# Patterns, Correlations, and Causality in Big Data of Materials: Analytics for Novel Materials Discovery

Matthias Scheffler [(*)]
Fritz Haber Institute of the Max Planck Society, Berlin, Germany
http://th.fhi-berlin.mpg.de/

***Big data of materials are structured*** in a way that is typically not visible by standard tools. Furthermore, with respect to a certain (desired) property, the "chemical compound space", which contains a practically infinite amount of different materials, is very sparsely populated. Thus, a key issue in data-driven materials science is to find the proper descriptive parameters (descriptors) that identify the materials-property related structures of this huge space.

We will show that and how compressed sensing, originally designed for representing a complex signal in the lowest possible dimensionality, can select, out of a huge-dimensional space of potential descriptors (features), a low dimensional descriptor. Examples are crystal-structure and stability prediction and the prediction of the band gap of binary and ternary compounds.

By applying sensitivity analysis, supervised pattern discovery, and causal inference techniques, we discuss the causal relationship between the selected descriptors and the predicted physical properties.

We will also address how the recently establishes *NOMAD (novel materials discovery) Laboratory*, a European Center of Excellence (http://NOMAD-CoE.eu), will address these issues.

(*)In collaboration with Luca Ghiringhelli, Claudia Draxl, Mario Boley, Niklas Menzel, and others.