Final Technical Report

DOE Award DE-FG02-96ER45571
Massachusetts Institute of Technology

First Principles Prediction of Structure, Structure Selectivity, and Thermodynamic Stability under Realistic Conditions

PI: Professor Gerbrand Ceder, Department of Materials and Engineering
Massachusetts Institute of Technology. gceder@mit.edu

01/28/2018

Program Objectives and Overview of Accomplishments

The objectives of this program were to develop first-principles theory to predict the structure and thermodynamic stability of materials. Since its inception the program focused on the development of the cluster expansion to deal with the increased complexity of complex oxides. This research led to the incorporation of vibrational degrees of freedom in ab-initio thermodynamics, developed methods for multi-component cluster expansions, included the explicit configurational degrees of freedom of localized electrons, developed the formalism for stability in aqueous environments, and culminated in the first ever approach to produce exact ground state predictions of the cluster expansion. Many of these methods have been disseminated to the larger theory community through the Materials Project, *pymatgen* software, or individual codes. We summarize three of the main accomplishments below.

Accomplishments

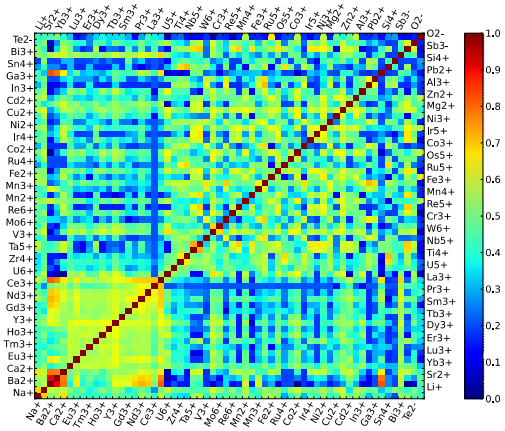

Figure 1: *Ionic similarity* indicates how likely two ions substitute for each other, thereby retaining the same crystal structure

Data-Mining Algorithms: A distance semi-metric in compound space

In this program we have rigorously developed data mining and machine learning algorithms as a demonstration that some physical properties can be "learned" solely from the basis of data. In 2006 we demonstrated on ef the first applications of data mining by machine learning the prediction of crystal structures [Nature Materials 2016]. We also developed a machine-learned "substation predictor" in which machine learning was used to predict which elements can easily substitute for each other while retaining the crystal structure. More recently, we have developed more formal theory to facilitate machine learning by developing distance metrics between crystal structures. Machine learning of materials behavior requires the use of feature vectors or descriptors that capture the

essential compositional or structural information that is most likely to influence a property. While humans tend to use intuitive –and imprecise – ideas to evaluate whether two chemistries or crystal structures are the same, machine learning requires quantitative metrics to evaluate "similarity". Rigorous structure descriptions including concepts such as unit cells and symmetry, but chemists and materials scientists often use more intuitive and less well-defined terminology such as local environment, coordination, and polyhedral connectivity, to explain the properties of a compound. Indeed, unit cells and symmetry vary discontinuously with small changes in the atomic coordinates, and therefore do not provide a good set of descriptors in which to understand or expand materials properties. We have developed a mathematical description of structure and chemistry that can be used to create a distance semi-metric between compounds. This allows us to rigorously define *similarity* between compounds, and we expect this formalism to form the basis for machine learning approaches in materials science.

Our novel semi-metric captures both topological and chemical information of the local environment around atoms (i.e. an octahedral site formed by oxygen ions is different from one formed by metal atoms). We combine a distance metric in chemical space (the *ionic similarity*) and in topological space (the *coordination similarity*), and coarse-grain the environment observed in known compounds (e.g. as documented in the ICSD database.

Figure 1 shows an *ionic similarity* metric. The matrix indicates how likely two elements are to substitute for each other in a given crystal structure. We have integrated this *ionic similarity* with a *topological similarity* between two atomic environments, e_i and e_j, obtained by creating a weighted Voronoi polyhedron as defined by O'Keefe.[i] From this construction, every ion i in a crystal structure X has a set of associated neighbors $\{n_i, w_i\}$, where n_i refers to a neighboring ion and w_i is its associated weight. A similarity score can then be defined as:

$$\text{score}\left(n_i \in e_i, n_j \in e_j\right) = \text{sim}_{\text{ion}}\left(n_i, n_j\right)\min\left(w_i, w_j\right)e^{-\frac{\left(w_i - w_j\right)^2}{2c^2}}.$$

The variable c tunes differences in topological or geometric similarity. The product of two atomic environments e_i and e_j, can then be defined via a one to one weighted bipartite graph matching between the vertices of $\{n_i\}$ to the vertices $\{n_j\}$. Each vertex in $\{n_i\}$ is matched to the vertex in $\{n_j\}$ that maximizes the sum of the similarity scores. If one neighborhood has more ions than the other, ions to complete the matching are inserted with similarities equal to zero:

$$\text{prod}\left(e_i, e_j\right) = \max_{\text{matchings}} \sum_{n_i, n_j \in \text{matchings}} \text{score}\left(n_i, n_j\right).$$

Finally, we define the similarity between two local environments:

$$\text{sim}_{\text{env}}\left(e_i, e_j\right) = \frac{\text{prod}\left(e_i, e_j\right)}{\sqrt{\text{prod}\left(e_i, e_i\right)\text{prod}\left(e_j, e_j\right)}}.$$

Figure 2 shows how our new distance measure can be used to construct a dendrogram, which groups compounds by similarity.

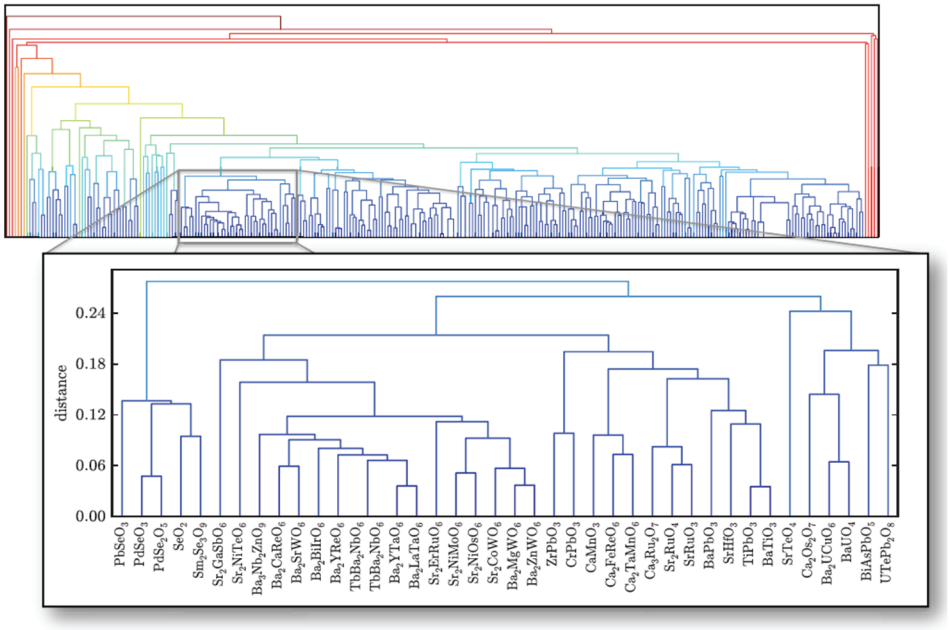

Figure 2: Dendogram showing which compounds are *similar* to each other

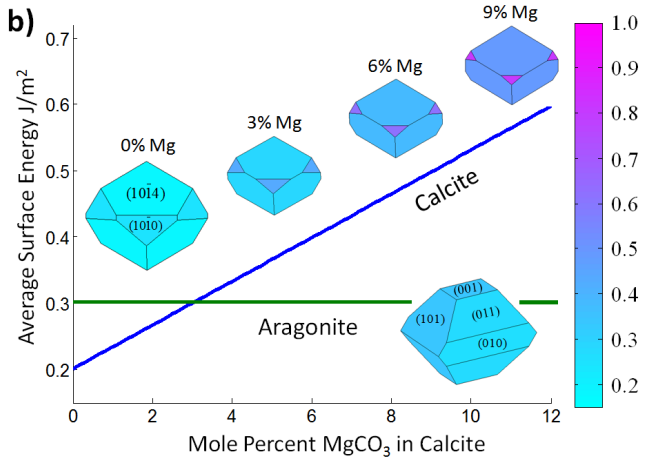

Figure 3: Surface energy of calcite and aragonite as function of Mg content

Size-dependent thermodynamics and nucleation preferences

Predicting which phases can form and under which conditions they can form is one of the great unsolved problems of materials science, geology, and solid state chemistry. While it is amply clear that the lowest energy state for a given composition can easily form, no guidance exists as to how and when metastable polymorphs form instead. Solving this problem could lead to predictive and rational materials synthesis, and would significantly enhance the reach of computational materials development.

Metastable phases can form when they nucleate preferentially over the ground state phase. Predicting nucleation energies requires knowledge of relevant surface energies in the medium in which the phases form. In a first attempt to rationalize metastable phase formation we have studied the nucleation selection of the calcite or aragonite form of $CaCO_3$. This is an important problem in nature. Although calcite is the equilibrium phase under ambient conditions, metastable aragonite is the precipitated polymorph in modern seawater. Using *ab initio* density functional theory calculations, we demonstrated that in marine environments, calcite spontaneously incorporates Mg^{2+} in solid-solution, which increases its surface energy, and thus nucleation barrier, so that it greatly exceeds that of aragonite. Figure 3 shows the surface energy of both

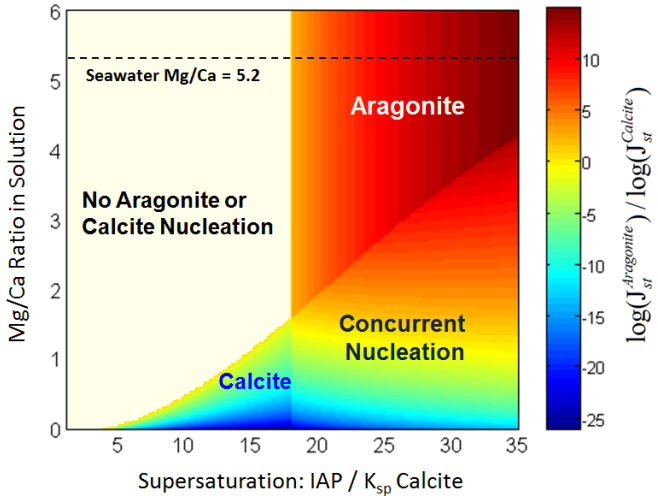

Figure 4: Nucleation preference as function of Mg/Ca ratio in solution and supersaturation

calcite and aragonite as a function of Mg concentration in the water the $CaCO_3$ precipitates from. Figure 4 shows the steady-state nucleation rates as function of supersaturation and Mg/Ca ratio in solution, clearly showing that aragonite is the preferred nucleation polymorph under high Mg/Ca ratios, including those that are found in modern seawater. More recently, we have applied similar theory to predict the pathways that MnO_x polymorphs take when they form in aqueous solution.

Exact Ground state prediction of lattice model via Cluster expansion, Maximum Satisfiability and Convex Optimization

Lattice models have wide applicability in science, and are commonly used in a wide range of applications, such as magnetism, alloy thermodynamics, fluid dynamics, phase transitions in oxides, and thermal conductivity. In first-principles thermodynamics, lattice models take on a particularly important role as they appear naturally through a course graining of the partition function of systems with substitutional degrees of freedom. As such, they have been used to predict the structure and phase diagrams of crystalline solids from ab-initio calculations. However, the procedure to find the exact ground state of a lattice model, also referred to as a generalized Ising model or cluster expansion, defined on an arbitrary lattice, with any interaction range and number of species remains an unsolved problem, with only a limited number of special-case solutions known in the literature. Thus, in light of the wide applicability and success of the generalized Ising model, an efficient approach to finding the true ground state of such a Hamiltonian would give significant insight into the behavior of these models, and facilitate their use in ab-initio alloy theory.

In this project we derive an algorithm to find exact ground states of lattice models. The algorithm not only finds the ground state but also proves that it is an absolute minimum. Combinatorial optimization (MAX-SAT) and non-smooth convex optimization (MAX-MIN) are combined to provide upper and lower bounds respectively on the ground state energy. By systematically converging upper and lower bounds to each other, we find and prove the exact ground state for realistic Hamiltonians whose solutions are otherwise intractable via traditional methods. The underlying mathematical relationship between the exact ground state problem and tiling suggests that this approach can also be useful in identifying aperiodic ground states.

There are two key elements in this problem. The first element is to efficiently find the ground state given a fixed periodicity of the solution. Calculating this periodic ground state is equivalent to solving the finite optimization problem:

$$\min_{\bar{s}} \sum_{\alpha \in \tilde{\mathbf{C}}} J_\alpha \prod_{(x,y,z,p,t)\in\alpha} \bar{s}_{x,y,z,p,t}$$

subject to:

$$\sum_{t \in \mathbf{c}(p)} \overline{s}_{x,y,z,p,t} = 1 \; \forall (x,y,z,p) \in \mathbf{F}_{finite}$$

where J is the effective cluster interactions parameters, $\overline{s}_{x,y,z,p,t}$ is the indicator variable of specie t on site (x,y,z,p), $\tilde{\mathbf{C}}$ is the set of all interacting clusters within the fixed periodic system and $\mathbf{F}_{finite}$ is the set of sites within the fixed periodic unit cell. Such an optimization over discrete $\{0,1\}$ variables can be equivalently posed as a logic problem, replacing the discrete variables by Boolean analogs. Following this insight, the finite Hamiltonian is exactly in the form of pseudo-Boolean optimization (PBO), allowing us to solve this optimization problem using weighted partial maximum satisfiability (MAX-SAT). The essence of MAX-SAT is to model the discrete optimization problem by maximizing the number of logical clauses that can be satisfied in a Boolean formula of conjunctive normal form.

The second element of our algorithm is the optimization of a lower bound to the ground state energy. We prove that minimization of the Hamiltonian on a finite group of sites without any periodicity constraints provides a lower bound for the ground state energy. For example:

$$H = \lim_{N \to \infty} \frac{1}{(2N+1)} \sum_{(i) \in \{-N,\ldots,N\}^3} \left(J_0 s_i + J_1 s_i s_{i+1} + J_2 s_i s_{i+2} \right) \geq \min_{s_0, s_1, s_2} \left(J_0 s_0 + J_1 s_0 s_1 + J_2 s_0 s_2 \right)$$

The lower bound introduced directly in this way is very loose. To improve it, we introduced the notion of a translational equivalent ECI's parametrized by λ by which leaves the global Hamiltonian unchanged while improving the lower bound. Formally, we can maximize the lower bound energy over the free parameters λ to obtain the tightest lower bound on the ground state energy:

$$H \geq \max_{\lambda} \min_{\mathbf{s} \in \{0,1\}^{\mathbf{B}}} E_{\lambda,\mathbf{s}}$$

To test the performance of this approach on practically relevant systems, we measure the runtime of this algorithm on binary 1D, 2D square, and 3D cubic lattices over random sets of asymmetric ECIs across a spectrum of interaction ranges. First, we restrict ourselves to only pair interactions, calculating runtimes for up to 28 pair interactions on unit cells up to 50 sites, where the energy of each interaction takes on a random value. In the 1D, 2D and 3D cases, this limit corresponds to all interactions up to and including the 28th, 10th and 5th nearest neighbors respectively. In all cases, our implementation gives a very promising single-core runtime on the order of hours for realistic Hamiltonians, which typically include fewer than 100 interactions.

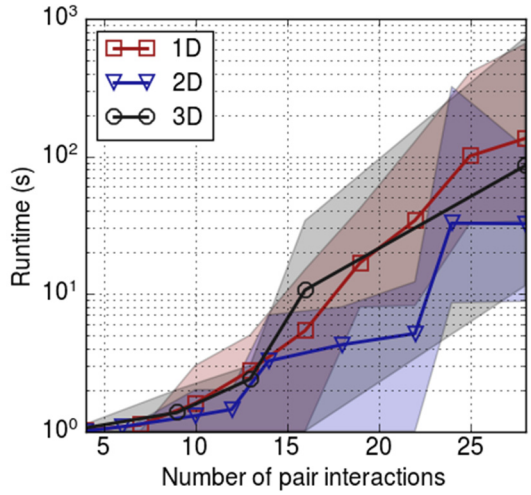

Figure 5: Single-core computation time needed to find and prove the ground state of a 1D, 2D, and 3D pair-interaction Hamiltonian for unit cells up to 50 sites in size across an increasing range of pair-interactions. In all cases, the solver finds the ground state for all unit cells up to 50 atoms in size, and calculates a tight lower bound on the true ground state energy without enlarging $|\mathbf{B}|$. Each point corresponds to the geometric average runtime of 100 such calculations with random interaction coefficients, while the shading gives the spread between the 20th and 80th percentiles.

We have also applied our method to obtain the exact ground state of a cluster expansion Hamiltonian used to model sodium-vacancy orderings in the layered Na_xNiO_2 compound as a function of composition. The J interactions for this system are determined from DFT calculations on 400 structures, through standard approaches. Our algorithm finds new ground states at x=2/5, 1/2, and 3/5, compared to the DFT input structures initially used to derive the cluster expansion. The advantage of our approach is that we can guarantee that there are no other configurations of any unit cell size that are lower in energy. The inset of **Figure** shows the ground state predicted at x=1/2 which is unusual and unlikely to be proposed from intuition.

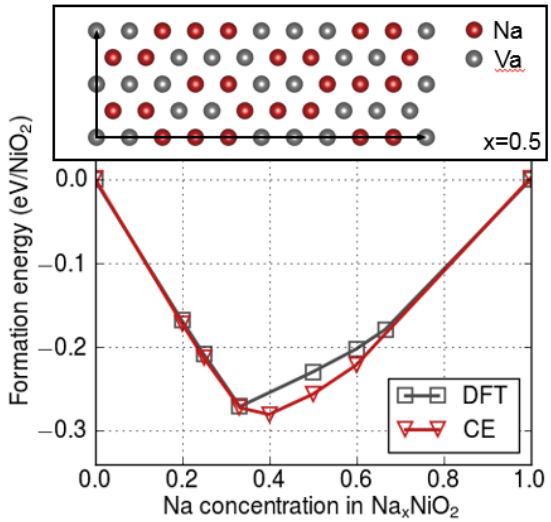

Figure 6: Enumeration of ground states found for a cluster expansion of sodium-vacancy orderings in layered Na_xNiO_2. The red triangles indicate the mathematically proven ground states, whereas the gray squares are the originally proposed ground states from DFT calculations of 400 possible Na-vacancy arrangements. The ground state configuration for x=1/2 is shown in the inset.

To conclude, we have introduced a MAX-MIN procedure to obtain the exact ground state of a generalized Ising model. Our procedure relies on converging an upper and lower bound on the ground state energy, where the upper bound is obtained from MAX-SAT finite optimization, while the lower bound is given by convex optimization over translationally-equivalent clusters. Mathematically, our approach relies on the tilability of minimum-energy local configurations to generate the exact global ground state. In practice, this procedure performs very well and has made it possible to determine the exact ground state of many formerly intractable systems. Finally, we envision that the MAX-MIN procedure introduced here can serve as a tractable approach for resolving ground states on aperiodic lattices.

Papers for this grant period (2013 – Current)

1. S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K. Persson, G. Ceder, Python Materials Genomics (Pymatgen): A Robust, Open-source Python Library for Materials Analysis, Computational Materials Science, 68, 314-319 (2013). *This work was supported in part by the Department of Energy's Basic Energy Sciences program under Grant No. DE-FG02-96ER45571.*
2. Y. Wu, P. Lazic, G. Hautier, K. Persson, G. Ceder, First Principles High Throughput Screening of Oxynitrides for Water-Splitting Photocatalysts, Energy & Environmental Science 6 (1), 157-168 (2013). *Some methodological work has been supported by the Department of Energy under Grant No. DE-FG02-96ER4551.*
3. A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation, APL Materials 1, 011002, 1-11 (2013). *Work at MIT on an early version of the Materials Project was supported by the Department of Energy under Grant No. DE-FG02-96ER45571.*
4. W. Sun, G. Ceder, Efficient Creation and Convergence of Surface Slabs, Surface Science, 617, 53-59 (2013). *This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under Grant No. DE-FG02-96ER45571.*
5. Y. Wu, G. Ceder, First Principles Study on Ta3N5:Ti3O3N2 Solid Solution as a Water-Splitting Photocatalyst, Journal of Physical Chemistry, 117 (47), 24710-24715 (2013). *Some methodological work has been supported by the Department of Energy under Grant No. DE-FG02-96ER45571.*
6. L. Yang, G. Ceder, Data-Mined Similarity Function between Material Compositions, Physical Review B, 88 (22), 224107 (2013). *This work was supported in part by the US Department of Energy under Grant No. DE-FG02-96ER45571.*
7. S.Y. Kang, Y.F. Mo, S.P. Ong, G. Ceder, Nanoscale Stabilization of Sodium Oxides: Implications for Na-O-2 Batteries, Nano Letters, 14 (2), 1016-1020 (2014). *This work was partially supported by DOE-Chicago under Grant No. DE-FG02-96ER45571.*
8. Y. Mo, S.P. Ong, G. Ceder, Insights into Diffusion Mechanisms in P2 Layered Oxide Materials by First-Principles Calculations, Chemistry of Materials, 26 (18), 5208-5214 (2014). *This work was partially supported by the U.S. Department of Energy (DOE) under Grant No. DE-FG02-96ER45571.*
9. W. Sun, S. Jayaraman, W. Chen, K. Persson, G. Ceder, Nucleation of Metastable Aragonite CaCO3 in Seawater, PNAS, 112 (11), 3199-3204 (2015). *This work was supported by the US Department of Energy, Office of Basic Energy Sciences, under Grant No. DE-FG02-96ER45571, and the National Science Foundation Graduate Research Fellowship (to W.S.).*
10. W. Huang, D. Kitchaev, S. Dacek, Z. Rong, A. Urban, S. Cao, C. Luo, G. Ceder, Finding the Ground State of a Generalized Ising Model by Convex Optimization and MAX-SAT, Physical Review B, 94, 13, 134424 (2016). *This work was supported primarily by the U.S. Department of Energy (DOE) under Grant No. DE-FG02-96ER45571.*
11. W. Huang, A. Urban, Z. Rong, Ziwhei Ding, S. Cao and G. Ceder, Construction of Ground-State Preserving Sparse Lattice Models for Predictive Materials Simulations, NPJ Computational Materials, 3(1) 30 (2017). *This work was supported primarily by the U.S. Department of Energy (DOE) under Grant No. DE-FG02-96ER45571.*

12. A. Urban, D-W Seo and G. Ceder, Computational Understanding of Li-ion Batteries, npj Computational Materials, 2, 16002 (2016). *This work was supported primarily by the U.S. Department of Energy (DOE) under Grant No. DE-FG02-96ER45571.*

13. H. Das, A. Urban, W. Huang, G. Ceder, First-Principles Simulation of the (Li-Ni-Vacancy)O Phase Diagram and Its Relevance for the Surface Phases in Ni-Rich Li-Ion Cathode Materials, Chemistry of Materials, 29, 18, 7840-7851 (2017). *W.H. also thanks the U.S. Department of Energy (DOE) for support under Grant No. DE-FG02-96ER45571*.