Data-Driven Materials Design

Janine George[1,2]
[1]BAM (Federal Institute for Materials Research and Testing), Materials Chemistry Department, Berlin, Germany
[2]Friedrich Schiller University Jena, IFTO, Jena, Germany.

janine.george@bam.de

Bonds and local atomic environments are crucial descriptors of material properties. They have been used to create design rules and heuristics, as descriptors in machine-learned interatomic potentials and general machine learning of material properties.[1]

Implementations and algorithms, such as ChemEnv and LobsterEnv, for identifying local atomic environments based on geometrical characteristics and quantum-chemical bonding analysis are now available.[2,3] Fully automatic workflows and analysis tools have been developed to use quantum-chemical bonding analysis on a large scale.[3-5] The first part of the lecture will demonstrate how our tools, which assess local atomic environments and perform automatic bonding analysis, help develop new machine-learning models and a new intuitive understanding of material properties.[6]

Many new universal machine-learned interatomic potentials, such as MACE-MP-0, have been developed.[7] The second part of the lecture will showcase how these potentials, in combination with DFT, could significantly accelerate our research. The focus will be on the interplay of DFT and machine-learned interatomic potentials, presenting new fully automated workflows for training, fine-tuning, and benchmarking these potentials, implemented in our software *autoplex* (https://github.com/autoatml/autoplex). Additionally, I will show how to train new interatomic potentials from scratch by exploring potential energy surfaces extensively, offering a method to enhance current universal machine-learned potentials.[8]

Beyond this, the general trend toward automation in computational materials science and some of our recent contributions will be discussed.[5,9]

References

[1] George, J. & Hautier, G. Chemist versus Machine: Traditional Knowledge versus Machine Learning Techniques. Trends Chem. 3, 86–95 (2021).
[2] Waroquiers, D. et al. ChemEnv: a fast and robust coordination environment identification tool. Acta Cryst B 76, 683–695 (2020).
[3] George, J. et al. Automated Bonding Analysis with Crystal Orbital Hamilton Populations. ChemPlusChem 87, e202200123 (2022).
[4] Naik, A. A., Ueltzen, K., Ertural, C., Jackson, A. J. & George, J. LobsterPy: A package to automatically analyze LOBSTER runs. JOSS 9, 6286 (2024).
[5] Ganose, A. M. et al. Atomate2: Modular workflows for materials science. ChemRxiv (2025) doi:10.26434/chemrxiv-2025-tcr5h.
[6] Naik, A. A., Ertural, C., Dhamrait, N., Benner, P. & George, J. A Quantum-Chemical Bonding Database for Solid-State Materials. Sci Data 10, 610 (2023).
[7] Batatia, I. et al. A foundation model for atomistic materials chemistry. Preprint at https://doi.org/10.48550/arXiv.2401.00096 (2024).
[8] Liu, Y. et al. An automated framework for exploring and learning potential-energy surfaces. Preprint at https://doi.org/10.48550/arXiv.2412.16736 (2024).
[9] Rosen, A. S. et al. Jobflow: Computational Workflows Made Simple. JOSS 9, 5995 (2024).