

Bootstrapping time correlation functions of molecular dynamicsNicolas Desbiens^{✉*} and Philippe Arnault^{✉†}
CEA, DAM, DIF, 91297 Arpajon, FranceWilliam Weens
CEA, DAM, DIF, 91297 Arpajon, France
*Laboratoire en Informatique Haute Performance pour le Calcul et la Simulation, 91680 Bruyères-le-Châtel, France*Vincent Dubois[‡]
CEA, DAM, DIF, 91297 Arpajon, FranceGuillaume Perrin
COSYS, Université Gustave Eiffel, 77420 Champs-sur-Marne, France

(Received 20 July 2021; accepted 4 November 2021; published 22 November 2021)

Molecular dynamics is often considered as a numerical experiment. The error bars on the results are therefore mandatory, but sometimes difficult to determine and computationally demanding. As a low-cost approach, we describe the application of the bootstrap (BS) method to the quantification of uncertainties pertaining to the time correlation functions. We chose the autocorrelation functions of velocity and interdiffusion current for a binary ionic mixture as a test bed, and we assessed the merit of the Darken approximation relating both of them. The intrinsic errors related to phase space sampling is investigated comparing the BS method with the reference method of replica. We also study how the BS method can assist in addressing the finite-size effects.

DOI: [10.1103/PhysRevE.104.055310](https://doi.org/10.1103/PhysRevE.104.055310)**I. INTRODUCTION**

Molecular dynamics is one of the most realistic simulation tools available to study the behavior of matter. Since its inception in the 1950s, with the high computing power of the new electronic devices, it has laid down new perspectives for the advancement of science [1]. Nowadays, this method reaches a mature state with still ongoing research and developments. This is accompanied by an emerging plea to take full advantage of statistical methods to quantify uncertainties of simulation results [2–11]. Such error analysis becomes a critical issue to advance progress in the theoretical framework as well as in the applications of the method. This is a particularly acute issue for transport coefficients in the extreme conditions met in inertial confinement fusion (ICF) experiments [12], but a proper account of uncertainties could also be beneficial to the modeling of equation of state [13].

Actually, the quantification of uncertainties is more numerically demanding for dynamical quantities, like transport coefficients and sound speed, than for static quantities, like pressure and energy. Indeed, the static properties are related to average of the instantaneous values of observables, whereas the dynamical quantities are related to their time correlation functions. This dichotomy is not always the rule when en-

trophy is computed from the velocity correlation function, for instance [14–16].

The time correlation functions of molecular dynamics simulations incorporate the relaxation of fluctuations around equilibrium, as if they were small external excitations of the system. Their analysis gives access to the preferential modes of excitation of the system such as sound waves [17]. The fluctuation-dissipation theorem also allows us to relate some of these functions to the transport coefficients of the hydrodynamic equations [18].

It is computationally demanding to quantify the uncertainty on these time correlation functions, since the estimate based on grouping time sequences by blocks and performing a statistical analysis of the deviations between blocks requires very long duration of the simulation [19,20]. Within this “block averaging” method [21,22], the blocks can also be replaced by several simulations, considered as “replica,” starting with different initial values to enhance the sampling of phase space. This requirement of long simulation and/or several replicate simulations is even more challenging for costly *ab initio* simulations, where the potential energy surface is recomputed at each time step.

Besides the need to sample thoroughly phase space, the molecular simulations also face an issue of finite-size effects due to the practical limitation on the number of atoms to evolve in the simulation box. This is particularly acute in critical-point conditions [23]. When periodic boundary conditions are used as a remedy to approximate the thermodynamic limit of large number of particles, artifacts incompatible with

*nicolas.desbiens@cea.fr

†philippe.arnault@cea.fr

‡vincent-jp.dubois@cea.fr

a fluid behavior may appear [24]. Furthermore, spatial length scale of correlation can also be squeezed in a too small simulation box [9,25]. To our knowledge, the only way to control the finite-size effect is to perform simulations with larger and larger number of atoms. Since this solution can become prohibitively costly, extrapolation procedures have been proposed with more or less success [24,26–31].

As a less expensive alternative to the computationally expensive quantification of uncertainty related to phase space sampling, we investigate in this paper the bootstrap (BS) method [32,33]. It is very simple to implement as a postprocessing. With only one single simulation, it provides the poor man’s tool to begin with. For *ab initio* simulations, it might be the only path to uncertainties quantification at a reasonable cost. As compared to other methods, the bootstrap approach is a very economical and fast way to estimate the error bars, providing the metric to investigate finite-size effects.

The BS method is based on resampling without adding new information. To approximate the sampling distribution of a statistic, one gets “surrogate population” by resampling with replacement from the sample data at hand, creating a large number of synthetic samples known as bootstrap samples. The sample summary (mean value of some observable, for instance) is then computed on each of these BS samples. The histogram of the set of computed values is the bootstrap distribution of the statistic. This approach to uncertainty quantification makes no assumption about the particular statistic of the observable of interest, apart for independence of the samples.

Since the BS method assumes that the sample data are uncorrelated, it needs some adaptation when applied to time series, such as the time correlation functions of molecular dynamics. In this paper, we adapt to the autocorrelation functions (ACFs) of molecular dynamics the scheme of moving-block bootstrap [34,35], where the resampling is performed on blocks of consecutive data, in order to preserve the underlying structure of correlation.

BS technics is rarely applied in the context of molecular dynamics for the properties of systems in thermodynamic equilibrium. Only recently its efficiency was probed by Crysup and Shanbhag [36] for the self-diffusion coefficient as obtained from the mean square displacement, according to Einstein’s relation [18]. To our knowledge, a bootstrap analysis of time correlation functions has not yet been performed. The uncertainty pertaining to their determination by molecular dynamics is often estimated according to Zwanzig’s analysis [7,9,37] under the assumption that the observable follows a Gaussian process.

In this paper, we assess the merit of the BS method at providing with reasonable numerical resources some reliable estimates of error for the interdiffusion current autocorrelation function (JACF) and the velocity autocorrelation function (VACF) giving respectively the mutual diffusion and the self-diffusion coefficients in the Green-Kubo formalism [18]. The former is an essential transport property for the hydrodynamic modeling of mixing layers between two different media, with numerous applications in technology [38], ICF [39,40], and astrophysics [41]. The latter is useful in obtaining approximation to the mutual diffusion through the Darken relation [42,43], which relates this coefficient solely to the self-diffusion coefficients of each component in the mixture.

This approximation is tested in many recent papers [44–49] as a robust estimate.

As an illustration of the BS method, we chose to reproduce the results obtained by Hansen, Joly, and McDonald [50] in 1985 for a binary ionic mixture (BIM) of ionized hydrogen and helium. Since the Darken relation can be derived by neglecting all cross-correlations between different atoms, there is a relation between the JACF and the VACFs, that Hansen *et al.* tested in three situations of weak, intermediate, and strong correlations among the ions. These correlations are quantified by only one dimensionless coupling parameter Γ , which gives the magnitude of the Coulomb potential energy of neighboring particles relative to their kinetic energy, defined later. Interestingly, the Coulomb potential is the power-law interaction with the longest range exacerbating the possibility for the large spatial correlations not to accommodate within the finite size of the simulation box. For potentials with shorter range, cooperative effects lead to hydrodynamic-like interaction that are also long-range [51]. Nowadays, it is possible to perform many replica simulations in order to investigate how the BS method is efficient at providing error estimates. It is also possible to increase the number of atoms used in the study of Hansen *et al.* to evaluate finite-size effects and the convergence to the thermodynamic limit using BS errors. This investigation can serve as a reference for extrapolating the efficiency of the BS method to more numerically demanding *ab initio* simulations.

We first recall the basics, the definition of the VACF and the JACF, and the method of replica in Sec. II. This serves to introduce the correlation time of ACF and how it is used in the partition of the simulation into uncorrelated sequences. The application of the BS method to the ACFs of molecular dynamics is described in Sec. III with specific issues for the average in time and the influence of particles. Our molecular dynamics simulations of the BIM are described in Sec. IV, and the results of our analysis in Sec. V.

II. THEORY

A. Zwanzig’s analysis of ACF

For the sake of completeness, and to introduce the concepts and notations, we recall Zwanzig’s analysis [7,9,37] of the statistical uncertainty quantification for ACF under the Gaussian process approximation.

Consider a dynamical variable $a(t)$ whose evolution in time is given by molecular dynamics. In classical statistical mechanics, its time correlation function $C(t)$ is defined as

$$C(t) = \langle a(s) a(s+t) \rangle, \quad (1)$$

where the average denoted by $\langle \rangle$ is an ensemble average over an equilibrium distribution of initial system points in phase space. According to Liouville’s theorem, $C(t)$ does not depend on the arbitrary time origin s [18].

When using only one molecular simulation, the ensemble average is replaced by a time average under the ergodic assumption

$$C(t) = \lim_{T \rightarrow \infty} C_T(t), \quad (2a)$$

$$C_T(t) = \frac{1}{T - t_{\max}} \int_0^{T-t_{\max}} ds a(s) a(s+t), \quad (2b)$$

where t_{\max} is the maximum time lag one wants to compute, for which $C(t_{\max}) \approx 0$. In the time average, the limiting factor to convergence is the duration T of the simulation.

When the duration T of a simulation is not long enough to completely sample the phase space, the ACF is estimated with a residual error $\epsilon_T(t)$ given by

$$\begin{aligned} \epsilon_T(t) &= C_T(t) - C(t), \\ &= \frac{1}{T} \int_0^T ds [a(s)a(s+t) - \langle a(s)a(s+t) \rangle], \end{aligned} \quad (3)$$

where the invariance of $C(t)$ with the time origin s has been used, and we redefine the duration ($T - t_{\max}$) to T for the sake of readability. The ensemble average of this deviation vanishes, $\langle \epsilon_T(t) \rangle = 0$, but its second moment gives the standard error $\sigma_T(t)$ in the estimation of $C(t)$

$$\sigma_T^2(t) = \langle \epsilon_T^2(t) \rangle. \quad (4)$$

This variance involves computing a fourth-order correlation function of $a(t)$ since

$$\begin{aligned} \sigma_T^2(t) &= \frac{1}{T^2} \int_0^T ds_1 \int_0^T ds_2 [\langle a(s_1)a(s_1+t)a(s_2)a(s_2+t) \rangle \\ &\quad - \langle a(s_1)a(s_1+t) \rangle \langle a(s_2)a(s_2+t) \rangle]. \end{aligned} \quad (5)$$

Zwanzig and Ailawadi [37] proposed to assume that the observable $a(t)$ follows a Gaussian process to simplify the expression of the standard error $\sigma_T(t)$. Indeed, in this case, a fourth-order correlation is given by second-order correlations according to $\langle a_1 a_2 a_3 a_4 \rangle = \langle a_1 a_2 \rangle \langle a_3 a_4 \rangle + \langle a_1 a_3 \rangle \langle a_2 a_4 \rangle + \langle a_1 a_4 \rangle \langle a_2 a_3 \rangle$. Since the second-order correlations are nothing else than the ACFs, $C(t)$, which are independent of the time origin, it is possible to perform one of the two integrations with the change of variable $s = s_2 - s_1$ giving rise to the following approximation to $\sigma_T(t)$, using the subscript ‘‘Z’’ for Zwanzig:

$$\sigma_Z^2(t) = \frac{2}{T} \int_0^T ds [C^2(s) + C(t+s)C(t-s)], \quad (6)$$

where the parity of the integrand has been used. The estimator of the standard error $\sigma_Z(t)$ can be evaluated by using Eq. (6) with the ensemble average $C(t)$ replaced by the time average $C_T(t)$ obtained in the molecular simulation at hand.

Zwanzig and Ailawadi [37] defined a mean relaxation timescale τ by

$$\tau = 2 \int_0^{+\infty} ds \frac{C^2(s)}{C^2(0)} \quad (7)$$

and concluded that the standard error $\sigma_T(t)$ is proportional to $\sqrt{2\tau/T}$. Indeed, for T chosen high enough, it varies from $\sqrt{2\tau/T} C(0)$ at $t = 0$ to $\sqrt{\tau/T} C(0)$ at $t = T$ according to Eq. (6). The defining equation (7) of τ is an exact result if the ACF decays exponentially as $\exp(-t/\tau)$. We shall use it in the numerical analysis of molecular dynamics simulations.

Zwanzig and Ailawadi [37] also evaluated the standard error $\tilde{\sigma}_T(t)$ of the ACF, $\tilde{C}_T(t)$, normalized by its value at null time lag $C_T(0)$, with the use of the time average instead of the ensemble average $C(0) = \langle a(s)a(s) \rangle$

$$\tilde{C}_T(t) = \frac{C_T(t)}{C_T(0)}. \quad (8)$$

They showed that this standard error $\tilde{\sigma}_T(t)$ is greatly reduced as compared to $\sigma_T(t)$ for the ACF values at short time range. They proposed the following approximation to $\tilde{\sigma}_T(t)$:

$$\tilde{\sigma}_Z^2(t) \approx \frac{2\tau}{T} \left[1 - \frac{C(t)}{C(0)} \right]^2. \quad (9)$$

In obtaining this result, the main approximation is to replace $C_T(0)$ by $C(0) + \epsilon_T(0)$ in order to Taylor-expand the denominator of $\tilde{C}_T(t)$ to first order in $\epsilon_T(0)/C(0)$.

One could argue that no gain in accuracy is obtained by the latter estimator, $\tilde{C}_T(t)$, unless a better estimate of $C(0)$ than $C_T(0)$ is available. Nevertheless, this partition of the uncertainty between $C_T(0)$ and $\tilde{C}_T(t)$ is useful if an analytical fit of $\tilde{C}_T(t)$ is required as is often the case to compute transport coefficients from the integral in time of ACFs following the Green-Kubo formulas [11]. Indeed, this time integral suffers from propagation and accumulation of errors [9], which can be alleviated using $\tilde{C}_T(t)$ instead of $C(t)$ to find an appropriate fit to the ACF.

B. Uncorrelated sequences

The ACF obtained by time average, $C_T(t)$, can be given a more practical expression that emphasizes the role of the correlation time τ , Eq. (7) [11],

$$\begin{aligned} C_T(t) &= \frac{1}{M_T} \sum_{k=1}^{M_T} a(s_k) a(s_k + t), \\ s_k &= (k-1) \alpha \tau, \end{aligned} \quad (10)$$

where the possibility is open to omit the values of time origin s , which are redundant due to correlation.

The parameter α is the number of correlation times τ between consecutive time origins, s_k and s_{k+1} . It controls how the molecular dynamics trajectory is partitioned into more or less independent sequences. Apart from the time savings afforded by this reduction in time origins to use, it is important to realize that, if the value of α is too low, the redundant time blocks do not bring information, whereas, if it is too large, some information is missing in the time average. We shall see in Sec. VB that the best compromise is given by $\alpha = 1$ when a benchmark is done against the replica method.

The number M_T of time origins s_k depends on the ratio between the duration T of the trajectory and the lag, equal to $\alpha \tau$, between them, but it depends also on the largest time value t_{\max} for which the ACF $C_T(t)$ is computed, otherwise $(s_{M_T} + t_{\max})$ may exceed the total duration T . It is practical to introduce a parameter β by $t_{\max} = \beta \tau$. It controls the loss of correlation where $C_T(t_{\max}) \approx 0$. It should exceed at least 4τ . In this study, we adopt the conservative choice of $\beta = 15\tau$. In Fig. 1 a scheme shows the partition of the duration of the simulation into a sequence of time origins, and explains why the value of M_T must fulfill the following constraint:

$$\alpha (M_T - 1) + \beta < T/\tau < \alpha M_T + \beta. \quad (11)$$

C. Replica sampling

A reference value of the standard error $\sigma_T(t)$ can be obtained by an ensemble average over R replicate simulations.

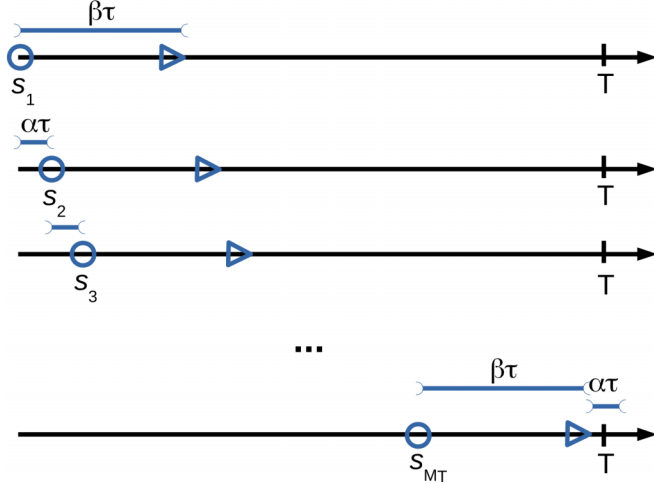


FIG. 1. Positions of the time sequences of duration $\beta\tau$ used in the calculation of the ACF. The successive time origins s_k are separated by $\alpha\tau$. The last time sequence verifies $s_{M_T} + \beta\tau < T$ and $s_{M_T} + \beta\tau + \alpha\tau > T$, where T is the duration of the simulation.

For each replicate simulation (r), a new value of the ACF is computed

$$C_T^{(r)}(t) = \frac{1}{M_T} \sum_{k=1}^{M_T} a^{(r)}(s_k) a^{(r)}(s_k + t),$$

$$s_k = (k - 1) \alpha \tau. \quad (12)$$

The distribution of replicas is characterized by its mean $C_T^{(R)}(t)$ and its variance $\sigma_R^2(t)$ given by

$$C_T^{(R)}(t) = \frac{1}{R} \sum_{r=1}^R C_T^{(r)}(t), \quad (13a)$$

$$\sigma_R^2(t) = \frac{1}{R-1} \sum_{r=1}^R [C_T^{(r)}(t) - C_T^{(R)}(t)]^2. \quad (13b)$$

As the number R of replicas increases, the convergence of phase space sampling improves and $C_T^{(R)}(t)$ tends to the ensemble average $C(t)$, while the standard deviation over replicas $\sigma_R(t)$ divided by \sqrt{R} approaches the standard deviation of the mean $C(t)$. In this limit of large number R of replica, $\sigma_R(t)$ tends to the intrinsic error $\sigma_T(t)$ on each replica simulation due to the finite duration T of the simulation

$$\sigma_T(t) = \lim_{R \rightarrow \infty} \sigma_R(t). \quad (14)$$

To a given value of the parameter α , corresponds a partition in time origins, and Eq. (14) provides the corresponding error. In Sec. VB varying α unveils the partition into uncorrelated sequences that leads to the lowest error without redundancy.

D. Velocity and interdiffusion current ACFs

We shall focus our investigation on two qualitatively different dynamical variables $a(t)$ of a binary mixture: the velocity $\mathbf{v}_i^{(m)}(t)$ of each particle i of the component m as an individual property of the particles, and the interdiffusion current $\mathbf{J}(t)$

between the two components as a collective property of particles. A time correlation function is defined for each spatial component $v_{i,\chi}^{(m)}$ and J_χ with χ standing for the Cartesian coordinates x, y, z :

$$C_{v_{i,\chi}^{(m)}}(t) = \langle v_{i,\chi}^{(m)}(s) v_{i,\chi}^{(m)}(s+t) \rangle, \quad (15a)$$

$$C_{J_\chi}(t) = \langle J_\chi(s) J_\chi(s+t) \rangle, \quad (15b)$$

with

$$J_\chi = \frac{1}{\sqrt{N}} [x_2 J_{1,\chi} - x_1 J_{2,\chi}], \quad (15c)$$

where

$$J_{m,\chi} = \sum_{i=1}^{N_m} v_{i,\chi}^{(m)}. \quad (15d)$$

Here the binary mixture contains N_1 atoms of one species and N_2 atoms of the other species. Each species has an atomic concentration $x_m = N_m/N$, with $N = N_1 + N_2$. The factor $1/\sqrt{N}$ ensures that $C_{J_\chi}(t)$ does not scale with N [52].

In a fluid, the isotropy can be exploited to reduce fluctuations by averaging over the directions [11]

$$C_{\mathbf{v}_i^{(m)}}(t) = \frac{1}{3} [C_{v_{i,x}^{(m)}}(t) + C_{v_{i,y}^{(m)}}(t) + C_{v_{i,z}^{(m)}}(t)], \quad (16a)$$

$$C_{\mathbf{J}}(t) = \frac{1}{3} [C_{J_x}(t) + C_{J_y}(t) + C_{J_z}(t)]. \quad (16b)$$

The mutual diffusion coefficient is given by the Green-Kubo relation [18]

$$D_{12} = \frac{\mathcal{J}}{x_1 x_2} \int_0^\infty dt C_{\mathbf{J}}(t), \quad (17)$$

where \mathcal{J} is a thermodynamic factor

$$\mathcal{J} = x_1 x_2 \left. \frac{\partial^2 (G/Nk_B T)}{\partial^2 x_1} \right|_{P,T,N}, \quad (18)$$

where T is the temperature (k_B , Boltzmann constant), P the pressure, and G the Gibbs free energy of the mixture. This factor expresses the fact that diffusion is driven by the gradient of chemical potential, and that this gradient is driven by the gradient of atomic concentration. Other sources of diffusion include chemical potential gradients induced by temperature and pressure gradients [53].

At contrast to the interdiffusion current ACF, the velocity ACF can be further averaged over the particles

$$C_{\mathbf{v}^{(m)}}(t) = \frac{1}{N_m} \sum_{i=1}^{N_m} C_{\mathbf{v}_i^{(m)}}(t). \quad (19)$$

This VACF is then used to compute a characteristic self-diffusion coefficient D_m of each component, using the Green-Kubo relation [18]

$$D_m = \int_0^\infty dt C_{\mathbf{v}^{(m)}}(t). \quad (20)$$

At this stage, it is worthwhile noting the different way the numbers N_m of particles enters the calculation of the self-diffusion and mutual diffusion coefficients, D_m and D_{12} .

It is often stated that the VACF, respectively D_m , is more accurately determined than the JACF, resp. D_{12} , due to the average over N_m particles, with a statistical error reduced by a factor $1/\sqrt{N_m}$. This is warranted only if the trajectories of the particles are uncorrelated. Unfortunately, the interactions among the particles create such correlation, and the gain in accuracy is not obvious as demonstrated by Kim *et al.* [7]. Moreover, an accurate determination of these coefficients using molecular dynamics simulations presupposes that the thermodynamic limit is reached, i.e., that the results do not vary as N increases with N/V constant, where V is the volume of the system.

Using the BS method, we shall see in Sec. VB and VC that the VACF is indeed far more accurately determined than the JACF. We shall also investigate in Sec. VC how bootstrapping over the time origins and the particles provides a metrics to assess the convergence toward the thermodynamic limit as the number of atoms increases.

III. BOOTSTRAPPING

The BS method is applied whenever a complete exploration of the statistic is impractical. It provides information on the statistic from a limited sample of data. In this paper, we apply it to a single molecular dynamics simulation, and try to assess its merit as compared to more expensive approaches, which use many replica simulations with different initial conditions to improve the sampling of phase space, or simulations with larger and larger number of particles to get rid of finite-size effects and reach the thermodynamic limit.

Since phase space sampling encompasses the trajectories of every particle, a complete quantification of uncertainties is obtained by a simultaneous BS in time and particles. From the distribution of BS samples of an ACF, its uncertainty can be quantified as well as the uncertainty of any functional of this ACF, in particular, the transport coefficient given by its time integral.

A. Bootstrapping time origins

Bootstrapping the values of time origin s_k allows one to preserve the underlying correlation of the time series $a(t)$ in the spirit of the moving-block BS method [34,35]. A number B of BS samples is generated by drawing with replacement M_T time origins $s_k^{(b)}$ from the original set. Each BS sample (b) is used to compute a new value of the ACF

$$C_T^{(b)}(t) = \frac{1}{M_T} \sum_{k=1}^{M_T} a(s_k^{(b)}) a(s_k^{(b)} + t),$$

$$s_k = (k - 1) \alpha \tau. \quad (21)$$

The BS distribution is characterized by its mean $C_T^{(B)}(t)$ and its variance $\sigma_B^2(t)$ given by

$$C_T^{(B)}(t) = \frac{1}{B} \sum_{b=1}^B C_T^{(b)}(t), \quad (22a)$$

$$\sigma_B^2(t) = \frac{1}{B-1} \sum_{b=1}^B [C_T^{(b)}(t) - C_T^{(B)}(t)]^2. \quad (22b)$$

As the number B of BS samples increases, the bias between $C_T^{(B)}(t)$ and $C_T(t)$ is expected to vanish, while the BS standard deviation $\sigma_B(t)$ approaches the standard error $\sigma_T(t)$. We shall see in Sec. VB that this is warranted only for values of α greater than unity that exclude correlated sequences.

B. Bootstrapping particles

Bootstrapping the particles allows one to estimate the influence on the ACFs of the limited number of particles in the simulation. Doing so, we address several interrelated questions: how do fluctuations among the trajectories of particles affect the average over N_m individual VACFs to produce a global VACF, Eq. (19)? How do they affect the collective variable \mathbf{J} , Eq. (15d), and as a consequence the JACF? Are there significant artifacts due to the periodic boundary conditions? At constant density, is the number of particles enough to get a large enough simulation box to accommodate all the correlation length scales? The brute-force approach to tackle globally these issues is to perform simulations with increasing number of particles and to monitor the convergence of the ACFs. We shall see in Sec. VC that the BS method provides a metrics to answer these questions.

A number B of BS samples is generated by drawing with replacement N particles from the original set. Each BS sample (b) is used to compute a new value $C^{(b)}(t)$ of an ACF $C(t)$. The BS distribution is characterized by its mean $C^{(B)}(t)$ and its variance $\sigma_B^2(t)$ given by analogs of Eqs. (22a) and (22b).

In the case of the VACF, $C = C_{v^{(m)}}$, the BS method provides an estimator of the mean, since according to Eq. (19) the VACF is the average of N_m individual VACFs of each atom of species m . As the number B of BS samples increases, the bias between $C_{v^{(m)}}^{(B)}(t)$ and $C_{v^{(m)}}(t)$ is expected to vanish, while the BS standard deviation $\sigma_B(t)$ is a measure of the fluctuations among the particles present in the simulation box.

In the case of the JACF, $C = C_{\mathbf{J}}$, the BS method is no longer an estimator of the mean, and the expectation of this estimator $C_{\mathbf{J}}^{(B)}(t)$ does not coincide with $C_{\mathbf{J}}(t)$. In order to interpret its variance $\sigma_B^2(t)$ as a measure of the fluctuations among the particles, it is necessary to renormalize each BS sample by a factor which restore the correct expectation.

To illustrate the difference between individual and collective correlations, and explain how the bootstrap method is adapted in each case, we consider a simplified situation corresponding to a set of N independent random variables x_i of vanishing expectation and equal variance σ^2 ,

$$E(x_i) = 0, \quad (23a)$$

$$E(x_i x_j) = \sigma^2 \delta_{ij}. \quad (23b)$$

This last assumption of independence of the random variables x_i amounts to neglect cross-correlations between different particles. This is the very hypothesis leading to the Darken relation.

The correlation function of an individual variable is represented by the summation S_2 of the square random variables

$$S_2 = \sum_{i=1}^N x_i^2, \quad (24)$$

whereas the correlation function of a collective variable is represented by the square of the summation S_1 of the random variables

$$S_1^2 = \left[\sum_{i=1}^N x_i \right]^2. \quad (25)$$

The expectation values of S_2 and S_1^2 are the same in this oversimplified example:

$$E(S_2) = E(S_1^2) = N \sigma^2. \quad (26)$$

The goal of the bootstrap method is to provide a simple and economical way to estimate the variance of S_2 and S_1^2 by sampling with replacement the random variables x_i . This gives rise to a distribution of bootstrap samples whose expectation and variance are expected to approximate those of S_2 and S_1^2 .

Selecting bootstrap samples can be represented by a set of independent random variables I_i with values ranging from 1 to N , and equal probability

$$P(I_i = j) = \frac{1}{N}. \quad (27)$$

Here i is the index of the selection of one particle, $i = 1, \dots, N$, and j is the index of the picked particle, $j = 1, \dots, N$.

When bootstrapping, the quantity S_2 is represented by

$$B_2 = \sum_{i=1}^N x_{I_i}^2 \quad (28a)$$

and the quantity S_1^2 by

$$B_1^2 = \left[\sum_{i=1}^N x_{I_i} \right]^2 = \sum_{i=1}^N \sum_{j=1}^N x_{I_i} x_{I_j}. \quad (28b)$$

Since the random variables I_i are independent, it is advantageous to separate the diagonal terms from the others in the above double sum, leading to

$$B_1^2 = \sum_{i=1}^N x_{I_i}^2 + \sum_{i=1}^N \sum_{j=1, j \neq i}^N x_{I_i} x_{I_j}. \quad (28c)$$

For the quantity B_2 , associated with correlation functions of individual variables, the bootstrap method provides an estimation for the error of the mean. In particular, it is assumed that there is no bias between the expectation values of B_2 and S_2 . Indeed, in the present simple case,

$$E(B_2) = \sum_{i=1}^N E(x_{I_i}^2), \quad (29)$$

with, according to the law of total expectation,

$$E(x_{I_i}^2) = \sum_{k=1}^N E(x_k^2 | I_i = k) P(I_i = k), \quad (30)$$

where $E(x_k^2 | I_i = k) = E(x_k^2) = \sigma^2$ is the conditional expectation of x_k^2 knowing that the value of $I_i = k$. Therefore

$$E(x_{I_i}^2) = \sum_{k=1}^N \sigma^2 P(I_i = k) = \sigma^2 \quad (31)$$

and

$$E(B_2) = N \sigma^2 = E(S_2). \quad (32)$$

The situation is different for the quantity B_1^2 , associated with the correlation functions of collective variables, where the expectation value $E(B_1^2)$ is different from $E(S_1^2)$. A correction to this bias involves scaling each bootstrap sample by the following factor:

$$f_B = \frac{E(S_1^2)}{E(B_1^2)}, \quad (33)$$

with

$$E(B_1^2) = \sum_{i=1}^N E(x_{I_i}^2) + \sum_{i=1}^N \sum_{j=1, j \neq i}^N E(x_{I_i} x_{I_j}). \quad (34)$$

The first term on the right-hand side is equal to $E(B_2)$. To compute the second term, we apply again the law of total expectation

$$E(x_{I_i} x_{I_j}) = \sum_{k=1}^N \sum_{l=1}^N E(x_{I_i} x_{I_j} | I_i = k, I_j = l) \times P(I_i = k) P(I_j = l), \quad (35)$$

using the independence of the random variables I_i . Again, the conditional expectation can then be evaluated,

$$E(x_{I_i} x_{I_j} | I_i = k, I_j = l) = E(x_k x_l) = \sigma^2 \delta_{kl}, \quad (36)$$

leading to

$$E(x_{I_i} x_{I_j}) = \sum_{k=1}^N \sigma^2 P(I_i = k) P(I_j = k) = \frac{\sigma^2}{N}, \quad (37)$$

and the final expression

$$E(B_1^2) = (2N - 1) \sigma^2. \quad (38)$$

The scaling factor $f_B = N/(2N - 1)$ tends to the value of 1/2 as N increases.

The calculations corresponding to the interdiffusion current ACF follow the same line of arguments, leading to a scaling factor

$$f_B = \frac{x_2^2 N_1 \sigma_1^2 + x_1^2 N_2 \sigma_2^2}{x_2^2 (2N_1 - 1) \sigma_1^2 + x_1^2 (2N_2 - 1) \sigma_2^2}, \quad (39)$$

which also tends to 1/2 as N_1 and N_2 increase.

Although the correlations between the atoms in the simulation are more intricate than in the preceding derivation, the ratio between $C_J^{(B)}(t)$ (computed without the renormalization) and $C_J(t)$ leads to the same value of 1/2.

IV. BIM SIMULATIONS

As a test bed to the BS method, we chose a binary mixture of ions in Coulomb interaction.

A. BIM system

The binary ionic mixture (BIM) model is an extension for mixtures of the one component plasma (OCP) model. The OCP is the simplest model of a Coulomb system, that

consists of a single species of ions immersed in a neutralizing background of electrons [54–58]. This corresponds to a limiting case of real matter under the extreme conditions of white dwarfs interiors [59], for instance, where the electrons are fully degenerate. Although this model is an oversimplification when electrons pile up around ions, an effective OCP can still be defined to serve as a prototype similar to the hard-sphere model in the theory of simple liquids [58]. Since its properties are either analytical or tabulated, it is used as a practical representation of Coulomb coupling in many situations encountered in hot dense plasmas. Its static and dynamical properties depend on only one dimensionless parameter, the Coulomb coupling parameter Γ ,

$$\Gamma = \frac{Q^2}{k_B T a}, \quad (40)$$

in atomic units, where $a = (4\pi n/3)^{-1/3}$ is the Wigner-Seitz radius, n is the ionic density, Q is the ionization, and T is here the temperature. As Γ increases, the OCP transits continuously from a nearly collisionless, gaseous regime for $\Gamma \ll 1$ to an increasingly correlated, liquidlike regime when $\Gamma \gtrsim 1$, up to the Wigner crystallization near $\Gamma = 175$.

The modeling of the BIM introduces three additional parameters: the atomic concentration x_1 ($x_2 = 1 - x_1$), the mass ratio M_2/M_1 , and the charge ratio Q_2/Q_1 .

There is a practical OCP system of units. Using the Wigner-Seitz radius a as the length scale, a natural timescale is the inverse of the plasma frequency ω_p , defined in atomic units by

$$\omega_p^2 = \frac{4\pi n Q^2}{M}, \quad (41)$$

where M is the mass of ions. Indeed, Newton's equations of motion are then independent of the mass M and the charge Q . Moreover, the density $n = 3/4\pi$ is constant and the temperature $T = 1/3\Gamma$ is a measure of the coupling. In these OCP units, all properties depend only on the Coulomb coupling parameter Γ (see Appendix A for details).

For the BIM of hydrogen (H) and helium (He), we adopted the plasma frequency ω_p computed with the mass M_1 and the charge $Q_1 = 1$ of H, leading to Newton's equations of motion parameterized by Γ , $M_2/M_1 = 4$, and $Q_2/Q_1 = 2$.

B. Molecular dynamics

The simulations comprise N_1 H ions and N_2 He ions in a cubic box periodically replicated in all directions. The initial particle positions are on a body-centered cubic lattice. Their initial velocities are assigned randomly from a Maxwell-Boltzmann distribution at the desired temperature.

In a microcanonical (NVE) simulation, at constant energy E , the particle trajectories are determined by solving Newton's equations of motion with the velocity Verlet integrator [60]. Since the range of the Coulomb interaction is larger than the simulation box, the force on each ion needs to include its interaction with the ions in the simulation box and in the periodically replicated cells. To this end, it is calculated using the Ewald summation technique [21].

Two thermostats are used: a simple velocity scaling is first performed at every time step of the equilibration phase

TABLE I. Conditions of molecular dynamic simulations: Γ is the coupling parameter, Δt the time step in unit of inverse plasma frequency ω_p , T_{eq} the duration of the equilibration phase, T_{prod} the duration of the production phase, and Δt_{ACF} the time step used to compute the ACFs.

Γ	$\omega_p \Delta t$	$T_{\text{eq}}/\Delta t$	$T_{\text{prod}}/\Delta t$	$\Delta t_{\text{ACF}}/\Delta t$
0.397	0.02	5000	15 000	6
3.992	0.025	6000	30 000	6
39.738	0.05	3000	15 000	3

to maintain the desired temperature until the spatial structure of the ions reaches equilibrium configurations; then the simulation transitions to the production phase in either NVE ensemble or the Gaussian isokinetic ensemble (NVK), at constant kinetic energy K [61–64]. The NVK thermostat is useful to compare the results with the microcanonical simulations of Hansen *et al.* [50] at given coupling parameter Γ , i.e., at given temperature T in OCP units (see Appendix A). It is also useful when comparing simulations with different numbers of atoms to avoid fluctuations of temperature (see Sec. VC). We adapted the velocity Verlet algorithm to the NVK ensemble using an even simpler discretization than Zhang's one [65], which is based on the leap-frog algorithm (Appendix B).

Hansen *et al.* [50] presented simulations of equimolar mixture of H and He at three values of the coupling parameter Γ equal to 0.4, 4, and 40, spanning a large range of correlations from weak to strong couplings. We tried to keep their conditions of simulation with a number of 250 atoms but the NVK thermostat needed smaller time steps Δt listed in Table I with the duration of the equilibration phase, T_{eq} , and of the production phase, T_{prod} . The time step Δt_{ACF} used to compute the ACFs is closer to Hansen's values.

The replica simulations, presented in Sec. VA, were performed keeping the number of 250 atoms with different random seeds to generate the initial velocities. Some replica simulations were also conducted at larger number of atoms, corresponding to the evaluation of finite-size effects presented in Sec. VC, where the number of atoms was increased to: 432, 1024, 2000, and 3456. These larger simulations used the same time steps as at 250 atoms.

V. RESULTS

We start the presentation of results with the reference method of replica. The error estimations from replica is then compared with what can be obtained from only one simulation using the BS method. The merit of the BS method is further assessed as concerns the finite-size effects. Finally, as an illustration of the use of the BS approach, we reexamine the comparison of Hansen *et al.* [50] between the ACF of the interdiffusion current and the Darken relation involving the VACFs of each species.

A. Replica

We performed 200 replica simulations for each value of the coupling parameter Γ to obtain an estimate of the intrinsic error σ_r of an ACF determination due to the finite duration

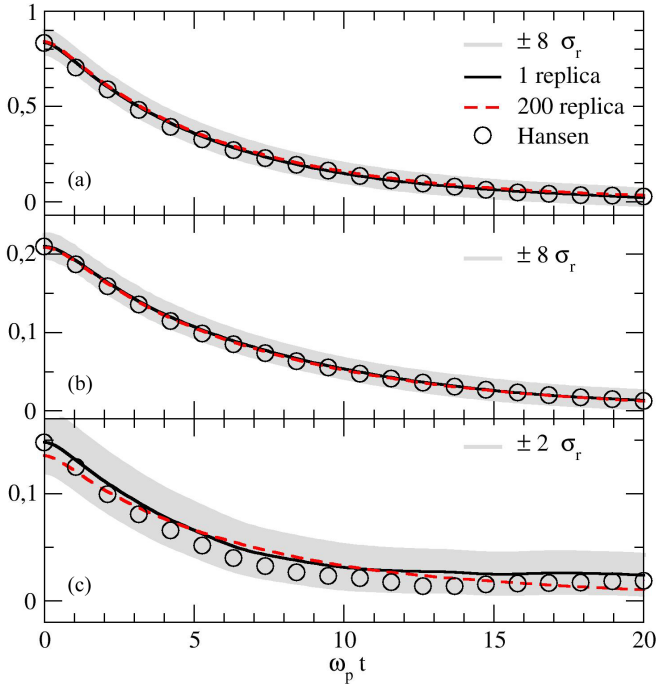


FIG. 2. ACFs at $\Gamma = 0.4$ as functions of time lag t , in unit of inverse plasma frequency ω_p , with error bars estimated using replica standard deviation σ_r : (a) VACF of H, (b) VACF of He, (c) JACF. All simulations use 250 atoms. Solid black lines with error bars represent only one replica, dashed red lines the average over 200 replicas, open circles the data of Hansen *et al.* [50].

T of a simulation, according to Eq. (14). In agreement with Zwanzig’s analysis presented in Sec. II A, the error is almost constant for all time lags, within a factor of 2. It suffices only to monitor its mean value. Whatever the ACF and the coupling value, by 50 replicas, the errors begin to converge and they no longer vary after 100 replicas. The case of weak coupling, at $\Gamma = 0.4$, is the slowest to converge.

The VACFs of H and He, and the JACF of the mixture are represented in Figs. 2–4 for one replica with the error bars derived from 200 replica. Since we are primarily interested in the intrinsic errors, the error bars comprise sometimes a very large multiple of the standard deviation σ_r of Eq. (14), for the three cases of coupling at $\Gamma = 0.4, 4$, and 40. This renders visible the errors when they are too small, here and in the following, even though there is no statistical sense in providing error bars larger than $2 - 4\sigma$.

At weak coupling ($\Gamma = 0.4$), the ACFs decrease with time lag monotonously with a somewhat larger correlation times than at stronger coupling (see Table II). Oscillations of the ACFs in time lag appear at intermediate coupling ($\Gamma = 4$) and develop into features of anticorrelation, where the ACFs exhibit negative values, at strong coupling ($\Gamma = 40$). These anticorrelations originate from a caging effect when a particle bounces against its nearest neighbors. We also plotted the data of Hansen *et al.* [50], which are reproduced with small but significant deviations, that we found difficult to interpret in absence of error bars in their paper. The averages over 200 replica are also plotted. Overall, they lie within $2\sigma_r$ from the results of only one replica. Comparing the error bars, it is

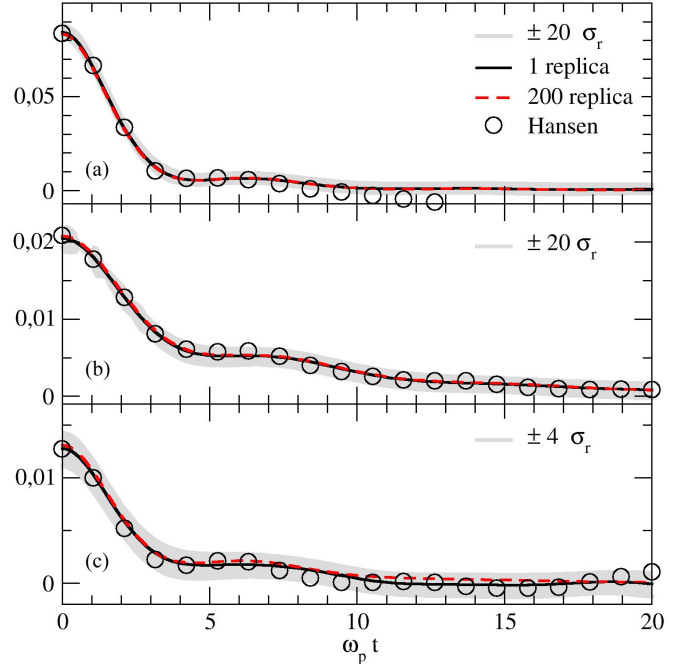


FIG. 3. ACFs at $\Gamma = 4$ as functions of time lag t , in unit of inverse plasma frequency ω_p , with error bars estimated using replica standard deviation σ_r : same conventions as in Fig. 2.

evident that the VACFs are determined with higher accuracy than the JACF.

Table II gathers the mean errors σ_r obtained with replica. As expected, the ratio between the errors on the VACFs, as ACF of individual observables, and the ones on the JACF,

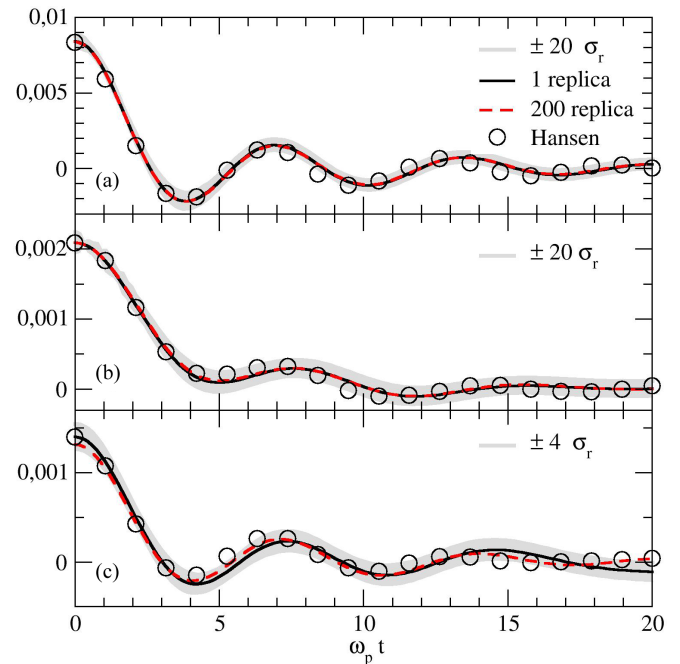


FIG. 4. ACFs at $\Gamma = 40$ as functions of time lag t , in unit of inverse plasma frequency ω_p , with error bars estimated using replica standard deviation σ_r : same conventions as in Fig. 2.

TABLE II. Comparison between the estimates of the mean intrinsic errors obtained with the replica method, σ_r , the BS on time origins, σ_t for $\alpha = 1$, the BS on time origins and atoms, σ_{ta} , and Zwanzig's Gaussian process approximation, σ_Z , for the ACFs, $C(t)$, of velocities for species H and He, and the JACF of the mixture. Γ is the coupling parameter. τ is Zwanzig's correlation time, Eq. (7), in unit of inverse plasma frequency ω_p . In the case of the VACFs, $\sigma_Z = \sqrt{2\tau/N_i T} C(0)$ whereas in the case of the JACF $\sigma_Z = \sqrt{2\tau/T} C(0)$, where T is the duration of the simulation. All simulations of the equimolar mixture used 250 atoms ($N_i = 125$).

$C(t)$	Γ	$\omega_p \tau$	$C(0)$	$\sigma_r/C(0)$	$\sigma_t/C(0)$	$\sigma_{ta}/C(0)$	$\sigma_Z/C(0)$
				(%)	(%)	(%)	(%)
VACF H	0.397	6.1	0.833	1.2	0.8	1.5	1.8
	3.992	2.7	0.0843	0.3	0.3	0.6	0.7
	39.738	2.7	0.0084	0.4	0.3	0.6	0.7
VACF He	0.397	7.7	0.210	1.3	1.0	1.9	2.0
	3.992	4.8	0.0205	0.5	0.4	0.8	1.0
	39.738	3.6	0.0021	0.5	0.4	0.7	0.8
JACF	0.397	7.1	0.148	13.	11.	15.	21.
	3.992	3.3	0.0128	4.5	3.6	5.7	9.3
	39.738	3.1	0.0014	3.8	3.8	5.1	9.2

as ACF of collective observable, is close to $1/\sqrt{N_i} \sim 1/10$, where $N_i = 125$ is the number of H or He atoms. The errors on ACFs are larger at $\Gamma = 0.4$ where the correlation time is longer. Zwanzig's Gaussian process approximation, σ_Z , is included. It overestimates the replica results by around a factor of 2. The mean errors σ_r obtained with BS on time origins are included anticipating the results of Sec. VB. The latter are close to the replica ones. The mean errors σ_{ta} obtained with BS on time origins and atoms are also included anticipating the results of Sec. VC. Bootstrapping also the atoms leads to mean errors larger by a factor between 1.3 and 2.

B. BS on time origins

The BS method is applied here as a post-processing of the simulation where the ACFs are recomputed from different sets of time origins, drawn with replacement from the origin set. The convergence of the error estimate on the VACFs and the JACF is rapid. Whatever the ACF and the coupling value, by 20 samples, the errors begin to converge and they no longer vary after 40 samples. The case of weak coupling, at $\Gamma = 0.4$, is the slowest to converge. From now on, all the results are presented for 50 samples.

As the BS method assumes that each time origin contributing to the average ACF is independent of the others, the resulting error estimate depends on the choice of the parameter α which controls the partition of the simulation into uncorrelated sequences (Sec. IIB). This parameter represents the number of Zwanzig's correlation time τ between consecutive time origins. The higher it is, the less time origins are used to compute the ACF, with a possible loss of information. The lower it is, the more time origins are used with possible redundancies.

We compare in Figs. 5–7 the BS estimates with the replica results for the mean errors on the VACF of H and He, and on

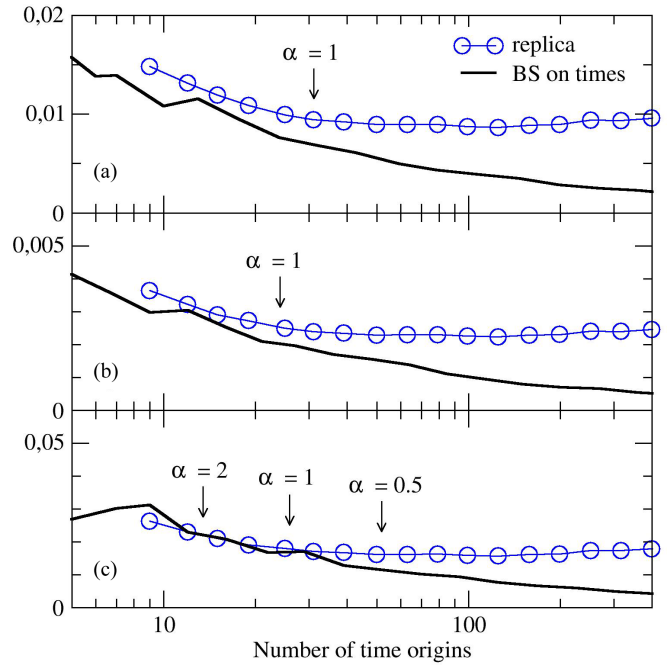


FIG. 5. Comparison at $\Gamma = 0.4$ between the BS estimates and the replica results for the mean intrinsic errors on ACF as functions of the number of time origins which depends on the value of α (the number of correlation time τ between successive time origins): (a) VACF of H, (b) VACF of He, (c) JACF. All simulations use 250 atoms. Solid black lines represent the BS estimates using only one replica, solid blue lines with open circles the results using 200 replicas.

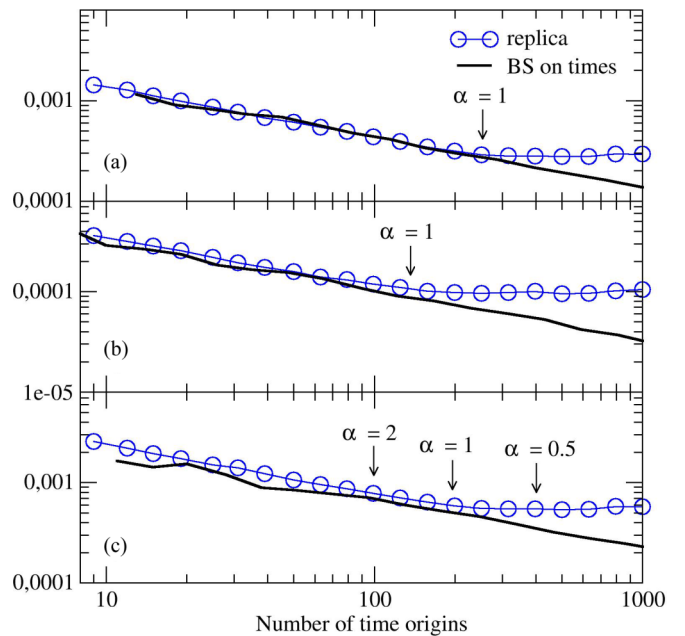


FIG. 6. Comparison at $\Gamma = 4$ between the BS estimates and the replica results for the mean intrinsic errors on ACF as functions of the number of time origins which depends on the value of α : same conventions as in Fig. 5.

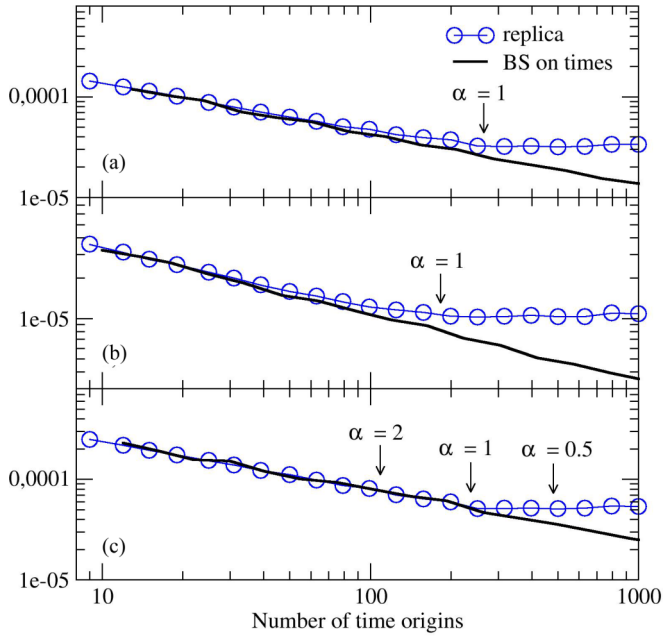


FIG. 7. Comparison at $\Gamma = 40$ between the BS estimates and the replica results for the mean intrinsic errors on ACF as functions of the number of time origins which depends on the value of α : same conventions as in Fig. 5.

the JACF of the mixture as functions of the number of time origins, corresponding to varying values of α , for the three Γ values of 0.4, 4, and 40.

For the low numbers of time origins, corresponding to values of α greater than 1, the errors of the methods of BS and replica are comparable. They decrease with increasing numbers of origins, corresponding to decreasing values of α , until α reaches unity, where a bifurcation occurs between the BS results and the replica ones.

For higher numbers of origins, corresponding to values of α lower than 1, the replica results plateau, whereas the BS estimates continue to decrease. This behavior originates from the inherent neglect of correlation between time origins in the BS approach, leading to a scaling of the BS error close to $1/\sqrt{M}$, where M is the number of time origins. This scaling is clearly apparent in Fig. 6 and 7. In the replica approach, the correlation between successive time origins, that are closer than a correlation time τ , does not grant more information when comparing several replica. From a practical point of view, although one could think using all the time origins in the BS sampling brings all the available information, this choice leads to an error estimate 3–4 times smaller than what is obtained using the reference method of replica.

This comparison between the methods of BS and replica highlights the need to control the number of time origins used in the BS approach. As expected, the origins must be separated by at least one correlation time τ . This leads to the prescription of the value of the control parameter $\alpha = 1$. With this prescription, the BS method provides reliable estimates of the intrinsic error of the ACFs from only one simulation, avoiding to perform several replica simulations.

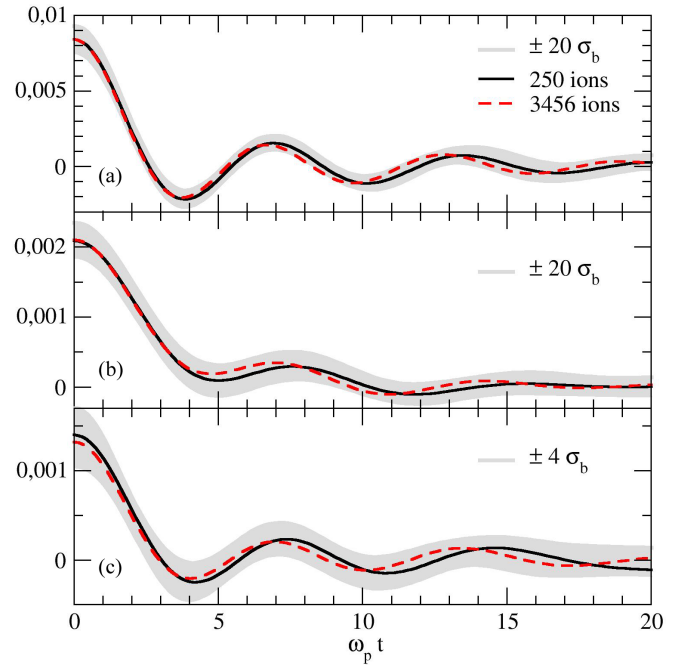


FIG. 8. Comparison at $\Gamma = 40$ between the ACFs from the simulations using 250 (solid black lines) and 3456 atoms (dashed red lines) as functions of time lag t , in unit of inverse plasma frequency ω_p : (a) VACF of H, (b) VACF of He, (c) JACF. The error bars around the ACFs of the simulation with 250 atoms represent a multiple of the standard error σ_b estimated bootstrapping both the time origins, with $\alpha = 1$, and the atoms.

C. BS on time origins and atoms

We now add to the BS on time origins a BS on atoms to get error estimates covering all the facets of phase space sampling. As a result, the error bars increase by a factor around 2 (see Table II). This provides a useful metrics to examine the finite-size effects and on how far from the thermodynamic limit a simulation is.

The case with the strongest finite-size effects is at the coupling of $\Gamma = 40$ as can be seen in Fig. 8, where the ACFs computed from a simulation with 250 atoms are compared with the ones computed with 3456 atoms. Error bars are drawn around the results corresponding to 250 atoms as a multiple of the standard errors σ_b obtained bootstrapping the time origins (with $\alpha = 1$) and the atoms. The deviations of the VACFs of H and He between 250 and 3456 atoms are large reaching $20\sigma_b$, whereas the deviation of the JACF stay within $3\sigma_b$.

It is instructive to describe a synthesis of the convergence of the results with the number of atoms, as displayed in Fig. 9. The maximum deviation of an ACF, obtained from a simulation with N atoms, from its values when $N = 3456$ is compared with the mean error σ_b determined by a bootstrap of time origins (with $\alpha = 1$) and of the atoms. For each ACF, all the data for different N are normalized to the corresponding value of σ_b at $N = 3456$.

The results of Fig. 8 are reported as red crosses symbols and red dashed dotted lines at $N = 250$. The maximum deviations between the VACF of H and He reach more than $10\sigma_b$, whereas it is only around $2\sigma_b$ for the JACF. Actually,

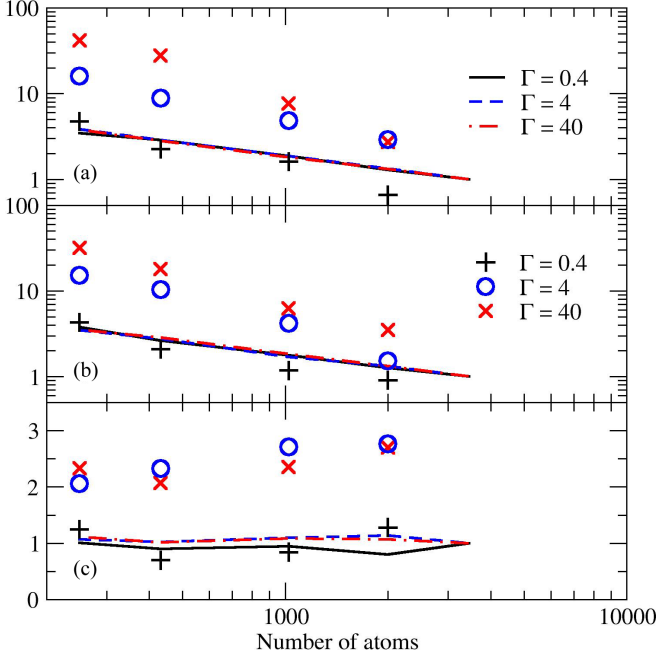


FIG. 9. Convergence of the ACFs as the number of atoms increases: (a) VACF of H, (b) VACF of He, (c) JACF. The symbols represent the maximum absolute deviations between the ACF from the simulation at the given number of atoms and the one from the simulation at 3456 atoms: black pluses correspond to $\Gamma = 0.4$, blue open circles $\Gamma = 4$, red crosses $\Gamma = 40$. The lines represent the mean errors σ_b estimated bootstrapping both the time origins, with $\alpha = 1$, and the atoms: black solid lines correspond to $\Gamma = 0.4$, blue dashed lines $\Gamma = 4$, red dashed dotted lines $\Gamma = 40$. For each ACF, all values are normalized to the value of σ_b corresponding to 3456 atoms.

the inspection of Fig. 8(c) shows that the deviations for the JACF are often less than this latter factor.

In the case of the JACF of the mixture, the results obtained at $N = 250$ do not vary when the number of atoms increases. Given the level of statistical errors obtained with the BS method, the convergence with N is already reached at $N = 250$. The interest in the BS determination of σ_b is to provide a metrics that can relate the results of only two simulations with increased number of atoms. Comparing $N = 250$ with $N = 432$, for instance, indicates that there is little to expect by increasing further the number of atoms. Any attempt to improve the accuracy of the JACF must pass by an increase in the duration of the simulation or by the production of replica.

In the cases of the VACF of H and He, both the maximum deviations and the BS mean error σ_b decrease as the number of atoms N increases. The decrease of σ_b is close to the scaling of $1/\sqrt{N}$ expected for the ACF of an individual observable that is averaged over the number of atoms. It was already observed in Table II when comparing the intrinsic errors of the VACF and the JACF.

At the lowest coupling of $\Gamma = 0.4$, the maximum deviations are close to σ_b , and consequently all the deviations for different values of N are within the statistical errors. Therefore, the finite-size effects are not significant at this level of statistical errors. Gains in accuracy can be obtained by

TABLE III. Self- and mutual diffusion coefficients, D_H , D_{He} , and D_{H-He} (with $\mathcal{J} = 1$), for simulations with $N = 250$ and 3456 ions, and the different coupling parameters Γ . In parentheses are given the relative standard deviations σ , in percentage, obtained bootstrapping the time origins and the atoms for the corresponding ACFs. The unit of the diffusion coefficients is $\omega_p a^2$, where ω_p is the inverse plasma frequency, and a is the Wigner-Seitz radius (see Appendix A).

Γ	N	D_H (%)	D_{He} (%)	D_{H-He} (%)
0.397	250	4.60 (9)	1.55 (13)	6.20 (62)
	3456	4.93 (3)	1.48 (3)	4.24 (83)
3.992	250	0.200(4)	0.108 (5)	0.129 (43)
	3456	0.198 (1)	0.112(1)	0.158 (37)
39.738	250	0.0105(5)	0.00600(4)	0.00804 (47)
	3456	0.0110(1)	0.00675(1)	0.00872 (30)

increasing the number of atoms, but also the duration of the simulation or the production of replica.

At the strong couplings of $\Gamma = 4$ and 40, the ratio between the maximum deviation and σ_b also decreases from around 10 at $N = 250$ to around 2 at $N = 2000$. This suggests that finite-size effects vanish at $N = 2000$. When comparing only two simulations with increased number of atoms, the discrepancies between both VACFs can be compared with the values of BS error estimates σ_b . It is large when comparing simulations at $N = 250$ and 432, and comparable for the simulations at $N = 1000$ and 2000. With such kind of comparisons, the BS method can assist in assessing the convergence of the ACFs with the number of atoms.

Finally, using the distribution of BS samples of the ACFs, we evaluated the self and mutual diffusion coefficients of the different BIMs for $N = 250$ and 3456 ions, gathered in Table III with their BS standard deviations. These results reflect the findings already obtained about the ACFs. However, they deserve a more comprehensive analysis, that is out of the present scope. We plan to investigate in a subsequent paper how the ACFs can be represented by analytic functions within a Bayesian approach. Then statistical errors on the transport coefficients will be examined thoroughly.

D. Darken relation revisited

As an illustration of the BS method in a typical application, we reexamine the accuracy of the Darken approximation giving the JACF of the mixture as a function of the VACFs of the components. This comparison was performed by Hansen *et al.* [50] without error bars. The BS method provides such error quantification, which informs about the quality of the approximation, but also how improvements on the conditions of simulation can lower the errors so as to better constrain the comparison.

The Darken relation, at the level of the ACFs, reads [50]

$$\tilde{C}_J(t) = \frac{C_J(t)}{C_J(0)} \approx x_2 m_2 \tilde{C}_{v(1)}(t) + x_1 m_1 \tilde{C}_{v(2)}(t), \quad (42)$$

with $m_i = M_i/(x_1 M_1 + x_2 M_2)$.

Since the VACFs are much more accurately determined than the JACF, we decided to neglect the intrinsic error of the Darken expression. In cases where the expressions to compare are determined with comparable accuracy, the BS method

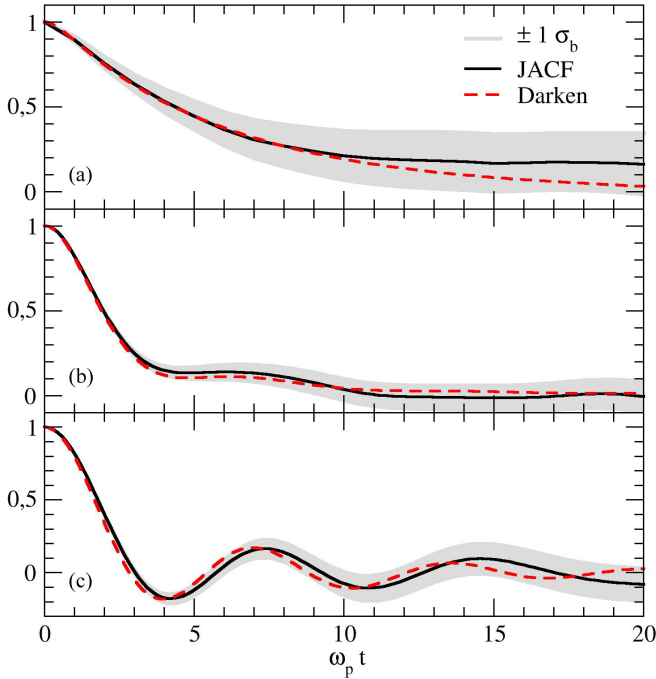


FIG. 10. Comparison between the JACF (black solid lines), normalized to its value at vanishing time lag, and the Darken approximation (red dashed lines) as functions of time lag t , in unit of inverse plasma frequency ω_p , at (a) $\Gamma = 0.4$, (b) $\Gamma = 4$, (c) $\Gamma = 40$. The error bars around the JACFs represent one standard error σ_b estimated bootstrapping both the time origins, with $\alpha = 1$, and the atoms.

can be easily applied to any function of the ACFs, naturally propagating the errors. The Darken expression is compared in Fig. 10 to the JACF with error bars obtained by bootstrapping both its time origins (with $\alpha = 1$) and the atoms.

As predicted by Zwanzig [37] and recalled in Sec. II A, the intrinsic error on the normalized JACF are greatly reduced at short time lag. It suffices to compare Fig. 10 with panels c) of Figs. 2–4.

The Darken approximation is within $1 - 2\sigma$ from the JACF, confirming its pertinence, but at the weakest coupling of $\Gamma = 0.4$ the error bars are much larger than at stronger coupling. There is room for improvement in this low coupling case to reduce the error bars by performing longer simulation and/or additional replica, as was discussed in the preceding sections.

The finite-size effects must also be examined since the VACFs are more sensitive to them than the JACF (see Sec. V C). We checked that the picture obtained with $N = 250$, does not change when increasing N .

VI. CONCLUSION

The bootstrap (BS) method is a promising approach for the quantification of uncertainty pertaining to the determination by molecular dynamics of time correlation functions. The point is that the usual methods involve long duration of the simulation and/or production of a large number of replica simulations to perform a statistical analysis by blocks [21]. At contrast, the BS method represents a tremendous short-

cut since it relies on only one simulation to produce error estimates. This direct uncertainty quantification can lead to important decisions on how to improve the accuracy through longer simulation or a larger number of atoms.

We have determined the conditions of application of the BS method for the autocorrelation functions (ACF) of individual and collective observables:

(1) The time sequences, that are averaged under the ergodic assumption, must be separated by a timescale of correlation to avoid redundancies. A comparison with the replica method indicates that Zwanzig’s correlation time defines an appropriate timescale [37].

(2) When bootstrapping atoms in the expression of a collective dynamic variable, the BS method is no longer an estimator of the mean and we have demonstrated that a renormalization of each BS sampling is necessary.

These findings were confirmed by the analysis of molecular simulations of a binary mixture of ions in Coulomb interaction [50]. As an individual observable, we chose the particle velocity and, as a collective one, the interdiffusion current. Their ACFs provide the transport coefficients of self-diffusion and mutual diffusion [18], that are important ingredients of the modeling of mixing layers [39]. This example of application is instructive and pertinent since the BS method allowed us to highlight every pieces of uncertainty affecting the powerful Darken approximation [42,43] relating the interdiffusion to the self-diffusions of each component of the mixture. We have addressed the issues of phase space sampling and finite-size effects within this framework. The monitoring of the convergence to the thermodynamic limit as the number of atoms increases is largely eased with error bars obtained from only one simulation via the BS method.

More applications of the BS method to time correlation functions of molecular dynamics would be welcome to study it in other contexts: shorter range potentials, different observables, *ab initio* simulations, etc.

In a second step of uncertainties quantification, we intend to use the BS error estimates on ACFs to better control the determination of transport coefficients within a Bayesian inference of the parameters of analytic fitting functions.

ACKNOWLEDGMENT

Luc Kazandjian is acknowledged for fruitful discussions.

APPENDIX A: OCP SYSTEM OF UNITS

The OCP units of space and time are respectively the Wigner-Seitz radius a and the inverse plasma frequency $1/\omega_p$. Let us note $\tilde{r} = r/a$ and $\tilde{t} = \omega_p t$ the reduced variables of space and time. Newton’s equations of motion

$$M \frac{d^2 \mathbf{r}_i}{dt^2} = - \sum_{j \neq i} \frac{Q^2}{r_{ij}^2} \frac{\mathbf{r}_{ij}}{r_{ij}} \quad (\text{A1a})$$

read in reduced variables

$$\frac{d^2 \tilde{\mathbf{r}}_i}{d\tilde{t}^2} = - \frac{1}{3} \sum_{j \neq i} \frac{1}{\tilde{r}_{ij}^2} \frac{\tilde{\mathbf{r}}_{ij}}{\tilde{r}_{ij}}, \quad (\text{A1b})$$

since

$$\omega_p^2 = \frac{4\pi n Q^2}{M} \quad (\text{A2})$$

and

$$\frac{4\pi}{3} a^3 n = 1, \quad (\text{A3})$$

where M is the particle mass and Q its charge.

As a consequence, there is no parameter in the equations of motion in reduced variables. An alternative view is to consider that the mass $M = 1$ is constant, as well as the charge $Q = 1/\sqrt{3}$. Moreover, the density is constant equal to

$$\tilde{n} = \frac{3}{4\pi}. \quad (\text{A4})$$

The only free parameter is the reduced temperature \tilde{T} that is given from T by the following calculation of the mean kinetic energy K per particle:

$$K = \frac{1}{N} \sum_i \frac{1}{2} M \left(\frac{d\mathbf{r}_i}{dt} \right)^2 = \frac{3}{2} k_B T, \quad (\text{A5a})$$

which reads in reduced variables

$$\tilde{K} = \frac{1}{N} \sum_i \frac{1}{2} \left(\frac{d\tilde{\mathbf{r}}_i}{d\tilde{t}} \right)^2 = \frac{3}{2} k_B \tilde{T}, \quad (\text{A5b})$$

with

$$k_B \tilde{T} = \frac{1}{3\Gamma}, \quad (\text{A5c})$$

where $\Gamma = Q^2/ak_B T$ is the Coulomb coupling parameter.

It is therefore evidence that every static or dynamic properties of the OCP depend only on the coupling parameter Γ (through the reduced temperature) when these properties are expressed in OCP units.

APPENDIX B: ALGORITHM FOR NVK DYNAMICS

The velocity Verlet algorithm has the advantage to start from a sole iteration where the positions \mathbf{r}_i , the velocities \mathbf{v}_i , and the forces \mathbf{F}_i acting on each particle, of mass M_i , are simultaneously given. It reads

$$\mathbf{r}_i^{(n+1)} = \mathbf{r}_i^{(n)} + \mathbf{v}_i^{(n)} \Delta + \frac{\mathbf{F}_i^{(n)}}{2M_i} \Delta^2, \quad (\text{B1a})$$

$$\mathbf{F}_i^{(n+1)} = \mathbf{F}(\mathbf{r}_1^{(n+1)}, \dots, \mathbf{r}_N^{(n+1)}), \quad (\text{B1b})$$

$$\mathbf{v}_i^{(n+1)} = \mathbf{v}_i^{(n)} + \frac{\mathbf{F}_i^{(n)} + \mathbf{F}_i^{(n+1)}}{2M_i} \Delta, \quad (\text{B1c})$$

where Δ is the time step, and the value of a quantity X at time $t = n\Delta$ is denoted $X^{(n)} = X(n\Delta)$. At the next time step, one computes first the new positions, second the new forces, and third the new velocities.

In the NVK ensemble, the trajectories of the atoms do not follow Newton's equations of motion but the following:

$$M_i \frac{d\mathbf{v}_i}{dt} = \mathbf{F}_i - \alpha M_i \mathbf{v}_i, \quad (\text{B2a})$$

with

$$\alpha = \frac{1}{2K} \sum_{j=1}^N \mathbf{F}_j \cdot \mathbf{v}_j, \quad (\text{B2b})$$

where

$$K = \frac{1}{2} \sum_{j=1}^N M_j v_j^2. \quad (\text{B2c})$$

The velocity Verlet algorithm then reads

$$\mathbf{r}_i^{(n+1)} = \mathbf{r}_i^{(n)} + \mathbf{v}_i^{(n)} \Delta + \frac{\mathbf{F}_i^{(n)}}{2M_i} \Delta^2 - \frac{1}{2} \alpha^{(n)} \mathbf{v}_i^{(n)} \Delta^2, \quad (\text{B3a})$$

$$\mathbf{F}_i^{(n+1)} = \mathbf{F}(\mathbf{r}_1^{(n+1)}, \dots, \mathbf{r}_N^{(n+1)}), \quad (\text{B3b})$$

$$\mathbf{v}_i^{(n+1)} = \mathbf{v}_i^{(n)} + \frac{\mathbf{F}_i^{(n)} + \mathbf{F}_i^{(n+1)}}{2M_i} \Delta - \frac{1}{2} [\alpha^{(n)} \mathbf{v}_i^{(n)} + \alpha^{(n+1)} \mathbf{v}_i^{(n+1)}] \Delta. \quad (\text{B3c})$$

Although $\alpha^{(n+1)}$ can be Taylor-expanded around $\mathbf{v}_i^{(n)}$ at first order in Δ , we made the following approximation:

$$\alpha^{(n+1)} = \alpha^{(n)}, \quad (\text{B4})$$

leading to a simpler expression of $\mathbf{v}_i^{(n+1)}$.

Since the present study is more focused on uncertainty quantification than on performance, this very simple algorithm is adopted. It reaches accurate results at the expense of a smaller time step than the more efficient algorithm presented in Ref. [64]. One difference between both algorithms is that it is adapted from the leapfrog scheme in Ref. [65], whereas the present algorithm modifies the velocity Verlet scheme.

- [1] G. Battimelli, G. Ciccotti, and P. Greco, *Computer Meets Theoretical Physics—The New Frontier of Molecular Simulation* (Springer, Cham, 2020).
- [2] A. Grossfield and D. M. Zuckerman, *Annu. Rep. Comput. Chem.* **5**, 23 (2009).
- [3] T. D. Romo and A. Grossfield, *J. Chem. Theory Comput.* **7**, 2464 (2011).
- [4] A. Nicholls, *J. Comput. Aided Mol. Des.* **28**, 887 (2014).

- [5] Y. Zhang, A. Otani, and E. J. Maginn, *J. Chem. Theory Comput.* **11**, 3537 (2015).
- [6] G. Pranami and M. H. Lamm, *J. Chem. Theory Comput.* **11**, 4586 (2015).
- [7] C. Kim, O. Borodin, and G. E. Karniadakis, *J. Comput. Phys.* **302**, 485 (2015).
- [8] P. N. Patrone, A. Dienstfrey, A. R. Browning, S. Tucker, and S. Christensen, *Polymer* **87**, 246 (2016).

- [9] K.-S. Kim, M. H. Han, C. Kim, Z. Li, G. E. Karniadakis, and E. K. Lee, *J. Chem. Phys.* **149**, 044510 (2018).
- [10] A. Grossfield, P. N. Patrone, D. R. Roe, A. J. Schultz, D. W. Siderius, and D. M. Zuckerman, *Living J. Comput. Mol. Sci.* **1**, 5067 (2018).
- [11] E. J. Maginn, R. A. Messerly, D. J. Carlson, D. R. Roe, and J. R. Elliott, *Living J. Comput. Mol. Sci.* **1**, 6324 (2018).
- [12] P. Grabowski, S. Hansen, M. Murillo, L. Stanton, F. Graziani, A. Zylstra, S. Baalrud, P. Arnault, A. Baczewski, L. Benedict, C. Blancard, O. Čertík, J. Clérouin, L. Collins, S. Copeland, A. Correa, J. Dai, J. Daligault, M. Desjarlais, M. Dharma-wardana *et al.*, *High Energy Density Phys.* **37**, 100905 (2020).
- [13] J. Gaffney, S. Hu, P. Arnault, A. Becker, L. Benedict, T. Boehly, P. Celliers, D. Ceperley, O. Certik, J. Clérouin, G. Collins, L. Collins, J.-F. Danel, N. Desbiens, M. Dharma-wardana, Y. Ding, A. Fernandez-Pañella, M. Gregor, P. Grabowski, S. Hamel *et al.*, *High Energy Density Phys.* **28**, 7 (2018).
- [14] S.-T. Lin, M. Blanco, and W. A. Goddard, *J. Chem. Phys.* **119**, 11792 (2003).
- [15] M. P. Desjarlais, *Phys. Rev. E* **88**, 062145 (2013).
- [16] G. Robert, P. Legrand, P. Arnault, N. Desbiens, and J. Clérouin, *Phys. Rev. E* **91**, 033310 (2015).
- [17] N. P. Kryuchkov, L. A. Mistryukova, V. V. Brazhkin, and S. O. Yurchenko, *Sci. Rep.* **9**, 10483 (2019).
- [18] J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 3rd ed. (Academic Press, Cambridge, 2006).
- [19] I. Bitsanis, M. Tirrell, and H. T. Davis, *Phys. Rev. A* **36**, 958 (1987).
- [20] R. Chitra and S. Yashonath, *J. Phys. Chem. B* **101**, 5437 (1997).
- [21] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 2017).
- [22] H. Flyvbjerg and H. G. Petersen, *J. Chem. Phys.* **91**, 461 (1989).
- [23] S. K. Das, M. E. Fisher, J. V. Sengers, J. Horbach, and K. Binder, *Phys. Rev. Lett.* **97**, 025702 (2006).
- [24] I.-C. Yeh and G. Hummer, *J. Phys. Chem. B* **108**, 15873 (2004).
- [25] J. Petrávic, *J. Chem. Phys.* **120**, 10188 (2004).
- [26] S. H. Jamali, L. Wolff, T. M. Becker, A. Bardow, T. J. H. Vlucht, and O. A. Moulτος, *J. Chem. Theory Comput.* **14**, 2667 (2018).
- [27] D. P. Sellan, E. S. Landry, J. E. Turney, A. J. H. McGaughey, and C. H. Amon, *Phys. Rev. B* **81**, 214305 (2010).
- [28] K.-S. Kim, C. Kim, G. E. Karniadakis, E. K. Lee, and J. J. Kozak, *J. Chem. Phys.* **151**, 104101 (2019).
- [29] C. J. Leverant, J. A. Harvey, and T. M. Alam, *J. Phys. Chem. Lett.* **11**, 10375 (2020).
- [30] A. T. Celebi, S. H. Jamali, A. Bardow, T. J. H. Vlucht, and O. A. Moulτος, *Mol. Simul.* **47**, 831 (2020).
- [31] T. Dornheim and J. Vorberger, *J. Chem. Phys.* **154**, 144103 (2021).
- [32] B. Efron, *Ann. Stat.* **7**, 1 (1979).
- [33] M. R. Chernick, *Bootstrap Methods: A Guide for Practitioners and Researchers* (John Wiley & Sons, New York, 2008).
- [34] E. Carlstein, *Ann. Stat.* **14**, 1171 (1986).
- [35] H. R. Kunsch, *Ann. Stat.* **17**, 1217 (1989).
- [36] B. Crysip and S. Shanbhap, *J. Comput. Methods Sci. Eng.* **19**, 387 (2019).
- [37] R. Zwanzig and N. K. Ailawadi, *Phys. Rev.* **182**, 280 (1969).
- [38] R. Krishna, *Chem. Eng. Sci.* **195**, 851 (2019).
- [39] K. K. Mackay and J. E. Pino, *Phys. Plasmas* **27**, 092704 (2020).
- [40] G. Viviconte, B.-J. Gréa, F. S. Godefert, P. Arnault, and J. Clérouin, *Phys. Rev. E* **100**, 063205 (2019).
- [41] D. Arnett, *Astrophys. J. Suppl. Ser.* **127**, 213 (2000).
- [42] L. S. Darken, *Trans. AIME* **175**, 184 (1948).
- [43] X. Liu, T. J. Vlucht, and A. Bardow, *Ind. Eng. Chem. Res.* **50**, 10350 (2011).
- [44] R. Krishna and J. M. van Baten, *Ind. Eng. Chem. Res.* **44**, 6939 (2005).
- [45] R. Krishna and J. M. van Baten, *Chem. Eng. Sci.* **153**, 174 (2016).
- [46] L. Wolff, S. H. Jamali, T. M. Becker, O. A. Moulτος, T. J. H. Vlucht, and A. Bardow, *Ind. Eng. Chem. Res.* **57**, 14784 (2018).
- [47] A. J. White, C. Ticknor, E. R. Meyer, J. D. Kress, and L. A. Collins, *Phys. Rev. E* **100**, 033213 (2019).
- [48] J. Clérouin, P. Arnault, B.-J. Gréa, S. Guisset, M. Vandenkoonaerde, A. J. White, L. A. Collins, J. D. Kress, and C. Ticknor, *Phys. Rev. E* **101**, 033207 (2020).
- [49] D. Rosenberger, N. Lubbers, and T. C. Germann, *Phys. Plasmas* **27**, 102705 (2020).
- [50] J. Hansen, F. Joly, and I. McDonald, *Physica A* **132**, 472 (1985).
- [51] K. H. Han, C. Kim, P. Talkner, G. E. Karniadakis, and E. K. Lee, *J. Chem. Phys.* **148**, 024506 (2018).
- [52] J. L. Lebowitz, J. K. Percus, and L. Verlet, *Phys. Rev.* **153**, 250 (1967).
- [53] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon Press, Oxford, 1992), p. 733.
- [54] J. P. Hansen, *Phys. Rev. A* **8**, 3096 (1973).
- [55] J. P. Hansen, I. R. McDonald, and E. L. Pollock, *Phys. Rev. A* **11**, 1025 (1975).
- [56] M. Baus and J.-P. Hansen, *Phys. Rep.* **59**, 1 (1980).
- [57] J. Daligault, *Phys. Rev. Lett.* **96**, 065003 (2006).
- [58] J. Clérouin, P. Arnault, C. Ticknor, J. D. Kress, and L. A. Collins, *Phys. Rev. Lett.* **116**, 115003 (2016).
- [59] D. Koester and G. Chanmugam, *Rep. Prog. Phys.* **53**, 837 (1990).
- [60] D. Frenkel and B. Smit, *Understanding Molecular Simulation*, 2nd ed. (Academic Press, San Diego, 2002).
- [61] D. J. Evans, W. G. Hoover, B. H. Failor, B. Moran, and A. J. C. Ladd, *Phys. Rev. A* **28**, 1016 (1983).
- [62] D. J. Evans and G. Morriss, *Phys. Lett. A* **98**, 433 (1983).
- [63] D. J. Evans and G. Morriss, *Chem. Phys.* **87**, 451 (1984).
- [64] P. Minary, G. J. Martyna, and M. E. Tuckerman, *J. Chem. Phys.* **118**, 2510 (2003).
- [65] F. Zhang, *J. Chem. Phys.* **106**, 6102 (1997).