Crystal Structure Prediction

We develop and implement methods that identify correlations and structure in big data of materials. This will enable scientists and engineers to decide which materials are useful for specific applications or which new materials should be the focus of future studies.

The first step to enable this is to make the data available for analysis. Then we want to provide interactive tools to find trends and anomalies to discover novel materials. Here the NOMAD knowledge in scaling big data analysis enriches the expertise of the NOMAD Center of Excellence.

The preparation, synthesis, and characterization of new materials is a complex and costly aspect of materials design. About 200,000 materials are “known” to exist, but the basic properties (e.g., optical gap, elasticity constants, plasticity, piezoelectric tensors, conductivity, etc.) have been determined for very few of them. Connecting organic and inorganic materials, surfaces, interfaces, and nanostructures, as well as inorganic/gas-solids, the number of possible materials is practically infinite. It is therefore highly likely that new materials with superior (but currently unknown) properties exist but different yet unknown, so that new materials with exceptional properties can be exploited for a number of widespread fields such as energy storage and transformation, mobility, safety, information, and health.

Despite a huge number of possible materials, we note that “the chemical compound space” is sparsely populated when the focus is on selected properties or functions. Our aim is to develop big-data analytics tools that will help to sort all the available material data to identify trends and anomalies.

To do so, we have embarked on a number of initiatives. Several of these are based on the use of linear and non-linear low-dimensional embedding methods, proper symmetries of a well-behaved physical representation and makes the similarity-recognition algorithm, based on descriptors that encode the same version on the same data should give the same result.

A web-based implementation (via notebooks) of a data-analysis tool for the investigation of the similarity among crystal structures and for the prediction of the difference in formation-energy among them. The tool performs the data for the analysis by a fine-tuned query in the NOMAD-Archive that contains several millions of crystal configurations.

The similarity-recognition algorithm, based on descriptors that encode the crystal structure, enables the identification of similar structures and makes use of linear and non-linear low-dimensional embedding methods, produces a 2-dimensional map that assigns to separate regions perfect and distorted configurations, for given pairs of crystal structures. The algorithm predicts the difference in formation energies selects the models out of thousands of candidates, by means of a compressed-sensing model out of thousands of candidates, by means of a compressed-sensing method. The algorithm predicting the difference in formation energies selects the model out of thousands of candidates, by means of a compressed-sensing method.