The AFLOW Fleet for Materials Discovery

Cormac Toher, Corey Oses, David Hicks, Eric Gossett, Frisco Rose, Pinku Nath, Demet Usanmaz, Denise C. Ford, Eric Perim, Camilo E. Calderon, Jose J. Plata, Yoav Lederer, Michal Jahnátek, Wahyu Setyawan, Shidong Wang, Junkai Xue, Kevin Rasch, Roman V. Chepulskii, Richard H. Taylor, Geena Gomez, Harvey Shi, Andrew R. Supka, Rabih Al Rahal Al Orabi, Priya Gopal, Frank T. Cerasoli, Laalitha Liyanage, Haihang Wang, Ilaria Siloi, Luis A. Agapito, Chandramouli Nyshadham, Gus L. W Hart, Jesús Carrete, Fleur Legrain, Natalio Mingo, Eva Zurek, Olexandr Isayev, Alexander Tropsha, Stefano Sanvito, Robert M. Hanson, Ichiro Takeuchi, Michael J. Mehl, Aleksey N. Kolmogorov, Kesong Yang, Pino D'Amico, Arrigo Calzolari, Marcio Costa, Riccardo De Gennaro, Marco Buongiorno Nardelli, Marco Fornari, Ohad Levy, and Stefano Curtarolo

Contents

C. Toher (✉) · C. Oses · D. Hicks · E. Gossett · F. Rose · P. Nath · D. Usanmaz · D. C. Ford · E. Perim · C. E. Calderon · M. Jahnátek · W. Setyawan · S. Wang · J. Xue · K. Rasch · R. V. Chepulskii · G. Gomez
Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA
Center for Materials Genomics, Duke University, Durham, NC, USA
e-mail: toherc@gmail.com

J. J. Plata
Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA
Center for Materials Genomics, Duke University, Durham, NC, USA
Departamento de Química Física, Universidad de Sevilla, Sevilla, Spain

W. Andreoni, S. Yip (eds.), *Handbook of Materials Modeling*,
https://doi.org/10.1007/978-3-319-42913-7_63-2

Abstract

The traditional paradigm for materials discovery has been recently expanded to incorporate substantial data-driven research. With the intent to accelerate the development and the deployment of new technologies, the AFLOW Fleet for computational materials design automates high-throughput first-principles calculations and provides tools for data verification and dissemination for a

Y. Lederer · O. Levy
Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA
Center for Materials Genomics, Duke University, Durham, NC, USA
Department of Physics, NRCN, Beer-Sheva, Israel

R. H. Taylor
Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA
Center for Materials Genomics, Duke University, Durham, NC, USA
Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

H. Shi
Center for Materials Genomics, Duke University, Durham, NC, USA

A. R. Supka
Department of Physics and Science of Advanced Materials Program, Central Michigan University, Mount Pleasant, MI, USA
Science of Advanced Materials Program, Central Michigan University, Mount Pleasant, MI, USA

R. Al Rahal Al Orabi
Department of Physics and Science of Advanced Materials Program, Central Michigan University, Mount Pleasant, MI, USA
Solvay, Design and Development of Functional Materials Department, AXEL'ONE Collaborative Platform – Innovative Materials, Saint Fons Cedex, France

P. Gopal
Department of Physics and Science of Advanced Materials Program, Central Michigan University, Mount Pleasant, MI, USA

broad community of users. AFLOW incorporates different computational modules to robustly determine thermodynamic stability, electronic band structures, vibrational dispersions, thermomechanical properties, and more. The AFLOW data repository is publicly accessible online at `aflow.org`, with more than 1.8 million materials entries and a panoply of queryable computed properties. Tools to programmatically search and process the data, as well as to perform online machine learning predictions, are also available.

F. T. Cerasoli · L. Liyanage · H. Wang · I. Siloi · L. A. Agapito
Department of Physics and Department of Chemistry, University of North Texas, Denton, TX, USA

C. Nyshadham · G. L. W. Hart
Department of Physics and Astronomy, Brigham Young University, Provo, UT, USA

J. Carrete
Institute of Materials Chemistry, TU Wien, Vienna, Austria

F. Legrain
Universitè Grenoble Alpes, Grenoble, France
CEA, LITEN, Grenoble, France

N. Mingo
CEA, LITEN, Grenoble, France

E Zurek
Department of Chemistry, State University of New York at Buffalo, Buffalo, NY, USA

O. Isayev · A. Tropsha
Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC, Chapel Hill, NC, USA
Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, USA

S. Sanvito
Center for Materials Genomics, Duke University, Durham, NC, USA
School of Physics, AMBER and CRANN Institute, Trinity College, Dublin, Ireland

R. M. Hanson
Department of Chemistry, St. Olaf College, Northfield, MN, USA

I. Takeuchi
Center for Nanophysics and Advanced Materials, University of Maryland, College Park, MD, USA

Department of Materials Science and Engineering, University of Maryland, College Park, MD, USA

M. J. Mehl
United States Naval Academy, Annapolis, MD, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

A. N. Kolmogorov
Department of Physics, Binghamton University, State University of New York, Binghamton, NY, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

1 Introduction

The AFLOW Fleet is an integrated software infrastructure for automated materials design (Curtarolo et al. 2013) centered around the Automatic Flow (AFLOW) (Curtarolo et al. 2012a) framework for computational materials science. It features multiple scientific software packages, including the AFLOW high-throughput framework, the AFLOWπ (Supka et al. 2017) medium-throughput framework, and the PAOFLOW (Buongiorno Nardelli et al. 2017) utility for electronic structure analysis, along with the AFLOW.org data repository (Curtarolo et al. 2012b), its associated representational state transfer application programming interface (REST-API)) (Taylor et al. 2014), and the AFLUX Search-API (Rose et al. 2017). These elements are well integrated with one another: a Python+JSON (JavaScript Object Notation)

K. Yang
Department of NanoEngineering, University of California San Diego, La Jolla, CA, USA
Center for Materials Genomics, Duke University, Durham, NC, USA

P. D'Amico
CNR-NANO Research Center S3, Modena, Italy
Dipartimento di Fisica, Informatica e Matematica, Universitá di Modena and Reggio Emilia, Modena, Italy

A. Calzolari
Center for Materials Genomics, Duke University, Durham, NC, USA
CNR-NANO Research Center S3, Modena, Italy
Department of Physics and Department of Chemistry, University of North Texas, Denton, TX, USA

M. Costa
Brazilian Nanotechnology National Laboratory (LNNano), CNPEM, Campinas, Brazil

R. De Gennaro
Dipartimento di Fisica, Universit'a di Roma Tor Vergata, Roma, Italy

M. Buongiorno Nardelli
Center for Materials Genomics, Duke University, Durham, NC, USA
Department of Physics and Department of Chemistry, University of North Texas, Denton, TX, USA
e-mail: Marco.BuongiornoNardelli@unt.edu

M. Fornari
Center for Materials Genomics, Duke University, Durham, NC, USA
Department of Physics and Science of Advanced Materials Program, Central Michigan University, Mount Pleasant, MI, USA
Science of Advanced Materials Program, Central Michigan University, Mount Pleasant, MI, USA
e-mail: forna1m@cmich.edu

S. Curtarolo (✉)
Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA
Center for Materials Genomics, Duke University, Durham, NC, USA
Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin-Dahlem, Germany
e-mail: stefano@duke.edu

interface connects AFLOW, AFLOWπ, and PAOFLOW; and all software packages access the AFLOW.org repository via the REST-API and the Search-API.

Similar infrastructure has been developed by initiatives such as the Materials Project (Jain et al. 2013), NoMaD (Scheffler et al. 2014), OQMD (Saal et al. 2013), the Computational Materials Repository (Landis et al. 2012), and AiiDA (Pizzi et al. 2016). The Materials Project uses the `pymatgen` (Ong et al. 2013) Python-language data generation software infrastructure, and their repository is available at `materialsproject.org`. The Novel Materials Discovery (NoMaD) Laboratory maintains an aggregate repository available at `nomad-repository.eu`, incorporating data generated by other frameworks including AFLOW. The Open Quantum Materials Database (OQMD) (Saal et al. 2013) uses tools such as `qmpy` to generate their database, which can be accessed at `oqmd.org`. The Atomic Simulation Environment (ASE) (Bahn and Jacobsen 2002) is used to generate the Computational Materials Repository, available at `cmr.fysik.dtu.dk`. The ASE scripting interface is also used by the Automated Interactive Infrastructure and Database (AiiDA) framework available at `aiida.net`, which revolves around relational databases for its overall design and data storage. Additional materials design utilities include the High-Throughput-Toolkit (HTTK) and the associated Open Materials Database, `httk.openmaterialsdb.se`, as well as the Materials Mine database available at `www.materials-mine.com`, while computationally predicted crystal structures can be obtained from the Theoretical Crystallography Open Database at `www.crystallography.net/tcod/`.

The AFLOW Fleet employs density functional theory (DFT) to perform the quantum mechanical calculations required to obtain materials properties from first principles. These DFT calculations are carried out by external software packages, namely, the Vienna Ab initio Simulation Package (VASP) (Kresse and Hafner 1993; Kresse and Furthmüller 1996) in the case of AFLOW and QUANTUM ESPRESSO (Giannozzi et al. 2009, 2017) in the case of AFLOWπ. Results are stored in the AFLOW.org repository (Curtarolo et al. 2012b) and made freely available online via the `aflow.org` web portal, which is programmatically accessible and searchable via the AFLOW Data REST-API (Taylor et al. 2014) and AFLUX Search-API (Rose et al. 2017), respectively. The repository currently contains calculated properties for over 1.8 million materials entries, including both experimentally observed and theoretically predicted structures, and new results are continuously being added. This AFLOW data is successfully applied to (i) formulate descriptors for the formation of disordered materials such as metallic glasses (Perim et al. 2016), (ii) find new magnetic materials (Sanvito et al. 2017) and superalloys (Nyshadham et al. 2017), (iii) generate phase diagrams for alloy systems (Barzilai et al. 2016, 2017a; Lederer et al. 2018) and identify new ordered compounds (Levy et al. 2010a,b,c; Barzilai et al. 2017b), and (iv) train machine learning models to identify potential superconductors (Isayev et al. 2015) and predict electronic and thermomechanical properties (Isayev et al. 2017).

2 AFLOW: Efficient Data Generation

The AFLOW framework for computational materials science automates the full workflow for materials properties calculations (Curtarolo et al. 2012a). Using a standard set of calculation parameters (Calderon et al. 2015), input files are automatically generated for the VASP (Kresse and Hafner 1993; Kresse and Furthmüller 1996) DFT software package with projector-augmented-wave pseudopotentials (Blöchl 1994) and the PBE parameterization of the generalized gradient approximation to the exchange-correlation functional (Perdew et al. 1996). Calculations are monitored as they run to detect and correct for errors without the need for any user intervention. Useful materials data is then extracted and processed for dissemination through the AFLOW.org repository. The entire framework is written in the C++ programming language (more than 400,000 highly integrated lines, as of version 3.1.153), providing a robust platform for continuous infrastructure development with reliable high performance.

2.1 AFLOW: Automated Workflows

AFLOW offers several automated workflows, each dedicated to a specific type of characterization yielding a set materials properties. For electronic properties, AFLOW performs four DFT calculations: two rounds of geometry relaxation (stage name: "RELAX") using the VASP conjugate gradient optimization algorithm, a static run (i.e., fixed geometry; stage name: "STATIC") with a denser **k**-point mesh to obtain an accurate density of states, and a band structure calculation (stage name: "BANDS") following the AFLOW Standard path through the high-symmetry **k**-points in the Brillouin zone (Setyawan and Curtarolo 2010).

Other workflows in AFLOW manage ensembles of DFT calculations, all offering the same automated error-correction procedures for high-throughput processing. For thermal and elastic properties, the Debye-Grüneisen model (Automatic GIBBS Library, AGL) (Toher et al. 2014) is combined with the Automatic Elasticity Library (AEL) (Toher et al. 2017) as described in Sect. 2.6. A more accurate thermal characterization can be resolved with the finite displacement method for phonon calculations (Automatic Phonon Library, APL) (Nath et al. 2016) and its associated extensions, i.e., the quasi-harmonic approximation (QHA-APL) (Nath et al. 2016) and Automatic Anharmonic Phonon Library (AAPL) (Plata et al. 2017), as described in Sect. 2.7. AFLOW also extends beyond ideal crystalline materials characterization, offering modules to investigate off-stoichiometric materials (AFLOW-POCC, Sect. 2.5) (Yang et al. 2016) and to predict metallic glass formation as a function of composition (Perim et al. 2016).

2.2 AFLOW: Prototype Library

The AFLOW framework uses decorated crystal structure prototypes for materials discovery (Mehl et al. 2017). Structural prototypes are specific arrangements of atoms which are commonly observed in nature, such as the rock salt, zinc blende, and wurtzite structures. The atomic sites in these prototypes are populated with different elemental species to generate materials structures, for which the properties and thermodynamic stability are then obtained from DFT calculations. An extensive list of the structural prototypes included in AFLOW has been published in Mehl et al. (2017) and is available online at `http://aflow.org/CrystalDatabase`.

Pages within the website display a curated list of data for each structural prototype, including materials exhibiting this structure, various symmetry descriptions, the primitive and atomic basis vectors, and original references where the structure was observed. Accompanying these descriptions is an interactive `Jmol` visualization of the prototype, as described in Sect. 2.8. The page also contains a prototype generator, where the structural degree(s) of freedom and atomic species are defined to create new materials by leveraging the AFLOW prototypes module. This generates the corresponding input file for one of many ab initio software packages, including VASP (Kresse and Hafner 1993; Kresse and Furthmüller 1996), QUANTUM ESPRESSO (Giannozzi et al. 2009, 2017), ABINIT (Gonze et al. 2002), and FHI-AIMS (Blum et al. 2009).

2.3 AFLOW-SYM: Symmetry Analyzer

The AFLOW framework automatically analyzes the symmetry of materials structures and returns a complete symmetry description. To address numerical tolerance issues, AFLOW employs an atom mapping procedure that is reliable even for non-orthogonal unit cells and uses an adaptive tolerance scheme to ensure symmetry results are commensurate with crystallographic principles (see Fig. 1). These routines—referred to as AFLOW-SYM (Hicks et al. 2018)—are robust and have been used to successfully determine the symmetry properties of over 1.8 million materials in the AFLOW repository.

Structural isometries are identified by determining the set of symmetry operators that lead to isomorphic mappings between the original and transformed atoms. The structure exhibits symmetry under a particular operation if the set of closest mapping distances are all below a tolerance threshold ϵ_0. Periodic boundary conditions introduce complexity for finding the minimum mapping vector, necessitating the exploration of neighboring cells. This is achieved via the method of images through either (i) a unit cell expansion, yielding the globally optimal distance or (ii) a bring-in-cell method (generally performed in fractional coordinates) that reduces each component of the distance vector independently. While computationally inexpensive compared to the unit cell expansion, the bring-in-cell method is only exact for orthogonal lattices (i.e., described by a diagonal metric tensor), since it

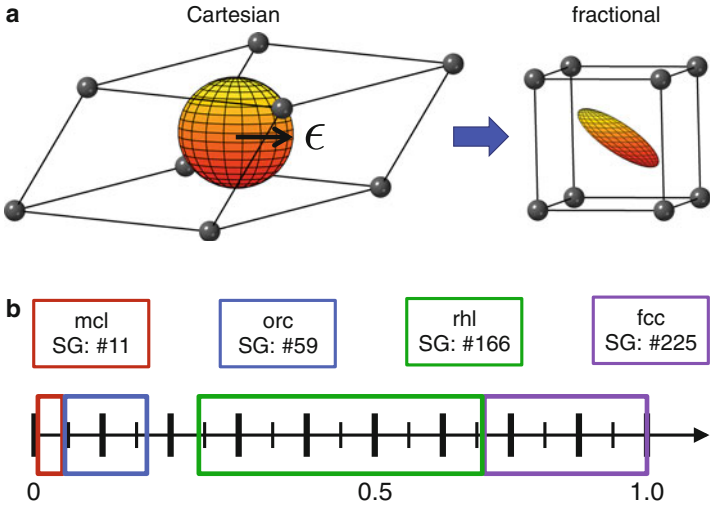

Fig. 1 Visualization of tolerance-sphere warping and adaptive tolerance method. (**a**) Illustration of the warping of space when transforming from cartesian to fractional coordinates in the general case. (**b**) Spectrum of space groups identified by AFLOW-SYM with different tolerance choices

does not consider overlap between lattice vectors (see Fig. 1a). To safely exploit the bring-in-cell approach, AFLOW-SYM employs a heuristic maximum tolerance $\epsilon_{\max}$ based on the maximum lattice skewness with a threshold which guarantees consistent and accurate results (Hicks et al. 2018).

Given a particular tolerance value, different symmetry operations can be realized in or excluded from the description of a crystal. Figure 1b highlights how the tolerance value affects the possible space groups for AgBr (ICSD #56551 with a reported space group #11). The neighboring space group regions are consistent with non-isomorphic subgroup relations, namely, between space groups #59 and #11 and between #225 and #166. However, a gap or "confusion" tolerance region occurs between space groups #59 and #166 (with no direct subgroup relations). The problematic regions stem from noise in the structural data, impeding the identification of operations consistent with symmetry principles. This problem is solved by using a radial tolerance scan extending from the input tolerance ϵ_0. Given a change in tolerance, the algorithm recalculates and verifies all symmetry properties until a globally consistent description is identified.

AFLOW-SYM is compatible with many established ab initio input files, including those for VASP (Kresse and Furthmüller 1996), QUANTUM ESPRESSO (Giannozzi et al. 2009, 2017), ABINIT (Gonze et al. 2002), and FHI-AIMS (Blum et al. 2009). From the structural information, AFLOW-SYM delivers the symmetries of the lattice, crystal (lattice + atoms), reciprocal lattice, superlattice (equally decorated sites), and crystal-spin (lattice + atoms + magnetic moment). This affords a multitude of symmetry descriptions to be presented, such as the space group number/symbol(s),

Pearson symbol, point group symbol(s), Wyckoff positions, and Bravais lattice type/variation (Setyawan and Curtarolo 2010). Moreover, the operators of the different symmetry groups—including the point groups, factor groups, space group, and site symmetries—are provided to users in rotation matrix, axis-angle, matrix generator, and quaternion representations for easy manipulation. All symmetry functions support the option to output in JSON format. This allows AFLOW-SYM to be leveraged from other programming languages such as Java, Go, Ruby, Julia, and Python—facilitating the incorporation of AFLOW-SYM into numerous applications and workflows.

2.4 AFLOW-CHULL: Convex Hull Analysis

Structure and energy data from the AFLOW.org repository are used to resolve the low-temperature/low-pressure thermodynamic stability of compound systems. For a given stoichiometry, the AFLOW.org repository provides the DFT energies of various crystal polymorphs. By exploring representative structures over the full range of stoichiometries, AFLOW-CHULL (Oses et al. 2018) constructs the minimum energy surface, i.e., the lower-half convex hull (Barber et al. 1996), defining thermodynamic stability for the system (at zero temperature and pressure). Structures on the hull are thermodynamically stable (ground state), while those far from the hull will decompose into a combination of stable phases, dictating synthesizability at these conditions. Any analysis of the hull requires sufficient statistics to ensure convergence, i.e., enough representative structures have been included in the alloy system calculations such that any additional entries are not expected to change the minimum energy surface.

The geometric construction offers several key properties critical for synthesizability. For a specific composition, the energetic distance to the hull quantifies the energy released during the decomposition, while the ground state phases defining the tie-line/facet below the compound are the products of the reaction. The distance from the hull also measures the "severity" of instability, i.e., structures near the hull may stabilize at higher temperatures or pressures. Similarly, a robust stability criterion can be quantified for ground state phases by removing the phase from the set and measuring the distance of the compound from the new hull. The larger the distance, the less likely the ground state phase will become unstable at higher temperatures/pressures (Sanvito et al. 2017). The generalized tie-lines (facet ridges) dictate which phases can coexist in equilibrium and play a role in determining the feasibility of synthesis/treatment techniques, such as precipitation hardening (Nyshadham et al. 2017).

Given a compound system, AFLOW-CHULL automatically queries the AFLOW.org database, constructs the hull, calculates the aforementioned properties, and delivers the information in one of the following formats: PDF, plaintext, and JSON. AFLOW-CHULL can also visualize the 2D and 3D hulls, as illustrated in Fig. 2. In the case of the PDF output format, hyperlinks are included to allow for additional queries of the full properties set offered through the AFLOW.org repository. Links are also

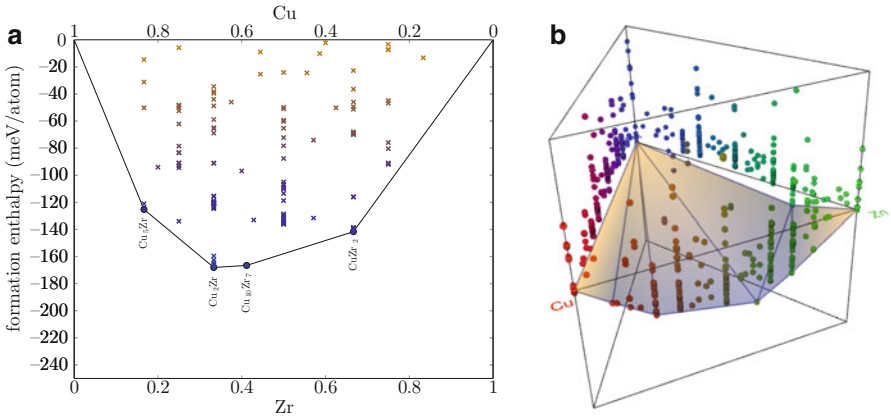

Fig. 2 Example convex hull illustrations offered by AFLOW. (**a**) 2D convex hull of the CuZr system generated automatically by AFLOW. (**b**) 3D convex hull of the CuMnZn system presented through the AFLOW-CHULL application online: http://aflow.org/aflow-chull

added connecting the hull visualization to relevant properties for easy navigation of the full PDF document.

A separate online application, available at `aflow.org/aflow-chull`, has been created to showcase the results of AFLOW-CHULL and provides interactive binary and ternary convex hull visualizations. The application consists of four components: the periodic table, visualization viewport, selected entries list, and the comparison page. The periodic table component is the entry point of the application and provides the interface to search for convex hulls of different alloy systems. Elements within the periodic table respond when selected to display information to the user. As a selection is made, the color of each border will change to green, yellow, and red based on hull reliability. A reliability threshold of 200 compounds for a binary hull has been heuristically defined. Selections highlighted in green are well above this threshold, while those in yellow/red are near/below the cutoff.

When a hull is selected, the application transitions to the visualization viewport component. Depending on the number of elements selected, a 2D plot (binary) or 3D plot (ternary) will appear. Each plot is interactive, allowing points to be selected, where each point represents an entry in the AFLOW repository. Information for each point is displayed in the selected entries list component, which is accessible through the navigation bar. Selected hulls will appear on the comparison page as a grid of cards, and selected points are highlighted across all hulls containing those entries.

2.5 AFLOW-POCC: Partial Occupations

The AFLOW Partial Occupation module (AFLOW-POCC) (Yang et al. 2016) models configurational and structural disorder including substitutions, vacancies, and ran-

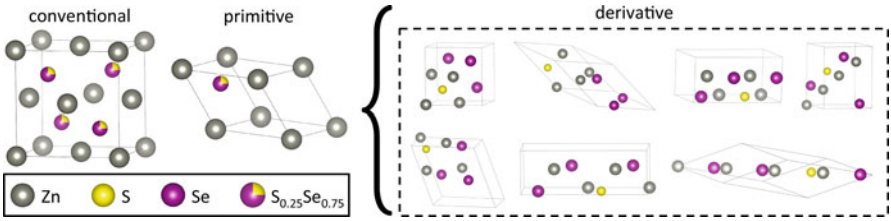

Fig. 3 Structure enumeration for off-stoichiometric materials modeling. For the off-stoichiometric material $ZnS_{0.25}Se_{0.75}$, a superlattice of size $n = 4$ accommodates the stoichiometry exactly. By considering all possibilities of decorated supercells and eliminating duplicates by UFF energies, seven structures are identified as unique. These representative structures are fully characterized by AFLOW and VASP and are ensemble-averaged to resolve the system-wide properties

dom lattice site occupation, by generating a set of representative configurations. First, a supercell size is determined that accommodates the fractional stoichiometry to within a user-defined tolerance. Given the supercell size, n, superlattices are generated using Hermite Normal Form matrices (Hart and Forcade 2008), which are then decorated in accordance with the stoichiometry to generate all possible configurations, as illustrated in Fig. 3. Duplicate configurations are rapidly identified and eliminated by estimating the energy of each structure using the Universal Force Field (UFF) model (Rappe et al. 1992). The properties of the remaining unique configurations are calculated with DFT, and ensemble-averaged to resolve system-wide properties of the disordered material. The ensemble-average employs a Boltzmann distribution weight which is a function of a disorder parameter (temperature), energy relative to the ground state configuration, and degeneracy as determined by the UFF model. Ensemble-average properties include the electronic band gap, density of states, and magnetic moment.

2.6 AEL and AGL: Thermomechanical Properties

The AFLOW Automatic Elasticity Library (AFLOW-AEL Toher et al. 2017) and the AFLOW Automatic GIBBS Library (AFLOW-AGL Toher et al. 2014) modules determine thermomechanical materials properties from calculations of strained primitive cells. These methods are generally computationally less costly than the phonon (APL and AAPL) calculations described in Sect. 2.7, although APL and AAPL generally give more quantitatively accurate results, particularly for properties where anharmonic effects are important. AEL and AGL have been combined into a single automated workflow, which has been used to calculate the thermomechanical properties for over 5000 materials in the AFLOW repository.

The AEL module applies a set of independent normal and shear strains to the primitive cell of a material (de Jong et al. 2015; Toher et al. 2017) as depicted in Fig. 4a and uses DFT to calculate the resulting stress tensors. This set of strain-stress data is used to generate the elastic stiffness tensor, i.e., the elastic constants:

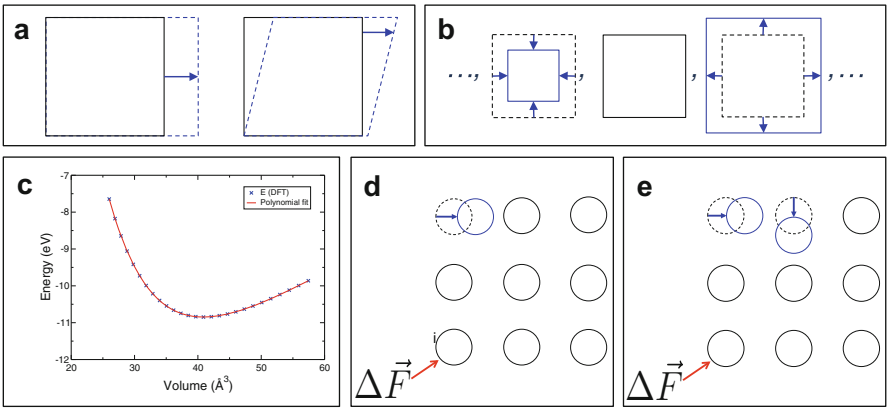

Fig. 4 Ensembles of distorted geometries for the calculation of thermomechanical properties with AFLOW. AEL uses the stresses from a set of (**a**) normal and shear strained cells to obtain the elastic constants. AGL calculates the energies of a set of (**b**) isotropically compressed and expanded unit cells and (**c**) fits the resulting $E(V)$ data by a numerical polynomial or by an empirical equation of state. APL obtains the (**d**) second-order harmonic IFCs from a set of single-atom displacements and the (**e**) third-order anharmonic IFCs from a set of two-atom displacements

$$\begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{13} \\ \sigma_{12} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ c_{12} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\ c_{13} & c_{23} & c_{33} & c_{34} & c_{35} & c_{36} \\ c_{14} & c_{24} & c_{34} & c_{44} & c_{45} & c_{46} \\ c_{15} & c_{25} & c_{35} & c_{45} & c_{55} & c_{56} \\ c_{16} & c_{26} & c_{36} & c_{46} & c_{56} & c_{66} \end{pmatrix} \begin{pmatrix} \epsilon_{11} \\ \epsilon_{22} \\ \epsilon_{33} \\ 2\epsilon_{23} \\ 2\epsilon_{13} \\ 2\epsilon_{12} \end{pmatrix} \tag{1}$$

written in the 6×6 Voigt notation using the mapping (Poirier 2000): $11 \mapsto 1$, $22 \mapsto 2$, $33 \mapsto 3$, $23 \mapsto 4$, $13 \mapsto 5$, $12 \mapsto 6$. These elastic constants are combined to calculate the bulk, B, and shear, G, elastic moduli in the Voigt, Reuss, and Voigt-Reuss-Hill (VRH, B_{VRH} and G_{VRH}) approximations. The Poisson ratio ν is then given by:

$$\nu = \frac{3B_{\mathrm{VRH}} - 2G_{\mathrm{VRH}}}{6B_{\mathrm{VRH}} + 2G_{\mathrm{VRH}}}. \tag{2}$$

The AGL module is based on the GIBBS (Blanco et al. 1996, 2004) quasi-harmonic Debye-Grüneisen model and calculates the energy as a function of volume, $E(V)$, for a set of isotropically compressed and expanded strains of the primitive cell, as illustrated in Fig. 4b. The $E(V)$ data are fitted by either a numerical polynomial or an empirical equation of state to obtain the adiabatic bulk modulus $B_{\mathrm{S}}(V)$, as shown in Fig. 4c. The Debye temperature $\theta_{\mathrm{D}}(V)$ as a function of volume is then calculated using the expression:

$$\theta_{\mathrm{D}} = \frac{\hbar}{k_{\mathrm{B}}}[6\pi^2 V^{1/2} n]^{1/3} f(\nu)\sqrt{\frac{B_{\mathrm{S}}}{M}}, \tag{3}$$

where n is the number of atoms per unit cell, M is the unit cell mass, and $f(\nu)$ is a function of the Poisson ratio ν:

$$f(\nu) = \left\{3\left[2\left(\frac{2}{3}\cdot\frac{1+\nu}{1-2\nu}\right)^{3/2} + \left(\frac{1}{3}\cdot\frac{1+\nu}{1-\nu}\right)^{3/2}\right]^{-1}\right\}^{\frac{1}{3}}, \tag{4}$$

where ν can be obtained from Eq. 2 using AEL or set directly by the user. The vibrational contribution to the free energy, F_{vib}, is given by:

$$F_{\mathrm{vib}}(\theta_{\mathrm{D}}; T) = nk_{\mathrm{B}}T\left[\frac{9}{8}\frac{\theta_{\mathrm{D}}}{T} + 3\log\left(1 - \mathrm{e}^{-\theta_{\mathrm{D}}/T}\right) - D\left(\frac{\theta_{\mathrm{D}}}{T}\right)\right], \tag{5}$$

where $D(\theta_{\mathrm{D}}/T)$ is the Debye integral:

$$D\left(\theta_{\mathrm{D}}/T\right) = 3\left(\frac{T}{\theta_{\mathrm{D}}}\right)^3 \int_0^{\theta_{\mathrm{D}}/T} \frac{x^3}{e^x - 1} dx. \tag{6}$$

The Gibbs free energy is obtained from:

$$\mathsf{G}(V; p, T) = E_{\mathrm{DFT}}(V) + F_{\mathrm{vib}}(\theta_{\mathrm{D}}(V); T) + pV. \tag{7}$$

The volume which minimizes $\mathsf{G}(V; p, T)$ at a given pressure p and temperature T is the equilibrium volume V_{eq}, which is used to evaluate $\theta_{\mathrm{D}}(V_{\mathrm{eq}})$ and the Grüneisen parameter γ as defined by:

$$\gamma = -\frac{\partial \log(\theta_{\mathrm{D}}(V))}{\partial \log V}. \tag{8}$$

Finally, θ_{D} and γ are used to calculate other thermal properties including C_{V}, C_{p}, α_{V}, and κ_{L} (Toher et al. 2014; Blanco et al. 2004).

2.7 AFLOW-APL: Phonons

The AFLOW Automatic Phonon Library (AFLOW-APL) (Curtarolo et al. 2012a) calculates the harmonic vibrational properties of a crystal using the finite displacement method. Computed properties include the phonon dispersion and density of states, vibrational entropy (S_{vib}), and the heat capacity (at constant volume, C_{V}) as a function of temperature. These features are determined through an analysis of the phonon modes, accessed through the Interatomic Force Constants (IFCs) (Maradudin et al. 1971). To first approximation, the harmonic (second-order)

IFC $C_{i,j;\alpha,\beta}$ is the negative of the force exerted in the α direction on the atom i when the atom j is displaced in the β direction, with all other atoms maintaining their equilibrium position. To determine the forces, the atoms of the structure (supercell) are individually perturbed as illustrated in Fig. 4d. The forces are obtained with DFT from the derivative of the total energy using the Hellmann-Feynman theorem. Supercells are used to sufficiently capture/isolate the impact of the distortion on the structure; distortions on small cells create forces on all atoms as well as their periodic images.

Given an input structure, AFLOW creates the full set of distorted supercell structures for the calculation of the forces. To minimize the number of expensive DFT calculations (primary computational bottleneck), AFLOW-SYM (see Sect. 2.3) is employed to determine which distortions are symmetrically equivalent using the site symmetry. Only inequivalent distortions are applied and explicitly calculated. Symmetry is then used to appropriately construct the IFC matrix, from which the dynamical matrix is constructed and the phonon modes, energies, and group velocities are derived.

APL has been extended to include the calculation of quasi-harmonic (quasi-harmonic approximation APL, QHA-APL Nath et al. 2016, 2017) and anharmonic (Automatic Anharmonic Phonon Library, AAPL Plata et al. 2017) effects in order to obtain properties such as the heat capacity at constant pressure C_p, coefficient of volumetric thermal expansion α_V, and lattice thermal conductivity κ_L.

QHA-APL performs harmonic APL calculations at multiple different volumes and extracts the Grüneisen parameter from the change of the phonon frequencies with respect to volume:

$$\gamma_i = -\frac{V}{\omega_i}\frac{\partial \omega_i}{\partial V}. \tag{9}$$

The Grüneisen parameter can be used in combination with harmonic properties such as C_V to calculate C_p, α_V (Nath et al. 2016), and κ_L (Nath et al. 2017).

AAPL obtains the third-order anharmonic IFCs by distorting two atoms in a supercell structure at a time as depicted in Fig. 4e and then calculating the change in forces on the other atoms (Plata et al. 2017). These IFCs are used to calculate the three-phonon scattering rates, and thus the scattering time and mean free displacement. These quantities are combined with the group velocities obtained from harmonic APL to solve the Boltzmann transport equation and calculate κ_L with quantitative accuracy (Plata et al. 2017).

2.8 AFLOW: Visualization Tools

AFLOW leverages a panoply of visualization tools for materials data, including standard software such as `gnuplot`, `latex`, and `xmgrace` for plots of phonon dispersions, electronic band structures, electronic density of states, and convex hull visualization. These plots are served publicly through the AFLOW.org repository.

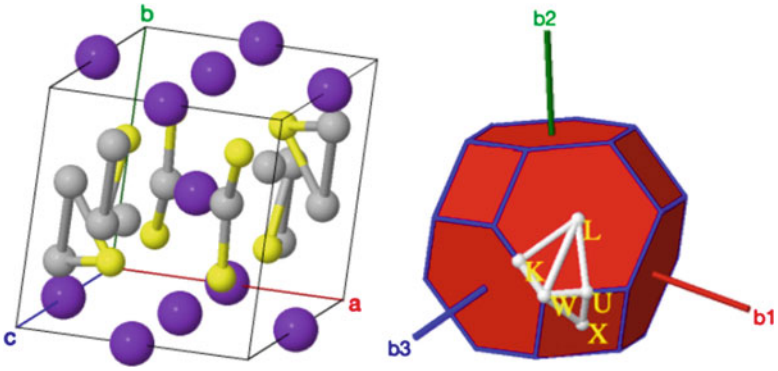

Fig. 5 Side-by-size visualization of the crystal structure and Brillouin zone using Jmol. The structure highlighted is Ag_3KS_2 (ICSD #73581): http://aflow.org/material.php?id=Ag6K2S4_ICSD_73581. The AFLOW Standard path of high-symmetry **k**-points is illustrated in the Brillouin zone (Setyawan and Curtarolo 2010)

To visualize crystal structures, AFLOW employs the `Jmol` software, which has incorporated substantial functionality for AFLOW-specific application. The `JSmol` branch of the software powers the online crystal structure visualizations in the AFLOW.org repository entry pages and AFLOW Prototype Library pages. With its recently added POSCAR reader, `JSmol` provides an assortment of capabilities ranging from different view perspectives, supercell expansions, and varying unit cell representations. A similar visualization application showing the AFLOW Standard high-symmetry paths in the Brillouin zone (Setyawan and Curtarolo 2010) is currently being incorporated, as illustrated in Fig. 5. Additionally, the `Jmol` desktop client offers a specialized macro (`aflow`) for visualization of alloy systems, which leverages the AFLUX Search-API.

3 AFLOWπ: Minimalist High-Throughput

The AFLOWπ (Supka et al. 2017) framework has been originally implemented as a minimalist software to perform verification tasks (see Sect. 5.7) on data published on AFLOW.org. By design, AFLOWπ is easy to install and to extend to a variety of electronic structure codes (currently only the QUANTUM ESPRESSO (Giannozzi et al. 2009, 2017) DFT package is implemented). AFLOWπ builds on the versatility of Python, providing a module to prepare, run, and analyze large sets of first-principles calculations and includes tools for the automatic projection on pseudo-atomic orbitals (PAO; see Sect. 4) and the self-consistent calculation of Hubbard U corrections within the Agapito, Curtarolo and Buongiorno Nardelli (ACBN0) approach (Agapito et al. 2015; Andrade et al. 2015). In addition, workflows for the calculation of elastic constants, diffusive transport coefficients, optical spectra, and phonon dispersions with DFT+U (see Fig. 6a for assessing the

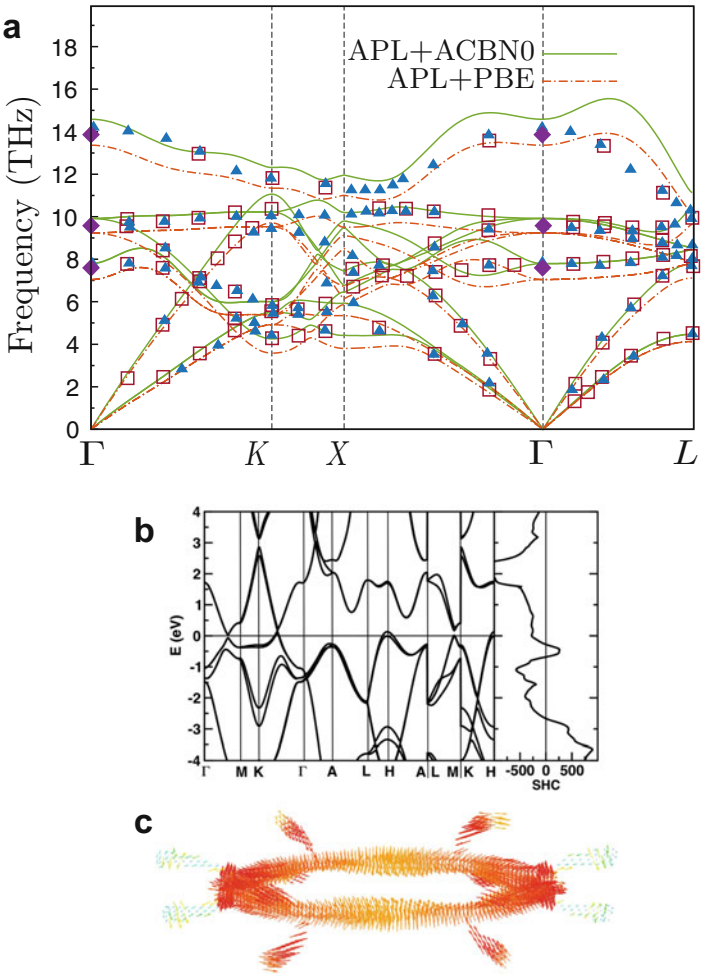

Fig. 6 Vibrational spectrum calculated with AFLOWπ and AFLOW-APL (left) and electronic properties computed with PAOFLOW (right). (**a**) Phonon dispersion of CaF_2 calculated with APL (Plata et al. 2017), using the ACBN0 method as implemented within AFLOWπ (green lines). The results obtained using PBE are shown by the broken orange lines for comparison. The blue triangles and red unfilled squares represent neutron scattering data from Schmalzl et al. (2003) and Elcombe and Pryor (1970), respectively, while the purple diamonds represent Raman and infrared data from Kaiser et al. (1962). (**b**) Electronic band structure, spin Hall conductivity (SHC), and (**c**) spin texture of the nodal line and Weyl points in HfC, as calculated using the PAOFLOW utility

effect of the Hubbard U corrections on the phonon dispersion calculated using APL) are included. When possible, AFLOWπ exploits the tight-binding hamiltonians as in D'Amico et al. (2016). Calculation results can be easily packaged and prepared for incorporation into the AFLOW.org data repository (see Sect. 5).

4 PAOFLOW: Fast Characterization

PAOFLOW (Buongiorno Nardelli et al. 2017) is a stand-alone tool to efficiently post-process standard DFT pseudo-potential plane-wave calculations to generate tight-binding (TB) Hamiltonians which faithfully reproduce the calculated electronic structure (eigenvalues and eigenvectors) with arbitrary precision in reciprocal space (Agapito et al. 2013, 2016a,b) (see Fig. 6b for PAOFLOW generated band structure for HfC (ICSD #169399, space group #187, AFLOW prototype: `AB_hP2_187_d_a`)). By exploiting the simplicity of the TB formalism and the efficiency of fast Fourier transforms, PAOFLOW interpolates the band structure and computes the matrix elements of the momentum operator, $\mathbf{p}_{m,n}$. These are used to improve the quality of integrated quantities such as the density of states (adaptive smearing), to compute electronic transport coefficients within the constant relaxation time approximation, and to compute the dielectric constants (D'Amico et al. 2016). In addition, the $\mathbf{p}_{m,n}$ matrix elements facilitate the calculation of the Berry curvature and related properties (anomalous Hall conductivity, spin Hall conductivity (see Fig. 6b), magnetic circular dichroism, spin circular dichroism; see spin texture of the nodal line and Weyl points in HfC shown in Fig. 6c). Starting from a well-interpolated band structure, it is also possible to compute topological invariants.

Because of the local representation of the electronic structure provided by the PAOFLOW software, surface-projected band structure and Landauer ballistic transport are also computable within PAOFLOW.

The software is implemented in Python, is portable and easy to install, and is parallel by design (on both CPUs and GPUs). PAOFLOW is also an integral part of the AFLOWπ framework.

5 AFLOW: Data Repository

The AFLOW data repository (Curtarolo et al. 2012b) contains the calculated properties for over 1.8 million materials entries, obtained using the AFLOW framework. These properties are available through the `aflow.org` web portal, which includes online search/sort and data analysis applications. The repository is programmatically accessible through the AFLOW Data REST-API (Taylor et al. 2014) and the AFLUX Search-API (Rose et al. 2017).

5.1 AFLOW: Web Portal

The AFLOW repository (Curtarolo et al. 2012b) is available online via the `aflow.org` web portal (Fig. 7a). It contains multiple online applications for data access, processing, and visualization, including the advanced "MendeLIB" search application at `http://aflow.org/advanced.php` which facilitates searching for materials entries with filters for elemental composition and calculated properties (see Fig. 7b), the interactive convex hull application at

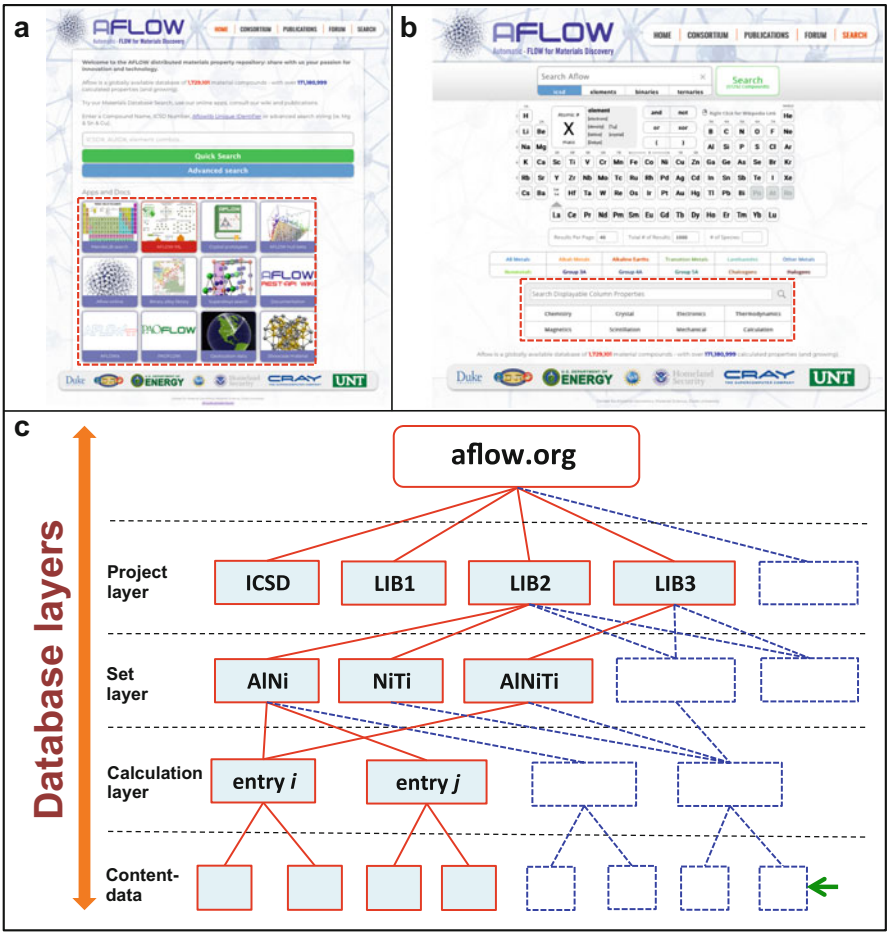

Fig. 7 AFLOW web portal and data repository. (**a**) Online applications and documentation are accessible via the "Apps and Docs" set of buttons surrounded by the dashed red rectangle. (**b**) The advanced search application can be used to search for specific compositions and also includes property search filters, as highlighted by the dashed red rectangle. (**c**) The AFLOW data repository is organized into project, set (i.e., alloy system), and calculation (i.e., materials entry) layers

`http://aflow.org/aflow-chull` (see Sect. 2.4), the online machine learning model at `http://aflow.org/aflow-ml` (see Sect. 5.2), and the AFLOW online tool at `http://aflow.org/aflow_online.html` which gives access to AFLOW crystal structure analysis and processing functions. These applications and more are accessible from the main `aflow.org` web page, via the "Apps and Docs" set of buttons highlighted by the dashed red rectangle in Fig. 7a.

5.2 AFLOW-ML: Online Machine Learning

The AFLOW machine learning (AFLOW-ML) online application provides a user interface to leverage machine learning models trained on AFLOW data. It accepts a standard structure file (POSCAR or QE) and outputs predictions for properties such as the band gap, elastic moduli, heat capacity, Debye temperature, vibrational free energy, and thermal expansion coefficient. Additionally, structures within the AFLOW repository can be imported via the sidebar. This application provides an accessible medium to retrieve machine learning predictions without the need to install a software library or machine learning package.

Currently, AFLOW-ML supports two different machine learning models. The first model, property-labeled materials fragments (Isayev et al. 2017), `plmf`, has been trained using data from the AFLOW repository and predicts properties such as the electronic band gap, specific heat capacities, and bulk/shear moduli. The second method is the molar fraction descriptor model (Legrain et al. 2017), `mfd`, which predicts vibrational properties such as vibrational free energy and entropy and is based only on the chemical composition of the material.

The AFLOW-ML API (Gossett et al. 2018) offers programmatic access to the AFLOW-ML online application and provides a simplified abstraction that facilitates leveraging powerful machine learning models. This distills the prediction process down to its essence: from a structure file, return a prediction. Using the API is a two-step process: first a structure file, in POSCAR 5 format (structure input for version 5 of VASP), is posted (i.e., uploaded) to the endpoint `/<model>/prediction` on the `aflow.org` server using standard HTTP libraries or dedicated programs such `curl` or `wget`, where `<model>` specifies the machine learning model to use in the prediction (current options: `plmf` and `mfd`). When a prediction is submitted, a JSON response object is returned that includes a task id. The results of the prediction are then retrieved from the `/prediction/result/` endpoint on the `aflow.org` server by appending the task id to the end of the URL, i.e., `/prediction/result/{id}/`. This endpoint monitors the prediction task and responds with a JSON object that details its status. When complete, the endpoint responds with the results of the prediction, represented as a JSON object containing a key-value pair for each predicted property.

5.3 AFLOW: Database Organization

The AFLOW data repository (Curtarolo et al. 2012b) is organized into project, set, and calculation layers as illustrated in Fig. 7c. At the project layer, the calculations are divided into different catalogs based on the origin and composition of the entries (Taylor et al. 2014; Rose et al. 2017). Within each catalog, entries are grouped into sets based on shared lattice type or alloy system. The entries within each set contain the results of DFT calculated properties for particular structures.

The AFLOW-ICSD catalog contains the DFT calculated properties for over 57,000 experimentally observed materials listed in the Inorganic Crystal Structure Database (ICSD) (Bergerhoff et al. 1983; Karen and Hellenbrandt 2002). Internally, this catalog is organized by lattice type and then by individual materials entry. Since the materials in this catalog are already known to exist, the primary interest is in accurately calculating electronic structure and thermomechanical properties. Therefore, calculations for this catalog are generally performed using the Hubbard U correction to the DFT exchange-correlation functional (Liechtenstein et al. 1995; Dudarev et al. 1998) where appropriate, using a set of standardized U values (Calderon et al. 2015). Within this catalog, entries are grouped by Bravais lattice type into 14 sets: "BCC", "BCT", "CUB", "FCC", "HEX'", "MCL"', "MCLC"', "ORC"', "ORCC"', "ORCF"', "ORCI"', "RHL"', "TET"' and "TRI"'. The name of each materials entry is generated using the format `<composition>_ICSD_<ICSD number>`.

The entries in the other catalogs, such as "LIB1," "LIB2," and "LIB3", are generated by decorating crystal structure prototypes to predict new hypothetical compounds and contain unary, binary, and ternary materials, respectively. Additional catalogs, "LIB4," "LIB5," and "LIB6," are currently being generated for quaternary, quinary, and hexenary materials. Within each catalog, the entries are grouped by element and exchange-correlation functional in the case of "LIB1" and by alloy system in the cases of "LIB2" and "LIB3." "LIB1" contains a total of 3068 entries, while "LIB2" currently has 329,192 entries and "LIB3" has over 1.4 million. Within each alloy system, the individual materials entries are named according to the relevant crystal prototype. For these catalogs, the emphasis is on the discovery of new thermodynamically stable or metastable materials and on their use to generate the thermodynamic density of states for the prediction of the formation of disordered materials such as metallic glasses (Perim et al. 2016) or high-entropy alloys (Lederer et al. 2018). Therefore, calculations in these catalogs are performed using the GGA-PBE exchange-correlation functional (Perdew et al. 1996) without Hubbard U corrections (Calderon et al. 2015) so as to produce consistent energy differences, enabling the calculation of accurate formation enthalpies.

5.4 AFLOW: Database Properties

Materials properties within the AFLOW repository (Curtarolo et al. 2012b) are indexed as keyword-value pairs which are programmatically accessible via the AFLOW Data REST-API (Taylor et al. 2014) and programmatically searchable via the AFLUX Search-API (Rose et al. 2017). Search filters for these properties are also available in the advanced search application of the `aflow.org` web portal as highlighted by the dashed red rectangle in Fig. 7b, where they are grouped into chemistry (e.g., chemical species, stoichiometry), crystal (e.g., space group, Bravais lattice type), electronic (e.g., band gaps), thermodynamic (energetic and thermal properties, e.g., formation enthalpies and Debye temperatures), magnetic, scintillation, mechanical (elastic moduli and pressure-related properties), and calculation (e.g., **k**-point mesh, AFLOW version) parameters. In total there are in excess of 170

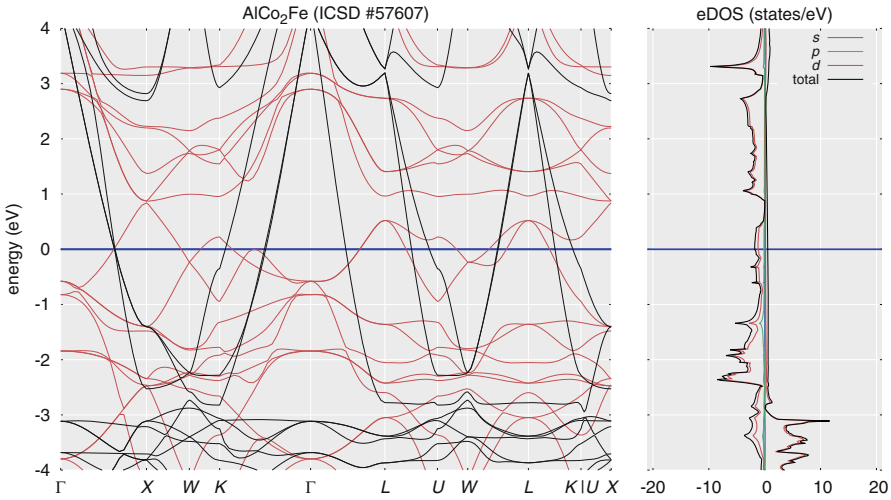

Fig. 8 Example band structure and density of states images automatically generated and served through the AFLOW.org data repository. The structure highlighted is $AlCo_2Fe$ (ICSD #57607): http://aflow.org/material.php?id=Al1Co2Fe1_ICSD_57607. The results of the spin-polarized calculation are differentiated by color on the band structure plot (black/red for majority/minority spin) and sign on the density of states plot (positive/negative for majority/minority spin). The band structure is calculated following the AFLOW Standard path of high-symmetry **k**-points (Setyawan and Curtarolo 2010)

million individual materials properties indexed in the AFLOW database (∼100 per materials entry). Lists of the keywords corresponding to the materials properties are provided in Taylor et al. (2014), Rose et al. (2017), and Toher et al. (2017).

Systems for which the "STATIC" and "BANDS" calculations have been performed are supplemented with automatically generated images of the density of states, projected density of states, and band structure. Both low (PNG)- and high (EPS)-quality variants of the images are available for download. An example band structure and density of states image is displayed in Fig. 8.

5.5 AFLOW: Data REST-API

The full data set generated by the high-throughput AFLOW process (Curtarolo et al. 2012b) is backed by a disk store of (at this time) over 12 TB of input criteria, calculated results, and derivative output. The backing store is exposed via the AFLOW Data REST-API (Curtarolo et al. 2012b) in a hierarchical organization. This direct exposure of our results not only grants the end user a high degree of utility via direct access, but, more importantly, guarantees data provenance that promotes reproducibility. The hierarchy of the AFLOW Data REST-API categorizes this abundance of information into meaningful high-level classifications allowing

for exploration of self-similar materials that are related by stoichiometric and/or crystallographic properties. Once a selection of materials has been determined, the full range of available properties and procedural data are retrievable.

The organizational hierarchy of both the underlying data store and the REST-API is project dependent, as described in Sect. 5.3. Each project is equivalent to one of the catalogs listed in Sect. 5.3 and in the REST-API is denoted by the project layers "ICSD_WEB," "LIB1_RAW," "LIB2_RAW," and "LIB3_RAW." Each project layer contains multiple set layers, which correspond to specific alloy systems in the case of "LIB1_RAW," "LIB2_RAW," and "LIB3_RAW." For instance, `http://aflowlib.duke.edu/AFLOWDATA/LIB2_RAW/` exposes the set layer for binary entries, where each set corresponds to different binary alloy systems, allowing for pairwise atomic species examination. Within each set is the entry layer, consisting of decorated structural prototypes which provide stoichiometric and structural variation for each alloy system. Each entry contains the calculated results for a particular structure and composition, organized as keyword-value pairs. The calculated values of thermodynamical, mechanical, electronic, magnetic, chemical, and crystallographic properties can be directly accessed by querying a Uniform Resource Identifier (URI) of the form `<server>/<project>/<set>/<entry>/?<keyword>`, where `<server>` is `http://aflowlib.duke.edu/AFLOWDATA`, `<project>` is the appropriate project layer, `<set>` is the alloy system, `<entry>` is the structural prototype, and `<keyword>` corresponds to the materials property of interest. A full description of the REST-API keywords is provided in Taylor et al. (2014), along with additions in the appendices of Rose et al. (2017) and Toher et al. (2017).

The ability to explore related entries predicated on a multitude of properties leads directly to novel materials discovery and use. The AFLOW Data REST-API disseminates our methods and results, without restriction, to a global research audience in order to promote scientific and engineering advancement.

5.6 AFLUX: Search-API

The Automatic Flow of LUX or AFLUX Search-API (Rose et al. 2017) is a human usable remote data search API. LUX is designed to be a domain agnostic solution to the outstanding problem of programmatically searching remote data that typically is either exposed via a capriciously limited utility or requires a-priori knowledge of the internal organization of the remote repository. The LUX query concept flattens the exposed data, while simultaneously providing arbitrarily complex query capability, allowing an end user full freedom in constraining the requested data. LUX is designed to operate in the nearly ubiquitous web URI context while minimizing any potentially conflicting interactions with existing URI functionality.

AFLUX is the domain-specific implementation of LUX and is available at the `<AFLUX-URI>`: `http://aflowlib.duke.edu/search/API/?`. At this time, the AFLUX API freely exposes over 180 million keyword-value properties without any requirements or restrictions on the end user. Specific properties and

compositions can be searched for by appending the appropriate keywords to the `<AFLUX-URI>`. Search results can be restricted by including specific values or value ranges in parentheses after the appropriate keyword. For example, a search can be restricted to entries that contain both of the elements Na and Cl, and have a calculated electronic band gap in excess of 1.0eV, by including the search parameters `species(Na,Cl),Egap(1.0*)` in the query part of the URI. In LUX, "," corresponds to the logical AND operator, and "*" is the `loose` operator which extends the search to entries in a specified value range. If no parameters are provided for a particular keyword, then the values of that property are returned for all entries which satisfy the remaining search criteria. A full list of all LUX logical operators can be found in Rose et al. (2017), along with descriptions of their functionality and appropriate usage.

In addition to materials properties keywords, LUX also accepts directives, which behave as pseudo property keywords. They are used to provide additional information on LUX usage and control the format and quantity of the returned data. Note that any directives included in a search query must come after all of the materials properties keywords. In particular, the `schema` directive can be used to retrieve the most current and canonical list of keywords using the AFLUX summons: `<AFLUX-URI>schema,format(json)`.

5.7 AFLOW: Data Quality Control

Data quality control, including validation of methodologies and verification of calculated data, is vital when constructing large databases such as the AFLOW repository (Curtarolo et al. 2012b) in order to guarantee the reliability of the results. Methodological validation involves quantifying the accuracy of calculation models with respect to experiment, while data verification includes checking the robustness of calculation parameters and the satisfaction of convergence criteria.

Physical models incorporated into the AFLOW framework are validated by comparison to benchmark sets of experimental data. This helps determine the predictive accuracy of the methods for real materials, as well as the regimes in which they are reliable. For example, the AEL and AGL modules were validated by comparison to a benchmark set of ∼75 experimentally well-characterized compounds of various structural types (Toher et al. 2014, 2017), and the accuracy was quantified by the Pearson and Spearman correlations and the root-mean-square deviations. Similar validation analyses were performed for the QHA-APL (Nath et al. 2016, 2017) and AAPL methods (Plata et al. 2017), as well as the property labeled materials fragments machine learning model (Isayev et al. 2017).

The AFLOW-POCC methodology has been validated by comparing the band gap as a function of composition for $ZnS_{1-x}Se_x$ and $Mg_xZn_{1-x}O$, and the magnetic moment per atom as a function of composition for $Fe_{1-x}Cu_x$, to experimental values (Yang et al. 2016).

The ACBN0 functional (Agapito et al. 2015), implemented in AFLOWπ (see Sect. 3) and PAOFLOW (see Sect. 4), has been validated by comparing the lattice

parameters, bulk moduli, electronic band gaps, phonon modes, high-frequency dielectric constants, and Born effective changes it produces to the experimentally measured values for the Zn and Cd chalcogenides (Gopal et al. 2015).

The convergence of both the charge density optimization and the ionic structural relaxation are automatically verified for all AFLOW calculations prior to incorporation into the data repository. This includes, for example, checking that the charge density has converged in accordance with the AFLOW Standard settings (Calderon et al. 2015) and verifying the relaxation of the cell size and shape by ensuring that all elements of the stress tensor are less than 10kB. The convergence level for any individual calculation can be verified by querying appropriate keywords for the stress tensor: `stress_tensor`, Pulay stress: `Pulay_stress`, residual external pressure on the relaxed cell: `pressure_residual`, and the δE value for the final electronic convergence step: `delta_electronic_energy_convergence`, using the AFLOW Data REST-API (Taylor et al. 2014) or the AFLUX Search-API (Rose et al. 2017). Initial calculation parameters can similarly be obtained using the appropriate keywords, such as the k-point grid, `kpoints`, or the electronic energy convergence threshold, `delta_electronic_energy_threshold`.

Conclusion

The AFLOW Fleet for computational materials design automates first-principles calculations of materials properties. AFLOW incorporates a wide range of different modules, including applications for symmetry and thermodynamic stability analysis, generation of ordered and disordered materials structures, and calculation of thermomechanical properties, in a single integrated framework. AFLOWπ is a versatile minimalist framework that includes tools for projection onto pseudo-atomic orbitals (PAO) and the self-consistent calculation of Hubbard U corrections using ACBN0. PAOFLOW generates tight-binding Hamiltonians which reproduce the electronic structure calculated using first-principles methods, facilitating the rapid calculation of electronic and magnetic properties such as transport coefficients and the Berry curvature. All results are stored in, and disseminated through, the AFLOW data repository, which is available online at `aflow.org` and is programmatically accessible via the AFLOW Data REST-API and the AFLUX Search-API.

Acknowledgements The authors acknowledge support from DOD-ONR (N00014-13-1-0030, N00014-13-1-0635, N00014-17-1-2090, N00014-16-1-2781, N00014-15-1-2583, N00014-15-1-2266), DOE (DE-AC02-05CH11231, specifically BES Grant # EDCBEE), and the Duke University Center for Materials Genomics. SC acknowledges support by the Alexander von Humboldt-Foundation – Max Planck Society (Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin-Dahlem, Germany). CO acknowledges support from the National Science Foundation Graduate Research Fellowship under Grant No. DGF-1106401. AFLOW calculations were performed at the Duke University Center for Materials Genomics and at the Fulton Supercomputer

Lab – Brigham Young University. The authors thank Amir Natan, Matthias Scheffler, Luca Ghiringhelli, Kenneth Vecchio, Don Brenner, and Jon-Paul Maria for helpful discussions.

References

Agapito LA, Ferretti A, Calzolari A, Curtarolo S, Buongiorno Nardelli M (2013) Effective and accurate representation of extended Bloch states on finite Hilbert spaces. Phys Rev B 88:165127

Agapito LA, Curtarolo S, Buongiorno Nardelli M (2015) Reformulation of DFT + U as a pseudohybrid Hubbard density functional for accelerated materials discovery. Phys Rev X 5:011006

Agapito LA, Fornari M, Ceresoli D, Ferretti A, Curtarolo S, Buongiorno Nardelli M (2016a) Accurate tight-binding Hamiltonians for two-dimensional and layered materials. Phys Rev B 93:125137

Agapito LA, Ismail-Beigi S, Curtarolo S, Fornari M, Buongiorno Nardelli M (2016b) Accurate tight-binding Hamiltonian matrices from *ab initio* calculations: minimal basis sets. Phys Rev B 93:035104

Andrade X, Strubbe D, De Giovannini U, Larsen AH, Oliveira MJT, Alberdi-Rodriguez J, Varas A, Theophilou I, Helbig N, Verstraete MJ, Stella L, Nogueira F, Aspuru-Guzik A, Castro A, Marques MAL, Rubio A (2015) Real-space grids and the Octopus code as tools for the development of new simulation approaches for electronic systems. Phys Chem Chem Phys 17:31371–31396

Bahn SR, Jacobsen KW (2002) An object-oriented scripting interface to a legacy electronic structure code. Comput Sci Eng 4:56–66

Barber CB, Dobkin DP, Huhdanpaa H (1996) The quickhull algorithm for convex hulls. ACM Trans Math Soft 22:469–483

Barzilai S, Toher C, Curtarolo S, Levy O (2016) Evaluation of the tantalum-titanium phase diagram from *ab-initio* calculations. Acta Mater 120:255–263

Barzilai S, Toher C, Curtarolo S, Levy O (2017a) The effect of lattice stability determination on the computational phase diagrams of intermetallic alloys. J Alloys Compd 728:314–321

Barzilai S, Toher C, Curtarolo S, Levy O (2017b) Molybdenum-titanium phase diagram evaluated from *ab initio* calculations. Phys Rev Mater 1:023604

Bergerhoff G, Hundt R, Sievers R, Brown ID (1983) The inorganic crystal structure data base. J Chem Inf Comput Sci 23:66–69

Blanco MA, Martín-Pendás A, Francisco E, Recio JM, Franco R (1996) Thermodynamical properties of solids from microscopic theory: applications to MgF_2 and Al_2O_3. J Mol Struct Theochem 368:245–255

Blanco MA, Francisco E, Luaña V (2004) GIBBS: isothermal-isobaric thermodynamics of solids from energy curves using a quasi-harmonic Debye model. Comput Phys Commun 158:57–72

Blöchl PE (1994) Projector augmented-wave method. Phys Rev B 50:17953–17979

Blum V, Gehrke R, Hanke F, Havu P, Havu V, Ren X, Reuter K, Scheffler M (2009) Ab initio molecular simulations with numeric atom-centered orbitals. Comput Phys Commun 180:2175–2196

Calderon CE, Plata JJ, Toher C, Oses C, Levy O, Fornari M, Natan A, Mehl MJ, Hart GLW, Buongiorno Nardelli M, Curtarolo S (2015) The AFLOW standard for high-throughput materials science calculations. Comput Mater Sci 108 Part A:233–238

Curtarolo S, Setyawan W, Hart GLW, Jahnátek M, Chepulskii RV, Taylor RH, Wang S, Xue J, Yang K, Levy O, Mehl MJ, Stokes HT, Demchenko DO, Morgan D (2012a) AFLOW: an automatic framework for high-throughput materials discovery. Comput Mater Sci 58:218–226

Curtarolo S, Setyawan W, Wang S, Xue J, Yang K, Taylor RH, Nelson LJ, Hart GLW, Sanvito S, Buongiorno Nardelli M, Mingo N, Levy O (2012b) AFLOWLIB.ORG: a distributed materials properties repository from high-throughput *ab initio* calculations. Comput Mater Sci 58:227–235

Curtarolo S, Hart GLW, Buongiorno Nardelli M, Mingo N, Sanvito S, Levy O (2013) The high-throughput highway to computational materials design. Nat Mater 12:191–201

D'Amico P, Agapito LA, Catellani A, Ruini A, Curtarolo S, Fornari M, Buongiorno Nardelli M, Calzolari A (2016) Accurate *ab initio* tight-binding Hamiltonians: effective tools for electronic transport and optical spectroscopy from first principles. Phys Rev B 94:165166

de Jong M, Chen W, Angsten T, Jain A, Notestine R, Gamst A, Sluiter M, Ande CK, van der Zwaag S, Plata JJ, Toher C, Curtarolo S, Ceder G, Persson KA, Asta MD (2015) Charting the complete elastic properties of inorganic crystalline compounds. Sci Data 2:150009

Dudarev SL, Botton GA, Savrasov SY, Humphreys CJ, Sutton AP (1998) Electron-energy-loss spectra and the structural stability of Nickel oxide: an LSDA+U study. Phys Rev B 57:1505–1509

Elcombe MM, Pryor AW (1970) The lattice dynamics of calcium fluoride. J Phys C Solid State Phys 3:492

Giannozzi P, Baroni S, Bonini N, Calandra M, Car R, Cavazzoni C, Ceresoli D, Chiarotti GL, Cococcioni M, Dabo I, Dal Corso A, de Gironcoli S, Fabris S, Fratesi G, Gebauer R, Gerstmann U, Gougoussis C, Kokalj A, Lazzeri M, Martin-Samos L, Marzari N, Mauri F, Mazzarello R, Paolini S, Pasquarello A, Paulatto L, Sbraccia C, Scandolo S, Sclauzero G, Seitsonen AP, Smogunov A, Umari P, Wentzcovitch RM (2009) QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. J Phys Condens Matter 21:395502

Giannozzi P, Andreussi O, Brumme T, Bunau O, Buongiorno Nardelli M, Calandra M, Car R, Cavazzoni C, Ceresoli D, Cococcioni M, Colonna N, Carnimeo I, Dal Corso A, de Gironcoli S, Delugas P, DiStasio Jr RA, Ferretti A, Floris A, Fratesi G, Fugallo G, Gebauer R, Gerstmann U, Giustino F, Gorni T, Jia J, Kawamura M, Ko H-Y, Kokalj A, Küçükbenli E, Lazzeri M, Marsili M, Marzari N, Mauri F, Nguyen NL, Nguyen H-V, Otero-de-la-Roza A, Paulatto L, Poncé S, Rocca D, Sabatini R, Santra B, Schlipf M, Seitsonen AP, Smogunov A, Timrov I, Thonhauser T, Umari P, Vast N, Wu X, Baroni S (2017) Advanced capabilities for materials modelling with quantum ESPRESSO. J Phys Condens Matter 29:465901

Gonze X, Beuken JM, Caracas R, Detraux F, Fuchs M, Rignanese GM, Sindic L, Verstraete M, Zerah G, Jollet F, Torrent M, Roy A, Mikami M, Ghosez P, Raty JY, Allan DC (2002) First-principles computation of material properties: the ABINIT software project. Comput Mater Sci 25:478–492

Gopal P, Fornari M, Curtarolo S, Agapito LA, Liyanage LSI, Buongiorno Nardelli M (2015) Improved predictions of the physical properties of Zn- and Cd-based wide band-gap semiconductors: a validation of the ACBN0 functional. Phys Rev B 91:245202

Gossett E, Toher C, Oses C, Isayev O, Legrain F, Rose F, Zurek E, Carrete J, Mingo N, Tropsha A, Curtarolo S (2018) AFLOW-ML: a RESTful API for machine-learning predictions of materials properties. Comput Mater Sci 152:134–145. Submitted arxiv.org/1711.10744

Hart GLW, Forcade RW (2008) Algorithm for generating derivative structures. Phys Rev B 77:224115

Hicks D, Oses C, Gossett E, Gomez G, Taylor RH, Toher C, Mehl MJ, Levy O, Curtarolo S (2018) AFLOW-SYM: platform for the complete, automatic and self-consistent symmetry analysis of crystals. Acta Cryst A 74:184–203

Isayev O, Fourches D, Muratov EN, Oses C, Rasch K, Tropsha A, Curtarolo S (2015) Materials cartography: representing and mining materials space using structural and electronic fingerprints. Chem Mater 27:735–743

Isayev O, Oses C, Toher C, Gossett E, Curtarolo S, Tropsha A (2017) Universal fragment descriptors for predicting electronic properties of inorganic crystals. Nat Commun 8:15679

Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G, Persson KA (2013) Commentary: the materials project: a materials genome approach to accelerating materials innovation. APL Mater 1:011002

Kaiser W, Spitzer WG, Kaiser RH, Howarth LE (1962) Infrared Properties of CaF_2, SrF_2, and BaF_2. Phys Rev 127:1950–1954

Karen VL, Hellenbrandt M (2002) Inorganic crystal structure database: new developments. Acta Cryst A58:c367

Kresse G, Furthmüller J (1996) Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. Phys Rev B 54:11169–11186

Kresse G, Hafner J (1993) *Ab initio* molecular dynamics for liquid metals. Phys Rev B 47:558–561

Landis DD, Hummelshøj JS, Nestorov S, Greeley J, Dułak M, Bligaard T, Nørskov JK, Jacobsen KW (2012) The computational materials repository. Comput Sci Eng 14:51–57

Lederer Y, Toher C, Vecchio KS, Curtarolo S (2018) The search for high entropy alloys: a high-throughput *ab-initio* approach. Acta Mater 159:354–383. Submitted arxiv.org/1711.03426

Legrain F, Carrete J, van Roekeghem A, Curtarolo S, Mingo N (2017) How chemical composition alone can predict vibrational free energies and entropies of solids. Chem Mater 29:6220–6227

Levy O, Hart GLW, Curtarolo S (2010a) Uncovering compounds by synergy of cluster expansion and high-throughput methods. J Am Chem Soc 132:4830–4833

Levy O, Hart GLW, Curtarolo S (2010b) Hafnium binary alloys from experiments and first principles. Acta Mater 58:2887–2897

Levy O, Chepulskii RV, Hart GLW, Curtarolo S (2010c) The new face of Rhodium alloys: revealing ordered structures from first principles. J Am Chem Soc 132:833–837

Liechtenstein AI, Anisimov VI, Zaanen J (1995) Density-functional theory and strong interactions: orbital ordering in Mott-Hubbard insulators. Phys Rev B 52:R5467–R5470

Maradudin AA, Montroll EW, Weiss GH, Ipatova IP (1971) Theory of lattice dynamics in the harmonic approximation. Academic Press, New York

Mehl MJ, Hicks D, Toher C, Levy O, Hanson RM, Hart GLW, Curtarolo S (2017) The AFLOW library of crystallographic prototypes: part 1. Comput Mater Sci 136:S1–S828

Buongiorno Nardelli M, Cerasoli FT, Costa M, Curtarolo S, De Gennaro R, Fornari M, Liyanage L, Supka AR, Wang H (2017) PAOFLOW: a utility to construct and operate on ab initio Hamiltonians from the projections of electronic wavefunctions on atomic orbital bases, including characterization of topological materials. Comput Mater Sci 143:462–472

Nath P, Plata JJ, Usanmaz D, Al Rahal Al Orabi R, Fornari M, Buongiorno Nardelli M, Toher C, Curtarolo S (2016) High-throughput prediction of finite-temperature properties using the quasi-harmonic approximation. Comput Mater Sci 125:82–91

Nath P, Plata JJ, Usanmaz D, Toher C, Fornari M, Buongiorno Nardelli M, Curtarolo S (2017) High throughput combinatorial method for fast and robust prediction of lattice thermal conductivity. Scr Mater 129:88–93

Nyshadham C, Oses C, Hansen JE, Takeuchi I, Curtarolo S, Hart GLW (2017) A computational high-throughput search for new ternary superalloys. Acta Mater 122:438–447

Ong SP, Richards WD, Jain A, Hautier G, Kocher M, Cholia S, Gunter D, Chevrier VL, Persson KA, Ceder G (2013) Python materials genomics (pymatgen): a robust, open-source python library for materials analysis. Comput Mater Sci 68:314–319

Oses C, Gossett E, Hicks D, Rose F, Mehl MJ, Perim E, Takeuchi I, Sanvito S, Scheffler M, Lederer Y, Levy O, Toher C, Curtarolo S (2018, in press) AFLOW-CHULL: Cloud-oriented platform for automonous phase stability analysis. J Chem Inf Model. https://doi.org/10.1021/acs.jcim.8b00393

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865–3868

Perim E, Lee D, Liu Y, Toher C, Gong P, Li Y, Simmons WN, Levy O, Vlassak JJ, Schroers J, Curtarolo S (2016) Spectral descriptors for bulk metallic glasses based on the thermodynamics of competing crystalline phases. Nat Commun 7:12315

Pizzi G, Cepellotti A, Sabatini R, Marzari N, Kozinsky B (2016) AiiDA: automated interactive infrastructure and database for computational science. Comput Mater Sci 111:218–230

Plata JJ, Nath P, Usanmaz D, Carrete J, Toher C, de Jong M, Asta MD, Fornari M, Buongiorno Nardelli M, Curtarolo S (2017) An efficient and accurate framework for calculating lattice thermal conductivity of solids: AFLOW-AAPL automatic anharmonic phonon library. NPJ Comput Mater 3:45

Poirier J-P (2000) Introduction to the physics of the earths interior, 2nd edn. Cambridge University Press, Cambridge

Rappe AK, Casewit CJ, Colwell KS, Goddard WA, Skiff WM (1992) UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. J Am Chem Soc 114:10024–10035

Rose F, Toher C, Gossett E, Oses C, Buongiorno Nardelli M, Fornari M, Curtarolo S (2017) AFLUX: the LUX materials search API for the AFLOW data repositories. Comput Mater Sci 137:362–370

Saal JE, Kirklin S, Aykol M, Meredig B, Wolverton C (2013) Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). JOM 65:1501–1509

Sanvito S, Oses C, Xue J, Tiwari A, Zic M, Archer T, Tozman P, Venkatesan M, Coey JMD, Curtarolo S (2017) Accelerated discovery of new magnets in the Heusler alloy family. Sci Adv 3:e1602241

Scheffler M, Draxl C (2014) Computer center of the Max-Planck society, garching. The NoMaD Repository. http://nomad-repository.eu

Schmalzl K, Strauch D, Schober H (2003) Lattice-dynamical and ground-state properties of CaF_2 studied by inelastic neutron scattering and density-functional methods. Phys Rev B 68:144301

Setyawan W, Curtarolo S (2010) High-throughput electronic band structure calculations: challenges and tools. Comput Mater Sci 49:299–312

Supka AR, Lyons TE, Liyanage LSI, D'Amico P, Al Rahal Al Orabi R, Mahatara S, Gopal P, Toher C, Ceresoli D, Calzolari A, Curtarolo S, Buongiorno Nardelli M, Fornari M (2017) AFLOWπ: a minimalist approach to high-throughput *ab initio* calculations including the generation of tight-binding Hamiltonians. Comput Mater Sci 136:76–84

Taylor RH, Rose F, Toher C, Levy O, Yang K, Buongiorno Nardelli M, Curtarolo S (2014) A RESTful API for exchanging materials data in the AFLOWLIB.org consortium. Comput Mater Sci 93:178–192

Toher C, Plata JJ, Levy O, de Jong M, Asta MD, Buongiorno Nardelli M, Curtarolo S (2014) High-throughput computational screening of thermal conductivity, Debye temperature, and Grüneisen parameter using a quasiharmonic Debye model. Phys Rev B 90:174107

Toher C, Oses C, Plata JJ, Hicks D, Rose F, Levy O, de Jong M, Asta MD, Fornari M, Buongiorno Nardelli M, Curtarolo S (2017) Combining the AFLOW GIBBS and elastic libraries to efficiently and robustly screen thermomechanical properties of solids. Phys Rev Mater 1:015401

Yang K, Oses C, Curtarolo S (2016) Modeling off-stoichiometry materials with a high-throughput *ab-initio* approach. Chem Mater 28:6484–6492