

Machine learning time-local generators of open quantum dynamicsPaolo P. Mazza¹,² Dominik Zietlow,² Federico Carollo¹, Sabine Andergassen¹,
Georg Martius,² and Igor Lesanovsky^{1,3,4}¹*Institut für Theoretische Physik and Center for Quantum Science, Universität Tübingen, Auf der Morgenstelle 14, 72076 Tübingen, Germany*²*Max Planck Institute for Intelligent Systems, Max-Planck-Ring 4, 72076 Tübingen, Germany*³*School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, United Kingdom*⁴*Centre for the Mathematics and Theoretical Physics of Quantum Non-Equilibrium Systems, University of Nottingham, Nottingham NG7 2RD, United Kingdom*

(Received 27 January 2021; accepted 18 March 2021; published 30 April 2021)

In the study of closed many-body quantum systems, one is often interested in the evolution of a subset of degrees of freedom. On many occasions it is possible to approach the problem by performing an appropriate decomposition into a bath and a system. In the simplest case the evolution of the reduced state of the system is governed by a quantum master equation with a time-independent, i.e., Markovian, generator. Such evolution is typically emerging under the assumption of a weak coupling between the system and an infinitely large bath. Here we are interested in understanding to which extent a neural network function approximator can predict open quantum dynamics—described by time-local generators—from an underlying unitary dynamics. We investigate this question using a class of spin models, which is inspired by recent experimental setups. We find that indeed time-local generators can be learned. In certain situations they are even time independent and allow to extrapolate the dynamics to unseen times. This might be useful for situations in which experiments or numerical simulations do not allow to capture long-time dynamics and for exploring thermalization occurring in closed quantum systems.

DOI: [10.1103/PhysRevResearch.3.023084](https://doi.org/10.1103/PhysRevResearch.3.023084)**I. INTRODUCTION**

The investigation of isolated quantum systems out of equilibrium is a central problem in physics [1,2]. Often, partial information that concerns a spatially localized subsystem S is sufficient to study dynamical effects related to relaxation or thermalization. This information is encoded in the so-called reduced quantum state ρ_S , [c.f. Figs. 1(a) and 1(b)], whose time dependence is obtained by “integrating out” the remainder of the evolving many-body system, which is often referred to as bath B [see Fig. 1(c)]. In certain settings, e.g., for systems that are weakly interacting with a large environment [3,4], the dynamics of the reduced state is effectively described by a quantum master equation [5–7]. In the simplest manifestation, the dynamics of the reduced quantum state ρ_S is then governed by a time-independent generator \mathcal{L} acting only on S :

$$\rho_S(t) := \text{Tr}_B(U_t(\rho_S \otimes \rho_B)U_t^\dagger) \approx e^{t\mathcal{L}}[\rho_S]. \quad (1)$$

Here, U_t is the unitary evolution operator of the full many-body system, the state $\rho_S \otimes \rho_B$ is the initial state of the many-body system in product form, and Tr_B indicates the average over the bath B . However, the generator does not need to be

time-independent and, in most generic instances [8], it may depend locally on time, \mathcal{L}_t , as illustrated in Fig. 1(c).

In this paper, we show how simple neural network architectures can learn time-local generators that govern the local dynamics of a closed quantum spin system. The input of the networks consists of the time-dependent average values of the reduced system observables from which the dynamical generator is estimated. First, we consider an architecture which provides a time-averaged generator $\bar{\mathcal{L}}$ corresponding to an effectively Markovian description of the system dynamics. Our findings indicate that this can yield indeed a valid approximation for the generator even beyond typical scenarios justifying a Markovian weak-coupling assumption [3,4]. Once the generator is known, the network can further be exploited to make predictions for times which go beyond the previously analyzed time frame, as is illustrated in Fig. 2. To study settings with a time-dependent generator, \mathcal{L}_t , we use a different neural network architecture which is based on so-called hypermodels and allows to assess the time dependence of the generator of the reduced dynamics. Our work links to recent efforts aiming at understanding quantum dynamics with neural networks [9–24]. Our approach, which uses machine learning tools and provides a directly interpretable object such as the physical generator of the dynamics, highlights a possible route for the application of neural networks in the study of the long-time dynamics of local observables in closed quantum systems. Moreover, it may find applications in the context of quantifying the degree of non-Markovianity in reduced quantum dynamics [8].

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

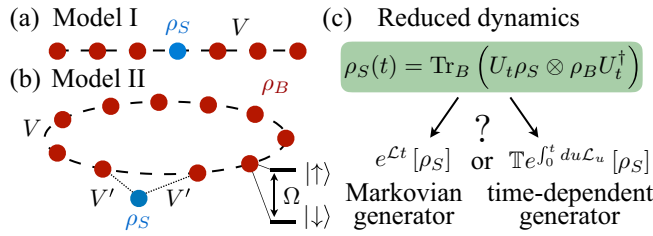


FIG. 1. Reduced local dynamics of a many-body quantum system. The reduced state ρ_S describes local degrees of freedom of a quantum spin system. The remainder of the system, B , takes the role of a “bath” with initial state ρ_B . We consider two scenarios: (a) Model I, spin chain with open boundary conditions, where the interaction strength V and range α are varied [see Eq. (2)]; (b) Model II, spin chain with closed boundaries and nearest-neighbor interaction, where the system-bath coupling V' and the initial temperature β of the bath are modified. (c) The dynamics of the reduced state, $\rho_S(t)$, is obtained by tracing out the degrees of freedom of B and to be learned by a neural network. The learned generator may be either Markovian (time independent) or time dependent.

II. MANY-BODY SPIN MODELS

For illustrating our ideas, we consider two one-dimensional systems (in the following referred to as Models I and II) consisting of N interacting spins, which are inspired by recent state-of-the-art quantum simulator experiments with Rydberg atoms [25–29] or trapped ions [30–32]. Model I has open boundary conditions [see Fig. 1(a)]. It features power-law interactions and is described by the Hamiltonian

$$H_I = \Omega \sum_{i=1}^N \sigma_i^x + V \sum_{i<j}^N \frac{n_i n_j}{|i-j|^\alpha}. \quad (2)$$

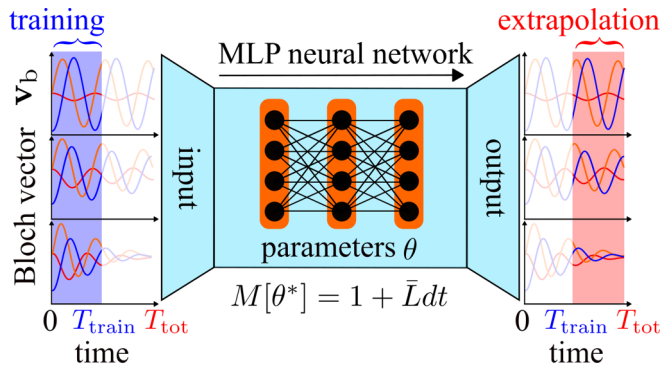


FIG. 2. Effective time-independent generator: Training and extrapolation. To obtain a time-independent generator, we use a MLP. The inputs are the time-dependent expectation values of the reduced system observables—for a single spin contained in the Bloch vector \mathbf{v}_b —in a given training time window (blue-shaded region) and for several initial Bloch vectors (top to bottom). The MLP learns the “propagator” $M[\theta^*]$, which depends on the optimized network parameters θ^* but not on the actual time. Considering a Euler integration step of length dt , this provides a matrix \bar{L} representing the action of an “averaged” generator on the Bloch vector. When the reduced dynamics is Markovian, this learned generator allows for the network to make predictions for unseen future times (red-shaded region).

Here, $n_i = (1 + \sigma_i^z)/2$ is a projector on the “up”-state of the i -th spin and $\sigma_i^{x,y,z}$ are Pauli matrices. The parameter α accounts for the interaction range, i.e., the power-law decay of the interaction potential, and V controls the overall interaction strength with respect to the strength Ω of a transverse field term. The reduced system S , for this model, is the middle spin of the chain, as shown in Fig. 1(a). Model II is a closed spin chain [illustrated in Fig. 1(b)], with nearest-neighbor interactions and Hamiltonian

$$H_{II} = \Omega \sum_{i=2}^N \sigma_i^x + V \sum_{j=2}^{N-1} n_j n_{j+1} + \Omega' \sigma_1^x + V' n_1 (n_2 + n_N). \quad (3)$$

Here, the spin with label 1 is singled out as reference spin (forming the system S) and the interaction strength with its neighbors is given by V' , while the interaction strength among all other spins is V . Moreover, we assume that the strength Ω' of the transverse field acting on the reference spin can be controlled. Both models capture a whole variety of different scenarios, including short- and long-range interactions, absence and presence of translational invariance, as well as weak and strong coupling between local and bath degrees of freedom. They moreover encompass a number of standard scenarios often explored in quantum many-body physics, such as the so-called PXP-model [33–35], the Ising model in the presence of longitudinal and transverse fields, and all-to-all connected spin systems with resemblance to the Dicke model [36] or the Lipkin-Meshkov-Glick model [37].

Our aim is to investigate the reduced dynamics of local degrees of freedom, as for instance a single spin S , as depicted in Figs. 1(a) and 1(b). To this end, we assume that the system is initialized at time $t = 0$ in the state $\rho = \rho_S \otimes \rho_B$. In our studies concerning Model I, we assume the bath to be in the infinite temperature state, $\rho_B \propto \mathbb{1}_B$. When considering Model II, instead, we assume a finite temperature situation with $\rho_B \propto e^{-\beta \tilde{H}_B}$, where the Hamiltonian \tilde{H}_B is the one of Eq. (3) with $\Omega' = V' = 0$, and β is the inverse temperature. We let the state ρ evolve according to the unitary dynamics $\rho(t) = U_t \rho U_t^\dagger$, with $U_t = e^{-iHt}$ from which we obtain the full dynamical information about the subsystem S . Throughout we will make use of the Bloch vector representation of the subsystem’s density matrix: $\rho_S(t) = [\mathbf{v}_b(t) \cdot \boldsymbol{\sigma}]/2$ with $\boldsymbol{\sigma} = (\mathbb{1}, \sigma^x, \sigma^y, \sigma^z)$. Here the Bloch vector $\mathbf{v}_b(t) = (1, \langle \sigma_k^x \rangle_t, \langle \sigma_k^y \rangle_t, \langle \sigma_k^z \rangle_t)$ is given in terms of the expectation $\langle \cdot \rangle_t$ with respect to the full quantum state at time t , and k labels the system spin [see Figs. 1(a) and 1(b)].

In the following, we will characterize the generator of this reduced dynamics; that is, we want to find a generator acting solely on S that reproduces (or approximates) the dynamics of $\mathbf{v}_b(t)$ for the two models under consideration [Figs. 1(a) and 1(b)]. For Model I we explore for which parameter combinations (V, α) , the local generator, which is in principle time-dependent \mathcal{L}_t , can be approximated by an effective “averaged” time-independent one, $\mathcal{L}_t \approx \bar{\mathcal{L}}$. We remark that, while B acts as a fictitious bath for subsystem S , the situation described here is far from that of typical weak-coupling limits. Indeed, B is a finite quantum system and its Hamiltonian has a discrete spectrum. As such, B can hardly be thought of as an infinite Markovian bath of bosonic oscillators, whose state is unaffected by the interaction with S . The only aspect in

common with a weak-coupling limit is that the interaction strength can be made small. However, this is not sufficient to argue that the local dynamics may be described by time-independent generators.

For Model II we consider the case in which an “isolated” spin (on site 1) interacts with B with an independently tunable strength V' . Experimentally such a situation can be realized, e.g., by exploiting the distance dependence of dipolar interactions of atoms in Rydberg states [28]. For Model II we also investigate how well the generator of the reduced dynamics is approximated by a time-independent model when the initial state of B (parametrized by the inverse temperature β) and the interaction strength V' are varied. In addition, for this scenario, we use a hypermodel in order to analyze regimes that necessitate a time-dependent generator \mathcal{L}_t .

III. NETWORK ARCHITECTURES, TRAINING, AND RESULTS

In order to learn dynamical generators for the reduced dynamics of subsystem S , we employ machine learning algorithms. Our approach is completely data-driven, i.e., the network has no prior information about the physical system. For learning time-independent generators, we use a linear multilayer perceptron (MLP) architecture (see Fig. 2) which turns out to be the simplest possible one. Every data point of our data-set \mathcal{D} contains the triple $(\mathbf{v}_b, \mathbf{v}'_b, t)$, where $\mathbf{v}_b := \mathbf{v}_b(t)$ is the value of the Bloch vector at a given time-point t , and \mathbf{v}'_b is the value of the Bloch vector after a further discrete time step dt , i.e., $\mathbf{v}'_b := \mathbf{v}_b(t + dt)$.

The input of our model is \mathbf{v}_b , and the output is the vector $\mathbf{o}(t) = M[\theta]\mathbf{v}_b$, where $M[\theta]$ is a 4×4 matrix (matching the dimension of the Bloch vector) that depends on the parameters θ of the network. To optimize the network, we introduce the following loss function, which is given by the norm of the difference between the next time step \mathbf{v}'_b and the output of the network:

$$C(\theta) = \mathbb{E}_{(\mathbf{v}_b, \mathbf{v}'_b) \sim \mathcal{D}} [\|M[\theta]\mathbf{v}_b - \mathbf{v}'_b\|], \quad (4)$$

where the expectation is taken over the training data \mathcal{D} . Minimizing the above function thus provides a model $M[\theta^*]$ for the propagation of the Bloch vector over an infinitesimal time step dt . This model is related to the generator of the local dynamics; indeed, we have $M[\theta^*] = \mathbb{1} + \bar{L}dt$, where \bar{L} is a representation of the time-averaged generator \bar{L} acting on Bloch vectors. The training data set is constructed by evolving the Bloch vector \mathbf{v}_b over a time interval $[0, T_{\text{train}}]$. The trajectories are generated evolving 100 randomly chosen initial states of the form

$$\rho_0 = \frac{1}{2} \begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix} \otimes \rho_B, \quad (5)$$

with $x^2 + y^2 + z^2 < 1$, i.e., we consider nonpure initial states. We split our data set such that we consider the 80% of each trajectory for training, i.e., finding the best parameters of our model. The remaining 20% form the validation set to assess the performance of the model built in the training phase. We anticipate already here that, after having trained and validated the model as discussed, we will then test its capability of predicting the time-evolution of the reduced system by choosing

initial states which are diagonal [see Eq. (7) below]. This is not at all a limitation for assessing the performance of the time-independent generator since even though we start from a diagonal matrix, the dynamics will quickly generate off-diagonal terms.

To encode and quantify the time dependence of dynamical generators, \mathcal{L}_t , we use a different architecture and consider a so-called hypermodel, which computes network weights based on a context input. In our case, the context input is the time t and a MLP F with nonlinear activation functions transforms this input to the 4×4 matrix of the generator $M[F(t; \theta)]$.

The data set is organized in the same way as for the time-independent model. The only difference with respect to the previous case is that we also pass the information about the actual time step t . The network parameters θ are optimized by minimizing the loss function

$$C_H(\theta) = \mathbb{E}_{(\mathbf{v}_b, \mathbf{v}'_b, t) \sim \mathcal{D}} [\|M[F(t; \theta)]\mathbf{v}_b - \mathbf{v}'_b\|]. \quad (6)$$

By training the model in this way we obtain a time-dependent representation of the propagator of the local dynamics, i.e., $M[F(t; \theta^*)] = \mathbb{1} + L_t dt$. For the training set, we choose quantum states as in the previous case. Since in this case the model approximates a time-dependent generator, we need to pass to the network information about all the time points of the different trajectories. Therefore, we cannot split the time-series as we have done previously. To test our model, instead, we use unseen trajectories that we obtain by starting from initial conditions that are well outside of the training set. For the test data set, we indeed consider trajectories generated by starting from the initial state

$$\rho_0 = \begin{pmatrix} 1-c & 0 \\ 0 & c \end{pmatrix} \otimes \rho_B, \quad (7)$$

with c being a random number, $c \in (0.01, 0.7)$ (see Appendix for details). This choice of the class of initial states is completely arbitrary. However, we do not expect such choice to play a relevant role in the approximation of the time-dependent generator.

A. Time-independent generators

With the above numerical approach, we can now investigate under which circumstances the generator of the dynamics of the subsystem S is time independent. We consider first the open spin chain [see Fig. 1(a) and Eq. (2)] and explore different values of V and α . To test the performance of the model after the training, we study the *time-averaged* norm difference between the exact Bloch vector \mathbf{v}_b^{ex} —obtained by simulating the full many-body spin chain dynamics—and the time-evolution of the same quantity as predicted by our model $\mathbf{v}_b^{\text{mod}}$:

$$\epsilon = \frac{1}{T_{\text{tot}}} \int_0^{T_{\text{tot}}} dt \|\mathbf{v}_b^{\text{mod}}(t) - \mathbf{v}_b^{\text{ex}}(t)\|, \quad (8)$$

computed by numerical integration.

As shown in Fig. 3(a), the error ϵ is small for short-range (large α) and weakly interacting (with small ratio V/Ω) spin chains. This means that our time-independent model captures the relevant features of the reduced dynamics, and suggests that for finite systems with sufficiently weak interactions, the dynamics of local degrees of freedom can indeed be

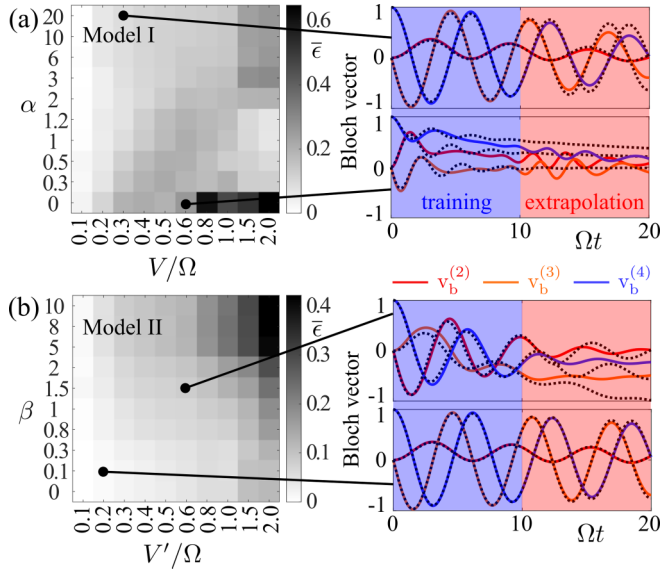


FIG. 3. Time-independent generator. (a) Model I: Comparison between the exact dynamics of the Bloch vector and the one obtained by learning an effective time-independent generator with neural networks. The density plot shows the time-averaged norm difference, $\bar{\epsilon}$ [Eq. (8) averaged over 5 randomly selected values c of the initial conditions; see Eq. (7)], for various parameter choices (V , α). The data are calculated for $N = 9$ spins with a training time of $T_{\text{train}} = 10/\Omega$, $dt = 0.01/\Omega$, and a total time $T_{\text{tot}} = 20/\Omega$ (see main text for details). We also show two examples for the Bloch vector evolution, where we choose $c = 0$ and $T_{\text{tot}} = 20/\Omega$ for illustration. Here solid curves correspond to the components of the Bloch vector propagated with the learned time-averaged generator. The dashed lines correspond to the numerically exact solution. (b) Same as (a), but for Model II and the parameter set (V' , β).

effectively described by a *time-independent* generator $\bar{\mathcal{L}}$. Furthermore, for this Markovian regime of the dynamics, our model allows to make predictions for times exceeding the training time.

We now turn to the discussion of Model II [see Fig. 1(b) and Eq. (3)]. Using the same procedure explained above, we can learn the time-averaged dynamical generator for subsystem S . In Fig. 3(b) we summarize the results of this analysis. We display the time-averaged error, as defined in Eq. (8), for different parameter choices, which in this case are the (inverse) temperature β of the initial state ρ_B and the interaction strength V' . We furthermore take $\Omega' = \Omega$. As shown in the figure, the dynamics is well approximated by a time-independent generator, $\bar{\mathcal{L}}$, for weak interactions V' and for small β . Here also the extrapolation to times that exceed T_{train} is possible. The data in Fig. 3 show that in certain parameter regimes the approximation of the dynamics through a time-independent generator does not work well. Here significant deviations are even observable during the training period.

B. Time-dependent time-local generators

We are now interested in analyzing how strongly the dynamical generator depends on time in those instances where the model introduced in the previous section fails. To this end, we adopt a hypermodel which can take as input the informa-

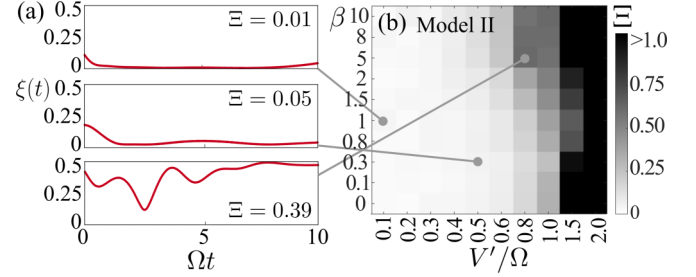


FIG. 4. Hypermodel and time-dependent generator. Hypermodels allow to encode time-dependent generators \mathcal{L}_t . The time dependence is quantified through the derivative of L_t with respect to time, expressed by the positive quantity $\xi(t)$ (see text for definition). (a) Three example curves corresponding to Model II ($N = 9$). The function $\xi(t)$ depends only on the parameters of the model and is independent from the initial condition chosen. Generally, the stronger the interaction V' , the more pronounced the time dependence. This trend is clearly visible in (b), where we show the averaged value of $\xi(t)$ in the interval 0 to $T = 10/\Omega$ ($dt = 0.01/\Omega$), denoted as Ξ .

tion about the running time. In this way, the neural network is able to learn an optimal parametrization of the “propagator” $M[F(t; \theta^*)] = \mathbb{1} + L_t dt$ for the Bloch vector dynamics which explicitly depends on time. Here, the matrix L_t encodes the action of the generator on the Bloch vector and, essentially, implements the differential equations obeyed by the entries of the Bloch vector \mathbf{v}_b . This architecture allows the ability to learn and accurately reproduce the dynamics of Model II for all studied parameter regimes (see Appendix for examples). This is not surprising, but gives us a handle for analyzing the time dependence of the dynamical generator L_t . To this end we consider the positive quantity $\xi(t) = \sqrt{\text{Tr}(\dot{L}_t^\dagger \dot{L}_t)}$, where the dot denotes the derivative with respect to time. This can only be zero when $\dot{L}_t = 0$, i.e., when the action of the generator is not depending on the running time. To quantify an overall time dependence within a time window $[0, T]$, we define the time-averaged value $\Xi = \frac{1}{T} \int_0^T dt \xi(t)$. In Fig. 4 we show the corresponding data. For strong interactions V' , $\xi(t)$ is nonzero throughout which indicates a strong explicit time dependence. This is also reflected in a large Ξ . For weak interactions, on the other hand, $\xi(t)$ remains small for all times considered. The oscillation we attribute to the finite size of the system. This confirms that, in such parameter regime, it is indeed possible to approximate the dynamics of the Bloch vector by means of a time-independent matrix, as considered in the previous section.

IV. CONCLUSIONS

We have presented two simple examples of neural network architectures that can learn the dynamical features of reduced quantum states. When such time evolution is effectively Markovian, the network can find a suitable approximation for the generator of the local dynamics. Here one can extrapolate the dynamics of reduced degrees of freedom to times that were unexplored during the training procedure. This possibility is particularly promising for applications in combination with tensor networks, which perform extremely well for short times. A neural network could learn the time-

independent generator during this time interval and then possibly extrapolate to long times. In principle, this might enable the exploration of the onset of stationary or thermalization regimes. When the dynamics is not of Markovian type, we have shown that hypermodels can recover the generator of the reduced quantum dynamics. This allows to quantify non-Markovian effects [38], which manifest in an explicit time dependence of the dynamical generator Ref. [8]. Furthermore, being able to reconstruct the generator of the reduced dynamics, as we have done here with hypermodels, makes it possible to explore different measures of non-Markovianity based on nondivisibility criteria for quantum dynamical time evolutions [39–42].

A possible future development in this regard could be the application of more advanced machine learning methods for learning time correlations in time series. This can be achieved with models borrowed from language studies {e.g., long short-term memory recurrent neural networks (LSTM) [43] and transformers [44]} or with algorithms geared toward more interpretable models by either learning analytical expressions of the differential equation [45] or modeling the time-evolution dynamics directly [46].

ACKNOWLEDGMENTS

We acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy EXC-Number 2064/1–Project number 390727645. I.L. acknowledges funding from the “Wissenschaftler Rückkehrprogramm GSO/CZS” of the Carl-Zeiss-Stiftung and the German Scholars Organization e.V. and through the DFG projects number 428276754 (SPP GiRyd) and 435696605. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting D.Z.

APPENDIX

In this Appendix, we report a more detailed analysis concerning the effectiveness of the hypermodel in predicting the

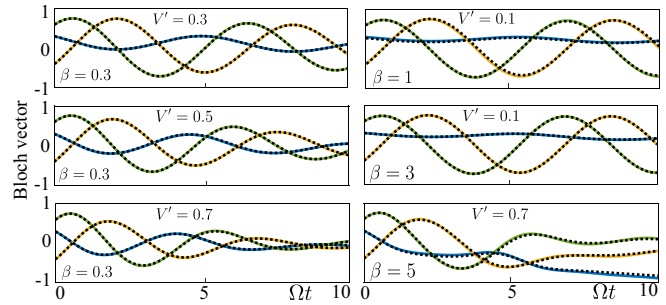


FIG. 5. Hypermodel accuracy: Comparison between the exact dynamics and the dynamics predicted by the hypermodel for initial state ρ_0 reported in Eq. (5) and for $N = 7$. As it can be observed the agreement between the model prediction (dotted line) and the exact dynamics (full line) is excellent. This fact enables us to use the local operator learned using the hypermodel for studying the time-dependence of L_t .

dynamics of the Bloch vector within the training time window. In Table I, we provide instead the hyperparameters of the used models.

Hypermodel vs exact dynamics

In the main text we analyze the time-dependence of time-local generators parametrizing the “propagator” $M[F(t; \theta^*)] = \mathbb{1} + dtL_t$ for the Bloch vector, using a hypermodel architecture. The matrix L_t encodes all the information concerning the dynamics of local (one-site) observables. As stated in the main text, a suitable measure to quantify the dependence on time of this matrix is given by $\Xi = \frac{1}{T} \int_0^T dt \xi(t)$, where $\xi(t) = \sqrt{\text{Tr}(L_t^\dagger L_t)}$. However, this can be a faithful measure of the time-dependence of the physical generator, only if L_t correctly implements the dynamics of the Bloch vector. It is thus important to check that the prediction of the hypermodel matches the exact dynamics, in the time window $[0, T]$ in which we want to analyze the generator L_t . To this end, we report in Fig. 5 a detailed comparison between the

TABLE I. Network architectures. Overview over the architecture, optimizers, training, and system parameters for Model I and Model II (top) and the hypermodel (bottom).

Setting	Architecture	Optimizer	Dataset
Time-independent	Linear perceptron (no hidden layers)	Adam Learning rate: 10^{-3} Betas: 0.9, 0.999 Epsilon: 10^{-8} Batch size: 256 Batches per epoch: 512 Epochs: 5	Trajectories (train/val): 100/20 Total time (ΩT): 10 Time step (Ωdt): 0.01
Time-dependent (hyper-model)	Hidden layers: 3 Nonlinearity: <i>tanh</i> Output nonlinearity: <i>none</i>	Adam Learning rate: 10^{-3} Betas: 0.9, 0.999 Epsilon: 10^{-8} Batch size: 256 Batches per epoch: 256 Epochs: 500	Trajectories (train/val): 100/20 Total time (ΩT): 10 Time step (Ωdt): 0.01

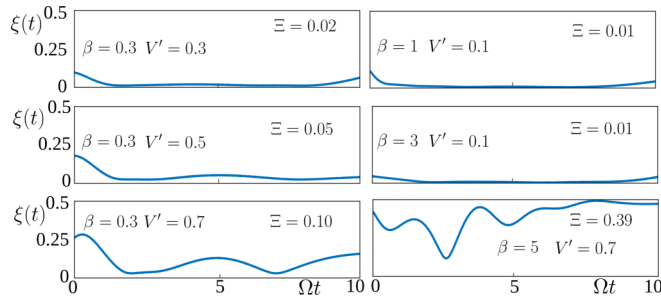


FIG. 6. **Time-dependence:** Behavior of the quantity $\xi(t)$ for the same parameter regimes of Fig. 5. The quantity Ξ grows with the interaction, meaning that the generator of the local dynamics L_t becomes more time-dependent.

predicted time evolution of the Bloch vector components and the exact ones for six different parameter regimes. We start from a generic initial state with form given in Eq. (5) with $x^2 + y^2 + z^2 < 1$, in particular we choose $x = 0.6$, $y = 0.3$ and $z = 0.4$, and we consider different interactions and initial bath temperatures. In the training time-window $[0, T]$, we

observe an excellent agreement between the two curves for $T = 10/\Omega$. In Fig. 6 we report the time evolution of the quantity $\xi(t)$ for the same parameter regimes. As it can be observed for weakly interacting systems the generator L_t is almost time-independent. In this regime, as discussed in the main text, it is possible to predict the local dynamics using simple time-independent architectures. Increasing the interaction, the parameter Ξ grows accordingly signaling that the dependence on time of L_t becomes stronger.

The training procedure for the hyper-model is slightly different with respect what we have done for the time-independent case, in particular it is different the differentiation between training and evaluation set. In this case, in fact, we generate two different types of data-sets, the first data-set for training consists in 100 trajectories with randomly chosen initial state ρ_0 of the the same form reported in Eq. (5). For evaluation we select, instead, randomly data-points from trajectories generated from initial states of type given in Eq. (7), with $0.01 < c < 0.7$. In this way we are performing a separation of the data-sets in the Hilbert space. This is contrary to what we had for the time-independent case where the data-set was given by separated time-intervals of the same trajectory.

-
- [1] L. D’Alessio, Y. Kafri, A. Polkovnikov, and M. Rigol, From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics, *Adv. Phys.* **65**, 239 (2016).
- [2] A. Polkovnikov, K. Sengupta, A. Silva, and M. Vengalattore, Nonequilibrium dynamics of closed interacting quantum systems, *Rev. Mod. Phys.* **83**, 863 (2011).
- [3] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press on Demand, 2002).
- [4] C. Gardiner and P. Zoller, *Quantum Noise: A Handbook of Markovian and non-Markovian Quantum Stochastic Methods with Applications to Quantum Optics* (Springer Science & Business Media, 2004).
- [5] G. Lindblad, On the generators of quantum dynamical semigroups, *Commun. Math. Phys.* **48**, 119 (1976).
- [6] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, Completely positive dynamical semigroups of n-level systems, *J. Math. Phys.* **17**, 821 (1976).
- [7] R. Alicki and K. Lendi, *Quantum Dynamical Semigroups and Applications*, Vol. 717 (Springer, 2007).
- [8] D. Chruściński and A. Kossakowski, Non-Markovian Quantum Dynamics: Local Versus Nonlocal, *Phys. Rev. Lett.* **104**, 070406 (2010).
- [9] N. Yoshioka and R. Hamazaki, Constructing neural stationary states for open quantum many-body systems, *Phys. Rev. B* **99**, 214306 (2019).
- [10] M. J. Hartmann and G. Carleo, Neural-Network Approach to Dissipative Quantum Many-Body Dynamics, *Phys. Rev. Lett.* **122**, 250502 (2019).
- [11] A. Nagy and V. Savona, Variational Quantum Monte Carlo Method with a Neural-Network Ansatz for Open Quantum Systems, *Phys. Rev. Lett.* **122**, 250501 (2019).
- [12] S. Krastanov, K. Head-Marsden, S. Zhou, S. T. Flammia, L. Jiang, and P. Narang, Unboxing quantum black box models: Learning non-Markovian dynamics, [arXiv:2009.03902](https://arxiv.org/abs/2009.03902).
- [13] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [14] S. Krastanov, S. Zhou, S. T. Flammia, and L. Jiang, Stochastic estimation of dynamical variables, *Quantum Sci. Technol.* **4**, 035003 (2019).
- [15] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [16] Z. Liu, L.-M. Duan, and D.-L. Deng, Solving quantum master equations with deep quantum neural networks, [arXiv:2008.05488](https://arxiv.org/abs/2008.05488).
- [17] 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).
- [18] T. Fösel, P. Tighineanu, T. Weiss, and F. Marquardt, Reinforcement Learning with Neural Networks for Quantum Feedback, *Phys. Rev. X* **8**, 031084 (2018).
- [19] S. Martina, S. Gherardini, and F. Caruso, Machine learning approach for quantum non-Markovian noise classification, [arXiv:2101.03221](https://arxiv.org/abs/2101.03221).
- [20] C. Miles, A. Bohrdt, R. Wu, C. Chiu, M. Xu, G. Ji, M. Greiner, K. Q. Weinberger, E. Demler, and E.-A. Kim, Correlator Convolutional Neural Networks: An interpretable architecture for image-like quantum matter data, [arXiv:2011.03474](https://arxiv.org/abs/2011.03474).
- [21] H. Buraou and M. Heyl, Unitary long-time evolution with quantum renormalization groups and artificial neural networks, [arXiv:2009.04473](https://arxiv.org/abs/2009.04473).
- [22] D. Luo, Z. Chen, J. Carrasquilla, and B. K. Clark, Autoregressive neural network for simulating open quantum systems via a probabilistic formulation, [arXiv:2009.05580](https://arxiv.org/abs/2009.05580).
- [23] S. Barison, F. Vicentini, and G. Carleo, An efficient quantum algorithm for the time evolution of parameterized circuits, [arXiv:2101.04579](https://arxiv.org/abs/2101.04579).

- [24] J. Reyes, S. Dhara, and E. R. Mucciolo, Machine learning regression for operator dynamics, [arXiv:2102.11868](#).
- [25] I. Bloch, J. Dalibard, and S. Nascimbène, Quantum simulations with ultracold quantum gases, *Nat. Phys.* **8**, 267 (2012).
- [26] H. Kim, Y. Park, K. Kim, H.-S. Sim, and J. Ahn, Detailed Balance of Thermalization Dynamics in Rydberg-Atom Quantum Simulators, *Phys. Rev. Lett.* **120**, 180502 (2018).
- [27] S. Ebadi, T. T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, W. W. Ho *et al.*, Quantum phases of matter on a 256-atom programmable quantum simulator, [arXiv:2012.12281](#).
- [28] A. Browaeys and T. Lahaye, Many-body physics with individually controlled rydberg atoms, *Nat. Phys.* **16**, 132 (2020).
- [29] I. Bloch, J. Dalibard, and W. Zwerger, Many-body physics with ultracold gases, *Rev. Mod. Phys.* **80**, 885 (2008).
- [30] R. Islam, E. E. Edwards, K. Kim, S. Korenblit, C. Noh, H. Carmichael, G.-D. Lin, L.-M. Duan, C.-C. J. Wang, J. K. Freericks *et al.*, Onset of a quantum phase transition with a trapped ion quantum simulator, *Nat. Commun.* **2**, 377 (2011).
- [31] J. G. Bohnet, B. C. Sawyer, J. W. Britton, M. L. Wall, A. M. Rey, M. Foss-Feig, and J. J. Bollinger, Quantum spin dynamics and entanglement generation with hundreds of trapped ions, *Science* **352**, 1297 (2016).
- [32] M. K. Joshi, A. Elben, B. Vermersch, T. Brydges, C. Maier, P. Zoller, R. Blatt, and C. F. Roos, Quantum Information Scrambling in a Trapped-Ion Quantum Simulator with Tunable Range Interactions, *Phys. Rev. Lett.* **124**, 240505 (2020).
- [33] B. Sun and F. Robicheaux, Numerical study of two-body correlation in a 1d lattice with perfect blockade, *New J. Phys.* **10**, 045032 (2008).
- [34] C. Ates, J. P. Garrahan, and I. Lesanovsky, Thermalization of a Strongly Interacting Closed Spin System: From Coherent Many-Body Dynamics to a Fokker-Planck Equation, *Phys. Rev. Lett.* **108**, 110603 (2012).
- [35] C. Turner, A. Michailidis, D. Abanin, M. Serbyn, and Z. Papić, Weak ergodicity breaking from quantum many-body scars, *Nat. Phys.* **14**, 745 (2018).
- [36] C. Emary and T. Brandes, Chaos and the quantum phase transition in the dicke model, *Phys. Rev. E* **67**, 066203 (2003).
- [37] S. Dusuel and J. Vidal, Finite-Size Scaling Exponents of the Lipkin-Meshkov-Glick Model, *Phys. Rev. Lett.* **93**, 237204 (2004).
- [38] F. Benatti, F. Carollo, R. Floreanini, and H. Narnhofer, Non-Markovian mesoscopic dissipative dynamics of open quantum spin chains, *Phys. Lett. A* **380**, 381 (2016).
- [39] H.-P. Breuer, E.-M. Laine, and J. Piilo, Measure for the Degree of Non-Markovian Behavior of Quantum Processes in Open Systems, *Phys. Rev. Lett.* **103**, 210401 (2009).
- [40] E.-M. Laine, J. Piilo, and H.-P. Breuer, Measure for the non-markovianity of quantum processes, *Phys. Rev. A* **81**, 062115 (2010).
- [41] H.-P. Breuer, Foundations and measures of quantum non-markovianity, *J. Phys. B* **45**, 154001 (2012).
- [42] Á. Rivas, S. F. Huelga, and M. B. Plenio, Quantum non-markovianity: characterization, quantification and detection, *Rep. Prog. Phys.* **77**, 094001 (2014).
- [43] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Comput.* **9**, 1735 (1997).
- [44] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need, [arXiv:1706.03762](#).
- [45] S. S. Sahoo, C. H. Lampert, and G. Martius, Learning equations for extrapolation and control, in *Proceedings of the 35th International Conference on Machine Learning (ICML 2018, Stockholm, Sweden, 2018)* (PMLR, 2018), Vol. 80, pp. 4442–4450.
- [46] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, Neural ordinary differential equations, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2018), Vol. 31, pp. 6571–6583.