Challenges for automated materials discovery

Pere Miro,[a] Damien E. Coupry[b] and Thomas Heine[b]

[a] *Department of Chemistry, University of South Dakota, USA*
[b] *Faculty of Chemistry and Food Chemistry, TU Dresden, Germany*
thomas.heine@tu-dresden.de

This talk focuses on bottlenecks that we have identified in data-driven materials prediction.

In the first part, we describe a data-driven exporation of 2D materials, which was launched 5 years before the first serious fully automated high-throughput efforts and came up with some remarkable new discoveries that have been realized in experiment shortly afterwards, most notably noble metal dichalcogenides. The data repositories are poorly digitized and the striking question is: how can we extract information of some 100 years of laborous experimental work to make it useful for materials discovery?

In the second part I will focus on metal-organic frameworks (MOFs), the most prominent class of molecular-based framework materials. Built out of well-defined building blocks, machine-assisted data discovery appears to be straight-forward. We will show the components of our machinery, that allow to design a virtually infinite number of MOFs and other framework materials using a library-based building block approach. For property prediction, we suggest an alternative for structure encoding and decoding. We test the approach for predicting properties such as geometrical features, gas uptake and band gaps.