

Techniques for the Analysis of Clusters and Aggregations within Atom Probe Tomography Data

L.T. Stephenson, M. Moody, and S.P. Ringer

Australian Key Centre for Microscopy & Microanalysis, University of Sydney, NSW, 2006, Australia

Advanced alloys may possess a discrete nanostructure or nanoscale microstructure such as atomic co-clustering, nanoscale precipitates, solute segregation and engineered interfaces. These can significantly influence the materials' macroscopic characteristics, such as mechanical and electromagnetic properties [1]. Atom probe tomography (APT) provides three-dimensional (3D) data on the spatial positions of all atom species in solids [2] and so is emerging as a key tool for nanostructural analysis into the next decade and beyond. A challenge is now to harness the rich source of data available via APT through visualisation and data mining techniques so as to quantify the structural and chemical characteristics of the nanostructure in 3D. This work is a report on our recent studies of the visualisation and analysis of atomic clustering and solute-solute interactions in dilute multi-component alloy systems. Our techniques aim to discriminate unequivocally the observed solute-solute atom dispersion and interactions with those that might be expected from a truly random distribution.

We study 1st and higher order nearest-neighbour distances (see Fig. 1), and the variation in distributions representing mean cluster size and mean cluster nearest neighbour. These statistics serve heuristically in selecting parameters needed for clustering algorithms. Furthermore, coordinate independent measures of localised solute density are also regarded, ensuring both a physical meaning to parameter selection and a solid quantitative meaning to clusters found.

There are three clustering algorithm variants that we have implemented {and/or developed}. Friends-of-Friends (FoF) is widely used within science (see e.g. studies in astrophysics [5]). The FoF thesis is that clustered solute atoms are more closely spaced than solute atoms randomly distributed within the matrix [2-4]. The FoF clustering parameter is the maximum separation between each cluster atom - d_{\max} . Chained and compact clusters are identified by this method, though some clusters are falsely identified and can be non-physical because of bridging effects and insensitivities of parameters. A density-based clustering algorithm (DBScan [6]) has been successfully implemented such that suppressing the addition of outlying atoms to clusters eliminates artefacts arising in the FoF implementation. An atom and its K nearest neighbours are clustered if the distance to the K^{th} nearest neighbour is less than the DBScan clustering parameter $d_{\max K}$. There exists a minimum constraint on cluster size and this may cause the formation of spherical non-physical clusters if K is too large. Regardless, density-based clustering is the natural algorithmic alternative when the FoF algorithm fails to perform. The discrepancy between clustering algorithms is obvious when this is the case (see Fig. 2), and any material statistics that are calculated via such techniques of APT data need careful consideration. Density-based clustering highlights regions of localised density and so can be trusted to not produce the artefacts of non-physical chained and sparse clusters.

[1] S.P. Ringer, K. Hono and T. Sakurai, *Met. Mater. Trans. A.* **26** 2207 (1995)

- [2] J.M. Hyde, Doctoral Thesis, University of Oxford, Oxford, 1993.
 [3] M.K. Miller and E.A. Kenik, *Microsc. Microanal.* 10, (2004) 336-341
 [4] D. Vaumousse, A. Cerezo, P.J. Warren, *Ultramicroscopy* 95 (2003) 215-22.
 [5] Davis, M., Efstathiou G., Frenk C. S., & White S. D. M. (1985) *ApJ*, 292, 371
 [6] M. Ester, H.P. Kriegel, I. Sander, X. Xu, *Proc. 2nd Int. Conf. on KDD* (1996) 226-23

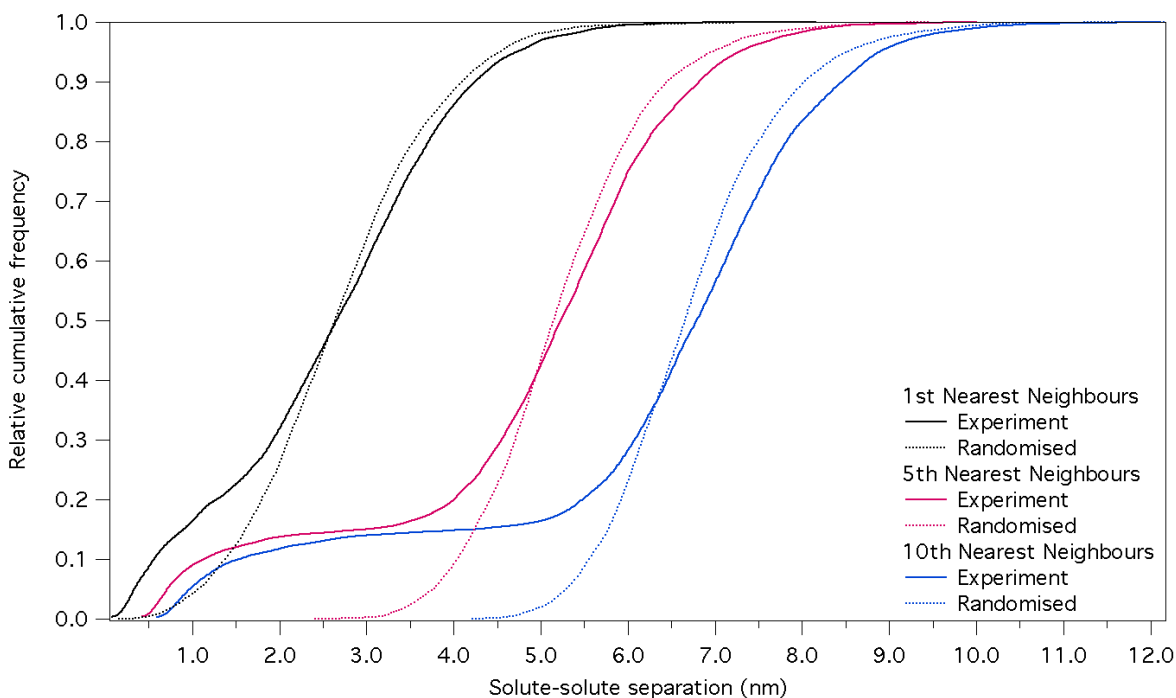


Fig. 2. Sn clustering in Al-1.7Cu-0.01Sn (at.%) aged for 30s at 200C. There is obvious clustering in the 1st nearest neighbour distribution however parameter selection from this curve is difficult. As seen, clustering becomes more defined in higher order distributions and choosing $d_{\max K}$ for density-based clustering is a case of choosing a distance corresponding with features (e.g. slope inflexions) of the appropriate curve, e.g. 5th nearest neighbour curve for choosing $d_{\max 5}$. Data courtesy of Dr. Tomoyuki Honma.

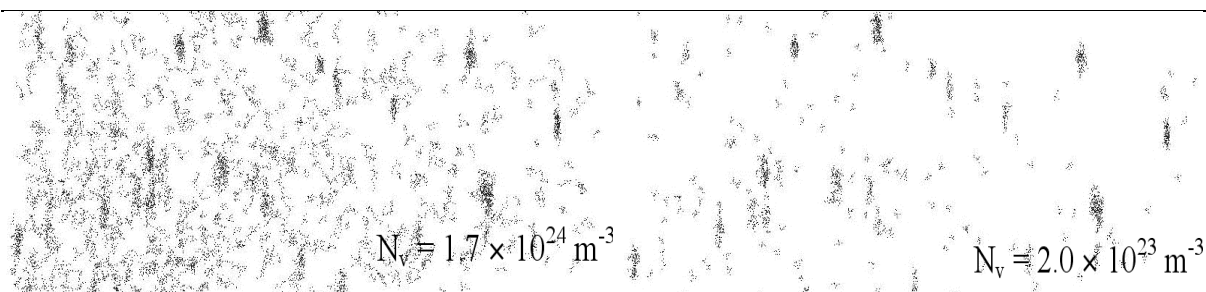


Fig. 2. Zn clustering in Al-2Zn-1.8Mg (aged for 60s at 150C) - the two 2D projections each represent the clustering found by Friends-of-friends ($d_{\max} = 0.7\text{nm}$ and $N_{\min} = 10$ atoms) and by DBScan ($d_{\max 5} = 0.7\text{nm}$ the maximum 5th nearest neighbour distance and $N_{\min} = 8$ atoms). For each case we calculate an estimated number density of clusters of size between 10 and 20 atoms, finding a disagreement by a factor of 10. Data courtesy of Peter V. Liddicoat.