Data-Driven Learning of the Generalized Langevin Equation with State-Dependent Memory

Pei Ge,[1] Zhongqiang Zhang,[2] and Huan Lei[1,3,*]
[1]Department of Computational Mathematics, Science, and Engineering, Michigan State University, East Lansing, Michigan 48824, USA
[2]Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, Massachusetts 01609, USA
[3]Department of Statistics and Probability, Michigan State University, East Lansing, Michigan 48824, USA

(Received 27 October 2023; revised 27 February 2024; accepted 12 July 2024; published 13 August 2024)

We present a data-driven method to learn stochastic reduced models of complex systems that retain a state-dependent memory beyond the standard generalized Langevin equation with a homogeneous kernel. The constructed model naturally encodes the heterogeneous energy dissipation by jointly learning a set of state features and the non-Markovian coupling among the features. Numerical results demonstrate the limitation of the standard generalized Langevin equation and the essential role of the broadly overlooked state-dependency nature in predicting molecule kinetics related to conformation relaxation and transition.

DOI: 10.1103/PhysRevLett.133.077301

Predicting the collective behavior of complex multiscale systems is often centered around projecting the full-dimensional dynamics onto a set of resolved variables. However, an accurate construction of such a reduced model remains a practical challenge for real applications such as molecular modeling. While model reduction frameworks such as the Koopman operator [1] and the Mori-Zwanzig projection formalism [2,3] enable us to write down the dynamic equations in terms of the resolved variables, the reduced model generally becomes non-Markovian with a memory term that may further depend on the resolved variables; the direct numerical evaluation involves solving the expensive full-dimensional orthogonal dynamics. In practice, one common approximation is to ignore such state dependency; the reduced model is simplified as the standard generalized Langevin equation (GLE) [4] with a memory kernel that only depends on time. Several approaches [5–20] have been developed to construct the memory kernel such that certain dynamic properties (e.g., the two-point correlations) can be properly reproduced. Despite its broad application, the validity of the standard GLE for real multiscale systems remains less understood [21,22].

Intuitively, the above model reduction problem is somewhat analogous to hiking on a mountain where the landscape map and the path roughness represent the free energy and the memory term, respectively. In general, we should not expect homogeneous path roughness at the different locations (e.g., the valleys and the ridges), which, conversely, needs to be inferred from the hiking records. Indeed, studies based on full molecular dynamics (MD) simulations [23–33] and sophisticated projection operator construction [34–42] show that the extracted memory term can exhibit a pronounced state-dependent nature, where the implications for the collective behaviors remain underexplored. For extensive MD systems, a recent study [43] on reduced modeling of polymer melt shows that the heterogeneous intermolecular energy dissipation (i.e., the memory) can be crucial for transport on the hydrodynamic scale. However, for canonical nonextensive problems such as biomolecule systems, a quantitative understanding of the state-dependent memory effect on the reduced dynamics remains an open problem. Several recent works [10,13,18,44–48] model the non-Markovian effect for transition dynamics based on the standard GLE. While elegant semianalytical studies [49–58] on idealized 1D double-well potential provide theoretical insights into the state-dependent nature, quantitative modeling that retains the reduced dynamics consistent with the full MD model, including collective properties such as transition and conformation relaxation, relies on accurate construction and efficient simulation of a reduced model beyond the standard GLE.

This Letter presents a data-driven approach for learning a new stochastic reduced model that retains a state-dependent memory for nonextensive systems. Instead of dealing with the orthogonal dynamics [6,40,43], the training only relies on the trajectory samples and does not directly solve the Mori-Zwanzig projection formalism. The main idea is to seek a generalized representation of the memory as the composition of a set of state-dependent features, which encodes the coupling between the resolved and unresolved variables and will be learned using three-point correlation functions. Efficient training is achieved by constructing the encoders using a set of sparse bases, whose correlations can be efficiently precomputed. The time-dependent

*Contact author: leihuan@msu.edu

component is directly learned in the Fourier space that enables the efficient evaluation of the convolution term via the FFT and meanwhile ensures non-negative energy dissipation (i.e., model stability). To simulate the model, coherent noise can be introduced that strictly satisfies the second fluctuation-dissipation theorem and retains a consistent invariant distribution. The present model, with a new memory form, essentially reveals a caveat in model reduction of multiscale systems and provides a reliable approach for simulating the stochastic reduced dynamics beyond empirical models. It enables us to probe open problems such as the effect of state-dependent memory on molecular kinetics. Numerical results show that the broadly overlooked state dependency can play a crucial role. In particular, the standard GLE is insufficient to capture the collective properties such as conformation relaxation and transition rate distribution, which, fortunately, can be reproduced by the present model.

Let $(\mathbf{q},\mathbf{p})\in\mathbb{R}^{2m}$ represent the resolved variables of a high-dimensional Hamiltonian system, where $\mathbf{q}$ denotes the coarse-grained coordinates as a function of the position variables of the full model, and $\mathbf{p}$ denotes the coarse-grained momenta. Following the Zwanzig's formalism [4,59], the reduced dynamics takes the form

$$\begin{aligned}\dot{\mathbf{q}} &= \mathbf{M}^{-1}\mathbf{p},\\ \dot{\mathbf{p}} &= -\nabla U(\mathbf{q}) - \int_0^t \mathbf{K}[\mathbf{q}(\tau), t-\tau]\mathbf{v}(\tau)\mathrm{d}\tau + \boldsymbol{\mathcal{R}}_t, \quad (1)\end{aligned}$$

where $\mathbf{M}$ is the mass matrix, $U(\mathbf{q})$ is the free energy, $\mathbf{v}\coloneqq\dot{\mathbf{q}}$ is the velocity, $\mathbf{K}(\mathbf{q},t)$ is the memory, and $\boldsymbol{\mathcal{R}}_t$ is the noise whose covariance function is related to the memory following the second fluctuation-dissipation theorem [40]. Before proceeding to the construction of $\mathbf{K}(\mathbf{q},t)$, we note that the rigorous form based on Zwanzig's formalism depends on both $\mathbf{q}$ and $\mathbf{p}$. Here, we focus on the state dependence on $\mathbf{q}$ and assume it is independent of $\mathbf{p}$ (see Ref. [59]). Furthermore, $\mathbf{M}$ generally depends on $\mathbf{q}$; the current choice of $\mathbf{q}$ leads to a constant mass matrix (see Refs. [13,41] and Supplemental Material (SM) [60]). Also, the construction of the free energy $U(\mathbf{q})$ can be nontrivial; several canonical methods based on enhanced sampling [69–72] and temperature acceleration [73–76] have been developed to facilitate the phase space exploration. We assume the phase space can be effectively explored and $U(\mathbf{q})$ is known *a priori*.

Instead of rigorously constructing $\mathbf{K}(\mathbf{q},t)$ from the full model, we ask the question of which forms of $\mathbf{K}$ can generate a memory effect. One common approach is to embed the memory in a larger Markovian dynamics with a set of auxiliary variables. An essential observation is that the memory term can be generally written as

$$\mathbf{K}(\mathbf{q}(\tau), t-\tau) \approx \mathcal{C}^+ \circ \exp((t-\tau)\mathcal{L}_{\mathrm{aux}}) \circ \mathcal{C}^-, \quad (2)$$

where $\mathcal{L}_{\mathrm{aux}}$ is the Liouville operator corresponding to the auxiliary dynamics and $\mathcal{C}^{\pm}$ are channels representing the coupling of the resolved and auxiliary variables. As a special case, if the coupling and the auxiliary dynamics take a linear form, the embedded memory recovers the standard GLE kernel, i.e., $\mathbf{K}(\mathbf{q},t)=\mathbf{K}(t)$ (e.g., see Refs. [19,77]). Therefore, to construct the reduced model beyond the standard GLE, the coupling channels need to properly retain certain kinds of state-dependency nature. This motivates us to represent $\mathcal{C}^{\pm}$ by seeking a set of state-dependent features $\boldsymbol{\phi}(\mathbf{q})=[\phi_1(\mathbf{q}),\ldots,\phi_n(\mathbf{q})]$, where $\boldsymbol{\phi}\colon \mathbb{R}^m\to\mathbb{R}^{n\times m}$ essentially encode the nonlinear coupling between the resolved and unresolved variables and the detailed form will be specified later. $\exp(t\mathcal{L}_{\mathrm{aux}})$ induces the non-Markovian interactions among the features with a time lag t characterized by a kernel function, i.e., $\mathcal{C}^+\circ\exp((t-\tau)\mathcal{L}_{\mathrm{aux}})\circ\mathcal{C}^- = \boldsymbol{\phi}(\mathbf{q}(t))^T\Theta(t-\tau)\boldsymbol{\phi}(\mathbf{q}(\tau))$, where $\Theta\colon \mathbb{R}^+\to\mathbb{R}^{n\times n}$ and component $\Theta_{ij}(t-\tau)$ represents the dissipation between features $\phi_i(\mathbf{q}(t))$ and $\phi_j(\mathbf{q}(\tau))$. In the remainder of this Letter, we use $\boldsymbol{\phi}_t$ to denote $\boldsymbol{\phi}(\mathbf{q}(t))$.

With the above observation, we propose the following form to model the reduced dynamics (1), i.e.,

$$\begin{aligned}\dot{\mathbf{q}} &= \mathbf{M}^{-1}\mathbf{p},\\ \dot{\mathbf{p}} &= -\nabla U(\mathbf{q}) - \int_0^t \boldsymbol{\phi}_t^T\Theta(t-\tau)\boldsymbol{\phi}_\tau\mathbf{v}(\tau)\mathrm{d}\tau + \boldsymbol{\mathcal{R}}_t, \quad (3)\end{aligned}$$

where encoders $\{\phi_i(\mathbf{q})\}_{i=1}^n$ and kernel $\Theta(t)$ need to be determined. As a special case, at the Markovian limit $\Theta(t)\propto\delta(t)$, Eq. (3) recovers the Langevin dynamics and the quadratic form $\boldsymbol{\phi}^T\boldsymbol{\phi}$ ensures positive energy dissipation. Also, by choosing $\Theta(t)$ to be diagonal with individual components corresponding to certain frequency modes, Eq. (3) reduces to the heat bath model [35] with a nonlinear coupling of bath coordinates. On the other hand, the present model enables an adaptive choice of the number of spatial features and a more general form of $\Theta(t)$ with the off-diagonal components capturing the non-Markovian coupling among the features, which turns out to be crucial for reproducing the collective dynamics (see SM [60]).

We emphasize that Eq. (3) should not be viewed as a direct approximation of Zwanzig's projection formalism. Rather, it serves as a reduced model that faithfully retains the state-dependent memory effect. To construct the model, we represent encoders $\{\phi_i(\mathbf{q})\}_{i=1}^n$ and kernel $\Theta(t)$ in form of

$$\begin{aligned}\phi_i(\mathbf{q}) &= \mathbf{H}_i^T\psi(\mathbf{q}),\\ \Theta(t) &= \mathrm{e}^{-\alpha t}\sum_{k=0}^{N_\omega}\hat{\Theta}_k\cos(\omega_k t), \quad (4)\end{aligned}$$

where $\psi(\mathbf{q})=[\psi_1(\mathbf{q}),\ldots,\psi_{N_b}(\mathbf{q})]$ is a set of sparse bases and $\mathbf{H}=[\mathbf{H}_1^T,\ldots,\mathbf{H}_n^T]$ are trainable coefficients. In this

work, we choose the piece-wise linear bases such that the correlation between ϕ_i and ϕ_j can be efficiently evaluated; other localized basis can be also used. $\Theta(t)$ needs to preserve positive semidefiniteness. Hence, we represent Fourier modes $\hat{\Theta}_k = \Gamma_k \Gamma_k^T$, where $\Gamma_k \in \mathbb{R}^{n\times n}$ is a low-triangular matrix to be determined along with $\alpha \geq 0$. For this study, the full dynamics is reversible; α approaches 0 and serves as a regularization parameter. We note that $\Theta(t)$ in Eq. (4) can be generalized by introducing an antisymmetry part and refer to SM [60] for discussion.

To learn the reduced model (3), we need to choose appropriate metrics such that the state-dependent non-Markovian nature can be manifested. While autocorrelations such as $c_{vv}(t) = \langle \mathbf{v}(t)\mathbf{v}(0)^T \rangle$ merely characterize the overall memory effect, a crucial observation is that the correlations conditional with different initial state $\mathbf{q}_0$ further depends on the local energy dissipation and therefore naturally encodes the signatures of the heterogeneous memory effect. Accordingly, we right-multiply the second equation of (3) by $\mathbf{v}(0)$ and take the conditional expectation on $\mathbf{q}_0 = \mathbf{q}^*$, i.e.,

$$\begin{aligned}\mathbf{g}(t;\mathbf{q}^*) &= \int_0^t \langle \boldsymbol{\phi}_t^T \Theta_{t-\tau} \boldsymbol{\phi}_\tau \mathbf{v}_\tau \mathbf{v}_0^T | \mathbf{q}_0 = \mathbf{q}^* \rangle \mathrm{d}\tau \\ &= \int_0^t \langle \mathrm{Tr}[\Theta_{t-\tau} \mathbf{H} \boldsymbol{\psi}_\tau \mathbf{v}_\tau \mathbf{v}_0^T \boldsymbol{\psi}_t^T \mathbf{H}^T] | \mathbf{q}_0 = \mathbf{q}^* \rangle \mathrm{d}\tau \\ &= \int_0^t \mathrm{Tr}[\Theta_{t-\tau} \mathbf{H} \mathbf{C}_{\psi,\psi}(t,\tau;\mathbf{q}^*) \mathbf{H}^T] \mathrm{d}\tau,\end{aligned}$$

where $\mathbf{g}(t;\mathbf{q}^*) \coloneqq \langle [\dot{\mathbf{p}}_t + \nabla U(\mathbf{q}_t)] \mathbf{v}_0^T | \mathbf{q}_0 = \mathbf{q}^* \rangle$ and $\mathbf{C}_{\psi,\psi}(t,\tau;\mathbf{q}^*) \coloneqq \langle \boldsymbol{\psi}_\tau \mathbf{v}_\tau \mathbf{v}_0^T \boldsymbol{\psi}_t^T | \mathbf{q}_0 = \mathbf{q}^* \rangle$ is a three-point correlation characterizing the coupling among the bases. Since $\boldsymbol{\psi}(\mathbf{q})$ is sparse, $\boldsymbol{\psi}_\tau \boldsymbol{\psi}_t^T$ can be evaluated with $O(1)$ complexity and hence $\mathbf{C}_{\psi,\psi}(t,\tau;\mathbf{q}^*)$ can be efficiently precomputed. Accordingly, we can train the reduced model in terms of coefficients $\mathbf{H}$ for encoders $\boldsymbol{\phi}(\mathbf{q})$ as well as matrices $\{\Gamma_k\}_{k=1}^{N_\omega}$ and α for kernel $\Theta(t)$ by minimizing the empirical loss

$$\begin{aligned}L &= \sum_{l=1}^{N_q} \sum_{k=1}^{N_t} \left\| \tilde{\mathbf{g}}(t_k;\mathbf{q}^{(l)}) - \mathbf{g}(t_k;\mathbf{q}^{(l)}) \right\|^2, \\ \tilde{\mathbf{g}}(t_k;\mathbf{q}^{(l)}) &= \sum_{j=1}^{k} \mathrm{Tr}\left[\Theta(t_k - t_j) \mathbf{H} \mathbf{C}_{\psi,\psi}(t_k,t_j;\mathbf{q}^{(l)}) \mathbf{H}^T \right] \delta t,\end{aligned} \tag{5}$$

where $\{\mathbf{q}^{(l)}\}_{l=1}^{N_q}$ represent configuration samples within the phase space. For systems with pronounced free energy barriers, $\mathbf{q}^{(l)}$ can be collected along with free energy construction (e.g., see Ref. [76]), whereas the conditional correlations for each $\mathbf{q}^{(l)}$ need to be sampled from unbiased equilibrium trajectories. $\tilde{\mathbf{g}}(\cdot)$ represents the prediction by the reduced model that depends on the trainable variables and the precomputed correlation $\mathbf{C}_{\psi,\psi}$ (see SM [60]). δt is the time step.

To simulate the reduced model (3) and (4) on $t \in [0,T]$, we generate coherent noise $\boldsymbol{\mathcal{R}}_t = \boldsymbol{\phi}_t^T \tilde{\mathbf{R}}(t)$, where $\tilde{\mathbf{R}}\colon \mathbb{R}^+ \to \mathbb{R}^n$ is a Gaussian random process. Specifically, we can show that by choosing $\langle \tilde{\mathbf{R}}(t) \tilde{\mathbf{R}}(\tau)^T \rangle = k_B T \Theta(t-\tau)$, the reduced model retains a consistent equilibrium density, i.e., $\rho_{\mathrm{eq}}(\mathbf{q},\mathbf{p}) \propto \exp\{-\beta[U(\mathbf{q}) + \frac{1}{2}\mathbf{p}^T \mathbf{M}^{-1} \mathbf{p}]\}$ (see proof in SM [60]). Accordingly, we can generate $\{\tilde{\mathbf{R}}(t_i)\}_{i=0}^N$ by

$$\tilde{\mathbf{R}}(t_i) = \beta^{-1/2} \sum_{k=0}^{2N} \tilde{\Theta}_k^{1/2} [\cos(\omega_k t_i) \xi_k + \sin(\omega_k t_i) \eta_k], \tag{6}$$

where $\tilde{\Theta}_k$ are the Fourier (essentially cosine) modes of $\Theta(|t|)$ on $[-T,T]$ (see Refs. [78,79] and SM [60] for the analytical form); ξ_k and η_k are independent Gaussian random vectors. In practice, $\tilde{\mathbf{R}}(t)$ by Eq. (6) can be generated using FFT [80] and the convolution term $\int_0^t \boldsymbol{\phi}_t^T \Theta(t-\tau) \boldsymbol{\phi}_\tau \mathbf{v}(\tau) \mathrm{d}\tau$ in Eq. (3) can be efficiently evaluated using the fast convolution algorithm [81], both of which only require $O(N \log N)$ complexity.

The present model enables us to simulate the reduced dynamics beyond the standard GLE and systematically investigate open problems like the state-dependent memory effect on the collective dynamics of complex systems such as molecule kinetics. In this Letter, we consider the molecule benzyl bromide in an aqueous environment. The MD model consists of one benzyl bromide molecule and 2400 water molecules. The isothermal-isobaric thermostat [82] is used to equilibrate the system at 298 K and 1 bar; a canonical ensemble with a Nosé-Hoover thermostat [83,84] is used in the production stage (see SM [60] for MD setup and units). The resolved variable $\mathbf{q}$ characterizes the interplay between the substituent and the benzene group and is defined as the distance between the bromine and the ipso-carbon atom.

Let us start with the standard GLE by setting features $\boldsymbol{\phi}(\mathbf{q}) \equiv \mathbf{I}$ in Eq. (3), which capture the dynamics on the resolved scale considered in Refs. [47,48]. We right-multiply $\mathbf{q}(0)$ to Eq. (3) and compute the correlation functions, i.e., $h(t) = \int_0^t \Theta(t-\tau) c_{vq}(\tau) \mathrm{d}\tau$, where $h(t) = \langle [\dot{\mathbf{p}}_t + \nabla U(\mathbf{q}_t)] \mathbf{q}_0^T \rangle$. The standard GLE kernel $\Theta(t)$ [i.e., $\mathbf{K}(t)$ in Eq. (1)] can be obtained using the Fourier transform of the integral equation. If the reduced dynamics (1) can be simplified as the standard GLE, then $c_{vq}(t)$ should be accurately reproduced. Figure 1 shows the prediction of $c_{vq}(t)$ from the standard GLE and the full MD model. The apparent deviations imply non-negligible state dependency. To further probe this effect, we compute $h''(t;\mathbf{q}^*) = g'(t;\mathbf{q}^*)$ conditional with different initial states $\mathbf{q}^*$. Unlike a unified short-time correlation

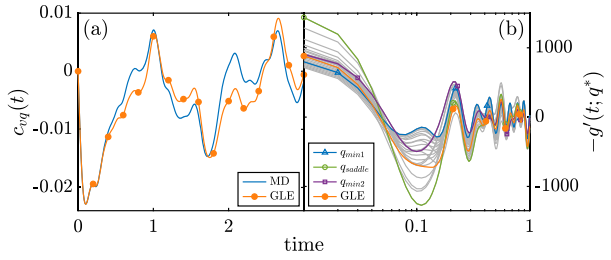

FIG. 1. Correlation functions predicted by the standard GLE and the full MD: (a) overall $c_{vq}(t)$ and (b) $-g'(t;\mathbf{q}^*)$ conditional with $\mathbf{q}^*$ representing various initial states (gray lines), including two local minima and the saddle point [see inset of Fig. 2(a)]. The large dispersion implies the limitation of the standard GLE, which predicts a single curve in short time.

[i.e., $g'(0;q^*) = -k_BT\Theta(0)/m$] predicted by the standard GLE, the large dispersion reveals the heterogeneous nature of the energy dissipation process.

To capture the state-dependent memory, we train the present model (3) with a different number of features. Figures 2(a) and 2(b) shows the obtained encoder $\phi(\cdot)$ using one feature and $\Theta(t)$ is scaled with $\Theta(0) = 1$. ϕ exhibits apparent deviation from a uniform distribution. In particular, it shows a peak value near the saddle point $q = 3.65$, implying a larger effective friction near the regime. This result supports a similar assumption in semi-analytical studies [55] on improving Kramers' theory [85]. Also, it explains the short-time dispersion shown in Fig. 1, where $g'(t;\mathbf{q}^*)$ at the saddle point is significantly larger than the local minima. Figures 2(c) and 2(d) shows the obtained encoders $\{\phi_i(\cdot)\}_{i=1}^n$ with $n = 4$ features and the diagonal components of $\Theta(t)$. Compared with the case of $n = 1$, the larger variation of ϕ_i enables a better representation of the state-dependent memory.

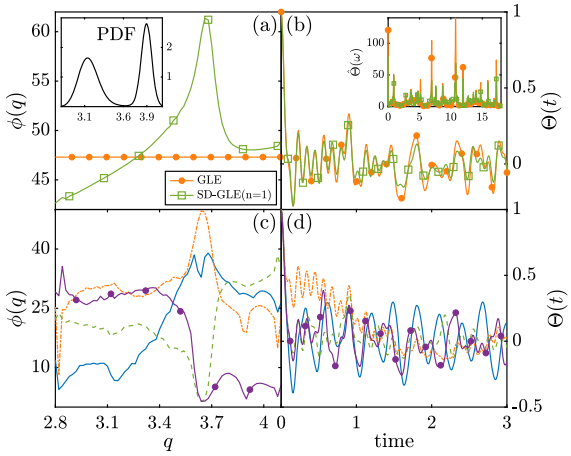

FIG. 2. The state features ϕ and diagonal components of the matrix-valued kernel $\Theta(t)$ for the present model with state-dependent memory (SD-GLE) trained using (a),(b) one feature and (c),(d) four features. Inset plots: (a) probability density function (PDF) of q, where $\phi(q)$ near the saddle point shows a peak; (b) Fourier modes of $\Theta(t)$.

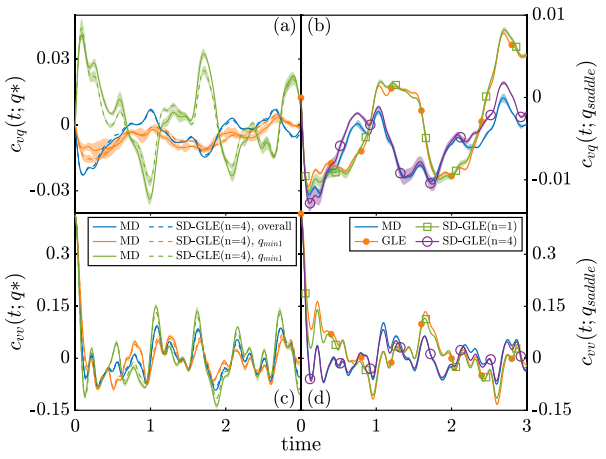

FIG. 3. Overall and conditional correlation functions predicted by the full MD and various reduced models for two local minima and the saddle point: (a),(b) c_{vq} and (c),(d) c_{vv}. Shaded regimes represent the 95% confidence interval; same for Fig. 4.

Next, we examine the conditional correlations $c_{vq}(t;\mathbf{q}^*)$ and $c_{vv}(t;\mathbf{q}^*)$. As shown in Fig. 3, for both local minima and the saddle point, the predictions of the present model using four features agree well with the MD results. In contrast, the predictions of the standard GLE show apparent deviations for $\mathbf{q}^*$ at the saddle point. Also, the present model using four features with a diagonal $\Theta(t)$ (see SM [60]) shows improved short-time predictions but remains insufficient for long-time correlations. This indicates the complex global variation of the memory term, which can not be represented by a simple state-dependent rescaling of a kernel function; the non-Markovian coupling among multiple features is crucial to capture the heterogeneous energy dissipation over the full space.

Finally, we examine the collective behavior related to molecule kinetics. Figure 4(a) shows the position correlation $c_{qq}(t)$ characterizing the molecule conformation relaxation. Compared with the MD results, the standard GLE shows a significant underestimation of the relaxation time. This discrepancy is possibly due to the larger effective friction near the saddle point [see Fig. 2(a)], which essentially dampens the transition between the two local

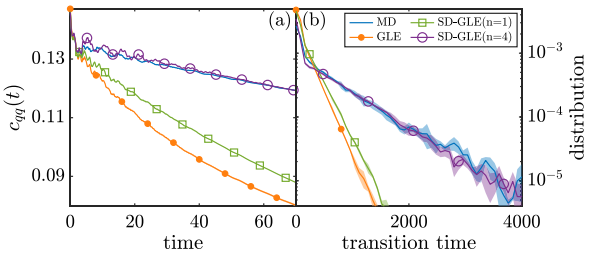

FIG. 4. Collective molecule behaviors predicted by the full MD and the various reduced models: (a) overall conformation relaxation and (b) distribution of the transition time between the two local minima.

minima. The standard GLE overlooks such state dependency and therefore yields a faster relaxation. This limitation is consistently reflected in the distribution of the transition time. For this system, the free energy barrier is approximately 3.5 k_BT (see SM [60]); the transition time is obtained from the simulation trajectories of the MD and various reduced models. As shown in Fig. 4(b), the standard GLE predicts a larger probability for the short transition time, indicating a smaller overall friction than the local (i.e., saddle point) value. Fortunately, the heterogeneous non-Markovianity can be faithfully retained in the present model. In particular, the constructed model using one feature yields a better prediction than the standard GLE. As we increase to four features, the predictions recover the MD results.

In summary, to plan an optimal hiking trail on a mountain, a landscape map is generally insufficient; the local path roughness needs to be properly considered. Similarly, to predict the reduced dynamics of a multiscale system, the state-dependent memory may need to be modeled to account for the heterogeneous energy dissipation arising from the unresolved dynamics, which, however, has been broadly overlooked. While the crucial role of the non-Markovian effect that complements the conservative free energy has been gradually recognized, the formulation of the memory term remains largely empirical (e.g., the standard GLE). The current work focuses on this caveat and presents a data-driven approach to learning such a stochastic reduced model beyond the standard GLE, where the complex state-dependent memory can be naturally encoded in the non-Markovian interactions among a set of features in terms of the resolved variables. The training does not rely on the explicit knowledge of the full model and only utilizes the trajectory samples, where the three-point correlations can be efficiently precomputed. Numerical results of a molecule system demonstrate the crucial role of the state-dependent non-Markovianity on collective behavior, where the standard GLE shows limitations due to the oversimplified assumption of a homogeneous memory kernel. In contrast, the present model accurately predicts the molecule kinetics including the transition time distribution, and provides a reliable approach to simulate stochastic reduced dynamics of multiscale problems that faithfully retains the collective behaviors and rare event properties [86] beyond empirical models.

Acknowledgments—We acknowledge helpful discussions from Zhaopeng Hao, Liyao Lyu, and Peiyuan Gao. The work is partly supported by NSF under Grants No. DMS-2110981, No. DMS-2143739 and the ACCESS program through allocation MTH210005.

Data availability—The code is public on [87].

[1] B. O. Koopman, Proc. Natl. Acad. Sci. U.S.A. **17**, 315 (1931).
[2] H. Mori, Prog. Theor. Phys. **33**, 423 (1965).
[3] R. Zwanzig, Lect. Theor. Phys. **3**, 106 (1961).
[4] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, New York, 2001).
[5] O. F. Lange and H. Grubmüller, J. Chem. Phys. **124**, 214903 (2006).
[6] E. Darve, J. Solomon, and A. Kia, Proc. Natl. Acad. Sci. U.S.A. **106**, 10884 (2009).
[7] M. Ceriotti, G. Bussi, and M. Parrinello, Phys. Rev. Lett. **102**, 020601 (2009).
[8] A. D. Baczewski and S. D. Bond, J. Chem. Phys. **139**, 044107 (2013).
[9] A. Davtyan, J. F. Dama, G. A. Voth, and H. C. Andersen, J. Chem. Phys. **142**, 154104 (2015).
[10] H. Lei, N. A. Baker, and X. Li, Proc. Natl. Acad. Sci. U.S.A. **113**, 14183 (2016).
[11] A. Russo, M. A. Durán-Olivencia, I. G. Kevrekidis, and S. Kalliadasis, IEEE Trans. Neural Networks Learn. Syst. **35**, 6531 (2024).
[12] G. Jung, M. Hanke, and F. Schmid, J. Chem. Theory Comput. **13**, 2481 (2017).
[13] H. S. Lee, S.-H. Ahn, and E. F. Darve, J. Chem. Phys. **150**, 174113 (2019).
[14] L. Ma, X. Li, and C. Liu, J. Comput. Phys. **380**, 170 (2019).
[15] S. Wang, Z. Ma, and W. Pan, Soft Matter **16**, 8330 (2020).
[16] Y. Zhu and D. Venturi, J. Stat. Phys. **178**, 1217 (2020).
[17] V. Klippenstein and N. F. A. van der Vegt, J. Chem. Phys. **154**, 191102 (2021).
[18] H. Vroylandt, L. Goudenège, P. Monmarché, F. Pietrucci, and B. Rotenberg, Proc. Natl. Acad. Sci. U.S.A. **119**, e2117586119 (2022).
[19] Z. She, P. Ge, and H. Lei, J. Chem. Phys. **158**, 034102 (2023).
[20] P. Xie, R. Car, and W. E, Proc. Natl. Acad. Sci. U.S.A. **121**, e2308668121 (2023).
[21] P. Hänggi, in *Stochastic Dynamics*, edited by L. Schimansky-Geier and T. Pöschel (Springer, Berlin, Heidelberg, 1997), pp. 15–22.
[22] V. Klippenstein, M. Tripathy, G. Jung, F. Schmid, and N. F. van der Vegt, J. Phys. Chem. B **125**, 4931 (2021).
[23] H. A. Posch, U. Balucani, and R. Vallauri, Physica (Amsterdam) **123A**, 516 (1984).
[24] J. E. Straub, M. Borkovec, and B. J. Berne, J. Phys. Chem. **91**, 4995 (1987).
[25] J. E. Straub, B. J. Berne, and B. Roux, J. Chem. Phys. **93**, 6804 (1990).
[26] S. S. Plotkin and P. G. Wolynes, Phys. Rev. Lett. **80**, 5015 (1998).
[27] G. Luo, I. Andricioaei, X. S. Xie, and M. Karplus, J. Phys. Chem. B **110**, 9363 (2006).
[28] R. B. Best and G. Hummer, Phys. Rev. Lett. **96**, 228104 (2006).
[29] R. B. Best and G. Hummer, Proc. Natl. Acad. Sci. U.S.A. **107**, 1088 (2010).
[30] M. Hinczewski, Y. von Hansen, J. Dzubiella, and R. R. Netz, J. Chem. Phys. **132**, 245103 (2010).
[31] R. Satija, A. Das, and D. E. Makarov, J. Chem. Phys. **147**, 152707 (2017).

[32] J. A. Morrone, J. Li, and B. J. Berne, J. Phys. Chem. B **116**, 378 (2012).
[33] J. O. Daldrop, B. G. Kowalik, and R. R. Netz, Phys. Rev. X **7**, 041065 (2017).
[34] J. M. Deutch and I. Oppenheim, J. Chem. Phys. **54**, 3547 (1971).
[35] R. Zwanzig, J. Stat. Phys. **9**, 215 (1973).
[36] R. Zwanzig, J. Phys. Chem. **96**, 3926 (1992).
[37] A. Berezhkovskii and A. Szabo, J. Chem. Phys. **135**, 074108 (2011).
[38] F. Glatzel and T. Schilling, Europhys. Lett. **136**, 36001 (2022).
[39] H. Vroylandt, Europhys. Lett. **140**, 62003 (2022).
[40] H. Vroylandt and P. Monmarché, J. Chem. Phys. **156**, 244105 (2022).
[41] C. Ayaz, L. Scalfi, B. A. Dalton, and R. R. Netz, Phys. Rev. E **105**, 054138 (2022).
[42] B. Jung and G. Jung, J. Chem. Phys. **159**, 084110 (2023).
[43] L. Lyu and H. Lei, Phys. Rev. Lett. **131**, 177301 (2023).
[44] R. Satija and D. E. Makarov, J. Phys. Chem. B **123**, 802 (2019).
[45] F. Grogan, H. Lei, X. Li, and N. A. Baker, J. Comput. Phys. **418**, 109633 (2020).
[46] D. Singh, K. Mondal, and S. Chaudhury, J. Phys. Chem. B **125**, 4536 (2021).
[47] C. Ayaz, L. Tepper, F. N. Brünig, J. Kappler, J. O. Daldrop, and R. R. Netz, Proc. Natl. Acad. Sci. U.S.A. **118**, e2023856118 (2021).
[48] B. A. Dalton, C. Ayaz, H. Kiefer, A. Klimek, L. Tepper, and R. R. Netz, Proc. Natl. Acad. Sci. U.S.A. **120**, e2220068120 (2023).
[49] J. E. Straub, M. Borkovec, and B. J. Berne, J. Chem. Phys. **89**, 4833 (1988).
[50] S. Singh, R. Krishnan, and G. Robinson, Chem. Phys. Lett. **175**, 338 (1990).
[51] B. Carmeli and A. Nitzan, Chem. Phys. Lett. **102**, 517 (1983).
[52] G. Tarjus and D. Kivelson, Chem. Phys. **152**, 153 (1991).
[53] R. Krishnan, S. Singh, and G. W. Robinson, J. Chem. Phys. **97**, 5516 (1992).
[54] G. A. Voth, J. Chem. Phys. **97**, 5908 (1992).
[55] J. B. Straus, J. M. Gomez Llorente, and G. A. Voth, J. Chem. Phys. **98**, 4082 (1993).
[56] G. R. Haynes, G. A. Voth, and E. Pollak, Chem. Phys. Lett. **207**, 309 (1993).
[57] G. R. Haynes, G. A. Voth, and E. Pollak, J. Chem. Phys. **101**, 7811 (1994).
[58] P. Cossio, G. Hummer, and A. Szabo, Proc. Natl. Acad. Sci. U.S.A. **112**, 14248 (2015).
[59] C. Hijón, P. Español, E. Vanden-Eijnden, and R. Delgado-Buscalioni, Faraday Discuss. **144**, 301 (2010).
[60] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.133.077301 for the discussion on the general mass matrix, the coherent noise and proof of the consistent equilibrium distribution, the details of the training process, the simulation of the reduced model, the full MD model of the benzyl bromide system, the additional numerical results predicted by the present and other forms of reduced models, and the generalization of the present model with an antisymmetry part of the memory kernel, which includes Refs. [61–68].
[61] D. Kingma and J. Ba, *International Conference on Learning Representations (ICLR)* (Academic Press, Cambridge, 2015).
[62] C. Ayaz, L. Tepper, and R. R. Netz, Turk. J. Phys. **46**, 194 (2022).
[63] Z. Li, X. Bian, X. Li, and G. E. Karniadakis, J. Chem. Phys. **143**, 243128 (2015).
[64] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, J. Comput. Chem. **25**, 1157 (2004).
[65] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, J. Phys. Chem. **97**, 10269 (1993).
[66] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[67] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[68] S. Miyamoto and A. Kollman Peter, J. Comput. Chem. **13**, 952 (2004).
[69] G. Torrie and J. Valleau, J. Comput. Phys. **23**, 187 (1977).
[70] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J. Comput. Chem. **13**, 1011 (1992).
[71] E. Darve and A. Pohorille, J. Chem. Phys. **115**, 9169 (2001).
[72] A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. **99**, 12562 (2002).
[73] L. Rosso, P. Mináry, Z. Zhu, and M. E. Tuckerman, J. Chem. Phys. **116**, 4389 (2002).
[74] L. Maragliano and E. Vanden-Eijnden, Chem. Phys. Lett. **426**, 168 (2006).
[75] J. B. Abrams and M. E. Tuckerman, J. Phys. Chem. B **112**, 15742 (2008).
[76] L. Maragliano and E. Vanden-Eijnden, J. Chem. Phys. **128**, 184110 (2008).
[77] H. Lei and X. Li, J. Chem. Phys. **154**, 184108 (2021).
[78] M. Berkowitz, J. Morgan, and J. A. McCammon, J. Chem. Phys. **78**, 3256 (1983).
[79] V. A. Ogorodnikov and S. M. Prigarin, *Numerical Modelling of Random Processes and Fields: Algorithms and Applications* (De Gruyter, Berlin, Boston, 1996).
[80] J. W. Cooley and J. W. Tukey, Math. Comput. **19**, 297 (1965).
[81] A. Schädle, M. López-Fernández, and C. Lubich, SIAM J. Sci. Comput. **28**, 421 (2006).
[82] G. J. Martyna, D. J. Tobias, and M. L. Klein, J. Chem. Phys. **101**, 4177 (1994).
[83] S. Nosé, Mol. Phys. **52**, 255 (1984).
[84] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[85] H. Kramers, Physica **7**, 284 (1940).
[86] W. E and E. Vanden-Eijnden, Annu. Rev. Phys. Chem. **61**, 391 (2010).
[87] P. Ge and H. Lei (2023), 10.5281/zenodo.13186786.