A LETTERS JOURNAL EXPLORING THE FRONTIERS OF PHYSICS

Società Italiana di Fisica

edp sciences IOP Institute of Physics

LETTER

Non-Markovian out-of-equilibrium dynamics: A general numerical procedure to construct time-dependent memory kernels for coarse-grained observables

To cite this article: Hugues Meyer *et al* 2019 *EPL* **128** 40001

View the article online for updates and enhancements.

Recent citations

- Bottom-up Construction of Dynamic Density Functional Theories for Inhomogeneous Polymer Systems from Microscopic Simulations
Sriteja Mantha *et al*

- Projection operators in statistical mechanics: a pedagogical approach
Michael te Vrugt and Raphael Wittkowski

IOP ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection–download the first chapter of every title for free.

This content was downloaded from IP address 158.64.79.10 on 25/08/2020 at 13:23

A Letters Journal Exploring the Frontiers of Physics

November 2019

EPL, **128** (2019) 40001
doi: 10.1209/0295-5075/128/40001

www.epljournal.org

Non-Markovian out-of-equilibrium dynamics: A general numerical procedure to construct time-dependent memory kernels for coarse-grained observables

Hugues Meyer[1,2], Philipp Pelagejcev[2] and Tanja Schilling[2]

[1] *Research Unit in Engineering Science, Université du Luxembourg - L-4364 Esch-sur-Alzette, Luxembourg*
[2] *Physikalisches Institut, Albert-Ludwigs-Universität - D-79104 Freiburg, Germany*

received 19 July 2019; accepted in final form 26 November 2019
published online 21 January 2020

PACS 05.10.Gg – Stochastic analysis method (Fokker-Planck, Langevin, etc.)
PACS 02.70.-c – Computational techniques; simulations
PACS 05.45.Tp – Time series analysis

Abstract – We present a numerical method to compute non-equilibrium memory kernels based on experimental data or molecular dynamics simulations. The procedure uses a recasting of the non-stationary generalized Langevin equation, in which we expand the memory kernel in a series that can be reconstructed iteratively. Each term in the series can be computed based solely on knowledge of the two-time auto-correlation function of the observable of interest. We discuss how to optimize this method in order to be the most numerically convenient. As a proof of principle, we test the method on the problem of crystallization from a super-cooled Lennard-Jones melt. We analyze the nucleation and growth dynamics of crystallites and observe that the memory kernel has a time extent that is about one order of magnitude larger than the typical timescale needed for a particle to be attached to the crystallite in the growth regime.

Copyright © EPLA, 2020

Introduction. – Complex, dynamical many-body processes are often modelled in terms of the effective dynamics of a small set of relevant observables. Depending on the context, these observables are called "reaction coordinates" or "order parameters" [1–3]. For instance in biophysics one might be interested in the evolution of the shape of a protein during a folding experiment, but not in the motion of every single water molecule. Then a set of geometrical parameters that characterize this shape would be suitable reaction coordinates. Or in materials science, one might model the dynamics of a phase transition in terms of a suitable mesoscopic density without resolving the details of the microscopic motion of the atoms. In the 1960's Mori and Zwanzig developed a projection operator formalism to derive the equation of motion of such averaged observables, the Generalized Langevin equation (GLE) [4,5]. The GLE is valid only if the density of microstates is stationary. Based on the same formalism, but for non-stationary densities of microstates [6], we have recently derived a general structure for the equation of motion of reaction coordinates [7,8]. The structure applies to any dynamical process for which the microscopic equations of motion are deterministic and for any phase-space observable, even if it contains an explicit dependence on time. The resulting non-stationary Generalized Langevin equation (nsGLE) is thus the equation that needs to be solved, if one wishes to derive reaction coordinate dynamics outside of thermal equilibrium.

The GLE and the nsGLE contain an effective friction term that is non-local in time and integrates over the history of the process. The function that controls these history effects is called "the memory kernel" and is nowadays subject to extensive research in various contexts [9–11]. However the non-locality in time makes the analysis of the GLE and the nsGLE mathematically inconvenient. In applications the memory kernel is therefore often approximated by a Dirac delta distribution such that a time local Langevin equation is recovered [12–16]. This assumption, that we refer to as the "Markovian approximation", is however in practice often not verified before it is used, which potentially leads to inaccurate or wrong results. To go beyond this approximation, there are two possible routes. The first one consists in searching (or constructing) a set of new reaction coordinates for

which a Markovian description is correct [17–21]. This method is often useful, but the reaction coordinates constructed may either be abstract quantities, which are not guaranteed to be accessible in experiment, or a large set of them might be required, which implies that the Markovian Langevin equation needs to be solved in a high-dimensional space [22–24]. The second route consists in keeping the original reaction coordinate despite its non-Markovian dynamics, and finding a way to evaluate the corresponding memory kernel. Various methods in this spirit do already exist, but they are restricted to processes with stationary microstate distributions, *i.e.*, they assume that the memory kernel is invariant under translations in time. We present here a general method to compute memory kernels for arbitrary processes and for arbitrary observables from MD simulations or experimental data, *i.e.*, a procedure to analyze the out-of-equilibrium dynamics of reaction coordinates and to construct their equation of motion.

The main idea of the procedure is to recast the nsGLE in a form in which the memory kernel is expressed as a sum of convolution products. Each term is determined recursively from the previous ones, and it probes the behaviour of memory at successively longer times. The discrepancies from the Markovian limit can be assessed by analyzing the functional shape of the computed terms and their relative timescales. The method neither requires modeling nor approximation and uses as a single input the two-time auto-correlation function of the observable under study, which is easily accessible in simulations or experiments. We also note that the computational time of the procedure scales linearly with the number of terms computed in the expansion.

As a realistic example of reaction coordinate analysis, we apply the method to the process of crystallization of a super-cooled Lennard-Jones melt. Crystallization processes are usually described in the framework of Transition State Theory, which is based on a Markovian approximation. We test this assumption here and reconstruct the memory kernel of the nsGLE which governs the evolution of nucleation and growth of crystalline clusters. We observe significant memory effects.

How to measure memory kernels. –

Numerical method. Consider a system of $N \gg 1$ degrees of freedom $\{\Gamma_i\}_{i\leq N}$ that evolve according to deterministic microscopic equations of motion (*e.g.*, Hamilton's equations of motion), and a phase-space observable $A(\mathbf{\Gamma})$ that is fully determined by the location $\mathbf{\Gamma}$ in phase-space. Next, take an ensemble, *i.e.*, a large number of copies of the system, and allow these copies to be initialized according to any phase-space distribution, in particular a non-stationary one. The microscopic equations of motion as well as the observable A can also be explicitly time-dependent. These formal definitions are general enough to encompass a very broad spectrum of processes (even including quantum mechanical microscopic processes [25]). We showed in ref. [7] that for any such observable and for any dynamical process, regardless of how far from equilibrium it evolves, one can always define functions $\omega(t)$, $K(t',t)$ and η_t such that the equations of motion for A_t and for its auto-correlation function $C(t',t) = \langle A^*(t')A(t)\rangle$ are

$$\frac{\mathrm{d}A_t}{\mathrm{d}t} = \omega(t)A_t + \int_0^t \mathrm{d}\tau K(\tau,t)A_\tau + \eta_t, \tag{1}$$

$$\frac{\partial C(t',t)}{\partial t} = \omega(t)C(t',t) + \int_{t'}^t \mathrm{d}\tau C(t',\tau)K(\tau,t), \tag{2}$$

where $K(\tau,t)$ is the memory kernel and the average is taken over the ensemble of non-equilibrium trajectories. The time dependence as a subscript denotes the dependence on a single trajectory, whereas the time dependence between parentheses indicates a fixed function of time independent of the trajectory. In particular we showed $\omega(t) = \mathrm{d}(\ln\sqrt{C(t,t)})/\mathrm{d}t$. If the timescale on which A_t evolves is much longer than the typical microscopic timescale, η_t can be interpreted as a noise. Finally, we recall the identity

$$\langle \eta^*(t')\eta(t)\rangle = -K(t',t)\langle |A(t')|^2\rangle, \tag{3}$$

which is valid for any times t, t'. This has the structure of an out-of-equilibrium fluctuation-dissipation relation, although its meaning can be broader than this. Note that other non-equilibrium fluctuation-dissipation theorems were recently published in different contexts and with different structures [26,27].

We emphasize that eqs. (1) and (2) are the result of an exact derivation based on projection operator techniques, but they are different in structure from other classes of GLE due to a different choice of projection operator. In modern statistical physics the GLE is often interpreted as the evolution equation of an observable within a "free-energy landscape" the drift term being in this context the thermodynamic force associated to this landscape [28,29]. Our approach is different. We intend to study processes far from thermal equilibrium, and we therefore do not refer to any equilibrium quantities such as the free energy, as averages over stationary distributions of microstates do not make sense in this framework. ω, K and η are here "kinetic" quantities that reproduce the dynamics of an observable, and their physical interpretation can, in general, not be motivated by thermodynamic arguments.

Several methods are already available to infer memory kernels from simulation data, *e.g.*, Fourier-Laplace analysis [30–33], projection operator analysis [34–36], parametrization techniques [37,38], and iterative numerical inference schemes [39]. However, these methods are applicable only to stationary processes where the memory kernel $K(t',t)$ effectively depends only on the difference $t'-t$. We propose here a method to solve the non-stationary case. (The method can, of course, also be applied to the stationary case.) The basic idea is to use

a measured auto-correlation function $C(t',t)$ to construct the memory kernel by means of an iterative procedure:

We formally integrate eq. (2) into

$$C(t',t) = C(t',t') + \int_{t'}^{t} \mathrm{d}\tau C(t',\tau)\mathcal{J}(\tau,t), \tag{4}$$

where we have defined

$$\mathcal{J}(t',t) = \omega(t') + \int_{t'}^{t} \mathrm{d}\tau K(t',\tau), \tag{5}$$

such that $K(t',t) = \partial_t \mathcal{J}(t',t)$. Taking the derivative of eq. (4) with respect to t' and rearranging terms allows to write

$$\mathcal{J}(t',t) = j_0(t',t) + \int_{t'}^{t} \mathrm{d}\tau \mathcal{S}_0(t',\tau)\mathcal{J}(\tau,t), \tag{6}$$

where we have defined

$$\mathcal{S}_0(t',t) = C(t',t')^{-1}\partial_{t'}C(t',t), \tag{7}$$
$$j_0(t',t) = C(t',t')^{-1}\partial_{t'}[C(t',t') - C(t',t)]. \tag{8}$$

(Note that the first time-derivative of the auto-correlation function $\partial_{t'}C(t',t)$ is all that is required for the following steps. We do not need to take any further time-derivatives of the input data.) By iteratively substituting $\mathcal{J}(t,\tau)$ on the right-hand side of eq. (6) by eq. (6) itself, we obtain

$$\mathcal{J}(t',t) = j_0(t',t) + \sum_{n=0}^{\infty} \int_{t'}^{t} \mathrm{d}\tau \mathcal{S}_n(t',\tau) j_0(\tau,t), \tag{9}$$

where the functions $\mathcal{S}_n$ are defined recursively via the identity

$$\int_{t'}^{t} \mathrm{d}\tau \mathcal{S}_n(t',\tau)\mathcal{S}_m(\tau,t) = \mathcal{S}_{n+m+1}(t',t), \tag{10}$$

which is valid for any $(n,m) \in \mathbb{N}^2$. In the limit $\omega = 0$, it is easy to show that $\mathcal{J}(t',t) \overset{\omega=0}{\to} -\mathcal{S}(t',t)$. Higher orders in the expansion have impact on longer times, and the number of terms needed for the sum $\mathcal{S} = \sum_{n=0}^{\infty}\mathcal{S}_n$ to converge at a certain time yields information about the strength and the time extent of the memory effects.

As all terms on the right-hand side of eq. (9) are expressed in terms of the auto-correlation function $C(t',t)$, we propose the following numerical scheme to compute $K(t',t)$:

1. Carry out a set of simulations (or experiments) and measure the observable A_t for each trajectory[1].
2. Compute the two-time auto-correlation function $C(t',t) = N_{\mathrm{traj}}^{-1}\sum_{i\in\mathrm{traj}} A_{t'}^{*(i)}A_t^{(i)}$.
3. Compute $\mathcal{S}_0$ and j_0 using eqs. (7) and (8).
4. Compute $\mathcal{S}_{n+1}$ recursively using eq. (10) with $m = 0$. Stop if for any times t' and t, $\mathcal{S}_n(t',t) \ll \sum_{j=0}^{n}\mathcal{S}_j(t',t)$.
5. Compute $\mathcal{J}$ using eq. (9) and finally $K(t',t) = \partial_t\mathcal{J}(t',t)$.

This method is general enough to be applied to any dynamical process and any observable, even far from equilibrium, because it relies only on the structure of eqs. (1) and (2). Note that once the memory kernel has been constructed, the corresponding nsGLE can be solved to predict the dynamics of the process. One has thus obtained an effective coarse-grained description in terms of one coordinate rather than N.

Markovian limit. The Markovian approximation is frequently used to model coarse-grained processes, because its mathematical structure is rather convenient. However, users of multi-scale modelling tools often do not check the validity of this approximation systematically. In this subsection, we derive four relations which can help to assess or disprove the Markovian assumption in practical studies.

A) We define the Markovian limit as $K(t',t) = \gamma(t')\delta(t'-t)$ (with $\gamma(t) < 0$ in most practical cases). This straightforwardly yields

$$C(t',t) = C(t',t')\exp\left(\int_{t'}^{t}\mathrm{d}\tau\bar{\gamma}(\tau)\right), \tag{11}$$

for $t' < t$, with $\bar{\gamma}(t) = \gamma(t)/2 + \omega(t)$. Therefore, for any time $s \in [t',t]$ a Markovian process fulfills the decomposition $C(t',t) = C(t',s)C(s,t)$. To test the validity of this decomposition, we define a dimensionless quantity $\epsilon(s)$ as

$$\epsilon(s) := \frac{1}{s(T-s)}\int_0^s \mathrm{d}t' \int_s^T \mathrm{d}t \left|1 - \frac{C(t's)C(s,t)}{C(t',t)}\right|, \tag{12}$$

where T is the total duration of the process. For a given value of s, $\epsilon(s)$ measures the deviation from the exponential decomposition at any point in the 2D domain $[0,s]\times[s,T]$ and averages over it. By definition, $\epsilon(s)$ vanishes in the Markovian case. The dependence on s allows to quantify how non-Markovian behaviour evolves throughout the process.

B) Next, we derive semi-analyitcal results regarding the functions $\mathcal{S}_n$ presented in the previous paragraph. In the Markovian limit $\mathcal{S}_0(t',t) = -\bar{\gamma}(t')e^{\int_{t'}^{t}\mathrm{d}\tau\bar{\gamma}(\tau)}$. The recursive relation (10) allows to infer that each order $\mathcal{S}_n(t',t)$ is proportional to the n-th–order term in the Taylor expansion of the exponential function $\exp(-\int_{t'}^{t}\mathrm{d}\tau\bar{\gamma}(\tau))$, the global prefactor being $-\bar{\gamma}(t')\exp(\int_{t'}^{t}\mathrm{d}\tau\bar{\gamma}(\tau))$ for all. Their sum becomes then

$$\mathcal{S}(t',t) = \sum_{n=0}^{\infty}\mathcal{S}_n(t',t) = -\bar{\gamma}(t'), \tag{13}$$

which is valid for $t' < t$. In the case $t' > t$, the same calculation would yield $\mathcal{S}(t',t) = \underline{\gamma}(t') = \gamma(t')/2 - \omega(t')$.

[1] The number of trajectories must be chosen such that the initial phase-space distribution $\rho_0(\mathbf{\Gamma})$ is probed with sufficient precision.

In other words, for any t and t', we would have

$$\mathcal{S}(t',t) = \omega(t') + \gamma(t')[\Theta(t'-t) - 1/2], \tag{14}$$

where Θ is the Heaviside function. This condition can be tested on experimental or simulation data.

C) Another interesting property of the functions $\mathcal{S}_n(t',t)$ in the Markovian limit can be obtained if one considers that their derivatives with respect to t can be written as

$$\partial_t \mathcal{S}_n(t',t) = \gamma(t)(\mathcal{S}_n(t',t) - \mathcal{S}_{n-1}(t',t)). \tag{15}$$

In other words, we have $\partial_t \mathcal{S}_n(t',t) = 0 \Leftrightarrow \mathcal{S}_n(t',t) = \mathcal{S}_{n-1}(t',t)$: the extremum of the function $\mathcal{S}_n(t',t)$ as a function of t is located where $\mathcal{S}_n(t',t)$ crosses the previous order $\mathcal{S}_{n-1}(t',t)$. This is a graphical check that can easily be performed to test the Markovian assumption.

D) Finally, in the stationary limit where $\gamma(t)$ becomes a constant, we have for $t' < t$

$$\mathcal{S}_n(t',t) = (-\gamma)^{n+1}(t-t')^n e^{\gamma(t-t')}/n!. \tag{16}$$

At constant t', $\mathcal{S}_n(t',t)$ presents an extremum at $t_n^* = t' - n/\gamma$, and $\mathcal{S}_n(t',t_n^*) = -\gamma n^n e^n/n! \overset{n\to\infty}{\propto} n^{-1/2}$. In a stationary process, one can hence test the Markovian assumption by checking if the extrema of the functions $\mathcal{S}_n(t',t)$ are located with an equal spacing and if their amplitude decay as $1/\sqrt{n}$.

Complementary formulations. Reconsidering the numerical method described previously, we note that the case where the drift term $\omega(t)$ vanishes is particularly convenient to deal with. The limit $\mathcal{J} = -\mathcal{S}$ is reached in this case, which implies less numerical computation. Also, we see from eq. (3) that the symmetry property $K(t',t) = K^*(t,t')$ is obtained if $\langle |A_t|^2 \rangle$ is constant, which holds for vanishing $\omega(t)$. We therefore now show how to reversibly remove the drift term in order to derive a more efficient implementation of the numerical procedure.

Recall that $\omega(t) = d[\ln(\sqrt{\langle |A(t)|^2 \rangle})]/\mathrm{d}t$ and let us divide eq. (1) by the function $f(t) = \sqrt{\langle |A(t)|^2 \rangle}$. The resulting term involving $\omega(t)$ can be written as $f(t)\omega(t)A_t = -\dot{f}(t)A_t$, such that the equation of motion, eq. (1), becomes

$$\frac{\mathrm{d}\tilde{A}_t}{\mathrm{d}t} = \int_0^t \mathrm{d}\tau \tilde{A}_\tau \tilde{K}(\tau,t) + \tilde{\eta}_t, \tag{17}$$

where we have defined a modified time-dependent phase-space observable

$$\tilde{A}(\mathbf{\Gamma},t) = A(\mathbf{\Gamma})/\sqrt{\langle |A(t)|^2 \rangle} \tag{18}$$

and its related functions

$$\tilde{K}(t',t) = K(t',t)\sqrt{\langle |A(t')|^2 \rangle / \langle |A(t)|^2 \rangle}, \tag{19}$$

$$\tilde{\eta}_t = \eta_t/\sqrt{\langle |A(t)|^2 \rangle}. \tag{20}$$

Since the structure of eq. (17) is similar to eq. (1), we can apply the numerical method presented above directly to the modified variable $\tilde{A}$, in order to compute $\tilde{K}(t',t)$ and finally obtain $K(t',t)$ using eq. (19).

If one intends to study the fluctuations of A_t independently of the evolution of the average trajectory, one can go one step further: We introduce another modified, explicitly time-dependent phase-space observable $\Delta A(\mathbf{\Gamma},t)$

$$\Delta A(\mathbf{\Gamma},t) \equiv A(\mathbf{\Gamma}) - \langle A(t) \rangle . \tag{21}$$

This observable measure deviations from the average trajectory, thus its auto-correlation function measures actual fluctuations. Since eqs. (1) and (2) are also valid for time-dependent observables, the numerical procedure to compute the corresponding memory kernel can still be applied, however one would still have to deal with the drift term $\Delta\omega(t) = d\ln\left(\sqrt{\langle |\Delta A(t)|^2 \rangle}\right)/\mathrm{d}t$ which measures how fast the variance of the time-dependent probability distribution of A evolves. In order to remove this term, we use again the transformations introduced above and finally define the modified observable

$$\Delta\tilde{A}(\mathbf{\Gamma},t) \equiv \Delta A(\mathbf{\Gamma})/\sqrt{\langle |\Delta A(t)|^2 \rangle}, \tag{22}$$

that measures deviations from the average trajectory normalized by the variance of the process. It is thus a unitless number indicating whether a particular trajectory is delayed or advanced compared to the average one. The corresponding auto-correlation function $\Delta\tilde{C}(t',t) = \langle \Delta\tilde{A}^*(t')\Delta\tilde{A}(t) \rangle$ can be used to obtain a memory kernel $\Delta\tilde{K}$, to which is associated a fluctuating term $\Delta\tilde{\eta}_t$. This is then used to rewrite the equation of motion for A_t as

$$\frac{\mathrm{d}A_t}{\mathrm{d}t} = \langle \dot{A}(t) \rangle + \Delta\omega(t)\Delta A_t + \int_{t'}^t \mathrm{d}\tau \Delta A_\tau \Delta K(\tau,t) + \Delta\eta_t, \tag{23}$$

with $\Delta K(t',t) = \Delta\tilde{K}(t',t)\sqrt{\langle |\Delta A(t)|^2 \rangle / \langle |\Delta A(t')|^2 \rangle}$ and $\Delta\eta_t = \langle |\Delta A(t)|^2 \rangle \Delta\tilde{\eta}_t$. When studying memory effects and deriving an equation of motion for an observable A, one can hence chose between three equivalent descriptions, eq.(1), eq. (17) and eq. (23), which involve slighlty different drifts, memory kernels and fluctuating forces. The numerical method that we have introduced can be implemented in all three cases.

Example —Lennard-Jones crystallization. – The crystallization process in simple systems was recently shown to exhibit non-trivial non-equilibrium features, potentially due to memory effects [40,41], hence we decided to investigate it using our method.

We carried out molecular dynamics simulations of $N = 32000$ particles interacting via a 6–12 Lennard-Jones potential. We used a cubic box with periodic boundary conditions and we ran the dynamics in the NVT ensemble, using a Nosé-Hoover thermostat to control the temperature. We first equilibrated the liquid phase at density $\rho = 1$ and temperature $T = 2$ (in Lennard-Jones reduced units), and then instantaneously quenched the temperature to $T = 0.75$, for which the equilibrium phase is known

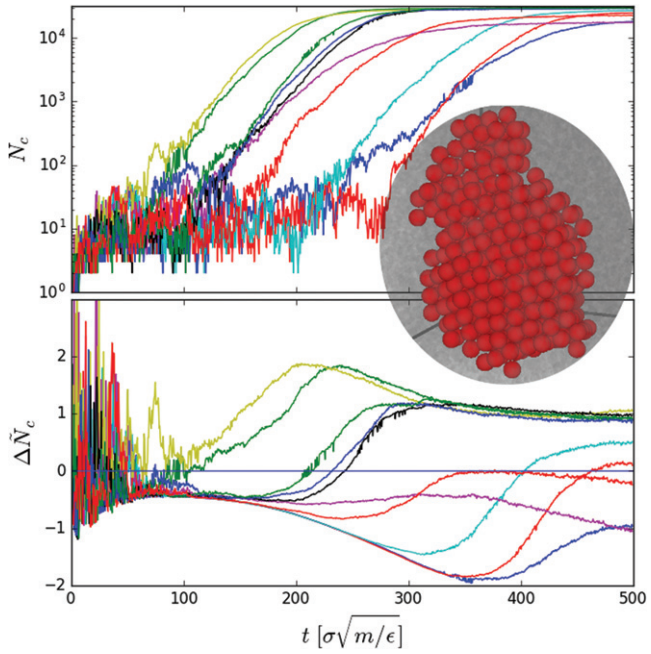

Fig. 1: Evolution of the size of the largest cluster N_c (top), and its associated modified observable $\Delta\tilde{N}_c$, for several simulation trajectories. We also show an example snapshot of a crystallite surrounded by the liquid phase.

to be a crystal. The time of the sudden temperature drop is considered as our intial time $t = 0$. Each trajectory starts from a different liquid-like configuration drawn from the high-temperature Boltzmann distribution. We then let the system evolve freely (*i.e.*, we did not use any biasing scheme to speed up sampling). In order to monitor the formation and growth of crystallites, we used the method introduced by ten Wolde *et al.* [42] based on orientational bond order parameters [43]. A crystalline cluster is defined as a set of neighbouring crystal-like particles. The observable that we used as a reaction coordinate is the number N_c of particles in the largest crystalline cluster. We observed the evolution of $N_c(t)$, see fig. 1, upper panel, where we also show a typical snapshot of a crystallite. A total of 4019 trajectories were used for the analysis.

We computed for each trajectory the modified variable $\Delta\tilde{N}_c$ as defined in eq. (21) (see fig. 1, lower panel), and its auto-correlation function $\Delta\tilde{C}(t',t)$ (all the associated quantities are then written with the notation $\Delta\tilde{\cdot}$ in order to be consistent). We then applied the numerical method presented above. As we used $\Delta\tilde{C}(t',t)$, the drift term $\Delta\tilde{\omega}$ vanishes, which implies $j_0 = -\Delta\tilde{\mathcal{S}}_0$ and $\Delta\tilde{\mathcal{J}} = -\Delta\tilde{\mathcal{S}}$. We computed $\Delta\tilde{\mathcal{S}}_0(t',t) = \partial_{t'}\Delta\tilde{C}(t',t)$, and we then iteratively applied the recursion relation equation (10) with $m = 0$, *i.e.*, $\int_{t'}^{t} \mathrm{d}\tau \Delta\tilde{\mathcal{S}}_0(t',\tau)\Delta\tilde{\mathcal{S}}_n(\tau,t) = \Delta\tilde{\mathcal{S}}_{n+1}(t',t)$. We show in fig. 2 the functions $\Delta\tilde{\mathcal{S}}_n$, as well as the function $\Delta\tilde{\mathcal{J}} = -\sum_n \Delta\tilde{\mathcal{S}}_n$. If the nucleation process were Markovian, $\Delta\tilde{\mathcal{J}}$ would be a step function of t, constant for $t < t'$ and $t > t'$ and discontinuous at $t = t'$. Although we observe this discontinuity, the behaviour around $t \neq t'$ is far from being constant. In addition, we note that

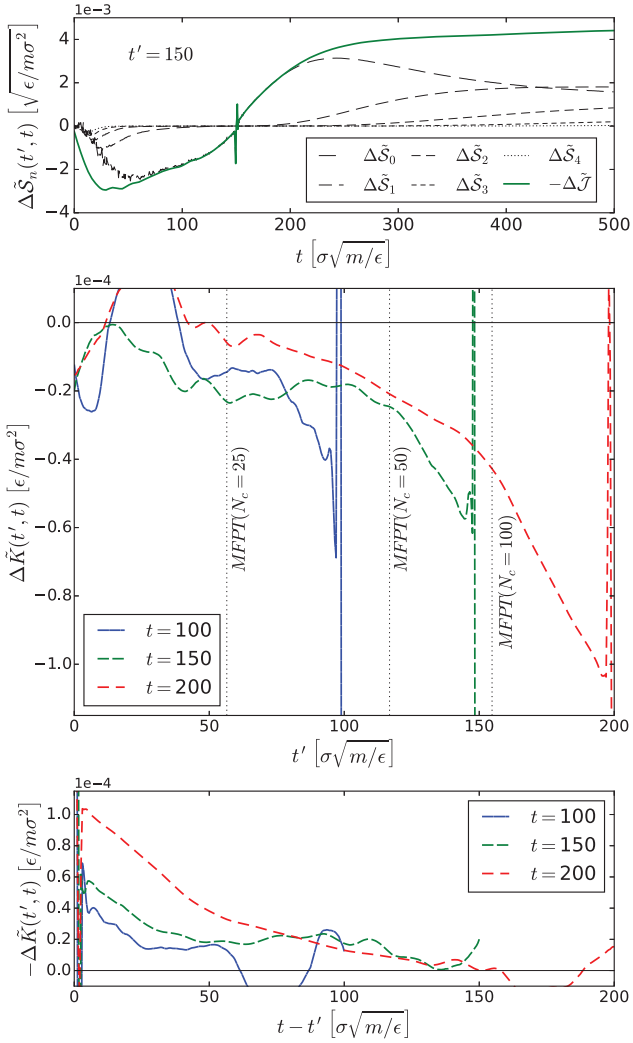

Fig. 2: Top: $\Delta\tilde{\mathcal{S}}_n(t',t)$ for $0 \leq n \leq 4$ as well of their sum $\Delta\tilde{\mathcal{S}}(t',t) = \sum_n \Delta\tilde{\mathcal{S}}_n(t',t) = -J(t',t)$, as a function of t for a fixed $t' = 150$. $\Delta\tilde{\mathcal{J}}$ would be a step function of t in the Markovian case. Center: memory kernel $\Delta\tilde{K}(t',t)$ as a function of t'. We also show the values of the mean first passage times for different values of N_c as vertical lines. According to committor analysis the critical nucleus has $N_c \approx 40$, thus $t = 100$ is in the induction time regime and $t = 200$ in the growth regime. Bottom: negative of the memory kernel as a function of $t - t'$.

$\Delta\tilde{N}_c(t)$ vanishes for each trajectory in the limit $t \to 0$, which implies that $\Delta\tilde{\mathcal{S}}_0(t',t) \sim 0$ for $t \to 0$, and thus $J(t',t) = -\Delta\tilde{\mathcal{S}}(t',t) \sim 0$ for $t \to 0$. This is observed numerically, and suggests a non-trivial behaviour for the memory kernel $\Delta\tilde{K}$ at short times. The algorithm needed 15 iterations in $\Delta\tilde{\mathcal{S}}$ to properly converge. The functions $\Delta\tilde{\mathcal{S}}_n(t',t)$ probe the behaviour of memory effects for different times. Each new order n presents extrema at increasing values of $|t' - t|$. The summation of all the orders results in the smooth function $\Delta\tilde{\mathcal{S}} = -\Delta\tilde{\mathcal{J}}$ (see fig. 2).

We checked the validity of the function $\Delta\tilde{\mathcal{J}}$ obtained in this way by using it as an input to compute the right-hand

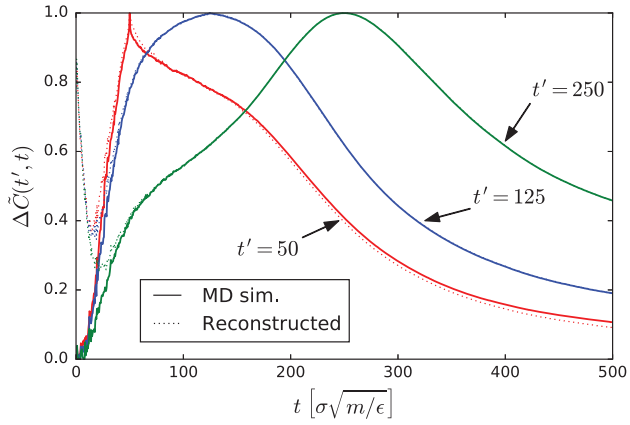

Fig. 3: Modified auto-correlation function $\Delta\tilde{C}(t',t)$ for various values of t' as a function of t. The solid lines are directly computed from the MD simulations, the dotted lines are the right-hand side of eq. (4), where $\Delta\tilde{\mathcal{J}}$ is computed via the method presented. The overlap is nearly perfect.

side of eq. (4), and we compared the result to the left-hand side, *i.e.*, $\Delta\tilde{C}(t',t)$ itself. As is shown in fig. 3, the overlap is very good[2]. This test confirms that the method presented is able to reconstruct the dynamics of a reaction coordinate, and that it can be used to develop numerical coarse-graining schemes for dynamics out of thermal equilibrium.

Once this check on $\Delta\tilde{\mathcal{J}}(t',t)$ was performed, we computed its derivative with respect to t, yielding $\Delta\tilde{K}(t',t)$ which is also shown in fig. 2. The singularities at $t'=t$ correspond to Markovian contributions to the dynamics, whereas the tails for $t'<t$ quantify memory effects. In addition we show vertical lines indicating the mean first passage times for various values of N_c. A committor analysis of the trajectories showed that the critical crystallite contains ca. 40 particles. Hence we observe memory both in the nucleation and in the growth regime. The average time needed for a particle to be attached to a cluster surface area of σ_{LJ}^2 was $\tau_{att}\simeq 5$ (in LJ units), which is about one order of magnitude smaller that the time extent of the memory kernel.

As mentioned above, these results cannot be completely understood in terms of equilibrium physics. In particular, the memory kernel should not be interepreted as a friction coefficient acting on a variable that evolves in a free-energy landscape. We have proven this approach to be inappropriate in the context of nucleation in a recent article [44] (see also [40,45] for other studies arguing the same point). The memory kernel that we find should rather be interpreted as a dynamical coefficient, whose extended support in time shows that the evolution of the size of crystalline clusters effectively depends on the history of the process. A possible reason for this could be, *e.g.*, transitions between different crystalline structures. However a detailed study of these effects is not the purpose of this paper. We simply conclude that memory effects are not negligible in the Lennard-Jones crystallization process.

Conclusion. – We have introduced a numerical method to construct memory kernels for any process for which the non-stationary Generalized Langevin Equation is relevant, regardless of how far from equilibrium the system is. This procedure requires little computational effort and no modeling assumption since its only input is the two-time auto-correlation function of the observable under study. The method can also be applied to a modified version of the nsGLE, where the memory kernel and the fluctuating force contribute only to the fluctuations about the average of the observable of interest. We investigated the crystallization process as a proof of principle, and we have shown that the procedure allows to reconstruct the dynamics of the problem. In this particular example, we demonstrate that memory effects play a significant role for nucleation and growth dynamics.

* * *

We thank G. Stock for valuable feedback on the manuscript. This project has been financially supported by the National Research Fund Luxembourg (FNR) within the AFR-PhD programme. Computer simulations presented in this paper were carried out using the HPC facility of the University of Luxembourg.

REFERENCES

[1] Peters B., *Annu. Rev. Phys. Chem.*, **67** (2016) 669.
[2] Rohrdanz M. A., Zheng W. and Clementi C., *Annu. Rev. Phys. Chem.*, **64** (2013) 295.
[3] Sittel F. and Stock G., *J. Chem. Phys.*, **149** (2018) 150901.
[4] Zwanzig R., *Phys. Rev.*, **124** (1961) 983.
[5] Mori H., *Prog. Theor. Phys.*, **33** (1965) 423.
[6] Grabert H., *Projection Operator Techniques in Nonequilibrium Statistical Mechanics*, Vol. **95** (Springer) 1982.
[7] Meyer H., Voigtmann T. and Schilling T., *J. Chem. Phys.*, **150** (2019) 174118.
[8] Meyer H., Voigtmann T. and Schilling T., *J. Chem. Phys.*, **147** (2017) 214110.
[9] Wolf S. and Stock G., *J. Chem. Theory Comput.*, **14** (2018) 6175.
[10] Kappler J., Daldrop J. O., Brünig F. N., Boehle M. D. and Netz R. R., *J. Chem. Phys.*, **148** (2018) 014903.
[11] Español P., de la Torre J. and Duque-Zumajo D., *Phys. Rev. E*, **99** (2019) 022126.
[12] Huopaniemi I., Luo K., Ala-Nissila T. and Ying S.-C., *J. Chem. Phys.*, **125** (2006) 124901.

[2]The discrepancies at very short times are due to numerical errors in the computation of the derivative of $\Delta\tilde{C}(t',t)$. Due to the initial vanishing width of the crystllite size distribution, the number of trajectories needed to eliminate statistical fluctuations at short times is very large. Here our aim is to demonstrate the validity of our method and not capture the details of the nucleation process, therefore we did not run additional simulations to resolve the early stage dynamics.

[13] Hörger I., Velasco E., Mingorance J., Rivas G., Tarazona P. and Vélez M., *Phys. Rev. E*, **77** (2008) 011902.
[14] Espanol P. and Löwen H., *J. Chem. Phys.*, **131** (2009) 244101.
[15] Shea P. and Kreuzer H. J., *Surf. Sci.*, **605** (2011) 296.
[16] Knoch F. and Speck T., *New J. Phys.*, **17** (2015) 115004.
[17] Peters B. and Trout B. L., *J. Chem. Phys.*, **125** (2006) 054108.
[18] Rohrdanz M. A., Zheng W., Maggioni M. and Clementi C., *J. Chem. Phys.*, **134** (2011) 03B624.
[19] Pérez-Hernández G., Paul F., Giorgino T., De Fabritiis G. and Noé F., *J. Chem. Phys.*, **139** (2013) 015102.
[20] Peters B., *Annu. Rev. Phys. Chem.*, **67** (2016) 669.
[21] Chodera J. D. and Noé F., *Curr. Opin. Struct. Biol.*, **25** (2014) 135.
[22] Hegger R. and Stock G., *J. Chem. Phys.*, **130** (2009) 034106.
[23] Schaudinnus N., Bastian B., Hegger R. and Stock G., *Phys. Rev. Lett.*, **115** (2015) 050602.
[24] Schaudinnus N., Lickert B., Biswas M. and Stock G., *J. Chem. Phys.*, **145** (2016) 184114.
[25] te Vrugt M. and Wittkowski R., *Phys. Rev. E*, **99** (2019) 062118.
[26] Altaner B., Polettini M. and Esposito M., *Phys. Rev. Lett.*, **117** (2016) 180601.
[27] Netz R. R., *J. Chem. Phys.*, **148** (2018) 185101.
[28] Grote R. F. and Hynes J. T., *J. Chem. Phys.*, **73** (1980) 2715.
[29] Bussi G., Laio A. and Parrinello M., *Phys. Rev. Lett.*, **96** (2006) 090601.
[30] Schnurr B., Gittes F., MacKintosh F. and Schmidt C., *Macromolecules*, **30** (1997) 7781.
[31] Mokshin A. V., Yulmetyev R. M. and Hänggi P., *Phys. Rev. Lett.*, **95** (2005) 200601.
[32] Shin H. K., Kim C., Talkner P. and Lee E. K., *Chem. Phys.*, **375** (2010) 316.
[33] Córdoba A., Schieber J. D. and Indei T., *Phys. Fluids*, **24** (2012) 073103.
[34] Darve E., Solomon J. and Kia A., *Proc. Natl. Acad. Sci. U.S.A.*, **106** (2009) 10884.
[35] Carof A., Vuilleumier R. and Rotenberg B., *J. Chem. Phys.*, **140** (2014) 124103.
[36] Li Z., Lee H. S., Darve E. and Karniadakis G. E., *J. Chem. Phys.*, **146** (2017) 014104.
[37] Lei H., Baker N. A. and Li X., *Proc. Natl. Acad. Sci. U.S.A.*, **113** (2016) 14183.
[38] Ma L., Li X. and Liu C., *J. Chem. Phys.*, **145** (2016) 204117.
[39] Jung G., Hanke M. and Schmid F., *J. Chem. Theory Comput.*, **13** (2017) 2481.
[40] Jungblut S. and Dellago C., *J. Chem. Phys.*, **142** (2015) 064103.
[41] Dorosz S., Voigtmann T. and Schilling T., *EPL*, **113** (2016) 10004.
[42] ten Wolde P. Rein, Ruiz-Montero M. J. and Frenkel D., *J. Chem. Phys.*, **104** (1996) 9932.
[43] Steinhardt P. J., Nelson D. R. and Ronchetti M., *Phys. Rev. B*, **28** (1983) 784.
[44] Kuhnhold A., Meyer H., Amati G., Pelagejcev P. and Schilling T., *Phys. Rev. E*, **100** (2019) 052140.
[45] Kuipers J. and Barkema G., *Phys. Rev. E*, **82** (2010) 011128.