![The Journal of Open Source Software](JOSS logo)

Correlation: An Analysis Tool for Liquids and for Amorphous Solids

Isaías Rodríguez[1], **Renela M. Valladares**[2], **Alexander Valladares**[2], **David Hinojosa-Romero**[1], **Ulises Santiago**[3], and **Ariel A. Valladares**[1]

1 Departamento de Materia Condensada, Instituto de Investigaciones en Materiales, Universidad Nacional Autónoma de México, Coyoacán, Ciudad de México, México. **2** Departamento de Física, Facultad de Ciencias, Universidad Nacional Autónoma de México, Coyoacán, Ciudad de México, México. **3** Department of Computational and Systems Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA.

Summary

For almost a century, since Bernal's attempts at a molecular theory of liquid structure (Bernal, 1937), correlation functions have been the bridge to compare theoretical calculations with experimental measurements in the study of disordered materials.

Pair Distribution Functions ($g(r)$), Radial Distribution Functions ($J(r)$), Plane Angle Distributions ($g(\theta)$) and Coordination Numbers (n_c) have been widely used to characterize amorphous and liquid materials (Elliott, 1986; Valladares et al., 2011; Waseda, 1980) and, in particular Bulk Metallic Glasses (Galván-Colín et al., 2015; Miller & Liaw, 2007).

Correlation is an Open-Source software designed to analyze liquid structures and amorphous solids; the software is user-friendly, the modular design makes it easy to integrate in High-Throughput Computing (HTC) to process structures with a large number of constituents in a standardized fashion. **Correlation** is ready to be used in Windows, Linux and Mac. Currently, we support DMol3 (CAR), CASTEP (CELL), ONETEP (DAT) and VASP (POSCAR) structure files. The code can handle up to 100,000 atoms, so it can be used to analyze both classical and first-principles simulations. At the end, the output of every single correlation function is exported to the corresponding comma-separated value file (CSV), to further analyze the results.

Statement of Need

As time goes by, the number of atoms in theoretical calculations has grown from a few dozens to hundreds of thousands of atoms, and with this increment the complexity to calculate correlation functions that represent the structure of materials has steadily increased. To answer this need, there have been several tools developed to calculate some of the more used correlation functions like: pair distribution functions, radial distribution functions, and plane angle distributions. Here, we present an incomplete list of these tools:

- **Forcite Plus**: Forcite Plus is part of the Materials Studio suite (Materials Studio 2016, 2016), the program includes analyzing tools to compute structure properties like RDF and PAD.

- **PTRAJ/CPPTRAJ**: A tool designed to analyze Amber Molecular Dynamics trajectories and related properties including RDF and time correlation functions, included in the AmbarTools suite (Roe & Cheatham, 2013).

- **VASPKIT**: A post-processing tool for VASP calculated data (Wang et al., 2021), the code includes tools to analyze structural properties and dynamics trajectories.

- **rdfpy**: An open Python library for fast computation of 2D and 3D radial distribution functions of crystalline structures in the Crystallographic Information File (CIF).

- **RadialDistributionFunction** : Computes the Radial Distribution Function (RDF) of a group of positions in a file, averages it for all snapshots in the file, atom positions must be in a custom format.

However, the use of these tools has been limited, either by a prohibiting cost (**Forcite Plus**), or has been restricted to private academic groups, or geopolitical limitations introduced by the licensing (**CASTEP** postprocessing tools), or by being specially designed to specific software for material simulation (**PTRAJ/CPPTRAJ**, **VASPKIT**), or by having a narrow scope of input formats and correlation functions calculated (**rdfpy**, **RadialDistributionFunction**). With these limitations in mind, we decided to develop software that could calculate several correlation functions of materials, as well as the more interesting properties derived from these functions, while making the software accessible to as many people as possible.

Mathematical Background

Pair Distribution Functions

The structure factor is one of the most useful tools in analyzing the scattering patterns obtained from X-ray, electron and neutron diffraction experiments of crystalline, disordered and amorphous materials. The Fourier transform of the scattering intensity given by the structure factor $S(Q)$, yields the pair distribution function (PDF) $g(r)$ defined by:

$$g(r) - 1 = \frac{1}{2\pi^2 \rho_0 r} \int_0^\infty Q[S(Q) - 1]sin(Qr)dQ, \tag{1}$$

The pair distribution function could also be seen like a distance map inside the material, the $g(r)$ function gives how feasible is finding two atoms separated by the distance (r) as can be seen in Figure 1 (Ziman, 1979).

The PDF is normalized so that, as $r \to \infty$, $g(r) \to 1$, Also, as $r \to 0$ (for r shorter than the distance of the closest approach of pair of atoms), $g(r)$ becomes zero. The main advantage of the PDF and the related functions, is that they emphasize the short-range order of a material.

Reduced pair distribution function $G(r)$

One of the most widely used pair correlation function is the reduced pair distribution function (rPDF), also known in the literature as differential pair distribution function $(D(r))$. This function is defined as $G(r) = 4\pi\rho_0 r(g(r) - 1)$. From this definition, and the previously discussed tendency at large r of the PDF, it's clear that the reduced pair distribution function (rPDF) oscillates around zero as $r \to \infty$. It also becomes evident that as $r \to 0$ (for r smaller than the closest pair of atoms), the rPDF behaves like $-4\pi\rho_0$.
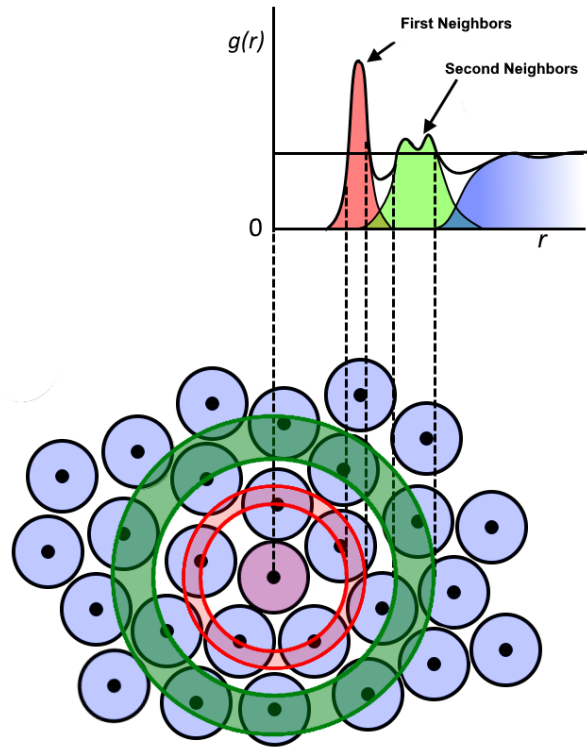
Figure 1: Schematic depiction of the first and second neighbor's coordination spheres for an amorphous metallic alloy and the corresponding pair distribution function. Design inspired by J. M. Ziman, Models of disorder (Cambridge University Press, 1979).

While the PDF $(g(r))$ has an intuitive geometric definition, the rPDF $(G(r))$ can be directly obtained by a Fourier transform of the structure factor $(S(Q))$ as can be seen in Equation 1; this close relation with the experimental data explains the popularity that this function has. Also, it has another advantage over other correlation functions [like PDF $(g(r))$] as the numerical density $\rho_0 = N/V$ needs to be estimated in order to normalize the functions. This is not necessary in rPDF $(G(r))$; where this information is already contained in $G(r)$ as the slope of the function when $r \to 0$.

Radial distribution function $J(r)$

The last correlation function we shall discuss is also one of the most physically intuitive, the RDF, $J(r)$, which is related to the Pair Distribution Function (PDF) by:

$$J(r) = 4\pi r^2 \rho_0 g(r). \tag{2}$$

The RDF has the useful property that the quantity $J(r)dr$ gives the number of atoms in a spherical shell with inner radius r and thickness dr around every atom as depicted in Figure 1. For example, the coordination number, or the number of neighbors (n_c), is given by:

$$n_c = \int_{r_1}^{r_2} J(r)dr, \tag{3}$$

where r_1 and r_2 define the RDF peak corresponding to the coordination shells in question.

Rodríguez et al., (2021). Correlation: An Analysis Tool for Liquids and for Amorphous Solids. *Journal of Open Source Software*, 6(65), 2976. 3
https://doi.org/10.21105/joss.02976

Plane Angle Distribution

The use of higher order correlation functions to analyze the structure of liquids and amorphous solids has been proposed in the literature (Galvan-Colin et al., 2016; Hafner, 1982), trying to reproduce the success obtained by Bernal in the analysis of the structure of liquids (Bernal, 1964).

In particular, the Plane Angle Distribution, also known as the Bond Angle Distribution $f(\theta)$ has been used to characterize the short-range order of amorphous and liquid structures (Galvan-Colin et al., 2016; Galván-Colín et al., 2015; Mata-Pinzón et al., 2016). Here we propose to substitute the term "Bond Angle Distribution" (BAD), by the term "Plane Angle Distribution" (PAD), also frequently used, since in condensed matter, proximity does not necessarily imply bonding.

Benchmarks

In order to assess the performance of **Correlation**, we calculated the PDF and PAD for two well known structures (Crystalline Silicon and a Graphene Layer), and compared the results with the commercially available software Forcite included in the Materials Studio suite (*Materials Studio 2016,* 2016); to test **Correlation** in amorphous and liquid materials we selected amorphous palladium (Rodríguez et al., 2019), amorphous palladium hydride (Rodríguez, 2019) and liquid bismuth. Because of the complexity to calculate PAD of amorphous and liquids in Forcite, we chose to compare them with the code developed by U. Santiago within our group. (Santiago, 2011).

The results of these benchmarks are shown in Figure 2, and the structures used to calculate these figures are included in the code as tests 1 to 5. The last structure included as test 6 is a 2x2x2 supercell of amorphous palladium hydride included in test 4, to benchmark memory and CPU performance in a structure with thousands of atoms.
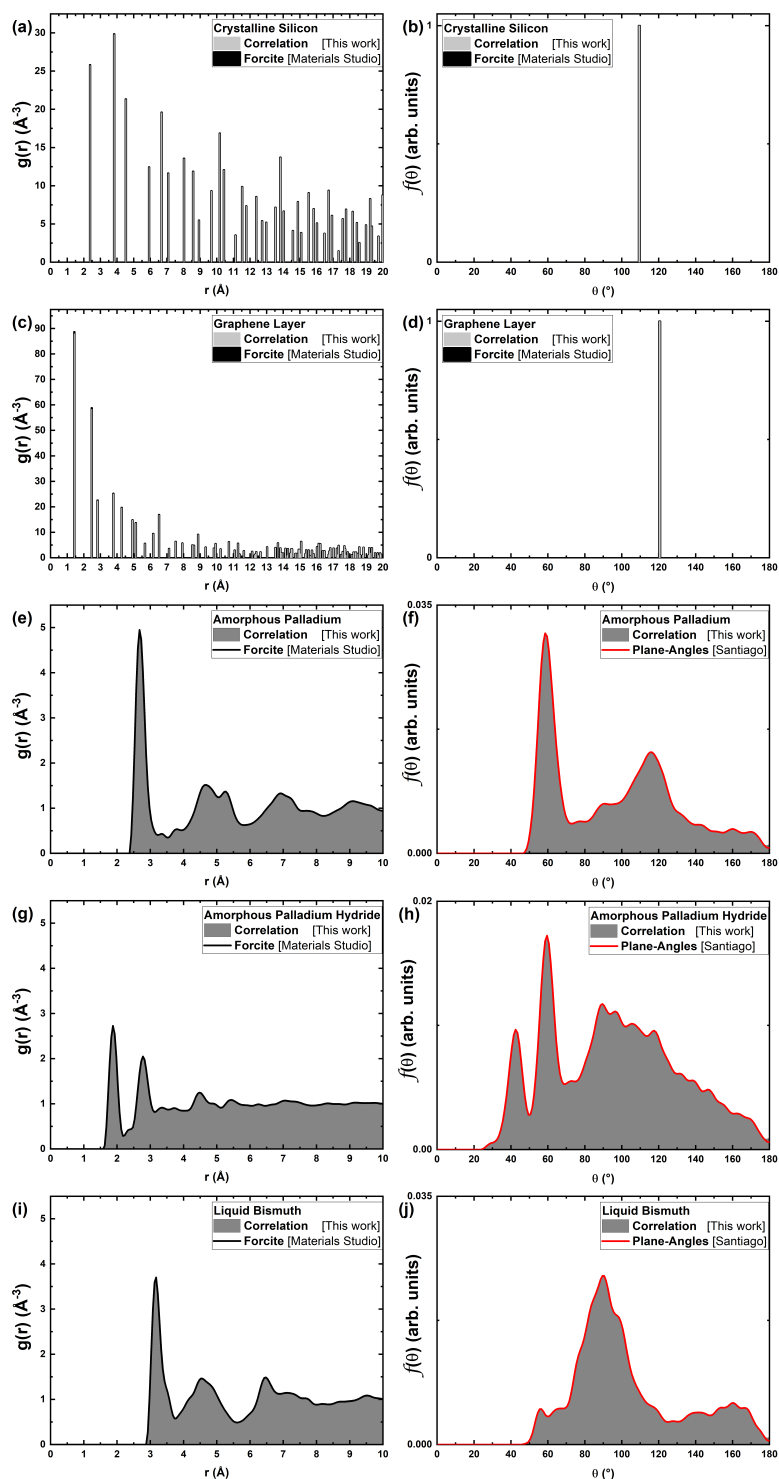
Figure 2: Pair Distribution Functions $g(r)$ on the left, Plane Angle Distributions on the right for: crystalline silicon, graphene Layer, amorphous palladium, amorphous palladium hydride and liquid bismuth. Correlation in gray, Forcite in black, Plane Angles in red. Similarity is remarkable between **Correlation** and Forcite as can be seen in all PDFs. Figures (a) to (d) indicate that the coincidence in the two results overlap completely.

Conclusion & Perspective

Correlation is a lightweight, modular software that can be used in HTC and adapted to analyze the main correlation functions used to characterize crystalline and amorphous solids, as well as liquids.

Correlation has been used in previously published work (Galvan-Colin et al., 2016; Galván-Colín et al., 2015; Mata-Pinzón et al., 2016) as well as several Ph.D. Theses (Galvan-Colin, 2016; Mata-Pinzón, 2016; Mejía-Mendoza, 2014; Rodríguez, 2019; Romero-Rangel, 2014; Santiago, 2011) done in our group. We will continue to support and enrich the software in the foreseeable future. Here we list the features planned to be added in the future:

- Support for additional output files, like hdf5 standard.

- Inclusion of other correlation functions like Velocity Correlation Functions, to further improve the analysis of liquids and phase transitions.

- Inclusion of structure factors and x-ray diffraction, to facilitate the comparison with experimental results.

- Parallelization of the main loop, to further improve the code by switching to a 'divide-and-conquer paradigm.

We are open to receive suggestions that would further improve the functionality of the software. Address all comments and observations to the GitHub: Correlation, or to the email of the first author, I.R: isurwars@ciencias.unam.mx

Acknowledgements

References

Bernal, J. D. (1964). The Bakerian Lecture, 1962 The structure of liquids. *Proc. R. Soc. Lond. A*, *280*, 299. https://doi.org/10.1098/rspa.1964.0147

Bernal, J. D. (1937). An attempt at a molecular theory of liquid structure. *Trans. Faraday Soc.*, *33*, 27. https://doi.org/10.1039/tf9373300027

Elliott, S. R. (1986). *Physics of Amorphous Materials*. John Wiley & Sons, Incorporated. ISBN: 978-0-470-20472-6

Galvan-Colin, J. (2016). *Atomic structure and properties of amorphous, liquid and amorphous porous Cu-Zr alloys by* ab initio *simulations* [PhD Thesis, Universidad Nacional Autónoma de México]. http://alephv23.cichcu.unam.mx/ptd2016/noviembre/0752730/Index.html

Galvan-Colin, J., Valladares, A., Valladares, R., & Valladares, A. (2016). *Ab initio* generation of binary alloy foams: The case of amorphous $Cu_{64}Zr_{36}$. *Proceedings of 2nd International Electronic Conference on Materials*, B001. https://doi.org/10.3390/ecm-2-B001

Galván-Colín, J., Valladares, A. A., Valladares, R. M., & Valladares, A. (2015). Short-range order in *ab initio* computer generated amorphous and liquid Cu–Zr alloys: A new approach. *Physica B: Condensed Matter*, *475*, 140. https://doi.org/10.1016/j.physb.2015.07.027

Hafner, J. (1982). Triplet correlation functions in metallic glasses. *J. Phys. F: Met. Phys.*, *12*, L205. https://doi.org/10.1088/0305-4608/12/11/001

Mata-Pinzón, Z. (2016). *Propiedades electrónicas y vibracionales y su influencia en la superconductividad del bismuto amorfo y sus aleaciones con plomo, talio y antimonio* [PhD Thesis, Universidad Nacional Autónoma de México]. http://alephv23.cichcu.unam.mx/ptd2016/noviembre/0753741/Index.html

Mata-Pinzón, Z., Valladares, A. A., Valladares, R. M., & Valladares, A. (2016). Superconductivity in Bismuth. A New Look at an Old Problem. *PLoS ONE*, *11*, e0147645. https://doi.org/10.1371/journal.pone.0147645

Materials Studio 2016: Forcite, CASTEP. (2016). Dassault Systèmes BIOVIA.

Mejía-Mendoza, L. M. (2014). *Estudio Computacional de Aleaciones Amorfas Basadas en Silicio-Carbono y Silicio-Germanio* [PhD Thesis, Universidad Nacional Autónoma de México]. http://alephv23.cichcu.unam.mx/ptd2014/junio/0714309/Index.html

Miller, M., & Liaw, P. (2007). *Bulk Metallic Glasses: An Overview*. Springer US. ISBN: 978-0-387-48921-6

Rodríguez, I. (2019). *Cálculo* ab initio *de propiedades estructurales, electrónicas y vibracionales en aleaciones de paladio* [PhD Thesis, Universidad Nacional Autónoma de México]. http://alephv23.cichcu.unam.mx/ptd2019/agosto/0793234/Index.html

Rodríguez, I., Valladares, R. M., Hinojosa-Romero, D., Valladares, A., & Valladares, A. A. (2019). Emergence of magnetism in bulk amorphous palladium. *Phys. Rev. B*, *100*, 024422. https://doi.org/10.1103/PhysRevB.100.024422

Roe, D. R., & Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.*, *9*(7), 3084–3095. https://doi.org/10.1021/ct400341p

Romero-Rangel, C. (2014). *Simulaciones Computacionales de Materiales Nanoporosos. El Caso de Carbono* [PhD Thesis, Universidad Nacional Autónoma de México]. http://alephv23.cichcu.unam.mx/ptd2014/junio/0714308/Index.html

Santiago, U. (2011). *Simulación de sistemas metálicos amorfos y porosos de elementos nobles* [PhD Thesis, Universidad Nacional Autónoma de México]. http://alephv23.cichcu.unam.mx/ptd2012/marzo/0677980/Index.html

Valladares, A. A., Díaz-Celaya, J. A., Galván-Colín, J., Mejía-Mendoza, L. M., Reyes-Retana, J. A., Valladares, R. M., Valladares, A., Alvarez-Ramirez, F., Qu, D., & Shen, J. (2011). New Approaches to the Computer Simulation of Amorphous Alloys: A Review. *Materials*, *4*, 716. https://doi.org/10.3390/ma4040716

Wang, V., Xu, N., Liu, J.-C., Tang, G., & Geng, W.-T. (2021). VASPKIT: A user-friendly interface facilitating high-throughput computing and analysis using VASP code. *Computer Physics Communications*, *267*, 108033. https://doi.org/https://doi.org/10.1016/j.cpc.2021.108033

Waseda, Y. (1980). *The structure of non-crystalline materials: Liquids and amorphous solids*. McGraw-Hill International Book Co. ISBN: 978-0-07-068426-3

Ziman, J. M. (1979). *Models of disorder: The theoretical physics of homogeneously disordered systems*. Cambridge University Press. ISBN: 978-0-521-21784-2 978-0-521-29280-1