A Generalized Convex Hull Construction for Materials Discovery

A. Anelli,[a] E. A. Engel[a*], C. J. Pickard,[b,c] and Michele Ceriotti[a]

a. Laboratory of Computational Science and Modeling, IMX, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

b. Department of Materials Science and Metallurgy, 27 Charles Babbage Road, Cambridge CB3 0FS, UK

c. Advanced Institute for Materials Research, Tohoku University, 2-1-1 Katahira, Aoba, Sendai 980-8577, Japan

* edgar.engel@epfl.ch

Searching for novel materials involves identifying potential candidates and selecting those that have desirable properties and facile synthesis. The identification of synthesizable compounds is a needle-in-a-haystack problem, best exemplified by the case of high-throughput computational structure searches. The typical size of the databases compiled in such searches generally prohibits bulk determination of accurate structural properties and viable synthetic pathways for all locally-stable structures generated. This renders the identification of synthesizable compounds and the relevant experimentally-realizable constraints a crucial step towards experimental realization of novel materials.

Conventionally, the screening is based on a convex hull constructions, which identify structures that (in the absence of kinetic effects) can be stabilized by manipulating a particular thermodynamic constraint (such as pressure or composition) chosen on the basis of experimental evidence or intuition. This is neither agnostic, nor capable of identifying structures stabilized by more complex sets of constraints. We therefore introduce a generalized convex hull (GCH) framework based on an abstract representation of structural similarity. The GCH is constructed on data-driven coordinates, such that the GCH represents the full structural diversity of the candidate compounds in an unbiased way. Moreover, we rigorously account for the inevitable uncertainty in input structures data, which renders the GCH probabilistic in nature. Not only is the resultant framework robust with respect to errors in the input data, it also facilitates automatic removal of redundant structures. It moreover provides a measure of the probability that a given structure is experimentally-realizable. Compared to the input (free) energies this represents a superior measure of stability, which is computed at negligibe computational cost and can for instance also be used to assist experimental crystal structure determination. The GCH framework efficiently identifies candidates with high probabilities of being synthesizable and suggests the relevant experimentally-realizable constraints, thereby providing a much needed starting point for the determination of viable synthetic pathways. We demonstate that our framework allows us to single out structures that can be stabilized by general thermodynamic constraints, ranging from pressure to the substitution of portions of organic compounds.