

Information theory-based plane symmetry classifications: revealing pseudo-symmetries in the presence of noise

Peter Moeck

Portland State University, Portland, Oregon, USA;

pmoeck@pdx.edu

Information theory-based methods [1-4] for the classification of more or less 2D periodic images from real-world crystals with atomic or molecular resolution into plane symmetry groups and projected Laue classes are described and applied to both synthetic patterns and experimental transmission electron microscope images of intrinsic membrane proteins. The synthetic patterns are per design either free of noise or contain approximately Gaussian distributed noise and are also highly pseudo-symmetric so that it is, for human beings on the basis of a visual inspection alone, very difficult to identify the underlying plane symmetry of the noisy patterns. This is because all genuine symmetries and pseudo-symmetries are broken by the added noise so that their differences are diminished. The new information theory-based classification methods, on the other hand, overcomes such challenges.

The methods enable the objective, i.e. researcher independent, identification of the plane symmetry group and 2D Laue class that provides the best possible separation of structure and generalized noise at the given noise level of a processed image. This identification enables the most meaningful averaging of the image in the spatial frequency domain in support of subsequent crystallographic analyses. This kind of averaging is over all correctly identified asymmetric units in the image and removes noise much more effectively than traditional Fourier filtering. Ratios of numerically obtained geometric Akaike Information Criterion values, i.e. first-order geometric-model-bias corrected sums of squared residuals between the complex-valued Fourier coefficients of the translation averaged image intensity and their counterparts from plane symmetry models of this data, are utilized for the crystallographic symmetry classifications.

Numerically obtained confidence levels are assigned to the classification of noisy images into minimal supergroups over their translationengleiche maximal subgroups. The resulting plane symmetry classifications are always generalized noise level dependent, which allows for better classification results in the future as noise decreases with future improvements to the imaging and image-processing procedures. The information theory-based methods deliver only probabilistic classifications as it is fundamentally unsound to assign an abstract mathematical concept, such as a plane symmetry group, with 100 % certainty to the record of a noisy real-world imaging experiment of an imperfect real-world crystal.

- [1] P. Moeck, *Symmetry* 10, paper 133 (46 pages) (2018), open access, DOI: 10.3390/sym10050133
- [2] P. Moeck, *IEEE Transactions on Nanotechnology* 18, 1166–1173 (2019), DOI: 10.1109/TNANO.2019.2946597, see also <http://arxiv.org/abs/1902.04155>, May 30, 2021 for an expanded version of this review
- [3] A. Dempsey and P. Moeck, Objective, Probabilistic, and Generalized Noise Level Dependent Classifications of sets of more or less 2D Periodic Images into Plane Symmetry Groups, <http://arxiv.org/abs/2009.08539>, December 15, 2020
- [4] P. Moeck and A. Dempsey, *Microscopy and Microanalysis* 25 (Suppl. 2), 1936–1937 (2019), DOI: 10.1017/S1431927619010419

Keywords: Crystallographic Image Processing, Information Theory