182
doi:10.1017/S1431927621001264
Microsc. Microanal. 27 (Suppl 1), 2021

Inter-Experiment Machine Learning on APT experiments: New Insights from Meta-Analysis

Martin Meier[1], Paul Bagot[1], Michael Moody[2] and Daniel Haley[1]

[1]University of Oxford, England, United Kingdom, [2]University of Oxford, Oxford, England, United Kingdom

Atom Probe Tomography (APT) is a rapidly expanding field, with many new research groups being established. We estimate that the number of APT experiments that have been conducted worldwide now well exceeds 1 million.

Due to the nature of research using APT, individual researchers only analyse small numbers of samples which limits our ability for detailed statistical analysis. However, having access to a large collection of data allows for new avenues of research. Two components are needed to enable this class of research: large collections of APT experiments, stored in a format that makes the data accessible to analysis software, and tools that allow for the analysis of such data. Software that enables efficient and (semi-) automated analysis of APT data has previously been developed, and this work is ongoing [1-3].

Current data analysis methods are therefore limited by the lack of an effective and automatable method for building databases of APT data, particularly that are accessible without human intervention. To date, user input is still required, e.g. for calibrating ("fitting") mass-charge spectra or for setting parameters for spatial reconstruction algorithms. Furthermore, core APT data is often stored in proprietary file formats, hindering the development of alternative novel data processing software. This has largely suppressed the availability of computation in APT, and the subsequent insights and improved efficacy that machine learning can provide.

Nevertheless, we demonstrate how data from raw APT data can be used to build, and derive new insights form such an experimental database. Using a machine learning approach, we draw inferences from a large number of APT experiments. A collection of APT spectra, all of which use the same binning, can be regarded as a high dimensional point cloud, where each histogram bin is one dimension. Spectral clustering on this cloud is used to examine ionic fingerprints and enable meta-analysis of spectra, such as for hydrogen behaviours, by matching large numbers of similar experiments.

To showcase the potential usefulness of the spectrum point cloud approach in the analysis of statistical phenomena, we demonstrate a proof-of-concept study on hydrogen. Hydrogen is a common contaminant in APT datasets [4, 5], and its behaviour is strongly dependant on the sample material and experimental parameters, which we now quantify.

Our work promises to greatly reduce costly searches for analysis conditions for materials and enable metanalysis techniques that are common within fields such as computational drug design, and high energy physics.

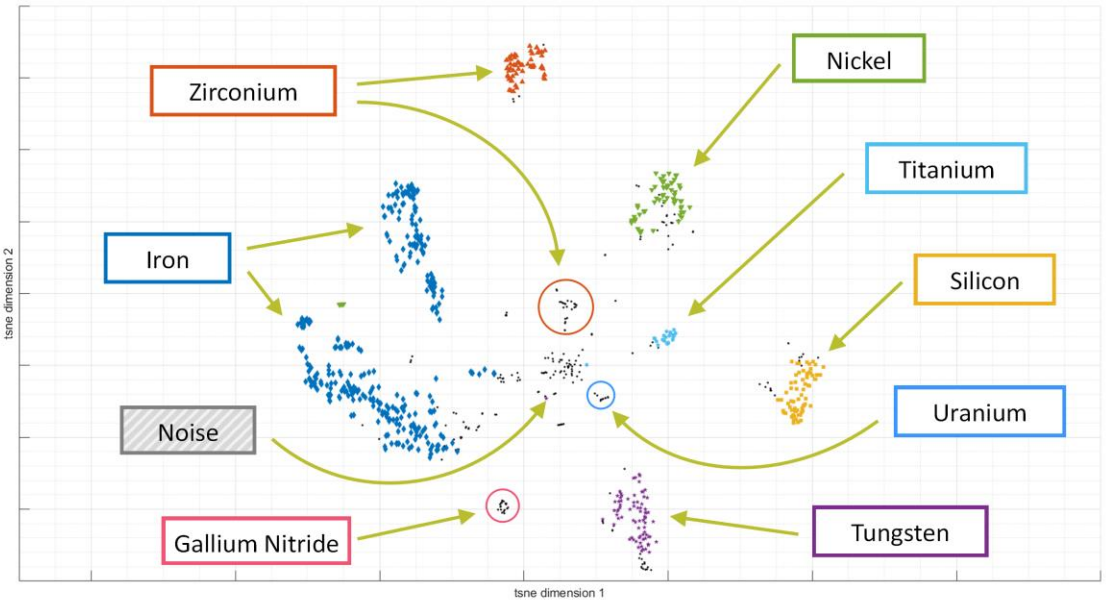

Figure 1. t-distributed stochastic neighbour embedding (tsne)6 of 848 APT spectra generated by Oxford APT group between 2013 and 2020. Clear clusters corresponding to different material types are visible.

References

1. Kühbach, M.; Bajaj, P.; Zhao, H.; Çelik, M. H.; Jägle, E. A.; Gault, B. *npj Computational Materials* **2021,** 7, (1), 21.
2. Haley, D.; Choi, P.; Raabe, D. *Ultramicroscopy* **2015,** 159, 338-345.
3. Madireddy, S.; Chung, D.-W.; Loeffler, T.; Sankaranarayanan, S. K. R. S.; Seidman, D. N.; Balaprakash, P.; Heinonen, O. *Scientific Reports* **2019,** 9, (1), 20140.
4. Kellogg, G. L. *The Journal of Chemical Physics* **1981,** 74, (2), 1479-1487.
5. Sundell, G.; Thuvander, M.; Andren, H. O. *Ultramicroscopy* **2013,** 132, 285-289.
6. van der Maaten, L.; Hinton, G. *J. Mach. Learn. Res.* **2008,** 9, 2579-2605.

https://doi.org/10.1017/S1431927621001264 Published online by Cambridge University Press