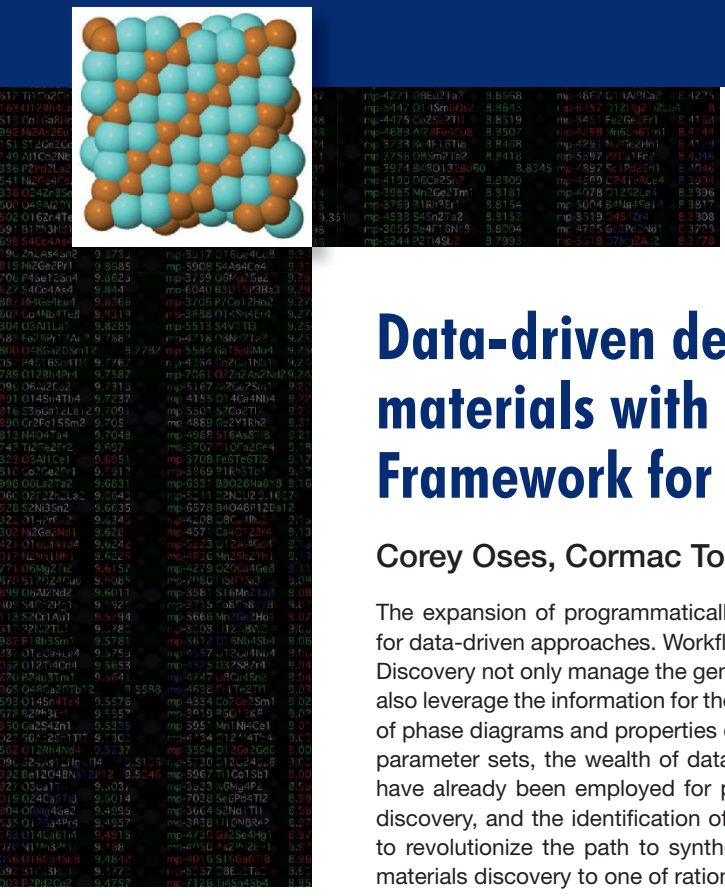


Data-driven design of inorganic materials with the Automatic Flow Framework for Materials Discovery

Corey Oses, Cormac Toher, and Stefano Curtarolo

The expansion of programmatically accessible materials data has cultivated opportunities for data-driven approaches. Workflows such as the Automatic Flow Framework for Materials Discovery not only manage the generation, storage, and dissemination of materials data, but also leverage the information for thermodynamic formability modeling, such as the prediction of phase diagrams and properties of disordered materials. In combination with standardized parameter sets, the wealth of data is ideal for training machine-learning algorithms, which have already been employed for property prediction, descriptor development, design rule discovery, and the identification of candidate functional materials. These methods promise to revolutionize the path to synthesis, and ultimately transform the practice of traditional materials discovery to one of rational and autonomous materials design.

Introduction

Density functional theory (DFT) implementations[1–7] offer a reasonable compromise between cost and accuracy in *ab initio* materials science calculations,[8] stimulating rapid development of automated frameworks and corresponding data repositories. Prominent examples include the Automatic Flow Framework for Materials Discovery (AFLOW),[9–12] Novel Materials Discovery Laboratory (NOMAD),[13] Materials Project,[14] Open Quantum Materials Database (OQMD),[15] Computational Materials Repository and its associated scripting interface Atomic Simulation Environment (ASE),[16] and Automated Interactive Infrastructure and Database for Computational Science (AiiDA).[17] Such repositories house an abundance of materials data. For instance, the AFLOW.org database contains more than 1.8 million compounds, each characterized by about 100 different properties.[11,18–20] Investigations employing this data have not only led to advancements in modeling electronics,[21–24] thermoelectrics,[25,26] superalloys,[27] and metallic glasses,[28] but also the synthesis of new materials: for example, two new magnets, Co_2MnTi and Mn_2PtPd, which are the first discovered by computational approaches.[29]

Further advancements and discoveries are contingent on continued development and expansion of these materials repositories. New entries are generated both by (1) calculating the properties of previously observed compounds from sources such as the Inorganic Crystal Structure Database,[30] and (2) decorating structural prototypes[31]—populating crystal sites of existing structures with atoms of different elements—to predict new materials. Accurate computation of materials properties—including electronic, magnetic, chemical, crystallographic, thermomechanical, and thermodynamic features—demands a combination of reliable calculation parameters/thresholds[11] and robust algorithms that scale with the size/diversity of the database. For example, convenient definitions for the primitive cell representation[9] and high-symmetry Brillouin zone path[10] have optimized and standardized electronic structure calculations. Careful treatment of spatial tolerance and proper validation schemes have finally enabled accurate and fully autonomous determination of the complete symmetry profile of crystals,[32] which is essential for elasticity[33] and phonon[9,34–36] calculations.

Beyond descriptions of simple crystals, exploration of complex properties[33,37] and materials[28,38] typically warrants advanced (and expensive) characterization techniques.[39–41] Fortunately, state-of-the-art workflows[33,37,38] and careful descriptor development[28] have enabled experimentally validated modeling within a DFT framework. The combination of plentiful and diverse materials data[11,18–20] and its programmatic accessibility[19,20] justify the application of data-mining techniques. These methods can quantitatively resolve subtle

Corey Oses, Department of Mechanical Engineering and Materials Science, Duke University, USA; corey.oses@duke.edu
Cormac Toher, Department of Mechanical Engineering and Materials Science, Duke University, USA; cormac.toher@duke.edu
Stefano Curtarolo, Duke University, USA; stefano@duke.edu
doi:10.1557/mrs.2018.207

trends and correlations among materials and their properties,[22,23,25,26,42,43] as well as motivate the formulation of novel property descriptors.[28,44] These "black-box" models, for which the learned, internal logic is largely obfuscated, are surprisingly accurate and quite valuable, particularly when few practical alternative modeling schemes exist—as is the case for predicting superconducting critical temperatures.[22,45]

Ultimately, the power in machine learning (ML) lies in the speed of its predictions, which outpaces DFT calculations by orders of magnitude.[46] Given that the number of currently characterized materials pales in comparison to the full space of hypothetical structures, methods to filter/screen the most interesting candidate materials[47]—powered by ML models—will undoubtedly become integral to future materials discovery workflows.

Thermodynamic formability modeling

Prediction of phase diagrams

Descriptions of thermodynamic stability and structural/chemical disorder are resolved through statistical analyses of aggregate sets of structures. Thermodynamic stability largely governs synthesizability, which can be determined by an analysis of how structures of similar compositions compete energetically (i.e., determination of the minimum Gibbs free energy surface). The procedure is algorithmically equivalent to finding the lower-half convex hull of all the relative free-energy minima,[48] as illustrated by the series of connected tie lines in **Figure 1**a. Composition and energy information from relevant AFLOW.org calculations are plotted, and the phases defining the minimum energy surface are identified.[49] Assuming sufficient sampling, the ground-state structures on the minimum energy surface form the low-temperature phase diagram.[50]

The convex hull construction offers a wealth of related thermodynamic properties. For near-hull structures (entries close to the tie lines), the energetic distance from the minimum energy surface is treated as a metric for synthesizability, as only small perturbations in temperature or pressure may be needed for it to be realized. In fact, this distance is equivalent to the amount of energy driving the decomposition of an unstable state to a linear combination of nearby ground-state structures. A similar distance—that of a stable phase from the pseudo-hull formed by neglecting that phase—quantifies the impact of a structure on the minimum energy surface and characterizes the robustness of stable structures (i.e., the stability criterion).[29,51]

AFLOW offers a module for autonomous calculation of the convex hull, which retrieves the set of relevant structure calculations from the repository[19,20] and delivers a thorough thermodynamic characterization for each.[51] Filtering schemes based on these thermodynamic properties, including the stability criterion and tie-line construction, played key roles in the discovery of new magnets[29] and in modeling superalloys.[27] The module powers an online web application for enhanced visualization of two-/three-dimensional hulls (available at www.aflow.org/aflow-chull).

Modeling disordered materials

Incorporating the effects of disorder is a necessary, albeit difficult, step in materials modeling. Not only is disorder intrinsic to all materials, but it also offers a route to enhanced and even otherwise inaccessible functionality. Disordered materials range from chemically disordered high-entropy materials and solid solutions, in which sites on a periodic crystal lattice are randomly occupied, to structurally disordered amorphous glasses exhibiting no crystalline periodicity. Materials such as high-entropy alloys[52,53] containing four to five metallic elements in equicomposition are currently being investigated for their enhanced thermomechanical properties,[54–58] and have also been reported to display superconductivity.[59] Research interest has recently expanded beyond metallic alloys to include high-entropy ceramics such as entropy stabilized oxides[60,61] and high-entropy borides,[62]

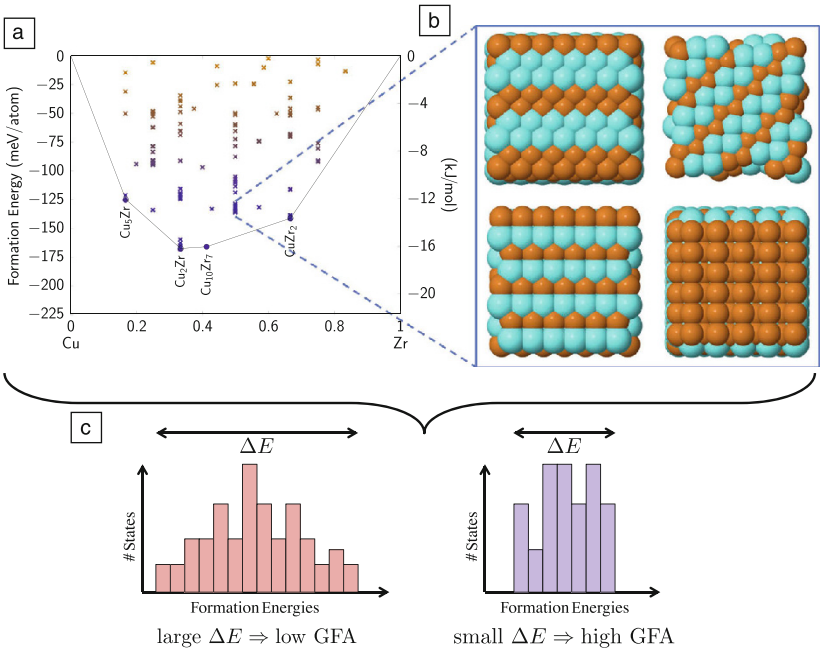


Figure 1. Descriptor for glass-forming ability (GFA). The GFA of metal-alloy systems can be predicted from the spread of formation energies of relevant ordered structures. Different structural phases with similar energies compete against each other during solidification, frustrating crystallization and promoting glass formation. A broad distribution of energies implies a low GFA, while a narrow distribution indicates a high GFA. (a) Energy-composition plot of CuZr. The series of connected tie lines define the minimum energy surface (convex hull).[51] (b) An illustration of distinct structural phases (with composition $Cu_{0.5}Zr_{0.5}$) that are very close in energy, causing frustration during solidification. (c) Illustrations of two examples of energy distributions characterizing low and high GFA, respectively.[28]

which display promising behavior including colossal dielectric constants[63] and superionic conductivity.[64]

Ab initio modeling of chemical/substitutional disorder—including vacancies and random site occupations—is a notoriously formidable problem, since it results in systems that cannot be described directly by a single unit cell with periodic boundary conditions. Rigorous statistical treatment of chemical disorder leverages a set of representative ordered supercells in thermodynamic competition. System-wide properties are resolved through ensemble averages of these supercells. This approach has been implemented in AFLOW for autonomous characterization,[38] and successfully validated for a number of technologically significant systems, including a zinc chalcogenide ($ZnS_{1-x}Se_x$), a wide-gap oxide semiconductor ($Mg_xZn_{1-x}O$), and an iron alloy ($Fe_{1-x}Cu_x$), recovering characteristic trends as a function of composition and offering additional insight into underlying physical mechanisms. The module determines the smallest superlattice size that accommodates the required stoichiometry to within a user-defined tolerance, and then generates the corresponding superlattices using Hermite Normal Form matrices.[65] All allowed decoration permutations are considered for each superlattice variant, generating the full set of possible supercell configurations. Degeneracies are rapidly identified by comparing approximate structure energies calculated with the Universal Force Field method.[66] Only unique supercells are individually characterized using standard *ab initio* packages.[1–7] The ensemble average values of properties such as the electronic bandgap, density of states, and the magnetic moment—weighted according to a Boltzmann distribution for a particular temperature—are then calculated to resolve the behavior of the disordered material.

Metallic glasses lack an ordered lattice and associated defects, which endow them with a unique combination of superb mechanical properties[67] and plastic-like processability,[68–70] and render them of great interest for several potential commercial and industrial applications.[71–73] To predict the glass-forming ability (GFA) of metal-alloy systems,[28] statistical approaches have been employed that blend the concept of thermodynamically competing ordered structures with the large quantities of precalculated data available in the AFLOW.org repository. The proposed physical mechanism is that ordered phases, which have similar energies, but are structurally distinct, compete against each other during solidification, frustrating crystal nucleation and thus promoting glass formation, as illustrated in Figure 1. The energy distribution of the different structures can be considered as forming a thermodynamic density of states (Figure 1c). A narrow distribution indicates a high GFA, while a wider distribution implies a low GFA. Atomic environment[74,75] comparisons determine the similarity of ordered crystalline phases, enabling the formulation of a quantitative descriptor that can be applied to the entire AFLOW.org database. The different structures are weighted according to a Boltzmann distribution to create the GFA descriptor. The model is found to successfully predict 73% of the glass-forming compositions for a set of 16 experimentally well-characterized alloy systems, and also indicates that about 17% of binary-alloy systems should be capable of glassification. By exploiting the precalculated data in the AFLOW.org repository, this model can be leveraged to rapidly predict GFA as a function of composition for thousands of alloy systems, demonstrating the power of applying intelligently constructed descriptors to computational materials data.

The AFLOW formation energy data is also employed to train cluster expansion models to compute the energies of multicomponent alloys,[76] which can be combined with thermodynamic modeling to predict the order–disorder transition temperature for solid solutions in high-entropy alloys.[44] Order–disorder transitions in the form of spinodal decomposition have also been proposed as a mechanism to (1) embed topologically protected conducting interface states in an insulating matrix[77] and (2) self-assemble nanostructures (such as thermoelectric devices[78]). The boundaries between different layers act as phonon scatterers, suppressing thermal conductivity and thus improving efficiency.

Exploiting ML algorithms

Model development

ML is rapidly emerging as a powerful tool for computational materials design.[42,79–81] Given sufficient training data, algorithms such as neural networks,[82] random forests,[83] gradient boosting decision trees,[84] and support vector machines[85] can learn to (1) identify the structures that are thermodynamically accessible for a given composition[42] and (2) accurately predict materials properties, such as the electronic bandgap,[23] elastic moduli,[23,86] vibrational energies,[26] and lattice thermal conductivity.[87]

The successful training of ML models depends on the set of features characterizing the material (i.e., the set of descriptors that form the feature vector).[42] Optimal descriptors are resolved by exploring different linear and nonlinear combinations of properties, and extracting the most efficient feature vector via compressive sensing.[43] Compressive sensing finds a sparse solution to the equations mapping the set of observable materials properties to the large set of possible test features—effectively reducing the dimensionality of the problem. The algorithm also filters for physically meaningful combinations of properties, based on dimensional analysis, to maximize interpretability of the final descriptor set.

Several different ML frameworks are leveraging data from the AFLOW.org repository. The materials fingerprinting model[22] codifies aspects of the electronic structure[10] to serve as unique markers for each material. In particular, the number of bands that intersect high-symmetry Brillouin zone points at discretized energy values form the band-structure fingerprint (illustrated in **Figure 2**), while simple discretization of the density of states form the density of states fingerprint. The Tanimoto coefficient—a distance metric[88]—between fingerprint vectors quantifies the similarity of the electronic structure between different materials. These fingerprints are employed

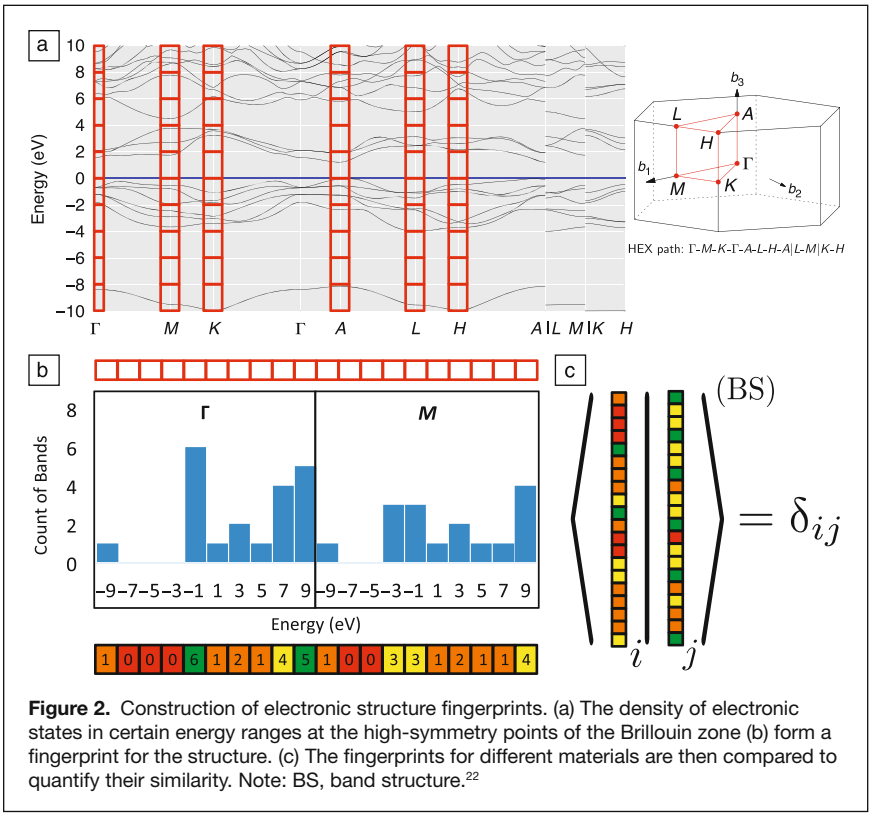


Figure 2. Construction of electronic structure fingerprints. (a) The density of electronic states in certain energy ranges at the high-symmetry points of the Brillouin zone (b) form a fingerprint for the structure. (c) The fingerprints for different materials are then compared to quantify their similarity. Note: BS, band structure.[22]

for the construction of networks (i.e., materials cartography),[22] where materials are represented by nodes and similarity correlates with relative positioning. When applied to compounds in the Inorganic Crystal Structure Database, significant clustering and structure can be identified for these networks, particularly with respect to material complexity (binaries versus ternaries), type (metal versus insulator), and, surprisingly, superconducting critical temperature—a particularly elusive phenomenon in which the driving mechanisms are still hotly contested.[22]

In the case of high-temperature superconductors, significant clustering suggests strong correlations among the electronic structure of these materials; although, as expected, these features alone are not enough to quantitatively resolve critical temperatures. Indeed, modeling improves with integration of more experimental observations[45,89] and properties, such as structural features and partial charges.[90] Incorporating additional relevant and physically meaningful training data, such as phonon spectra, should offer an applicability domain expansion and higher-fidelity predictions.

Thermomechanical properties calculated using the elastic constants[33] and Debye–Grüneisen[37] modules of AFLOW have been employed to train a gradient boosting decision trees framework[23] to predict quantities such as the bulk and shear modulus, Debye temperature, and heat capacity. Indicative of its versatility, the same model[23] has also been trained on AFLOW electronic structure data to classify materials as metals or insulators, and to predict the electronic bandgap for compounds identified as nonmetals. Model development is based on a fragment construction approach: each crystal is represented by a graph where nodes are decorated with corresponding atomic properties and connectivity is dictated by distance and the geometry of the local environment (Voronoi polyhedra adjacency). Path and circular fragments—representative of linear geometry and coordination polyhedra within the crystal—form the basis for feature development. To train the models, the gradient boosting decision tree algorithm is employed, which amalgamates a series of weak, easily constructed prediction rules to resolve a single, highly predictive function.

The resulting models have been thoroughly validated with simulated and real test sets, showing predictive metrics at 90% or higher against existing calculated and experimental measurements. Beyond property value prediction, feature-importance analyses of the models recover meaningful ways to tune the bandgap and Debye temperature, offering practical design rules for device engineering. The development of such models achieves the greatest impact on thermomechanical properties, where characterizations demand many single *ab initio* calculations, and thus presents a substantial boost in prediction speed at a fraction of the resources.

Workflow integration

ML approaches are expected to become indispensable in two specific scenarios: prediction of complex properties and screening of large sets of materials. Unfortunately, widespread exploitation of ML techniques in materials science has been hindered by the difficulty of setting up and interfacing the models with materials design infrastructures. To streamline this process, the AFLOW-ML API[91] has been created to provide programmatic access to the ML models described in References 23 and 26, with plans to extend it with additional models as they are developed. By posting a structure file to the API, users can retrieve ML predictions of thermal, mechanical, and electronic properties in the JavaScript Object Notation (JSON) data format.[92] In this way, all technical details of the ML algorithms are abstracted away, rendering a simple interface no more complicated than that of a standard API. This procedure can be easily incorporated into materials design workflows, due to its use of ubiquitous HTTP commands, along with the JSON format that is easily interpreted by a wide range of modern programming languages.

Source code and online forum

Since 2018, the AFLOW software (V3.1.204) has been made available for download/redistribution under the terms of the

GNU License, http://www.gnu.org/licenses. The source code/license/readme files can be found at www.aflow.org/src/aflow. Though some of the aforementioned modules are conveniently interfaced through the website, only the executable offers full and unabridged functionality. Additionally, the Forum (www.aflow.org/forum) advertises updates and new functionality, and hosts discussion boards for registered members to post questions.

Conclusion

Broad-scale thermodynamic formability modeling and exploitation of ML algorithms are the current frontiers in computational materials design. Recent progress in these fields has been enabled by large, programmatically accessible materials databases generated by automated computational infrastructure. Ensembles of ordered phases are being successfully employed to (1) construct phase diagrams and (2) formulate descriptors and models to predict the formation and properties of disordered materials. ML models have the potential to rapidly accelerate materials design as tools for predicting properties and identifying subtle/hidden trends—thus leading to enhanced understanding of the physical mechanisms underlying materials behavior.

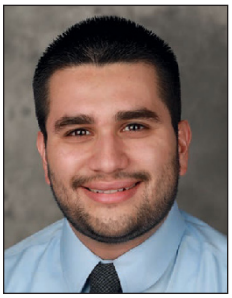
Acknowledgments

We thank E. Perim, Y. Lederer, O. Levy, O. Isayev, A. Tropsha, N. Mingo, J. Carrete, J.J. Vlassak, J. Schroers, D. Hicks, and E. Gossett for insightful discussions. C.O. acknowledges support from the NSF Graduate Research Fellowship No. DGF1106401. S.C. acknowledges support by the Alexander von Humboldt Foundation.

References

1. W. Kohn, L.J. Sham, *Phys. Rev.* **140** (4), A1133 (1965). doi:10.1103/PhysRev.140.A1133.
2. R.G. Parr, W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1994) pp. 142–197.
3. R.M. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, UK, 2004) pp. 119–184, doi:10.1017/CBO9780511805769.
4. G. Kresse, J. Fürthmüller, *Phys. Rev. B Condens. Matter* **54**, 11169 (1996).
5. P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G.L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A.P. Seitsonen, A. Smogunov, P. Umari, R.M. Wentzcovitch, *J. Phys. Condens. Matter* **21**, 395502 (2009).
6. X. Gonze, J.-M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G.-M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, P. Ghosez, J.-Y. Raty, D.C. Allan, *Comput. Mater. Sci.* **25**, 478 (2002).
7. V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, M. Scheffler, *Comput. Phys. Commun.* **180**, 2175 (2009).
8. P. Haas, F. Tran, P. Blaha, *Phys. Rev. B Condens. Matter* **79**, 085104 (2009).
9. S. Curtarolo, W. Setyawan, G.L.W. Hart, M. Jahnátek, R.V. Chepulskii, R.H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M.J. Mehl, H.T. Stokes, D.O. Demchenko, D. Morgan, *Comput. Mater. Sci.* **58**, 218 (2012).
10. W. Setyawan, S. Curtarolo, *Comput. Mater. Sci.* **49**, 299 (2010).
11. C.E. Calderon, J.J. Plata, C. Toher, C. Oses, O. Levy, M. Fornari, A. Natan, M.J. Mehl, G.L.W. Hart, M. Buongiorno Nardelli, S. Curtarolo, *Comput. Mater. Sci.* **108** (Pt. A), 233 (2015).
12. A.R. Supka, T.E. Lyons, L.S.I. Liyanage, P. D'Amico, R. Al Rahal Al Orabi, S. Mahatara, P. Gopal, C. Toher, D. Ceresoli, A. Calzolari, S. Curtarolo, M. Buongiorno Nardelli, M. Fornari, *Comput. Mater. Sci.* **136**, 76 (2017).
13. M. Scheffler, C. Draxl, Computer Center of the Max Planck Society, Garching, NOMAD Repository (2014), http://nomad-repository.eu.
14. 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, *APL Mater.* **1**, 011002 (2013).
15. J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, *JOM* **65**, 1501 (2013).
16. D.D. Landis, J.S. Hummelshøj, S. Nestorov, J. Greeley, M. Dułak, T. Bligaard, J.K. Nørskov, K.W. Jacobsen, *Comput. Sci. Eng.* **14**, 51 (2012).
17. G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, B. Kozinsky, *Comput. Mater. Sci.* **111**, 218 (2016).
18. S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R.H. Taylor, L.J. Nelson, G.L.W. Hart, S. Sanvito, M. Buongiorno Nardelli, N. Mingo, O. Levy, *Comput. Mater. Sci.* **58**, 227 (2012).
19. R.H. Taylor, F. Rose, C. Toher, O. Levy, K. Yang, M. Buongiorno Nardelli, S. Curtarolo, *Comput. Mater. Sci.* **93**, 178 (2014).
20. F. Rose, C. Toher, E. Gossett, C. Oses, M. Buongiorno Nardelli, M. Fornari, S. Curtarolo, *Comput. Mater. Sci.* **137**, 362 (2017).
21. K. Yang, W. Setyawan, S. Wang, M. Buongiorno Nardelli, S. Curtarolo, *Nat. Mater.* **11**, 614 (2012).
22. O. Isayev, D. Fourches, E.N. Muratov, C. Oses, K. Rasch, A. Tropsha, S. Curtarolo, *Chem. Mater.* **27**, 735 (2015).
23. O. Isayev, C. Oses, C. Toher, E. Gossett, S. Curtarolo, A. Tropsha, *Nat. Commun.* **8**, 15679 (2017).
24. G. Ceder, Y.-M. Chiang, D.R. Sadoway, M.K. Aydinol, Y.-I. Jang, B. Huang, *Nature* **392**, 694 (1998).
25. A. van Roekeghem, J. Carrete, C. Oses, S. Curtarolo, N. Mingo, *Phys. Rev. X* **6**, 041061 (2016).
26. F. Legrain, J. Carrete, A. van Roekeghem, S. Curtarolo, N. Mingo, *Chem. Mater.* **29**, 6220 (2017).
27. C. Nyshadham, C. Oses, J.E. Hansen, I. Takeuchi, S. Curtarolo, G.L.W. Hart, *Acta Mater.* **122**, 438 (2017).
28. E. Perim, D. Lee, Y. Liu, C. Toher, P. Gong, Y. Li, W.N. Simmons, O. Levy, J.J. Vlassak, J. Schroers, S. Curtarolo, *Nat. Commun.* **7**, 12315 (2016).
29. S. Sanvito, C. Oses, J. Xue, A. Tiwari, M. Zic, T. Archer, P. Tozman, M. Venkatesan, J.M.D. Coey, S. Curtarolo, *Sci. Adv.* **3**, e1602241 (2017).
30. G. Bergerhoff, R. Hundt, R. Sievers, I.D. Brown, *J. Chem. Inf. Comput. Sci.* **23**, 66 (1983).
31. M.J. Mehl, D. Hicks, C. Toher, O. Levy, R.M. Hanson, G.L.W. Hart, S. Curtarolo, *Comput. Mater. Sci.* **136**, S1 (2017).
32. D. Hicks, C. Oses, E. Gossett, G. Gomez, R.H. Taylor, C. Toher, M.J. Mehl, O. Levy, S. Curtarolo, *Acta Crystallogr. A Found. Adv.* **74**, 184 (2018).
33. C. Toher, C. Oses, J.J. Plata, D. Hicks, F. Rose, O. Levy, M. de Jong, M.D. Asta, M. Fornari, M. Buongiorno Nardelli, S. Curtarolo, *Phys. Rev. Mater.* **1**, 015401 (2017).
34. P. Nath, J.J. Plata, D. Usanmaz, R. Al Rahal Al Orabi, M. Fornari, M. Buongiorno Nardelli, C. Toher, S. Curtarolo, *Comput. Mater. Sci.* **125**, 82 (2016).
35. P. Nath, J.J. Plata, D. Usanmaz, C. Toher, M. Fornari, M. Buongiorno Nardelli, S. Curtarolo, *Scr. Mater.* **129**, 88 (2017).
36. J.J. Plata, P. Nath, D. Usanmaz, J. Carrete, C. Toher, M. de Jong, M.D. Asta, M. Fornari, M. Buongiorno Nardelli, S. Curtarolo, *NPJ Comput. Mater.* **3**, 45 (2017).
37. C. Toher, J.J. Plata, O. Levy, M. de Jong, M.D. Asta, M. Buongiorno Nardelli, S. Curtarolo, *Phys. Rev. B Condens. Matter* **90**, 174107 (2014).
38. K. Yang, C. Oses, S. Curtarolo, *Chem. Mater.* **28**, 6484 (2016).
39. L. Hedin, *Phys. Rev.* **139**, A796 (1965).
40. F. Aryasetiawan, O. Gunnarsson, *Rep. Prog. Phys.* **61**, 237 (1998).
41. H. Jiang, R.I. Gomez-Abal, P. Rinke, M. Scheffler, *Phys. Rev. B Condens. Matter* **82**, 045108 (2010).
42. L.M. Ghiringhelli, J. Vybiral, S.V. Levchenko, C. Draxl, M. Scheffler, *Phys. Rev. Lett.* **114**, 105503 (2015).
43. R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler, L.M. Ghiringhelli, *Phys. Rev. Mater.* **2**, 083802 (2018).
44. Y. Lederer, C. Toher, K.S. Vecchio, S. Curtarolo, *Acta Mater.* (2018), doi:10.1016/j.actamat.2018.07.042.
45. V. Stanev, C. Oses, A.G. Kusne, E. Rodriguez, J. Paglione, S. Curtarolo, I. Takeuchi, *NPJ Comput. Mater.* **4**, 29 (2018).
46. J.S. Smith, O. Isayev, A.E. Roitberg, *Chem. Sci.* **8**, 3192 (2017).
47. A. Walsh, *Nat. Chem.* **7**, 274 (2015).
48. S. Curtarolo, D. Morgan, G. Ceder, *Calphad* **29**, 163 (2005).
49. C.B. Barber, D.P. Dobkin, H. Huhdanpaa, *ACM Trans. Math. Softw.* **22**, 469 (1996).
50. G.L.W. Hart, S. Curtarolo, T.B. Massalski, O. Levy, *Phys. Rev. X* **3**, 041035 (2013).
51. C. Oses, E. Gossett, D. Hicks, F. Rose, M.J. Mehl, E. Perim, I. Takeuchi, S. Sanvito, M. Scheffler, Y. Lederer, O. Levy, C. Toher, S. Curtarolo, "AFLOW-CHULL: Cloud-Orientated Platform for Autonomous Phase Stability Analysis," submitted arXiv:1806.06901 (2018).
52. M.C. Gao, J.-W. Yeh, P.K. Liaw, Y. Zhang, Eds., *High-Entropy Alloys: Fundamentals and Applications* (Springer, Cham, Switzerland, 2015).
53. O.N. Senkov, J.D. Miller, D.B. Miracle, C. Woodward, *Nat. Commun.* **6**, 6529 (2015).
54. X. Lim, *Nature* **533**, 306 (2016).

55. O.N. Senkov, G.B. Wilks, D.B. Miracle, C.P. Chuang, P.K. Liaw, *Intermetallics* **18**, 1758 (2010).
56. B. Gludovatz, A. Hohenwarter, D. Catoor, E.H. Chang, E.P. George, R.O. Ritchie, *Science* **345**, 1153 (2014).
57. O.N. Senkov, G.B. Wilks, J.M. Scott, D.B. Miracle, *Intermetallics* **19**, 698 (2011).
58. Z. Li, K.G. Pradeep, Y. Deng, D. Raabe, C.C. Tasan, *Nature* **534**, 227 (2016).
59. F. von Rohr, M.J. Winiarski, J. Tao, T. Klimczuk, R.J. Cava, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E7144 (2016).
60. C.M. Rost, E. Sachet, T. Borman, A. Moballegh, E.C. Dickey, D. Hou, J.L. Jones, S. Curtarolo, J.-P. Maria, *Nat. Commun.* **6**, 8485 (2015).
61. Z. Rak, C.M. Rost, M. Lim, P. Sarker, C. Toher, S. Curtarolo, J.-P. Maria, D.W. Brenner, *J. Appl. Phys.* **120**, 095105 (2016).
62. J. Gild, Y. Zhang, T. Harrington, S. Jiang, T. Hu, M.C. Quinn, W.M. Mellor, N. Zhou, K. Vecchio, J. Luo, *Sci. Rep.* **6**, 37946 (2016).
63. D. Bérardan, S. Franger, D. Dragoe, A.K. Meena, N. Dragoe, *Phys. Status Solidi RRL* **10**, 328 (2016).
64. D. Bérardan, S. Franger, A.K. Meena, N. Dragoe, *J. Mater. Chem. A* **4**, 9536 (2016).
65. G.L.W. Hart, R.W. Forcade, *Phys. Rev. B Condens. Matter* **77**, 224115 (2008).
66. A.K. Rappe, C.J. Casewit, K.S. Colwell, W.A. Goddard, W.M. Skiff, *J. Am. Chem. Soc.* **114**, 10024 (1992).
67. W. Chen, J. Ketkaew, Z. Liu, R.M. Ojeda Mota, K. O'Brien, C.S. da Silva, J. Schroers, *Scr. Mater.* **107**, 1 (2015).
68. J. Schroers, N. Paton, *Adv. Mater. Proc.* **164**, 61 (2006).
69. J. Schroers, T.M. Hodges, G. Kumar, H. Raman, A.J. Barnes, Q. Pham, T.A. Waniuk, *Mater. Today* **14**, 14 (2011).
70. G. Kaltenboeck, M.D. Demetriou, S. Roberts, W.L. Johnson, *Nat. Commun.* **7**, 10576 (2016).
71. J. Schroers, *Adv. Mater.* **22**, 1566 (2010).
72. W.L. Johnson, J.H. Na, M.D. Demetriou, *Nat. Commun.* **7**, 10313 (2016).
73. M.F. Ashby, A.L. Greer, *Scr. Mater.* **54**, 321 (2006).
74. J.L.C. Daams, P. Villars, *Eng. Appl. Artif. Intell.* **13**, 507 (2000).
75. J.L.C. Daams, J.H.N. van Vucht, P. Villars, *J. Alloys Compd.* **182**, 1 (1992).
76. A. van de Walle, *Calphad* **33**, 266 (2009).
77. D. Usanmaz, P. Nath, C. Toher, J.J. Plata, R. Friedrich, M. Fornari, M. Buongiorno Nardelli, S. Curtarolo, *Chem. Mater.* **30**, 2331 (2018).
78. D. Usanmaz, P. Nath, J.J. Plata, G.L.W. Hart, I. Takeuchi, M. Buongiorno Nardelli, M. Fornari, S. Curtarolo, *Phys. Chem. Chem. Phys.* **18**, 5005 (2016).
79. H.K.D.H. Bhadeshia, *ISIJ Int.* **39**, 966 (1999).
80. E.O. Pyzer-Knapp, K. Li, A. Aspuru-Guzik, *Adv. Funct. Mater.* **25**, 6495 (2015).
81. R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T.D. Hirzel, D. Duvenaud, D. Maclaurin, M.A. Blood-Forsythe, H.S. Chae, M. Einzinger, D.-G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W. Huang, S.I. Hong, M. Baldo, R.P. Adams, A. Aspuru-Guzik, *Nat. Mater.* **15**, 1120 (2016).
82. B.G. Sumpter, D.W. Noid, *Annu. Rev. Mater. Sci.* **26**, 223 (1996).
83. L. Breiman, *Mach. Learn.* **45**, 5 (2001).
84. J.H. Friedman, *Ann. Stat.* **29**, 1189 (2001).
85. C. Cortes, V. Vapnik, *Mach. Learn.* **20**, 273 (1995).
86. M. de Jong, W. Chen, R. Notestine, K.A. Persson, G. Ceder, A. Jain, M.D. Asta, A. Gamst, *Sci. Rep.* **6**, 34256 (2016).
87. J. Carrete, W. Li, N. Mingo, S. Wang, S. Curtarolo, *Phys. Rev. X* **4**, 011019 (2014).
88. D. Bajusz, A. Rácz, K. Héberger, *J. Cheminform.* **7**, 20 (2015).
89. National Institute of Materials Science (NIMS), SuperCon (2011), http://supercon.nims.go.jp/index_en.html.
90. G. Henkelman, A. Arnaldsson, H. Jónsson, *Comput. Mater. Sci.* **36**, 354 (2006).
91. E. Gossett, C. Toher, C. Oses, O. Isayev, F. Legrain, F. Rose, E. Zurek, J. Carrete, N. Mingo, A. Tropsha, S. Curtarolo, *Comput. Mater. Sci.* **152**, 134 (2018).
92. D. Crockford, JavaScript Object Notation (JSON) Format (2017), http://www.json.org. □



Corey Oses is a doctoral candidate in the Department of Mechanical Engineering and Materials Science at Duke University, and a National Science Foundation Graduate Fellow. He received his BS degree in applied and engineering physics from Cornell University in 2013. His research includes design of data-driven thermodynamic descriptors for magnetic and disordered materials and the development of autonomous frameworks for chemical and crystallographic materials properties. Oses can be reached by email at corey.oses@duke.edu and corey.oses@gmail.com.

Cormac Toher is an assistant research professor in the Department of Mechanical Engineering and Materials Science at Duke University, and a member of the Center for Materials Genomics. He received his PhD degree in physics from Trinity College Dublin, Ireland, in 2008. His research includes the automation of calculations of the thermal and mechanical properties of materials, the development of descriptors for the formation and stability of disordered materials, and the development of machine-learning algorithms trained on computational materials science data. Toher can be reached by email at cormac.toher@duke.edu and toherc@gmail.com.

Stefano Curtarolo has been on the faculty of materials science, physics, and chemistry at Duke University, since receiving his PhD degree in materials science from the Massachusetts Institute of Technology in 2003. His research interests focus on the interface between fundamental and applied materials science: disordered materials (glasses and high-entropy systems), super high-temperature structural materials, automatic materials discovery, magnetic/vibrational states of materials, thermal conductivity, and thermoelectric properties of advanced energy systems. He is a Fellow of the American Physical Society and the Institute of Physics. He is a recipient of research awards from IUPAP, DOD, NSF, and the Alexander von Humboldt Foundation. He is the founder of the AFLOW.org consortium. Curtarolo can be reached by email at stefano@duke.edu.