Automated computation of materials properties

Cormac Toher, Corey Oses, and Stefano Curtarolo

1Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708, United States
2Materials Science, Electrical Engineering, Physics and Chemistry, Duke University, Durham North Carolina, 27708, United States

(Dated: May 15, 2018)

Materials informatics offers a promising pathway towards rational materials design, replacing the current trial-and-error approach and accelerating the development of new functional materials. Through the use of sophisticated data analysis techniques, underlying property trends can be identified, facilitating the formulation of new design rules. Such methods require large sets of consistently generated, programmatically accessible materials data. Computational materials design frameworks using standardized parameter sets are the ideal tools for producing such data. This work reviews the state-of-the-art in computational materials design, with a focus on these automated ab-initio frameworks. Features such as structural prototyping and automated error correction that enable rapid generation of large datasets are discussed, and the way in which integrated workflows can simplify the calculation of complex properties, such as thermal conductivity and mechanical stability, is demonstrated. The organization of large datasets composed of ab-initio calculations, and the tools that render them programmatically accessible for use in statistical learning applications, are also described. Finally, recent advances in leveraging existing data to predict novel functional materials, such as entropy stabilized ceramics, bulk metallic glasses, thermoelectrics, superalloys, and magnets, are surveyed.

1. INTRODUCTION

Materials informatics requires large repositories of materials data to identify trends in and correlations between materials properties, as well as for training machine learning models. Such patterns lead to the formulation of descriptors that guide rational materials design. Generating large databases of computational materials properties requires robust, integrated, automated frameworks [1]. Built-in error correction and standardized parameter sets enable the production and analysis of data without direct intervention from human researchers. Current examples of such frameworks include AFLOW (Automatic FLOW) [2–10], Materials Project [11–14], OQMD (Open Quantum Materials Repository) [15–17], the Computational Materials Repository [18] and its associated scripting interface ASE (Atomic Simulation Environment) [19], AiiDA (Automated Interactive Infrastructure and Database for Computational Science) [20–22], and the Open Materials Database at http://openmaterialsdb.se with its associated High-Throughput Toolkit (HTTK). Other computational materials science resources include the aggregated repository maintained by the Novel Materials Discovery (NoMaD) Laboratory [23], the Materials Mine database available at www.materials-mine.com, and the Theoretical Crystallography Open Database (TCOD) [24]. For this data to be consumable by automated machine learning algorithms, it must be organized in programmatically accessible repositories [4, 5, 7, 11, 12, 15, 23]. These frameworks also contain modules that combine and analyze data from various calculations to predict complex thermomechanical phenomena, such as lattice thermal conductivity and mechanical stability.

Computational strategies have already had success in predicting materials for applications including photovoltaics [25], water-splitters [26], carbon capture and gas storage [27, 28], nuclear detection and scintillators [29–32], topological insulators [33, 34], piezoelectrics [35, 36], thermoelectric materials [37–40], catalysis [41], and battery cathode materials [42–44]. More recently, computational materials data has been combined with machine learning approaches to predict electronic and thermomechanical properties [45, 46], and to identify superconducting materials [47]. Descriptors are also being constructed to describe the formation of disordered materials, and have recently been used to predict the glass forming ability of binary alloy systems [48]. These successes demonstrate that accelerated materials design can be achieved by combining structured data sets generated using autonomous computational methods with intelligently formulated descriptors and machine learning.

2. AUTOMATED COMPUTATIONAL MATERIALS DESIGN FRAMEWORKS

Rapid generation of materials data relies on automated frameworks such as AFLOW [2–6], Materials Project’s pymatgen [13] and atomate [14], OQMD [15–17], ASE [19], and AiiDA [21]. The general automated workflow is illustrated in Figure 1. These frameworks begin by creating the input files required by the electronic structure codes that perform the quantum-mechanics level calculations, where the initial geometry is generated by decorating structural prototypes (Figure 1(a, b)). They execute and monitor these calculations, reading any error messages written to the output files and diagnosing calculation failures. Depending on the nature of the errors, these frameworks are equipped with a catalog of prescribed solutions — enabling them to adjust the appropriate parameters and restart the calculations (Figure 1(c)). At the end of a successful calculation, the frameworks parse the output files to extract the relevant materials data.
such as total energy, electronic band gap, and relaxed cell volume. Finally, the calculated properties are organized and formatted for entry into machine-accessible, searchable and sortable databases.

In addition to running and managing the quantum-mechanics level calculations, the frameworks also maintain a broad selection of post-processing libraries for extracting additional properties, such as calculating x-ray diffraction (XRD) spectra from relaxed atomic coordinates, and the formation enthalpies for the convex hull analysis to identify stable compounds (Figure 1(d)). Results from calculations of distorted structures can be combined to calculate thermal and elastic properties [2, 49–51], and results from different compositions and structural phases can be amalgamated to generate thermodynamic phase diagrams.

A. Generating and using databases for materials discovery

A major aim of high-throughput computational materials science is to identify new, thermodynamically stable compounds. This requires the generation of new materials structures, which have not been previously reported in the literature, to populate the databases. The accuracy of analyses involving sets of structures, such as that used to determine thermodynamic stability, is contingent on sufficient exploration of the full range of possibilities. Therefore, autonomous materials design frameworks such as AFLow use crystallographic prototypes to generate new materials entries consistently and reproducibly.

Crystallographic prototypes are the basic building blocks used to generate the wide range of materials entries involved in computational materials discovery. These prototypes are based on i. structures commonly observed in nature [52–54], such as the rocksalt, zincblende, wurzite or Heusler structures illustrated in Figure 1(b), as well as ii. hypothetical structures, such as those enumerated by the methods described in Refs. 55 and 56. The AFLow Library of Crystallographic Prototypes [54] is also available online at aflow.org/CrystalDatabase/, where users can choose from hundreds of crystal prototypes with adjustable parameters, and which can be decorated to generate new input structures for materials science calculations.

New materials are then generated by decorating the various atomic sites in the crystallographic prototype with different elements. These decorated prototypes serve as the structural input for ab-initio calculations. A full relaxation of the geometries and energy determination follows, from which phase diagrams for stability analyses can be constructed. The resulting materials data are then stored in an online data repository for future consideration.

The phase diagram of a given alloy system can be approximated by considering the low-temperature limit in which the behavior of the system is dictated by the ground state [57, 58]. In compositional space, the lower-half convex hull defines the minimum energy surface and the ground-state configurations of the system. All non-ground-state stoichiometries are unstable, with the decomposition described by the hull facet directly below it. In the case of a binary system, the facet is a tie-line as illustrated in Figure 2(a). The energy gained from this decomposition is geometrically represented by the (vertical)distance of the compound from the facet and quantifies the excitation energy involved in forming this compound. While the minimum energy surface changes at finite temperature (favoring disordered structures), the $T = 0$ K excitation energy serves as a reasonable descriptor for relative thermodynamic stability [59]. This analysis generates valuable information such as ground-state structures, excitation energies, and phase coexistence for storage in the online data repository. This stability data can be visualized and displayed by online modules, such as those developed by AFLow [59], the Materials Project [60], and the OQMD [16, 61]. An example visualization from AFLow is shown in Figure 2(b).

Convex hull phase diagrams have been used to discover new thermodynamically stable compounds in a wide range of alloy systems, including hafnium [62, 63], rhodium [64], rhenium [65], ruthenium [66], and technetium [67] with various transition metals, as well as the Co-Pt system [68]. Magnesium alloy systems such as the lightweight Li-Mg system [69] and 34 other Mg-based systems [70] have also been investigated. This approach has also been used to calculate the solubility of elements in titanium alloys [71], to study the effect of hydrogen on phase separation in iron-vanadium [72], and to find new superhard tungsten nitride compounds [73]. The data has been employed to generate structure maps for hcp metals [74], as well as to search for new stable compounds with the Pt$_4$Ti phase [75], and with the $L_1_1$ and $L_1_3$ crystal structures [76]. Note that even if a structure does not lie on the ground state convex hull, this does not rule out its existence. It may be synthesizable under specific temperature and pressure conditions, and then be metastable under ambient conditions.

B. Standardized protocols for automated data generation

Standard calculation protocols and parameters sets [6] are essential to the identification of trends and correlations among materials properties. The workhorse method for calculating quantum-mechanically resolved materials properties is density functional theory (DFT). DFT is based on the Hohenberg-Kohn theorem [77], which proves that for a ground state system, the potential energy is a unique functional of the density: $V(\rho) = V(\rho(\mathbf{r}))$. This allows for the charge density $\rho(\mathbf{r})$ to be used as the central variable for the calculations rather than the many-body wave function $\Psi(r_1, r_2, ..., r_N)$, dramatically reducing the number of degrees of freedom in
The Kohn-Sham equations \[ 78 \] map the \( n \) coupled equations for the system of \( n \) interacting particles onto a system of \( n \) independent equations for \( n \) non-interacting particles:

\[
\left[-\frac{\hbar^2}{2m} \nabla^2 + V_s(\vec{r})\right] \phi_i(\vec{r}) = \varepsilon_i \phi_i(\vec{r}),
\]

where \( \phi_i(\vec{r}) \) are the non-interacting Kohn-Sham eigenfunctions and \( \varepsilon_i \) are their eigenenergies. \( V_s(\vec{r}) \) is the Kohn-Sham potential:

\[
V_s(\vec{r}) = V(\vec{r}) + \int \frac{\rho_s(\vec{r}')}{|\vec{r} - \vec{r}'|} d^3 \vec{r}' + V_{XC}[\rho_s(\vec{r})],
\]

where \( V(\vec{r}) \) is the external potential (which includes influences of the nuclei, applied fields, and the core electrons when pseudopotentials are used), the second term is the direct Coulomb potential, and \( V_{XC}[\rho_s(\vec{r})] \) is the exchange-correlation term.

The mapping onto a system of \( n \) non-interacting particles comes at the cost of introducing the exchange-correlation potential \( V_{XC}[\rho_s(\vec{r})] \), the exact form of which is unknown and must be approximated. The simplest approximation is the local density approximation (LDA) \[ 79 \], in which the magnitude of the exchange-correlation energy at a particular point in space is assumed to be proportional to the magnitude of the density at that point in space. Despite its simplicity, LDA produces realistic results for atomic structure, elastic and vibra-
appropriately normed (80). Beyond LDA materials, even putting crystal bulk phases in the wrong energetic order tends to overestimate the binding energies of materials. Treatments include the Hubbard $U$ corrections (84, 85), self-interaction corrections (79) and hybrid functionals such as Becke’s 3-parameter modification of LYP (B3LYP, [86]), and that of Heyd, Scuseria and Ernzerhof (HSE, 87).

Within the context of ab-initio structure prediction calculations, GGA-PBE is the usual standard since it tends to produce accurate geometries and lattice constants [57]. For accounting for strong correlation effects, the DFT+$U$ method (84, 85) is often favored in large-scale automated database generation due to its low computational overhead. However, the traditional DFT+$U$ procedure requires the addition of an empirical factor to the potential [84, 85]. Recently, methods have been implemented to calculate the $U$ parameter self-consistently from first-principles, such as the ACBN0 functional [88].

DFT also suffers from an inadequate description of excited/unoccupied states, as the theory is fundamentally based on the ground state. Extensions for describing excited states include time-dependent DFT (TDDFT) [89] and the GW correction [90]. However, these methods are typically much more expensive than standard DFT, and are not generally considered for large scale database generation.

At the technical implementation level, there are many DFT software packages available, including VASP [91–94], QUANTUM ESPRESSO [95, 96], ABINIT [97, 98], FHI–AIMS [99], SIESTA [100] and GAUSSIAN [101]. These codes are generally distinguished by the choice of basis set. There are two principle types of basis sets: plane waves, which take the form $\psi(\mathbf{r}) = \sum e^{i\mathbf{k}\cdot\mathbf{r}}$, and local orbitals, formed by a sum over functions $\phi_n(\mathbf{r})$ localized at particular points in space, such as gaussians or numerical atomic orbitals [102]. Plane wave based packages include VASP, QUANTUM ESPRESSO and ABINIT, and are generally better suited to periodic systems such as bulk inorganic materials. Local orbital based packages include FHI–AIMS, SIESTA and GAUSSIAN, and are generally better suited to non-periodic systems such as organic molecules. In the field of automated computational materials science, plane wave codes such as VASP are generally preferred: it is straightforward to automatically and systematically generate well-converged basis sets since there is only a single parameter to adjust, namely the cut-off energy determining the number of plane waves in the basis set. Local orbital basis sets tend to have far more independently adjustable degrees of freedom, such as the number of basis orbitals per atomic orbital as well as their respective cut-off radii, making the automated generation of reliable basis sets more difficult. Therefore, a typical standardized protocol for automated materials science calculations [6] relies on the VASP software package with a basis set cut-off energy higher than that recommended by the VASP potential files, in combination with the PBE formulation of GGA.

Finally, it is necessary to automate the generation of the $k$-point grid and pathways in reciprocal space used for the calculation of forces, energies and the electronic band structure. In general, DFT codes use standardized methods such as the Monkhorst-Pack scheme [103] to generate reciprocal lattice $k$-point grids, although optimized grids have been calculated for different lattice types and are

![FIG. 2. Convex hull phase diagrams for multicomponent alloys systems. (a) Schematic illustrating construction of convex hull for a general binary alloy system $A_iB_{1-x}$. Ground state structures are depicted as red points, with the minimum energy surface outlined with blue lines. The minimum energy surface is formed by connecting the lowest energy structures with tie lines which form a convex hull. Unstable structures are shown in green, with the decomposition reaction indicated by orange arrows, and the decomposition energy indicated in purple. (b) Example ternary convex hulls as generated by AFLOW.](image-url)
FIG. 3. Standardized paths in reciprocal space for calculation of the electronic band structures for the 25 different lattice types [3].

available online [104]. Optimizing $k$-point grid density is a computationally expensive process that is difficult to automate, so instead standardized grid densities based on the concept of “$k$-points per reciprocal atom” (KPPRA) are used. The KPPRA value is chosen to be sufficiently large to ensure convergence for all systems. Typical recommended values used for KPPRA range from 6,000 to 10,000 [6], so that a material with two atoms in the calculation cell will have a $k$-point mesh of at least 3,000 to 5,000 points. Standardized directions in reciprocal space have also been defined for the calculation of the band structure as illustrated in Figure 3 [3]. These paths are optimized to include all of the high-symmetry points of the lattice.

3. INTEGRATED CALCULATION OF MATERIALS PROPERTIES

Automated frameworks such as AFLOW combine the computational analysis of properties including symmetry, electronic structure, elasticity, and thermal behavior into integrated workflows. Crystal symmetry information is used to find the primitive cell to reduce the size of DFT calculations, to determine the appropriate paths in reciprocal space for electronic band structure calculations (see Figure 3, [3]), and to determine the set of inequivalent distortions for phonon and elasticity calculations. Thermal and elastic properties of materials are important for predicting the thermodynamic and mechanical stability of structural phases [105–108] and assessing their importance for a variety of applications. Elastic properties such as the shear and bulk moduli are important for predicting the hardness of materials [109, 110], and thus their resistance to wear and distortion. Elasticity ten-
sors can be used to predict the properties of composite materials [111, 112]. They are also important in geophysics for modeling the propagation of seismic waves in order to investigate the mineral composition of geological formations [106, 113, 114]. The lattice thermal conductivity ($\kappa_\text{s}$) is a crucial design parameter in a wide range of important technologies, such as the development of new thermoelectric materials [39, 115, 116], heat sink materials for thermal management in electronic devices [117], and rewritable phase-change memories [118]. High thermal conductivity materials, which typically have a zincblende or diamond-like structure, are essential in microelectronic and nanoelectronic devices for achieving efficient heat removal [119], and have been intensively studied for the past few decades [120]. Low thermal conductivity materials constitute the basis of a new generation of thermoelectric materials and thermal barrier coatings [121].

The calculation of thermal and elastic properties offer an excellent example of the power of integrated computational materials design frameworks. With a single input file, these frameworks can automatically set-up and run calculations of different distorted cells, and combine the resulting energies and forces to calculate thermal and mechanical properties.

A. Autonomous symmetry analysis

Critical to any analysis of crystals is the accurate determination of the symmetry profile. For example, symmetry serves to i. validate the forms of the elastic constants and compliance tensors, where the crystal symmetry dictates equivalence or absence of specific tensor elements [50, 107, 122], and ii. reduce the number of ab-initio calculations needed for phonon calculations, where, in the case of the finite-displacement method, equivalent atoms and distortion directions are identified through factor group and site symmetry analyses [123].

Autonomous workflows for elasticity and vibrational characterizations therefore require a correspondingly robust symmetry analysis. Unfortunately, standard symmetry packages [124–127], catering to different objectives, depend on tolerance-tuning to overcome numerical instabilities and atypical data — emanating from finite temperature measurements and uncertainty in experimentally reported observations. These tolerances are responsible for validating mappings and identifying isometries, such as the $n$-fold operator depicted in Figure 4(a).

Some standard packages define separate tolerances for space, angle [127], and even operation type [124–126] (e.g., rotation versus inversion). Each parameter introduces a factorial expansion of unique inputs, which can result in distinct symmetry profiles as illustrated in Figure 4(b). By varying the spatial tolerance $\epsilon$, four different space groups can be observed for AgBr (ICSD #56551 [128]), if one is found at all. Gaps in the range, where no consistent symmetry profile can be resolved, are peculiarly problematic in automated frameworks, triggering critical failures in subsequent analyses.

Cell shape can also complicate mapping determinations. Anisotropies in the cell, such as skewness of lattice vectors, translate to distortions of fractional and reciprocal spaces. A uniform tolerance sphere in cartesian space, inside which points are considered mapped, generally warps to a sheared spheroid, as depicted in Figure 4(c). Hence, distances in these spaces are direction-dependent, compromising the integrity of rapid minimum-image determinations [129] and generally warranting prohibitively expensive algorithms [130]. Such failures can result in incommensurate symmetry profiles, where the real space lattice profile (e.g., bcc) does not match that of the reciprocal space (fcc).

The new AFLOW-SYM module [130] within AFLOW offers careful treatment of tolerances, with extensive validation schemes, to mitigate the aforementioned challenges. Although a user-defined tolerance input is still available, AFLOW defaults to one of two pre-defined tolerances, namely tight (standard) and loose. Should any discrepancies occur, these defaults are the starting values of a large tolerance scan, as shown in Figure 4(b). A number of validation schemes have been incorporated to catch such discrepancies. These checks are consistent with crystallographic group theory principles, validating operation types and cardinalities [131]. From considerations of different extreme cell shapes, a heuristic threshold has been defined to classify scenarios where mapping failures are likely to occur — based on skewness and mapping tolerance. When benchmarked against standard packages for over 54,000 structures in the Inorganic Crystal Structure Database, AFLOW-SYM consistently resolves the symmetry characterization most compatible with experimental observations [130].

Along with accuracy, AFLOW-SYM delivers a wealth of symmetry properties and representations to satisfy injection into any analysis or workflow. The full set of operators — including that of the point-, factor-, crystallographic point-, space groups, and site symmetries — are provided in matrix, axis-angle, matrix generator, and quaternion representations in both cartesian and fractional coordinates. A span of characterizations, organized by degree of symmetry-breaking, are available, including those of the lattice, superlattice, crystal, and crystal-spin. Space group and Wyckoff positions are also resolved. The full dataset is made available in both plaintext and JSON formats.

B. Elastic Constants

There are two main methods for calculating the elastic constants based on the response of either the stress tensor or the total energy to a set of applied strains [50, 51, 132–134]. Automated implementations of these methods are included in the AFLOW (referred to as the Automatic Elasticity Library, AEL [51]) and Materials
FIG. 4. **Challenges in autonomous symmetry analysis.** (a) An illustration of a general $n$-fold symmetry operation. (b) Possible space group determinations with mapping tolerance $\epsilon$ for AgBr (ICSD #56551). (c) Warping of mapping tolerance sphere with a transformation from cartesian to fractional basis.

Project frameworks [50].

To calculate the elastic tensor several different normal and shear strains should be applied to the calculation cell in each independent direction [50, 51], as illustrated in Figure 5(a). The resulting stress tensor elements $\sigma_{ij}$, obtained from the directional forces on the cell calculated with DFT, can then be fitted to the applied strains $\epsilon_{ij}$ to obtain the corresponding elastic constants $c_{ij}$ in the form of the stiffness tensor:

$$
\begin{pmatrix}
\sigma_{11} \\
\sigma_{22} \\
\sigma_{33} \\
\sigma_{23} \\
\sigma_{13} \\
\sigma_{12}
\end{pmatrix} =
\begin{pmatrix}
\epsilon_{11} \\
\epsilon_{22} \\
\epsilon_{33} \\
\epsilon_{23} \\
\epsilon_{13} \\
\epsilon_{12}
\end{pmatrix}
$$

(3)

written in the $6 \times 6$ Voigt notation using the mapping [106]: $11 \mapsto 1$, $22 \mapsto 2$, $33 \mapsto 3$, $23 \mapsto 4$, $13 \mapsto 5$, $12 \mapsto 6$. Symmetry analysis such as that provided by AFLOW-SYM can be used to reduce the number of required calculations by up to a factor of three in the case of cubic systems, as well as for verification of the computed tensors [107].

The elastic constants can then be used in the Voigt or Reuss approximations, which for polycrystalline materials correspond to assuming uniform strain and uniform stress respectively, and give the upper and lower bounds on the elastic moduli. In the Voigt approximation, the bulk modulus is given by

$$
B_{\text{Voigt}} = \frac{1}{9} \left[ (c_{11} + c_{22} + c_{33}) + 2(c_{12} + c_{23} + c_{13}) \right],
$$

(4)

and the shear modulus is given by

$$
G_{\text{Voigt}} = \frac{1}{15} \left[ (c_{11} + c_{22} + c_{33}) - (c_{12} + c_{23} + c_{13}) \right] + \frac{1}{5} (c_{44} + c_{55} + c_{66}).
$$

(5)
The Reuss approximation uses the elements of the compliance tensor $s_{ij}$ (the inverse of the stiffness tensor) to calculate the bulk modulus

$$\frac{1}{B_{\text{Reuss}}} = (s_{11} + s_{22} + s_{33}) + 2(s_{12} + s_{23} + s_{13}),$$

while the shear modulus is given by

$$\frac{15}{G_{\text{Reuss}}} = 4(s_{11} + s_{22} + s_{33}) - 4(s_{12} + s_{23} + s_{13}) + 3(s_{44} + s_{55} + s_{66}).$$

The two approximations are combined to obtain the Voigt-Reuss-Hill (VRH) averages [135] for the bulk modulus

$$B_{\text{VRH}} = \frac{B_{\text{Voigt}} + B_{\text{Reuss}}}{2};$$

and the shear modulus

$$G_{\text{VRH}} = \frac{G_{\text{Voigt}} + G_{\text{Reuss}}}{2}.$$  

The Poisson ratio $\nu$ is then given by

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

C. Quasi-harmonic Debye-Grüneisen model

Thermal properties can be predicted by several different methods, such as the quasi-harmonic Debye-Grüneisen model which uses volume as a proxy for temperature [49], and by calculating the phonon dispersion from the dynamical matrix of IFCs [123].

The energy versus volume data from a set of simple static primitive cell calculations can be fitted to a quasiharmonic Debye-Grüneisen model such as the “GIBBS” method [49, 51, 136] to obtain thermal properties, as demonstrated in Figure 5(b). This method has been implemented in the AFLOW framework in the form of the Automatic GIBBS Library (AGL).

First, the adiabatic bulk modulus $B_s$ as a function of cell volume $V$ is obtained either i. by fitting the $E_{\text{DFT}}(V)$ data to an equation of state (EOS), or ii. by taking the numerical second derivative of a polynomial fit of $E_{\text{DFT}}(V)$, which gives the static bulk modulus $B_{\text{static}}$:

$$B_s(V) \approx B_{\text{static}}(\bar{x}) \approx B_{\text{static}}(\bar{x}_\text{opt}(V)) = \frac{\partial^2 E(\bar{x}_\text{opt}(V))}{\partial \bar{x}^2} = \frac{\partial^2 E(V)}{\partial V^2}. $$

Three different empirical EOS have been implemented within AGL: the Birch-Murnaghan EOS [106, 136, 137]; the Vinet EOS [136, 138]; and the Banon-Cáceres-Núñez spinodal EOS [136, 139]. However, these EOS often introduce an additional source of error into the results.
since they are calibrated for specific sets of systems and pressure-temperature regimes. Recent studies have found the numerical calculation of $B$ to be just as, if not more, reliable as the empirical EOS [51]. Therefore, the numerical method is the default for the automated generation of thermomechanical properties for the AFLOW database.

The bulk modulus can then be used to calculate the Debye temperature as a function of volume:

$$\theta_d(V) = \frac{h}{k_b} [6\pi^2 V^{1/2} n]^{1/3} f(\nu) \sqrt{\frac{B_b}{M}},$$  \hspace{1cm} (12)

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

$$f(\nu) = \left\{ 3 \left[ \frac{2}{3} \left( \frac{1 + \nu}{1 - 2\nu} \right)^{3/2} + \left( \frac{1}{3} \frac{1 + \nu}{1 - \nu} \right) \right] - 1 \right\}^{\frac{1}{3}}.$$  \hspace{1cm} (13)

The integration offered by the AFLOW framework allows the value of $\nu$ required by this expression to be obtained directly and automatically from the AEL calculation (Equation 10).

To obtain the equilibrium volume at a particular $(p, T)$ point, the Gibbs free energy is minimized with respect to volume. In the quasi-harmonic approximation, the vibrational component of the free energy, $F_{\text{vib}}(\vec{x}, T)$, is given by

$$F_{\text{vib}}(\vec{x}; T) = \int_0^\infty \left[ \frac{\hbar \omega}{2} + \frac{1}{\beta} \log \left( 1 - e^{-\beta \hbar \omega} \right) \right] g(\vec{x}; \omega) d\omega,$$  \hspace{1cm} (14)

where $\beta = (k_b T)^{-1}$ and $g(\vec{x}; \omega)$ is the phonon density of states, which depends on the sytem geometry $\vec{x}$. In the Debye-Gruneisen model, $F_{\text{vib}}$ can be written as

$$F_{\text{vib}}(\theta_0; T) = \frac{n}{\beta} \left[ \frac{9}{8} \frac{\theta_0}{T} + 3 \log \left( 1 - e^{-\theta_0/T} \right) - D \left( \frac{\theta_0}{T} \right) \right],$$  \hspace{1cm} (15)

where $D(\theta_0/T)$ is the Debye integral

$$D(\theta_0/T) = 3 \left( \frac{T}{\theta_0} \right)^3 \int_0^{\theta_0/T} \frac{x^3}{e^x - 1} dx.$$  \hspace{1cm} (16)

Next, the full Gibbs free energy as a function of temperature and pressure is calculated

$$G(V; p, T) = E_{\text{DF}}(V) + F_{\text{vib}}(\theta_0(V); T) + pV,$$  \hspace{1cm} (17)

and fitted by a polynomial in $V$, the minimum of which gives the equilibrium volume, $V_{\text{eq}}$. Note that the symbol $G$ is used for shear modulus while $G$ is used for the Gibbs free energy.

$\theta_0$ is then determined from its value at $V_{\text{eq}}$, while other thermal properties such as the Gruneisen parameter can be calculated using the expression

$$\gamma = -\frac{V}{\theta_0} \frac{\partial \theta_0(V)}{\partial V}. $$  \hspace{1cm} (18)

The specific heat capacity at constant volume can be obtained using the expression

$$C_{V,\text{vib}} = 3nk_b \left[ 4D \left( \frac{\theta_0}{T} \right) - \frac{3\theta_0/T}{e^{\theta_0/T} - 1} \right],$$  \hspace{1cm} (19)

while the specific heat capacity at constant pressure is given by

$$C_{p,\text{vib}} = C_{V,\text{vib}}(1 + \alpha \gamma T),$$  \hspace{1cm} (20)

where $\alpha$ is the coefficient of thermal expansion

$$\alpha = \frac{\gamma C_{V,\text{vib}}}{B_1 V}.$$  \hspace{1cm} (21)

The lattice thermal conductivity can be calculated using the Leibfried-Schlömann equation [140–142] using the Debye temperature and the Gruneisen parameter

$$\kappa_L(\theta_\alpha) = \frac{0.849 \times 3\sqrt{3}}{20\pi^3(1 - 0.514\gamma^{-1}_\alpha + 0.228\gamma^{-2}_\alpha)} \times \left( \frac{k_\alpha \theta_\alpha}{\hbar} \right)^2 \frac{k_\beta mV^{3/2}}{h\gamma^{3/2}_\alpha}.$$  \hspace{1cm} (22)

where $V$ is the volume of the unit cell, $m$ is the average atomic mass, while $\theta_\alpha$ and $\gamma_\alpha$ are the acoustic Debye temperature and Gruneisen parameter obtained by only considering the acoustic modes, based on the assumption that the optical phonon modes in crystals do not contribute to heat transport [141, 142]. $\theta_\alpha$ and $\gamma_\alpha$ can be derived directly from the phonon DOS by only considering the acoustic modes [141, 143]. $\theta_\alpha$ can also be estimated from the traditional Debye temperature $\theta_\text{D}$ using the expression $\theta_\alpha = \theta_\text{D} e^{-\frac{3}{2}}$ [141, 142]. There is no simple way to extract $\gamma_\alpha$ from the traditional Gruneisen parameter, so the approximation $\gamma_\alpha = \gamma$ is used in the AEL-AGL approach to calculating the thermal conductivity.

The thermal conductivity at temperatures other than $\theta_\alpha$ is estimated using the expression [141, 142, 144]:

$$\kappa_L(T) = \kappa_L(\theta_\alpha) \theta_\alpha/T.$$  \hspace{1cm} (23)

D. Harmonic Phonons

Thermal properties can also be obtained by directly calculating the phonon dispersion from the dynamical matrix of IFCs. The approach is implemented within the AFLOW Phonon Library (APL) [2]. The IFCs are determined from a set of supercell calculations in which the atoms are displaced from their equilibrium positions [123] as shown in Figure 5(c).

The IFCs derive from a Taylor expansion of the potential energy, $V$, of the crystal about the atoms’ equilibrium
where the dynamical matrix and phonon eigenvectors have dimensions \( (3n_a \times 3n_a) \) and \( (3n_a \times 1) \), respectively, and \( n_a \) is the number of atoms in the cell. Hence, Equation 28 has \( \lambda = 3n_a \) solutions/modes referred to as branches. In practice, Equation 28 is solved for discrete sets of \( \vec{q} \)-points to compute the phonon density of states (grid over all possible \( \vec{q} \)) and dispersion (along the high-symmetry paths of the lattice [3]). Thus, the phonon eigenvalues and eigenvectors are appropriately denoted \( \omega_\lambda(\vec{q}) \) and \( \vec{v}_\lambda(\vec{q}) \), respectively.

Similar to the electronic Hamiltonian, the dynamical matrix is Hermitian, i.e., \( \mathbf{D}(\vec{q}) = \mathbf{D}^*(\vec{q}) \). Thus \( \omega_\lambda^2(\vec{q}) \) must also be real, so \( \omega_\lambda(\vec{q}) \) can either be real or purely imaginary. However, a purely imaginary frequency corresponds to vibrational motion of the lattice that increases exponentially in time. Therefore, imaginary frequencies, or those corresponding to soft modes, indicate the structure is dynamically unstable. In the case of a symmetric, high-temperature phase, soft modes suggest there exists a lower symmetry structure stable at \( T = 0 \) K. Temperature effects on phonon frequencies can be modeled with

\[
\tilde{\omega}_\lambda^2(\vec{q}, T) = \omega_\lambda^2(\vec{q}, T = 0) + \eta T^2,
\]

where \( \eta \) is positive in general. The two structures, the symmetric and the stable, differ by the distortion corresponding to this “frozen” (non-vibrating) mode. Upon heating, the temperature term increases until the frequency reaches zero, and a phase transition occurs from the stable structure to the symmetric [146].

In practice, soft modes [147] may indicate: i. the structure is dynamically unstable at \( T \), ii. the symmetry of the structure is lower than that considered, perhaps due to magnetism, iii. strong electronic correlations, or iv. long range interactions play a significant role, and a larger supercell should be considered.

With the phonon density of states computed, the following thermal properties can be calculated: the internal vibrational energy

\[
U_{\text{vib}}(\vec{x}, T) = \int_0^\infty \left( \frac{1}{2} + \frac{1}{e^{(\beta \hbar \omega)} - 1} \right) \hbar \omega g(\vec{x}; \omega) d\omega,
\]

the vibrational component of the free energy \( F_{\text{vib}}(\vec{x}; T) \) (Equation 14), the vibration entropy

\[
S_{\text{vib}}(\vec{x}, T) = \frac{U_{\text{vib}}(\vec{x}, T) - F_{\text{vib}}(\vec{x}; T)}{T},
\]

and the isochoric specific heat

\[
C_{V_{\text{vib}}}(\vec{x}, T) = \int_0^\infty \frac{k_B (\beta \hbar \omega)^2 g(\vec{x}; \omega)}{(1 - e^{-(\beta \hbar \omega)}) (e^{(\beta \hbar \omega)} - 1)} d\omega.
\]

### E. Quasi-harmonic phonons

The harmonic approximation does not describe phonon-phonon scattering, and so cannot be used to calculate properties such as thermal conductivity or thermal expansion. To obtain these properties, either the
quasi-harmonic approximation can be used, or a full calculation of the higher order anharmonic IFCs can be performed. The quasi-harmonic approximation is the less computationally demanding of these two methods, and compares harmonic calculations of phonon properties at different volumes to predict anharmonic properties. The different volume calculations can be in the form of harmonic phonon calculations as described above [148, 149], or simple static primitive cell calculations [49, 136]. The Quasi-Harmonic Approximation is implemented within APL and referred to as QHA-APL [49]. In the case of the quasi-harmonic phonon calculations, the anharmonicity of the system is described by the mode-resolved Grüneisen parameters, which are given by the change in the phonon frequencies as a function of volume

\[
\gamma_\lambda(\vec{q}) = -\frac{V}{\omega_\lambda(\vec{q})} \frac{\partial \omega_\lambda(\vec{q})}{\partial V},
\]

where \(\gamma_\lambda(\vec{q})\) is the parameter for the wave vector \(\vec{q}\) and the \(\lambda^{th}\) mode of the phonon dispersion. The average of the \(\gamma_\lambda(\vec{q})\) values, weighted by the specific heat capacity of each mode \(C_{V,\lambda}(\vec{q})\), gives the average Grüneisen parameter:

\[
\gamma = \frac{\sum_{\lambda,\vec{q}} \gamma_\lambda(\vec{q}) C_{V,\lambda}(\vec{q})}{C_V}.
\]

The specific heat capacity, Debye temperature and Grüneisen parameter can then be combined to calculate other properties such as the specific heat capacity at constant pressure \(C_p\), the thermal coefficient of expansion \(\alpha\), and the lattice thermal conductivity \(\kappa_L\) [149], using similar expressions to those described in Section 3C.

F. Anharmonic phonons

The full calculation of the anharmonic IFCs requires performing supercell calculations in which pairs of inequivalent atoms are displaced in all pairs of inequivalent directions [150–159] as illustrated in Figure 5(c). The third order anharmonic IFCs can then be obtained by calculating the change in the forces on all of the other atoms due to these displacements. This method has been implemented in the form of a fully automated integrated workflow in the AFLOW framework, where it is referred to as the AFLOW Anharmonic Phonon Library (APL) [159]. This approach can provide very accurate results for the lattice thermal conductivity when combined with accurate electronic structure methods [159], but quickly becomes very expensive for systems with multiple inequivalent atoms or low symmetry. Therefore, simpler methods such as the quasi-harmonic Debye model tend to be used for initial rapid screening [49, 51], while the more accurate and expensive methods are used for characterizing systems that are promising candidates for specific engineering applications.

4. ONLINE DATA REPOSITORIES

Rendering the massive quantities of data generated using automated ab-initio frameworks available for other researchers requires going beyond the conventional methods for the dissemination of scientific results in the form of journal articles. Instead, this data is typically made available in online data repositories, which can usually be accessed both manually via interactive web portals, and programmatically via an application programming interface (API).

A. Computational materials data web portals

Most computational data repositories include an interactive web portal front end that enables manual data access. These web portals usually include online applications to facilitate data retrieval and analysis. The front page of the AFLOW data repository is displayed in Figure 6(a). The main features include a search bar where information such as ICSD reference number, AFLOW unique identifier (AUID) or the chemical formula can be entered in order to retrieve specific materials entries. Below are buttons linking to several different online applications such as the advanced search functionality, convex hull phase diagram generators, machine learning applications [45, 160, 161] and AFLOW-online data analysis tools. The link to the advanced search application is highlighted by the orange square, and the application page is shown in Figure 6(b). The advanced search application allows users to search for materials that contain (or exclude) specific elements or groups of elements, and also to filter and sort the results by properties such as electronic band structure energy gap (under the “Electronics” properties filter group) and bulk modulus (under the “Mechanical” properties filter group). This allows users to identify candidate materials with suitable materials properties for specific applications.

Another example online application available on the AFLOW web portal is the convex hull phase diagram generator. This application can be accessed by clicking on the button highlighted by the orange square in Figure 6(a), which will bring up a periodic table allowing users to select two or three elements for which they want to generate a convex hull. The application will then access the formation enthalpies and stoichiometries of the materials entries in the relevant alloy systems, and use this data to generate a two or three dimensional convex hull phase diagram as depicted in Figure 6(c). This application is fully interactive, allowing users to adjust the energy axis scale, rotate the diagram to view from different directions, and select specific points to obtain more information on the corresponding entries.
B. Programmatically accessible online repositories of computed materials properties

In order to use materials data in machine learning algorithms, it should be stored in a structured online database and made programmatically accessible via a representational state transfer API (REST-API). Examples of online repositories of materials data include AFLOW [4, 5], Materials Project [11], and OQMD [15]. There are also repositories that aggregate results from multiple sources such as NoMaD [23] and Citrine [162].

REST-APIs facilitate programmatic access to data repositories. Typical databases such as AFLOW are organized in layers, with the top layer corresponding to a project or catalog (e.g., binary alloys), the next layer corresponding to data sets (e.g., all of the entries for a particular alloy system), and then the bottom layer corresponding to specific materials entries, as illustrated in Figure 7(a).

In the case of the AFLOW database, there are currently four different “projects”, namely the “ICSD”, “LIB1”, “LIB2” and “LIB3” projects; along with three more under construction: “LIB4”, “LIB5” and “LIB6”. The “ICSD” project contains calculated data for previously observed compounds [52], whereas the other three projects contain calculated data for single elements, binary alloys, and ternary alloys respectively, and are constructed by decorating prototype structures with combinations of different elements. Within “LIB2” and “LIB3”, there are many different data sets, each corresponding to a specific binary or ternary alloy system. Each entry in the set corresponds to a specific prototype structure and stoichiometry. The materials properties values for each of these entries are encoded via keywords, and the data can be accessed via URLs constructed from the different layer names and the appropriate keywords. In the case of the AFLOW database, the location of each layer and entry is identified by an AFLOW uniform resource locator (AURL) [5], which can be converted to a URL providing the absolute path to a particular layer, entry or property. The AURL takes the form server: AFLOWDATA/project/set/entry/?keywords, for example aflowlib.duke.edu:AFLOWDATA/LIB2_RAW/Cu_pvVs/15/?energy_atom, where aflowlib.duke.edu is the
The AFLOW database is organized as a multilayered system. (b) Example of an AURL which enables direct programmatic access to specific materials entry properties in the AFLOW database.

The automated approach to computational materials science has been used to accelerate the design of materials for structural applications such as metallic glasses and superalloys, and for functional applications including thermoelectrics, magnets, catalysts, batteries, photovoltaics and superconductors.

5. MATERIALS APPLICATIONS
A. Disordered Materials

Section 2 describes how the thermodynamic stability of ordered compounds at zero temperature can be predicted from the convex hull phase diagrams generated using the formation enthalpies available in computational materials data repositories such as AFLOW [2, 4–6]. At finite temperature, however, entropic contributions due to thermally driven disorder play an important role, and lead to the formation of disordered materials such as metallic glasses and solid solutions. The thermodynamically favored phase at a given temperature and pressure is the phase with the lowest Gibbs free energy. Since the entropy term in the Gibbs free energy is multiplied by the temperature \(T\), the entropic contribution to the Gibbs free energy becomes increasingly important at higher temperatures. The entropy of materials has two main components: the vibrational entropy, \(S_{\text{vib}}\), which can be calculated from the phonon dispersion or the Debye model as described in Section 3; and the configuration entropy, \(S_{\text{config}}\), due to the disorder in the atomic positions or site occupations. Configurational entropy originates from chemical disorder as in the case of high entropy alloys in which all of the atoms are arranged on a regular lattice, but the specific lattice sites are randomly occupied by different chemical species; or structural disorder as in the case of metallic glasses, where the atoms no longer occupy regular lattice sites resulting in an amorphous material.

High entropy materials. High entropy materials display structural order (i.e., all of the atoms are arranged on a periodic crystal lattice) but chemical disorder (i.e., the actual occupation of these lattice sites is random) [164].

In the ideal entropy limit in which the occupation of the atomic sites is completely random, the configuration entropy per atom is given by \(S_{\text{config}} = k_B \sum_i x_i \log_2 (x_i)\) [164], where \(x_i\) is the fractional composition of each species component. Note that this expression increases with increasing numbers of species, and is also maximized when all of the values of \(x_i\) are equal, i.e. for equimolar compositions.

The expression for the ideal entropy can be combined with calculations for special quasirandom structures (SQS) [165], which are special structural configurations where the radial correlation functions mimic those of a perfectly random structure, to estimate the Gibbs free energy for high entropy alloys. This can then be used in conjunction with the energies of the ordered phases obtained from computational materials data repositories such as AFLOW [2, 4–6] to generate structural phase diagrams as a function of temperature and composition, predicting the phase transition boundaries between ordered compounds, phase separation regions, and single phase solid solutions [108, 166, 167]. The calculated ordered structure energies in AFLOW can also be used to train cluster expansion models [168] to predict the energies of large ensembles of configurations, which can be combined with thermodynamic descriptors to estimate the transition temperature and miscibility gaps for solid solutions and high entropy alloys [169].

The concept of entropy stabilization has recently been extended beyond metallic alloys to include multi-component ceramics, such as high entropy oxides [170, 171]. High entropy oxides consist of an ordered anion sublattice occupied by oxygen ions, with a disordered cation sublattice randomly occupied by five different metal ions, such as Co, Cu, Mg, Ni and Zn [170, 171]. The oxygen ions screen the metal ions from each other, reducing the energy cost associated with forming a random configuration of the metal ions, enabling the formation of a single-phase, entropy-stabilized ceramic.

Metallic glasses. Metallic glasses are alloys in which the atoms do not occupy the sites of a regular periodic lattice, but instead form a structurally disordered amorphous phase. These materials are of great commercial and industrial interest due to their unique combination of superb mechanical properties [172] and plastic-like processability [173–175] for several potential applications [176–180].

Several different attempts have been made to understand the formation of metallic glasses and predict the glass forming ability (GFA) of different alloy compositions. Most of these efforts center around maximizing the packing density of the different atoms [181], which requires elements with a range of different atomic radii [182–186]. Other efforts have been made to use phase diagram data on liquidus temperatures to predict GFA [187–189]. Work is also underway to use machine learning techniques to predict potential glass formers [190].

Much of the theoretical work described above relies on the use of experimental rather than \textit{ab-initio} computational data to predict new materials, due to the difficulty of modeling amorphous structures using first-principles techniques. However, Perim \textit{et al.} [48] recently demonstrated that the energies of different structural phases can be combined into a descriptor to predict the formation of metallic glasses. If there are many different structural phases with similar formation enthalpy, this will frustrate crystallization during solidification and thus promote glass formation [48]. This frustration can be quantified to formulate a spectral descriptor for GFA using the structural and energetic information available in computational materials data repositories such as AFLOW [2, 4–6]. The differences in the geometry between two structures are quantified by describing each structure in terms of its atomic environments [191–193], while the formation enthalpy differences between the respective structures are expressed in the form of Boltzmann factors. The energetic and structural descriptors can then be combined with appropriate normalization factors to formulate a spectral descriptor for GFA as function of composition \(x\): \(GFA (\{x\})\). Comparisons with known glass forming compositions available in the literature can then be used to define a threshold, such that if \(GFA (\{x\})\) exceeds this threshold, then the composition \(x\) would be expected to
be glass forming.

The \( GFA \) \( \{ x \} \) descriptor has been used to perform an automated analysis of the GFA of over 1,400 binary alloy systems from the AFLOW data repository [48]. While over half of all binary alloy systems are predicted to have a GFA below that of the threshold, nevertheless some 17% of alloy systems display a maximum value of \( GFA \) \( \{ x \} \) greater than the maximum value for the Cu-Zr system, a well-known good glass former. These included several alloy systems for which glass formation had never previously been observed or sometimes even investigated, suggesting that there are many possible glass forming compositions which remain to be discovered. This success demonstrates the power of combining descriptors based on the easily calculated properties of periodic crystalline phases with large pre-calculated databases for predicting the synthesizability of complex disordered materials.

**Modeling off-stoichiometry 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, as demonstrated by its ubiquity in technological applications. Prominent examples include fuel cells [194], high-temperature superconductors [195, 196], and low thermal conductivity thermoelectrics [197].

Specifically, chemical disorder can arise in the form of doping, vacancies, and even in the occupation of lattice sites themselves (random), which cannot inherently be modeled using periodic systems. One approach for modeling such effects includes SQS [165]. These quasirandom approximates are very computationally effective, but only offer a single representation of the disordered states, i.e., that with the lowest site correlations. Instead of reducing down to a single representation, AFLOW treats such systems as an ensemble of ordered supercells [198]. Properties are resolved through ensemble averages of the representative states, with opportunities to optimize computation (via supercell size/site error) and tune the level of disorder explored (via parameter \( T \)). AFLOW-POCC (AFLOW partial occupation module) has already resolved significant stoichiometric trends in wide-gap semiconductors and magnetic systems while offering additional insight into underlying physical mechanisms. Ultimately, the screening criteria and property predictions generated by these *bona fide* thermodynamic models and descriptors are accelerating design of new, technologically-significant materials, including advanced ceramics [199] and metallic glasses [48].

**B. Superalloys**

Superalloys are characterized by their extraordinary mechanical properties, particularly at temperatures near their melting point. Such traits make them the ideal candidates for applications in the aerospace and power generation industries. Among the more common examples, many have a face-centered cubic structure with base elements nickel, cobalt, and iron, though nickel-based superalloys dominate the market. Surprisingly, a novel cobalt-based superalloy, \( \text{Co}_3(\text{Al,W}) \) was discovered in 2006 that exhibits mechanical properties better than many nickel-based superalloys. This inspired a thorough computational investigation with AFLOW of alloys containing 40 different elements, yielding over 2,224 relevant ternary systems [59]. The search offered 102 systems shown to i. be more stable than \( \text{Co}_3[\text{Al}_{0.5},\text{W}_{0.5}] \) — the \( \text{L}_12 \)-like random structure previously characterized thermodynamically [200] and very close to the compositions reported by experiments [201], ii. have a relevant concentration \( \{X_3[A_7B_{1-x}]\} \) that is in two-phase equilibrium with the host matrix, and iii. exhibit only small deviations from the host matrix lattice (within 5% relative mismatch).

For these 102 candidates, additional pertinent properties were extracted, including the density and bulk modulus (as a proxy for hardness). Low density materials are preferred to mitigate the stress on turbine components. Significant trends for the bulk modulus are elucidated when plotting with respect to component \( B \) on a Petti-for scale: Ni-based materials show a peak at or before Ni, whereas Co-based materials monotonically increase. Additionally, Co-based materials are generally more resistant to compression compared to Ni-based materials.

Of the 102 candidates, 37 materials have no reported phase diagrams in standard databases, and are thus expected to be unexplored or new. Additional screening based on the toxicity and (low) melting temperature of components uncovered six priority candidates for experimental validation.

**C. Thermoelectrics**

Thermoelectric materials generate an electric voltage when subjected to a temperature gradient, and can also generate a temperature gradient when a voltage is applied [202, 203]. Their lack of moving parts and resulting scalability means that they have potential applications in power generation for spacecraft, energy recovery from waste heat in automotive and industrial facilities [204, 205] and in spot cooling for nanoelectronics using the Peltier cooling effect [204, 205]. However, most of the available thermoelectric materials have low efficiency, only converting a few percent of the available thermal energy into electricity. Therefore, a major goal of thermoelectrics research is to develop new materials that have higher thermoelectric efficiency.

The thermoelectric efficiency of a material is determined by the figure of merit \( zT \), which is obtained from [202, 203]

\[
zT = \frac{\sigma S^2 T}{\kappa_L + \kappa_e},
\]

where \( S \) is the Seebeck coefficient, \( \sigma \) is the electrical conductivity, \( \kappa_L \) is the lattice thermal conductivity, and \( \kappa_e \) is
is the electronic thermal conductivity. The lattice thermal conductivity $\kappa_L$ can be calculated using the methods described in Section 3. Most of the electronic thermal conductivity $\kappa_e$ will depend directly on the electrical conductivity $\sigma$ through the Wiedemann-Franz law [202]

$$\kappa_e = L \sigma T,$$  \hspace{1cm} (36)

where $L$ is the Lorenz factor, which has a value of $2.4 \times 10^{-8} \text{ J}^2 \text{ K}^{-2} \text{ C}^{-2}$ for free electrons. The Seebeck coefficient $S$ is given by [202]

$$S = \frac{8\pi^2 k_B^2}{3\hbar^2 m^* T} \left( \frac{\pi}{3h} \right)^4, $$  \hspace{1cm} (37)

where $n$ is the charge carrier concentration, $e$ is the electronic charge, and $m^*$ is the density of states effective mass of the charge carriers in the material. The effective mass tensor $m_{ij}$ can be calculated from the curvature of electronic band structure dispersion $E(\vec{k})$

$$m_{ij}^{-1} = \frac{1}{\hbar^2} \frac{d^2E}{dk_i dk_j}, $$  \hspace{1cm} (38)

where $k_i$ and $k_j$ are components of the wave vector $\vec{k}$. Larger curvature of the band implies a lower effective mass, while flat narrow bands tend to result in a large effective mass. Charge carrier mobility and thus electrical conductivity tend to reduce with increasing effective mass. However, as can be seen from Equation 37, the Seebeck coefficient increases with effective mass, and $\kappa_e$ also increases with $\sigma$. Therefore, a compromise should be found between high effective mass to maximize $S$ and high charge carrier mobility to give high $\sigma$ in order to optimize the thermoelectric efficiency of the device.

Several computational high-throughput searches have been performed for thermoelectric materials [37–40, 116, 206–208]. Many of the efforts towards developing more efficient thermoelectric materials have focused on either lowering the lattice thermal conductivity $\kappa_L$, or finding materials in which the electronic properties are highly directional, allowing for a narrow energy band distribution while simultaneously having a low effective mass, thus increasing the power factor $\sigma S^2$. High-throughput searches for materials with low lattice thermal conductivity have focused on materials such as half-Heusler structures [39, 40, 209], which have lower densities and thus lower thermal conductivities than the full Heusler structures. Other promising materials include structures such as clathrates [210–214] and skutterudites [208, 215–217], which contain hollow voids that can be filled with “rattler” atoms to reduce the lattice thermal conductivity. Filled skutterudites in particular, such as $R_x\text{Co}_4\text{Sb}_{12}$, are excellent thermoelectric materials because of their combination of a high effective mass with high carrier mobility due to the existence of a secondary conduction band with 12 conducting charge carrier pockets [208].

Searches of large databases of inorganic materials to find new thermoelectrics include the study of 48,000 materials from the Materials Project database [206], where the power factor was calculated using the BoltzTraP code [218] and the thermal conductivity was estimated using the Clarke [219] and Cahill-Pohl [220] models. Almost 600 oxides, nitrides and sulfides from the ICSD were investigated by Garrity [116], where the lattice thermal conductivity was calculated at the quasiharmonic phonon level of approximation, with particular attention being paid to degeneracies in the conduction band minimum, or materials with strongly anisotropic conduction bands, that result in an effective low-dimensional conductor with a corresponding increase in the power factor. The thermoelectric material LiZnSb was proposed by an automated search of the calculated band structures of 1,640 compounds in the ICSD containing Sb [38], although later experimental measurements did not find a high thermoelectric efficiency for this compound [221].

Other strategies to increase the power factor include engineering the band structure [222] through volume changes by alloying different materials to create solid solutions, such as antifluorite Mg$_2$Si and Mg$_2$Ge with Mg$_x$Sn, or orthorhombic Ca$_2$Si and Ca$_2$Ge with Ca$_2$Sn [223]. Tuning the composition of alloys can also be used to converge the valence and conduction bands, enabling high valley degeneracy to be achieved in materials such as PbTe$_{1-x}$Se$_x$ alloys [224]. Solid solutions can also produce local anisotropic structural disorder, increasing phonon scattering and thus improving the thermoelectric efficiency [225, 226].

The exploitation of thermodynamic phenomena such as spinodal decomposition to self-assemble heterostructures with increased phonon scattering [227] has also been proposed to enhance the efficiency of thermoelectric devices. In this approach, materials such as PbSe and PbTe, that are miscible at high temperatures, undergo phase separation when the mixture is cooled slowly, creating a layered heterostructure with a network of boundaries between the different components, which scatter phonons and thus suppress the thermal conductivity. This concept has also been extended to other nanotechnology applications, e.g. as a means to embed a network of electrically conducting nanowires, in the form of topologically protected interface states, within an insulating matrix [228].

The combination of different competing materials properties that must be optimized to maximize the thermoelectric efficiency highlights the importance of integrated frameworks such as AFLLOW, which can automatically calculate different types of materials properties such as thermal conductivity and electronic band structures. Having all of these electronic and thermal properties calculated and available in an integrated, searchable, sortable data repository such as AFLLOW.org accelerates the design of new, high-efficiency thermoelectric materials.
D. Magnetic materials

The search for new magnetic systems remains a long-standing challenge despite their ubiquity in modern technology [229]. Magnetism demonstrates remarkable sensitivity to a number of properties, including electronic configuration, bond length/angle, and magnetic ion valence, and thus its presence is rather uncommon and difficult to predict. In fact, only two percent of the known inorganic compounds [52] exhibit magnetic or- derness, and thus its presence is rather uncommon and difficult to predict. Consumer applications place additional practical restrictions for magnets, with the current global market effectively populated by only two dozen compounds. These obstacles motivated a large-scale computational search with AFLOW for new magnets among the Heusler structure family. Heusler structures are of particular interest for a number of reasons: i. several are known high-performance magnets, ii. the breadth of distinct compounds offers an excellent chance for discovery, iii. the full set of materials will likely offer other types of interesting materials (aside from magnets), and iv. they are metallic and thus well described by DFT. There are three types of Heusler structures, i.e., the regular-XYZ (Cu2MnAl-type), inverse- (XY)XZ (Hg2CuTi-type), and the half-Heuslers XYZ (MgCuSb-type). By decorating these prototypes with ternary combinations, a total of 236,115 compounds were generated and added to the AFLOW.org repository.

As a first attempt, the analysis is limited to Heuslers containing elements of the 3d, 4d, and 5d periods, i.e., a subset of 36,540 compounds. Of this set, 248 are determined to be thermodynamically stable and 22 have a magnetic ground state compatible with the unit cells considered. Among these 22 magnetic ground state compounds, a few prominent classes can be identified, including Co2YZ and Mn2YZ. Upon further analysis of these classes, four materials became of particular interest. In the first class Co2YZ, there already exists 25 known compounds all lying on the Slater-Pauling curve (magnetic moment per formula unit versus number of valence electrons) [230]. The regression predicts Co2MnTi to have the notably high Curie transition temperature TC of 940 K — a feature shared by only two dozen known magnets. The second class Mn2YZ is of interest because of their high TC and potentially large magnetocrystalline anisotropy [231]. Two known examples from this class, Mn2VAl and Mn2VGa, show ferrimagnetic ordering, matching two candidates from the list of 22, Mn2PtCo and Mn2PtV. One more compound was highlighted for satisfying a stringent thermodynamic constraint. Mn2PdPt shows to be robustly stable by at least 30 meV, where the criterion derives from the distance of the stable phase from the pseudo convex hull that neglects it. This criterion quantifies the impact of the structure on the minimum energy surface.

Following an attempt to synthesis these four candidates, two were successful (Co2MnTi and Mn2PtPd) and the other two decomposed into binary compounds. In fact, Co2MnTi shows a TC of 938 K, almost exactly as predicted by the Slater-Pauling curve. Surprisingly, Mn2PdPt shows antiferromagnetic ordering and tetragonal distortion (c/a ∼ 1.8), a result corroborated by calculation upon further analysis. Beyond the synthesis of these two systems, this investigation offers a new, accelerated pathway to discovery over traditional trial-and-error approaches.

6. CONCLUSION

Automated computational materials design frameworks have the capability to rapidly generate materials data without the need for laborious human intervention. They are being used to construct large repositories of programmatically accessible materials properties, calculated in a standardized, consistent fashion so as to facilitate the identification of trends and the training of machine learning models to predict electronic, thermal and mechanical behavior. When combined with physical models and intelligently formulated descriptors, the data becomes a powerful tool to accelerate the discovery of new materials for applications ranging from high-temperature superalloys to thermoelectrics and magnets.

7. ACKNOWLEDGMENTS

We thank Drs. S. Barzilai, Y. Lederer, O. Levy, F. Rose, P. Nath, D. Usannaz, D. Hicks, E. Gossett, D. Ford, R. Friedrich, M. Esters, P. Colinet, E. Perim, C. Calderon, K. Yang, M. Mehl, M. Buongiorno Nardelli, M. Fornari, G. Hart, I. Takeuchi, E. Zurek, P. Avery, R. Hanson, A. Kolmogorov, A. Natan, N. Mingo, J. Carrete, S. Sanvito, D. Brenner, K. Vecchio, M. Scheffler, L. Ghiringhelli, O. Isayev, A. Tropsha, J. Schroers and J. J. Vlassak for insightful discussions. This work is supported by DOD-ONR (N00014-16-1-2326, N00014-16-1-2583, N00014-17-1-2090, N00014-17-1-2876), by NSF (DMR-1436151), and by the Duke University—Center for Materials Genomics. SC acknowledges support by the Alexander von Humboldt-Foundation for financial support. CO acknowledges support from the NSF Graduate Research Fellowship #DGF1106401.

* stefano@duke.edu


