INTRODUCTION AND MAIN RESULTS

In different branches of Science, the interpretation and mathematical modeling of both experimental and computational data requires the analysis of the system dynamics in terms of a reduced set of collective variables (CVs), or order parameters. Prominent examples include chemical reactions in solution, conformational changes in biomolecules or phase transitions in condensed matter systems. A standard approach is to approximate the dynamical properties with long but cheap simulations of more general systems, can be estimated from a data-driven approach provides a reduced dynamical model for multidimensional collective variables, enabling the accurate sampling of their long-time dynamical properties at a computational cost drastically reduced with respect to all-atom numerical simulations. We illustrate the potential of this method on several model systems, both in and out of equilibrium.

The most widespread model for this task is the Langevin equation, which can be derived – in some particular cases – from the Hamiltonian dynamics of a small system interacting with a large environment. It describes the evolution of a Markov process, which requires that the decorrelation time of the environment is short compared to the characteristic times of the reduced system. However many cases do not enter the validity range of this approximation, displaying memory effects [1–8]. To go beyond the Markovian approximation, a popular class of processes is given by the generalized Langevin equation [9–15]

\[ \begin{align*}
\dot{x}(t) &= v(t) \\
M\dot{v}(t) &= F_{\text{eff}}(x(t)) - \int_{0}^{t} K(t - \tau)v(\tau)d\tau + R(t),
\end{align*} \]

(1)

where \( x(t) \) is the value of the \( d \)-dimensional collective variable at time \( t \), \( v(t) \) its time derivative, \( M \) is an effective mass, \( F_{\text{eff}} \) is a mean force, usually deriving from a potential \( V \) identified with the free energy, \( K \) a memory kernel (which could more generally depend on the CVs) and \( R(t) \) a (colored) noise.

Whereas the general form of the GLE can be derived from the dynamics of the original full system following the Mori-Zwanzig formalism [9, 16–18] (see however Ref. 19 for caveats), there are few cases where an analytical derivation of the memory kernel is possible [20]. For more general systems, \( K \) can be estimated from a data-driven approach. In most cases, the goal is to extract the memory kernel from trajectories of the CV computed with all-atom simulations [4, 21–36].

As already mentioned, the solutions of (1) are not Markov processes, except when \( K \) is the Dirac \( \delta \) function and \( R \) is a white noise. Both for fitting the model and then for generating new trajectories of the effective dynamics, it is convenient to consider the subclass of models where an extended process \( (x, v, h) \) is Markovian, with \( h \) some hidden auxiliary variables [12, 37–44]. Restricting further to the case where the evolution of the hidden variables and the coupling with the observed variables

Likelihood-based parametric estimator for memory kernel in molecular dynamics

Hadrien Vroylandt,1 Ludovic Goudenège,2 Pierre Monmarché,3,4 Fabio Pietrucci,5 and Benjamin Rotenberg6

1 Sorbonne Université, institut des sciences du calcul et des données, ISCD, F-75005 Paris, France
2 CNRS, FR 3487, Fédération de Mathématiques de CentraleSupélec, CentraleSupélec, 91190 Gif-sur-Yvette, Université Paris-Saclay, France
3 Sorbonne Université, Laboratoire Jacques-Louis Lions, LJLL, F-75005 Paris
4 Sorbonne Université, Laboratoire de Chimie Théorique, LCT, F-75005 Paris
5 Sorbonne Université, Muséum National d’Histoire Naturelle, UMR CNRS 7590, Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, IMPMC, F-75005 Paris, France
6 Physicochimie des électrolytes et Nanosystèmes Interfaciaux, Sorbonne Université, CNRS, 4 Place Jussieu F-75005 Paris, France

We introduce a new method to accurately and efficiently estimate the effective dynamics of collective variables in molecular simulations. Such reduced dynamics play an essential role in the study of a broad class of processes, ranging from chemical reactions in solution to conformational changes in biomolecules or phase transitions in condensed matter systems. The memory kernel and random noise are correctly recovered by this procedure. This data-driven approach provides a reduced dynamical model for multidimensional collective variables, enabling the accurate sampling of their long-time dynamical properties at a computational cost drastically reduced with respect to all-atom numerical simulations. We illustrate the potential of this method on several model systems, both in and out of equilibrium.
are linear, this leads to an equation of the form
\[
\begin{align*}
\dot{x} &= v \\
\dot{v} &= M^{-1}F_{\text{eff}}(x) - A_{vh}h - A_{vv}v + \sigma_{vv}\xi(t) + \sigma_{vh}W(t) \\
\dot{h} &= -A_{hh}h - A_{hv}v + \sigma_{vh}\xi(t) + \sigma_{hh}W(t)
\end{align*}
\]
where \(A_{vh}, A_{vv}, A_{hh}, A_{hv}, \sigma_{vv}, \sigma_{vh}, \sigma_{hh}\) are constant matrices and \(\xi\) and \(W\) are independent standard white noises. This gives a convenient class of models parametrized by the dimension \(d_h\) of \(h\), the corresponding matrices and the (rescaled) effective force \(M^{-1}F_{\text{eff}}\). For equilibrium processes, the coefficients of (2) are related by the so-called Fluctuation-Dissipation relation [39]. Although we could enforce this condition, thereby reducing the number of parameters, we do not since we also consider non-equilibrium systems in the following.

Integrating over the hidden variables, we recover (1) with a memory kernel of the form of a finite Prony series [39, 40]
\[
K(\tau) = w_0\delta(\tau) + \sum_{k=1}^{d_h} w_k e^{-\lambda_k \tau}
\]
where \(w_k\) and \(\lambda_k\) are (possibly complex) coefficients of the series derived from the matrices \(A_{vh}, A_{vv}, A_{hh}, A_{hv}\). In principle, on all finite time intervals, any kernel given as the sum of a Dirac function at zero and of a continuous function can be approximated arbitrarily accurately by a sum of the form (3). However, in practice \(d_h\) is relatively small and memory kernel with e.g. algebraic tail can only be approximated on small time interval [37, 44].

The use of auxiliary variables in the form of (2) has been abundantly used and studied, as it allows efficient integration of GLE (1) [39, 41, 45]. The estimation of GLE parameters from simulations is an active field of research. The main method consists in a non-parametric estimation of the memory kernel via the Volterra integral equation [12, 24, 33, 36, 38, 45–47], but other methods have also been proposed [23, 31, 43, 44, 48]. In the present work, we i) introduce a novel parametric estimator of GLE coefficients, based on a maximum likelihood approach and ii) show that it allows building faithful coarse-grained models of MD simulations in a cost-effective way (i.e., starting from a relatively small training data set), such that the dynamics is well reproduced.

**DATA-DRIVEN APPROACH ON EXTENDED DYNAMICS**

In statistics, a standard method to deal with hidden variables is the Expectation-Maximization (EM) algorithm, which belongs to the category of likelihood maximization algorithms [49, 50]. It is of frequent use to estimate parameters of time series models in the case of partial or noisy observations of the system, either for hidden Markov models [51] or state-space models [52]. A first application in the context of GLE was proposed in Ref. 37 to reconstruct the memory kernel in the absence of effective force \(F_{\text{eff}}(x)\) and under more restrictive conditions than the method presented below.

The algorithm proceeds by alternating steps: In the E-step, one determines the conditional probability law of the hidden variables given the observed ones at fixed parameters; in the M-step, one optimizes the parameters to maximize the log-likelihood averaged with respect to these conditional laws. In the following we denote as \(\Theta_j\) the whole set of parameters estimated after \(j\) iterations of the algorithm, which includes the mean force projected on some functional basis (which can be very large in general, or reduced if prior knowledge on the system is available), the coefficients of the matrices \(A, \sigma\) of (2) and, for technical reasons discussed below, the mean value at time zero of the hidden variables, \(\langle h_0\rangle\).

**EM algorithm**

The available data, obtained from all-atom simulations, consists of a set of independent trajectories. For simplicity of the notation, we introduce the algorithm with only one trajectory \(\{x\}_{0:N} = \{x(k\Delta t), k \in [0, N]\}\) for some time-step \(\Delta t\) and simulation time \(T = N\Delta t\), the extension to the general case being straightforward. The statistical models we consider are Euler-Maruyama discretizations of (2) with the same time-step \(\Delta t\), for a fixed dimension \(d_h\) of auxiliary variables \(h\). The state of the system at time \(t = k\Delta t\) will be denoted \((x_k, v_k, h_k) = (X, h)_k = X_k\) and we write \(\{X\}_{0:N}\) a complete trajectory of the system. Hence, \(X\) is the value of the known variables since, from the choice of the Euler-Maruyama scheme, the velocity can be computed as \(v_k = (x_{k+1} - x_k)/\Delta t\). In the following we write \(\pi(z)\) the probability density of a variable \(z\), \(\pi(z|u)\) the conditional probability density of \(z\) with respect to \(u\) and, in both cases, \(\pi_0\) to explicit the value of the parameters if needed.

As the extended system is Markovian, we have for the probability density of a trajectory
\[
\pi(\{X\}_{0:N}) = \pi(X_0) \times \prod_{k=0}^{N-1} \pi(X_{k+1}|X_k) \tag{4}
\]
and the form of (2) and of the Euler-Maruyama scheme lead to a Gaussian transition kernel, characterized by its mean \(\mu\) and variance \(\Sigma\) (see Appendix).

**E-step** The first step is to compute the conditional law of the hidden variables given the observed variables at the current guess of the parameters, i.e. \(\pi_{\Theta_j}(h|\{X\}_{0:N})\). Due to the Markovianity of the extended system, it is sufficient to compute the mean and variance of the Gaussian marginal laws
\[
\pi(h_k, h_{k+1} | \{\tilde{X}\}_{0:N}) \quad \text{for all } k \in [0, N - 1].
\]

Taking advantage of the explicit form of the transition probability (Eq. (3) in the S.I.), we apply an iterative predictor-corrector-smoother approach (also known as Kalman filter and Rauch-Tung-Striebel smoother) [53]. Starting from the trajectory up to step \(k-1\), we determine the law of the hidden variable \(h_k\) conditioned on the past information \(\{\tilde{X}\}_{0:k-1}\). We then use the expression of the transition probability \(\pi(X_k | X_{k-1})\) to determine the current value of \(\pi(h_k | \{\tilde{X}\}_{0:k})\). These are the prediction and correction parts that are run forward on the trajectory, i.e., from \(k = 0 \) to \(k = N\) (arrow (1) on Fig. 1). The initial guess at \(k = 0\) of \(\pi(h_0)\) uses the measured \(\langle h_0 \rangle\) vector as the mean and an arbitrary variance (identity matrix). Such initial guess could be optimized, but we did not observe any influence on the final results. The second part of the E-step, called the smoother part, computes \(\pi(h_{k-1} | h_k, \{\tilde{X}\}_{0:N})\) and is run backward, i.e. from \(k = N\) to \(k = 0\) (arrow (2) on Fig. 1) which finally gives the required probability law of \(h_{k-1}, h_k\) conditioned to the full observed trajectory. Detailed formula are presented in Appendix.

**M-step** For any set of parameters \(\Theta\), introduce the evidence lower bound \(\mathbb{L}_{LB}^j\) after the \(j\)-th iteration of the algorithm as the expectation with respect to \(\pi_{\Theta_j}(\{h\}_{0:N} | \{\tilde{X}\}_{0:N})\) of the log-likelihood of the full trajectory \(\{X\}_{0:N}\) with parameter \(\Theta\), namely (see derivation in Appendix)

\[
\mathbb{L}_{LB}^j(\Theta) = \int \pi_{\Theta_j}(\{h\}_{0:N} | \{\tilde{X}\}_{0:N}) \ln \pi_{\Theta}(\{X\}_{0:N}) \, dh_{0:N} \\
= \int \pi_{\Theta_j}(h_0 | \{\tilde{X}\}_{0:N}) \ln \pi_{\Theta}(X_0) dh_0 \\
+ \sum_{k=0}^{N-1} \int \pi_{\Theta_j}(h_k, h_{k+1} | \{\tilde{X}\}_{0:N}) \\
\times \ln \pi_{\Theta}(X_{k+1} | X_k) dh_k dh_{k+1}. \\
(5)
\]

The M-step consists in setting \(\Theta_{j+1}\) to be the maximizer of this quantity. Notice that due to the particular form of (2), \(\mathbb{L}_{LB}^j(\Theta)\) is an explicit function of \(\Theta\), that can be easily optimized as described in Appendix.

**Full algorithm** The algorithm then run as follows. An initial random or informed guess \(\Theta_0\) is taken for the parameters. Such informed guess could come from a previous execution of the algorithm with a different number of hidden dimensions. From parameters \(\Theta_j\), a new set of parameters \(\Theta_{j+1}\) is computed through an iteration of E and M steps. Since maximizing the evidence lower bound \(\mathbb{L}_{LB}^j\) increases the observed likelihood, the method is iterated until either a prescribed maximum number of EM steps or a convergence criterion is reached.

**Assessing the quality of a given model** The number of hidden dimensions \(d_h\) is an important parameter of the algorithm. It can be chosen using a model validation approach, classically by dividing the set of trajectories between a training and a validation set. However here, we simply compute the optimal parameters for several values of \(d_h\) and compare the predictions of the corresponding models for a number of observable properties, such as the memory kernel, velocity-autocorrelation functions (VACF) or mean first passage times. Similarly, the quality of the model depends on the time step used for the coarse-grained dynamics. This choice depends among other things on the numerical scheme for the propagator. For a given underlying dynamics of the full system, the most accurate choice for the coarse-grained one is to use the same time step \(\Delta t_{full}\), but as a compromise with the amount of data one can also use \(\Delta t = m \Delta t_{full}\) (i.e. using only every \(m\) step), with \(m\) a small integer.

**Efficient sampling of new trajectories** Once the model has been optimized by the EM algorithm, it can be used to generate new trajectories in the CV-space. Due to their limited computational cost compared to MD trajectories, such synthetic data grants easier access to well-converged average properties, in the form of static and dynamic observables. As an example, in section the mean first passage times (as well as their probability densities) of a Lennard-Jones dimer in a bath are estimated based on the GLE model and compared with the corresponding ones extracted from expensive MD simulations.

**Code availability** A python package to perform the analysis introduced in the present work is available at https://github.com/HadrienNU/GLE_AnalysisEM.

**RESULTS** We first present the result of the algorithm on a simple yet non-trivial test case with a 1D system following the extended dynamics of (2), with 5 hidden dimensions and a quadratic potential well \(V(x) = x^2/2\), using 20 trajectories of \(25 \cdot 10^3\) steps and a time-step of \(5 \cdot 10^{-3}\). The effective force is fitted as a linear function of \(x\). Fig. 2a compares the result of our algorithm to the true memory function that can be computed from (3) and the one obtained by the inverse Volterra method [54]. It demonstrates that the present EM method is able to reproduce
the true memory kernel. Furthermore, the parametric structure of the fitted model enforces the decay to zero of the memory kernel, whereas the Volterra method is unstable at long time. Fig. 2b finally shows that the method accurately reproduces the VACF.

![Graph](https://via.placeholder.com/150)

**Figure 2.** Equilibrium 1D case. (a) Memory kernel \(K(\tau)/M\) divided by the mass \(M\): The true kernel used to generate the reference trajectories (dark blue line) is compared with the predictions of the inverse Volterra method (cyan solid line, with shaded area indicating uncertainties computed from a bootstrap analysis) and of the present EM method (dashed-dotted yellow line). (b) Velocity autocorrelation function, from the reference trajectories (dark blue line) and from new trajectories sampled using the fitted EM model (dashed-dotted yellow line).

The algorithm also applies to multidimensional and nonequilibrium systems. This is illustrated on Fig. 3 for a 2D system with two different thermal noises along each axis, with temperatures \(T_x = 1\) and \(T_y = 5\) and a quadratic potential \(V(x, y) = \frac{1}{2} \left( x^2 + \frac{3}{4} xy + y^2 \right) \) whose principal axes are not aligned with the \(x\) and \(y\) axes, leading to non-equilibrium conditions. This setup is inspired by a similar Markovian model used to describe non-equilibrium experiments on cold atoms [55]. We run 20 trajectories of \(30 \times 10^3\) steps with a time-step of \(5 \times 10^{-3}\). The effective 2D force is fitted as a linear combination of \(x\) and \(y\). The corresponding quadratic potential, illustrated in Fig. 3a, is in good agreement with the one used to generate the trajectories. Fig. 3b then shows that the algorithm correctly estimates the memory kernel (in the present case, a simple one with a single hidden dimension for each visible dimension). In particular, the presence of strong Markovian component is captured by the algorithm but missed by the inverse Volterra method. Finally, the dynamics of the system is well reproduced, as demonstrated for the VACF on Fig. 3c.

As a final illustration, we apply our algorithm to a more realistic 3D system composed of 512 Lennard-Jones (LJ) particles at reduced temperature \(T = k_B T/\epsilon = 1\) and reduced density \(\rho = \rho_0^3 = 1\). Two of the LJ particles are singled out to form a dimer [23], the others constituting the solvent. The CV of interest is the distance \(r\) between the two particles forming the dimer. LJ parameters for all interactions are taken as \(\epsilon = 1\) and \(\sigma = 1\) (in LJ units), except between the two particles forming the dimer, with \(\epsilon_d = 2\) and \(\sigma = 1\). The size of the cubic simulation box is \(8\sigma\), with periodic boundary conditions in all directions. The dynamics is integrated with a time step of \(\Delta t_{MD} = 0.001\) (in LJ units) in the NVE ensemble using the LAMMPS simulation package [56]. We run 20 trajectories with length of \(10^6\) time-steps and CV values are extracted every 2 steps.

We fit GLE models (2) with the EM algorithm for a number of hidden dimensions ranging from 3 to 6. In all cases, the effective force \(F_{\text{eff}}(r)\) determined from the MD trajectory is used as the single function of the above-mentioned functional basis, so that fitting this part reduces to determining a single prefactor. Our aim is to test the ability of these models to reproduce, in the statistical sense, the properties of the original simulations. In order to check the importance of the hidden variables, we also provide an analysis for a Markovian model, fitted using a maximum likelihood algorithm with 0 hidden dimensions (corresponding to the M-step of the above EM algorithm). For each fitted GLE model, we generate 75 new trajectories of length \(10^5\) time-steps using (2), to compute the observable properties and compare them with those obtained from the original set of MD trajectories.

We first compare the stationary distribution for the various GLE models in Fig. 4a, which shows the free energy as a function of the \(r\) coordinate computed from the histogram of each set of new trajectories (i.e. not the one corresponding to the fitted effective force). The good agreement with the MD free energy profile demonstrates (i) that the coefficient multiplying the model free energy profile of each model is fitted precisely and (ii) that the numerical integration of the GLE models is performed accurately. Notice that free energy beyond \(r = 4\sigma\) is affected by the size of the periodic box. The free energy displays two potential wells at \(r = 1.12\sigma\) and \(r = 2.00\sigma\), corresponding to the contact pair (CP, i.e. the dimer) and the solvent shared pair (SSP, with solvent atoms belonging to the solvation shells of both solutes), whose dynamics is investigated below.

We then consider dynamical observables in Fig. 4b,
Non-equilibrium 2D case. (a) Locus of $V(x,y) = 1$ for the original potential and the one estimated by the EM algorithm. (b) $xx$ component of the reconstructed memory kernel $K_{xx}(\tau)$ divided by the mass $M$: The true kernel used to generate the reference trajectories (dark blue line) is compared with the predictions of the inverse Volterra method (cyan solid line, with shaded area indicating uncertainties computed from a bootstrap analysis) and of the present EM method (dashed-dotted yellow line); the left peak represents the Dirac function of (3). (c) Velocity autocorrelation function (for the line, with shaded area indicating uncertainties computed from a bootstrap analysis) and of the present EM method (dashed framework for the prediction of thermodynamic and kinetic properties of experimental interest. However, the key problems in this context consist in the identification of a suitable dynamic equation and its parametrization. We developed a novel approach combining generalized Langevin equations, their numerically-efficient representation via Markovian equations including hidden variables, and a powerful machine-learning algorithm borrowed from the field of statistical modeling and data science. Starting from non-Markovian trajectories (e.g. projected all-atom molecular dynamics trajectories in condensed-matter applications), we maximize the likelihood of an extended Markovian model employing the expectation-maximization algorithm. The advantage of obtaining an explicit parametrization allows for inexpensive sampling of synthetic trajectories, that can be used for the direct computation of quantitative observables such as the first passage times between metastable states, generally hard to access through atomistic simulations.

Several features distinguish our approach from others existing in the literature. Firstly, the model we optimize includes an explicit parametrization of both the friction and the noise, ensuring consistency between the analysis of the MD trajectories and the generation of new projected trajectories. Secondly, our method is based on a maximum likelihood procedure, which is well justified from a mathematical perspective. In particular, instead of estimating a non-parametric kernel which is then parametrized (as e.g. in Volterra-based approaches), the parametric model is directly fitted on the data; this should limit the accumulation of errors. Thirdly, we do not enforce equilibrium conditions (such as the fluctuation dissipation theorem) on the model, so that the present approach offers the possibility to investigate non-

CONCLUSIONS

In this work we addressed the construction of reduced mathematical models of the dynamics of complex molecular systems. Projecting the phase-space trajectories on a reduced set of collective variables leads to a powerful framework for the prediction of thermodynamic and kinetic properties of experimental interest. However, the key problems in this context consist in the identification of a suitable dynamic equation and its parametrization. We developed a novel approach combining generalized Langevin equations, their numerically-efficient representation via Markovian equations including hidden variables, and a powerful machine-learning algorithm borrowed from the field of statistical modeling and data science. Starting from non-Markovian trajectories (e.g. projected all-atom molecular dynamics trajectories in condensed-matter applications), we maximize the likelihood of an extended Markovian model employing the expectation-maximization algorithm. The advantage of obtaining an explicit parametrization allows for inexpensive sampling of synthetic trajectories, that can be used for the direct computation of quantitative observables such as the first passage times between metastable states, generally hard to access through atomistic simulations.

Several features distinguish our approach from others existing in the literature. Firstly, the model we optimize includes an explicit parametrization of both the friction and the noise, ensuring consistency between the analysis of the MD trajectories and the generation of new projected trajectories. Secondly, our method is based on a maximum likelihood procedure, which is well justified from a mathematical perspective. In particular, instead of estimating a non-parametric kernel which is then parametrized (as e.g. in Volterra-based approaches), the parametric model is directly fitted on the data; this should limit the accumulation of errors. Thirdly, we do not enforce equilibrium conditions (such as the fluctuation dissipation theorem) on the model, so that the present approach offers the possibility to investigate non-
equilibrium systems. Finally, the present approach readily applies to multidimensional CVs and corresponding matrix memory kernels.

The maximum likelihood approach offers a versatile strategy to implement various extended Markovian models, which could be extended in particular to position-dependent generalized Langevin equations and higher order discretization schemes. Overall, the present work provides an efficient way to generate reduced dynamical models for multidimensional collective variables, with the same memory kernels as the underlying complex system, enabling the accurate sampling of the long-time dynamics of the latter at a dramatically reduced computational cost.

We thank Sara Bonella, Michele Casula, Arthur France-Lanord, Marco Saïta, Mathieu Salanne, and Rodolphe Vuilleumier for fruitful discussions within the MAESTRO collaboration.

Figure 4. Lennard-Jones fluid: two solutes in an explicit solvent. In all panels, results are shown for the reference MD trajectory (dark blue line) and for trajectories generated by the estimated Markov model (dashed cyan lines) and EM models with 3 to 6 hidden dimensions (from dark green to purple). Unless specified, all quantities are in LJ units. (a) Free energy (in units of the thermal energy, $k_B T$) estimated from histograms of the distance $r$ between the two solutes; the various cases are shifted by physically irrelevant constants for clarity. (b) Memory kernel compared and (c) velocity autocorrelation function; the inset in panel (b) shows a zoom on intermediate times. (d) Mean first passage time to reach $r = 2^{1/6}$ and ending at $r = 2$ (the inset shows a zoom on the tail of the distributions, on semi-logarithmic scale).

---

[1] J. T. Hynes, Annual Review of Physical Chemistry 36, 573 (1985), https://doi.org/10.1146/annurev.pc.36.100185.003041.
[2] J. P. Bergsma, B. J. Gertner, K. R. Wilson, and J. T. Hynes, The Journal of Chemical Physics 86, 1356 (1987), https://doi.org/10.1063/1.452224.
[3] L. Bocquet and J. Piascki, Journal of Statistical Physics 87, 1005 (1997).
[4] J. O. Daldrop, J. Kappler, F. N. Brünig, and R. R. Netz, Proceedings of the National Academy of Sciences 115, 5169 (2018).
[5] W. Min, G. Luo, B. J. Cherayil, S. C. Kou, and X. S. Xie, Physical Review Letters 94, 198302 (2005).
[6] S. Kheifets, A. Simha, K. Melin, T. Li, and M. G. Raizen, Science (2014).
[7] M. Lysy, N. S. Pillai, D. B. Hill, M. G. Forest, J. W. R. Mellnik, P. A. Vasquez, and S. A. McKinley, Journal of the American Statistical Association 111, 1413 (2016).
[8] B. G. Mitterwallner, C. Schreiber, J. O. Daldrop, J. O. Rädler, and R. R. Netz, Physical Review E 101, 032408 (2020).
[9] R. Zwanzig, Nonequilibrium Statistical Mechanics (Oxford University Press, 2001).
[10] A. J. Chorin, O. H. Hald, and R. Kupferman, Proceedings of the National Academy of Sciences 97, 2968 (2000).
[11] A. J. Chorin, O. H. Hald, and R. Kupferman, Physica D: Nonlinear Phenomena 166, 239 (2002).
[12] L. Ma, X. Li, and C. Liu, The Journal of Chemical Physics 145, 204117 (2016).
[13] S.-H. Chung and M. Roper, Biophysical Reviews and Letters 14, 171 (2019).
[14] E. Darve, J. Solomon, and A. Kia, Proceedings of the National Academy of Sciences 106, 10884 (2009).
[15] S. Izvekov, The Journal of Chemical Physics 138, 134106 (2013).
[16] R. Zwanzig, Journal of Statistical Physics 9, 215 (1973).
[17] H. Mori, Progress of Theoretical Physics 34, 399 (1965).
[18] H. Mori, Progress of Theoretical Physics 33, 423 (1965).
[19] F. Glattzel and T. Schilling, arXiv:2107.01111 [cond-mat] (2021), arXiv:2107.01111 [cond-mat].
[20] T. J. Doerries, S. A. M. Loos, and S. H. L. Klapp, arXiv:2010.12806 [cond-mat] (2020), arXiv:2010.12806 [cond-mat].
[21] M. Berkowitz, J. D. Morgan, D. J. Kouri, and J. A. McCammon, The Journal of Chemical Physics 75, 2462 (1981).
[22] B. J. Berne and G. D. Harp, in Advances in Chemical Physics (John Wiley & Sons, Ltd, 1970) pp. 63–227.
[23] B. J. Berne, M. E. Tuckerman, J. E. Straub, and A. L. R.
The transition probability density in Eq. (4) of the main text is

\[
\pi(X_{k+1}|X_k) = \mathcal{N}\left( \begin{pmatrix} v_{k+1} \\ h_{k+1} \end{pmatrix}; \mu_{\Delta t}(X_k) + \sigma_{\Delta t} G_v, \sigma_{\Delta t} G_h \right) \delta \left( v_k - \frac{x_{k+1} - x_k}{\Delta t} \right)
\]
where
\[
\mathcal{N}(y; \mu, \Sigma) = \frac{e^{-\frac{1}{2}[y-\mu]^T \Sigma^{-1}[y-\mu]}}{\sqrt{2\pi}^D \det \Sigma} = \frac{e^{-\frac{1}{2} \text{Tr}[\Sigma^{-1}[y-\mu][y-\mu]^T]}}{\sqrt{2\pi}^D \det \Sigma}
\]
is a notation for a \((D\text{-variate})\) Gaussian distribution density for the variable \(y\) with mean \(\mu\) and variance \(\Sigma\). Since the presence of the Dirac function imposes the velocity as \(v_k = (x_{k+1} - x_k) / \Delta t\), we always assume this condition to be satisfied and consider in the following only the non-degenerate part of the transition probability.

From Eq. (4) in the main text, the log-likelihood of a trajectory is then given by
\[
\ln \pi_\Theta (\{X\}_{0:N}) = \ln \pi(X_0) - \frac{1}{2} \ln \left(2\pi \Delta t \sigma_{\Delta t}^T \sigma_{\Delta t}\right) - \frac{1}{2} \sum_{k=0}^{N-1} \text{Tr} \left[ \left(\sigma_{\Delta t} \sigma_{\Delta t}^T\right)^{-1} \left(X_{k+1} - \mu_{\Delta t}(X_k)\right) \left(X_{k+1} - \mu_{\Delta t}(X_k)\right)^T \right].
\]

**M step**

Our ultimate objective is to maximize, with respect to the parameters \(\Theta = \{F, A_vv, A_{vh}, A_{hh}, \sigma_vv, \sigma_{vh}, \sigma_{hh}, (h_0)\}\), the log-likelihood of the observed trajectory \(\{\tilde{X}\}_{0:N}\) given by
\[
\ln L(\Theta) = \ln \int \pi_\Theta (\{X\}_{0:N}) d\{h\}_{0:N} = \ln \int \pi_\Theta (\{h\}_{0:N} | \{\tilde{X}\}_{0:N}) \pi_\Theta (\{\tilde{X}\}_{0:N}) d\{h\}_{0:N}.
\]

Using a convexity inequality, this can be shown to be lower bounded by the evidence lower bound (Eq. (5) in the main text) \[50\]
\[
L_{LB}^I(\Theta) = \int \pi_{\Theta_j}(\{h\}_{0:N} | \{\tilde{X}\}_{0:N}) \ln \pi_\Theta (\{X\}_{0:N}) d\{h\}_{0:N} = \int \pi_{\Theta_j}(\{h_0\} | \{\tilde{X}\}_{0:N}) \ln \pi_\Theta(X_0) d\{h_0\} + \sum_{k=0}^{N-1} \int \pi_{\Theta_j}(h_k, h_{k+1} | \{\tilde{X}\}_{0:N}) \ln \pi_\Theta(X_{k+1} | X_k) d\{h_k, dh_{k+1}\}.
\]

From the transition probability (7), the evidence lower bound is then given by
\[
L_{LB}^I(\Theta) = \ln \pi(X_0) - \frac{1}{2} \ln \left(2\pi \Delta t \sigma_{\Delta t}^T \sigma_{\Delta t}\right) - \frac{1}{2} \sum_{k=0}^{N-1} \text{Tr} \left[ \left(\sigma_{\Delta t} \sigma_{\Delta t}^T\right)^{-1} \left(v_{k+1} \right) \left(h_{k+1}\right)^T - 2 \left(\left(v_{k+1} \right) \left(h_{k+1}\right)^T \right) \right]
\]
where all averages are with respect to \(\pi_{\Theta_j}(h_k, h_{k+1} | \{\tilde{X}\}_{0:N})\). Since \(L_{LB}^I\) has the form of the log-likelihood of a Gaussian distribution, a standard maximum likelihood procedure \[57, 58\] can be used to optimize it and estimate \(\Theta_{j+1}\).

**E-step**

The prediction-correction part of the E-step computes iteratively, forwards and for the whole trajectory, the probability distribution of hidden variables conditioned on the past trajectory of the visible variables as
\[
\pi \left(h_k | \{\tilde{X}\}_{0:k}\right) = \frac{f \pi \left(h_{k-1} | \{\tilde{X}\}_{0:k-1}\right) \pi \left((\tilde{X}, h)_k | (\tilde{X}, h)_{k-1}\right) dh_{k-1}}{\pi \left(\tilde{X}_k | \{\tilde{X}\}_{0:k-1}\right)}.
\]

This is followed by the Rauch-Tung-Striebel smoother part of the E-step, which makes use iteratively, backwards and for the compete trajectory, of:
\[
\pi \left(h_k, h_{k+1} | \{\tilde{X}\}_{0:N}\right) = \pi \left(h_k | h_{k+1}, \{\tilde{X}\}_{0:N}\right) \pi \left(h_{k+1} | \{\tilde{X}\}_{0:N}\right).
\]
The second term is the marginal of the previous iteration (from time step $k + 2$ to $k + 1$)

$$
\pi (h_{k+1}|\{X\}_{0:N}) = \int \pi (h_{k+1}, h_{k+2}|\{X\}_{0:N}) \, dh_{k+2}
$$

whereas the first is obtained using the result of the prediction-correction part of the E-step for time step $k$ and the transition probability distribution as

$$
\pi (h_k|h_{k+1}, \{X\}_{0:k}) \propto \frac{\pi ((\tilde{X}, h)_{k+1}|h_k, \{X\}_{0:k}) \pi (h_k|\{X\}_{0:k})}{\int \pi ((\tilde{X}, h)_{k+1}|h_k, \{X\}_{0:k}) \pi (h_k|\{X\}_{0:k}) \, dh_k} \pi (h_{k+1}|\{X\}_{0:N}) \, dh_{k+1}.
$$

The E-step is illustrated in Fig. 5: here, we sample a trajectory $\{X\}_{0:N}$ with known parameters $\Theta$ (with a single auxiliary variable, i.e. $d_h = 1$), and our goal is to reconstruct the law $\pi_{\Theta}(\{h\}_{0:N}|\{X\}_{0:N})$ of the trajectory of the hidden variable, using only the trajectory of the observed variables $\{\tilde{X}\}_{0:N}$. This conditional law, represented in Fig. 5 by its mean $k \mapsto \langle h_k \rangle$ and twice its standard deviation (blue area), is concentrated on the original realization.

![Figure 5](image.png)

**Figure 5.** Estimation by the E-Step of the law of the trajectory of a single hidden variable ($d_h = 1$), $\pi_{\Theta}(h_{k:0:N}|\{\tilde{X}\}_{0:N})$, reconstructed using only the trajectory of the observed variable $\{\tilde{X}\}_{0:N}$ for the one-dimensional case ($d = 1$) with a quadratic potential $V(x) = x^2/2$ discussed in the main text (with more hidden variables). The figure shows the original trajectory of the hidden variable (red line) and the conditional law, represented by its mean $k \mapsto \langle h_k \rangle$ (blue line) and twice its standard deviation (shaded area).

**Details of the examples**

**MD simulation details for the LJ dimer**  The dynamics is integrated with a time step of $\Delta t_{MD} = 0.001$ (in LJ units) in the NVE ensemble using the LAMMPS simulation package [56]. We run 20 trajectories of $10^6$ timesteps and CV values are extracted every 2 steps.

**EM convergence**  Initial values of the parameters are taken randomly. For all examples, we stop the EM iterations if the difference of log-likelihood between two EM steps is less than $10^{-8}$ or if the number of EM steps exceeds 2000.

**Mean first passage time estimation**  The FPT is estimated for molecular dynamics by constraining the initial position with restraints using PLUMED. 2000 trajectories per restrained position are launched. Kernel estimates of the mean first passage time as well as the FPT density are then obtained conditioned on the realized starting position. The FPT from the fitted models are computed using 1500 trajectories per initial value of the distance and kernel estimate is also used to estimate the MFPT and the FPT density.