



The University of Manchester Research

# Overcoming the Limitations of Cutoffs for Defining Atomic Coordination in Multicomponent Systems

DOI: 10.1002/jcc.25137

#### **Document Version**

Accepted author manuscript

Link to publication record in Manchester Research Explorer

#### Citation for published version (APA):

Higham, J., & Henchman, R. (2017). Overcoming the Limitations of Cutoffs for Defining Atomic Coordination in Multicomponent Systems. *Journal of Computational Chemistry*. https://doi.org/10.1002/jcc.25137

Published in: Journal of Computational Chemistry

#### Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

#### **General rights**

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

#### **Takedown policy**

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



## Overcoming the Limitations of Cutoffs for Defining Atomic Coordination in Multicomponent Systems

Author Jonathan Higham<sup>\*</sup>, Richard H. Henchman<sup>†</sup>

November 29, 2017

#### Abstract

A common way to understand structure in multimolecular systems is the coordination shell which comprises all the neighbors of an atom. Coordination, however, is non-trivial to determine because there is no obvious way to determine when atoms are neighbors. A common solution is to take all atoms within a cutoff at the first minimum of the radial distribution function, g(r). We show that such an approach cannot be consistently applied to model multicomponent systems: mixtures of atoms differing in size or charge. Coordination shells using the total g(r) are found to be too restrictive for atoms of different size while those using pairwise g(r)s are excessive for charged mixtures. The recently introduced relative angular distance (RAD) algorithm, however, which defines coordination instantaneously from atomic positions, is consistently able to define coordination shells containing the expected neighboring atoms for all these systems. A more robust way to determine coordination should in turn make coordination a more robust way to understand structure.

Keywords: Liquid, mixture, coordination shell, Lennard-Jones, molten salt.

<sup>\*</sup>Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom, and School of Chemistry, The University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom.

<sup>&</sup>lt;sup>†</sup>Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom, and School of Chemistry, The University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom, henchman@manchester.ac.uk



Relative angular distance Cutoff from radial distribution function

The relative angular distance (RAD) algorithm identifies the coordination shell of an atom's neighbors instantaneously from the coordinates of atoms in a series of mixtures. Standard methods rely on a cutoff placed at the first minimum in the radial distribution function, which has to be constructed as an average over many configurations, assumes a fixed, spherical boundary, and cannot be defined consistently for any one kind of radial distribution function for a mixture.

### INTRODUCTION

Structure is a fundamental quantity for any molecular system. However, beyond the cases of crystalline systems or single, rigid molecules whose atoms have well-defined positions, structure is a challenge to define. The most common approach makes use of the radial distribution q(r) which quantifies the average probability of two atoms at a distance r relative to the probability at any distance. While g(r) has the advantage of being accessible by experiment or simulation, it only gives a partial picture of local structure because it averages extensively in time and space. Another widely used way to understand structure is the coordination shell of an  $atom^{1,2}$  and its associated coordination number  $N_c$ . This gives discrete, instantaneous knowledge about the local structure of any atom. However, there is no general way to determine  $N_{\rm c}$  and multiple methods exist to define it. The simplest approach is the nearest-neighbor method whereby  $N_{\rm c}$  is fixed for all atoms at some pre-set value, such as the value in a crystal.  $N_c$  can be derived from  $g(r)^3$ , the most common method being to set  $N_{\rm c}$  equal to the number of atoms within a cutoff placed at the first minimum of q(r). While simple in principle, cutoffs require a smooth, well-converged q(r) with a well-defined minimum which requires averaging over many different configurations but at the cost of information about each individual configuration. Moreover, cutoffs are abrupt, making them sensitive to small fluctuations, they assume spherical symmetry, and must be derived for every atom pair at the conditions of interest because they are not transferable. Cutoffs do not have to be derived from g(r) and can be specified arbitrarily. A third method to determine coordination is Voronoi analysis<sup>4</sup> which defines neighbors instantaneously for a given configuration without having to specify cutoffs, from q(r) or otherwise. Atoms are defined as neighbors if their Voronoi polyhedra share a common face. While intuitive, this method yields excessive values of  $N_{\rm c}$  compared to g(r)-derived values and second-shell atoms commonly get included in the coordination shell, even for crystals<sup>5</sup>. The ideal method to determine coordination would yield  $N_{\rm c}$  instantaneously but not require the evaluation of q(r) or any ad hoc cutoffs. A recent method which does exactly this is the Relative Angular Distance (RAD) algorithm<sup>6</sup>. RAD defines two atoms as neighbors if no third atom blocks their interaction; the closer the third atom is to both atoms and the more collinear the atoms,

the more likely the third atom is to block.

Despite the limitations of cutoffs, whether or not they are derived from g(r), they are still widely used, especially in multicomponent systems in which the limitations discussed above are even more acute. For cutoffs derived from g(r), there are multiple types of g(r) to consider from which to derive a cutoff: either the total g(r) or the g(r) between particular pairs of atoms may be used. Here we compare  $N_c$  values determined by two different cutoff methods for two kinds of binary 50:50 mixture: Lennard-Jones atoms of either different size or different charge. The purpose of our study is not to examine the structure of mixtures, which is a long-standing and well-researched topic<sup>7-10</sup>, but rather to use the mixture as a model to demonstrate the limitations of cutoff methods and how these may be overcome. One cutoff method has the cutoff at the first minimum in the full radial distribution function of all atoms and the other method has individual cutoffs for each pairwise radial distribution function specific to the two types of atoms involved. Neither method is found to capture the coordination shell consistently for all systems, even to a qualitative level of accuracy. The RAD method, on the other hand, successfully picks out the neighboring atoms in all cases.

## **METHODOLOGY**

#### **Coordination-Number Algorithms**

Three methods are used to define  $N_{\rm c}$ :

1. Relative Angular Distance  $(RAD)^6$ . With respect to atom *i*, atom *j* is unblocked if

$$\frac{1}{r_{ij}^2} > \frac{1}{r_{ik}^2} \cos \theta_{jik} \tag{1}$$

for all atoms k, where  $r_{ij}$  and  $r_{ik}$  are the distances from i to j and i to k and  $\theta_{jik}$  is the angle subtended at i by j and k. Atom j is deemed to be in i's shell if j and every atom closer to i than j are unblocked. RAD efficiently and non-recursively works outwards along the sorted list of atom j's nearest neighbors, comparing only closer atoms k as potential blockers, and ending the search if a block occurs. RAD operates independent of atom type. Radial distribution functions of atoms in the RAD shell are accumulated over all configurations.

- 2. Total g(r) Cut-off (GC<sub>tot</sub>). The coordination shell is all atoms within the cutoff placed at the first minimum in the total g(r).
- 3. Pairwise g(r) Cut-off (GC<sub>pair</sub>). The coordination shells for each atom pair are evaluated separately up to the first minimum of the respective pairwise radial distribution functions g(r) and then combined to give the total.

 $N_{\rm c}$  values are calculated using C++ code which reads in a coordinate trajectory from a molecular dynamics simulation. Distribution functions are recorded with a bin-width of 5 pm.

#### **Molecular Dynamics Simulations**

Each system consisted of 600 atoms. For the neutral mixtures, half of the atoms had Lennard-Jones parameters of  $\epsilon = 0.9961$  kJ mol<sup>-1</sup> and  $\sigma = 0.3405$  nm to mimic argon<sup>11</sup> (A) and the other half (B) have  $\sigma$  scaled by a factor, which ranges from 1.0 to 3.2 in 0.2 increments, making twelve systems altogether. Lorentz-Berthelot mixing rules were used. For the charged mixture, the same Lennard-Jones parameters for argon were used for all atoms but with charges of +1 e for half the atoms (A) and -1 e for the other half (B). All simulations were performed with the sander module of AMBER 14.<sup>12</sup> The Lennard-Jones mixture with equal-sized atoms was made using the AMBER module leap, minimized with 500 steps of steepest descent minimisation, equilibrated for 100 ps at constant volume and temperature of 137.77 K using the Langevin thermostat, followed by a 1 ns at a constant pressure of 1400 bar using the Berendsen algorithm and a time constant of 2 ps (chosen to reproduce a temperature of 1.15  $\epsilon/k_{\rm B}$  and density of 0.9  $\sigma^3$ ). Subsequent Lennard-Jones mixtures had the same protocol and modified topology files, with the restart file coming from the mixture with the successively smaller scale factor. A similar protocol was used for the charged mixture except that the temperature was 1500 K to ensure a molten phase, the pressure was 1 bar, and the starting structure was taken from the equilibrated simulation with equal-sized neutral

atoms. In all simulations, a 2 fs timestep, a 0.8 nm cut-off, periodic boundary conditions, and Particle-Mesh Ewald with default AMBER parameters were used. Structures for analysis were collected every 20 ps in a further 10 ns simulation for each system under the same conditions as the final stage of equilibration.

## RESULTS

#### Lennard-Jones Mixtures

The total radial distribution functions g(r), together with the RAD g(r) and all GC cutoffs, are shown for the different mixtures in Figure 1. For  $\sigma_{\rm B}/\sigma_{\rm A} = 1$ , equivalent to the pure liquid, the plot is the same as in earlier work<sup>6</sup>. As  $\sigma_{\rm B}/\sigma_{\rm A}$  increases, the first peak gradually resolves itself into three peaks, one for each of the AA, AB/BA and BB interactions, where AB indicates B around A, and BA, conversely, A around B. The GC<sub>tot</sub> cutoff aligns closely with the AA cutoff because the AA interaction, being the shortest, brings about the first minimum in g(r). The RAD shell picks out all three peaks because all three kinds of interaction contribute to the coordination shell. The outer RAD peaks are a smaller component of the total g(r) because these peaks also comprise second-shell interactions, which RAD excludes, as well as first-shell interactions. For mixtures with the largest  $\sigma_{\rm B}/\sigma_{\rm A}$ , the middle RAD peak, corresponding to the AB/BA interactions, fades away because of a phase separation between the small and large atoms, discussed later.

Average coordination numbers  $N_c$  by the three methods are plotted in Figure 2 as a function of  $\sigma_B/\sigma_A$  for total, AX, BX, AA, AB/BA and BB coordinations, where "X" denotes either atom. Overall, total  $N_c$  values are highest for  $GC_{pair}$ , lowest for  $GC_{tot}$ , and intermediate for RAD. RAD values decrease smoothly from 9.6 at  $\sigma_B/\sigma_A = 1$  to 5 at  $\sigma_B/\sigma_A = 2.4$ because of reduced packing for differently sized atoms and then increase for larger  $\sigma_B/\sigma_A$ . The AA and BB values of  $N_c$  for RAD mirror the total trend, while AB and BA gradually drop away, indicating phase separation. RAD's asymmetry in that BA coordination exceeds AB coordination captures the intuitive idea that more smaller atoms can pack around a larger atom than larger atoms can fit around a smaller atom. The BB coordination exceeds that of AA at larger values of  $\sigma_{\rm B}/\sigma_{\rm A}$  because the larger size of B means that fewer BB interactions are disrupted than AA interactions. Strikingly, GC<sub>tot</sub> values start at 12.6 but drop precipitously to 0.7 at  $\sigma_{\rm B}/\sigma_{\rm A} = 1.4$  because of a marginally resolved first minumum in g(r) (Figure 1) owing to AA interactions. The AA coordination gradually increases as its first peak gets better resolved while the AB/BA and BB interactions contribute nothing to coordination because the larger size of B means they are beyond the AA cutoff. GC<sub>pair</sub> values, which use individual cutoffs for each pair, do not suffer from this problem. The total  $N_{\rm c}$  stays at ~12 for all values of  $\sigma_{\rm B}/\sigma_{\rm A}$ . GC<sub>pair</sub> and RAD share similar trends for AA and AB/BA neighbors but the BB coordination is notably smaller for RAD because of the blocking of BB interactions by A atoms.

At a higher level of detail, Figure 3 shows the probability distributions of  $N_{\rm c}$  for all three methods. The GC<sub>tot</sub> distributions are roughly Gaussian but for  $\sigma_B/\sigma_A \ge 1.4$  there is a large spike in  $N_{\rm c} = 0$  which is off the scale in each plot. This primarily represents isolated B atoms because all other atoms lie beyond the cutoff determined by the AA interaction. Consistent with the averages in Figure 2, non-zero coordinations for these systems are essentially due to AA interactions. The  $GC_{pair}$  and RAD distributions have no such spike at  $N_c = 0$  but both display some degree of bimodality. The bimodality is stronger for GC<sub>pair</sub> and extends over  $1.4 \le \sigma_{\rm B}/\sigma_{\rm A} \le 3.2$  while for RAD it occurs for  $2.4 \le \sigma_{\rm B}/\sigma_{\rm A} \le 3.2$ . For GC<sub>pair</sub> at low values of  $\sigma_{\rm B}/\sigma_{\rm A}$ , the first peak represents AX interactions and the second peak BX interactions. For both methods at higher  $\sigma_{\rm B}/\sigma_{\rm A}$ , the smaller contributions from AB/BA interactions mean that the peaks represent predominantly AA and BB interactions, respectively. The widely different  $N_{\rm c}$  values for the two GC methods suggest an intermediate method whose cutoff lies at the first minimum in the AX and BX g(r)s. While giving slightly larger  $N_c$  values because of the inclusion of BA neighbors, this still omits the AB and BB interactions because the first minima of AX and BX are dominated by the shorter AA and BA interactions, yielding  $N_{\rm c}$  values still below that of RAD (data not shown). In summary, RAD and  $GC_{\rm pair}$  yield plausible values of  $N_{\rm c}$  and agree moderately well for Lennard-Jones mixtures of differently sized atoms, whereas  $GC_{tot}$  is dominated by the shortest pairwise interaction and clearly fails.

#### **Charged Lennard-Jones Mixtures**

Coordination values  $N_c$  for the charged mixture using all three methods are listed in Table 1. The total and pairwise g(r)s are shown in Figure 4. GC<sub>tot</sub> and RAD give similar  $N_c$  values to each other for all types of interaction, with RAD  $N_c$  values being slightly larger than GC<sub>tot</sub> values. Symmetry in charge ensures that AA and BB coordinations are identical, as are AB and BA. Total  $N_c$  values deviate a small amount from the crystalline value of 6 because of thermal disorder, and mostly comprise counter-ion AB pairs as expected. The RAD-shell g(r)s tail off smoothly to zero versus the sudden drop for the GC<sub>tot</sub> shells, but nonetheless they closely coincide. However, for the GC<sub>pair</sub> method, while its  $N_c$  values are in close agreement with RAD for the AB coordination, the values for the AA and BB coordinations are excessively large. This is because the first peaks for these same-charged pairs lie at the second coordination shell overall due to electrostatic repulsion. The intermediate method considered earlier based on AX and BX g(r)s does not help here because it is identical to the GC<sub>tot</sub> method. In summary for charged mixtures, it is RAD and this time the GC<sub>tot</sub> which yield plausible values and agree moderately well, whereas GC<sub>pair</sub> comprises predominantly second-shell interactions and clearly fails.

## DISCUSSION

RAD consistently picks out the coordination shell of neighboring atoms for a range of mixtures, whether the atoms are of different size or charge. In contrast, methods based on cutoffs at the first minima of g(r) are unable to achieve this feat, performing well for one case but poorly for the other. A cutoff based on the total g(r) GC<sub>tot</sub> fails for atoms of different sizes while a cutoff based on the pairwise g(r) GC<sub>pair</sub> fails for charged mixtures. The failure for atoms with different sizes occurs because GC<sub>tot</sub> excludes larger atoms from shells, resulting in  $N_c$  values that are too small. If atoms are charged, the repulsion between identical ions places the first minimum at the second shell overall, resulting in  $N_c$  values that are too large. Without knowing in advance the behaviour of the system, it is impossible to know which to use. It could be argued that common sense could be used to decide which method is appropriate, but this is subjective and likely to fail in more complex cases. Moreover, extrapolating these findings for simple model mixtures, neither of the two methods should work at all for multicomponent systems having atoms of different size and charge. One possible alternative is to use contact distances from atom surfaces instead of distances between atom centers in the construction of g(r). However, this is strictly only valid for hard spheres and ignores the differing energetic interactions of atom pairs, is technically more complex, and the disadvantages of cutoffs remain.

To the best of our knowledge, no methods related to RAD have been used to determine neighbors in molecular systems. However, similar methods have been used elsewhere. Subsequent to developing RAD, we found a method in the geography literature, referred to as the Gabriel graph<sup>13</sup>, which is a version of blocking using inverse distances  $1/r_{ij}$  and  $1/r_{ik}$  in Eq. 1 instead of the inverse-square distances as in RAD. The Gabriel graph assigns two points as neighbors if the circle centered on the points' midpoint with diameter equal to their distance contains no other point. This circle is parametrised by point k in the equation  $r_{ik}/r_{ij} = \cos \theta_{jik}$ . Blocking occurs if any atom k lies within this circle such that  $r_{ik}/r_{ij} < \cos \theta_{jik}$ . RAD's blocking region is defined as  $(r_{ik}/r_{ij})^2 = \cos \theta_{jik}$ , which resembles a squashed circle. More blocking regions are possible such as the overlap of two intersecting circles or a lune<sup>14</sup>. A further point of interest is that the Gabriel graph is equivalent to the coordination derived from a modified Voronoi analysis in which points are not neighbors if their connecting line does not intersect their shared face  $^{14-16}$ . Both we<sup>6</sup> and Malins et al.<sup>16</sup> had tested these respective but equivalent methods earlier but discounted them because they gave too large a coordination for the important simple-cubic geometry by including the diagonal neighbors into the coordination shell when there was infinitesimal distortion. Regarding other possible variations of RAD, our original formulation<sup>6</sup> had the provision to account for atom size. However, the application of this to mixtures made little difference and so was not adopted here for simplicity. Charge could also be included in the blocking but it cancels on both sides of the equation in the cases considered here.

## CONCLUSIONS

We have demonstrated that methods to determine the coordination shell based on cutoffs at the first minimum in the radial distribution function cannot be consistently applied to different kinds of mixtures. The GC<sub>tot</sub> method with a cutoff at the first minimum in the total g(r) yields plausible values of  $N_c$  for charged mixtures but gives values that are much too small for mixtures of differently sized atoms; the GC<sub>pair</sub> method, on the other hand, with individual cutoffs at the first minima in each of the pair-specific g(r)s, yields plausible values of  $N_c$  for mixtures of differently sized atoms but gives excessively large values for charged mixtures. Thus, short of arbitrarily setting the cutoff, some other criterion would have to be established as to which method to use, and neither method may be appropriate. In contrast, RAD has shown itself to be a robust, general method to define an atom's coordination shell for both kinds of mixtures, and its application to mixtures of differently sized or charged atoms consistently picks out the neighboring atoms. While there is no exact, rigorous way to define coordination, the robustness and generality of RAD should help make coordination a more reliable way to probe structure in complex molecular systems.

#### ACKNOWLEDGMENTS

This work was funded by BBSRC grant BB/K001558/1.

## References

- 1. J. D. Bernal, Proc. R. Soc. London Ser. A 280, 299 (1964).
- 2. W. Brostow and Y. Sicotte, Physica A 80, 513 (1975).
- 3. P. G. Mikolaj and C. J. Pings, Phys. Chem. Liq. 1, 93 (1968).
- 4. G. Voronoi, J. Reine Angew. Math. 134, 198 (1908).
- 5. J. P. Troadec, A. Gervois, and L. Oger, Europhys. Lett. 42, 167 (1998).
- 6. J. Higham and R. H. Henchman, J. Chem. Phys. 145, 084108 (2016).
- I. R. McDonald, in *Statistical Mechanics: Volume 1*, edited by K. Singer (Royal Society of Chemistry, 1973), pp. 134–193.
- 8. T. A. Weber and F. H. Stillinger, Phys. Rev. B 31, 1954 (1985).
- 9. T. Biben and J. P. Hansen, Phys. Rev. Lett. 66, 2215 (1991).
- 10. M. Dijkstra, R. van Roij, and R. Evans, Phys. Rev. E 59, 5744 (1999).
- 11. A. Michels, H. Wijker, and H. Wijker, Physica 15, 627 (1949).
- D. A. Case, J. T. Berryman, R. M. Betz, D. S. Cerutti, T. E. Cheatham III, T. A. Darden,
  R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, et al., *AMBER 2015* (University of California, San Francisco, 2015).
- 13. K. R. Gabriel and R. R. Sokal, Syst. Zool. 18, 259 (1969).
- 14. G. T. Toussaint, Pattern Recogn. 12, 261 (1980).
- 15. J. W. Jaromczyk and G. T. Toussaint, Proc. Ieee 80, 1502 (1992).
- A. Malins, S. R. Williams, J. Eggers, and C. P. Royall, J. Chem. Phys. 139, 234506 (2013).

Figure 1: Radial distribution functions g(r) (solid gray), g(r) cutoffs (dashed gray), and RAD coordination shells (solid black) for a mixture of Lennard-Jones atoms with diameters  $\sigma_A$  and  $\sigma_B$ . The labels in the top right correspond to  $\sigma_B/\sigma_A$ .

Figure 2: Total, AX, BX, AA, AB/BA and BB  $N_c$  values determined by  $GC_{tot}$  (gray squares),  $GC_{pair}$  (gray triangles), and RAD (black circles) for a mixture of Lennard-Jones atoms of two different sizes  $\sigma_A$  and  $\sigma_B$ ; "X" denotes either atom. For RAD there are separate symbols for AB (small circles) and BA (large circles) coordination.

Figure 3: Probability distributions of  $N_c$  for  $GC_{tot}$  (gray squares),  $GC_{pair}$  (gray triangles), and RAD (black circles).

Figure 4: Total and pairwise AA, AB and BB g(r) (solid gray), GC<sub>tot</sub> cutoffs (left dashed gray) and GC<sub>pair</sub> cutoffs (right dashed gray), and the RAD shell (solid black) for the charged mixture.



Figure 1 Jonathan Higham and Richard H Henchman J. Comput. Chem.



Figure 2 Jonathan Higham and Richard H Henchman J. Comput. Chem.



Figure 3 Jonathan Higham and Richard H Henchman J. Comput. Chem.



Figure 4 Jonathan Higham and Richard H Henchman J. Comput. Chem.

Method	AA/BB	AB/BA	Total	
$\mathrm{GC}_{\mathrm{tot}}$	0.6	4.1	4.7	
$\mathrm{GC}_{\mathrm{pair}}$	15.1	5.3	20.4	
RAD	0.8	4.6	5.4	

Table 1: Charged Mixture  $N_{\rm c}$  by  ${\rm GC}_{\rm tot}$ ,  ${\rm GC}_{\rm pair}$  and RAD.