The Journal of Chemical Physics

RESEARCH ARTICLE | JUNE 09 2023

Getting over the hump with KAMEL-LOBE: Kernel-averaging method to eliminate length-of-bin effects in radial distribution functions

S. Arman Ghaffarizadeh ; Gerald J. Wang

Check for updates

J. Chem. Phys. 158, 224112 (2023)
https://doi.org/10.1063/5.0138068

View Online
Export Citation
CrossMark

Downloaded from http://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0138068/17987189/224112_1_5.0138068.pdf

AIP Publishing

The Journal of Chemical Physics
Special Topic: Adhesion and Friction
Submit Today!
AIP Publishing

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Getting over the hump with KAMEL-LOBE: Kernel-averaging method to eliminate length-of-bin effects in radial distribution functions

Cite as: J. Chem. Phys. 158, 224112 (2023); doi: 10.1063/5.0138068
Submitted: 8 December 2022 • Accepted: 16 March 2023 •
Published Online: 9 June 2023

View Online Export Citation CrossMark

S. Arman Ghaffarizadeh[1] and Gerald J. Wang[2,a)]

AFFILIATIONS

[1] Department of Mechanical Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213, USA
[2] Department of Civil and Environmental Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213, USA

[a)] **Author to whom correspondence should be addressed:** gjwang@cmu.edu

ABSTRACT

Radial distribution functions (RDFs) are widely used in molecular simulation and beyond. Most approaches to computing RDFs require assembling a histogram over inter-particle separation distances. In turn, these histograms require a specific (and generally arbitrary) choice of discretization for bins. We demonstrate that this arbitrary choice for binning can lead to significant and spurious phenomena in several commonplace molecular-simulation analyses that make use of RDFs, such as identifying phase boundaries and generating excess entropy scaling relationships. We show that a straightforward approach (which we term Kernel-Averaging Method to Eliminate Length-Of-Bin Effects) mitigates these issues. This approach is based on systematic and mass-conserving mollification of RDFs using a Gaussian kernel. This technique has several advantages compared to existing methods, including being useful for cases where the original particle kinematic data have not been retained, and the only available data are the RDFs themselves. We also discuss the optimal implementation of this approach in the context of several application areas.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0138068

I. INTRODUCTION

In any system consisting of discrete particles with known positions, there can be tremendous utility in knowing the probability distribution for inter-particle separation distances. This fact is well-known to practitioners of molecular simulation, who typically store this information within the radial distribution function (RDF),[1–3] but closely related ideas are frequently used in fields as far-ranging as granular and colloidal mechanics,[4] atmospheric science,[5] astronomy and cosmology,[6] animal behavior,[7] pedestrian mobility,[8,9] and urban systems modeling.[10]

A very typical approach to estimating this probability distribution from particle trajectory data involves *histogramming*. In particular, inter-particle separation distances are computed for all pairs of particles in the system, and these distances are then organized in a histogram ranging from a separation distance of zero to a maximum separation distance, r_{max}. In some cases, a normalization convention is subsequently applied (as is the case for the RDF, for which convention dictates that the RDF approaches unity at large separation distances). All such histograms necessarily require the introduction of an arbitrary length-scale, Δr, which is used to convert a continuous variable (the inter-particle separation distance) into a discrete variable amenable to binning. For the purposes of acronym facilitation, we refer to Δr, in this work, as the "length of bin" (this quantity also frequently goes by the name, "bin width"). It is natural to wonder whether this arbitrary choice of length-scale can introduce spurious length-of-bin effects. As we show throughout Sec. IV, there are indeed many molecular simulation analysis tools with a sensitive dependence on the length of bin.

Downloaded from http://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0138068/17987189/224112_1_5.0138068.pdf

Published under an exclusive license by AIP Publishing

In this manuscript, we address the following questions: "To what extent does the choice of length of bin affect several commonplace modeling and analysis procedures used in molecular simulation?" and "Given an existing RDF, constructed with some choice of Δr, how can we reduce or even eliminate these length-of-bin effects?" In a data-enthusiastic age, such questions take on outsize importance for the practice of molecular simulation as they impact optimal practices for on-the-fly storage and analysis of data, comparison against molecular simulation datasets generated in the past (often with the underlying kinematic data inaccessible or never stored in the first place), and post-processing and analysis of new datasets, including and especially for the purpose of using modern data-driven methods.

Several approaches have been proposed to mitigate spurious effects associated with arbitrary choices of the length of bin, which we broadly categorize into four (largely but not entirely disjoint) sets:

1. Averaging over multiple snapshots of the system: This is the most straightforward option available to a molecular simulation practitioner and is almost certainly the most widely used approach to generate smooth RDFs. This approach has two critical shortcomings, relative to the methods discussed below: (1) Despite well-characterized scalings for statistical error in particle-based simulations as a function of thermodynamic and/or hydrodynamic conditions,[11] it is essentially impossible, in practice, to predict *a priori* the number of snapshots needed to achieve an "acceptable" level of smoothness in the RDF (we leave the discussion of what "acceptable" means in the first place to Sec. IV). As a consequence, in practice, one must always collect a conservatively large number of snapshots in order to produce a smooth RDF via such averaging, a pronounced burden if each timestep of the simulation carries significant cost (as is the case, e.g., with *ab initio* molecular dynamics) or if the system features especially long time-scales for spatial decorrelation.[12] Moreover, this approach is a non-starter for historical datasets in the literature, for which it may be difficult (or impossible) to obtain additional snapshots of the system under study, due to a variety of practical constraints. (2) Averaging is altogether untenable for any application that requires *on-the-fly* estimates (however coarse) of the RDF, as might be the case for a molecular simulation integrated into a workflow with feedback control.
2. Augmenting the spatial configuration data: This family of approaches, unlike the one above, leverages additional data and identities from statistical physics—in particular, related to inter-particle forces[13–15]—as a strategy for reducing variance in the RDF. These techniques, including techniques developed explicitly within the framework of control variates,[16–18] are particularly promising in terms of efficient utilization of molecular simulation data. However, just as above, these techniques cannot readily be used with historical datasets if the forces were not originally retained; it is worth emphasizing that it was not broadly appreciated until recently that force data might be useful for improving estimates of the RDF. Moreover, although the framework of control variates is sufficiently general that this family of approaches might potentially be extensible beyond molecular simulation, at present, there are no extensions of this idea to athermal systems, for which there may be a complex relationship between inter-particle forces and the RDF not dictated by equilibrium statistical mechanics.
3. Fitting a smooth function to the data: This broad idea can take an enormous number of forms. In the simplest sense, this can entail fitting the RDF data with a (sufficiently high-order) polynomial (see, e.g., Ref. 3) or a composition of a large number of functions with physically or empirically motivated functional forms (see, e.g., Ref. 19 for an early example of this approach, Ref. 20 for an example using a generic set of orthogonal basis functions, or Refs. 21 and 22 for recent machine-learning-inspired implementations of this idea). This problem has also been tackled by fitting smooth functions to the empirically observed *cumulative* distribution function,[23,24] sidestepping issues related to sampling error sensitivity for the probability density itself. A core challenge with all of these approaches, as noted by Allen and Tildesley,[3] is that a suitable choice for the form of this function (or set of basis functions) must be chosen carefully. Without *a priori* knowledge of the general shape of the RDF for a particular material at a particular point on its phase diagram, it can be challenging to select a suitable form, and it may require a substantial number of terms to reasonably approximate the RDF.
4. Mollifying the RDF data itself: One can also make use of a variety of smoothing techniques, which can directly operate on the RDF data, to reduce length-of-bin effects. A (non-exhaustive) list of techniques includes nearest-neighbor averages, window averages, and Gaussian kernel smoothing; we refer the interested reader to a more complete accounting in Ref. 25. Such techniques avoid the shortcomings discussed above; in particular, they can be used to analyze any historical dataset that contains the RDF, can be used on the fly, can be readily extended to athermal systems, and do not require any prior assumptions about suitable basis functions. The technique described in this manuscript takes this approach and goes beyond any existing technique for three primary reasons: (1) this mollification technique explicitly accounts for the radial Jacobian factor in the RDF (as was done in Ref. 24, which describes a smooth-function-fitting approach); (2) the method is specifically calibrated and tested to reduce length-of-bin effects for *multiple* end-use applications involving the RDF instead of being optimal for a single application (as done, e.g., in Refs. 26 and 27); and (3) unlike the approach in Refs. 26 and 27, this method can function exclusively as a post-processing technique acting on the RDF and the RDF alone (it does not require access to the underlying particle positions, or the additional storage and computational costs associated with retaining and reprocessing the particle positions).

II. METHODOLOGY

A. Molecular-dynamics (MD) simulations

Throughout this work, we consider a bulk system of fluid atoms with mass m_{LJ}, number density ρ, and average temperature

Downloaded from http://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0138068/17987189/224112_1_5.0138068.pdf

Published under an exclusive license by AIP Publishing

T (situated within a cubic domain, with periodic boundary conditions applied in all three dimensions). Fluid atoms interact via the Lennard–Jones (LJ) potential,

$$u_{\mathrm{LJ}}(r) = 4\varepsilon_{\mathrm{LJ}}\left[\left(\frac{\sigma_{\mathrm{LJ}}}{r}\right)^{12} - \left(\frac{\sigma_{\mathrm{LJ}}}{r}\right)^{6}\right], \tag{1}$$

where $u_{\mathrm{LJ}}(r)$ is the interaction energy of two particles separated by a distance r, and $\varepsilon_{\mathrm{LJ}}$ and σ_{LJ} are energy- and length-scales for the LJ potential, respectively. In all discussions that follow, all quantities are non-dimensionalized against the length-scale σ_{LJ}, energy-scale $\varepsilon_{\mathrm{LJ}}$, mass-scale m_{LJ}, time-scale $\sqrt{m_{\mathrm{LJ}}\sigma_{\mathrm{LJ}}^2/\varepsilon_{\mathrm{LJ}}}$, density-scale $m_{\mathrm{LJ}}/\sigma_{\mathrm{LJ}}^3$, diffusivity-scale $\sigma_{\mathrm{LJ}}^2/\sqrt{m_{\mathrm{LJ}}\sigma_{\mathrm{LJ}}^2/\varepsilon_{\mathrm{LJ}}}$, entropy-scale k_{B}, and temperature-scale $\varepsilon_{\mathrm{LJ}}/k_{\mathrm{B}}$, where k_{B} is the Boltzmann constant.

Simulations are performed with the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS)[28] code using a timestep of 2×10^{-3}. All systems contain $N = 2000$ fluid atoms. Systems are initially run for a time of 200 in the NVT ensemble using the Langevin thermostat[29] to reach the desired temperature T; the desired temperatures studied herein span $0.9 \le T \le 3$. To vary the density ρ, the side length of the simulation box is varied within $13 \le L \le 15$, corresponding to densities in the range $0.59 \le \rho \le 0.91$. A total of 70 combinations of T and ρ are studied. All of the observables discussed below are subsequently sampled over a data collection period of 200 in the NVE ensemble.

- Radial distribution function: We compute the RDF for our (isotropic) systems as $g(r) = \left\langle \frac{N(r+\Delta r/2) - N(r-\Delta r/2)}{4\pi r^2 \rho \Delta r} \right\rangle$, where the angle brackets indicate an average computed over all particles serving as the reference particle, and $N(r + \Delta r/2) - N(r - \Delta r/2)$ is the number of particles whose spatial locations are between $r - \Delta r/2$ and $r + \Delta r/2$ removed from the reference particle, and ρ is the fluid density. For each simulation, several lengths of bin Δr were considered.
- Coefficient of self-diffusion: We measure the mean-squared displacement as $\langle \|\mathbf{r}(t) - \mathbf{r}(0)\|^2 \rangle$, where $\mathbf{r}$ is the position vector of a reference particle, t denotes the length of the measurement window, and, again, the angle brackets indicate an average computed over all particles serving as the reference particle. Using the Einstein relation, the self-diffusion coefficient (in three spatial dimensions) is estimated as $D = \frac{\langle \|\mathbf{r}^2(t)\| \rangle}{6t}$, which we obtain via least-squares regression to a (zero-intercept) line.

III. KAMEL-LOBE

In this section, we describe the Kernel-Averaging Method to Eliminate Length-Of-Bin Effects (KAMEL-LOBE) scheme [Fig. 1(a)]. The goal of this technique is to standardize RDFs such that:

1. The resulting RDFs are useful for scientific applications that make use of the RDF (specifically in the sense that these applications no longer exhibit a dependence on otherwise arbitrary choices for the RDF length of bin Δr).
2. We make no assumptions (implicitly or explicitly) about the general shape of the RDF (e.g., the sizes or locations of its peaks), eliminating the need for *a priori* knowledge of suitable basis functions to represent the RDF.
3. This standardization can be carried out in a computationally efficient manner *either* on the fly during a molecular simulation *or* as a post-processing step, including on already-computed RDFs (for which the underlying kinematic data are not available).

A. The method itself

In what follows, we assume that we are in possession of an RDF: In particular, we have a vector $\mathbf{g}$ supplying n values of the RDF at the radial coordinates $r \in \{\frac{1}{2}\Delta r, \frac{3}{2}\Delta r, \ldots, r_{\max}\}$, where (as a reminder) Δr is the length of each bin, and $r_{\max} \equiv (n - \frac{1}{2})\Delta r$ is the maximum radius at which the RDF is evaluated. The core idea of KAMEL-LOBE is to convert $\mathbf{g}$ into a smoothed version $\tilde{\mathbf{g}}$ through the composition of three linear operations, $\tilde{\mathbf{g}} = \mathbb{T}_3\mathbb{T}_2\mathbb{T}_1\mathbf{g}$, given as follows:

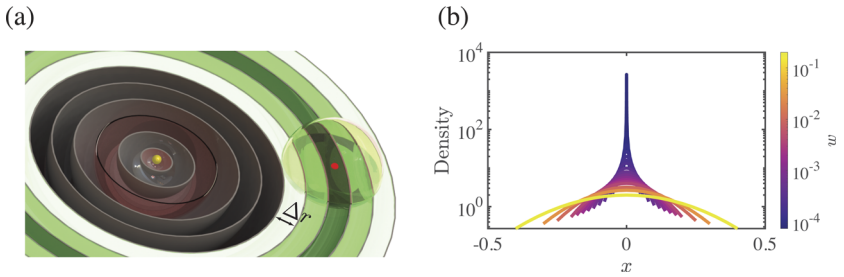

FIG. 1. (a) Schematic representation of the KAMEL-LOBE scheme, showing radial bins of constant width Δr around a reference particle (yellow). The particle number contribution made by the other particle (red) is the largest (dark green) for the bin within which its center is located, but non-zero for nearby bins (lighter shades of green). (b) Density for the Gaussian kernel used in KAMEL-LOBE as a function of x, the distance from the bin within which a particle's center is located. The color indicates the width of the Gaussian kernel w, with larger values (yellow) corresponding to spreading across a broader spatial extent.

Downloaded from http://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0138068/17987189/224112_1_5.0138068.pdf

Published under an exclusive license by AIP Publishing

1. Integration to obtain the cumulative radial distribution function: We compute the total number of particles within radius r as $N(r) = \int_0^r 4\pi r'^2 \rho g(r') dr'$. We emphasize again that our analysis is carried out under the assumption of an isotropic system. We can numerically approximate $N(r)$ via trapezoidal-rule integration as $\mathbf{N} = \mathbb{T}_1 \mathbf{g}$, where $\mathbb{T}_1$ is given by

$$\mathbb{T}_1 = 2\pi\rho\Delta r \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 2 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 1 & 2 & \ddots & 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & (\Delta r)^2 & 0 & \cdots & 0 \\ 0 & 0 & (2\Delta r)^2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \ddots & 0 & ((n-1)\Delta r)^2 \end{bmatrix}. \tag{2}$$

It is worth noting that this step is conceptually identical to the first step in the construction by van Zon and Schofield,[24] which follows in the footsteps of work by Berg and Harris.[23]

2. Application of a Gaussian kernel to mollify the cumulative radial distribution function: Rather than assuming each particle to be concentrated at a point (i.e., a Dirac delta "distribution"), we spread each particle's contribution to $N(r)$ over a range, following a(n approximately) Gaussian distribution centered at the particle's location. The degree of smoothing is set by the standard deviation (characteristic width) of this Gaussian kernel [Fig. 1(b)], which we denote as w (so as to avoid conflation with the Lennard–Jones length-scale σ_{LJ}). Although a Gaussian distribution has infinite support, we spread each particle over a finite range of $\sim\pm 2w$ from the location of its center, with no spreading beyond this range (i.e., no mass is spread outside of a width of $\sim 4w$). In what follows, we investigate the range $0 \le w \le 0.2$; the lower bound corresponds to no smoothing (i.e., the way that RDFs are typically presented, for which the Gaussian kernel takes on the limiting shape of a Dirac delta "distribution").
   To accomplish this smoothing for a particle with its center at a distance r from the reference particle, we first compute the number of bins k over which smoothing will be performed as $k = 2\lceil 2w/\Delta r \rceil - 1$, where $\lceil x \rceil$ is the ceiling function. We then compute a vector of Gaussian weights $\mathbf{u} = (u_1, u_2, \ldots, u_k)$. The weight u_j is calculated as the area within the jth bin of a Gaussian probability density function centered at the location of the particle of interest; the jth bin has a width of Δr and is centered at $(j-m)\Delta r$, where $m = \lceil 2w/\Delta r \rceil$. In other words, $u_j = \Phi((j-m+1/2)\Delta r) - \Phi((j-m-1/2)\Delta r)$, where Φ denotes the Gaussian cumulative distribution function. Smoothing is performed only for separation distances exceeding $2w$; the first m entries of $\mathbf{N}$ are left unchanged (this prevents us from computing $\mathbf{u}$ in a bin where $\mathbf{g}$ is not defined in the first place, i.e., where $r < 0$). Overall, we transform the cumulative radial distribution function $\mathbf{N}$ to its smoothed version $\tilde{\mathbf{N}}$ as $\mathbb{T}_2 \mathbf{N}$, where $\mathbb{T}_2$ is partitioned as

$$\mathbb{T}_2 = \frac{1}{\sum_{j=1}^k u_j} \begin{bmatrix} \mathbb{A} \\ \mathbb{B} \\ \mathbb{C} \end{bmatrix}, \tag{3}$$

and the two blocks $\mathbb{A}$ $(m \times n)$, $\mathbb{B}$ $((n-2m) \times n)$, and $\mathbb{C}$ $(m \times n)$ are

$$\mathbb{A} = [\mathbb{I}_m | 0_{n-m}], \tag{4}$$

$$\mathbb{B} = \begin{bmatrix} u_1 & u_2 & \cdots & u_k & 0 & 0 & \cdots & 0 \\ 0 & u_1 & u_2 & \cdots & u_k & \ddots & \ddots & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \ddots & 0 & u_1 & u_2 & \ddots & u_k \end{bmatrix}, \tag{5}$$

$$\mathbb{C} = [0_{n-m} | \mathbb{I}_m]. \tag{6}$$

Here, $\mathbb{I}_m$ denotes an $m \times m$ identity matrix, and 0_{n-m} denotes an $m \times (n-m)$ matrix of zeroes. The normalization $\frac{1}{\sum_{j=1}^k u_j}$ is included in $\mathbb{T}_2$ to conserve particle mass after smoothing; for reasonable choices of Δr, this normalizing factor should be ~1.05, reflecting the fact that ~95% of the mass of a Gaussian is contained within two standard deviations of its mean.

3. Differentiation to obtain a mollified radial distribution function: We compute the smooth $g(r)$ profile denoted by $\tilde{g}(r)$ as $\tilde{g}(r) = \frac{1}{4\pi r^2 \rho} \frac{d}{dr} \tilde{N}(r)$. We numerically evaluate this expression with centered finite differences, $\tilde{\mathbf{g}} = \mathbb{T}_3 \tilde{\mathbf{N}}$, where $\mathbb{T}_3$ is given by

Downloaded from http://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0138068/17987189/224112_1_5.0138068.pdf

Published under an exclusive license by AIP Publishing

$$\mathbb{T}_3 = \left(\frac{1}{4\pi\rho}\right)\left(\frac{2}{\Delta r}\right)\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{(\Delta r)^2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{(2\Delta r)^2} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \ddots & 0 & \frac{1}{((n-1)\Delta r)^2} \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -2 & 1 & 0 & \ddots & \vdots \\ 2 & -2 & 1 & \ddots & 0 \\ \ddots & \ddots & \ddots & \ddots & 0 \\ \ddots & \ddots & 2 & -2 & 1 \end{bmatrix}. \tag{7}$$

In particular, $\mathbb{T}_3$ is lower triangular, with an alternating pattern of 2's and −2's below the diagonal. Each $i,j \in \{1,2,\ldots,n\}$ element of $\mathbb{T}_3$ can be written as

$$[\mathbb{T}_3]_{ij} = \begin{cases} 0 & i < j \text{ or } i = j = 1 \\ \left(\frac{1}{4\pi\rho}\right)\left(\frac{2}{\Delta r}\right)\left(\frac{1}{(i-1)\Delta r}\right)^2 & i = j (\neq 1) \\ 2\left(\frac{1}{4\pi\rho}\right)\left(\frac{2}{\Delta r}\right)\left(\frac{1}{(i-1)\Delta r}\right)^2 (-1)^{i-j} & i > j \end{cases} \tag{8}$$

B. Statistical mechanics underlying an upper bound for *w*

Thus far, we have not discussed how to select w, the standard deviation of the Gaussian kernel. Purely from the perspective of statistics, it is clear that progressively larger choices of w will progressively reduce spurious high-frequency variations in the RDF associated with low statistics in each bin. As such, a more interesting question is: What is the *largest* reasonable choice of w? Before exploring this question empirically through extensive numerical simulation and analysis in Sec. IV, we first provide some rationalization for an upper bound on w grounded in the statistical mechanics of a Lennard–Jones material.

Because the minimum of the LJ potential (for two bare LJ atoms) is located at a separation distance of $2^{1/6}$, at zero temperature, one should not expect to find two atoms with any smaller separation distance than this (in a condensed phase due to long-range attractive interactions, the minimum is located at a separation distance slightly less than $2^{1/6}$ but not by more than a few percent for realistic densities). At finite temperature T (which, in our non-dimensional units, is also to say finite thermal energy T), atoms may access smaller separation distances than that corresponding to mechanical equilibrium but cannot come arbitrarily close due to short-range repulsion. We can obtain the scaling for the smallest thermodynamically plausible separation distance $r_{\min}$ as a function of temperature by comparing the interaction energy against the thermal energy,

$$4(\xi^2 - \xi) \sim T, \tag{9}$$

where $\xi \equiv r_{\min}^{-6}$. This implies

$$\xi \sim \frac{1+\sqrt{1+T}}{2}, \tag{10}$$

which, in turn, suggests

$$r_{\min} \sim \left(\frac{2}{1+\sqrt{1+T}}\right)^{1/6}. \tag{11}$$

We require that w respects $r_{\min}$; in other words, the choice of w must not cause a significant amount of particle weight to be shifted to distances less than $r_{\min}$. Since the majority of particle weight appears in the vicinity of a separation distance of $2^{1/6}$ on the unmollified RDF, this requirement is equivalent to requiring

$$2w < 2^{1/6} - \left(\frac{2}{1+\sqrt{1+T}}\right)^{1/6} \tag{12}$$

since the furthest distance that KAMEL-LOBE causes any particle weight to move is $2w$.

For T on the order of unity (characteristic of a LJ fluid, at least when the density is unity or less), (12) implies $w \lesssim 0.07$; the lower the temperature, the lower the maximum value of w that respects $r_{\min}$. Thus, balancing both statistics and statistical mechanics considerations, we expect that a suitable choice of w for fluid systems is on the order of 10^{-2} for realistic temperatures [even when $T = 10$, (12) demands that $w < 0.12$].

Before proceeding, two remarks are in order:

1. A strength of KAMEL-LOBE over existing techniques that employ Gaussian smoothing[26,27] is that KAMEL-LOBE truncates (and re-normalizes) its Gaussian, as opposed to using a full Gaussian with infinite support. This approach ensures that $r_{\min}$ can be better respected through suitably small choice of w.
2. For systems with $T \ll 1$, it is clear that only w's that are vanishingly small satisfy (12). As such, we caution that KAMEL-LOBE (or any other mollification technique) should be used with care for the analysis of RDFs obtained from simulations at very low temperatures.

Downloaded from http://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0138068/17987189/224112_1_5.0138068.pdf

Published under an exclusive license by AIP Publishing

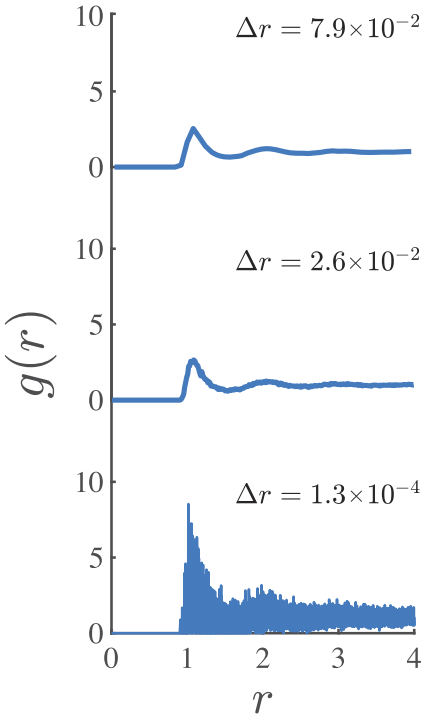

FIG. 2. Instantaneous RDFs computed for a Lennard–Jones fluid (ρ = 0.81, T = 1.1) with three different Δr values: $\Delta r = 7.9 \times 10^{-2}$, $\Delta r = 2.6 \times 10^{-2}$, and $\Delta r = 1.3 \times 10^{-4}$.

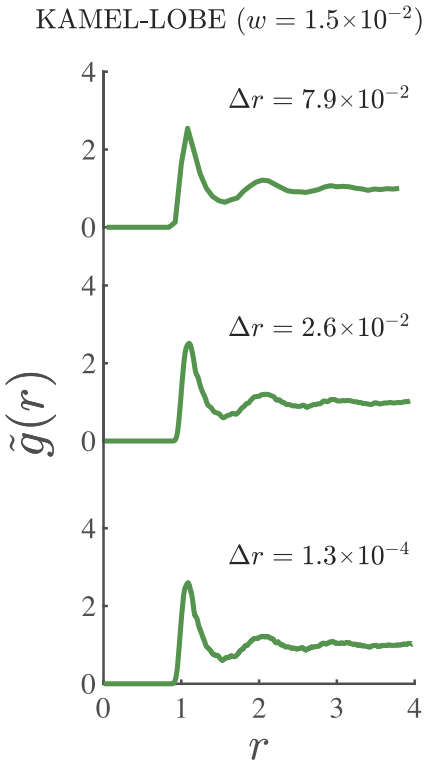

FIG. 3. Instantaneous RDFs, processed using KAMEL-LOBE ($w = 1.5 \times 10^{-2}$), for a Lennard–Jones fluid (ρ = 0.81, T = 1.1) with three different Δr values: $\Delta r = 7.9 \times 10^{-2}$, $\Delta r = 2.6 \times 10^{-2}$, and $\Delta r = 1.3 \times 10^{-4}$ (same conditions as Fig. 2).

IV. RESULTS AND DISCUSSION

A. The radial distribution function itself

We begin with the very simple question: "How does KAMEL-LOBE affect the RDF itself?" To calibrate expectations, we first report the effect of Δr on the RDF for systems of different temperatures and densities using the standard approach for computing $g(r)$. The maximum radius is set to $r_{\max} = 4.0$, and the number of bins is varied such that $1.3 \times 10^{-4} \leq \Delta r \leq 7.9 \times 10^{-2}$.

In the following, when $g(r)$ is computed using information from a single snapshot of a system at a fixed moment in time, we refer to it as an "instantaneous profile." Figure 2 illustrates instantaneous profiles for a system with $\rho = 0.81$ and $T = 1.1$. Qualitatively, we observe three regimes:

1. For the coarsest choices of Δr, $g(r)$ does not appear smooth for the reason that Δr is not much less than the intrinsic wavelength of the RDF. We expect *a priori* that this wavelength is of order unity since solvation shells in a dense fluid should be separated by roughly one molecular diameter.
2. For intermediate choices of Δr, $g(r)$ appears smooth; we do not comment at this junction on what exactly constitutes an "intermediate choice" of Δr, and we will show below that an exact definition is, in fact, immaterial for the applications of interest.
3. For the finest choices of Δr, $g(r)$ exhibits significant amounts of noise, a consequence of each individual bin housing a relatively small number of particles. We also note that since the traditional approach to computing the RDF assigns the entirety of a particle to the bin in which its center of mass falls, the finer Δr is, the higher the maximum value taken on by $g(r)$ within an instantaneous profile.

By visual inspection, these same RDFs are all rendered considerably smoother through the use of KAMEL-LOBE (Fig. 3). To check that the added smoothness is statistically meaningful, in Fig. 4, we show comparisons between RDFs obtained using KAMEL-LOBE and RDFs produced from averaging in time. We readily observe that the differences between KAMEL-LOBE RDFs and time-averaged RDFs are small. In fact, point-wise differences between KAMEL-LOBE and long-time-averaged results [shown in Fig. 4 as $g_{10\,000}(r)$] are 5% or less, suggesting that KAMEL-LOBE can generate reasonable estimates for RDFs in settings where long-time averages are impractical or infeasible to collect. Such estimates may be especially useful for molecular simulations that make use of computationally expensive interatomic interactions (as is the case for, e.g., *ab initio* molecular dynamics).

Downloaded from http://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0138068/17987189/224112_1_5.0138068.pdf

Published under an exclusive license by AIP Publishing

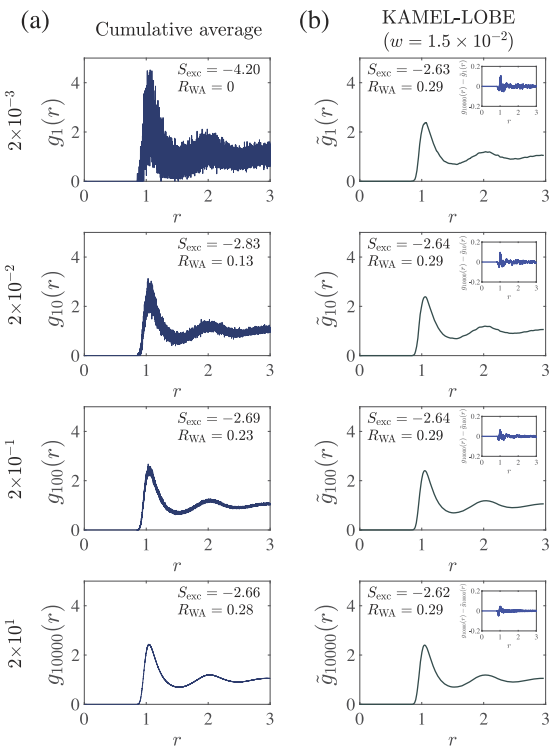

FIG. 4. For a system with $\rho = 0.86$ and $T = 2$, a comparison between (a) cumulative averages of the instantaneous RDF profiles, computed using $\Delta r = 1 \times 10^{-4}$ and a timestep of 2×10^{-3} and (b) that same cumulative RDF profile after use of KAMEL-LOBE with $w = 1.5 \times 10^{-2}$. The total amount of time used for each RDF is shown on the far left; the four rows thus make use of 10^0, 10^1, 10^2, and 10^4 instantaneous RDFs, respectively. The values of R_{WA} (Sec. IV B) and (per-particle) S_{exc} (Sec. IV C) are shown for each RDF. Each inset figure in (b) shows the difference between each RDF obtained from KAMEL-LOBE and the RDF obtained from long-time averaging, $g_{10\,000}(r)$.

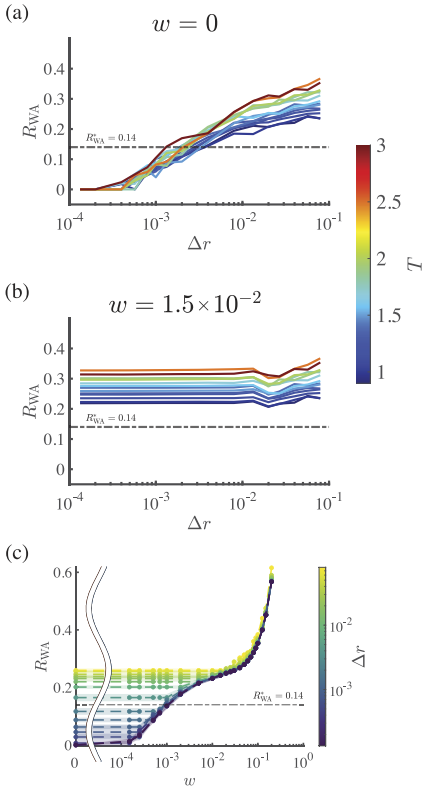

FIG. 5. The Wendt–Abraham parameter R_{WA} as a function of Δr for systems with $\rho = 0.81$ and temperatures in the range $0.9 \le T \le 3$, computed using (a) unmodified RDFs and (b) RDFs post-processed using KAMEL-LOBE. (c) R_{WA} as a function of the Gaussian kernel width w for a system with $\rho = 0.81$ and $T = 1.1$. Symbol colors indicate the choice of Δr; shading indicates one standard deviation of R_{WA} over all systems simulated. The left-most data ($w = 0$) correspond to no use of KAMEL-LOBE and are separated from the other data by a broken horizontal axis. In all panels, the threshold that is typically used to delineate the liquid phase from the amorphous solid phase, $R^*_{WA} = 0.14$, is indicated by a black dashed line.

B. Application of KAMEL-LOBE to simple analyses that use the radial distribution function

These observations have straightforward yet noteworthy consequences for several commonplace heuristics and analysis procedures used for interpreting molecular simulation results, of which we highlight two here:

1. Delineation of phases: Numerous techniques[30–35] make use of multiple points sampled from the RDF as a criterion to define the phase boundaries for a material. For example, the Wendt–Abraham parameter R_{WA} is defined as the ratio between the first minimum of the RDF (beyond the first maximum) and the first maximum of the RDF and is used to identify the boundary between the amorphous solid and liquid phases, with the value $R^*_{WA} = 0.14$ frequently used as the threshold between these phases.[30] It is a direct consequence of our earlier observations that the choice of Δr can affect R_{WA}. In Figs. 5(a) and 5(b), we show this dependence over the full range of temperatures studied for a system with $\rho = 0.81$, for which we do not expect the presence of a solid phase.[31,36]

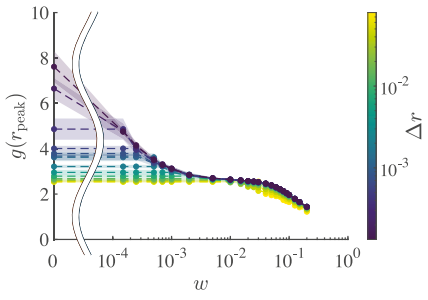

FIG. 6. The value of the first RDF peak $g(r_{peak})$ as a function of the Gaussian kernel width w for a system with $\rho = 0.81$ and $T = 1.1$. Symbol colors indicate the choice of Δr; shading indicates one standard deviation of $g(r_{peak})$ over all systems simulated. The left-most points ($w = 0$) correspond to no use of KAMEL-LOBE and are separated from the other data by a broken horizontal axis.

Downloaded from http://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0138068/17987189/224112_1_5.0138068.pdf

Published under an exclusive license by AIP Publishing

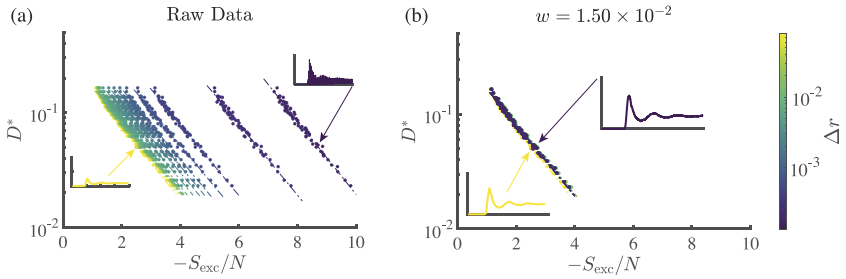

FIG. 7. For the same set of 70 systems of LJ fluid, scaled diffusivity D^* as a function of (negative per-particle) excess entropy using a variety of choices of Δr ($1.3 \times 10^{-4} \leq \Delta r \leq 7.9 \times 10^{-2}$, indicated by color). Thin dashed lines indicate the best fit of form Eq. (15) to systems analyzed using constant Δr. Radial distribution functions for the two extremal choices of Δr are shown as insets. (a) $w = 0$ (unmodified RDFs) and (b) $w = 1.5 \times 10^{-2}$.

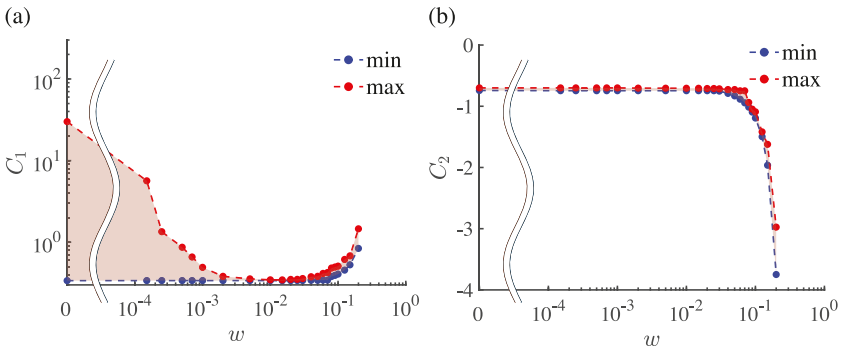

FIG. 8. Minimum (blue) and maximum (red) values of the excess entropy scaling relationship fitting parameters (a) C_1 and (b) C_2, as a function of the Gaussian kernel width w. For both figures, the range of lengths of bin used is $1.3 \times 10^{-4} \leq \Delta r \leq 7.9 \times 10^{-2}$. The leftmost points ($w = 0$) correspond to no use of KAMEL-LOBE and are separated from the other data by a broken horizontal axis.

In other words, we would expect that $R_{WA} > R^*_{WA}$ strictly. Unlike the unmodified RDFs (i.e., $w = 0$), which do not match this expectation—in fact, they lead to values of R_{WA} that cross the R^*_{WA} threshold at *temperature-dependent values* of Δr [Fig. 5(a)]—the RDFs processed using KAMEL-LOBE lead to R_{WA} values that are strictly above R^*_{WA} for all systems studied [Fig. 5(b)]. What's more, except for the coarsest values of Δr ($\Delta r \gtrsim 10^{-2}$), all R_{WA} curves show essentially no dependence on Δr. In Fig. 5(c), we show that for $w \gtrsim 2 \times 10^{-3}$, all choices of Δr lead to $R_{WA} > R^*_{WA}$. Moreover, for $w \gtrsim 10^{-2}$, R_{WA} depends weakly on the choice of Δr. Since it is undesirable to smooth to the extent that $g(r)$ is featureless (corresponding to the $R_{WA} \to 1$ limit), we conclude that $w \approx \mathcal{O}(10^{-2})$ is suitable for this application, a result that supports the analysis in Sec. III B. In other words, to return to the question posed in Sec. I, such a choice of w generates mollified RDFs that are, for this application, "acceptably" smooth.

2. Kinetic-theory-based analyses: Several approaches to modeling fluid transport properties grounded in kinetic theory[37,38] rely heavily upon the RDF. For example, within Enskog theory[38] or Prigogine–Nicolis–Misguich theory,[39] the values of various fluid transport coefficients (e.g., viscosity and thermal conductivity) can be expressed in terms of the RDF evaluated at the separation distance corresponding to inter-particle contact, typically taken as the first maximum of the RDF, $g(r_{peak})$. Naturally, any scheme for computing the RDF that leads to a Δr-dependent RDF will affect the specific value of $g(r_{peak})$. In Fig. 6, we show that $g(r_{peak})$ is only weakly dependent on Δr when KAMEL-LOBE is employed, with w selected to be greater than 2×10^{-3}. As in the previous example, we find that $w \approx \mathcal{O}(10^{-2})$ is suitable for this application, yet again in support of the analysis in Sec. III B.

We note that these are only two of many application areas where a consistent protocol for generating RDFs is important; several others are discussed by Torquato.[40]

C. Case study on the use of KAMEL-LOBE for excess entropy scaling relationships

In this section, we describe in detail the beneficial effects that KAMEL-LOBE has on a particular application that makes use of radial distribution functions, namely, the construction of excess entropy scaling relations. We provide a brief overview of excess entropy scaling that is sufficient to follow the remainder of this section but refer the interested reader to the comprehensive review of Dyre.[41] In brief, the excess entropy S_{exc} is defined as the difference between a system's entropy at a given density and temperature and the ideal gas entropy under the same conditions, $S_{ideal}(\rho, T)$,

$$S_{exc}(\rho, T) \equiv S(\rho, T) - S_{ideal}(\rho, T). \quad (13)$$

Downloaded from http://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0138068/17987189/224112_1_5.0138068.pdf

Published under an exclusive license by AIP Publishing

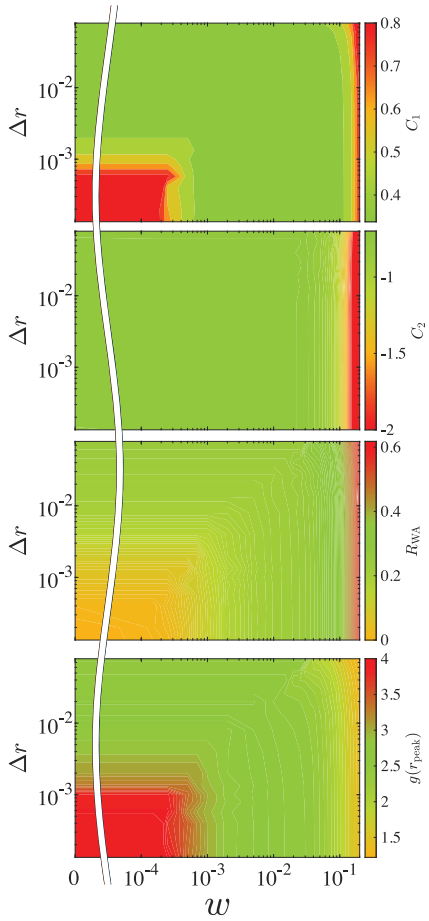

FIG. 9. Contour plots for four quantities of interest considered herein [C_1, C_2, R_{WA}, and $g(r_{peak})$] as a function of Δr and w. C_1 and C_2 are computed using the same 70 systems of LJ fluid as shown in Fig. 7; R_{WA} and $g(r_{peak})$ are shown for a system with $\rho = 0.81$ and $T = 1.1$.

By construction, this is a negative quantity. In all of the results that follow, we invoke the two-particle-correlation approximation, which has been shown to be valid for a broad range of fluids.[41] This approximation supplies an estimate for the excess entropy using the RDF[42–45] (hence the connection with this work),

$$S_{exc}/\mathrm{N} \approx -2\pi\rho \int_0^\infty (g(r)\ln g(r) - g(r) + 1)r^2 dr. \quad (14)$$

The essence of an excess entropy scaling relationship is that this quantity can be directly connected to transport coefficients. In particular, Rosenfeld[46] demonstrated the existence of a relationship between the excess entropy and (an appropriately scaled version of the) diffusion coefficient for a system of LJ particles, given by

$$D^* = C_1 \exp\left(-C_2 \frac{S_{exc}}{N}\right), \quad (15)$$

where $D^* \equiv \left(\rho^{1/3}\sqrt{m/k_B T}\right)D$, D is the fluid's coefficient of self-diffusion, and C_1 and C_2 are material-specific constants. The applicability of this scaling for diffusivity has been validated in a wide range of fluids beyond LJ fluids, including hard spheres;[47] supercooled, glassy, and/or binary mixtures;[48–51] ionic melts;[52] hydrocarbons;[53] a variety of coarse-grained fluids;[54] and active fluids.[55]

Given our earlier findings, it is natural to ask the question: "Do arbitrary binning choices also affect excess entropy scaling relations?" In what follows, we answer this question in the affirmative (and to a significant extent) and demonstrate that KAMEL-LOBE can substantially improve the consistency of excess entropy scaling relations obtained using RDFs within the two-particle-correlation approximation.

Based upon the MD dataset generated for LJ fluids ($0.59 \le \rho \le 0.91$, $0.9 \le T \le 3$), Fig. 7(a) demonstrates the broad range of excess entropy scaling relations that can be obtained simply by varying Δr. (It is worth noting that the choice of Δr does *not* affect the measured self-diffusion coefficient, and so this choice has no effect on D^*.) Although an excess entropy scaling relation manifests for each choice of Δr [i.e., an expression of the form given in Eq. (15) accurately describes the data], it is evident that there are substantial length-of-bin effects, namely, C_1 and C_2 both show a significant dependence on Δr. As such, these constants do not depend only on the material in question; they also depend upon the approach to binning when constructing the RDF. Of particular interest, even when the RDF is "quite smooth" to the eye, as is the case for all choices of Δr in the vicinity of 10^{-2}, C_1 and C_2 both still exhibit dependence on Δr. We also observe that the highest and lowest computed magnitudes for the excess entropy correspond to the finest and coarsest bin widths, respectively, which we can rationalize as a natural consequence of Jensen's inequality and the (negative) entropy being a convex function. The use of KAMEL-LOBE [Fig. 7(b)] collapses all of the data to a single curve, which is well described using a single value of C_1 and C_2.

This observation has major implications for the use of excess entropy scaling relationships: Dating back to Rosenfeld's initial work (which reported C_1 as 0.6 and C_2 as −0.8 for LJ fluids), values of C_1 and C_2 are often treated as quasi-universal for the fluid in question.[42] From a practical perspective, this observation has led to major efforts to compute these constants for a broad range of fluids of industrial interest.[56,57] Our work suggests that the generalizability of such efforts would benefit significantly from the use of the KAMEL-LOBE scheme since direct use of Eq. (14) on raw RDF data can lead to a wide range of values for the fitting parameters in excess entropy scaling relationships; as such, it is not possible to directly compare C_1 and C_2 values unless one also knows the corresponding choice of Δr used in the analysis. It is worth mentioning that this issue could be avoided altogether if one computed the exact excess entropy (see, e.g., approaches described in Refs. 58–60); nevertheless, since the use of the two-particle-correlation approximation is widespread, there is value in techniques that enhance the consistency and reproducibility of excess entropy results within the framework of this approximation.

To underscore the value of KAMEL-LOBE, in Fig. 8, we show the minimum and maximum values of C_1 and C_2 obtained as a function of w using the same range of choices for Δr. With no use of KAMEL-LOBE, it is possible for length-of-bin effects to give rise to values of C_1 that vary over nearly three decades. A desirable choice of w is one for which the distance between the maximum

Downloaded from http://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0138068/17987189/224112_1_5.0138068.pdf

Published under an exclusive license by AIP Publishing

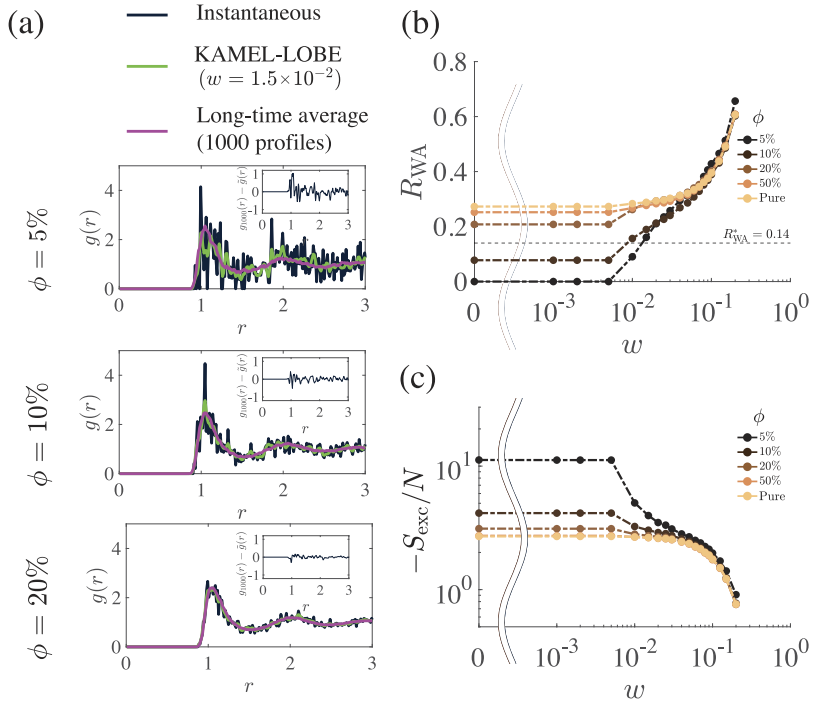

FIG. 10. (a) For three values of the solute fraction ϕ, comparison between instantaneous RDFs using $\Delta r = 1.3 \times 10^{-2}$, those same RDFs processed with KAMEL-LOBE ($w = 1.5 \times 10^{-2}$), and the long-time-averaged RDF computed over 1000 profiles. Each inset figure in (a) shows the difference between each RDF obtained from KAMEL-LOBE and the RDF obtained from long-time averaging, $g_{1000}(r)$. (b) For five values of the solute fraction ϕ, R_{WA} as a function of w. (c) For five values of the solute fraction, $-S_{exc}/N$ as a function of w.

and minimum (i.e., the magnitude of length-of-bin effects) is as small as possible, for both C_1 and C_2; this is true for $2 \times 10^{-3} \lesssim w \lesssim 2 \times 10^{-2}$. Reassuringly, the upper end of this range is on the same order of magnitude as the choice made in Refs. 26 and 27, which studied elemental sodium and aluminum, and made choices for w (which these references term as the "broadening parameter") that are $\sim 3 \times 10^{-2}$ in non-dimensional units. As with the examples in Sec. IV B, we find that $w \approx \mathcal{O}(10^{-2})$ is suitable for the application of excess entropy scaling, yet again supporting the analysis presented in Sec. III B.

The broad suitability of $w \approx \mathcal{O}(10^{-2})$ is underscored in Fig. 9, which reveals the presence of a vertical band of w values centered around $w \approx \mathcal{O}(10^{-2})$ for which all quantities of interest discussed herein show little to no dependence on the choice of Δr. In contradistinction, the far left side of Fig. 9 highlights the magnitude of length-of-bin effects that may arise in the absence of KAMEL-LOBE. Figure 9 promisingly suggests that the length-scale w may not require extensive calibration for each end-use application and that in fact a single choice may be suitable for most applications.

D. Case study on the use of KAMEL-LOBE for solvation structure

In this final section, we turn our attention to the use of RDFs to study solvation structure, inspired by work on similar problems in Refs. 17 and 61. We study a simple problem in this setting, which highlights the power of KAMEL-LOBE to overcome low statistics. For MD simulations with 8000 LJ particles ($\rho = 0.86$, $T = 2$), we designate a fraction ϕ of the particles (at random) to be "solute" particles. We now ask the question: "Using KAMEL-LOBE, how well can we reconstruct the results obtained using all particles and long-time averaging, equipped only with RDFs computed from the subset of solute particles?" Since all inter-particle interactions remain the same as before, we expect that any RDF obtained from solute–solute interactions should not differ (at least in the sense of a long-time average) from its counterpart obtained using all particles. In Fig. 10, for $\phi = 50\%$, we observe close agreement between results obtained using the solute–solute RDF and the result obtained using long-time averaging over all particles (i.e., $\phi = 100\%$). As expected, the quality of RDFs (and quantities of interest computed using these RDFs) deteriorates as ϕ (and the total number of solute–solute pairs) decreases; nevertheless, using KAMEL-LOBE, there is still reasonably strong point-wise agreement for the RDF for ϕ as low as 20%. Moreover, for $\phi \geq 20\%$, there is less than 10% variation in R_{WA} and S_{exc}/N when $w \gtrsim 10^{-2}$, yet again suggesting the broad suitability of $w \approx \mathcal{O}(10^{-2})$.

V. CONCLUSION

Histogram-based approaches are widely used for the computation of radial distribution functions (RDFs) and related quantities in molecular simulation and beyond. Such approaches necessarily make use of an artificial length-scale, namely, the size of each histogram bin (or length of bin). Here, we have demonstrated that numerous applications that make use of the RDF are plagued with length-of-bin effects. We introduce the Kernel-Averaging Method to Eliminate Length-Of-Bin Effects (KAMEL-LOBE), which systematically mollifies RDFs with a Gaussian kernel. As compared to existing

Downloaded from http://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0138068/17987189/224112_1_5.0138068.pdf

Published under an exclusive license by AIP Publishing

techniques, this approach has the benefit of simultaneously satisfying all of the following: KAMEL-LOBE (1) can be used for historical RDF datasets, for which the underlying kinematic data are inaccessible; (2) is able to perform on-the-fly smoothing for instantaneous RDFs that are collected during a simulation, generating RDFs that compare favorably with those obtained from averaging in time; (3) is agnostic to whether or not a system is thermal; (4) does not require *a priori* knowledge of suitable basis functions; and (5) explicitly accounts for the radial Jacobian factor in the computation of the RDF. We demonstrate that the mollified RDFs produced using KAMEL-LOBE substantially improve the consistency of numerous molecular simulation analyses that make use of RDFs, including the identification of phase boundaries, the calculation of quantities relevant to kinetic theory, and the generation of excess entropy scaling relationships. These strengths persist even when considering RDFs computed over a subset of all particles in the system. We have further shown that a non-dimensionalized Gaussian kernel width on the order of 10^{-2} (in particular, 1.5×10^{-2}) is broadly suitable for all of the applications studied herein.

Based upon these results, there is strong reason to believe that there are also significant length-of-bin effects in many data-driven models that incorporate RDFs as a feature (indeed, all of the applications presented in this work can be viewed as especially physically transparent data-driven models). As such, there is a compelling case to use KAMEL-LOBE as a standard pre-processing step in molecular simulation analyses that make use of an RDF. In an increasingly data-driven landscape, including the widespread use of machine-learning techniques, we believe that such standardized pre-processing steps are critical for ensuring the robustness and comparability of molecular simulation results as well as models trained using these results.

We close by noting several natural extensions for the ideas discussed in this manuscript: The Lennard–Jones interaction potential used throughout this work is steeply repulsive at inter-particle separation distances below unity; it would be interesting to assess the utility of KAMEL-LOBE for much softer and/or bounded potentials, which permit a greater degree of particle–particle overlap. For such potentials, it is well-known that the general framework of excess entropy scaling applies but requires modification in the scaling of the transport quantity.[62] It would also be interesting to investigate the extent to which the strategy described here may also be useful for histogram-based calculations beyond the radial distribution function (as one of many examples, inhomogeneous fluid density profiles in the vicinity of a solid interface[63–65]). All of these calculations share, in common, the problem that the full contribution of any given particle is made to the bin within which that particle's center resides, which amplifies the importance of the (entirely artificial) length of the bin, thereby generating spurious effects. It is reasonable to expect that judicious averaging with respect to a sensible choice of the kernel may reduce or eliminate such spurious effects.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the use of computing resources funded, in part, by the Carnegie Mellon University (CMU) College of Engineering and startup support from the CMU Department of Civil and Environmental Engineering, in addition to fellowship support from the Natural Sciences and Engineering Research Council of Canada and support from the National Science Foundation under Award Nos. 2021019 and 2133568. The authors also gratefully acknowledge excellent feedback and insightful suggestions from the two anonymous peer reviewers.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

S. Arman Ghaffarizadeh: Data curation (lead); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Gerald J. Wang**: Conceptualization (lead); Data curation (supporting); Formal analysis (equal); Funding acquisition (lead); Investigation (equal); Methodology (equal); Project administration (lead); Resources (lead); Software (equal); Supervision (lead); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

An implementation of and tutorial notebook for KAMEL-LOBE is available at https://github.com/M5-Lab/KAMEL-LOBE. The molecular simulation data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] M. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation* (Oxford University Press, 2010).
[2] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic Press, 2002).
[3] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, 2017).
[4] J. N. Israelachvili, *Intermolecular and Surface Forces*, 2nd ed. (Academic Press London, San Diego, 1991).
[5] M. L. Larsen, R. A. Shaw, A. B. Kostinski, and S. Glienke, Phys. Rev. Lett. **121**, 204501 (2018).
[6] P. J. E. Peebles, *The Large-Scale Structure of the Universe* (Princeton University Press, 1980).
[7] J. G. Puckett, R. Ni, and N. T. Ouellette, Phys. Rev. Lett. **114**, 258103 (2015).
[8] J. Cristín, V. Méndez, and D. Campos, Sci. Rep. **9**, 18488 (2019).
[9] K. B. Kramer and G. J. Wang, Phys. Fluids **33**, 103318 (2021).
[10] R. Rusali and G. J. Wang, in *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, BuildSys '20* (Association for Computing Machinery, New York, 2020), pp. 298–301.
[11] N. G. Hadjiconstantinou, A. L. Garcia, M. Z. Bazant, and G. He, J. Comput. Phys. **187**, 274 (2003).
[12] Y. Li and G. J. Wang, J. Chem. Phys. **156**, 114113 (2022).
[13] D. de las Heras and M. Schmidt, Phys. Rev. Lett. **120**, 218001 (2018).
[14] A. Purohit, A. J. Schultz, and D. A. Kofke, Mol. Phys. **117**, 2822 (2019).

Downloaded from http://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0138068/17987189/224112_1_5.0138068.pdf

Published under an exclusive license by AIP Publishing

[15]B. Rotenberg, J. Chem. Phys. **153**, 150902 (2020).
[16]D. Borgis, R. Assaraf, B. Rotenberg, and R. Vuilleumier, Mol. Phys. **111**, 3486 (2013).
[17]S. W. Coles, D. Borgis, R. Vuilleumier, and B. Rotenberg, J. Chem. Phys. **151**, 064124 (2019).
[18]S. W. Coles, E. Mangaud, D. Frenkel, and B. Rotenberg, J. Chem. Phys. **154**, 191101 (2021).
[19]E. Matteoli and G. A. Mansoori, J. Chem. Phys. **103**, 4672 (1995).
[20]P. N. Patrone and T. W. Rosch, J. Chem. Phys. **146**, 094107 (2017).
[21]A. Moradzadeh and N. R. Aluru, J. Phys. Chem. Lett. **10**, 7568 (2019).
[22]G. T. Craven, N. Lubbers, K. Barros, and S. Tretiak, J. Chem. Phys. **153**, 104502 (2020).
[23]B. A. Berg and R. C. Harris, Comput. Phys. Commun. **179**, 443 (2008).
[24]R. van Zon and J. Schofield, J. Chem. Phys. **132**, 154110 (2010).
[25]J. S. Simonoff, *Smoothing Methods in Statistics* (Springer, New York, 1996).
[26]P. M. Piaggi, O. Valsson, and M. Parrinello, Phys. Rev. Lett. **119**, 015701 (2017).
[27]P. M. Piaggi and M. Parrinello, J. Chem. Phys. **147**, 114112 (2017).
[28]A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, Comput. Phys. Commun. **271**, 108171 (2022).
[29]T. Schneider and E. Stoll, Phys. Rev. B **17**, 1302 (1978).
[30]H. R. Wendt and F. F. Abraham, Phys. Rev. Lett. **41**, 1244 (1978).
[31]J.-P. Hansen and L. Verlet, Phys. Rev. **184**, 151 (1969).
[32]H. J. Raveché, R. D. Mountain, and W. B. Streett, J. Chem. Phys. **61**, 1970 (1974).
[33]T. M. Truskett, S. Torquato, S. Sastry, P. G. Debenedetti, and F. H. Stillinger, Phys. Rev. E **58**, 3083 (1998).
[34]L. Costigliola, T. B. Schrøder, and J. C. Dyre, Phys. Chem. Chem. Phys. **18**, 14678 (2016).
[35]M. I. Ojovan and D. V. Louzguine-Luzgin, J. Phys. Chem. B **124**, 3186 (2020).
[36]H. Watanabe, N. Ito, and C.-K. Hu, J. Chem. Phys. **136**, 204102 (2012).
[37]H. van Beijeren and M. H. Ernst, J. Stat. Phys. **21**, 125 (1979).
[38]J. R. Dorfman, H. van Beijeren, and T. R. Kirkpatrick, *Contemporary Kinetic Theory of Matter* (Cambridge University Press, 2021).
[39]I. Prigogine, G. Nicolis, and J. Misguich, J. Chem. Phys. **43**, 4516 (1965).
[40]S. Torquato, Phys. Rev. E **51**, 3170 (1995).
[41]J. C. Dyre, J. Chem. Phys. **149**, 210901 (2018).
[42]R. E. Nettleton and M. S. Green, J. Chem. Phys. **29**, 1365 (1958).
[43]H. J. Raveché, J. Chem. Phys. **55**, 2242 (1971).
[44]A. Baranyai and D. J. Evans, Phys. Rev. A **40**, 3817 (1989).
[45]B. B. Laird and A. D. J. Haymet, Phys. Rev. A **45**, 5680 (1992).
[46]Y. Rosenfeld, Phys. Rev. A **15**, 2545 (1977).
[47]J. C. Dyre, J. Phys.: Condens. Matter **28**, 323001 (2016).
[48]C. Sachin and A. Joy, Physica A **588**, 126578 (2022).
[49]I. H. Bell, J. C. Dyre, and T. S. Ingebrigtsen, Nat. Commun. **11**, 4300 (2020).
[50]T. S. Ingebrigtsen, J. R. Errington, T. M. Truskett, and J. C. Dyre, Phys. Rev. Lett. **111**, 235901 (2013).
[51]W. P. Krekelberg, M. J. Pond, G. Goel, V. K. Shen, J. R. Errington, and T. M. Truskett, Phys. Rev. E **80**, 061205 (2009).
[52]M. Agarwal, M. Singh, B. Shadrack Jabes, and C. Chakravarty, J. Chem. Phys. **134**, 014502 (2011).
[53]R. Chopra, T. M. Truskett, and J. R. Errington, J. Phys. Chem. B **114**, 16487 (2010).
[54]J. Jin, K. S. Schweizer, and G. A. Voth, J. Chem. Phys. **158**, 034103 (2023).
[55]S. A. Ghaffarizadeh and G. J. Wang, J. Phys. Chem. Lett. **13**, 4949 (2022).
[56]M. Hopp and J. Gross, Ind. Eng. Chem. Res. **56**, 4527 (2017).
[57]W. A. Fouad, J. Chem. Eng. Data **65**, 5688 (2020).
[58]I. Saika-Voivod, P. H. Poole, and F. Sciortino, Nature **412**, 514 (2001).
[59]I. Saika-Voivod, F. Sciortino, and P. H. Poole, Phys. Rev. E **69**, 041503 (2004).
[60]J. R. Errington, T. M. Truskett, and J. Mittal, J. Chem. Phys. **125**, 244502 (2006).
[61]C. N. Nguyen, T. K. Young, and M. K. Gilson, J. Chem. Phys. **137**, 044101 (2012).
[62]W. P. Krekelberg, T. Kumar, J. Mittal, J. R. Errington, and T. M. Truskett, Phys. Rev. E **79**, 031203 (2009).
[63]F. F. Abraham, J. Chem. Phys. **68**, 3713 (1978).
[64]G. J. Wang and N. G. Hadjiconstantinou, Phys. Fluids **27**, 052006 (2015).
[65]G. J. Wang and N. G. Hadjiconstantinou, Phys. Rev. Fluids **2**, 094201 (2017).

Downloaded from http://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0138068/17987189/224112_1_5.0138068.pdf

Published under an exclusive license by AIP Publishing