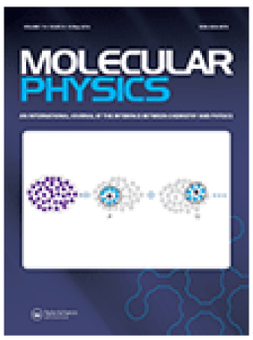


Taylor & Francis
Taylor & Francis Group

Molecular Physics

An International Journal at the Interface Between Chemistry and Physics

ISSN: 0026-8976 (Print) 1362-3028 (Online) Journal homepage: https://www.tandfonline.com/loi/tmph20

Force-sampling methods for density distributions as instances of mapped averaging

Apoorva Purohit, Andrew J. Schultz & David A. Kofke

To cite this article: Apoorva Purohit, Andrew J. Schultz & David A. Kofke (2019): Force-sampling methods for density distributions as instances of mapped averaging, Molecular Physics, DOI: 10.1080/00268976.2019.1572243

To link to this article: https://doi.org/10.1080/00268976.2019.1572243

MOLECULAR PHYSICS
https://doi.org/10.1080/00268976.2019.1572243

LIBLICE 2018 SPECIAL ISSUE

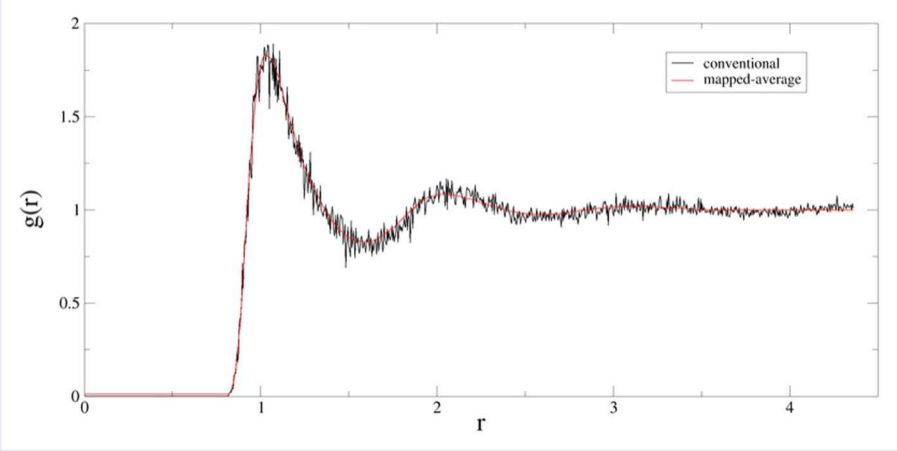
Force-sampling methods for density distributions as instances of mapped averaging

Apoorva Purohit, Andrew J. Schultz and David A. Kofke

Department of Chemical and Biological Engineering, The State University of New York, Buffalo, NY, USA

ABSTRACT

We show that two recently proposed methods for computing singlet and pair density distributions without histograms are particular implementations of the general mapped-averaging framework for deriving alternative ensemble averages for physical properties.



ARTICLE HISTORY
Received 13 November 2018
Accepted 11 January 2019

KEYWORDS
Statistical mechanics; density distributions; mapped averaging

1. Introduction

Spatial density functions are central quantities in the formalism of statistical mechanics, as they are used to characterise inhomogeneities and structural correlations in molecular systems. Singlet (one-body) density distributions are valuable when dealing with systems where interfaces, impurities, or inhomogeneous external fields generate non-uniformity, as well as systems that become inhomogeneous spontaneously, such as in crystals. Computational studies are often performed to measure singlet density distributions, which are used to study wetting/drying properties [1], phase coexistence [2], capillary effects [3], sedimentation [4], and adsorption [5]. Similarly, the pair distribution functions (or radial distribution functions) have been extensively used for disordered systems as a tool to understand the structure/property relationships like amorphization [6], solid dispersions [7], coarse graining [8], biomineralization processes [9], polymorphism and stability [10].

Given the wealth of knowledge available from these density distribution functions, their accurate measurement is important. However, the experimental uncertainty induced by quantum counting inefficiencies, experimental imprecision or sample inhomogeneities, can lead to errors in drawing meaningful conclusions from density distribution studies. Thus, computational studies are needed. Computationally, the standard approach to sample these density distribution functions is to discretise the space into a 3D grid (bins) and monitor the filling of the histograms in the course of the simulation. One issue with this approach is the tradeoff between precision and resolution: as the bin size decreases and the resolution of the measured density profile increases, the uncertainty in the calculated values increases because

CONTACT David A. Kofke kofke@buffalo.edu Department of Chemical and Biological Engineering, The State University of New York, Buffalo, NY 14260-4200, USA

the number of events contributing to each bin is proportional to the volume of the bin. Thus, the variance of the distributions diverges when the grid spacing tends to zero. Hence, there is a need for alternatives to the conventional counting based method conventionally used in literature.

In this regard, a method based on force sampling was recently developed that leads to a reduced variance of the results with respect to the counting based technique, achieving a finite variance even when the grid spacing tends to zero. The singlet density can be computed via sampling the instantaneous forces acting on the particles [11,12]:

$$\rho^{(1)}(\mathbf{r}) = \rho_o + \beta \int_{\gamma(\mathbf{r})} \mathrm{d}\mathbf{s} \cdot \mathbf{f}(\mathbf{s}) \tag{1}$$

where $\beta \equiv (k_B T)^{-1}$ with T the temperature and k_B the Boltzmann constant, ρ_o is the normalisation constant, $\gamma(\mathbf{r})$ represents a path that connects, say, the origin with position $\mathbf{r}$, and $\mathrm{d}\mathbf{s}$ is the differential line element. Also, $\mathbf{f}(\mathbf{r})$ is the force density at position $\mathbf{r}$, defined:

$$\mathbf{f}(\mathbf{r}) = \left\langle \sum_{i=1}^{N} \mathbf{f}_i \delta(\mathbf{r}_i - \mathbf{r}) \right\rangle \tag{2}$$

where $\mathbf{f}_i$ is the total force acting on molecule i, which is at position $\mathbf{r}_i$, δ is the Dirac delta function, and the angle brackets represent an ensemble average over configurations of the N molecules.

The calculation approach is to sample the force density during the simulation, and then compute $\rho(\mathbf{r})$ via spatial integration of it. This method significantly enhances the convergence of simulations compared to the conventional counting based method [12]. Additionally, the resolution used for recording the force density as a function of position has no effect on the statistical uncertainty of the calculated density, because only the integral of $\mathbf{f}$ is required in Equation (1).

Similarly, the pair distribution function can be computed via force sampling by [11]:

$$\rho^{(2)}(r) = \rho^2 - \frac{1}{V} \left\langle \sum_{i=1}^{N} \sum_{j<i} \frac{H(r_{ij} - r)}{4\pi r_{ij}^2} \beta(\mathbf{f}_j - \mathbf{f}_i) \cdot \hat{\mathbf{r}}_{ij} \right\rangle \tag{3}$$

$H(.)$ denotes the Heaviside function and $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ is the position vector between particles i and j, $r_{ij} = |\mathbf{r}_{ij}|$, and $\hat{\mathbf{r}}_{ij} = \mathbf{r}_{ij}/r_{ij}$ (Equation (3) differs slightly from the formula presented in [11], which omitted a factor of 2).

In a separate development, we introduced the 'mapped-averaging' framework, through which approximate theoretical results derived from statistical mechanics can be reintroduced into the underlying formalism, yielding new ensemble averages that give exactly the error in the theory [13]. Among other potential uses, this result allows direct evaluation of the correction to the theory by molecular simulation. To the extent that the theory is accurate, this correction will be small, and hence measured with small uncertainty. In favourable cases, the computational effort needed to evaluate a property to a given precision via mapped averaging may be many orders of magnitude less than that required by the conventional average [14–19].

In this paper, we extend the mapped-averging framework to derive histogram-free expressions for the singlet and pair density distribution functions. We show that particular cases of these expressions correspond to the force-based formulas reviewed above. This observation is interesting in itself, because it is often instructive to see how a given result can be generated in different ways. The development is also potentially useful, because as a general framework for deriving ensemble averages, mapped averaging opens the door to the formulation of new expressions that can further improve the performance of density calculations.

The outline of this paper is as follows. The next section provides a brief overview of the mapped averaging framework. In the two sections following that, we apply mapped averaging to the formulation of singlet and pair densities, respectively. We finish with some concluding remarks.

2. Overview of mapped averaging

The mapped averaging framework is derived from the idea of targeted perturbation, which was suggested by Jarzynski as a means to improve the calculation of free energy differences [20]. Let $\mathbf{X}$ be a vector of all relevant coordinates describing a configuration (hence for N monatomic molecules in a 3-dimensional space, $\mathbf{X}$ comprises $3N$ coordinate values). Then a coordinate mapping $\mathbf{X} \to \mathbf{x}$ couples with the parameter perturbation $\lambda \to \lambda'$, such that the transformed coordinates $\mathbf{x}$ are more representative of those encountered in the λ' state. This increases the overlap of the sets of configurations relevant to the λ and λ' systems, which consequently enhances the precision of calculations. When applied for a differential perturbation [17], the approach yields expressions for free-energy derivatives, and thereby provides a route to derive new ensemble averages for thermodynamic properties [13].

Specifically, we are interested in the change in the unitless free energy βA: $\Delta(\beta A) \equiv (\beta A)(\lambda') - (\beta A)(\lambda)$. Jarzynski showed that this can be expressed as a targeted-perturbation ensemble average in the λ system [20]:

$$\Delta(\beta A) = -\ln \left\langle J e^{-\Delta(\beta U)} \right\rangle_{\lambda} \tag{4}$$

where $\Delta(\beta U) \equiv (\beta u)(\mathbf{x};\lambda') - (\beta u)(\mathbf{X};\lambda)$, $J \equiv |\partial\mathbf{x}/\partial\mathbf{X}|$ is the Jacobian of the mapping, and $(\beta u)(\mathbf{x};\lambda)$ is the unitless energy for configuration $\mathbf{x}$ and parameter value λ.

Mapped averaging is the application of targeted perturbation for the infinitesimal change $\lambda \to \lambda + d\lambda$, which yields the thermodynamic free-energy derivatives (with μ representing a second parameter along with λ in the second derivative):

$$(\beta A)_\lambda \equiv \frac{\partial(\beta A)}{\partial\lambda} = -\langle J_\lambda\rangle + \langle(\beta U)_\lambda\rangle \tag{5a}$$

$$(\beta A)_{\lambda\mu} \equiv \frac{\partial^2(\beta A)}{\partial\lambda\partial\mu} = -\langle J_{\lambda\mu} - J_\lambda J_\mu\rangle + \langle(\beta U)_{\lambda\mu}\rangle - \mathrm{Cov}\left[J_\lambda - (\beta U)_\lambda, J_\mu - (\beta U)_\mu\right] \tag{5b}$$

where $\mathrm{Cov}[\mathrm{Y},\mathrm{Z}] \equiv \langle\mathrm{YZ}\rangle - \langle\mathrm{Y}\rangle\langle\mathrm{Z}\rangle$. In this limit, the mapping is defined as a mapping velocity $\dot{\mathbf{x}}^\lambda$, such that $\mathbf{x} = \mathbf{X} + \dot{\mathbf{x}}^\lambda d\lambda$ (the λ superscript on $\dot{\mathbf{x}}$ indicates the variable 'driving' the mapping). A good mapping velocity is prescribed by conserving the normalised probabilities (p/q) of configurations in the perturbed and unperturbed states [13]:

$$\frac{\delta}{\delta\lambda}\left(\frac{p}{q}\right) + \nabla_\mathbf{x}\cdot\left(\frac{p}{q}\dot{\mathbf{x}}^\lambda\right) = 0 \tag{6}$$

where q is the normalisation constant for the unnormalised probability p. This equation cannot be solved in general for the exact p and q, so approximations are employed with the aim of generating a mapping velocity $\dot{\mathbf{x}}^\lambda$ that is effective if not perfect. This mapping velocity is used to calculate the derivatives of J and βU using [13]:

$$J_\lambda = \nabla_\mathbf{x}\cdot\dot{\mathbf{x}}^\lambda \tag{7a}$$

$$J_{\lambda\mu} - J_\lambda J_\mu = \nabla_\mathbf{x}\cdot\dot{\mathbf{x}}^\lambda_\mu + \dot{\mathbf{x}}^\mu\cdot\nabla_\mathbf{x}(\nabla_\mathbf{x}\cdot\dot{\mathbf{x}}^\lambda) \tag{7b}$$

$$(\beta U)_\lambda = (\beta u)_\lambda - \beta\mathbf{f}\cdot\dot{\mathbf{x}}^\lambda \tag{7c}$$

$$(\beta U)_{\lambda\mu} = (\beta u)_{\lambda\mu} - \left(\dot{\mathbf{x}}^\lambda_\mu + \dot{\mathbf{x}}^\mu\cdot\nabla_\mathbf{x}\dot{\mathbf{x}}^\lambda\right)\cdot\beta\mathbf{f} + \dot{\mathbf{x}}^\mu\cdot\beta\boldsymbol{\phi}\cdot\dot{\mathbf{x}}^\lambda - \left(\dot{\mathbf{x}}^\lambda\cdot(\beta\mathbf{f})_\mu + \dot{\mathbf{x}}^\mu\cdot(\beta\mathbf{f})_\lambda\right), \tag{7d}$$

where $\mathbf{f} \equiv -\nabla_\mathbf{x}u$ is the force vector and $\boldsymbol{\phi} \equiv \nabla_\mathbf{x}\nabla_\mathbf{x}u$ is the force-constant matrix (Hessian) for a given configuration. Equation (7c) shows how we use $(\beta U)_\lambda$ to represent the variation of the energy due to its direct dependence on λ (if any), plus the effect of the mapping (i.e. U is in a Lagrangian frame; u is Eulerian).

We may combine Equations (5a), (6), (7a), and (7c) to obtain:

$$(\beta A)_\lambda = -(\ln q)_\lambda + \left\langle(\ln p)_\lambda + (\beta u)_\lambda + \dot{\mathbf{x}}^\lambda\cdot\left(\nabla_\mathbf{x}\ln p - \beta\mathbf{f}\right)\right\rangle. \tag{8}$$

Whereas Equations (5) and (7) are true in general—for any choice of $\dot{\mathbf{x}}^\lambda$—Equation (8) is correct only if $\dot{\mathbf{x}}^\lambda$ is given according to Equation (6).

We consider the potential $u(\mathbf{x})$ as a sum of singlet (ϕ_1), pair (ϕ_2), and possibly multibody contributions:

$$u(\mathbf{x}) = \sum_i^N \phi_1(\mathbf{r}_i) + \sum_{i,j}^N \phi_2(\mathbf{r}_i,\mathbf{r}_j) + \cdots \tag{9}$$

We use $\mathbf{r}$ to represent a spatial-coordinate vector, so $\mathbf{r}_i$ represents the position of molecule i.

3. Singlet density

3.1. General equations

The singlet density $\rho^{(1)}(\mathbf{r})$ is given as the functional derivative of the grand potential βA with respect to the single-particle potential $\phi_1(\mathbf{r})$ [21]:

$$\rho^{(1)}(\mathbf{r}) = \frac{1}{\beta}\left(\frac{\delta\beta A}{\delta\phi_1(\mathbf{r})}\right)_{\beta,V,\mu} = \left\langle\sum_i^N \delta(\mathbf{r}-\mathbf{r}_i)\right\rangle \tag{10}$$

This expression as an average of Dirac delta-functions prescribes the use of histograms to evaluate $\rho^{(1)}(\mathbf{r})$, which is the conventional approach.

Given that $\rho^{(1)}$ can be expressed as the first derivative of the free energy, the mapped-averaging framework can be applied to develop alternative ensemble averages for it. Accordingly, in this section we develop a general mapped average for the singlet density function.

Let $p(\mathbf{r}_i;\phi_1)$ be the approximate unnormalised density function for particle i, which we assume is independent of the other positions and with a Boltzmann dependence on ϕ_1:

$$p(\mathbf{r}_i;\phi_1(\mathbf{r})) = p_0(\mathbf{r}_i)\exp(-\beta\phi_1(\mathbf{r}_i)) \tag{11a}$$

$$q(\phi_1(\mathbf{r})) = \int p(\tilde{\mathbf{r}};\phi_1(\mathbf{r}))\,\mathrm{d}\tilde{\mathbf{r}} \tag{11b}$$

where $p_0(\mathbf{r})$ is a ϕ_1-independent contribution to p that allows for additional flexibility in choosing its form. Then $p(\mathbf{x})$ is given as a product of $p(\mathbf{r}_i)$ over all $i = 1\ldots N$. In this case, the balance equation for the mapping may be decomposed, yielding the same equation for each coordinate's mapping velocity $\dot{\mathbf{r}}_i^{\phi_1(\mathbf{r})}(\mathbf{r}_i)$, thus [13]:

$$\nabla_{\mathbf{r}_i}\cdot\left(p(\mathbf{r}_i;\phi_1(\mathbf{r}))\dot{\mathbf{r}}_i^{\phi_1(\mathbf{r})}\right) = -q\frac{\delta}{\delta\phi_1(\mathbf{r})}\left(\frac{p(\mathbf{r}_i;\phi_1(\mathbf{r}))}{q}\right) = \beta p(\mathbf{r})\left(\delta(\mathbf{r}-\mathbf{r}_i) - \frac{p(\mathbf{r}_i)}{q}\right) \tag{12}$$

For a mapping given according to Equations (11) and (12), the singlet density can be given via Equation (8), summed over all molecules:

$$\rho^{(1)}(\mathbf{r}) = \frac{1}{\beta}(\beta A)_{\phi_1(\mathbf{r})} = \frac{N}{q}p(\mathbf{r};\phi_1(\mathbf{r})) + \frac{1}{\beta}\sum_i \left\langle \dot{\mathbf{r}}_i^{\phi_1(\mathbf{r})} \cdot \left[\nabla_{\mathbf{r}_i} \ln p_0(\mathbf{r}_i) - \beta(\nabla_{\mathbf{r}_i}\phi_1(\mathbf{r}_i) + \mathbf{f}_i)\right]\right\rangle. \quad (13)$$

Equation (13) exhibits the general structure typically seen in a mapped average. The first term on the right-hand side shows that $p(\mathbf{r})$ forms a baseline estimate of $\rho^{(1)}(\mathbf{r})$, and the ensemble average is the correction to this estimate. In the average, note that the force on atom i, $\mathbf{f}_i$, includes a contribution from the gradient of the singlet potential ϕ_1, and this is cancelled by the explicit addition of $\nabla_{\mathbf{r}_i}\phi_1$ seen here. The ensemble average will be small to the extent that other atoms do not contribute much to $\mathbf{f}_i$ (e.g. at low density), and to the extent that p_0 is weakly varying (which means that p is well represented by $\exp(-\beta\phi_1)$).

We note also that the terms $(\ln p)_{\phi_1}$ and $(\beta u)_{\phi_1}$ that appear in Equation (8) both give rise to delta functions, but these cancel, leaving no averages that need to be evaluated using histograms. Instead, one simply selects a point in space $\mathbf{r}$ (or set of points to get a density profile), which specifies $p(\mathbf{r};\phi_1(\mathbf{r}))$ and the mapping velocity, $\dot{\mathbf{r}}_i^{\phi_1(\mathbf{r})}$ (see Equation (16) below for an example), then the average specified by Equation (13) is recorded to determine the singlet density there. The average involves summing over all other molecules, regardless of their position (i.e. it is not restricted to atoms in a bin centred on $\mathbf{r}$), and evaluating the total force on them (and perhaps the gradient of p_0 at their position), and accumulating as specified in (13).

3.2. One-dimensional Cartesian variation

It is easier to proceed further if we have in mind a specific geometry. Hence we consider the case where the density is inhomogeneous in only one dimension (say z), and the system is periodic in all dimensions. Then derivatives with respect to x and y can be made zero, and Equation (12) becomes (after trivial integration over x and y):

$$\frac{\mathrm{d}}{\mathrm{d}z_i}\left(p(z_i)\dot{z}_i^{\phi_1(z)}\right) = \beta p(z)\left(\frac{\delta(z-z_i)}{\mathcal{A}} - \frac{p(z_i)}{q}\right) \quad (14)$$

where $\mathcal{A}$ is the cross-sectional area and now

$$q = \mathcal{A}\int_{-L/2}^{L/2} p(\tilde{z})\,\mathrm{d}\tilde{z} \quad (15)$$

such that the simulation box extends from $-L/2$ to $L/2$ in the z dimension. On integrating Equation (14) from z^+ to z^- (via the periodic boundary) and using the boundary condition $\dot{z}_i^{\phi_1(z)}(z^+) = -\dot{z}_i^{\phi_1(z)}(z^-)$:

$$\dot{z}_i^{\phi_1(z)}(z_i) = \frac{\beta}{\mathcal{A}}\frac{p(z)}{p(z_i)}\left(\frac{1}{2} - H(z-z_i) - \frac{c(z_i)-c(z)}{c(L/2)}\right), \quad (16)$$

which introduces the cumulative (unnormalized) probability distribution function:

$$c(z_i) = \int_{-L/2}^{z_i} p(\tilde{z})\,\mathrm{d}\tilde{z}. \quad (17)$$

Also for the 1-D case, Equation (13) is written:

$$\rho(z) = \frac{N}{q}p(z) - \frac{1}{\beta}\left\langle \sum_i^N \dot{z}_i^{\phi_1(z)}\left(\beta f_{z,i} - \frac{\mathrm{d}\ln p(z_i)}{\mathrm{d}z_i}\right)\right\rangle \quad (18)$$

where $f_{z,i}$ is the z-component of the force on molecule i.

3.3. Uniform p and force-sampling method

For the special case of uniform $p(z)$, i.e. p independent of z, $q = V$ and Equation (16) for the mapping velocity becomes:

$$\dot{z}_i^{\phi_1(z)}(z_i) = \frac{\beta}{\mathcal{A}}\left(\frac{1}{2} - H(z-z_i) - \frac{z_i - z}{L}\right) \quad (19)$$

Substituting Equation (19) in Equation (18):

$$\rho(z) = \frac{N}{V} - \left\langle \frac{1}{\mathcal{A}}\sum_i^N \left(\frac{1}{2} - H(z-z_i) - \frac{z_i - z}{L}\right)\beta f_{z,i}\right\rangle \quad (20)$$

Equation (20) is the reformulated ensemble average for singlet density obtained by uniform mapped averaging. Using this method, the variation in density between any two z coordinates (say z_2 and z_1 such that $z_2 > z_1$) is ($\Delta\rho = \rho(z_2) - \rho(z_1)$):

$$\Delta\rho = \frac{1}{\mathcal{A}}\sum_i^N \beta f_{z,i}\left(H(z_2 - z_i) - H(z_1 - z_i) + \frac{z_1 - z_2}{L}\right) = \frac{\beta}{\mathcal{A}}\left(\sum_{z_1<z_i<z_2} f_{z,i} + \frac{z_1 - z_2}{L}\sum_i^N f_{z,i}\right) \quad (21)$$

When $\sum_i^N f_{z,i} = 0$, the variation in density ($\Delta\rho$) given by (21) is the same as the difference in singlet densities between z_2 and z_1 coordinates using Equation (1). However, when the sum of forces on all atoms in the simulation box is not zero, force sampling gives slightly different densities at both ends of the simulation box [12]. In contrast, in the mapped averaging development a

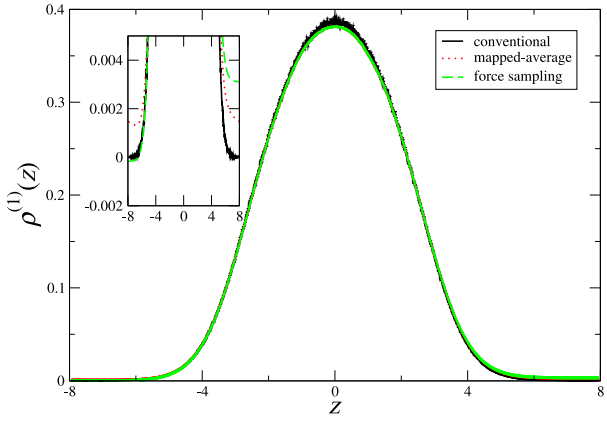


Figure 1. Monte Carlo simulation results of the singlet density profiles obtained using uniform mapped-averaging (Equation (20)), force sampling [12], and conventional histogram methods. Simulated system is a Lennard-Jones (LJ) fluid under the influence of parabolic external potential, $\phi_1 = \epsilon(z/\sigma)^2$, where σ and ϵ are the LJ size and energy parameters, respectively. Plotted quantities are in units such that $\sigma = 1$. Inset figure shows the same curves but with the ordinate scale greatly expanded to show the behaviour of the tails. Additional simulation details are provided in the Appendix.

linear term emerges (second term in Equation (21)) that has the effect of equalising the densities at the two ends of the box. This effect is demonstrated in Figure 1, which shows results from Monte Carlo simulations performed by us of density profiles for a Lennard-Jones fluid under the influence of a parabolic external potential. The results obtained using uniform mapped-averaging and force sampling are nearly indistinguishable from each other. However, near the ends of the simulation box where no particles are detected by the conventional approach, for sufficiently large systems the uniform mapped-averaging and force-sampling methods densities are statistically zero, but not identically so. The inset zooms in the density profiles near the ends of the box, showing an offset exhibited by the force-sampling method that is corrected by the mapped-averaging formula. The force-sampling results reported by de las Heras and Schmidt [12] exhibit the same offset behaviour, and also show that the effect is attenuated as more sampling is performed.

4. Pair density

4.1. General equations

For this development we assume a homogeneous system (in particular, $\phi_1 \equiv 0$), so that the pair density $\rho^{(2)}(\mathbf{r}_i, \mathbf{r}_j)$ is a function only of the separation $\mathbf{r}_j - \mathbf{r}_i$.

The pair distribution as a function of the intermolecular separation $\mathbf{r}$ can be given as a functional derivative with respect to the pair potential ϕ_2 [21], and this provides a route to a mapped-average formulation. We have then:

$$\rho^{(2)}(\mathbf{r}) = \frac{2}{V\beta}\left(\frac{\delta\beta A}{\delta\phi_2(\mathbf{r})}\right)_{\beta,V,N} \tag{22a}$$

$$\equiv \frac{2}{V\beta}\sum_{i=1}^{N}\sum_{j<i}\left(\frac{\delta\beta A}{\delta\phi_2^{(ij)}(\mathbf{r})}\right)_{\beta,V,N}$$

$$= \frac{2}{V}\sum_{i=1}^{N}\sum_{j<i}\left\langle\frac{\delta(\mathbf{r}-\mathbf{r}_{ij})}{4\pi r^2}\right\rangle \tag{22b}$$

where $r \equiv |\mathbf{r}|$; the division by V results from functional differentiation with respect to $\phi_2(\mathbf{r})$ rather than $\phi_2(\mathbf{r}_1, \mathbf{r}_2)$ [21]. Each atom i appears in a pair with every other atom j, and when we go to develop a mapping for i we do not want to consider all these interactions at once, which would be needed for a mapping based on (22a). Accordingly, in (22b) we introduce $\phi_2^{(ij)}$ as the pair potential for the ij pair, treating it as if it can be changed independently of all the other pair interactions. This is just a formal device for the development, as in the end we take $\phi_2^{(ij)} \equiv \phi_2$ for all i,j. Our aim then is to develop a mapped average for the functional derivative in (22b). In the presentation to follow, for clarity of notation we generally omit the (ij) superscript on ϕ_2, except in a few instances where the distinction needs to be emphasised.

As with the singlet case, let us define $p(\mathbf{r}_{ij})$ to depend on the pair potential but with an additional term that gives more flexibility to its form, thus:

$$p(\mathbf{r}_{ij}; \phi_2(\mathbf{r})) = p_0(\mathbf{r}_{ij})\exp(-\beta\phi_2(\mathbf{r}_{ij})) \tag{23a}$$

$$q(\phi_2(\mathbf{r})) = \int p(\tilde{\mathbf{r}}_{ij}; \phi_2(\mathbf{r}))d\tilde{\mathbf{r}}_{ij} \tag{23b}$$

Typically the system volume is much larger than the pair-interaction volume, and $q \approx V$. Similar to Equation (12), the mapping velocity $\dot{\mathbf{r}}_{ij}^{\phi_2(\mathbf{r})}$ is given by:

$$\nabla_{\mathbf{r}_{ij}} \cdot \left(p(\mathbf{r}_{ij}; \phi_2(\mathbf{r}))\dot{\mathbf{r}}_{ij}^{\phi_2(\mathbf{r})}\right)$$

$$= \beta p(\mathbf{r})\left(\delta(\mathbf{r}-\mathbf{r}_{ij}) - \frac{p(\mathbf{r}_{ij})}{q}\right) \tag{24}$$

We obtain from Equation (24) a specification of the mapping for the pair separation, $\dot{\mathbf{r}}_{ij}^{\phi_2(\mathbf{r})}$, which is insufficient to specify the movements of the actual molecules i and j. A natural choice is to move the molecules while keeping the pair center-of-mass fixed, which for equal-mass molecules specifies:

$$\dot{\mathbf{r}}_j^{\phi_2(\mathbf{r})} = -\dot{\mathbf{r}}_i^{\phi_2(\mathbf{r})} = \tfrac{1}{2}\dot{\mathbf{r}}_{ij}^{\phi_2(\mathbf{r})} \tag{25}$$

From Equation (8) we have, for mapping of just i and j,

$$\left(\frac{\delta\beta A}{\delta\phi_2^{(ij)}(\mathbf{r})}\right) = -(\ln q)_{\phi_2^{(ij)}(\mathbf{r})} + \left\langle \dot{\mathbf{r}}_j^{\phi_2(\mathbf{r})} \cdot \left(\nabla_{\mathbf{r}_j} \ln p - \beta \mathbf{f}_j\right) + \dot{\mathbf{r}}_i^{\phi_2(\mathbf{r})} \cdot \left(\nabla_{\mathbf{r}_i} \ln p - \beta \mathbf{f}_i\right)\right\rangle = \frac{\beta p(\mathbf{r})}{q} + \left\langle \dot{\mathbf{r}}_{ij}^{\phi_2(\mathbf{r})} \cdot \left(\nabla_{\mathbf{r}_{ij}} \ln p(\mathbf{r}_{ij}) - \frac{\beta}{2}\left(\mathbf{f}_j - \mathbf{f}_i\right)\right)\right\rangle \tag{26}$$

where in the first line we have already imposed the cancellation of $(\ln p)_\lambda$ and $(\beta u)_\lambda$. As with the singlet case, the $\nabla_{\mathbf{r}_{ij}} \ln p(\mathbf{r}_{ij})$ term includes a contribution from $\phi_2(\mathbf{r}_{ij})$, which cancels the contributions to $\mathbf{f}_j$ and $\mathbf{f}_i$ due to their direct mutual interaction.

Employing this result in Equation (22b), we arrive at the mapped-average expression for the pair densities:

$$\rho^{(2)}(\mathbf{r}) = \frac{N(N-1)}{Vq} p(\mathbf{r}) + \frac{2}{V}\left\langle \sum_{i=1}^{N}\sum_{j<i} \dot{\mathbf{r}}_{ij}^{\phi_2(\mathbf{r})} \cdot \left(\frac{1}{\beta}\nabla_{\mathbf{r}_{ij}} \ln p(\mathbf{r}_{ij}) - \frac{1}{2}\left(\mathbf{f}_j - \mathbf{f}_i\right)\right)\right\rangle \tag{27}$$

4.2. Isotropic distribution

For the case of only radial dependence, we will have

$$\dot{\mathbf{r}}_{ij}^{\phi_2(\mathbf{r})} = \dot{r}_{ij}^{\phi_2(\mathbf{r})} \hat{\mathbf{r}}_{ij} \tag{28}$$

Then Equation (24) can be integrated over angular coordinates and simplified as:

$$\frac{1}{r_{ij}^2}\frac{d}{dr_{ij}}\left(r_{ij}^2 p \dot{r}_{ij}^{\phi_2(r)}\right) = \beta p(r)\left(\frac{\delta(r - r_{ij})}{4\pi r^2} - \frac{p(r_{ij})}{q}\right). \tag{29}$$

Using a boundary condition that $r_{ij}^2 p(r_{ij}) \dot{r}_{ij}^{\phi_2(r)} = 0$ at $r_{ij} = 0$, and integrating from 0 to r_{ij}:

$$r_{ij}^2 p(r_{ij}) \dot{r}_{ij}^{\phi_2(r)} = \beta p(r) \int_0^{r_{ij}} d\tilde{r}_{ij} \tilde{r}_{ij}^2 \left(\frac{\delta(r - \tilde{r}_{ij})}{4\pi r^2} - \frac{p(\tilde{r}_{ij})}{q}\right)$$

$$\dot{r}_{ij}^{\phi(r)}(r_{ij}) = \beta \frac{p(r)}{r_{ij}^2 p(r_{ij})}\left(\frac{H(r_{ij} - r)}{4\pi} - \frac{c_r(r_{ij})}{q}\right) \tag{30}$$

where the radial-cumulative probability function is

$$c_r(r_{ij}) = \int_0^{r_{ij}} \tilde{r}^2 p(\tilde{r})\, d\tilde{r}. \tag{31}$$

Equation (27) becomes:

$$\rho^{(2)}(r) = \frac{N(N-1)}{Vq} p(r) + \frac{2}{V}\left\langle \sum_{i=1}^{N}\sum_{j<i} \dot{r}_{ij}^{\phi_2(r)} \left(\frac{1}{\beta}\frac{\partial \ln p(r_{ij})}{\partial r_{ij}} - \frac{1}{2}\left(\mathbf{f}_j - \mathbf{f}_i\right) \cdot \hat{\mathbf{r}}_{ij}\right)\right\rangle. \tag{32}$$

4.3. Uniform p and force-sampling method

For the uniform-reference treatment, when p is approximated as independent of r (which requires $p_0(r) = \exp(+\beta\phi_2(r))$):

$$\dot{r}_{ij}^{\phi(r)}(r_{ij}) = \beta\left(\frac{H(r_{ij} - r)}{4\pi r_{ij}^2} - \frac{r_{ij}}{3V}\right) \tag{33}$$

Therefore, pair densities can be calculated by:

$$\rho^{(2)}(r) = \rho^2 - \frac{\rho}{V} - \frac{\beta}{V}\left\langle \sum_{i=1}^{N}\sum_{j<i}\left(\frac{H(r_{ij} - r)}{4\pi r_{ij}^2} - \frac{r_{ij}}{3V}\right) \times \left(\mathbf{f}_j - \mathbf{f}_i\right) \cdot \hat{\mathbf{r}}_{ij}\right\rangle \tag{34}$$

Equation (34) is the reformulated ensemble average for pair densities obtained by uniform-reference mapped averaging, and it is very similar to Equation (3) obtained via the force sampling approach. The difference involves only the terms ρ/V and $r_{ij}/3V$, which will vanish in the thermodynamic limit $V \to \infty$.

We note that (34) can be rearranged as follows:

$$\rho^{(2)}(r) = \rho^2 - \frac{\rho}{V} - \frac{\beta}{V}\left\langle \sum_{i=1}^{N} \mathbf{f}_i \cdot \sum_{j\neq i} \hat{\mathbf{r}}_{ij}\left(\frac{H(r_{ij} - r)}{4\pi r_{ij}^2} - \frac{r_{ij}}{3V}\right)\right\rangle \tag{35}$$

It is interesting to consider $\rho^{(2)}$ inside the core, wherein $H(r_{ij} - r) = 1$ for all pairs. In this region the pair density is zero, requiring that the second term in Equation (35) must average to ρ^2. It is not obvious that the formula satisfies this, but empirically we observe it to be true (within statistical uncertainty). We can go a step further and remove the uncertainty in the core by asserting that the

pair density is zero there, and evaluate the r-dependent pair density via the difference $\rho^{(2)}(r) - \rho^{(2)}(0)$, yielding

$$\rho^{(2)}(r) = \frac{\beta}{V}\left\langle \sum_{i=1}^{N} \mathbf{f}_i \cdot \sum_{j\neq i} \hat{\mathbf{r}}_{ij} \frac{1 - H(r_{ij} - r)}{4\pi r_{ij}^2} \right\rangle, \qquad (36)$$

which uses $\rho^{(2)}(0) = 0$. In contrast to Equation (35), which involves sums over pairs $r_{ij} > r$, Equation (36) requires sums over pairs that are separated by less than r. This formulation appears to be advantageous as it requires contributions from fewer pairs (assuming small r is of interest), and it eliminates potentially problematic issues with long-range contributions.

5. Conclusion

A uniform-reference mapped-averaging formulation for the singlet density and pair densities has been presented, and shown to be equivalent to previous formulas which were derived based on considerations involving the force density. We do in fact identify one small difference in the formulas for the singlet density, but it turns out that the mapped-average formula provides a correction needed when the sum of forces over all molecules is not zero (which may occur due to fluctuations, but will average to zero with enough sampling). Previous studies [11,12] have shown that the statistical noise is significantly smaller in the density profiles obtained via these formulas, compared to the conventional histogram based approach. The force-based and mapped-averaging formulas offer another significant advantage, in that their uncertainty is insensitive to the grid size, allowing for evaluation of the densities with arbitrary resolution. The efficiency of these formulations is of course good in general, but it should be particularly valuable in applications involving ab initio molecular dynamics, where it is difficult to perform the extensive sampling needed to generate density distributions to good precision.

Mapped averaging is a general framework, and it is possible to develop other formulas by making different choices for the reference distribution p. A reasonable alternative would set it equal the form given by using $p_0 = 1$ in Equations (11) or (23), which represents the low-density estimates of $\rho^{(1)}$ and $\rho^{(2)}$, respectively. Another choice is to use an estimate of the density distribution at the conditions of interest, generated for example by a short molecular simulation. It is quite possible that these choices do not yield improvements in performance, because it appears that coordinated multi-atom motions are needed to form a good mapping at high density. Nevertheless, it is a worthwhile direction to investigate, and it provides an avenue for generating yet other ideas.

Finally, we note that $\rho^{(2)}$ can be given by a second-order functional derivative with respect to ϕ_1, and this can be used to generate a completely different mapped average involving the Hessian matrix as well as the forces via Equations (5b) and (7).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the U.S. National Science Foundation, grants OAC-1739145 (Division of Advanced Cyberinfrastructure) and CHE-1464581 (Division of Chemistry).

ORCID

Andrew J. Schultz http://orcid.org/0000-0002-1180-0156
David A. Kofke http://orcid.org/0000-0002-2530-8816

References

[1] R. Evans, M.C. Stewart, and N.B. Wilding, J. Chem. Phys. **147**, 044701 (2017).
[2] A. Parmeggiani, T. Franosch, and E. Frey, Phys. Rev. Lett. **90**, 086601 (2003).
[3] A. Macioek, R. Evans and N.B. Wilding, J. Chem. Phys. **119**, 8663 (2003).
[4] T. Biben, R. Ohnesorge and H. Löwen, EPL **28**, 665 (1994).
[5] D.D. Do and H.D. Do, Langmuir **20**, 7103 (2004).
[6] Z.-Q. Hu, *Modern Inorganic Synthetic Chemistry* (Elsevier, Amsterdam, 2011), pp. 455–478.
[7] G. de Araujo, C. Benmore and S. Byrn, Sci. Rep. **7**, 46367 (2017).
[8] Z. Li, X. Bian, X. Yang, and G.E. Karniadakis, J. Chem. Phys. **145**, 044102 (2016).
[9] R.J. Reeder and F.M. Michel, *Research Methods in Biomineralization Science*, Academic Press, city, 2013), pp. 477–500.
[10] T.-Q. Yu, P.-Y. Chen, M. Chen, A. Samanta, E. Vanden-Eijnden, and M. Tuckerman, J. Chem. Phys. **140**, 214109 (2014).
[11] D. Borgis, R. Assaraf, B. Rotenberg and R. Vuilleumier, Mol. Phys. **111**, 3486 (2013).
[12] D. de las Heras and M. Schmidt, Phys. Rev. Lett. **120**, 218001 (2018).
[13] A.J. Schultz, S.G. Moustafa, W. Lin, S.J. Weinstein and D.A. Kofke, J. Chem. Theory Comput. **12**, 1491 (2016).
[14] S.G. Moustafa, A.J. Schultz and D.A. Kofke, J. Chem. Theory Comput. **13**, 825 (2017a).
[15] S.G. Moustafa, A.J. Schultz, and D.A. Kofke, J. Chem. Phys. **149**, 124109 (2018). doi:10.1063/1.5043614.
[16] S.G. Moustafa, A.J. Schultz, E. Zurek, and D.A. Kofke, Phys. Rev. B **96**, 014117 (2017b). doi:10.1103/PhysRevB.96.014117.
[17] S.G. Moustafa, A.J. Schultz, and D.A. Kofke, Phys. Rev. E **92**, 043303 (2015).
[18] A. Purohit, A.J. Schultz, S.G. Moustafa, J.R. Errington and D.A. Kofke, Mol. Phys. **116**, 3027 (2018). doi:10.1080/00268976.2018.1481542.

[19] A.J. Schultz and D.A. Kofke, J. Chem. Phys. **149**, 204508 (2018). doi:10.1063/1.5053714.
[20] C. Jarzynski, Phys. Rev. E **65**, 046122 (2002).
[21] J.-P. Hansen and I. McDonald, *Theory of Simple Liquids*, 3rd ed. (Academic Press, London, 2006).

Appendix. Details of calculations

We employed the Lennard-Jones (LJ) model defined by the pair potential: $U(r) = 4\epsilon((\sigma/r)^{12} - (\sigma/r)^6)$, where σ and ϵ are the LJ size and energy parameters, respectively, and r is the pair separation. The LJ potential was truncated at $r_c = 3\sigma$, and σ and ϵ/k_B parameters were set to unity (LJ units). Periodic boundary conditions were employed in all three orthogonal directions, using a cubic simulation volume of edge length $L = 15.874$. Simulation runs of 10^8 MC steps (10^7 steps of equilibration) were performed on simulation boxes containing 500 atoms at density $\rho = 0.125$ and temperature $T = 5$. Data were recorded using a bin width $\Delta z = L/1000$.