \rangle$ is systematically smaller than $\langle \mathcal{D}_{\text{noisy}} \rangle$, then the denoising procedure works correctly. We plot $\langle \mathcal{D}_{\text{denoised}} \rangle$ and $\langle \mathcal{D}_{\text{noisy}} \rangle$ for $T = 200$, $K = 75$ and several different values of σ and d_E in Fig. 6(d). One can see that $\langle \mathcal{D}_{\text{denoised}} \rangle$ is systematically smaller than $\langle \mathcal{D}_{\text{noisy}} \rangle$. This

fact supports our claim on the correctness of the denoising procedure. This completes the validation of the proposed method.

B. Probing dynamics of dissipative Jaynes-Cummings model

As another example of non-Markovian quantum dynamics, we consider the dynamics of a two-level atom (the system) interacting with a decaying bosonic mode (the environment) via the Jaynes-Cummings (JC) interaction [94,95]. Note that in contrast to the previous example, the effective environment of a spin is a bosonic mode that is infinite-dimensional. However, in the previous example we considered a random GKSL generator driving the joint dynamics. Its randomness prevents further dimensionality reduction of the effective environment. In the given example, the bosonic mode can be truncated since we consider a case when the mode is not highly excited. This means that one can build an effective environment of finite dimension. The interesting question is whether the proposed algorithm can detect a low-dimensional effective environment from data. To address this question, we probe our method on this model and try to identify a low-dimensional effective environment. The details of the model are given below. The joint dynamics of the atom and the bosonic mode within the JC model is driven by the Lindblad equation that reads

$$\begin{aligned} \frac{d\varrho_{\text{SE}}}{dt} &= -i[H_{\text{JC}}, \varrho_{\text{SE}}] + \gamma \left(a\varrho_{\text{SE}}a^\dagger - \frac{1}{2}a^\dagger a\varrho_{\text{SE}} - \frac{1}{2}\varrho_{\text{SE}}a^\dagger a \right), \\ H_{\text{JC}} &= a^\dagger a + \frac{1}{2}\sigma_z + \frac{g}{2} \left(\frac{\sigma_x + i\sigma_y}{2} a + \text{H.c.} \right), \end{aligned} \quad (31)$$

where H_{JC} is the JC model Hamiltonian, $a(a^\dagger)$ is the bosonic mode annihilation (creation) operator, $\sigma_x, \sigma_y, \sigma_z$ are the Pauli matrices, g is the interaction strength, and γ is the bosonic mode dissipation rate. Taking the partial trace w.r.t. the bosonic mode, one obtains the density matrix of the atom, $\varrho(t) = \text{Tr}_E[\varrho_{\text{SE}}(t)]$, that experiences non-Markovian dynamics. As in the previous example, for $\gamma = 0.05$, $g = 2.5$ and various amplitudes of noise σ we prepare a number of noisy data sets $\{T_N^{(i)}(0)\}_{i=1}^L$ and noisy and noiseless versions of test trajectories $T_N^{\text{test}}(0)$ for each data set using the model above. We simulate the joint atom and mode dynamics driven by Eq. (31) at successive time steps of length $\tau = 0.03$, and compute the corresponding trajectories of the system by taking a partial trace over the environment. The initial joint atom and mode state is taken in the factorized form

$$\varrho_{\text{SE}}(0) = |\psi\rangle\langle\psi| \otimes |\alpha\rangle\langle\alpha|, \quad (32)$$

where $|\alpha\rangle$ is the coherent state of the bosonic mode, and $|\psi\rangle$ is sampled uniformly from the Bloch sphere. For all data sets we fix $\alpha = 1.1$. To proceed with numerical simulation of Eq. (31) we truncate the infinite-dimensional Hilbert space of the bosonic mode keeping such a number of eigenstates with the lowest energy that guarantees conservation of $>95\%$ of the initial environment state $\varrho_E(0)$ probability mass. The truncation is needed only for the tractability of numerical calculations necessary for data set generation. The limit of applicability of our data processing method is not sensitive to the genuine dimensionality of the environment; it can be either finite or infinite-dimensional. As mentioned before, only the

finiteness of the memory matters. We fix $L = 2, K = 100$, and $N = 1000$ for all data sets.

We apply our method to generated data sets and reconstruct r and M . As in the previous example, we use M to predict behavior of the test trajectory. The comparison of the prediction and the test trajectory for each amplitude of noise is shown in Fig. 7(b). One can see that one predicts dynamics of the atom even in presence of relatively high noise. We also note, that in all considered cases $r \leq 39$, while the dimension $d^2[d_E^{\text{eff}}]^2$ of the joint system and environment density matrix after truncation of the bosonic mode is 100, which is well above 39. This justifies the truncation of the bosonic mode and shows the finiteness of d_E^{eff} even though the genuine environment is infinite. The main outcome of this numerical experiment is that the proposed method is able to identify a finite-dimensional effective environment even in the case of infinite-dimensional genuine environment.

C. Probing the spin-boson model dynamics

The third example of non-Markovian quantum dynamics, which we analyze with our method, is the dynamics of a two-level atom coupled with a set of noninteracting bosonic modes. We refer to this model as the spin-boson model [14]. We consider the dynamics of a two-level atom in the limit of a continuum of bosonic modes. In this case there is no a straightforward way to extract a finite dimensional effective environment using first-principle numerical modeling. We test the ability of the proposed method to do this from a data set as well as the ability to predict the non-Markovian dynamics of the spin. We also study how the dimension r reconstructed from data depends on some parameters of the spin-boson model and show that r serves as a non-Markovianity or complexity measure in this case. The Hamiltonian of the spin-boson model reads

$$H_{\text{SB}} = \frac{1}{2}\sigma_z + \frac{1}{2}\Delta\sigma_x + \sum_k \omega_k a_k^\dagger a_k + \sigma_z X, \\ X = \sum_k \frac{g_k}{\sqrt{2\omega_k}}(a_k^\dagger + a_k), \quad (33)$$

where Δ is the tunneling matrix element, ω_k are frequencies of bosonic modes, $a_k(a_k^\dagger)$ are annihilation (creation) operators of bosonic modes, and g_k is the strength of interaction between the atom and the k th mode. As before, we consider the atom as a system and the set of bosonic modes as an environment. The atom affected by the bosonic modes experiences non-Markovian quantum dynamics that can be analyzed by our method. We consider the case when the initial state of bosonic modes is the ground state uncorrelated with the initial state of the atom. In this case, the influence of the environment of noninteracting bosonic modes on the system is fully described by the two-time correlation function of bosonic modes

$$C(t) = \langle \text{vac} | e^{i\sum_k \omega_k a_k^\dagger a_k t} X e^{-i\sum_k \omega_k a_k^\dagger a_k t} X | \text{vac} \rangle \\ = \int_0^\infty \frac{J(\omega)}{\pi} \exp(-i\omega t), \\ J(\omega) = \pi \sum_k \frac{g_k^2}{2\omega_k} \delta(\omega - \omega_k), \quad (34)$$

where $|\text{vac}\rangle$ is the ground state of bosonic modes, $J(\omega)$ is the spectral density. We consider the limit of continuum bosonic modes when the gap between frequencies of neighboring modes $\omega_{k+1} - \omega_k$ vanishes. In this limit we choose the following spectral density:

$$J(\omega) = \frac{\gamma g^2 \omega}{(\omega^2 - \omega_0)^2 + \gamma^2 \omega^2}, \quad (35)$$

where ω_0 is the resonance frequency, γ is the width of the spectral function, g is the aggregated interaction strength.

We simulate the dynamics of the atom with an arbitrary pure initial state $|\psi\rangle$ using the numerical approach and code developed in Ref. [96]. As before, we prepared a number of data sets $\{T_N^{(i)}(0)\}_{i=0}^L$ and corresponding test trajectories $T_N^{\text{test}}(0)$ specified by various amplitudes of noise and parameters of the spin-boson model. The initial state of the two-level atom $|\psi\rangle$ is sampled uniformly from the Bloch sphere for each trajectory of a data set. We fix $L = 2, \tau = 0.15, K = 100$, and $N = 1000$ for all data sets. As before, we reconstruct M from a data set and use it to predict the behavior of the test trajectory. The comparison of the prediction and the test trajectory is given in Fig. 7(c). One can see again, that even for the relatively high amplitude of noise, the approach is capable of predicting the atom dynamics.

To demonstrate on a concrete example that the trajectories space dimension r can be also considered as a measure of non-Markovianity and system's dynamics complexity, we study a relation between r and the value of the parameter γ of the considered spin-boson model. Note that γ determines the width of the two-time correlation function $C(t)$. The smaller γ is the bigger the width of the correlation function is. Therefore, the smaller γ is the stronger memory effects are, and the more challenging the atom dynamics simulation is. We also consider different values of K to understand how the memory depth of the dynamics depends on γ . We generate several noiseless data sets for $L = 4, T = 1000, \tau = 0.15$ and different values of γ . For each value of γ we process the corresponding data set by our method. We consider different values of the hyperparameter K and the fixed value of $\sigma = 10^{-6}$ to cut off the numerical simulation error. We start from the analysis of the prediction accuracy. In Fig. 9(a) we show how the accuracy of prediction $\mathcal{D}_{\text{noiseless}}$ depends on K and γ . One can observe that a steep improvement of the accuracy then changes to a smooth saturation regime with increasing K . Note that for smaller γ one has worse accuracy of prediction. This observation agrees well with the fact that for smaller γ , dynamics becomes more complicated. Then we turn to the analysis of r . In Fig. 9(b) we show how the minimal Markovian embedding dimension depends on different values of K and γ . One can see that value of r saturates starting from some sufficiently large value of K . Note also that for smaller γ the value of K , for which r saturates, is bigger. This supports the fact that for smaller γ one has deeper memory. But the most important observation is that for smaller γ one has r bigger. In Fig. 9(c), for $K = 500$ (which is well above the memory depth for all considered γ) we show the dependence of r on γ explicitly. One sees that r decreases with increasing γ , which is in accord with interpretation of r as the complexity

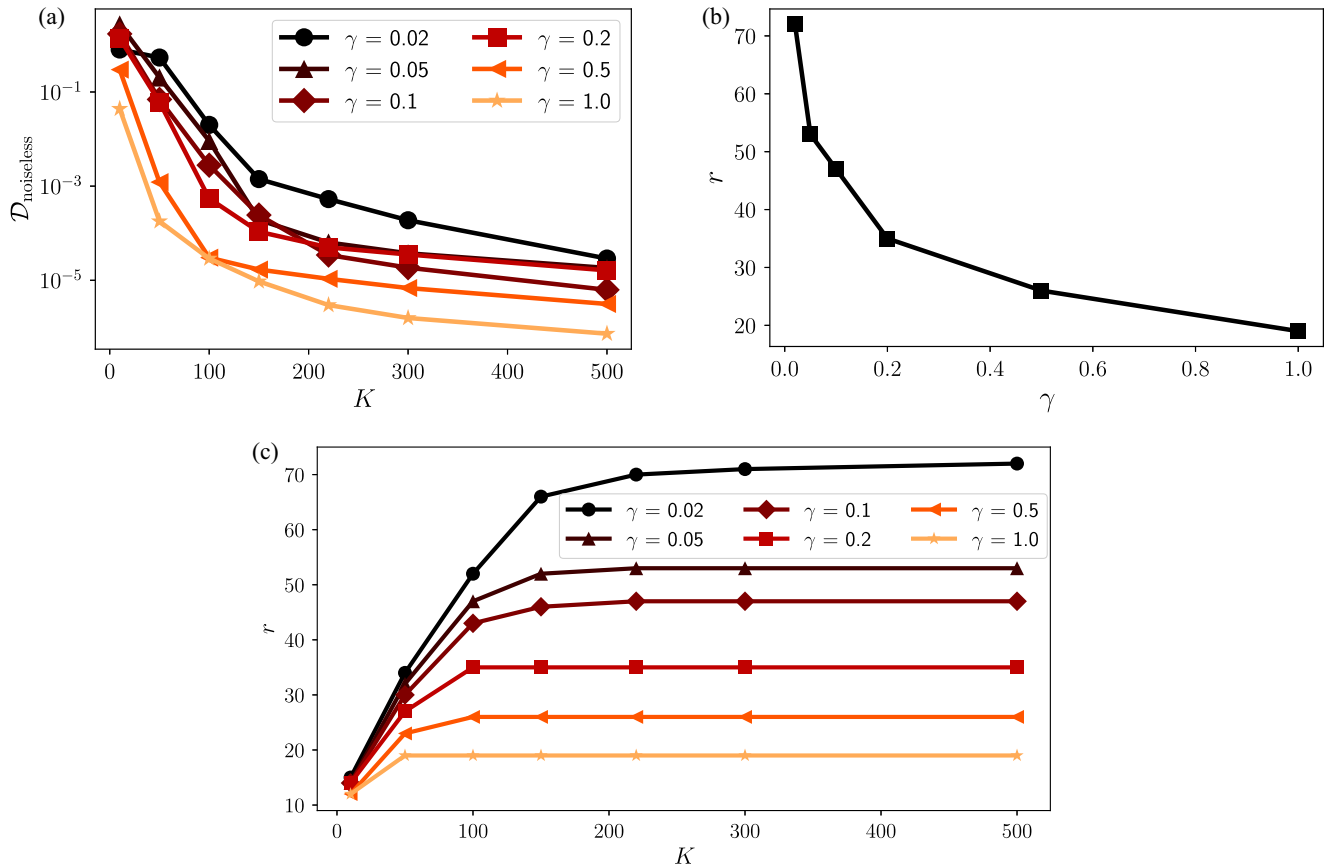


FIG. 9. Analysis of the spin boson model [Eq. (33)] within the proposed method. (a) Accuracy of prediction for different values of K and γ . (b) Trajectories subspace dimension for different values of K and γ . (c) Dependence of the trajectories subspace dimension on γ for $K = 500$.

of a non-Markovian quantum dynamics or a measure of non-Markovianity.

V. COMPARISON WITH EXISTING DATA-DRIVEN METHODS OF NON-MARKOVIAN QUANTUM DYNAMICS IDENTIFICATION

In this section, we compare the given method of non-Markovian quantum dynamics identification with previously proposed ones, especially focusing attention on the related “black-box” methods namely the TTM [45–51,54], and a method based on a recurrent neural network (RNN) parametrization of a predictive model [55]. We evaluate all the methods by means of the following metrics: ability of a method to extract information about the environment of a non-Markovian quantum system, data efficiency, and robustness to noise.

The first and foremost distinguishing feature of the proposed method is its ability to extract information about the quantum environment of a system by observing only quantum trajectories of a system. It estimates the effective dimension of the environment and reconstructs eigenvalues of the joint system and environment dynamics. The effective dimension of the environment shows the size of the environment fraction that mutually exchanges information with a system. The remaining fraction of the environment is seen as a decoupled

part, since it only absorbs information and does not bring any information back. The eigenvalues of the joint system and environment dynamics are essentially spectroscopic data that defines the physics of the system and environment as a whole. All previously proposed methods do not reconstruct any information about the system’s quantum environment, except maybe attempts to reconstruct the effective dimension of the environment by looking for the “elbow” in the validation curve plot [61,62] while fitting the observed data by process-tensor-based methods. However, this method requires a separate validation data set and multiple retraining of a process-tensor-based model, while the proposed method is free from these drawbacks.

The next indicators of performance are data efficiency and robustness to noise. Our method is equipped with the automatic model selection that selects the simplest possible model among describing observed data. This is achieved by identification of the low-dimensional linear subspace where all the quantum trajectories are placed. The dynamics of the system in this subspace can be described by a model with a fewer number of parameters, which implies less data for learning and better robustness against noise. To demonstrate the data efficiency and robustness to noise of our method, we compare the prediction of a non-Markovian quantum dynamics built using our method with that which is built using the TT method that serves as an example of a “black-box”

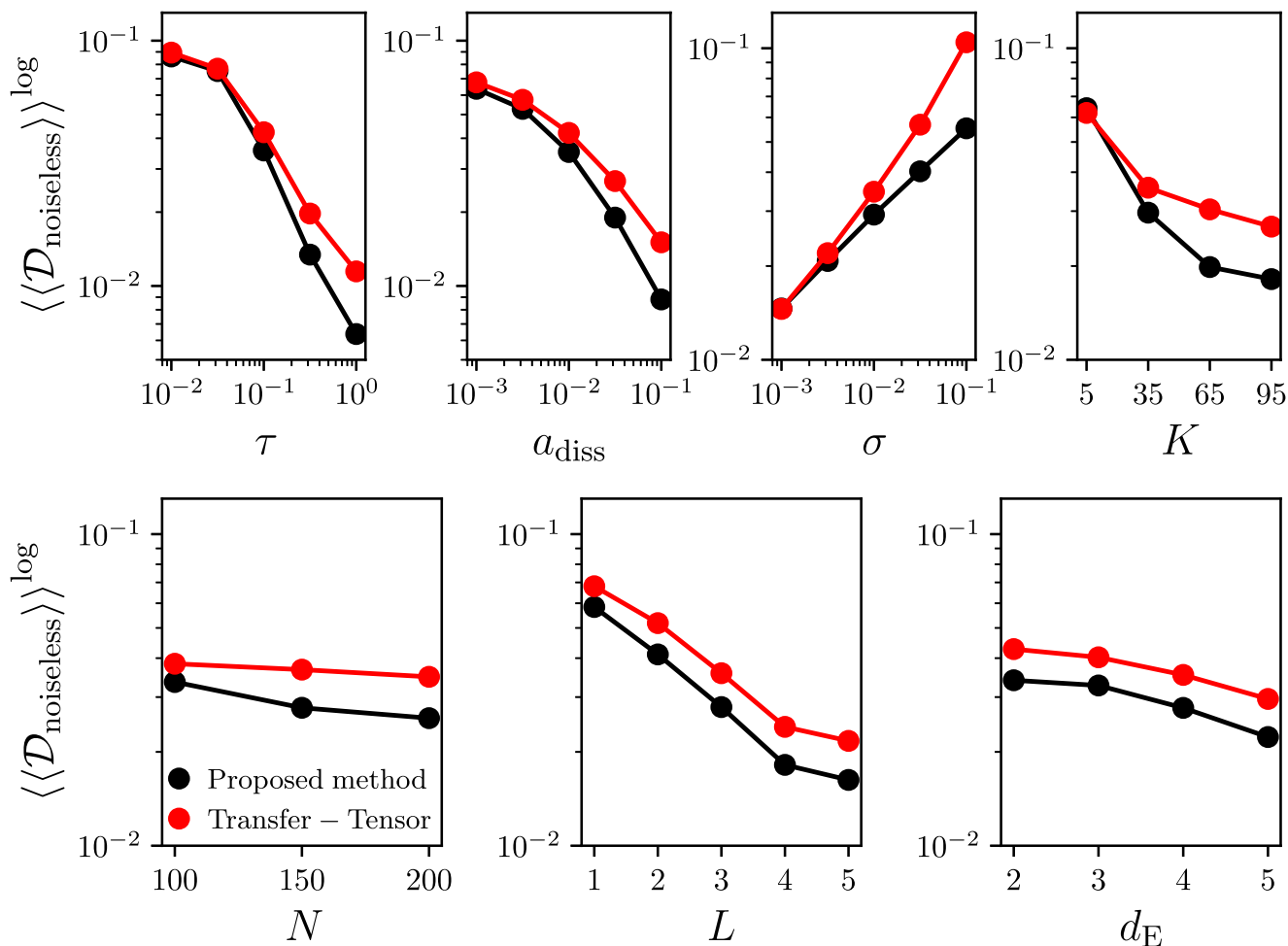


FIG. 10. The comparison of the averaged prediction accuracy of the proposed method and the TTM for the model with a predefined effective environment [Eq. (29)]. The value of $\mathcal{D}_{\text{noiseless}}$ is the distance Eq. (30) between the prediction built on top of noisy “prehistory” of a test trajectory and the noiseless remaining part of a trajectory. The first averaging sign in $\langle\langle \mathcal{D}_{\text{noiseless}} \rangle\rangle^{\log}$ means averaging the distance over all trajectories in the test set, the second averaging sign with superscript log means that for each particular plot one performs logarithmic averaging, i.e., $\langle\langle \mathcal{D}_{\text{noiseless}} \rangle\rangle = \exp(\log(\mathcal{D}_{\text{noiseless}}))$, over all parameters τ , σ , K , N , L , d_E , a_{diss} , κ but the parameter that is the x axis.

method, and which does not account for the physics under the hood of the observed data. To make our comparison systematic we performed multiple numerical experiments (300 000 in total) and compared the performance of both methods for each experiment. For the learning of the TTM model from the data, we utilized a ridge regression [66] to solve the following optimization problem:

$$\begin{aligned} & \text{minimize}_W \sum_{i=1}^L \sum_{t=0}^{N-K-1} \|\varrho^{(i)}(t+K+1) - W[T_K^{(i)}(t)]\|_F^2 \\ & + \theta \|W\|_F^2, \end{aligned} \tag{36}$$

where W is the kernel of the discrete Nakajima-Zwanzig equation that one reconstructs from the data and θ is the regularization coefficient taken equal to σ to mitigate the effect of noise. Each experiment consisted of a data preparation stage, when test and training sets of trajectories are generated within a particular model, a fitting stage, when both the proposed method and the TTM are applied to a training

trajectories set to build data-driven model of the dynamics, and a test stage, when both data-driven models are used to predict dynamics of trajectories from a test set. After the test stage, prediction errors of both methods are evaluated. As the error measure we used $\langle\mathcal{D}_{\text{noiseless}}\rangle$ that is distance Eq. (30) between the prediction built on top of noisy “prehistory” of a test trajectory and the noiseless remaining part of a trajectory averaged over all trajectories from a test set. As a model of non-Markovian dynamics we used the model with a predefined effective environment considered in the Sec. IV A. Each experiment is defined by a particular tuple of parameters $(\tau, \sigma, K, N, L, d_E, a_{\text{diss}}, \kappa)$ where τ is the time-step size, σ is the amplitude of noise, K is the guess of memory depth, N is the number of discrete time steps in test and training sets, L is the number of trajectories in a training set, d_E is the predefined dimension of the effective environment, a_{diss} is the dissipation rate (see Appendix B) and κ is the random seed that is used to generate the Lindbladian driving the dynamics of the system and the environment. We performed numerical experiments for all tuples of parameters from the

TABLE II. Comparison of different non-Markovian quantum dynamics identification methods with the method proposed in the present research.

	Information about the environment that can be reconstructed	Data efficiency and robustness to noise	Mathematical guarantees of convergence to the optimal point	Is able to predict the response of a non-Markovian quantum system on an external perturbation?
TTM		The most general model requires a large data set	Converges with mathematical guarantees	No
Methods based on learning of a matrix product state representation of a process tensor	Effective dimension [61]	Cross-validation based bond dimension tuning improves data efficiency	No mathematical guarantees of convergence	Yes
Methods based on tomography of a process tensor	Eigenfrequencies, effective dimension	Extremely data demanding (requires exponential in the number of discrete time steps amount of data points)	Converges with mathematical guarantees	Yes
RNN based method		“Black-box” model requires a large data set	No mathematical guarantees of convergence	No
Proposed method	Eigenfrequencies, effective dimension	The automatic model selection improves data efficiency and robustness to noise in contrast with “black-box” methods	Converges with mathematical guarantees	No

set

$$\begin{aligned}
& (\tau^{(1)}, \dots, \tau^{(5)}) \times (\sigma^{(1)}, \dots, \sigma^{(5)}) \times (K^{(1)}, \dots, K^{(4)}) \\
& \times (N^{(1)}, \dots, N^{(3)}) \times (L^{(1)}, \dots, L^{(5)}) \times (d_E^{(1)}, \dots, d_E^{(4)}) \\
& \times (a_{\text{diss}}^{(1)}, \dots, a_{\text{diss}}^{(5)}) \times (\kappa^{(1)}, \dots, \kappa^{(10)}), \quad (37)
\end{aligned}$$

where \times stands for the Cartesian product and superscripts mark a particular value of a parameter used in numerical experiments. In order not to list all values of parameters here we refer a reader to Fig. 10 where the x axis of subplots shows the ranges of parameters we used.

The obtained high-dimensional data needs to be visualized in a concise way. To do that we introduce a logarithmic averaging that we define as follows:

$$\langle A \rangle^{\log} = \exp((\log A)), \quad (38)$$

that is more preferable than the standard averaging since the value of $\langle \mathcal{D}_{\text{noiseless}} \rangle$ may vary several orders of magnitude for different parameters. For the sake of demonstration of what is done next, let us fix σ as a target parameter. For σ we calculated $\langle \langle \mathcal{D}_{\text{noiseless}} \rangle \rangle_{\tau, K, N, L, d_E, a_{\text{diss}}, \kappa}^{\log}$ that is the logarithmic averaged value of $\langle \mathcal{D}_{\text{noiseless}} \rangle$ over all parameters but σ . We did the same averaging for all parameters except κ and plot the results in Fig. 10. One can think of this plots as “projections” of the obtained high-dimensional data on a particular plane. They represent how the prediction accuracy for both methods depends on a particular parameter. One can note, that the proposed method outperforms the TTM almost always except cases with small noise, small memory depth, small

time-step size, where both methods perform equally. One can also note, that the proposed method compared to the TTM works especially good in case of high amplitudes of noise, that is especially practical one. These observations approve the proposed method’s improved data-efficiency and robustness to noise in comparison with previously propose “black-box” methods such as TTM of RNN based method. The overall comparison summary is given in Table II. To conclude the presented comparison.

To conclude the presented comparison, one can note that the proposed method is a substantial step forward toward interpretable data-driven methods of non-Markovian quantum dynamics analysis. This is the first method allowing automatic extraction of information about the environment from the measured system dynamics and it outperforms previously proposed “black-box” methods, such as TTM, in terms of robustness to noise and data efficiency.

VI. DISCUSSION AND OUTLOOK

We developed a data-driven method for non-Markovian quantum dynamics analysis. The input data processed by the method is a set of quantum trajectories, i.e., sequences of density matrices of the system at consecutive time steps. We note that quantum trajectories can be reconstructed with the use of quantum state tomography that is de facto the gold standard for the characterization of quantum information processing devices [88], and is a standard tool for studying the performance of various physical quantum computing platforms including photons [97], trapped ions [98], and superconduct-

ing circuits [99]. So, the method we developed is well-suited to the study of existing noisy intermediate-scale quantum (NISQ) devices including listed above. We also note that the developed approach takes into account inevitable discrepancies between true density matrices (which one would obtain with infinite number of perfect measurements) and practically reconstructed ones obtained with finite amount of experimental statistics. The estimated level of this reconstruction noise (the standard deviation σ), as well as the guess about the memory depth K , are the method's hyperparameters.

The analysis of the non-Markovian dynamics performed within our method consists of two major steps. At the first stage, the method determines the simplest model describing observed data, in conformity with Occam's razor principle, and outputs the hyperparameter r defining the complexity of the model. This simplest model selection corresponds to the best machine-learning practices and improves both data efficiency and noise robustness of the proposed method [66]. Moreover, r defines the dimension of the effective environment that is an important characteristic of the genuine environment. At the second stage, the method reconstructs the predictive model based on the results at the previous step. The obtained model allows one to predict the continuation of the quantum system's trajectory consisting of at least K elements.

Both stages rely on matrix decompositions guaranteeing convergence to the optimal solution while most of previously proposed methods rely on an approximate nonconvex optimization [55,61,62,68]. Due to the established relation between a quantum system's trajectories and the joint system and environment dynamics, the reconstructed predictive model is not just a "black box," but an interpretable physical model of non-Markovian quantum dynamics. It allows not only the system's dynamics prediction, but also provides additional information on the underlying open quantum system properties, in particular, the effective dimension of the environment d_E^{eff} and eigenfrequencies of the joint system and environment dynamics. As an additional valuable output, the proposed method performs quantum trajectories denoising by projecting them onto the r -dimensional principle subspace. We note that the introduced denoising procedure can be used in combination with other approaches for improving accuracy of quantum tomography, including compressive sensing [87] and employing tensor-product structures [88].

We have validated the performance of our approach for dynamics of two-level system interacting either with a known finite-dimensional effective environment or with an infinite dimensional environment (JC model and spin-boson model). For the first case, we have justified the ability of the method to recover the correct information on effective environment, and for the latter, we have validated the performance of the method with respect to the models related to actual experiments (see, e.g., Refs. [100,101]). We note that in both cases, the considered values of reconstruction noise σ , are taken to mimic realistic tomographic experiments.

Our technique is highly relevant for ongoing experiments with NISQ processors, where quantum state tomography of several-qubit systems is an available tool. Our method makes it possible to extract relevant information about the environment affecting a quantum processor, which is extremely hard

to measure directly, and build its dynamics prediction. It potentially allows building new types of data-driven controllers for the next generation of quantum processors taking into account non-Markovian dynamics and using non-Markovianity as a resource.

We also note that the method is directly applicable to the analysis of incomplete data (trajectories of some elements of a density matrix or trajectories of single observables). Indeed, one requires only the linearity of ϕ_K ; it can output either a trajectory of density matrices, a trajectory of certain elements of density matrices, or a trajectory of certain observables. However, in this case, the physical meaning of d_E^{eff} is different and requires further analysis. This is in the scope of the next works.

Further research is also required to determine whether our scheme is capable of predicting the response of a non-Markovian system to an external perturbation. Such a prediction capability would open a promising line of research with the development of new data-driven quantum control methods. We also need to study how the proposed approach could be generalized on the case of time dependent interaction between a system and its environment. Specifically, is it possible to capture the dynamics with varying K ? This generalization would be useful to detect transitions between different dynamics regimes and dynamical phase transitions.

Realization of the method in Python as well as numerical experiments for the model with predefined effective environment are available publicly in Ref. [102].

ACKNOWLEDGMENTS

The development of the data processing scheme, analysis of the spin-boson model, and analysis of the damped Jaynes-Cummings model are supported by the Russian Science Foundation (Grant No. 19-71-10092), by the Leading Research Center on Quantum Computing (Agreement No. 014/20; analysis of non-Markovian processes for NISQ devices), and by the Priority 2030 program at the National University of Science and Technology "MISIS" under the project K1-2022-027 (applications to various quantum models). The analysis of the finite-environment-induced non-Markovian quantum dynamics is supported by the Foundation for the Advancement of Theoretical Physics and Mathematics "BASIS" for support under Project No. 19-1-2-66-1. The authors thank Alexander Ryzhov and Georgiy Semin for fruitful discussions.

I.A.L. came up with the original idea of the presented approach. I.A.L., E.O.K., and M.A.G. performed verification of the approach on various physical models of non-Markovian dynamics. All authors analyzed and interpreted the data. I.A.L., E.O.K., and A.K.F. wrote the paper with critical input from M.A.G., H.O., and S.N.F.

APPENDIX A: SUFFICIENCY OF K

To introduce the criterion of K to be sufficient, i.e., to understand which value of K is enough for the Markovianity of trajectories dynamics, let us properly define the linear map ϕ_K that is not discussed in detail in the main text. We define

this linear map through its action on an arbitrary joint system and environment state that reads

$$\phi_K : \rho_{SE} \mapsto [\text{Tr}_E(\rho_{SE}), \text{Tr}_E(\Phi[\rho_{SE}]), \dots, \text{Tr}_E(\Phi^{K-1}[\rho_{SE}])]. \quad (\text{A1})$$

The object of our interest is the orthogonal complement of the kernel of ϕ_K , i.e., the relevant subspace that is introduced in the main text. One can note, that the relevant subspace is essentially an example of a Krylov subspace due to the structure of ϕ_K . Like for any Krylov subspace its dimension $r(K)$ increases with increasing K until some critical values of r and K are reached after which r saturates and remains constant (see, for example, Ref. [103], Lemma 1). The effect of saturation of r takes place when

$$\ker^\perp(\phi_{K+1}) = \ker^\perp(\phi_K), \quad (\text{A2})$$

which means that the relevant subspace does not change with increasing K anymore. More precisely it can be formulated as follows. There exist K_c such that $r(K) < r(K+1)$ if $K < K_c$ and $r(K) = r(K+1) = r_c$ if $K \geq K_c$. Let us show that this is indeed true. First of all r is bounded from above, since a linear span of any number of vectors from a finite dimensional linear space cannot be greater than the dimension of this space. Second, $r(K)$ is a nondecreasing function of K , because by adding new vectors to a linear span one can not decrease the dimension of this linear span. Finally, one needs to prove that once $\ker^\perp(\phi_{K+1}) = \ker^\perp(\phi_K)$, then for any $k \in \mathbb{N}$ the following is true $\ker^\perp(\phi_{K+k}) = \ker^\perp(\phi_K)$. The equality $\ker^\perp(\phi_{K+1}) = \ker^\perp(\phi_K)$ implies that $\ker^\perp(\phi_{K+1}\Phi) = \ker^\perp(\phi_K\Phi)$ and $\text{span}[\phi_1, \ker^\perp(\phi_{K+1}\Phi)] = \text{span}[\phi_1, \ker^\perp(\phi_K\Phi)]$. Now note that

$$\begin{aligned} \text{span}[\phi_1, \ker^\perp(\phi_K\Phi)] &= \ker^\perp(\phi_{K+1}), \\ \text{span}[\phi_1, \ker^\perp(\phi_{K+1}\Phi)] &= \ker^\perp(\phi_{K+2}), \end{aligned} \quad (\text{A3})$$

which leads us to the equality $\ker^\perp(\phi_{K+2}) = \ker^\perp(\phi_{K+1})$. Applying the same logic again and again we get $\ker^\perp(\phi_{K+3}) = \ker^\perp(\phi_{K+2})$, $\ker^\perp(\phi_{K+4}) = \ker^\perp(\phi_{K+3})$, etc. This proves the equality $\ker^\perp(\phi_{K+k}) = \ker^\perp(\phi_K)$. Gathering all together we get the described above behavior, i.e., $r(K)$ strictly increases up to some critical value with increasing K and then saturates and remains constant.

The equality Eq. (A2) is equivalent to the following relation:

$$\ker^\perp(\phi_K\Phi) \subseteq \ker^\perp(\phi_K). \quad (\text{A4})$$

This relation implies that there exists such a linear map M that

$$\phi_K\Phi = M\phi_K, \quad (\text{A5})$$

which immediately leads to

$$T_K(t+1) = \phi_K\Phi[\rho_{SE}(t)] = M\phi_K[\rho_{SE}(t)] = M[T_K(t)], \quad (\text{A6})$$

i.e., there exists a Markovian master equation driving the dynamics of trajectories of size K for any initial ρ_{SE} . One can note that the inverse statement is also true. This brings us to the criterion of K to be sufficient:

Criterion 1. The value of K is sufficient for Markovianity of quantum trajectories iff

$$\ker^\perp(\phi_K\Phi) \subseteq \ker^\perp(\phi_K). \quad (\text{A7})$$

Let us derive an important upper bound of the minimal sufficient K . Note, that if K is insufficient, then the following holds:

$$r(K+1) \geq r(K) + 1, \quad (\text{A8})$$

or in other words,

$$r(K) > K. \quad (\text{A9})$$

Note also that the dimension of the relevant subspace does not exceed $d^2 d_E^2$. Therefore, one has

$$d^2 d_E^2 \geq r(K) > K, \quad (\text{A10})$$

which states that any insufficient K is less than $d^2 d_E^2$. This means that the minimal sufficient K is not greater than $d^2 d_E^2$ and for any finite dimensional environment the minimal sufficient K is also finite. This also means that the minimal sufficient K is less or equal to r .

Equation (A10) implies that in the worst case the minimal sufficient K and r grow exponentially with the number of subsystems of the environment. In the thermodynamic limit the minimal sufficient K and r may even tend to infinity. Our conjecture is that this case corresponds to the dynamical chaos. Indeed, the same unbounded growth of r takes place when one builds time-delay embedding of classical chaotic dynamics [70]. This also is well compatible with understanding of r as the complexity measure. When r and the minimal K tend to infinity this means infinite complexity or chaotic behavior.

APPENDIX B: GKSL GENERATOR OF THE MODEL WITH PREDEFINED EFFECTIVE ENVIRONMENT

In this subsection we describe how we obtain a GKSL generator $\mathcal{L} : \mathcal{B}(\mathcal{H} \otimes \mathcal{H}_E) \rightarrow \mathcal{B}(\mathcal{H} \otimes \mathcal{H}_E)$ driving dynamics of the joint system and effective environment density matrix. The corresponding Lindblad equation reads

$$\begin{aligned} \frac{d\rho_{SE}}{dt} &= \mathcal{L}(\rho_{SE}) \\ &= -i a_{\text{unit}} [H, \rho_{SE}] \\ &\quad + a_{\text{diss}} \sum_{i,j=1}^{d^2 d_E^2 - 1} \gamma_{ij} \left(F_i \rho_{SE} F_j^\dagger - \frac{1}{2} \{F_j^\dagger F_i, \rho_{SE}\} \right). \end{aligned} \quad (\text{B1})$$

Amplitudes a_{unit} and a_{diss} determine contributions of the Hamiltonian and the dissipative parts of the equation into the dynamics. In all experiments we choose $a_{\text{unit}} = 1$, $a_{\text{diss}} = 0.003$, until otherwise stated. The Hamiltonian H is generated randomly as follows:

$$H = \frac{1}{2}(A + A^\dagger), \quad \text{Re}(A) \sim \mathcal{N}(0, I), \quad \text{Im}(A) \sim \mathcal{N}(0, I), \quad (\text{B2})$$

where $\mathcal{N}(0, I)$ is the matrix normal distribution with zero mean and the identity covariance matrix. The positive semi-definite matrix γ that characterizes dissipation in the system and environment is also generated randomly. It reads

$$\gamma = Q\Lambda Q^\dagger, \quad (\text{B3})$$

where Λ is a diagonal matrix with diagonal elements sampled uniformly from the interval $[0, 1]$ and Q is a unitary matrix generated in two steps: (1) one randomly generates a matrix

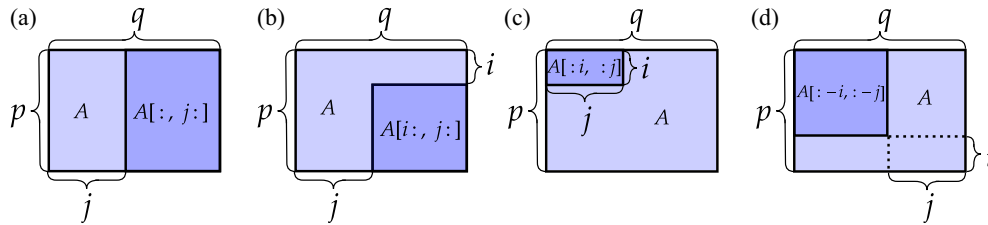


FIG. 11. Illustration of NumPy notations for matrices slicing. (a) First j columns are removed, (b) first j columns and first i rows are removed, (c) only first j columns and i rows are kept, (d) last i columns and last i rows are removed.

A such that $\text{Re}(A) \sim \mathcal{N}(0, I)$ and $\text{Im}(A) \sim \mathcal{N}(0, I)$, (2) one performs QR factorization of A and takes unitary factor as Q . The set of real, traceless matrices $\{F_1, F_2, \dots, F_{d^2-d-1}\}$ forms an orthonormal basis, i.e., $\text{Tr}(F_i^\dagger F_j) = \delta_{ij}$, where δ_{ij} stands for the Kronecker's symbol.

APPENDIX C: NUMPY NOTATIONS FOR TENSORS SLICING

To represent submatrices of a matrix we use notations that are standard in many programming packages for numerical

computation such as NumPy [104]. Let us consider a rectangular matrix A of the size $p \times q$, that can be of complex, real, or any other type. If we want to consider a truncated version of A for which we remove first j columns, then we write $A[:, j:]$. If we additionally want to remove first i rows, then we write $A[i:, j:]$. If we want to keep only first j columns and first i rows, then we write $A[:i, :j]$. It is also possible to count columns and rows starting from the other end using negative integers as indices. For example, if we want to remove last j columns and last i rows, then we write $A[:-i, :-j]$. All the examples are also illustrated in Fig. 11.

[1] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O'Brien, Quantum computers, *Nature (London)* **464**, 45 (2010).

[2] I. M. Georgescu, S. Ashhab, and F. Nori, Quantum simulation, *Rev. Mod. Phys.* **86**, 153 (2014).

[3] I. Bloch, J. Dalibard, and S. Nascimbene, Quantum simulations with ultracold quantum gases, *Nat. Phys.* **8**, 267 (2012).

[4] R. Blatt and C. F. Roos, Quantum simulations with trapped ions, *Nat. Phys.* **8**, 277 (2012).

[5] N. Friis, O. Marty, C. Maier, C. Hempel, M. Holzäpfel, P. Jurcevic, M. B. Plenio, M. Huber, C. Roos, R. Blatt, and B. Lanyon, Observation of Entangled States of a Fully Controlled 20-Qubit System, *Phys. Rev. X* **8**, 021012 (2018).

[6] S. Trotzky, Y.-A. Chen, A. Flesch, I. P. McCulloch, U. Schollwöck, J. Eisert, and I. Bloch, Probing the relaxation towards equilibrium in an isolated strongly correlated one-dimensional bose gas, *Nat. Phys.* **8**, 325 (2012).

[7] A. Mazurenko, C. S. Chiu, G. Ji, M. F. Parsons, M. Kanász-Nagy, R. Schmidt, F. Grusdt, E. Demler, D. Greif, and M. Greiner, A cold-atom Fermi–Hubbard antiferromagnet, *Nature (London)* **545**, 462 (2017).

[8] A. Keesling, A. Omran, H. Levine, H. Bernien, H. Pichler, S. Choi, R. Samajdar, S. Schwartz, P. Silvi, S. Sachdev *et al.*, Quantum Kibble–Zurek mechanism and critical dynamics on a programmable rydberg simulator, *Nature (London)* **568**, 207 (2019).

[9] J. T. Barreiro, M. Müller, P. Schindler, D. Nigg, T. Monz, M. Chwalla, M. Hennrich, C. F. Roos, P. Zoller, and R. Blatt, An open-system quantum simulator with trapped ions, *Nature (London)* **470**, 486 (2011).

[10] R. Vasseur and J. E. Moore, Nonequilibrium quantum dynamics and transport: From integrability to many-body localization, *J. Stat. Mech.: Theory Exp.* (2016) 064010.

[11] D. A. Abanin, E. Altman, I. Bloch, and M. Serbyn, Colloquium: Many-body localization, thermalization, and entanglement, *Rev. Mod. Phys.* **91**, 021001 (2019).

[12] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner *et al.*, Probing many-body dynamics on a 51-atom quantum simulator, *Nature (London)* **551**, 579 (2017).

[13] H.-P. Breuer, F. Petruccione *et al.*, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, UK, 2002).

[14] I. de Vega and D. Alonso, Dynamics of non-Markovian open quantum systems, *Rev. Mod. Phys.* **89**, 015001 (2017).

[15] R. Alicki and K. Lendi, *Quantum Dynamical Semigroups and Applications* (Springer, Berlin, 2007), Vol. 717.

[16] L. Accardi, Y. G. Lu, and I. Volovich, *Quantum Theory and Its Stochastic Limit* (Springer Science & Business Media, Cham, 2013).

[17] S. Nakajima, On quantum theory of transport phenomena: Steady diffusion, *Prog. Theor. Phys.* **20**, 948 (1958).

[18] R. Zwanzig, Ensemble method in the theory of irreversibility, *J. Chem. Phys.* **33**, 1338 (1960).

[19] G. M. Moy, J. J. Hope, and C. M. Savage, Born and Markov approximations for atom lasers, *Phys. Rev. A* **59**, 667 (1999).

[20] W. G. Unruh, Maintaining coherence in quantum computers, *Phys. Rev. A* **51**, 992 (1995).

[21] G. M. Palma, K.-A. Suominen, and A. Ekert, Quantum computers and dissipation, *Proc. R. Soc. London A* **452**, 567 (1996).

[22] B. Vacchini and H.-P. Breuer, Exact master equations for the non-Markovian decay of a qubit, *Phys. Rev. A* **81**, 042103 (2010).

[23] A. E. Teretenkov, Non-Markovian evolution of multi-level system interacting with several reservoirs: Exact and approximate, *Lobachevskii J. Math.* **40**, 1587 (2019).

- [24] J. Piilo, S. Maniscalco, K. Härkönen, and K.-A. Suominen, Non-Markovian Quantum Jumps, *Phys. Rev. Lett.* **100**, 180402 (2008).
- [25] L. Diósi and W. T. Strunz, The non-Markovian stochastic Schrödinger equation for open systems, *Phys. Lett. A* **235**, 569 (1997).
- [26] W. T. Strunz, L. Diósi, and N. Gisin, Open System Dynamics with Non-Markovian Quantum Trajectories, *Phys. Rev. Lett.* **82**, 1801 (1999).
- [27] L. Diósi, N. Gisin, and W. T. Strunz, Non-Markovian quantum state diffusion, *Phys. Rev. A* **58**, 1699 (1998).
- [28] Y. Tanimura, Nonperturbative expansion method for a quantum system coupled to a harmonic-oscillator bath, *Phys. Rev. A* **41**, 6676 (1990).
- [29] Y. Tanimura and R. Kubo, Time evolution of a quantum system in contact with a nearly Gaussian-Markovian noise bath, *J. Phys. Soc. Jpn.* **58**, 101 (1989).
- [30] A. Strathearn, P. Kirton, D. Kilda, J. Keeling, and B. W. Lovett, Efficient non-Markovian quantum dynamics using time-evolving matrix product operators, *Nat. Commun.* **9**, 3322 (2018).
- [31] F. Mascherpa, A. Smirne, A. D. Somoza, P. Fernández-Acebal, S. Donadi, D. Tamascelli, S. F. Huelga, and M. B. Plenio, Optimized auxiliary oscillators for the simulation of general open quantum systems, *Phys. Rev. A* **101**, 052108 (2020).
- [32] E. A. Polyakov and A. N. Rubtsov, Dressed quantum trajectories: Novel approach to the non-Markovian dynamics of open quantum systems on a wide timescale, *New J. Phys.* **21**, 063004 (2019).
- [33] L. Wang, Y. Fujihashi, L. Chen, and Y. Zhao, Finite-temperature time-dependent variation with multiple Davydov states, *J. Chem. Phys.* **146**, 124127 (2017).
- [34] A. W. Chin, Á. Rivas, S. F. Huelga, and M. B. Plenio, Exact mapping between system-reservoir quantum models and semi-infinite discrete chains using orthogonal polynomials, *J. Math. Phys.* **51**, 092109 (2010).
- [35] J. Prior, A. W. Chin, S. F. Huelga, and M. B. Plenio, Efficient Simulation of Strong System-Environment Interactions, *Phys. Rev. Lett.* **105**, 050404 (2010).
- [36] G. Schön and A. D. Zaikin, Quantum coherent effects, phase transitions, and the dissipative dynamics of ultra small tunnel junctions, *Phys. Rep.* **198**, 237 (1990).
- [37] P. W. Anderson, Localized magnetic states in metals, *Phys. Rev.* **124**, 41 (1961).
- [38] S. Saykin, D. Mozysky, and V. Privman, Relaxation of shallow donor electron spin due to interaction with nuclear spin bath, *Nano Lett.* **2**, 651 (2002).
- [39] R. Hanson, V. V. Dobrovitski, A. E. Feiguin, O. Gywat, and D. D. Awschalom, Coherent dynamics of a single spin interacting with an adjustable spin bath, *Science* **320**, 352 (2008).
- [40] J. Fischer and H.-P. Breuer, Correlated projection operator approach to non-Markovian dynamics in spin baths, *Phys. Rev. A* **76**, 052119 (2007).
- [41] A. Lerose, M. Sonner, and D. A. Abanin, Influence Matrix Approach to Many-Body Floquet Dynamics, *Phys. Rev. X* **11**, 021040 (2021).
- [42] M. Sonner, A. Lerose, and D. A. Abanin, Influence functional of many-body systems: Temporal entanglement and matrix-product state representation, *Annals Phys.* **435**, 168677 (2021).
- [43] V. M. Acosta, E. Bauch, M. P. Ledbetter, C. Santori, K. Fu, P. E. Barclay, R. G. Beausoleil, H. Linget, J. F. Roch, F. Treussart *et al.*, High nitrogen-vacancy density diamonds for magnetometry applications, *Phys. Rev. B* **80**, 115202 (2009).
- [44] V. V. Soshenko, O. R. Rubinas, V. V. Vorobyov, S. V. Bolshedvorskii, P. V. Kapitanova, V. N. Sorokin, and A. V. Akimov, Microwave antenna for exciting optically detected magnetic resonance in diamond nv centers, *Bull. Lebedev Phys. Inst.* **45**, 237 (2018).
- [45] J. Cerrillo and J. Cao, Non-Markovian Dynamical Maps: Numerical Processing of Open Quantum Trajectories, *Phys. Rev. Lett.* **112**, 110401 (2014).
- [46] M. Buser, J. Cerrillo, G. Schaller, and J. Cao, Initial system-environment correlations via the transfer-tensor method, *Phys. Rev. A* **96**, 062122 (2017).
- [47] R. Rosenbach, J. Cerrillo, S. F. Huelga, J. Cao, and M. B. Plenio, Efficient simulation of non-Markovian system-environment interaction, *New J. Phys.* **18**, 023035 (2016).
- [48] A. A. Kananenka, C.-Y. Hsieh, J. Cao, and E. Geva, Accurate long-time mixed quantum-classical Liouville dynamics via the transfer tensor method, *J. Phys. Chem. Lett.* **7**, 4809 (2016).
- [49] A. Gelzinis, E. Rybakovas, and L. Valkunas, Applicability of transfer tensor method for open quantum system dynamics, *J. Chem. Phys.* **147**, 234108 (2017).
- [50] F. A. Pollock and K. Modi, Tomographically reconstructed master equations for any open quantum dynamics, *Quantum* **2**, 76 (2018).
- [51] Y.-Q. Chen, K.-L. Ma, Y.-C. Zheng, J. Allcock, S. Zhang, and C.-Y. Hsieh, Non-Markovian Noise Characterization with the Transfer Tensor Method, *Phys. Rev. Appl.* **13**, 034045 (2020).
- [52] M. R. Jørgensen and F. A. Pollock, Exploiting the Causal Tensor Network Structure of Quantum Processes to Efficiently Simulate Non-Markovian Path Integrals, *Phys. Rev. Lett.* **123**, 240602 (2019).
- [53] M. R. Jørgensen and F. A. Pollock, Discrete memory kernel for multitime correlations in non-Markovian quantum processes, *Phys. Rev. A* **102**, 052206 (2020).
- [54] S. Gherardini, A. Smirne, S. Huelga, and F. Caruso, Transfer-tensor description of memory effects in open-system dynamics and multi-time statistics, *Quantum Sci. Technol.* **7**, 025005 (2022).
- [55] L. Banchi, E. Grant, A. Rocchetto, and S. Severini, Modelling non-Markovian quantum processes with recurrent neural networks, *New J. Phys.* **20**, 123030 (2018).
- [56] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, Recent advances in recurrent neural networks, [arXiv:1801.01078](https://arxiv.org/abs/1801.01078) (2017).
- [57] R. Iten, T. Metger, H. Wilming, L. del Rio, and R. Renner, Discovering Physical Concepts with Neural Networks, *Phys. Rev. Lett.* **124**, 010508 (2020).
- [58] M. Raissi, P. Perdikaris, and G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.* **378**, 686 (2019).
- [59] B. Lusch, J. N. Kutz, and S. L. Brunton, Deep learning for universal linear embeddings of nonlinear dynamics, *Nat. Commun.* **9**, 4950 (2018).
- [60] A. A. Budini, Embedding non-Markovian quantum collisional models into bipartite Markovian dynamics, *Phys. Rev. A* **88**, 032115 (2013).

- [61] I. A. Luchnikov, S. V. Vintskevich, D. A. Grigoriev, and S. N. Filippov, Machine-Learning Non-Markovian Quantum Dynamics, *Phys. Rev. Lett.* **124**, 140502 (2020).
- [62] C. Guo, K. Modi, and D. Poletti, Tensor-network-based machine learning of non-Markovian quantum processes, *Phys. Rev. A* **102**, 062414 (2020).
- [63] M. Gavish and D. L. Donoho, The optimal hard threshold for singular values is $\frac{4}{\sqrt{3}}$, *IEEE Trans. Inf. Theory* **60**, 5040 (2014).
- [64] P. J. Schmid, Dynamic mode decomposition of numerical and experimental data, *J. Fluid Mech.* **656**, 5 (2010).
- [65] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, On dynamic mode decomposition: Theory and applications, *J. Comput. Dyn.* **1**, 391 (2014).
- [66] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, Berlin, 2006).
- [67] I. A. Luchnikov, S. V. Vintskevich, H. Ouerdane, and S. N. Filippov, Simulation Complexity of Open Quantum Dynamics: Connection with Tensor Networks, *Phys. Rev. Lett.* **122**, 160401 (2019).
- [68] S. Shrapnel, F. Costa, and G. Milburn, Quantum Markovianity as a supervised learning task, *Int. J. Quantum. Inform.* **16**, 1840010 (2018).
- [69] S. Milz and K. Modi, Quantum stochastic processes and quantum non-Markovian phenomena, *PRX Quant.* **2**, 030201 (2021).
- [70] H. Arbabi and I. Mezic, Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the koopman operator, *SIAM J. Appl. Dyn. Syst.* **16**, 2096 (2017).
- [71] H. Arbabi and I. Mezić, Study of dynamics in post-transient flows using Koopman mode decomposition, *Phys. Rev. Fluids* **2**, 124402 (2017).
- [72] M. Kamb, E. Kaiser, S. L. Brunton, and J. N. Kutz, Time-delay observables for Koopman: Theory and applications, *SIAM J. Appl. Dyn. Syst.* **19**, 886 (2020).
- [73] S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, and J. N. Kutz, Chaos as an intermittently forced linear system, *Nat. Commun.* **8**, 19 (2017).
- [74] S. Pan and K. Duraisamy, Data-driven discovery of closure models, *SIAM J. Appl. Dyn. Syst.* **17**, 2381 (2018).
- [75] S. Chen and S. A. Billings, Representations of nonlinear systems: The Narmax model, *Int. J. Control* **49**, 1013 (1989).
- [76] R. Hegger, H. Kantz, and T. Schreiber, Practical implementation of nonlinear time-series methods: The TISEAN package, *Chaos* **9**, 413 (1999).
- [77] M. Hausknecht and P. Stone, Deep recurrent q-learning for partially observable MDPs, in *Proceedings of the AAAI Fall Symposium Series* (AAAI, Menlo Park, CA, 2015).
- [78] R. S. Bennink and P. Lougovski, Quantum process identification: A method for characterizing non-Markovian quantum dynamics, *New J. Phys.* **21**, 083013 (2019).
- [79] S. Campbell, F. Ciccarello, G. M. Palma, and B. Vacchini, System-environment correlations and Markovian embedding of quantum non-Markovian dynamics, *Phys. Rev. A* **98**, 012142 (2018).
- [80] S. Xue, T. Nguyen, M. R. James, A. Shabani, V. Ugrinovskii, and I. R. Petersen, Modeling for non-Markovian quantum systems, *IEEE Trans. Control Syst. Technol.* **28**, 2564 (2019).
- [81] S. Xue, M. R. James, A. Shabani, V. Ugrinovskii, and I. R. Petersen, Quantum filter for a class of non-Markovian quantum systems, in *Proceedings of the 54th IEEE Conference on Decision and Control (CDC)* (IEEE, Piscataway, NJ, 2015), pp. 7096–7100.
- [82] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, and K. Modi, Non-Markovian quantum processes: Complete framework and efficient characterization, *Phys. Rev. A* **97**, 012127 (2018).
- [83] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, and K. Modi, Operational Markov Condition for Quantum Processes, *Phys. Rev. Lett.* **120**, 040405 (2018).
- [84] I. L. Chuang and M. A. Nielsen, Prescription for experimental determination of the dynamics of a quantum black box, *J. Mod. Opt.* **44**, 2455 (1997).
- [85] J. B. Altepeter, D. Branning, E. Jeffrey, T. C. Wei, P. G. Kwiat, R. T. Thew, J. L. O’Brien, M. A. Nielsen, and A. G. White, Ancilla-Assisted Quantum Process Tomography, *Phys. Rev. Lett.* **90**, 193601 (2003).
- [86] D. H. Mahler, L. A. Rozema, A. Darabi, C. Ferrie, R. Blume-Kohout, and A. M. Steinberg, Adaptive Quantum State Tomography Improves Accuracy Quadratically, *Phys. Rev. Lett.* **111**, 183601 (2013).
- [87] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Quantum State Tomography via Compressed Sensing, *Phys. Rev. Lett.* **105**, 150401 (2010).
- [88] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Efficient quantum state tomography, *Nat. Commun.* **1**, 149 (2010).
- [89] P. Mehta, M. Bukov, C.-H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, A high-bias, low-variance introduction to machine learning for physicists, *Phys. Rep.* **810**, 1 (2019).
- [90] S. Arlot, A. Celisse *et al.*, A survey of cross-validation procedures for model selection, *Statist. Surv.* **4**, 40 (2010).
- [91] M. E. Tipping, Sparse Bayesian learning and the relevance vector machine, *J. Mach. Learn. Res.* **1**, 211 (2001).
- [92] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, Completely positive dynamical semigroups of n-level systems, *J. Math. Phys.* **17**, 821 (1976).
- [93] G. Lindblad, On the generators of quantum dynamical semigroups, *Commun. Math. Phys.* **48**, 119 (1976).
- [94] E. T. Jaynes and F. W. Cummings, Comparison of quantum and semiclassical radiation theories with application to the beam maser, *Proc. IEEE* **51**, 89 (1963).
- [95] J. M. Fink, M. Göppl, M. Baur, R. Bianchetti, P. J. Leek, A. Blais, and A. Wallraff, Climbing the Jaynes–Cummings ladder and observing its nonlinearity in a cavity QED system, *Nature (London)* **454**, 315 (2008).
- [96] N. Lambert, S. Ahmed, M. Cirio, and F. Nori, Modelling the ultra-strongly coupled spin-boson model with unphysical modes, *Nat. Commun.* **10**, 3721 (2019).
- [97] A. G. White, D. F. V. James, P. H. Eberhard, and P. G. Kwiat, Nonmaximally Entangled States: Production, Characterization, and Utilization, *Phys. Rev. Lett.* **83**, 3103 (1999).
- [98] C. F. Roos, G. P. T. Lancaster, M. Riebe, H. Häffner, W. Hänsel, S. Gulde, C. Becher, J. Eschner, F. Schmidt-Kaler, and R. Blatt, Bell States of Atoms with Ultralong Lifetimes and their Tomographic State Analysis, *Phys. Rev. Lett.* **92**, 220402 (2004).
- [99] A. J. Berkley, H. Xu, R. C. Ramos, M. A. Gubrud, F. W.

- Strauch, P. R. Johnson, J. R. Anderson, A. J. Dragt, C. J. Lobb, and F. C. Wellstood, Entangled macroscopic quantum states in two superconducting qubits, *Science* **300**, 1548 (2003).
- [100] J. Lee, M. J. Martin, Y.-Y. Jau, T. Keating, I. H. Deutsch, and G. W. Biedermann, Demonstration of the Jaynes-Cummings ladder with Rydberg-dressed atoms, *Phys. Rev. A* **95**, 041801(R) (2017).
- [101] L. Magazzù, P. Forn-Díaz, R. Belyansky, J.-L. Orgiazzi, M. A. Yurtalan, M. R. Otto, A. Lupascu, C. M. Wilson, and M. Grifoni, Probing the strongly driven spin-boson model in a superconducting quantum circuit, *Nat. Commun.* **9**, 1 (2018).
- [102] <https://github.com/LuchnikovI/Automatic-non-markovian-dynamics-learner>.
- [103] M. H. Gutknecht, A brief introduction to Krylov space methods for solving linear systems, in *Frontiers of Computational Science* (Springer, Berlin, 2007), pp. 53–62.
- [104] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith *et al.*, Array programming with NumPy, *Nature (London)* **585**, 357 (2020).