Structural Properties from Molecular Dynamics Simulation

David Keffer

Department of Materials Science & Engineering University of Tennessee, Knoxville date begun: April 7, 2016 date last updated: April 11, 2016

Table of Contents

I. Purpose of Document	2
II. Radial Distribution Functions	2
II.A. What is a radial distribution function?	2
II.A.1. Connection with Theory	4
II.A.2. Connection with Experiment	5
II.A.3. Connection with Simulation	6
II.B. Decompositions of the RDF	6
II.B.1. Decomposition by Species	6
II.B.2. Decomposition by Intermolecular & Intramolecular Components	7
II.B.3. Others Decompositions	8
II.C. What do RDFs look like	9
II.C.1. Gas Phase	9
II.C.2. Liquid Phase	10
II.C.3. Crystalline Solid Phase	10
II.D. Relation to thermodynamic Properties	11
II.E. Some Examples	12
II.E.1. Adsorption in an Amorphous Nanoporous Material	12
II.E.2. Flow-Induced Crystallization of Polymers	13
II.E.3. Carbon Composites	14
II.F. In LAMMPS	14
III. End-to-End Distribution Functions	14
IV. Density Distributions	16

I. Purpose of Document

The purpose of this document is to provide a practical introduction to the evaluation of structural properties from molecular dynamics simulation. This document currently over three distribution functions: the radial distribution function, the end-to-end distribution function of chains, and density distributions.

II. Radial Distribution Functions

The radial distribution function is a structural property of interest to experimentalists, theoreticians and simulators alike. It is probably the more important metric of local atomic structure.

II.A. What is a radial distribution function?

The radial distribution function (RDF) is an unnormalized probability distribution that a particle can be found at a separation, r, from another particle. RDFs exist for both equilibrium and non-equilibrium systems.

Several example RDFs are shown in Figure 1. They are taken from the following reference:

Gao, C.Y., Nicholson, D.M., Keffer, D.J., Edwards, B.J., "A multiscale modeling demonstration based on the pair correlation function", J. Non-Newtonian Fluid Mech. 152(1-3) 2008 pp. 140-147, doi: 10.1016/j.jnnfm.2007.05.003.



Figure 1. RDFs for a Lennard-Jones fluid in liquid and gas states.

The RDF is unnormalized because it is defined such that it goes to unity as the separation goes to infinity. (A properly normalized probability distribution is defined to integrate to unity.)

The RDF is a simplification of other properties in at least three respects. First, the RDF describes only the probability distribution between two pairs of atoms. (To be sure, the RDF exists regardless of whether an interaction potential is pairwise.) Thus we can consider the RDF a two-body correlation. Others have formulated expressions for arbitrary n-body correlations. The two-body correlation is often called the pair correlation function (PCF).

Second, the RDF has been reduced from six degrees of freedom to 1. In three-dimensional space, there is an x, y and z coordinate for both particles 1 and 2. Thus we should have six degrees of freedom. However, if we make the assumption that the fluid is homogeneous, then the position of particle 1 doesn't matter, only the position of particle 2 relative to particle 1. This allows us to place particle 1 at the origin, reducing our six degrees of freedom to three. In the presence of an inhomogeneous field or structure (like an interface), this assumption would be invalidated. If we further consider that the position of particle 2 relative to particle 1 can be expressed in polar coordinates, the RDF averages over the two angular coordinates, leaving only 1 degree of freedom, the radial separation between particles, *r*, as shown in Figure 1. When the PCF is reduced to a single degree of freedom, we rename it the RDF. Thus the RDF is a subset of PCFs. Certainly, where angular dependence is important, we can use a PCF with more degrees of freedom to describe the structure.

Third, the RDF as we have written it above is not a function of time. However, we are familiar with autocorrelation functions and we can regard the RDF as the value of a correlation function when the elapsed time is zero. This correlation function is called the van Hove correlation function. [PH YSI CAL REVIEW VOLUME 95, NUMBER 1 JULY 1, 1954 Correlations in Space and Time and Born Approximation Scattering in Systems of Interacting Particles LEON VAN HOVE] This correlation function, G(r,t) gives the probability of finding a particle at a separation, r, at time, t, given an initial RDF of G(r,t=0) = g(r). The RDF shown in Figure 1 is not a correlation function. An example of the correlation function, also known as the dynamic pair density function is shown below. This figure is taken from the following reference.

Walter Kob and Hans C. Andersen, "Testing mode-coupling theory for a supercooled binary Lennard-Jones mixture I: The van Hove correlation function", Phys. Rev. E 51, 4626 – Published 1 May 1995



FIG. 5. $4\pi r^2 G_s(r,t)$ for the A particles versus r for times $t \approx 0.32 \times 2^n$ with $n = 0, 1, 2, \ldots$ (a) T = 2.0, (b) T = 0.6, and (c) T = 0.466.

FIG. 10. $G_d(r,t)/\rho$ for the AA correlation function. (a) $T = 2.0, \ 0 \le t \le 400$; (b) $T = 0.6, 0 \le t \le 3000$; (c) $T = 0.466, \ 0 \le t \le 100\ 000$.

Figure 2. The van Hove correlation function split into self (s) and distinct (d) components for a Lennard-Jones fluid. Only the lowest temperature is shown.

II.A.1. Connection with Theory

The RDF distribution function can be defined a variety of ways. In Chapter 13 of McQuarrie [Statistical Mechanics, Donald A. McQuarrie, University Science books, Sausalito, CA, 1976, second Edition, 2000], the *n*-body correlation function in a system of *N* particles is defined as the ratio of two integrals. The numerator, an integral over all but *n* particles, represents the probability of finding a state with n particles at positions $\mathbf{r}_1, \dots, \mathbf{r}_n$. The denominator is the same integral over all *N* particles (the configuration integral) and serves as a normalization constant.

$$g^{(n)}(\mathbf{r}_{1},...,\mathbf{r}_{n}) = \frac{V^{n}N!}{N^{n}(N-n)!} \frac{\int \cdots \int \exp\left(-\frac{U_{N}}{k_{B}T}\right) d\mathbf{r}_{n+1} \dots d\mathbf{r}_{N}}{\int \cdots \int \exp\left(-\frac{U_{N}}{k_{B}T}\right) d\mathbf{r}_{1} \dots d\mathbf{r}_{N}}$$
(1)

The pair correlation function (PCF) results when n=2, and is denoted $g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = g(\mathbf{r}_1, \mathbf{r}_2)$. When the pair correlation function is evaluated in a homogeneous system and averaged over the angular degrees of freedom, it is called the radial distribution function (RDF) and is denoted $g(r) = g(|\mathbf{r}_2 - \mathbf{r}_1|)$. The correlation functions are dimensionless.

One useful relation of the RDF is given below.

$$4\pi\rho\int_{0}^{\infty}g(r)r^{2}dr = N - 1 \approx N$$
⁽²⁾

The radial integral provides the number of neighbors around the central particle. If the limits of integration are fixed to other values, then this integral provides the number of particles, N_{nbr} , in the spherical shell from r_{min} to r_{max} .

$$4\pi\rho \int_{r_{\min}}^{r_{\max}} g(r)r^2 dr = N_{nbr}$$
(3)

One can also consider the radial distribution function as a normalized density distribution function,

$$g(r) = \frac{\rho(r)}{\rho} \tag{4}$$

where $\rho(r)$ is the radial density profile (with units of density) and ρ is the average density.

II.A.2. Connection with Experiment

The radial distribution function is also very important from an experimental point of view. xray and neutron scattering experiments measure the atomic-level of structure in materials. These signatures are obtained in reciprocal space and are often denoted S(Q). A Fourier transform can be used to convert S(Q) to the radial distribution function.

$$\rho g(r) = \frac{1}{2\pi^2} \int_{0}^{\infty} dQ Q^2 \frac{\sin(Qr)}{Qr} (S(Q) - 1)$$
(5)

Alternatively, some scientists prefer to work in reciprocal space. The g(r) can be transformed into S(Q) via the inverse transform

$$S(Q) = 1 + 4\pi\rho \int_{0}^{\infty} dr r^2 \frac{\sin(Qr)}{Qr} g(r)$$
(6)

These equations were taken from Appendix D of "Computer Simulation of Liquids" by Allen & Tildesley. (See references section of course website.)

II.A.3. Connection with Simulation

Molecular-level simulation can routinely generate radial distribution functions. In a single component system,

$$g(r) = \frac{\rho(r)}{\rho} = \frac{\frac{n(r)}{V_{shell}(r)}}{\frac{N}{V_{sim}}} = \frac{V_{sim}}{V_{shell}(r)} \frac{1}{N} \left[\frac{1}{N} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \delta(r - r_{ij}) \right]$$
(7)

Practically speaking, equation (7) is evaluated by discretizing the radial dimension into bins. The number of bins chosen depends upon the number of particles and the duration of the simulation. To illustrate the issue, consider two asymptotes. At the very coarse scale, we have one large bin. All particle positions fall into that single bin and we have no useful information. At the opposite extreme, the very fine scale, each particle falls into a separate bin, making all bins have either a value of zero or one. This contains all the information, but the resulting plot is not what we are expecting. Therefore, an intermediate bin size is selected, typically no finer than 0.01 Å and often as coarse as 0.1 Å.

The maximum length of a pair correlation function is typically half the size of the smallest dimension of the simulation volume. If the cut-off is larger, you will start including periodic images in the RDF, giving rise to an artificial periodic structure.

II.B. Decompositions of the RDF

The RDF can be decomposed in several different ways. In simulation, a component of the RDF can be constructed just as easily as the total RDF. In experiment, only the total RDF is available.

II.B.1. Decomposition by Species

The RDF between two particle atom types, say between types A and B, (the distribution of atoms of type B around atoms of type A) is

$$g_{AB}(r) = \frac{\rho_{AB}(r)}{\rho_{B}} = \frac{\frac{n_{AB}(r)}{V_{shell}(r)}}{\frac{N_{B}}{V_{sim}}} = \frac{V_{sim}}{V_{shell}(r)} \frac{1}{N_{B}} \left[\frac{1}{N_{A}} \sum_{i=1}^{N_{A}} \sum_{j=1}^{N_{B}} \delta(r - r_{ij}) \right]$$
(8)

The distribution of atoms of type A around atoms of type B is

$$g_{BA}(r) = \frac{\rho_{BA}(r)}{\rho_{A}} = \frac{\frac{n_{BA}(r)}{V_{shell}(r)}}{\frac{N_{A}}{V_{sim}}} = \frac{V_{sim}}{V_{shell}(r)} \frac{1}{N_{A}} \left[\frac{1}{N_{B}} \sum_{i=1}^{N_{B}} \sum_{j=1}^{N_{A}} \delta(r - r_{ij}) \right]$$
(9)

A consequence of this is the following

$$g_{AB}(r) = g_{BA}(r) \tag{10}$$

$$\rho_{AB}(r) \neq \rho_{BA}(r) \tag{11}$$

$$\frac{\rho_{AB}(r)}{\rho_B} = \frac{\rho_{BA}(r)}{\rho_A}$$
(12)

II.B.2. Decomposition by Intermolecular & Intramolecular Components

The pair correlation function is often decomposed into intermolecular and intramolecular components. Part of the motivation for doing so lies in the fact that the intramolecular contribution is generally much narrower and sharper than the non-bonded contribution. See for example the following example, taken from

Wang, Q., Keffer, D.J., Nicholson, D.M., Thomas, J.B., "Use of the Ornstein-Zernike Percus-Yevick Equation to Extract Interaction Potentials from Pair Correlation Functions", Phys. Rev. E 81(6) 2010 article # 061204, doi: 10.1103/PhysRevE.81.061204

In Figure 3, the main plot shows the intermolecular contribution to the radial density function, $\rho(r) = \rho g(r)$, for N₂. The inset shows the intramolecular contribution. The reason that $\rho(r)$ rather than g(r) is plotted in this paper is because g(r) is normalized to go to unity as r goes to infinity, but such normalization is impossible for an intramolecular mode than goes to zero as r goes to infinity. No such scaling problem exists for $\rho(r)$. Note that the scale of the y-axis of the intramolecular plot is a factor of about 200 larger than the intermolecular plot. Also the intramolecular contribution is only non-zero for a very short radial distance, corresponding to the nitrogen-nitrogen bond distance.



FIG. 8. Bonded and nonbonded RDFs obtained from MD simulation of the diatomic fluid at T^* =8.3333, ρ^* =0.073 41.

Figure 3. Intermolecular (nonbonded) and intramolecular (bonded) contributions to the radial density function of molecular nitrogen.

II.B.3. Others Decompositions

Various other decompositions may be useful to understanding the structure of complex materials with hierarchical structure. For example, we examine a composite from the following reference.

McNutt, N.W., Rios, O., Feygenson, M., Proffen, T.E., Keffer, D.J., "Structural Analysis of Lignin-derived Carbon Composite Anodes", J. Appl. Crystallogr. 47(5) 2014 pp. 1577-1584, doi: 10.1107/s1600576714014666.

In this example, the material is composed of carbon in amorphous (A) and crystalline (C) domains. Six RDFs are shown, including the total, the amorphous-amorphous (AA), the amorphous-crystalline (AC) contribution. The crystalline-crystalline contribution is further split into two components—intercrystallite (shown) and intracrystallite (not shown). The intracrystallite is further decomposed into two components—interplanar and intraplanar. This decomposition is useful in showing the origin of all of the features in the total RDF, a capability not available from experiment alone.



Figure 3

Simulated PDFs with decomposition for two materials with common density and crystalline volume fraction ($\rho = 1.38 \text{ g cm}^{-3}$, $\Phi_c = 0.5$). (a) Smaller crystallites (r = 7 Å) and (b) larger crystallites (r = 17 Å).

Figure 4. Decomposition of the RDF of a carbon composite containing crystalline nanoparticles distributed in an amorphous domain.

II.C. What do RDFs look like.

We have already shown numerous RDFs. However, here we recap and add some additional information.

II.C.1. Gas Phase

A typical gas phase RDF is shown in the low density curves of Figure 1. In fact, at very low densities,

$$\lim_{\rho \to 0} g(r) = \exp\left(-\frac{U(r)}{k_b T}\right)$$
(13)

For a Lennard-Jones fluid



Figure 5. RDF of a Lennard-Jones fluid at a reduced temperature of unity and a density approaching infinite dilution.

II.C.2. Liquid Phase

Typical liquid phase RDFs are shown in the high density curves of Figure 1. The longer range structure is due to partial ordering of the fluid. The first trough is due to the excluded volume of particles occupying the first peak. The second peak is attraction to particles in the first peak. The second trough is due to the excluded volume of particles occupying the second peak. The third peak is attraction to particles in the second peak. And so on. The magnitude of the peaks diminish, reflecting no long range order in liquids.

II.C.3. Crystalline Solid Phase

The RDF of a crystalline materials shows distinct peaks corresponding to the separation between points of the crystal lattice. The magnitude of such peaks corresponds to the number of such neighbors at that distance. the width of the peak corresponds to the magnitude of the vibration around the lattice point. Often this vibration is due to thermal noise. An example is shown below, taken from the following reference.

Title: Role of cation choice in the radiation tolerance of pyrochlores[†] Author: Ram Devanathan *, Fei Gao and Christina J. Sundgren DOI: 10.1039/C2RA22745B (Paper) RSC Adv., 2013, 3, 2901-2909



Figure 5. The Ce-O component of the RDF shows the distinct crystal structure of the material.

The integral over any individual peak in Figure 5 corresponds to the number of particles at that distance.

II.D. Relation to thermodynamic Properties

The Henderson theorem states that the radial distribution corresponds to a unique interaction potential.

Title: UNIQUENESS THEOREM FOR FLUID PAIR CORRELATION-FUNCTIONS Author: Henderson, R.L. PHYSICS LETTERS A Volume: A 49 Issue: 3 Pages: 197-198 DOI: 10.1016/0375-9601(74)90847-0 Published: 1974 potential energy example

Obtaining the interaction potential from the RDF is generally known as the "inverse problem". It remains an open field of research.

Regardless, whether the potential is used to generate the RDF or vice versa, once both are known, one can compute thermodynamic properties. Knowing the RDF and another variable, such as temperature, is sufficient to define the thermodynamic state for a single-component, single-phase system. Thermodynamic properties can be obtained from the RDF as follows. Some properties, such as the potential energy and the pressure can be obtained directly from the

$$U = 2\pi N \rho \int_{0}^{\infty} u(r)g(r)r^{2}dr$$
(14)

$$p = \rho k_B T - \frac{2\pi\rho^2}{3} \int_0^\infty r \left(\frac{du(r)}{dr}\right) g(r) r^2 dr$$
(15)

Other properties, such as the heat capacity, can be determined from the temperature dependence of the pair correlation function.

$$C_{v} = \left(\frac{\partial E}{\partial T}\right)_{N,V} = \frac{3}{2}Nk_{B} + 2\pi N\rho \int_{0}^{\infty} u(r) \left(\frac{\partial g(r)}{\partial T}\right)_{N,V} r^{2} dr$$
(16)

Of course, in an MD simulation, the potential energy and pressure can be explicitly evaluated. These formula are presented in order to show a theoretical link between the thermodynamic and structural properties of the system.

II.E. Some Examples

II.E.1. Adsorption in an Amorphous Nanoporous Material

In amorphous or disordered systems the meaning of the RDF many not be immediately apparent. Therefore, snapshots from MD simulations can be included to illustrate the local environment captured in a particular peak. As an example we consider some RDFs from the following example.

Suraweera, N.S., Albert, A.A., Humble, J.R., Barnes, C.E., Keffer, D.J., "Hydrogen adsorption and diffusion in amorphous, metal-decorated nanoporous silica", Int. J. Hydrogen Energy 39 2014 pp. 9241-9253, doi: 10.1016/j.ijhydene.2014.03.247.

In Figure 6, H_2 is adsorbed in an amorphous silica material, based on an inorganic polymer of silsesquioxane units. The H_2 - H_2 and H_2 -Si are shown in Figure 6. Since the structure is amorphous, snapshots taken from configurations of the simulation are used to give a visual example of what the peaks correspond to.



Figure 6. H₂-H₂ and H₂-Si RDFs for adsorption of H₂ in amorphous silica frameworks.

II.E.2. Flow-Induced Crystallization of Polymers

A flow field can induce polymers to crystallize or partially crystallize above the melting temperature. Measuring the RDF in a flowing system experimentally requires that the rheometer be placed inside a beam line. If this is not possible experimentally, simulation is well suited to characterize the structural transition. This examples is taken from the following reference.

Ionescu, T.C., Baig, C., Edwards, B.J., Keffer, D.J, Habenschuss, A., "Structure Formation under Steady-State Isothermal Planar Elongational Flow of n-Eicosane: A Comparison between Simulation and Experiment", Phys. Rev. Lett. 96(3) 2006 article #037802, doi: 10.1103/PhysRevLett.96.037802.



FIG. 1. Comparison of the structure factor, s(k), between experiment (Ref. [21]), and simulation under quiescent conditions at T = 315 K and $\rho = 0.81$ g/cm³.



FIG. 3. (a) Structure comparison in terms of x-ray diffraction data between liquid (Ref. [21]) and crystalline *n*-eicosane; (b) Structure comparison in terms of total static structure factor between x-ray diffraction data for the crystalline *n*-eicosane and the simulated elongated structure.

Figure 7. Left: Structure factor s(k) (equivalently S(Q)) for eicosane from simulation and experiment at equilibrium. Right: s(k) for flowing system (simulation) or thermally induced crystal (experiment).

In Figure 7, the structure factor for the flowing system is compared to a thermally induced crystal from the experimental system. Thus the comparison should not be exact. However, the agreement at long k (short r) is excellent, showing that the flowing system has begun to adopt the short-range crystalline order due to flow-induced crystallization.

II.E.3. Carbon Composites

This example refers to Figure 4 above.

McNutt, N.W., Rios, O., Feygenson, M., Proffen, T.E., Keffer, D.J., "Structural Analysis of Lignin-derived Carbon Composite Anodes", J. Appl. Crystallogr. 47(5) 2014 pp. 1577-1584, doi: 10.1107/s1600576714014666.

II.F. In LAMMPS

LAMMPS has efficient functionality to compute the RDF up to the cut-off distance for the potential. The RDF command is invoked with a compute command. The following example computes four RDFs, for 1-1, 1-2, 2-1 and 2-2 pairs. The 1-2 and 2-1 RDF should be identical. However the integral should not since the scaling factors are different.

compute rdf all rdf 100 1 1 1 2 2 1 2 2

The rdf can be output using a fix statement.

```
fix rdf all ave/time ${Nsamplefreq} ${Nsamplesizerdf}
${Nblocksizerdf} c_rdf file output_rdf.txt mode vector
```

As written, this output file contains columns for both the RDF and the cumulative integral of the RDF, also known as the coordination number.

If you want the RDF out to a longer distance, then this must be accomplished based on postprocessing of saved configurations.

III. End-to-End Distribution Functions

Structure can be captured in distribution functions other than the RDF. The end-to-end distribution function is a measure often used to examine polymer chains. One example is taken from the following reference.

Title: Molecular Dynamics Simulation of Polyethylene Terephthalate Oligomers Authors: Wang, Q., Keffer, D.J., Petrovan, S., Thomas, J.B. Journal: *J. Phys. Chem. B* Vol. 114 Issue 2. pp. 786–795. Published 2010 doi: http://doi.org/10.1021/jp909762j



Figure 8. Left: end-to-end distribution as a function of chain length in PET. The corresponding correlation function used to measure the longest relaxation time (Rouse time) of the chain.

In Figure 8 (left), the monomer has a single peak. The dimer has two peaks, representing a folded and unfolded configuration. As the chain length increases, the probability of the folded configuration diminishes.

In Figure 8 (right), the relaxation time (the characteristic time associated with the exponential decay of these curves) increases with chain length. Ideally, to sample independent configurations, the simulation duration should be many multiples of this relaxation time.

The distribution of end-to-end distribution changes in the presence of a flow field. To illustrate this point, we take an example from the following reference.

Title: Single-chain dynamics of linear polyethylene liquids under shear flow Authors: Kim, J.M., Edwards, B.J., Keffer, D.J., Khomami, B. Journal: Phys. Lett. A Vol. 373 Issue 7 pp. 769-772 Published 2009 doi: <u>http://doi.org/10.1016/j.physleta.2008.12.062</u>



Figure 9. End-to-end distributions as a function of Weissenberg number, a dimensionless representation of flow field strength.

In Figure 9, the single peak at low Weissenberg numbers (weak flow) eventually changes to a bimodal distribution, corresponding to folded and extended polymers. Some polymers flow in this shear flow due to the vorticity in the flow field.

These end-to-end distribution functions are computed during post-processing of saved configurations.

IV. Density Distributions

Molecular dynamics simulations are also able to generate three-dimensional distributions. An example is shown below. The "clouds" in these plots correspond to iso-density surfaces. Volume inside the clouds corresponds to higher densities. Volume outside the clouds corresponds to lower densities. This example is taken from the following reference.

Suraweera, N.S., Xiong, R., Luna, J.P., Nicholson, D.M., Keffer, D.J., "On the Relationship between the Structure of Metal-Organic Frameworks and the Adsorption and Diffusion of Hydrogen", Molec. Simul. 37(7) 2011 pp. 621-639, doi: 10.1080/08927022.2011.561432



Figure 5. Density distributions at 77 K and 1 bar. The surface is drawn at a contour value of 0.015 mol/Å^3 . For each IRMOF, two views are shown, including (I) a view of the unit cell with a small cage in the centre and (II) a vertex. IRMOF legend: (a) IRMOF-1, (b) IRMOF-2, (c) IRMOF-3, (d) IRMOF-7, (e) IRMOF-8, (f) IRMOF-10, (g) IRMOF-10Br², (h) IRMOF-10Br³, (i) IRMOF-10NH²₂ and (j) IRMOF-10NH²₃. Colour legend: violet, Zn; red, O; green, C; white, H; brown, Br; blue, N (colour online).

Figure 10. Density distribution of RDX center of mass in a couple crystalline metal-organic frameworks.

In Figure 10, a couple examples of three-dimensional density distributions are shown for the center of mass of the explosive RDX in several iso-reticular metal-organic frameworks (IRMOFs). Depending on where the contour is drawn you convey different information. The low contour contains more volume and shows paths for movement from one cage to the next. The high value contour contains just the volume associated with the deepest adsorption site. Two views are shown, one of the entire cage and one of the cage vertex where the adsorption site is located.

These three-dimensional density distribution functions are computed during post-processing of saved configurations. In this example, the simulation volume contained many cages. The density distribution is a result of mapping the RDX in all cages onto a single cage. The volume of that cage is divided into bins in three-dimensions. Some external visualization software takes the three-dimensional histogram and generates the three-dimensional contour plots.