Rational Materials Design via Machine Learning

Trial-and-error research approaches are increasingly ill equipped to meeting the complex challenges involved in the discovery and design of next-generation materials and chemistry. Our work recognizes the great opportunities that are arising with the shift towards data-driven in silico research and a rational design paradigm. The notion to utilize modern data science – the 4th pillar of science – in the materials context is so recent that much of the basic infrastructure has not yet been developed, or is still in its infancy. The existing tools and expertise tend to be in-house, specialized, or otherwise unavailable to the community at large. Data science is thus in practice beyond the scope and reach of most researchers in the field.

Our work focuses on creating general-purpose tools designed to overcoming this situation, filling the prevalent infrastructure gap, and thus making data-driven research a viable and widely accessible proposition. Our work fuses materials modeling, high-throughput screening, and Big Data analytics into an integrated research infrastructure. It facilitates the large-scale exploration of chemical space, whose uncharted domains are expected to hold new classes of materials, compounds, and reactions with transformative properties. It employs informatics and machine learning to mine the resulting data sets. The goal is to develop an understanding of the hidden structure-property relationships that govern the behavior of chemical and materials systems. These insights are a prerequisite for rational design and inverse engineering capability as outlined in the White House Materials Genome Initiative."