The Journal of Chemical Physics

RESEARCH ARTICLE | OCTOBER 26 2023

Inferring free-energy barriers and kinetic rates from molecular dynamics via underdamped Langevin models FREE

David Daniel Girardier ; Hadrien Vroylandt ; Sara Bonella ; Fabio Pietrucci

Check for updates

J. Chem. Phys. 159, 164111 (2023)
https://doi.org/10.1063/5.0169050

View Online
Export Citation
CrossMark

06 November 2023 13:00:22

AIP Publishing

The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Inferring free-energy barriers and kinetic rates from molecular dynamics via underdamped Langevin models

Cite as: J. Chem. Phys. 159, 164111 (2023); doi: 10.1063/5.0169050
Submitted: 22 July 2023 • Accepted: 27 September 2023 •
Published Online: 26 October 2023

View Online Export Citation CrossMark

David Daniel Girardier,[1] Hadrien Vroylandt,[2] Sara Bonella,[3] and Fabio Pietrucci[1,a)]

AFFILIATIONS

[1] Sorbonne Université, Musée National d'Histoire Naturelle, UMR CNRS 7590, Institut de Minéralogie, de Physique des Materiaux et de Cosmochimie, IMPMC, F-75005 Paris, France
[2] Sorbonne Université, Institut des Sciences du Calcul et des données, ISCD, F-75005 Paris, France
[3] Centre Européen de Calcul Atomique et Moléculaire (CECAM), Ecole Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland

[a)] **Author to whom correspondence should be addressed:** fabio.pietrucci@sorbonne-universite.fr

ABSTRACT

Rare events include many of the most interesting transformation processes in condensed matter, from phase transitions to biomolecular conformational changes to chemical reactions. Access to the corresponding mechanisms, free-energy landscapes and kinetic rates can in principle be obtained by different techniques after projecting the high-dimensional atomic dynamics on one (or a few) collective variable. Even though it is well-known that the projected dynamics approximately follows – in a statistical sense – the generalized, underdamped or overdamped Langevin equations (depending on the time resolution), to date it is nontrivial to parameterize such equations starting from a limited, practically accessible amount of non-ergodic trajectories. In this work we focus on Markovian, underdamped Langevin equations, that arise naturally when considering, e.g., numerous water-solution processes at sub-picosecond resolution. After contrasting the advantages and pitfalls of different numerical approaches, we present an efficient parametrization strategy based on a limited set of molecular dynamics data, including equilibrium trajectories confined to minima and few hundreds transition path sampling-like trajectories. Employing velocity autocorrelation or memory kernel information for learning the friction and likelihood maximization for learning the free-energy landscape, we demonstrate the possibility to reconstruct accurate barriers and rates both for a benchmark system and for the interaction of carbon nanoparticles in water.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0169050

06 November 2023 13:00:22

I. INTRODUCTION

Several important phenomena, known as activated processes, are characterized by rare transitions between states corresponding to different values of some observable. Examples include chemical reactions, phase transitons, protein conformational changes. Computer simulations, particularly molecular dynamics (MD), have proved an important tool to complement experiments in the study of such rare events.

In typical MD simulations, condensed-matter systems formed by thousands or more atoms evolve in time via numerical integration of Hamilton's equations of motion, often supplemented by a thermostat in order to sample canonical averages.[1] Due to the chaotic nature of the dynamics, the complexity of the potential energy landscape and the number of degrees of freedom, it is essentially impossible to analyze and understand the system's behaviour without recourse to some form of coarse-graining in configuration space and in time. Microstates are thus lumped together into macrostates (e.g., reactant vs product, or liquid vs crystal, etc.) carrying relevant physical or chemical information. The characterization of the macrostates is performed via different mathematical tools, including those broadly referred to as "machine learning," and it is usually based upon similarity metrics (used for clustering) or upon order parameters, often referred to as collective variables (CVs).[2–4] The latter are typically functions of the atomic coordinates and can both resolve different macrostates and

Published under an exclusive license by AIP Publishing

parametrize steps along the progressive transformation between them.

In the context of activated processes, CVs are especially useful in combination with enhanced sampling techniques – that mitigate the timescale limitations of brute-force MD – to reconstruct mechanisms, free-energy landscapes and kinetic rates.[5] It has been demonstrated[6,7] that the dynamics projected on the space of CVs is approximated by Langevin equations, where the free energy gradient – i.e., the mean force – plays the role of the conservative force, supplemented with two non-conservative forces, a friction and a random noise. These two terms effectively model the influence of all the overlooked degrees of freedom on the CV dynamics. Langevin equations have all the advantages of a parsimonious, physically interpretable model. When optimally constructed (as explained in the following), these stochastic differential equations mimic the behavior of the projected Hamiltonian trajectories in a statistical sense, i.e. they can quantitatively reproduce averages, distributions, correlation functions and, in particular, kinetic rates of the original system. The small dimensionality of CV space results in a strongly reduced computational cost compared to Hamilton's equations.

In spite of these well-known advantages, some conceptual and practical difficulties still make the exploitation of Langevin models as quantiative tools to access the thermodynamic and kinetic properties of activated processes non-trivial.

Firstly, several different mathematical forms of Langevin equations are available. The most complex is the generalized Langevin equation (GLE), that features an inertial term together with non-Markovian, history-dependent friction and random force terms and is expected to be accurate at the smaller timescales, e.g., in the range $\delta t \sim 1 - 100$ fs for small solutes and chemical reactions in water.[8] For coarser time resolutions, the projected dynamics becomes amenable to a Markovian description, leading to the underdamped Langevin equation (ULE) that contains a time-independent friction, γ, and is expected to be valid at intermediate timescales, e.g., $\delta t \sim 100 - 10^3$ fs for small-to-medium molecules in water. For even slower time scales, $\delta t \gg \gamma^{-1}$, the overdamped Langevin equation (OLE), in which the inertial term is neglected, is customarily employed. These equations are well-known and their properties, including the transition between different forms depending on the problem and timescales under investigation, have been studied both formally and in simulations of model systems. However, in our experience, these analyses sometimes fail to capture the behavior of realistic models of condensed-matter systems, indicating that the choice of a particular form of the Langevin equation remains delicate and should be validated, case-by-case, against numerical evidence based on Hamiltonian trajectories.

Secondly, the Langevin equations contain a set of parameters or functions (effective mass, friction, and free energy landscape) that also need to be inferred from the underlying Hamiltonian evolution. Considerable work has been devoted to this aim, with numerous approaches put forward, in particular, for the GLE (often focused on extracting the non-Markovian friction)[8–24] and the OLE.[25–33] We emphasize the importance of developing parametrization schemes that make efficient use of computational resources: as an example (in the overdamped case), the recent likelihood-based technique of Ref. 33 exploits a small set of ~100 short trajectories relaxing from a barrier top to reconstruct at once the free-energy and diffusion profiles as well as kinetic rates.

Comparatively fewer approaches are available for the accurate parametrization of ULE models from MD trajectories, notwithstanding several important contributions.[11,34–41] It is fair to say that MD-based ULE models of CV dynamics are far from being a widespread tool for activated processes.

This is unfortunate because ULEs play an important role, as they address a time-resolution window relevant for many physico-chemical phenomena, while retaining a formal structure simpler than the one of GLEs. As pointed out in recent works[42,43] a non-trivial difficulty resides in the fact that customary definitions of finite-difference velocities in CV-space cause spurious velocity-related correlations (see also the insightful discussion of timescales in Ref. 39). As a result, the paramterization of accurate ULE models is far from straightforward.

In this work we explore both issues, the choice of the most suitable form of Langevin equation and its parametrization – in particular for the ULE – pointing out some difficulties related to the transition from model to realistic systems, and showcasing some interesting results that pave the way for future more broad applications. We start by proposing an assessment - based on MD data - of the boundaries between non-Markovian and underdamped, and between underdamped and overdamped behavior. This assessment relies on careful analysis of the memory kernel in the GLE, and on the study of different correlation functions of the CVs. We then propose a relatively inexpensive procedure for ULEs parametrization and validate it based on tests on benchmark Langevin systems as well as on MD trajectories of complex systems projected on a CV.

In the process, we quantify the accuracy of different approximate calculation schemes for effective mass, friction, free energy landscape, and rates in the context of ULEs. The results show that our approach allows to reconstruct fairly accurate free-energy profiles and kinetic rates also for rare events characterised by high barriers, starting from a limited amount of short unbiased MD trajectories.

II. THEORETICAL METHODS

A. Time resolution-dependent Langevin approximations

Three forms of Langevin equations are customarily considered in applications to model the dynamics of high-dimensional Hamiltonian systems after projection on a low-dimensional CV space. In this work we assume, for simplicity, that only a scalar CV function of the Cartesian coordinates of the N_{at} atoms in the system, $r \equiv \{r_i\}_{i=1,\dots,N_{at}}$, is employed. We note this CV as $q(r)$. The most suitable form of the Langevin equation depends, in general, (i) on the specific system studied, (ii) on the CV employed, and, crucially, (iii) on the the time resolution δt used to analyze the trajectory $\{q(t_k)\}_{k=1,\dots,M}$, $t_k = k\delta t$.

The first form is the GLE, non-Markovian, typically (but not always[44,45]) written as

$$\begin{cases} \dot{q} = v \\ \dot{v} = -\dfrac{1}{m}\dfrac{dF}{dq} - \displaystyle\int_0^t dt' K(t') v(t-t') + R(t) \end{cases} \tag{1}$$

where v is the velocity, m is an effective mass (see below), $F(q) = -k_b T \log \rho_{eq}(q)$ is the free energy profile (potential of mean

06 November 2023 13:00:22

Published under an exclusive license by AIP Publishing

force) corresponding to the equilibrium density $\rho_{eq}(q)$, $K(t)$ is a history-dependent friction kernel, and $R(t)$ is a random force (noise), related to friction via the fluctuation-dissipation theorem[46] (for a more extended discussion on the validity of the latter theorem see Refs. 7 and 47). This is one of the most general forms of Langevin equation, as it can be derived from the Hamiltonian dynamics using a projection operator formalism.[48] In this equation, both the memory kernel and the random force represent the action of the environment, i.e. the projected out coordinates, such that the time dependence of the memory kernel is representative of the main timescales of this environment.

The friction kernel K, and thus the random force R, can in principle vary also with the position, depending on the coupling between CV and bath.[49–52] However, we consider for simplicity the case of position-independent friction in the following.

When the timescale for the decay of the friction kernel (the memory duration) is small with respect to δt (and with respect to the velocity autocorrelation timescale, *vide infra*), the dynamics is well described by an ULE. The latter equation, also known as Kramers equation, can be written

$$\begin{cases} \dot{q} = v \\ \dot{v} = -\dfrac{1}{m}\dfrac{dF}{dq} - \gamma v + \sqrt{2c}\eta(t) \end{cases} \tag{2}$$

where γ is the time-independent friction, replacing the integral over the memory kernel (see below for more details about their relationship). By imposing the fluctuation-dissipation theorem, which guarantees canonical equilibrium averages in the long run, one finds $c = k_B T\gamma/m$, corresponding to a diffusion coefficient $D = k_B T/m\gamma = c/\gamma^2$. η is an uncorrelated Gaussian white noise with zero mean and unit variance (Wiener process).

For even larger time resolution δt, such that $\gamma\delta t \gg 1$, the CV velocity equilibrates quickly with the environment and a simpler – so-called overdamped – Langevin equation can give a good approximation of the projected dynamics:

$$\dot{q} = -\frac{D}{k_B T}\frac{dF}{dq} + \sqrt{2D}\eta(t) \tag{3}$$

We remark that, in the case of barrier-crossing processes, the aim of tracking the dynamics of the system along the transformation pathway imposes constraints on the choice of δt, that needs to be much smaller than the typical duration of a MD trajectory connecting reactants to products (such duration ranges from tens of femtoseconds in chemical reactions to hundreds of nanoseconds in ice nucleation).

As mentioned in the introduction, in this paper we focus on ULEs. The discussion above suggests that the lower and upper limits of δt bracketing the expected validity of the underdamped approximation can be estimated from equilibrium MD trajectories.

As for the lower limit, determining the transition from GLEs to ULEs, the key property to monitor is the temporal extent of the non-Markovian memory effects in friction and noise. The underdamped Langevin equation should in fact be a good approximation of the true dynamics *at least* for time resolutions δt larger than the memory. As an example, in Fig. 1 we show the memory kernel, $K(t)$, for a fullerene dimer in water with q as the distance between centers of mass (see Results for the system's details).

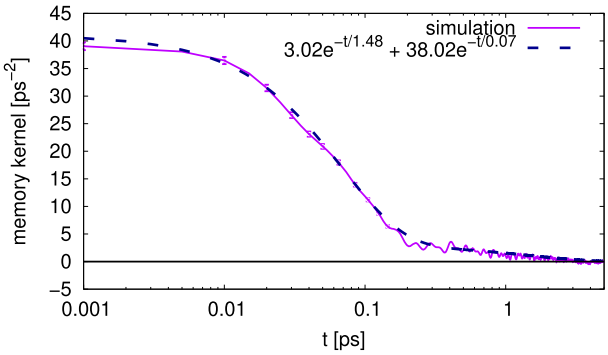

FIG. 1. Memory kernel of the GLE for the fullerene dimer in water. The kernel is computed using 40 trajectories of 250 ps with timestep $\tau = 0.001$ ps. All trajectories remain in the free-energy minimum (associated complex). Error bars are computed using bootstraps, averaging repeatedly over different random sets of trajectories and computing the standard deviation of the results. We also plot as dashed line a fit of the kernel by the double exponential.

The memory kernel was computed by inverting a Volterra equation built on the velocity autocorrelation function,[9,52–55]

$$\left\langle v(0)\left(\dot{v}(t) + \frac{1}{m}\frac{dF}{dq}(q(t))\right)\right\rangle = -\int_0^t dt' K(t')\langle v(0)v(t-t')\rangle \tag{4}$$

For the numerical inversion we use the trapezoidal method of Ref. 56 implemented in the VolterraBasis package (available at https://github.com/HadrienNU/VolterraBasis). Note that the estimation of the memory kernel is a delicate problem and a critical step for a faithful description of the dynamics via generalized Langevin equations. Besides the approach adopted here, several alternative methods are available,[16,17,57] notably based on likelihood maximization,[22,58] Laplace transforms,[14] or velocity autocorrelation functions.[19,59,60] These approaches are based on an expansion of the memory kernel as a sum of few exponential terms (Prony series), such that the environment is represented by a small number of characteristic timescales.

The upper limit of δt, enabling the transition from ULEs to OLEs, on the other hand, corresponds to $\approx\gamma^{-1}$, the characteristic friction (or inertial) timescale, that, in turn, can be estimated from the timescale for the decay of autocorrelations in the CV position or velocity. In fact, this is the time necessary to observe relaxation to the equilibrium distribution of an initially perturbed CV velocity, and the assumption underlying the OLE is that – at the chosen resolution δt – the velocity can be considered in equilibrium throughout the transition process. In the Results section, we investigate the validity of these theoretical limits for a model system and, more interestingly, for the fullerene dimer. Figure 2 summarizes our framework for the choice of a suitable GLE/ULE/OLE model based on the estimation of characteristic timescales of the system as well as on the choice of a resolution δt.

B. Estimation of the mass and friction of underdamped Langevin equations

Let us begin by providing two estimators for the mass appearing in the ULE that will be compared in the applications. The first is

06 November 2023 13:00:22

Published under an exclusive license by AIP Publishing

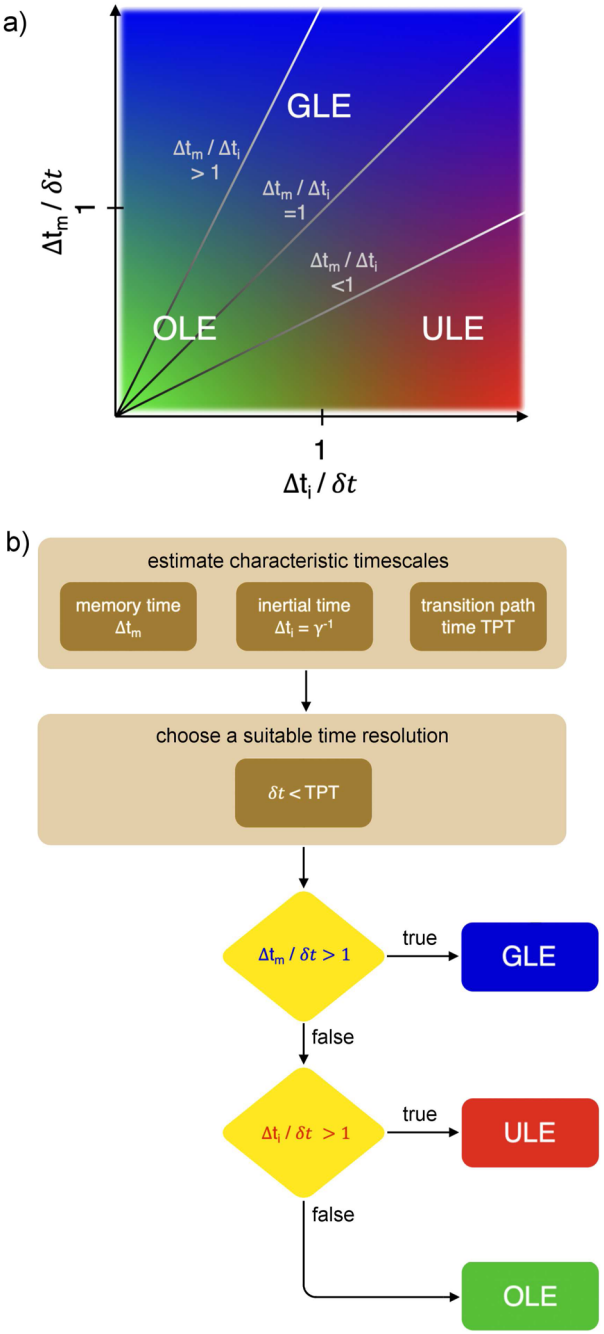

FIG. 2. (a) Schematic diagram showing where generalized, underdamped or overdamped Langevin models are expected to approximate well the projected high-dimensional dynamics. The pertinent parameters are the memory time Δt_m (time range of non-Markovian effects in the memory kernel), the inertial time $\Delta t_i = \gamma^{-1}$ (relaxation time of the CV velocity to equilibrium), the transition path time TPT (typical duration of reactive trajectories). The resolution δt of the Langevin model is a free parameter, determining the model location on the line with inclination $\Delta t_m/\Delta t_i$ characterizing a given system. b) Algorithmic view of the steps leading to the choice and parametrization of a Langevin model. For $\delta t <$ TPT the model is able to track transition paths.

based on the computation of the average force from Hamiltonian dynamics,[11,52,61,62] the effective mass corresponding to the CV q can be obtained as

$$m = \left\langle \sum_{i=1}^{N} \frac{1}{m_i} \sum_{\alpha=x,y,z} \left(\frac{\partial q}{\partial r_{i\alpha}} \right)^2 \right\rangle^{-1} \tag{5}$$

where m_i are the atomic masses and $\partial q/\partial r_{i\alpha}$ the derivative of the CV with respect to the atomic coordinate.

When, as in the applications considered in this work, the mass is position-independent (thus avoiding additional force terms in the Langevin equation[62]), the mass can also be estimated via the equipartition theorem applied to the CV velocity without the need to restrict to a local average with respect to q:[63]

$$m = \frac{k_B T}{\langle v^2 \rangle} \tag{6}$$

(see supplementary material for the theoretical relation between this and the previous expression). Note, however, that the velocity v appearing in the prevous formula is, typically, obtained via finite-differences from the q displacement at coarse resolution δt: for instance, in this work we adopt the convention $v(t_k) = [q(t_{k+1}) - q(t_{k-1})]/2\delta t$. As a result, the value of m depends on δt, and it coincides with the mass in Eq. (5) only for the finest time resolution, as shown numerically in the Results section.

We now focus on the estimation of the Markovian friction coefficient γ: we consider first estimators based on approximations of the non-Markovian friction in the GLE. The simplest approximation can be obtained in cases where $v(t)$ does not vary significantly over the time scale t_0 of the decay of $K(t)$. Here the velocity can be simply taken out of the integral so that $\gamma \approx \int_0^{t_0} K(t')dt'$. An alternative estimator can be derived by observing that, in essence, the underdamped approximation consists in assuming that the non-Markovian friction $-\int_0^t dt' K(t')v(t-t')$ is well-reproduced by the Markovian expression $-\gamma v(t)$. This opens the possibility to estimate an effective time-independent friction directly from a linear regression. Both approaches have been tested in the Results section.

On the other hand, γ can also be estimated from the following alternative approach that does not involve the memory kernel. A trajectory confined in a free-energy minimum is expected to yield position and velocity autocorrelation functions that approximately follow – for weak anharmonicity – the analytic Ornstein-Uhlenbeck forms. For a harmonic potential $F(q) = \frac{1}{2} m\omega_0^2 q^2$ the ULE autocorrelation functions are[64,65]

$$C_q(t) \equiv \frac{\langle q(0)q(t) \rangle}{\langle q^2 \rangle} = e^{-\frac{\gamma t}{2}} \left[\cos(\omega_1 t) + \frac{\gamma}{2\omega_1} \sin(\omega_1 t) \right] \tag{7}$$

$$C_v(t) \equiv \frac{\langle v(0)v(t) \rangle}{\langle v^2 \rangle} = e^{-\frac{\gamma t}{2}} \left[\cos(\omega_1 t) - \frac{\gamma}{2\omega_1} \sin(\omega_1 t) \right] \tag{8}$$

where $\omega_1 = \sqrt{\omega_0^2 - \gamma/4}$. Assuming that the motion in one of the metastable states can be approximated as harmonic, a numerical estimate of γ can be obtained by computing the autocorrelation above from MD and extracting the value of the friction from a fit.

06 November 2023 13:00:22

Published under an exclusive license by AIP Publishing

Instead of relying on the approximate harmonicity of the system's metastable states, in principle, a sufficiently-stiff parabolic bias potential could be added at selected positions in q-space, opening also the possibility to estimate the position dependence of γ by obtaining local harmonic fits as described above and scanning an opportune range of qs. However, as analyzed in Ref. 66 as well as in the specific case of the fullerene dimer in supplementary material, such bias potentials significantly modify the unperturbed value of γ and it is unclear how to remove such bias.

C. Inference of the free energy landscape

Having addressed the estimation of mass and friction in Sec. II B, the only remaining parameters for the construction of an optimal Langevin model are those determining the shape of the free energy landscape. In a similar way to the maximum-likelihood approach based on underdamped models of Ref. 33, we attempt to reconstruct the free energy using short MD trajectories spanning the transition region, as typically provided by committor analysis or transition path sampling techniques. This allows keeping an acceptable computational cost despite large barriers that render brute-force MD unfeasible, and it also avoids the use of biasing forces that complicate post processing.

We base our method on the likelihood to obtain the MD trajectories from the model, *i.e.* the transition probability for the whole trajectory. In the case of the ULE, the joint variation in both CV position and velocity between adjacent trajectory frames represents Markovian steps, so that the likelihood function is built as the product of short-time propagators

$$\mathcal{L}_\theta = \prod_{k=0}^{M-1} \mathcal{P}_\theta(q_{k+1}, v_{k+1}|q_k, v_k) \tag{9}$$

In the equation above, θ is the set of model parameters, which, in principle, include mass, friction, and free-energy landscape. The exact expression for the propagator is, however, not available and this quantity is determined via approximations,[67] constructed either from formal properties of the Fokker-Planck equation or starting from a specific form of numerical Langevin integrator. In the following, we consider both options and discuss the relationship between the resulting expressions.

Firstly, we follow the approach in Ref. 68, based on a cumulant generating-function expansion, that approximates the propagator up to the fourth order in the time-separation δt between frames with a simple Gaussian form:

$$\mathcal{P}_\theta(q', v'|q, v) = \frac{1}{\sqrt{4\pi^2 \det(M)}} \exp\left[-\frac{M_{qq}}{2\det M}(v' - \langle v\rangle)^2 - \frac{M_{vv}}{2\det M}(q' - \langle q\rangle)^2 + \frac{M_{qv}}{\det M}(q' - \langle q\rangle)(v' - \langle v\rangle)\right] \tag{10}$$

here averages $\langle q\rangle$, $\langle v\rangle$ and variances $M_{qq} = \langle q^2\rangle - \langle q\rangle^2$, $M_{qv} = \langle qv\rangle - \langle q\rangle\langle v\rangle$, $M_{vv} = \langle v^2\rangle - \langle v\rangle^2$ are computed on the trajectory and depend on the model parameters θ. We note that this approximation was also derived in Ref. 69, that studies its convergence.

Following Ref. 68, we seek the average of the dynamical variables $\Gamma = \{q, v, q^2, qv, v^2\}$; indicating $A(\Gamma, t) = \langle\Gamma(t)\rangle$, their time evolution is described by a backward Fokker-Planck equation

$$\partial_t A(\Gamma, t) = L^+ A(\Gamma, t) \tag{11}$$

whereas the probability density $\rho(q, v, t)$ is the solution of the Klein-Kramers form of the Fokker-Planck equation

$$\partial_t \rho = L\rho \equiv -\partial_q(v\rho) - \partial_v\left[\left(-\frac{1}{m}\frac{dF}{dq} - \gamma v\right)\rho\right] + \partial_v^2\left(\frac{\gamma}{m\beta}\rho\right) \tag{12}$$

In the equations above, the Fokker-Planck operator L and its "backward" counterpart L^+ (i.e., the generator of the stochastic process) are given by

$$L = -a\partial_q - b\partial_v + c\partial_v^2 \tag{13}$$

$$L^+ = a\partial_q + b\partial_v + c\partial_v^2 \tag{14}$$

where we set for simplicity $a \equiv v$, $b \equiv -\frac{1}{m}\frac{dF}{dq} - \gamma v$, $c \equiv \gamma/m\beta$. Note that the absence of ∂_q^2 and ∂_{qv}^2 in the previous formulas is related to the non-invertibility of the diffusion matrix, so that the Gaussian form of the propagator is valid only when going beyond the second order approximation in the time displacement δt: limiting ourselves to the third order $\langle\Gamma\rangle = A_0 + A_1\delta t + A_2\delta t^2 + A_3\delta t^3 + \mathcal{O}(\delta t^4)$, the moments are computed with the help of a recursive formula[68]

$$(j+1)A_{j+1}(\Gamma) = L^+ A_j(\Gamma) \tag{15}$$

After some tedious but straightforward steps, the cumulants are thus obtained:

$$\langle q\rangle = q + v\delta t + \frac{1}{2}b\delta t^2 + \frac{1}{6}(b_q v + b b_v)\delta t^3 \tag{16}$$

$$\langle v\rangle = v + b\delta t + \frac{1}{2}(b_q v + b b_v)\delta t^2 + \frac{1}{6}(b_{qq}v^2 + b_q b_v v + 2b b_{qv} v + b b_q + b b_v^2 + 2c b_{qv})\delta t^3 \tag{17}$$

$$M_{qq} = \langle q^2\rangle - \langle q\rangle\langle q\rangle = \frac{2}{3}c\delta t^3 \tag{18}$$

$$M_{vv} = \langle v^2\rangle - \langle v\rangle\langle v\rangle = 2c\delta t + (c_q v + 2c b_v)\delta t^2 + \frac{1}{3}(c_{qq}v^2 + 2c_q b_v v + 4c b_{qv} v + c_q b + 2c b_q + 4c b_v^2)\delta t^3 \tag{19}$$

$$M_{qv} = \langle qv\rangle - \langle q\rangle\langle v\rangle = c\delta t^2 + \frac{1}{3}(c_q v + 3c b_v)\delta t^3 \tag{20}$$

where subscripts q and v of b and c indicate partial derivatives with respect to position or velocity. These formulas can in principle be applied in the case of position-dependent friction, even though the applications presented in this work will be constant-friction ones.

The second approach to determine the short time propagator requires to choose a numerical integrator for the ULE. Once this is done, the noise in the integrator is isolated and the propagator is written as the probability distribution of the noise.[65,70] In this work,

06 November 2023 13:00:22

Published under an exclusive license by AIP Publishing

we adopt the algorithm proposed by Vanden-Eijnden and Ciccotti in Ref. 71:

$$q_{n+1} = q_n + v_n\tau + A_n \tag{21}$$

$$v_{n+1} = v_n + \frac{1}{2}\frac{f(q_{n+1}) + f(q_n)}{m}\tau - \gamma v_n\tau + \sigma G_n - \gamma A_n \tag{22}$$

$$A_n = \frac{1}{2}\tau^2\left(\frac{f(q_n)}{m} - \gamma v_n\right) + \frac{\sigma}{2}\tau\left(G_n + \frac{1}{\sqrt{3}}\tilde{G}_n\right) \tag{23}$$

with the force $f(q) = \frac{-\partial F}{\partial q}$, $\sigma = \sqrt{2c\tau}$. This is a second-order integrator with two independent Gaussian noises G_n and $\tilde{G}_n$ with zero mean, unit variance and uncorrelated in time:

$$\langle G_i G_j\rangle = \langle \tilde{G}_i \tilde{G}_j\rangle = \delta_{ij} \quad \langle G_i \tilde{G}_i\rangle = \langle G_i \tilde{G}_j\rangle = 0 \tag{24}$$

Contrary to simpler integrators, the two noises are estimated at each time step, affecting the evolution of both the position and the velocity. Interestingly, applying the procedure described above to this integrator would lead to a propagator identical to the previous one, except for the elimination of the terms of order τ^3 in $\langle q\rangle$ and $\langle v\rangle$. We benchmarked the two alternative forms of the propagator on our systems, finding negligible differences in the reconstructed free-energy profiles, therefore in the following we neglect the latter extra terms.

Likelihood maximization is performed employing the simple Metropolis Monte Carlo algorithm of Ref. 33, where parameters are randomly varied according to an adaptive scheme that preserves the desired acceptance, as detailed in supplementary material. The code is freely available from github https://github.com/physix-repo/optLE.

Note that the procedure described in this section can, in principle, be used to estimate all parameters in the ULE, i.e. mass, friction and free energy profile. Alternatively, the mass and friction can be estimated independently, following the routes in Subsection II B, and used as input parameters in the likelihood minimization. In the following, we explore both options.

III. RESULTS

A. Analytic double-well benchmark system

To benchmark our approach we first consider a one-dimensional double well with 10 k_BT barrier, setting constant friction and mass to $\gamma = 5$ ps^{-1} and $m = 1$ nm^{-2} ps^2, respectively. Note that we specify units of space and time, even though the system is an analytic benchmark, for easier comparison with the fullerene dimer in solution (*vide infra*).

We generate 500 reference trajectories by numerically integrating the corresponding underdamped Langevin with a time step $\tau = 0.001$ ps (see Methods section), the latter being chosen based on benchmarks on temperature and diffusion coefficient. Each trajectory has a duration of 2 ps, and is initiated from the top of the barrier with initial velocity randomly drawn from the Boltzmann distribution and tracking the out-of-equilibrium relaxation into either free-energy well (see Fig. 3), mimicking committor analysis or transition path sampling data. We consider in the following 5 sets of 100 trajectories each, attempting the reconstruction of m, γ and $F(q)$ for each set separately, so that we estimate error bars by comparing the results of the 5 sets.

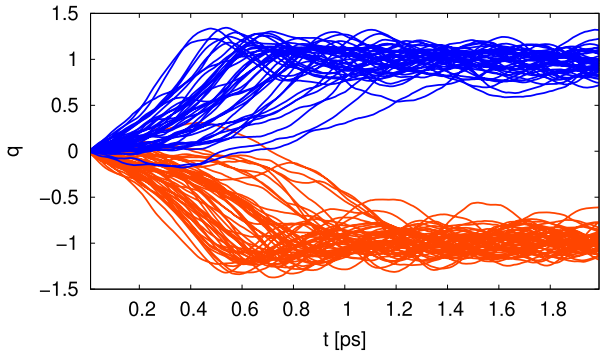

FIG. 3. Set of 100 trajectories generated with the Langevin integrator in Ref. 71 (time step $\tau = 0.001$ ps), started from the top of the barrier of the double-well potential in Fig. 6, with initial velocity randomly drawn from the Boltzmann distribution.

In order to infer the Langevin parameters from the trajectories, we analyze the latter at three different time resolutions $\delta t = 0.001$, 0.01 and 0.05 ps.

Comparing the exact mass with the one deduced from equipartition [Eq. (6)], a strong dependence on δt is evident in Fig. 4, with similar values only for $\delta t \approx \tau$. This can be understood by considering that an increase of δt has the effect of smoothing the fast variations of the noisy, fluctuating Langevin trajectories, so that $\langle v^2\rangle$ tends to be reduced – thus the effective mass increased – due to sub-sampling.

We estimate the friction by fitting the position- and velocity autocorrelation functions to the solutions of an Ornstein-Uhlenbeck process [Eq. (7)], using a trajectory started in a well (Fig. 5). As expected, since the wells are not exactly harmonic, the fit is imperfect: γ is overestimated with respect to the exact value, the deviation being however relatively small, within 7–14%. The C_v curve is used for the final estimate instead of C_q, because in the latter the error due to anharmonicity is larger.

By severely reducing the anharmonicity of the landscape through a stiff quadratic potential added to the original free-energy

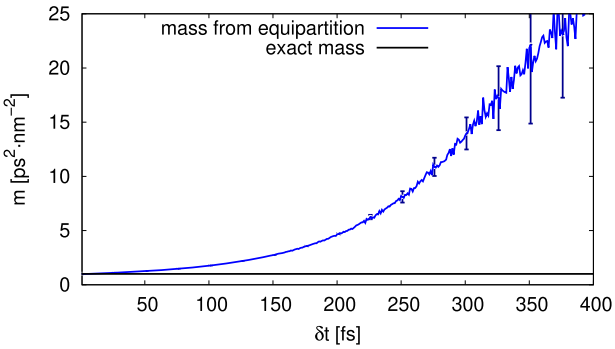

FIG. 4. Mass for the double-well benchmark system: the exact mass is compared with the effective one computed from equipartition as $m = k_BT/\langle v^2\rangle$, with the velocities estimated by finite-differences at different time resolutions δt from a 0.25 ns trajectory remaining in one well.

06 November 2023 13:00:22

Published under an exclusive license by AIP Publishing

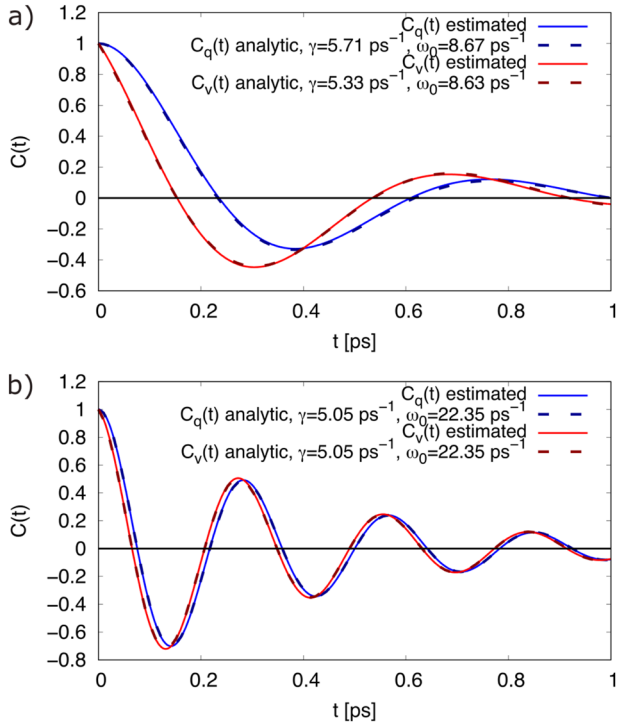

FIG. 5. Position and velocity autocorrelation functions for the double-well benchmark system, computed using a 50 ns trajectory confined in a well (a) without bias, and (b) with an additional parabolic bias $\frac{1}{2}m\omega_0^2(q-q_0)^2$ with q_0 = 1.00 nm, ω_0 = 22.36 ps^{-1} (corresponding to a Boltzmann distribution with standard deviation ≈0.045 nm). The parameters γ and ω_0 are fitted via the analytical formulas (7) and (8) in order to reproduce the trajectory data. The error bars are $<10^{-2}$ (not displayed).

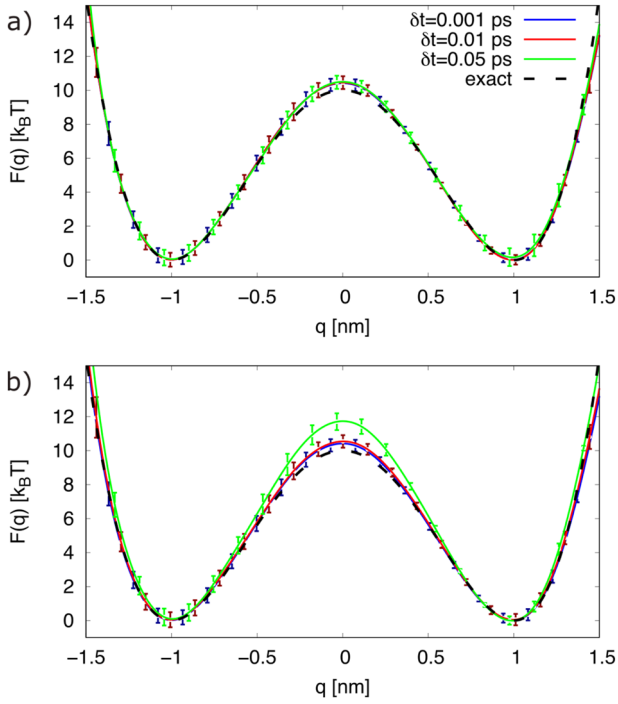

FIG. 6. Exact double-well free-energy landscape (dashed line) compared with the optimal Langevin models obtained by likelihood maximisation at different time resolutions δt, using (a) m = 1 ps^2 nm^{-2} (exact), γ = 5.33 ps^{-1} (fit at δt = 0.001 ps), (b) m, γ estimated at different δt (see supplementary material). The friction is always estimated from a fit of the v-autocorrelation in the absence of a parabolic bias.

landscape, the error on the estimated value of γ is strongly reduced, as expected, down to only 1% [Fig. 5(b)]. The unbiased and biased friction estimates will be both employed below to evaluate the effect of their respective error on barriers and rates.

We remark that in the case of the present benchmark, the exact γ is fixed *a priori* when generating the reference Langevin trajectories, therefore modifying $F(q)$ cannot affect γ; we anticipate that this does not hold in the case of MD trajectories projected on a CV q (see Sec. III B).

Finally, starting from different m, γ values as discussed above, the free-energy landscape $F(q)$ is statistically inferred by maximizing the likelihood in Eqs. (9) and (10) (with 10^6 iterations of the Monte Carlo algorithm detailed in Ref. 33, repeating 5 times the procedure and averaging the resulting profiles).

When employing the mass and (unbiased) friction estimated for the shortest resolution $\delta t = \tau$, as shown in Fig. 6(a), it is possible to infer rather precise free-energy profiles from 500 short, non-ergodic trajectories irrespective of the time resolution of the latter, at least in the range 0.001–0.05 ps. We computed mean first passage times (MFPTs) for the escape from the leftmost well by generating 100 brute-force Langevin trajectories for each set of optimized parameters, finding an overestimation of about 50% with respect to the reference MFPT $(22.1 \pm 1.0) \times 10^3$ ps corresponding to the original Langevin model with exact parameters (see Table I). The slow-down might be attributed to the combination of slightly higher barrier and slightly higher friction.

In fact, repeating the procedure employing, instead, the parabolically-biased estimate of γ, much more precise, leads to excellent agreement with the reference MFPT, with a marginally-improved free-energy barrier as well (see supplementary material).

On the other hand, estimating mass and friction from coarser time-resolution trajectories gives larger errors on the barrier and MFPT: the deterioration remains limited up to δt = 0.01 ps, while the MFPT gets significantly overestimated at δt = 0.05 ps [see Fig. 6(b) and the last column in Table I].

We note that the joint estimation of γ and $F(q)$ from likelihood maximization – by considering all of these quantities as variational parameters in Eq. (9) – does not lead to satisfactory results due to spurious correlations introduced by the finite-difference numerical approximation of the CV-velocity v (as discussed in the Introduction, see Refs. 42, 43, and 72 for similar conclusions and detailed analyses).

B. Fullerene dimer dissociation in water

We attempt to reconstruct mass, friction, free-energy landscape and kinetic rate for a realistic complex system, formed by two fullerene C_{60} molecules solvated by 2398 water molecules at

06 November 2023 13:00:22

Published under an exclusive license by AIP Publishing

TABLE I. Mean first passage time (MFPT) for the optimal Langevin models of the double-well benchmark system. Different combinations of mass and friction parameters are employed (see the main text for details). The friction is estimated from a fit of v-autocorrelation in the absence ("unbiased") or in presence ("biased") of a parabolic bias on the free-energy landscape. In the third column, the mass (nm^{-2} ps^2) is set to its exact value and friction (ps^{-1}) is estimated for the shortest $\delta t = \tau = 1$ fs. The last column represents the equipartition mass and the friction estimated at each resolution δt. The reference value corresponds to the original system (exact parameters).

		MFPT (10^3 ps)	
	δt (ps)	$m = 1, \gamma = 5.33$	$m(\delta t), \gamma(\delta t)$
Unbiased	0.001	33.4 ± 2.9	33.4 ± 2.9
	0.01	34.2 ± 2.4	36.0 ± 3.1
	0.05	37.2 ± 3.8	>50
		$m = 1, \gamma = 5.05$	$m(\delta t), \gamma(\delta t)$
Biased	0.001	22.3 ± 1.4	22.3 ± 1.4
	0.01	22.8 ± 1.7	27.5 ± 1.5
	0.05	22.1 ± 1.5	>50
	reference	22.1 ± 1.0	

$T = 298$ K and $p = 1$ atm. We adopted the SPC water model[73] and the OPLS-AA force-field[74] for carbon, as implemented in the code GROMACS version 2019.4.[75,76] We used the leapfrog algorithm with a timestep $\tau = 1$ fs to propagate the equations of motion: after 100 ps of initial NpT equilibration,[77,78] we generated the trajectories for further analyses in the NVT ensemble [periodic box vectors $(4.601, 0, 0)$, $(1.536, 4.338, 0)$, $(-1.534, 2.169, 3.757)$ in nm units]. All computational details can be found in Ref. 33.

We target the association/dissociation of the fullerene dimer in water: the CV q chosen to represent this process in the projected dynamics is the distance between centers of mass of the two fullerenes. The memory kernel $K(t)$ corresponding to the generalized Langevin Eq. (1) decays monotonously, well-fitted by a double exponential: it drops to about 10% of the initial value within ≈0.2 ps, and it tends subsequently towards zero with a slower queue extending for few picoseconds (see Fig. 1). Based on this information, one would expect that the lower limit on the time resolution δt for the validity of the underdamped Langevin approximation would be about 0.2 ps. An upper limit of $\delta t \approx 0.5$ ps can be deduced from the disappearance of oscillations and appearance of an approximately exponential shape in the velocity autocorrelation function in Fig. 7. This is consistent with the results of Ref. 33 where fairly accurate overdamped models of the same system are obtained for $\delta t \gtrsim 0.5$ ps. Note that this regime corresponds also to the disappearance of velocity autocorrelations.

The effective mass corresponding to the Langevin description of the projected dynamics is estimated both analytically, based on the definition of the CV q using Eq. (5), giving in this case a q-independent mass that does not require a thermal average (see supplementary material for the derivation), and from the equipartition relation Eq. (6) using trajectories at different time resolutions δt. In the following, both possibilities are considered in combination with the estimation of free-energy landscapes and rates, to determine which one gives more accurate results. As shown in Fig. 8, similarly to the case of the double-well benchmark system (see also the discussion in the related section), the second kind of mass coincides with the analytical one only for the smallest δt, increasing monotonously afterwards, as expected.

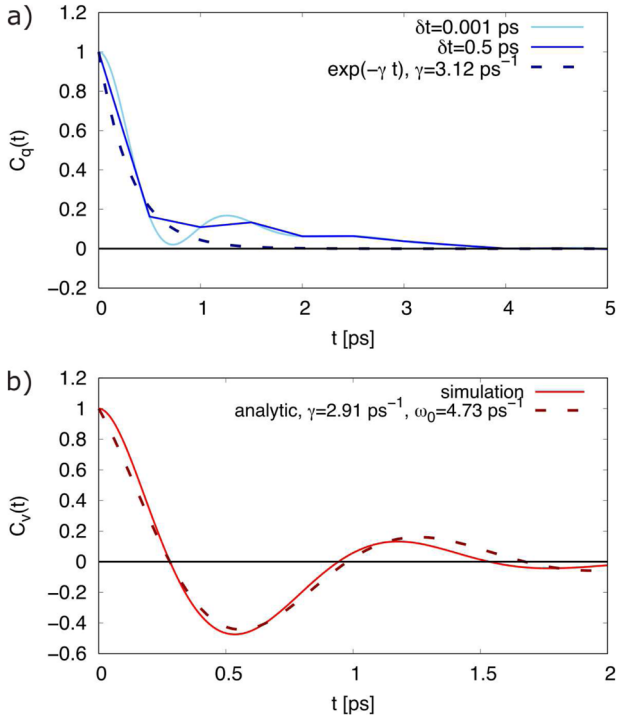

FIG. 7. Autocorrelation functions for the fullerene dimer in water, computed from a 4 ns trajectory in the well of the associated complex. (a) Position autocorrelation, sampled at a fine time resolution ($\delta t = 0.001$ ps) as well as at a coarse one ($\delta t = 0.5$ ps). The dashed line shows an exponential fit, i.e., to the analytical Ornstein-Uhlenbeck overdamped form. b) Velocity autocorrelation: the parameters γ and ω_0 are fitted via the analytical formula [Eq. (8)] in order to reproduce the trajectory data. The error bars are $<10^{-2}$ (not displayed).

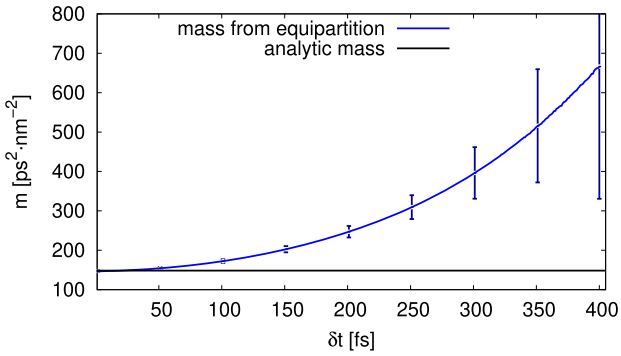

FIG. 8. Effective mass corresponding to the distance between centers of mass of two fullerenes in water: the mass computed analytically (see supplementary material) is compared with the one computed from equipartition as $m = k_B T / \langle v^2 \rangle$, with the velocities estimated by finite-differences at different time resolutions δt from a 10 ns trajectory remaining in the association well.

06 November 2023 13:00:22

Published under an exclusive license by AIP Publishing

As explained in the Methods section, friction can be estimated from projected MD trajectories following two alternative ways: either by fitting the v-autocorrelation function to a Ornstein-Uhlenbeck process, or by integrating over the non-Markovian friction.

Following the first route, considering trajectories in the associated well, a value of $\gamma = 2.91 \pm 0.05$ ps^{-1} is extracted, notwithstanding a small but visible deviation between the analytical and numerical functions, likely due to residual anharmonic and non-Markovian effects (for times smaller than ≈0.2 ps) on the dimer dynamics [Fig. 7(b)].

Following the second route, in the present system we cannot simply bring the velocity out of the non-Markovian friction integral and set $\gamma = \int_0^\infty K(t')dt'$, since the velocity has significant variations within the memory time scale[79] (see Fig. 7): in fact, we tested that direct integration of the kernel gives $\gamma \approx 7$ ps^{-1} and leads to a severely overestimated free-energy barrier for fullerenes dissociation (see also the results for the projected dynamics of a peptide in Ref. 11).

However, as shown in Fig. 9 there is an approximate linear correlation between the instantaneous velocity $v(t)$ and $\int_0^{t_0} K(t')v(t-t')dt'$, fulfilling the hypothesis of the underdamped approximation: this allows to estimate the value $\gamma = 2.79$ ps^{-1} by linear regression. We set $t_0 = 15$ ps, well beyond the major decay of the memory kernel (Fig. 1), and we tested that changing to $t_0 = 20$ ps the variation on the estimated γ is less than 0.05%. We remark that the present analysis indicates that, somehow counter-intuitively, the ULE approximation holds already for a time scale $\delta t = 0.01$ ps, much faster than the decay of memory effects.

With reference to the "phase diagram" in Fig. 2(a), the solvated fullerenes dimer is characterized by a ratio between memory time and inertial time ≈0.2/0.35 ps <1 and by a transition path time of several picoseconds: in the following, we considered model resolutions $0.01 \le \delta t \le 0.4$ ps to attempt the parametrization of ULE models.

We now proceed to estimate free-energy landscapes and kinetic rates by likelihood maximisation, keeping fixed the different mass and friction values estimated so-far. To this aim, we employ as input data a set of association-dissociation trajectories obtained with the aimless shooting technique:[80] 500 transition trajectories of 20 ps are generated (see Fig. 10) with a Monte Carlo procedure, where at each step an atomic configuration is randomly picked with a time displacement of ±1 ps from the previous transition trajectory, evolving forward and backward in time from Maxwell-Boltzmann random velocities and accepting the move if the resulting path joins reactants (here the associated dimer) and products (the dissociated state). The resulting acceptance ratio is 34.1%. The algorithm employed is the same as in Ref. 33.

The free-energy profiles estimated for different δt resolutions and for different choices of the mass and friction parameters are summarized in Fig. 11. For comparison, the reference brute-force free-energy profile was computed as a population histogram $F(q) = -k_B T \log \rho_{eq}(q)$ from an ergodic unbiased trajectory of 500 ns.

The main results we find are that (i) the best predictions of barriers and rates are obtained from the analytic mass combined with friction estimated at small δt (v-autocorrelation or linear regression of non-Markovian friction giving similar results), and (ii) the predictions are robust with respect to variations of the user-defined resolution δt in the broad range 0.01–0.4 ps.

In general, we can observe that the free-energy barrier, even in the most favorable cases, is underestimated by ≈1.5 $k_B T$: this deviation contributes to the underestimation of the MFPT (Table II) compared to the reference value $(6.0 \pm 0.7) \times 10^3$ ps estimated with brute-force MD.

Possibly, results could be improved by estimating a position-dependent friction landscape (here, γ is estimated in the local minimum only and kept constant elsewhere). Several previous works addressed the estimation of position-dependent friction coefficients in underdamped Langevin models:[38,40,81] typically, the dissociation process of small species in water appears to proceed with an increase of the friction towards the transition state. Specifically for the case of the fullerene dimer in water, Morrone and co-workers investigated the position dependence of the friction, albeit in the overdamped limit,[82] finding a similar trend. Unfortunately, the umbrella sampling approach based on CV autocorrelation that was shown effective in the overdamped regime in Refs. 33 and 83 cannot be applied to the underdamped case, where the parabolic bias significantly distorts γ,[66] so that further method development is needed. Notwithstanding the perturbation due to the bias, as we show in

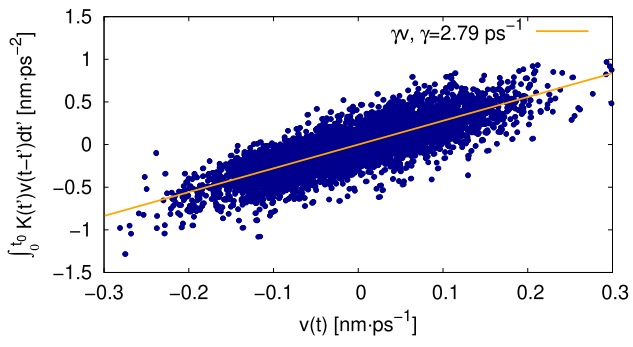

FIG. 9. Non-Markovian friction $\int_0^{t_0} K(t')v(t-t')dt'$ along aimless shooting trajectories, with $t_0 = 15$ ps, as a function of the instantaneous velocity $v(t)$ plotted every ps. The straight line represents the best linear fit ($R^2 = 0.67$) by $\gamma v(t)$ with $\gamma = 2.79$ ps^{-1}. The memory kernel $K(t)$ is shown in Fig. 1.

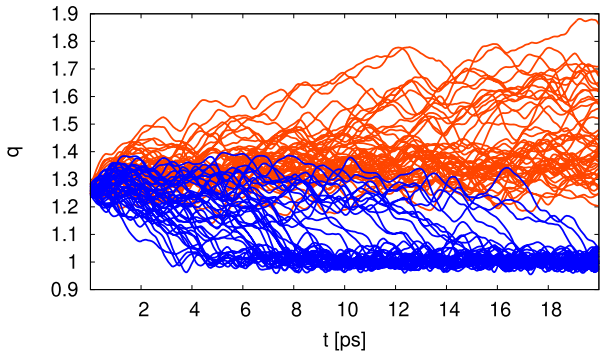

FIG. 10. Aimless shooting trajectories of the distance q between the centers of mass of two fullerenes in water, relaxing to the associated state (blue) or the dissociated one (red) starting from the transition state ensemble with initial velocities randomly sampled from the Boltzmann distribution.

06 November 2023 13:00:22

Published under an exclusive license by AIP Publishing

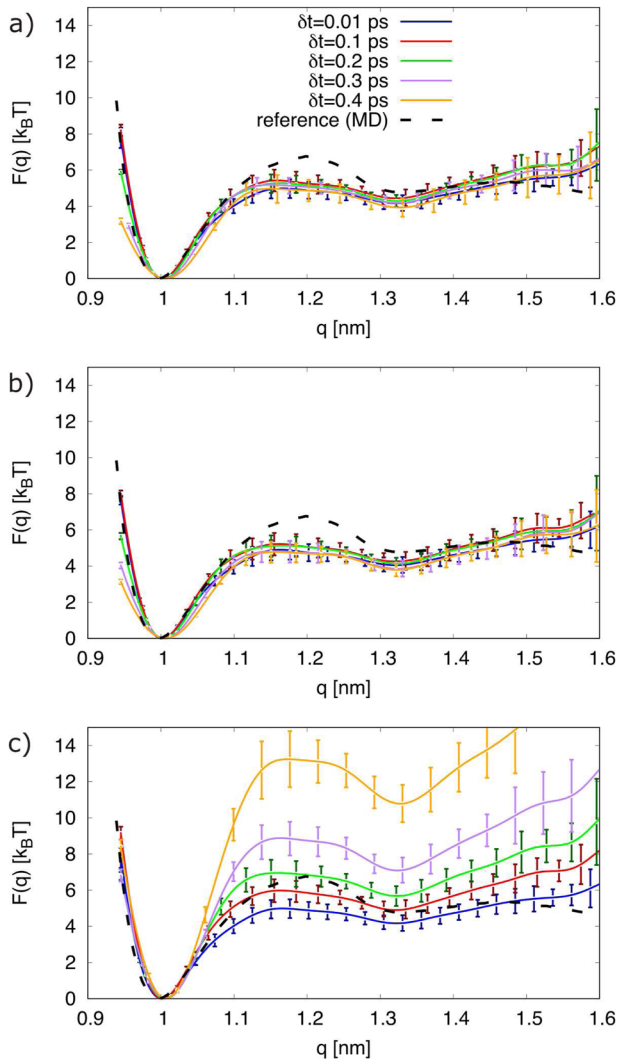

FIG. 11. Free-energy landscape for the interaction between fullerenes in water obtained by brute-force MD (dashed line) compared with the optimal Langevin models obtained by likelihood maximisation at different time resolutions δt, using (a) m = 147.7 nm^{-2} ps^2 (analytic), γ = 2.91 ps^{-1} (v-autocorrelation fit at δt = 0.001 ps), (b) same m, γ = 2.79 ps^{-1} (linear fit of non-Markovian friction force), (c) m and γ estimated at different δt (see supplementary material).

supplementary material the friction appears higher close to the transition state than in the free-energy minimum, suggesting that the barrier could be slightly increased considering a position-dependent friction.

It is important to remark that in principle, even in a best-case-scenario where a Langevin model could be "perfectly" parametrized for a given CV, it is reasonable to expect to estimate a kinetic rate generally faster than the exact one. The reason is that, according to a variational principle demonstrated in Ref. 84, when projecting the high-dimensional dynamics (supposed by the Authors to be overdamped) on a smaller CV space, the kinetics of the resulting overdamped Langevin models overestimates the transition rates unless

TABLE II. Mean first passage time (MFPT) for the dissociation of the fullerenes dimer in water, estimated from brute-force all-atom MD or from optimal Langevin models with different parametrization. Columns two and three refer to models where the mass (nm^{-2} ps^2) and friction (ps^{-1}) are computed at the finest resolution (integration time step) δt = 1 fs, but differ in the way the friction is computed (v-autocorrelation or linear fit of non-Markovian friction force, respectively). The last column refers to models with mass and friction estimated at time resolution δt, from equipartition and v-autocorrelation, respectively.

	MFPT (10^3 ps)		
δt (ps)	$m = 147.7, \gamma = 2.91$	$m = 147.7, \gamma = 2.79$	$m(\delta t), \gamma(\delta t)$
0.01	0.7 ± 0.1	0.6 ± 0.1	0.7 ± 0.1
0.1	1.0 ± 0.2	0.9 ± 0.2	1.8 ± 0.3
0.2	0.9 ± 0.1	0.8 ± 0.1	6.4 ± 1.4
0.3	0.9 ± 0.1	0.8 ± 0.2	42.8 ± 8.8
0.4	1.1 ± 0.3	0.8 ± 0.2	>100
MD		6.0 ± 0.7	

the CV corresponds to the committor function, i.e., the optimal CV. The principle remains to be mathematically demonstrated when projecting from high-dimensional MD trajectories (i.e., Hamilton's dynamics with a thermostat, typically) to an overdamped or underdamped CV model; however, it was shown numerically to hold at least in the case of C_{60} and C_{240} dimers in water in Ref. 85. In the latter work, the distance between centers of mass of the fullerenes is shown to be a fairly good CV, albeit not perfect, yielding a MFPT of $(2.9 \pm 0.2) \times 10^2$ ps from an overdamped model at resolution δt = 1 ps, improved to $(7.9 \pm 0.7) \times 10^3$ ps after optimizing the CV.

Taking into account all these factors, the MFPT values predicted by the models are rather satisfactory, being within one order of magnitude from the reference one and similar to those predicted by overdamped models in Refs. 33 and 85 Altogether, our results indicate that a small number of ~100 out-of-equilibrium transition path sampling trajectories are sufficient to infer, with some accuracy, orders-of-magnitude slower kinetic rates without passing through expensive umbrella sampling, forward flux sampling, or other enhanced sampling techniques.

IV. CONCLUSIONS

We investigated several non-trivial aspects of the parametrization of underdamped Langevin equations starting from relatively short, non-ergodic MD trajectories of systems exhibiting activated processes. We could identify a viable approach to estimate friction coefficients and to infer free-energy landscapes, leading to the prediction of kinetic rates. Importantly, the protocol is physics-based, exploiting theoretical results about the behavior of Hamiltonian dynamics after projection on a CV, and it yields predictions that are robust with respect to variation of the main user-defined parameter, i.e., the time resolution.

The optimal approach we identified has three steps: (i) the effective mass is computed analytically (or, equivalently, from short-time equipartition), (ii) the friction is estimated either from velocity autocorrelation (in the case of weak anharmonicity) or from a Markovian fit of non-Markovian friction, and (iii) the free-energy landscape is estimated by likelihood maximization. The different

06 November 2023 13:00:22

Published under an exclusive license by AIP Publishing

steps employ as input data unbiased MD trajectories of two types: wandering in local minima, or relaxing from barrier tops, the latter being generated from well-established transition path sampling techniques. We remark that basing our prediction of barriers and rates on transition path sampling ensures the best available starting data from the viewpoint of the microscopic mechanism. Since few hundreds such reactive trajectories are needed, the computational cost of the whole procedure is feasible, much lower than alternative techniques like transition interface sampling or forward flux sampling. Furthermore, the same source and amount of MD data could also allow to self-consistently optimize the CV for the activated process, as shown in Ref. 85 for the case of overdamped projected dynamics. The importance of the CV employed for the dimensionality reduction should not be underestimated, since low-quality CVs likely lead to Langevin models that underestimate the intrinsic timescales of the process.[84]

In the case of the solvated fullerenes dimer, we observed that the underdamped, instantaneous friction approximation can be extended – somehow counter-intuitively – to resolution times much smaller than the memory duration. This has an important implication: whenever the simple test we proposed to confirm such behavior is successful, a Markovian Langevin model can replace a more sophisticated non-Markovian model, with obvious practical advantages. The applicability of this idea remains to be tested on processes that are traditionally treated as non-Markovian, like chemical reactions and ion-ion association.

While the effective mass for the CV dynamics can be computed analytically, in principle it should be possible to include the friction as an extra parameter in the likelihood function, that in the present work is used for estimating only the free-energy landscape. This step would lead to a more seamless and unified protocol, however, to reach this goal further research is needed to clarify how to get rid of spurious velocity-related correlations that arise due to the use of finite-difference definitions of numerical CV derivatives.

Our results point to the need of high-precision friction estimation in order to infer accurate barriers and rates. In this respect, it appears important to focus future efforts on estimating in an efficient and robust way the position dependence of the friction $\gamma(q)$: once this information obtained, it will be possible to readily incorporate it into the propagator and likelihood expressions we derived in the present work. We expect that this amelioration will sizably improve the accuracy of the predicted barriers and rates in many applications.

SUPPLEMENTARY MATERIAL

The supplementary material document includes details about the calculation of the effective mass, the likelihood maximization algorithm, the umbrella sampling algorithm for estimating a position-dependent friction, and the integration of the memory kernel.

ACKNOWLEDGMENTS

This work was performed with the support of the Institut des Sciences du Calcul et des Données (ISCD) of Sorbonne University (Grant No. IDEX SUPER 11-IDEX-0004) and of the Franco-Swiss exchange Programme Germaine de Stael (project Grant No. 47901VL Campus France, Grant No. 2022-18 SATW). Calculations were performed on the GENCI-IDRIS French national supercomputing facility, under Grant Nos. A0110811069 and A0130811069. We gratefully acknowledge Line Mouaffac and Karen Palacio Rodriguez for suggestions related to code development and simulations, as well as Léon Huet, Rodolphe Vuilleumier, and Christoph Dellago for insightful discussions.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

David Daniel Girardier: Methodology (equal); Writing – review & editing (equal). **Hadrien Vroylandt**: Methodology (equal); Writing – review & editing (equal). **Sara Bonella**: Methodology (equal); Writing – review & editing (equal). **Fabio Pietrucci**: Methodology (equal); Project administration (lead); Writing – review & editing (lead).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] C. M. B. Leimkuhler, *Molecular Dynamics: With Deterministic and Stochastic Numerical Methods* (Springer, 2015).
[2] P. Gkeka, G. Stoltz, A. Barati Farimani, Z. Belkacemi, M. Ceriotti, J. D. Chodera, A. R. Dinner, A. L. Ferguson, J.-B. Maillet, H. Minoux *et al.*, J. Chem. Theory Comput. **16**, 4757 (2020).
[3] J. Rogal, Eur. Phys. J. B **94**, 223 (2021).
[4] L. Bonati, E. Trizio, A. Rizzi, and M. Parrinello, J. Chem. Phys. **159**, 014801 (2023).
[5] F. Pietrucci, Rev. Phys. **2**, 32 (2017).
[6] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, 2001).
[7] H. Vroylandt, Europhys. Lett. **140**, 62003 (2022).
[8] B. J. Gertner, K. R. Wilson, and J. T. Hynes, J. Chem. Phys. **90**, 3537 (1989).
[9] M. Berkowitz, J. D. Morgan, D. J. Kouri, and J. A. McCammon, J. Chem. Phys. **75**, 2462 (1981).
[10] B. J. Berne, M. E. Tuckerman, J. E. Straub, and A. L. R. Bug, J. Chem. Phys. **93**, 5084 (1990).
[11] O. F. Lange and H. Grubmüller, J. Chem. Phys. **124**, 214903 (2006).
[12] E. Darve, J. Solomon, and A. Kia, Proc. Natl. Acad. Sci. U. S. A. **106**, 10884 (2009).
[13] H. S. Lee, S.-H. Ahn, and E. F. Darve, MRS Online Proc. Libr. **1753**, 90 (2015).
[14] H. Lei, N. A. Baker, and X. Li, Proc. Natl. Acad. Sci. U. S. A. **113**, 14183 (2016).
[15] J. O. Daldrop, J. Kappler, F. N. Brünig, and R. R. Netz, Proc. Natl. Acad. Sci. U. S. A. **115**, 5169 (2018).
[16] A. Carof, R. Vuilleumier, and B. Rotenberg, J. Chem. Phys. **140**, 124103 (2014).
[17] G. Jung, M. Hanke, and F. Schmid, J. Chem. Theory Comput. **13**, 2481 (2017).
[18] Y. Yoshimoto, Z. Li, I. Kinefuchi, and G. E. Karniadakis, J. Chem. Phys. **147**, 244110 (2017).

06 November 2023 13:00:22

Published under an exclusive license by AIP Publishing

[19] A. V. Straube, B. G. Kowalik, R. R. Netz, and F. Höfling, Commun. Phys. **3**, 126 (2020).
[20] B. Lickert, S. Wolf, and G. Stock, J. Phys. Chem. B **125**, 8125 (2021).
[21] H. Meyer, S. Wolf, G. Stock, and T. Schilling, Adv. Theory Simul. **4**, 2000197 (2021).
[22] H. Vroylandt, L. Goudenège, P. Monmarché, F. Pietrucci, and B. Rotenberg, Proc. Natl. Acad. Sci. U. S. A. **119**, e2117586119 (2022).
[23] T. Schilling, Phys. Rep. **972**, 1 (2022).
[24] P. Xie, R. Car *et al.*, arXiv:2211.06558 (2022).
[25] R. Best, and G. Hummer, Phys. Chem. Chem. Phys. **13**, 16902 (2011).
[26] C. Micheletti, G. Bussi, and A. Laio, J. Chem. Phys. **129**, 074105 (2008).
[27] Q. Zhang, J. Brujić, and E. Vanden-Eijnden, J. Stat. Phys. **144**, 344 (2011).
[28] D. Crommelin and E. Vanden-Eijnden, Multiscale Model. Simul. **9**, 1588 (2011).
[29] D. Crommelin, J. Stat. Phys. **149**, 220 (2012).
[30] M. Baldovin, A. Puglisi, and A. Vulpiani, PLoS One **14**, e0212135 (2019).
[31] F. Sicard, V. Koskin, A. Annibale, and E. Rosta, J. Chem. Theory Comput. **17**, 2022 (2021).
[32] L. Donati, M. Weber, and B. G. Keller, J. Math. Phys. **63**, 123306 (2022).
[33] K. Palacio-Rodriguez and F. Pietrucci, J. Chem. Theory Comput. **18**, 4639 (2022).
[34] J. Gradišek, S. Siegert, R. Friedrich, and I. Grabec, Phys. Rev. E **62**, 3146 (2000).
[35] J. Timmer, Chaos, Solitons Fractals **11**, 2571 (2000).
[36] N. Schaudinnus, A. J. Rzepiela, R. Hegger, and G. Stock, J. Chem. Phys. **138**, 204106 (2013).
[37] N. Schaudinnus, B. Bastian, R. Hegger, and G. Stock, Phys. Rev. Lett. **115**, 050602 (2015).
[38] N. Schaudinnus, B. Lickert, M. Biswas, and G. Stock, J. Chem. Phys. **145**, 184114 (2016).
[39] B. Lickert and G. Stock, J. Chem. Phys. **153**, 244112 (2020).
[40] S. Wolf, B. Lickert, S. Bray, and G. Stock, Nat. Commun. **11**, 2918 (2020).
[41] S. Kieninger, S. Ghysbrecht, and B. G. Keller, arXiv:2303.14696 (2023).
[42] D. B. Brückner, P. Ronceray, and C. P. Broedersz, Phys. Rev. Lett. **125**, 058103 (2020).
[43] F. Ferretti, V. Chardes, T. Mora, A. M. Walczak, and I. Giardina, Phys. Rev. X. **10**, 031018 (2020).
[44] H. Meyer, T. Voigtmann, and T. Schilling, J. Chem. Phys. **150**, 174118 (2019).
[45] F. Glatzel, and T. Schilling, Europhys. Lett. **136**, 36001 (2021).
[46] R. Kubo, Rep. Prog. Phys. **29**, 255 (1966).
[47] G. Jung and F. Schmid, Soft Matter **17**, 6413 (2021).
[48] R. Zwanzig, J. Stat. Phys. **9**, 215 (1973).
[49] B. Carmeli and A. Nitzan, Chem. Phys. Lett. **102**, 517 (1983).
[50] E. Pollak and A. M. Berezhkovskii, J. Chem. Phys. **99**, 1344 (1993).
[51] G. R. Haynes, G. A. Voth, and E. Pollak, J. Chem. Phys. **101**, 7811 (1994).
[52] H. Vroylandt and P. Monmarché, J. Chem. Phys. **156**, 244105 (2022).
[53] 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).
[54] H. K. Shin, C. Kim, P. Talkner, and E. K. Lee, Chem. Phys. **375**, 316 (2010), part of special issue on Stochastic Processes in Physics and Chemistry (in Honor of Peter Hänggi).
[55] A. Obliger, J. Chem. Phys. **158**, 144101 (2023).
[56] P. Linz, Comput. J. **12**, 393 (1969).
[57] D. Lesnicki, R. Vuilleumier, A. Carof, and B. Rotenberg, Phys. Rev. Lett. **116**, 147804 (2016).
[58] J. Fricks, L. Yao, T. C. Elston, and M. G. Forest, SIAM J. Appl. Math. **69**, 1277 (2009).
[59] S. Wang, Z. Ma, and W. Pan, Soft Matter **16**, 8330 (2020).
[60] N. Bockius, J. Shea, G. Jung, F. Schmid, and M. Hanke, J. Phys.: Condens. Matter **33**, 214003 (2021).
[61] E. Darve and A. Pohorille, J. Chem. Phys. **115**, 9169 (2001).
[62] H. S. Lee, S.-H. Ahn, and E. F. Darve, J. Chem. Phys. **150**, 174113 (2019).
[63] A. Jain, I.-H. Park, and N. Vaidehi, J. Chem. Theory Comput. **8**, 2581 (2012).
[64] A. Nitzan, *Chemical Dynamics in Condensed Phases: Relaxation, Transfer and Reactions in Condensed Molecular Systems* (Oxford University Press, 2006).
[65] G. A. Pavliotis, *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations* (Springer, 2014), Vol. 60.
[66] J. O. Daldrop, B. G. Kowalik, and R. R. Netz, Phys. Rev. X. **7**, 041065 (2017).
[67] S. M. Iacus, *Simulation and Inference for Stochastic Differential Equations: With R Examples*, *Springer Series in Statistics Vol. 1* (Springer, New York, NY, 2008).
[68] A. N. Drozdov, Phys. Rev. E **55**, 2496 (1997).
[69] M. Kessler, Scand. J. Stat. **24**, 211 (1997).
[70] P. E. Kloeden, E. Platen, and H. Schurz, *Numerical Solution of SDE through Computer Experiments* (Springer Science & Business Media, 2012).
[71] E. Vanden-Eijnden and G. Ciccotti, Chem. Phys. Lett. **429**, 310 (2006).
[72] A. Frishman and P. Ronceray, Phys. Rev. X. **10**, 021009 (2020).
[73] H. J. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[74] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).
[75] H. J. Berendsen, D. van der Spoel, and R. van Drunen, Comput. Phys. Commun. **91**, 43 (1995).
[76] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, SoftwareX **1**, 19 (2015).
[77] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).
[78] M. Parrinello and A. Rahman, J. Appl. Phys. **52**, 7182 (1981).
[79] R. F. Grote and J. T. Hynes, J. Chem. Phys. **73**, 2715 (1980).
[80] B. Peters, G. T. Beckham, and B. L. Trout, J. Chem. Phys. **127**, 034109 (2007).
[81] C. Ayaz, L. Scalfi, B. A. Dalton, and R. R. Netz, Phys. Rev. E **105**, 054138 (2022).
[82] J. A. Morrone, J. Li, and B. J. Berne, J. Phys. Chem. B **116**, 378 (2012).
[83] G. Hummer, New J. Phys. **7**, 34 (2005).
[84] W. Zhang, C. Hartmann, and C. Schütte, Faraday Discuss. **195**, 365 (2016).
[85] L. Mouaffac, K. Palacio-Rodriguez, and F. Pietrucci, J. Chem. Theory Comput. **19**, 5701 (2023).

06 November 2023 13:00:22

Published under an exclusive license by AIP Publishing