doi:10.1017/S1431927619010560
Microsc. Microanal. 25 (Suppl 2), 2019

The Role of Cluster Analysis in Automated Serial Electron Crystallography

Stef Smeets[1,2], Bin Wang[2] and Xiaodong Zou[2]

[1.] Kavli Institute of Nanoscience, Department of Bionanoscience, Delft University of Technology, Delft, The Netherlands.
[2.] Department of Materials and Environmental Chemistry, Stockholm University, Stockholm, Sweden.
* Corresponding author: s.smeets@tudelft.nl

Electron crystallography methods have now reached a level where high-quality data can be collected quickly and routinely [1]. 3D Electron diffraction data suitable for structure determination can be obtained from crystals as small as 50 nm. Our recent efforts have focused on the development of serial electron crystallography, which combines computer-controlled stage translation with beam shift to automatically collect diffraction data on a large number of crystals. For each stage position, an overview image is collected at a low magnification using a parallel beam, crystals are detected using image recognition techniques, and the beam is focused and shifted to each crystal (Figure 1). Then there are two options: (1) for a serial electron diffraction experiment (SerialED), a single, still diffraction pattern is collected for each crystal, which is useful for screening and quantitative phase analysis [2], [3]. In this way, up to 4000 crystals can be screened per hour. (2) Alternatively, in a serial rotation electron diffraction (SerialRED) experiment, a 3D data set is collected by continuously rotating the crystal in the beam while tracking the position of the crystal automatically. In this way, data can be collected on approximately 100 crystals per hour [4]. These data are more suitable for detailed structure analysis and phase identification.

Because these methods are fully automated in the software that we have developed [5], a large amount of data can now be generated. This is driving a need to develop new algorithms to deal with the ensemble, rather than the individual data points. We recently started exploring the use of hierarchical cluster analyses (HCA), which are now central to our data reduction pipeline. HCA in the context of multi-crystal diffraction experiments was first developed for high-throughput X-ray beamlines at synchrotrons in the realm of structural biology [6], [7], but we find them equally well suited for electron diffraction. Merging data from different crystals serves to improve data completeness and redundancy, which is necessary for precise determination of the crystal structure to a high resolution.

Two types of metrics are used for the HCA. First, lattice-based clustering [7], is used to group crystals with similar lattices, which is particularly well suited for multi-phase materials. Here, the distance metric is based on the volume or the lattice parameters. Second, reflection-based clustering [6] is used to find the best matching data sets, as not all crystals may diffract equally well. The HCA help to remove those outliers to improve the quality of the merged data. The distance metric is derived from the correlation coefficients of the common reflection intensities (CC_I) between pairs of data sets. The advantage HCA is that the results can be neatly visualized in a so-called dendrogram (Figure 2). A cut distance is then defined to control to which level the data sets should be grouped.

The HCA methods have been tested on several SerialRED data sets collected on a series of polycrystalline aluminosilicate samples: ZSM-5 ($Pnma$, a = 20.07 Å, b = 19.92 Å, c = 13.42 Å), a mixture of ZSM-5 and Mordenite ($Cmcm$, a = 18.256 Å, b = 20.534 Å, c = 7.542 Å), and PST-20 ($Im\bar{3}m$, a = 55.0664 Å) containing an impurity of ZSM-25 ($Im\bar{3}m$, a = 45.0711 Å) [8]. The crystals (100-500 nm in size) were randomly distributed on a grid, and data were collected using the SerialRED method. Using HCA allowed us isolate the phases and to select the optimal data sets which, when merged, allowed us to determine and

refine the crystal structures using standard crystallographic software up to 0.8 Å for ZSM-5 and Mordenite and 1.5 Å for PST-20 and ZSM-25. The atomic coordinates obtained from the merged data are consistent with those established in literature. It is worth noting that the refinement results on ZSM-5 in particular are virtually indistinguishable from those of the manually collected electron diffraction data based on some recent publications on the same material.

References:

[1] A Brown and J Clardy, Nature **564** (2018), p. 348.
[2] S Smeets, X Zou and W Wan, J. Appl. Crystallogr. **51** (2018), p. 1262.
[3] S Smeets, J Ångström and C-O A Olsson, Steel Res. Int. **90** (2019), p. 1800300.
[4] B Wang, X Zou and S Smeets, Submitted for publication (2019).
[5] S Smeets et al., Instamatic **1.0.** Zenodo (2018), https://dx.doi.org/10.5281/zenodo.1090388.
[6] R Giordano et al., Acta Cryst D **68** (2012), p. 649.
[7] J Foadi et al., Acta Cryst D **69** (2013), p. 1617.
[8] P Guo et al., Nature **524** (2015), p. 74.

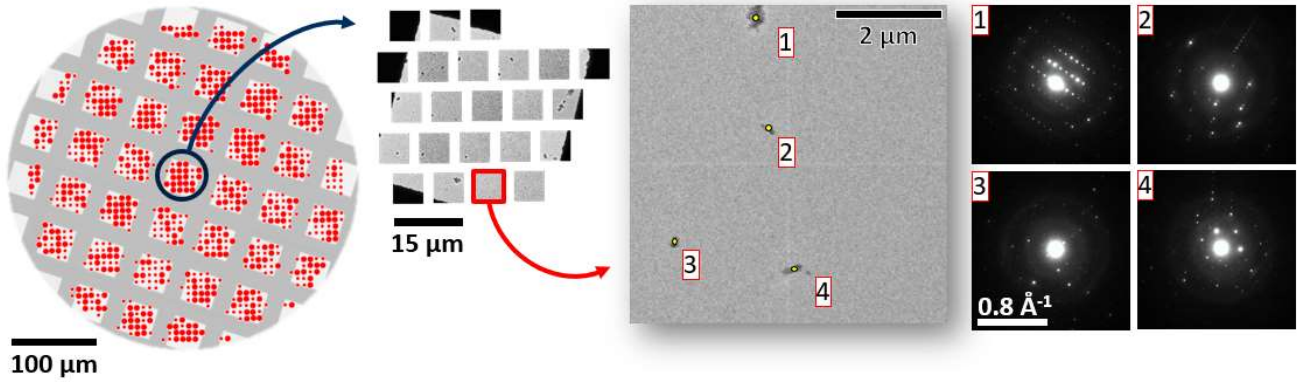

Figure 1. Overview of a serial electron crystallography experiment.

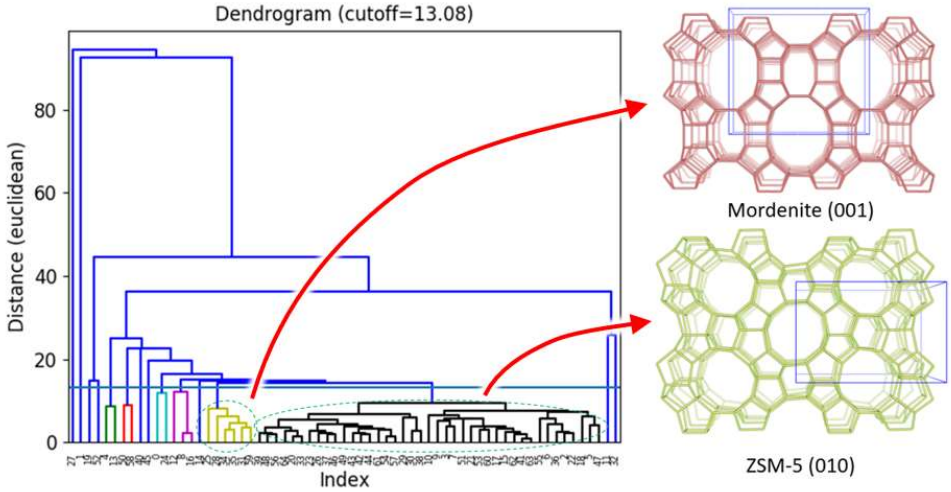

Figure 2. Dendrogram showing the lattice-based clustering for the ZSM-5 and Mordenite phase mixture. The black (41 data) and yellow (6 data) clusters used to determine their respective structures are indicated.