

Quantifying local coordination environment and structural similarity through order parameter-based site fingerprints and their application to machine learning

Nils E. R. Zimmermann,¹ Anubhav Jain¹

¹Energy Technology Area, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, U.S.A.

Structure characterization and classification is frequently based on coordination information of all or selected atomic sites in the crystal structure [1]. Therefore, reliable and robust procedures to find near(est) neighbors and to evaluate the resulting coordination pattern are critically important for machine learning applications that aim to exploit site or structure information for predicting materials properties. In this talk, we present new local structure order parameters (LoStOPs, q_i) [2, 3] that are specifically designed to rapidly detect rotationally invariant (e.g., $q_{\text{tetrahedron}}$, Figure 1) as well as asymmetric local coordination environments (e.g., $q_{\text{square_pyramid}}$) [4].

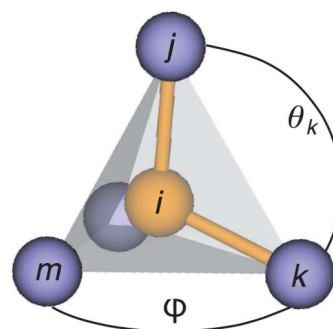


Figure 1: Tetrahedral order parameter [3]; figure taken from ref. 1.

We furthermore introduce an innovative optimization approach to ensure that the different LoStOPs are comparable with each other. We then apply the new local environment descriptors to define site fingerprints and measure similarity between structures. We showcase several application cases of these new tools, which are implemented in the open-source libraries pymatgen [5] and matminer [6]. They are, for example, useful as proxy variables for hydrothermal stability of zeolites [7], and the tools enable finding similar structures in a large materials database while, at the same time, carrying valuable coordination information via structure fingerprints [4]. In respect of the latter, we have computed all 2.45 billion structure similarity distances between each pair of the $\approx 70,000$ materials that are currently present in the Materials Project [8] database. Finally, we integrated the structure similarity data into the openly accessible Materials Project [8] website, thus, enabling easy exploration and discovery of related structures using our new innovative tools.

- [1] N. E. R. Zimmermann, M. K. Horton, A. Jain, M. Haranczyk, *Front. Mater* 4, 34, 2017.
- [2] B. Peters, *J. Chem. Phys.* 131, 244103, 2009.
- [3] N. E. R. Zimmermann, B. Vorselaars, D. Quigley, B. Peters, *J. Am. Chem. Soc.* 137, 13352-13361, 2015.
- [4] N. E. R. Zimmermann, A. Jain, in preparation, 2018.
- [5] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, G. Ceder, *Comput. Mater. Sci.* 68, 314-319, 2013; <https://github.com/materialsproject/pymatgen>.
- [6] L. Ward, A. Dunn, A. Faghaninia, N. E. R. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K. A. Persson, G. J. Snyder, I. Foster, A. Jain, in preparation, 2018; <https://github.com/hackingmaterials/MatMiner>.
- [7] N. E. R. Zimmermann, M. Haranczyk, *Cryst. Growth Des.* 16, 3043-3048, 2016.
- [8] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, *APL Mater* 1, 011002, 2013; <http://materialsproject.org>.