# AFLOW: Integrated infrastructure for computational materials discovery

Cormac Toher, Corey Oses, and David Hicks

*Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708, United States*

Stefano Curtarolo*

*Materials Science, Electrical Engineering, Physics and Chemistry, Duke University, Durham North Carolina, 27708, United States*

(Dated: July 30, 2018)

* stefano@duke.edu

## 1. INTRODUCTION

Generating databases of computational materials properties, sufficiently large to train machine learning models, requires robust, integrated, automated frameworks [1]. The AFLOW framework for computational materials design [2–7] automates the generation of computational materials science data: built-in error correction and a standardized parameter set enable the calculation and analysis of materials properties without direct intervention from human researchers. AFLOW also contains modules to perform symmetry analysis and generate phase diagrams. Data generated using AFLOW is available online via the AFLOW.org web portal [8, 9], and is programmatically accessible using the AFLOW REST-API [10] and the AFLUX Search-API [11].

## 2. GETTING STARTED WITH AFLOW

### A. Installing AFLOW

The AFLOW source code is freely available under version 3 of the GNU Public License. It can be downloaded and compiled using the `xaflow` script. `xaflow` must be installed first before installing `aflow`. Note that `xaflow` is a makefile.
Installing `xaflow` (perform only once):

1. Download `xaflow`: `wget materials.duke.edu/AFLOW/xaflow`.

   - Note that this may require you to first install `wget` on your machine, if it is not already available.
   - On mac: `brew install wget`.
   - On linux: `sudo apt-get install wget`.
   - If necessary, `brew` can be installed on mac using:

     ```
     ruby -e "$(curl -fsSL https://raw.githubusercontent.com/
     Homebrew/install/master/install)"
     ```

   - The compression software `xz` is also required, and if necessary can be installed on mac using: `brew install xz`

2. Move/copy `xaflow` somewhere in your path:

   ```
   mkdir ~/bin
   export PATH=~/bin:$PATH
   mv xaflow ~/bin/xaflow
   ```

   - Add the '`export PATH=~/bin:\$PATH`' line to your `~/.bashrc` so it always stays in your path.

3. Ensure the file can be executed: `chmod 555 ~/bin/xaflow`

4. Check that `xaflow` is in your path: `which xaflow`

   - A returned path means `xaflow` has been successfully installed.

Installing `aflow` (perform with every new version):

1. Run `xaflow`

   - `xaflow` downloads, unpacks, and makes AFLOW in `~/src/AFLOW`
   - The compilation can be redirected to another path:

     ```
     xaflow AWD=/home/src/AFLOW
     ```

2. Install the newly compiled executable to your path: `xaflow install`

   - By default, the executable is installed to `~/bin`
   - The installation can be redirected to another path:

     ```
     xaflow install ULB=/usr/local/bin
     ```

   - As with all makefiles, the `AWD` and `ULB` variables can be set simultaneously:

     ```
     xaflow install AWD=/home/src/AFLOW ULB=/usr/local/bin
     ```

3. Check that AFLOW is in your path: `which aflow`

   - A returned path means AFLOW has been successfully installed.

4. Run `aflow`

AFLOW also needs to be properly configured for your environment and desired workflow. These settings include **i.** machine settings, such as MPI options, binary paths, and database paths, **ii.** VASP settings, such as POTCAR (pseudo-potential) file paths, precision defaults, and run setting defaults, and **iii.** AFLOW settings, such as output file names and default graphic settings. These settings can be modified in the `.aflow.rc` file located in your home directory. For a full description of tunable options, run:

```
aflow --readme=aflowrc
```

Once AFLOW is installed, version information can be retrieved using the command:

```
aflow --version
```

A list of the commands to access the help and readme files for the different AFLOW modules can be retrieved using the command:

```
aflow --help
```

### B. Decorating crystal prototypes to set up new calculations

AFLOW contains thousands of built-in crystallographic prototypes that can be decorated to generate new materials structures and calculation input files. These prototypes are based on structures commonly observed in nature [12–14], such as the rocksalt, zincblende, wurtzite or Heusler structures. The AFLOW Library of Crystallographic Prototypes [14] is available online at `aflow.org/CrystalDatabase/`, where users can choose from hundreds of crystal prototypes with adjustable parameters, which are organized by space group and lattice type, Pearson symbol, Strukturbericht, and chemical symbol.

Prototypes implemented within AFLOW can be decorated to generate POSCARs, which are the structural input files for the VASP DFT package. Input files for QUANTUMESPRESSO, FHI–AIMS and ABINIT can also be generated by including the flags `--qe`, `--aims` and `--abinit`, respectively. Lists of the prototypes implemented within AFLOW can be retrieved using the command `aflow --protos`. The prototypes available at `aflow.org/CrystalDatabase/` require an additional `--params` command to be used, which controls internal and external degrees of freedom, such as cell size and the angles between the lattice vectors, as detailed on each prototype page. For example, a POSCAR file for the compound MgO in the rocksalt structure can be created using the command:

```
aflow --proto=AB_cF8_225_a_b:Mg:O --params=3.5
```

where the `--params` option in this case is used to set the length of the lattice vectors.

The `--params` option can also be used to alter the ratio for the length of the lattice vectors for some prototypes, altering the shape (and even the space group) of the unit cell. For example, the prototype `A_hR1_166_a` takes two parameters, $a$ and $c/a$, where $a$ and $c$ determine the length of the cell in different directions. Using a parameter value of $c/a = 2.0$ will give a rhombohedral structure, whereas $c/a = 2.45$ will give a cubic structure.

AFLOW prototypes can also be used to create `aflow.in` files, which are the general input file for AFLOW calculations. This can be done using the option `--aflow_proto` as in the following example for $Cu_2TiZn$ in the Heusler structure:

```
aflow --aflow_proto=T0001.A2BC:Cu:Ti:Zn
```

This will create the directory structure `AFLOWDATA/Cu_pvTi_svZn/T0001.A2BC/` with an `aflow.in` file inside, which contains everything needed to run a VASP relaxation.

There also exist options for high-throughput prototype creation, such as

1. Multiple prototype inputs:

   ```
   aflow --aflow_proto=T0001.A2BC,TFCC001.ABC:Cu:Ti:Zn
   ```

2. Multiple species combinations:

   ```
   aflow --aflow_proto=T0001.A2BC:Ag,Au,Cu,Fe:Ti:Zn
   ```

3. Pre-defined sets of species combinations:

   ```
   aflow --aflow_proto=T0001.A2BC:LIB3:LIB3:LIB3
   ```

Have fun!

**Exercises:**

1. Select a prototype from the Library of Crystallographic Prototypes at `aflow.org/CrystalDatabase/`. Use the online tool to create a POSCAR for that structure, decorated with elements of your choice.

2. Use the `aflow --proto` command with the appropriate options as described on the online prototype page to generate a POSCAR for the same structure selected in the previous exercise, decorated with elements of your choice.

3. Generate structures for the prototype `A_hR1_166_a` using an element and value of $a$ of your choice, with $c/a$ values of 0.5, 0.612, 1.0, 1.225, 1.5, 2.45, and 3.0. Save these structures with appropriate names, as they will be used as part of a later exercise.

4. Select an alloy system of your choice, and use the `aflow --aflow_proto` option to create an `aflow.in` file for the Heusler structure (prototype `T0001.A2BC`) for that alloy system. Save this file, as it will be used again in later exercises.

### C. The aflow.in file

The main input file for AFLOW calculations is by default called `aflow.in`, although other filenames can be specified using the `--use_aflow.in=` option, *e.g.* `--use_aflow.in=agl_aflow.in`. This file contains all of the information required to set up and run a VASP calculation, or the sets of VASP calculations required to obtain thermal and mechanical properties.

The start of the `aflow.in` file contains some lines with the system name, the VASP version (4.6 is the default for AFLOW), and some APL phonon calculation options that are commented out by default. Below these are the relevant options for VASP relaxations. The first is `[VASP_RUN]RELAX_STATIC_BANDS=2`, which instructs AFLOW to run two sequential VASP relaxations (using the relaxed coordinates from the first as input for the second), a static calculation with a denser kpoint grid to get the electronic DOS, and a band structure calculation along the high-symmetry paths. The rest of this section contains options for the relaxation type (*i.e.* energy or forces), spin polarization, pseudopotential and XC-functional type, etc.

Below this in the `aflow.in` file are the kpoint grid options: k-points per reciprocal atom (KPPRA) for relaxation and static; grid type (M for Monkhorst-Pack, G for Γ-centered); and path through the high-symmetry points in reciprocal space for electronic band structure calculations. Next is the decorated POSCAR, and finally the list of POTCAR files to use. The defaults are listed in the AFLOW standard paper [4], which also provides a useful overview of the types of VASP calculations that AFLOW performs.

### D. Running AFLOW

VASP calculations can be run through AFLOW with the command:

```
aflow --run --force --D <run directory path>
```

where the `--force` option ensures that AFLOW will run the calculation for given material even if it is already available in the AFLOW repository, and `--D` is used to specify the run directory path. A full list of the options to run VASP with AFLOW can be found by running:

```
aflow --readme=aflow
```

The `aflow --monitor` command is also worth using: this will monitor the calculation to see if it is using excessive resources, and will stop the calculation before it crashes the machine. A typical submission script for VASP calculations would include the following as the actual run commands:

```
touch log.$PBS_JOBID
aflow  --monitor >> log.$PBS_JOBID &
aflow --run --force --D ./ >> log.$PBS_JOBID
```

Note that the relaxed POSCAR files are automatically appended to the end of the `aflow.in` file. Since AFLOW reads the `aflow.in` file starting from the bottom, this file can be used as the input file for future calculations requiring relaxed structures.

Also, note that AFLOW automatically creates a file called `LOCK` when it runs VASP in a directory, which is a calculation log file containing progress information, warning and error messages, etc. AFLOW will not run again in that directory while the `LOCK` file is present. To run another calculation in the same directory (*e.g.* a phonon calculation with APL), the `LOCK` file needs to be either moved to a separate directory, renamed, or a different `LOCK` file name should be set using the `--use_LOCK=` option. For example, APL can be run in the same directory after a relaxation calculation by using the command:

```
aflow --use_LOCK=apl.LOCK --force --multi --D ./
```

In this example, the `LOCK` file for the APL calculation is now called `apl.LOCK`, and the existing `LOCK` file will be ignored, so that APL can be run in the same directory subsequent to a VASP relaxation. APL will set up a set of directories with distorted structures, and run

static VASP calculations in each to get the forces. The `--multi` option instructs AFLOW to descend into all of the subdirectories without the appropriately named `LOCK` file, so that it runs all of the necessary calculations with a single command. Finally, AFLOW will need to be run again in this directory (renaming the `apl.LOCK` file in the base directory first), so that APL can read the forces from the finished VASP calculations, and use these to obtain the interatomic force constants and thus generate the phonon dispersion. More details on APL options can be found by running:

```
aflow --readme=apl
```

### E. Symmetry analysis

The new AFLOW-SYM module [15] within AFLOW performs symmetry analysis of crystal structures. It defaults to one of two pre-defined tolerances, namely `tight` (standard) and `loose`. Two different algorithms, based on the International Tables for Crystallography standard representation and on a general input representation of the crystal structure, respectively, are employed. If there are inconsistencies between the two algorithms, or if the set of symmetry operations found does not match any of the 230 space groups, then a scan is performed to find a tolerance value that delivers a consistent result.

The AFLOW-SYM symmetry module can be run using the `--aflowSYM` command line option, for example:

```
aflow --aflowSYM < POSCAR
```

where `POSCAR` is the name (and path if in a different directory) of the file containing the structure in the VASP input format. Structural data in the QUANTUMESPRESSO, FHI–AIMS, or ABINIT input formats can also be read by AFLOW-SYM. The `--aflowSYM` command will print information on the symmetry operations for the lattice point group, reciprocal-lattice point group, coset representatives of the factor group, crystal point group, dual of the crystal point group, site symmetry and space group, and the unique and equivalent sets of atoms, which are are saved in the files `aflow.pgroup.out`, `aflow.pgroupk.out`, `aflow.fgroup.out`, `aflow.pgroup_xtal.out`, `aflow.pgroupk_xtal.out` `aflow.agroup.out`, `aflow.sgroup.out` and `aflow.iatoms.out`, respectively.

Tolerances can also be specified explicitly, *e.g.*:

```
aflow --aflowSYM=0.0001 < POSCAR
```

The option `--print` can be used to control the format of the output to be in plain text or in JSON format:

```
aflow --aflowSYM=0.0001 --print=json < POSCAR
```

The `--sgdata` option calculates and returns the space-group information (number/symbol, Wyckoff positions, cell setting, etc.), only validating that the symmetry descriptions match with the ITC conventions:

```
aflow --sgdata=0.0001 --print=txt < POSCAR
```

The `--edata` option calculates and returns the extended crystallographic symmetry data (crystal, lattice, reciprocal lattice and super-lattice symmetry), while incorporating the full set of checks for robust symmetry determination:

```
aflow --edata=0.0001 --print=txt < POSCAR
```

The command `--aflowSG` returns the space-group number and symbol in the ITC convention:

```
aflow --aflowSG < POSCAR
```

The command `--wyccar` lists the Wyckoff positions consistent with the ITC in a POSCAR format:

```
aflow --wyccar < POSCAR
```

The command `--pearson` provides the Pearson symbol:

```
aflow --pearson < POSCAR
```

The command `--sprim` returns the structure in the standard primitive cell:

```
aflow --sprim < POSCAR
```

Note: Only `--aflowSYM` and `--edata` incorporate the full set of consistency checks between the different symmetry descriptions. Using the standalone symmetry commands (*e.g.*,

`--aflowSG`, `--wyccar`, `--pearson`, etc.) does not guarantee commensurate symmetry descriptions across different functions (*e.g.*, the space-group may be inconsistent with the Pearson symbol).

The AFLOW-SYM module is also available online at `aflow.org/aflow_online.html/`. This application allows users to upload a POSCAR structure file and analyze its symmetry, as well as perform other operations such as finding the standard primitive or standard conventional cell, or finding the appropriate path through the high-symmetry points in reciprocal space for band structure calculations.

More details and a full list of options for the AFLOW-SYM module can be found by running:

```
aflow --readme=symmetry
```

**Exercises:**

1. Use AFLOW-SYM to determine the symmetry properties of the structure that you generated using the prototype that you selected from the Library of Crystallographic Prototypes at `aflow.org/CrystalDatabase/` in the previous set of exercises. What are the space group number and Pearson symbol for this material?

2. Use AFLOW-SYM to determine the symmetry properties of the structures that you generated for the prototype `A_hR1_166_a`. How do the choice of parameters affect the symmetry of the structure?

## 3. THERMODYNAMIC AND THERMO-MECHANICAL PROPERTIES

### A. Convex hull phase diagrams

The phase diagram of a given alloy system can be approximated by considering the low-temperature limit where the behavior of the system is dictated by the ground state [16, 17]. 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. Facets characterize regions of phase coexistence, with their shape (dimensionality) dictated by the

Gibbs phase rule. For example, the facets are tie-lines for binary systems and triangles for ternary systems.

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 [18, 19]. A similar distance metric can be used to quantify the robustness of a stable structure: the distance of the compound from the pseudo-hull constructed without it. The stability criterion played a pivotal role in the discovery of two new magnets — the first discovered by computational means [19, 20].

Leveraging data from the AFLOW.org repository, AFLOW can construct, analyze, and visualize convex hulls locally on any UNIX-like machine with an internet connection [19]:

```
aflow --chull --alloy=MnPdPt
```

AFLOW-CHULL performs a number of data curation steps automatically, including outlier-detection and removal as well as identification of duplicate structures. Thermodynamic descriptors are processed and printed in a PDF by default. JSON and plain-text versions of the output are also available:

```
aflow --chull --alloy=MnPdPt --print=json,txt,pdf
```

Included in the output are: a visualization of the hull, identification of ground-state systems, full decomposition reactions, stability criteria, equivalent ground-state structures (highlighted in **green**), similar ground-state structures by space group (highlighted in **orange**), and determination of coexisting phases. Systems of any dimensionality can be processed by AFLOW, though visualizations are only available for 2-3 dimensional systems

There are also options for directly querying specific thermodynamic descriptors. Using the AFLOW unique identifier (`auid`), the distance from the hull can be calculated directly:

```
aflow --chull --alloy=MnPd --d2h=aflow:9001c322296fc162
```

Similarly, for the stability criterion:

```
aflow --chull --alloy=MnPd --sc=aflow:4fc97e549ca17d4e
```

Additionally, the energy of the hull at a particular stoichiometry can be determined:

```
aflow --chull --alloy=MnPd --hull_energy=0.25
```

Similar to the prototype creation, high-throughput thermodynamic characterization is also possible with AFLOW:

```
aflow --chull --alloy=MnPt,PdPt
aflow --chull --alloy=Mn,Pd:Pt
aflow --chull --alloy=LIB2:LIB2
```

These thermodynamic characterizations also power an online web application for enhanced visualization and data manipulation available at AFLOW.org/aflow_chull.

**Exercises:**

1. What is the distance of $Al_{17}Co_{12}$ from the hull?
2. What is the stability criterion of $Te_2Zr$?
3. What is the energy of the CuZr hull at $x = 0.5$?
4. What is the most stable structure on the MnPdPt hull?

### B. AEL: Elastic Constants

The elastic properties of a material can be calculated using the Automatic Elasticity Library, AEL [21] module within AFLOW. AEL applies several different normal and shear strains to the calculation cell in each independent direction [21, 22], and uses the stress tensors calculated with DFT to obtain the elastic constants.

The elastic constants are used in the Voigt or Reuss approximations to calculate the upper and lower bounds on the bulk and shear moduli. The two approximations are combined to obtain the Voigt-Reuss-Hill (VRH) averages [23], and other properties such as the Poisson ratio are also calculated.

To perform an AEL calculation with AFLOW, the following line should be included in the `aflow.in` file:

```
[AFLOW_AEL]CALC
```

AEL can also be run from within other parts of AFLOW, such as from AGL (see subsection 3 C), in order to obtain properties such as the Poisson ratio to be used within these other workflows. In this case, AEL will run and the other AEL settings in the `aflow.in` file will be read, even if the `[AFLOW_AEL]CALC` line is not present.

AEL applies two types of strains to structures: normal strains and shear strains. The number and size of these strains can be controlled by including the following lines in the `aflow.in` file:

```
[AFLOW_AEL]NNORMALSTRAINS=<number>
[AFLOW_AEL]NSHEARSTRAINS=<number>
[AFLOW_AEL]NORMALSTRAINSTEP=<number>
[AFLOW_AEL]SHEARSTRAINSTEP=<number>
```

The default setting is 4 strains in each direction (2 compressive and 2 expansive) and a step size of 0.005, *i.e.*:

```
[AFLOW_AEL]NNORMALSTRAINS=4
[AFLOW_AEL]NSHEARSTRAINS=4
[AFLOW_AEL]NORMALSTRAINSTEP=0.005
[AFLOW_AEL]SHEARSTRAINSTEP=0.005
```

AEL can use the symmetry of the lattice to reduce the number of independent strains required for an elasticity calculation. For lattices with cubic symmetry, the three orthogonal directions are equivalent so that there is only one independent strain direction for each type of strain. The number of independent strain directions can be determined and set automatically by including the following line in the `aflow.in` file:

```
[AFLOW_AEL]STRAINSYMMETRY=option
```

The default is `OFF` so that all three directions are treated as independent for each strain type. Set `[AFLOW_AEL]STRAINSYMMETRY=ON` to automatically reduce the number of calculations.

AEL is run following the instructions for AFLOW calculations described above. To automatically run all of the subdirectories containing the strained structures, the following options should be used:

```
aflow --multi --D ./
```

Note that after the DFT calculations have finished, AFLOW should be run a second time to complete all of the post-processing.

More details and a full list of options for running AEL calculations can be found by running:

```
aflow --readme=ael
```

**Exercises:**

1. Add the appropriate line to run an AEL calculation to the `aflow.in` file created using the Heusler prototype in the previous exercise.

2. Use the command `aflow --run --generate_aflowin_only` to create the subdirectories for this AEL calculation. (Note: this will not perform any DFT calculations, which would require VASP to be installed on your computer, and can be computationally expensive. To run DFT calculations, leave out the option `--generate_aflowin_only`. Use the command `aflow --multi --D ./` to run all subdirectories). How many subdirectories are created?

3. Copy the `aflow.in` file to a new directory, and add the option to switch on the use of symmetry to determine number of independent directions. Re-run the command to generate the subdirectories. How many subdirectories are created now?

### C. AGL: Quasi-harmonic Debye-Grüneisen model

The thermal properties of materials can be calculated using the Automatic GIBBS Library (AGL) module [24] within AFLOW. AGL fits the energies obtained from a set of isotropically expanded and compressed structures to a quasi-harmonic Debye-Grüneisen model [25, 26].

First, the energy-volume data is fit either numerically or to an empirical equation of state to obtain the bulk modulus as a function of volume, $B(V)$. Then, the Debye temperature as a function of volume $\theta_{\mathrm{D}}(V)$ is calculated using the expression:

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

where $M$, $V$ and $n$ are the mass, volume and number of atoms in the unit cell, respectively, and $f(\nu)$ is a function of the Poisson ratio $\nu$ [25, 26]. The volume dependence of $\theta_{\mathrm{D}}(V)$ is

used to calculate the Grüneisen parameter $\gamma$. Next, $\theta_{\mathrm{D}}(V)$ is used to calculate the Gibbs free energy, which is then minimized with respect to volume for each temperature and pressure to determine the corresponding equilibrium volume. $\theta_{\mathrm{D}}$ and $\gamma$ at the equilibrium volume are used to calculate other thermal properties, such as the specific heat capacity, thermal expansion and the lattice thermal conductivity.

To perform an AGL calculation with AFLOW, the following line should be included in the `aflow.in` file:

```
[AFLOW_AGL]CALC
```

AGL applies a set of isotropic compressive and expansive strains to structures. The number and size of these strains can be controlled by including the following lines in the `aflow.in` file:

```
[AFLOW_AGL]NSTRUCTURES=<number>
[AFLOW_AGL]STRAINSTEP=<number>
```

The default setting is 28 structures and a step size of 0.01.

AGL allows the user to set the value of the Poisson ratio to be used in the calculation of the Debye temperature by including the following line in the `aflow.in` file:

```
[AFLOW_AGL]POISSON=<number>
```

The default value is 0.25. AGL can also call AEL to calculate the Poisson ratio from the elastic constants by including the following line in the `aflow.in` file:

```
[AFLOW_AGL]AELPOISSONRATIO=ON
```

This option will run AEL in full, calculating the Poisson ratio, bulk and shear moduli, and other elastic properties.

AGL allows the user to set the number of temperature and pressure points to be used in the calculation, as well as the size of the increments between each temperature or pressure point. These values can be set by including the following lines in the `aflow.in` file:

```
[AFLOW_AGL]NPRESSURE=<number>
[AFLOW_AGL]SPRESSURE=<number>
[AFLOW_AGL]NTEMP=<number>
[AFLOW_AGL]STEMP=<number>
```

NPRESSURE is the number of pressure points and SPRESSURE is the increment between pressure points. NTEMP is the number of temperature points and STEMP is the increment between temperature points. The default values are NPRESSURE=11, SPRESSURE=2.0, NTEMP=201, STEMP=10.0.

AGL can be run following the instructions for running AFLOW calculations described above. To automatically run all of the subdirectories containing the strained structures, the following options should be used:

```
aflow --multi --D ./
```

Note that after the DFT calculations have finished running, AFLOW should be run a second time to complete all of the post-processing.

More details and a full list of options for running AGL calculations can be found by running:

```
aflow --readme=agl
```

**Exercises:**

1. Add the appropriate line to run an AGL calculation to the `aflow.in` file created using the Heusler prototype in the previous exercise.

2. Use the command `aflow --run --generate_aflowin_only` to create the subdirectories for this calculation. (Note: this will not perform any DFT calculations, which would require VASP to be installed on your computer, and can be computationally expensive. To run DFT calculations, leave out the option `--generate_aflowin_only`. Use the command `aflow --multi --D ./` to run all subdirectories). How many subdirectories are created?

3. Copy the `aflow.in` file to a new directory, and add the option to use AEL to calculate the Poisson ratio. Use the command `aflow --run` to create the subdirectories for this calculation. What structures are created?

### D. Harmonic Phonons

Thermal properties can also be obtained directly by using the Automatic Phonon Library (APL) [2] to calculate the phonon dispersion from the dynamical matrix of interatomic force

constants (IFCs). The IFCs are determined from a set of supercell calculations in which the atoms are displaced from their equilibrium positions [27]. AFLOW-SYM is leveraged to identify the smallest number of distortions required to recover the full dynamical matrix [15].

To perform an APL calculation with AFLOW, the following line should be included (or uncommented) in the `aflow.in` file:

```
[AFLOW_APL]CALC
```

Alternatively, a fresh `aflow.in` can be created with the line uncommented by running:

```
aflow --aflow_proto=T0001.A2BC:Cu:Ti:Zn --module=apl
```

There are many tunable options available in the standard `aflow.in`. The primary engine for calculating phonons is the Direct (supercell) Method (DM, described above), but also available is the Linear Response (LR) method via VASP:

```
[AFLOW_APL]ENGINE=DM
```

For the direct method, the magnitude of the distortion is set to 0.015 Å by default:

```
[AFLOW_APL]DMAG=0.015
```

Specifying the necessary dimensions for the supercell is often not a trivial task, especially with anisotropic cells. However, a good starting size is about 100 atoms:

```
[AFLOW_APL]MINATOMS=100
```

Alternatively, the supercell size can be specified directly (overrides `MINATOMS`):

```
[AFLOW_APL]SUPERCELL=3x3x3
```

To calculate the dispersion curves, the following setting needs to be provided:

```
[AFLOW_APL]DC=y
```

A non-standard [3] path can be specified with

```
[AFLOW_APL]DCUSERPATH=G-X|X-U|K-G|G-L
```

It is generally good practice to calculate both the positive and negative distortions and average the forces (distortion plus minus), as specified here:

```
[AFLOW_APL]DPM=y
```

The default option (keyword commented out or not included) is to avoid the calculation of the negative distortion if it is found to be symmetrically equivalent to the positive distortion. Turning this option off (`DPM=n`) will skip calculation of the negative distortion altogether.

Another good practice is to calculate the forces of the undistorted supercell. These can be subtracted off from the forces of the distorted cell, effectively taring (*i.e.* zeroing) the forces. Though the input structure should be fully relaxed (and thus have no forces on the atoms), small non-zero forces are possible in practice and can have a significant impact on the dynamical matrix. The option to also calculate the forces of the undistorted supercell is

```
[AFLOW_APL]ZEROSTATE=y
```

The phonon density of states are calculated if the following setting is turned on:

```
[AFLOW_APL]DOS=y
```

To calculate the thermodynamic properties, the following two settings should be specified:

```
[AFLOW_APL]TP=y
[AFLOW_APL]TPT=0:2000:10
```

The first setting specifies that thermodynamic properties should be calculated, and the second specifies the temperature range: from 0-2000 K in 10 K increments. These properties include the zero point vibrational energy, the internal vibrational energy, the vibrational free energy, the vibrational entropy, and the specific heat at constant volume. APL also allows the user to override the standard KPPRA setting for the supercell calculations, since 6,000-10,000 (AFLOW Standard [4]) is far too high for supercells:

```
[AFLOW_APL]KPPRA=2000
```

More details and a full list of options for running APL calculations can be found by running:

```
aflow --readme=apl
```

**Exercises:**

1. Add/uncomment the appropriate line to run an APL calculation to the `aflow.in` file created using the Heusler prototype in the previous exercise.

2. Use the command `aflow --run --generate_aflowin_only` to create the subdirectories for this calculation. (Note: this will not perform any DFT calculations, which would require VASP to be installed on your computer, and can be computationally expensive. To run DFT calculations, leave out the option `--generate_aflowin_only`. Use the command `aflow --multi --D` to run all subdirectories). How many subdirectories are created? How many atoms do the supercells contain?

## 4. AFLOW ONLINE DATA REPOSITORIES

Rendering the massive quantities of data generated using AFLOW available for other researchers requires going beyond the conventional methods for the dissemination of scientific results in the form of journal articles. Instead, the data is made available through the AFLOW online data repository, which can be accessed both manually at the `aflow.org` web portal, and programmatically via an application programming interface (API).

### A. AFLOW.org computational materials data web portal

The front page of the AFLOW online data repository is available at `http://www.aflow.org/`. 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 the search bar are buttons linking to several different online applications such as the advanced search functionality, convex hull phase diagram generators, machine learning applications [28–30] and AFLOW-online data analysis tools. The advanced search application is in the top left corner. 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. Use the "# of Species:" box to control the number of species

in the materials returned for a search query.

For example, to find the band gaps for all of the entries for NaCl, select "Na" and "Cl" from the periodic table, and insert the number "2" in the box next to "# of Species:". Next, click on the "Electronics" button, and from this add the property filter for "Egap". Finally, click on the "Search" button to submit the search query.

To restrict returned results to a particular value range, select the box in the column under "Restrict Value" in the appropriate property filter. If this box is not selected for a particular property in the search filter list, then all results otherwise satisfying the search criteria will be returned, even if this property has not been calculated. To toggle between sorting in ascending or descending order for a particular property, click on the column header for that property. The full AFLOW entry page for a material in the search results can be generated by clicking on that material's entry in the "Chemical Formula" column.

**Exercises:**

1. Use the advanced search functionality to find the band gap for SiC in the zincblende structure (space group number 216).

2. Use the advanced search functionality to find the bulk moduli for materials containing Ti in the AFLOW database. How many results are returned?

3. Use the "Restrict Value" button to limit the search to only entries for which the bulk modulus has already been calculated for materials containing Ti. Open the entry page of the material with the highest bulk modulus. What is the lattice type and space group of this material? Is it a metal or an insulator?

### B. Programmatically accessible online repositories of computed materials properties

The AFLOW REST-API [10] facilitates programmatic access to the computed materials science data. The location of each layer and entry is identified by an AFLOW uniform resource locator (AURL) [10], convertible to a URL that provides the absolute path to a particular layer, entry or property. The AURL takes the form:

```
<server>:AFLOWDATA/<project>/<set>/<entry>/?<keyword>
```

The AFLOW database currently contains 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 [12], 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 materials entry are encoded via keywords, which are listed in Ref. 10, with additional keywords in Ref. 11.

The AURL for the formation enthalpy per atom of the ternary Heusler structure with the composition $Cu_2TiZn$ is:

```
aflowlib.duke.edu:AFLOWDATA/LIB3_RAW/Cu_pvTi_svZn/
T0001.A2BC/?enthalpy_formation_atom
```

Each AURL can be converted to a web URL by changing the ":" after the server name to a "/", so that the AURL described above becomes the URL

```
aflowlib.duke.edu/AFLOWDATA/LIB3_RAW/Cu_pvTi_svZn/
T0001.A2BC/?enthalpy_formation_atom
```

This URL, if queried via a web browser or using a UNIX utility such as `curl` or `wget`, returns the energy per atom in eV for entry `T0001.A2BC` of the Cu-Ti-Zn ternary alloy system.

Data in layers above the entry layer can also be accessed using the REST-API. For example, a full list of the entries available for the alloy system Cu-Ti-Zn can be retrieved using the AURL:

```
aflowlib.duke.edu:AFLOWDATA/LIB3_RAW/Cu_pvTi_svZn/?aflowlib_entries
```

**Exercises:**

1. Use the AFLOW REST-API to download the formation enthalpies for all of the entries in the ternary alloy system AlNiTi.

2. Use the AFLOW REST-API to download the full list of binary alloy systems that are available in the LIB2 project. How many alloy systems are available for download?

3. Use the AFLOW REST-API to download the stoichiometries and space groups for all of the entries in an alloy system of your choice. How many entries are present in this system?

### C. AFLUX Search-API

The search and sort functions of the `aflow.org` web portal is combined with the programmatic data access functionality of the AFLOW REST-API in the form of the AFLUX Search-API, which uses the LUX language to enable the embedding of logical operators within URI query strings [11]. The AFLUX Search-API allows users to construct and retrieve customized data sets, which they can feed into materials informatics machine learning packages to identify trends and correlations for use in rational materials design.

AFLUX search queries take the form

```
<query-server>/?<matchbook>,<directives>
```

The current location of the query server is `http://www.aflowlib.duke.edu/search/API`. The matchbook contains the logical search query and consists of AFLOW REST-API materials property keywords with arguments, while the directives provide control over the quantity and format of the returned data.

For example, the energy per cell of every entry in the AFLOW repository containing the elements Na or K, and the element Cl, with an electronic band gap between 1 and 5 eV, can be retrieved using the search query:

```
aflowlib.duke.edu/search/API/?species((Na:K),Cl),Egap(1*,*5),energy_cell
```

In this AFLUX search query, the comma "`,`" represents the logical AND operation, the colon "`:`" the logical OR operation, and the asterisk "`*`" is the "loose" operation that defines a range of values to search within.

AFLUX returns only the first 64 entries matching the search query by default. The number and set of entries can be controlled by appending the `paging` directive to the end of the search query as follows:

```
aflowlib.duke.edu/search/API/?species((Na:K),Cl),Egap(1*,*5),energy_cell,paging(0)
```

where calling the `paging` directive with the argument "0" instructs AFLUX to return all of the entries satisfying the search criteria (beware that this could potentially be a large amount of data, depending on the search query).

The AFLUX schema, including a full list of current keywords along with their description, status, and other information such as units, can be retrieved using the `schema` directive:

```
aflowlib.duke.edu/search/API/?schema
```

**Exercises:**

1. Use AFLUX to retrieve the formation enthalpies for all of the entries in a ternary alloy system of your choice.

2. Use AFLUX to retrieve the space groups of all materials that contain Cu and Ti but not V (the logical NOT operator is denoted by the exclamation mark "`!`").

3. Use AFLUX to retrieve the band gaps of all of the materials that have a calculated bulk modulus between 200GPa and 300GPa.

4. Use AFLUX to retrieve the materials in the ICSD catalog that contain the element Sn but not Pb, and that have a band gap between 1 and 3eV. How many such materials are available?

## 5. AFLOW-ML ONLINE

Materials data generated using AFLOW has been used to train machine-learning models to predict electronic band gaps and thermo-mechanical properties [28, 29]. These models are available online at `aflow.org/aflow-ml` where users can upload a POSCAR and generate predictions.

The AFLOW-ML API [30] enables remote programmatic access to these models. The API allows users to submit structural data for the material of interest using a utility such as `cURL`, and then retrieve the results of the model's predictions in JSON format.

The structural data should be submitted in the POSCAR format for version 5 of VASP: compared to the standard POSCAR format generated by the AFLOW prototype modules, this contains an extra line below the lattice vectors listing the chemical symbols of the

elements in the material in the same order as their atomic coordinates (alphabetical in the case of AFLOW generated POSCARs). The POSCAR data can be submitted using a command of the form:

```
curl <server>/<version>/<model>/prediction --data-urlencode file@<poscar file>
```

The server is located at `http://aflow.org/API/aflow-ml`, the current version is `v1.0`, and available models include property labelled materials fragments (`plmf`, [28]) and molar fraction descriptor (`mfd`, [29]). For example, a `plmf` prediction can be obtained for a structure file called POSCAR using the command:

```
curl http://aflow.org/API/aflow-ml/v1.0/plmf/prediction
--data-urlencode file@POSCAR
```

This will return a JSON file of the form:

```
{
  "id": "531c2817-ee51-4a23-be61-6e8ca6f31aec",
  "model": "plmf",
  "results_endpoint": "/prediction/result/531c2817-ee51-4a23-be61-6e8ca6f31aec"
}
```

where the task id is a unique identifier corresponding to that particular submission, and can be used to retrieve the status and results of the prediction:

```
curl http://aflow.org/API/aflow-ml/v1.0/prediction/result/
531c2817-ee51-4a23-be61-6e8ca6f31aec
```

The results of the model's predictions are returned in the form of keyword value pairs in JSON format. A full list and decription of the keywords can be found in Ref. [30].

**Exercise:**

1. Convert the POSCAR files, generated previously using the AFLOW prototype modules, to the format used by version 5 of VASP by adding the additional line with the element symbols. Upload the POSCARs to the `aflow.org/aflow-ml` page, and generate predictions. Are these materials metals or insulators?

**2.** Use the AFLOW-ML API to obtain predictions for both the `plmf` and `mfd` models for these POSCAR files, and retrieve the results in JSON format. For the structures generated using the prototype `A_hR1_166_a`, how does the value of $c/a$ affect the elastic moduli and electronic band gap? What properties can you obtain using the `mfd` model?

---


[1] S. Curtarolo, G. L. W. Hart, M. Buongiorno Nardelli, N. Mingo, S. Sanvito, and O. Levy, *The high-throughput highway to computational materials design*, Nat. Mater. **12**, 191–201 (2013).

[2] S. Curtarolo, W. Setyawan, G. L. W. Hart, M. Jahnátek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. J. Mehl, H. T. Stokes, D. O. Demchenko, and D. Morgan, *AFLOW: An automatic framework for high-throughput materials discovery*, Comput. Mater. Sci. **58**, 218–226 (2012).

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

[4] C. E. Calderon, J. J. Plata, C. Toher, C. Oses, O. Levy, M. Fornari, A. Natan, M. J. Mehl, G. L. W. Hart, M. Buongiorno Nardelli, and S. Curtarolo, *The AFLOW standard for high-throughput materials science calculations*, Comput. Mater. Sci. **108 Part A**, 233–238 (2015).

[5] C. Toher, C. Oses, D. Hicks, E. Gossett, F. Rose, P. Nath, D. Usanmaz, D. C. Ford, E. Perim, C. E. Calderon, J. J. Plata, Y. Lederer, M. Jahnátek, W. Setyawan, S. Wang, J. Xue, K. Rasch, R. V. Chepulskii, R. H. Taylor, G. Gomez, H. Shi, A. R. Supka, R. Al Rahal Al Orabi, P. Gopal, F. T. Cerasoli, L. Liyanage, H. Wang, I. Siloi, L. A. Agapito, C. Nyshadham, G. L. W. Hart, J. Carrete, F. Legrain, N. Mingo, E. Zurek, O. Isayev, A. Tropsha, S. Sanvito, R. M. Hanson, I. Takeuchi, M. J. Mehl, A. N. Kolmogorov, K. Yang, P. D'Amico, A. Calzolari, M. Costa, R. De Gennaro, M. Buongiorno Nardelli, M. Fornari, O. Levy, and S. Curtarolo, *The AFLOW Fleet for Materials Discovery*, vol. arXiv:1712.00422 (2018).

[6] A. R. Supka, T. E. Lyons, L. S. I. Liyanage, P. D'Amico, R. Al Rahal Al Orabi, S. Mahatara, P. Gopal, C. Toher, D. Ceresoli, A. Calzolari, S. Curtarolo, M. Buongiorno Nardelli, and M. Fornari, *AFLOW$\pi$: A minimalist approach to high-throughput ab initio calculations including the generation of tight-binding hamiltonians*, Comput. Mater. Sci. **136**, 76–84 (2017).

[7] M. Buongiorno Nardelli, F. T. Cerasoli, M. Costa, S. Curtarolo, R. De Gennaro, M. Fornari, L. Liyanage, A. R. Supka, and H. Wang, *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 (2017).

[8] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno Nardelli, N. Mingo, and O. Levy, *AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations*, Comput. Mater. Sci. **58**, 227–235 (2012).

[9] W. Setyawan and S. Curtarolo, *AflowLib: Ab-initio Electronic Structure Library Database*, http://www.aflow.org (2011).

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

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

[12] G. Bergerhoff, R. Hundt, R. Sievers, and I. D. Brown, *The inorganic crystal structure data base*, J. Chem. Inf. Comput. Sci. **23**, 66–69 (1983).

[13] M. J. Mehl, *Naval Research Laboratory Crystal Structure database*, http://cst-www.nrl.navy.mil/lattice/ (2011).

[14] M. J. Mehl, D. Hicks, C. Toher, O. Levy, R. M. Hanson, G. L. W. Hart, and S. Curtarolo, *The AFLOW Library of Crystallographic Prototypes: Part 1*, Comput. Mater. Sci. **136**, S1–S828 (2017).

[15] D. Hicks, C. Oses, E. Gossett, G. Gomez, R. H. Taylor, C. Toher, M. J. Mehl, O. Levy, and S. Curtarolo, *AFLOW-SYM: platform for the complete, automatic and self-consistent symmetry analysis of crystals*, Acta Crystallogr. Sect. A **74**, 184–203 (2018).

[16] S. Curtarolo, D. Morgan, and G. Ceder, *Accuracy of ab initio methods in predicting the crystal structures of metals: A review of 80 binary alloys*, Calphad **29**, 163–211 (2005).

[17] G. L. W. Hart, S. Curtarolo, T. B. Massalski, and O. Levy, *Comprehensive Search for New Phases and Compounds in Binary Alloy Systems Based on Platinum-Group Metals, Using a Computational First-Principles Approach*, Phys. Rev. X **3**, 041035 (2013).

[18] C. Nyshadham, C. Oses, J. E. Hansen, I. Takeuchi, S. Curtarolo, and G. L. W. Hart, *A computational high-throughput search for new ternary superalloys*, Acta Mater. **122**, 438–447 (2017).

[19] C. Oses, E. Gossett, D. Hicks, F. Rose, M. J. Mehl, E. Perim, I. Takeuchi, S. Sanvito, M. .Scheffler, Y. Lederer, O. Levy, C. Toher, and S. Curtarolo, *AFLOW-CHULL: Cloud-orientated platform for autonomous phase stability analysis*, arXiv:1806.06901 (2018).

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

[21] C. Toher, C. Oses, J. J. Plata, D. Hicks, F. Rose, O. Levy, M. de Jong, M. D. Asta, M. Fornari, M. Buongiorno Nardelli, and S. Curtarolo, *Combining the AFLOW GIBBS and Elastic Libraries to efficiently and robustly screen thermomechanical properties of solids*, Phys. Rev. Mater. **1**, 015401 (2017).

[22] M. de Jong, W. Chen, T. Angsten, A. Jain, R. Notestine, A. Gamst, M. Sluiter, C. K. Ande, S. van der Zwaag, J. J. Plata, C. Toher, S. Curtarolo, G. Ceder, K. A. Persson, and M. D. Asta, *Charting the Complete Elastic properties of Inorganic Crystalline Compounds*, Sci. Data **2**, 150009 (2015).

[23] R. Hill, *The elastic behaviour of a crystalline aggregate*, Proc. Phys. Soc. Sect. A **65**, 349 (1952).

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

[25] J.-P. Poirier, *Introduction to the Physics of the Earth's Interior* (Cambridge University Press, 2000), 2nd edn.

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

[27] A. A. Maradudin, E. W. Montroll, G. H. Weiss, and I. P. Ipatova, *Theory of Lattice Dynamics in the Harmonic Approximation* (Academic Press, New York, 1971).

[28] O. Isayev, C. Oses, C. Toher, E. Gossett, S. Curtarolo, and A. Tropsha, *Universal fragment*

*descriptors for predicting electronic properties of inorganic crystals*, Nat. Commun. **8**, 15679 (2017).

[29] F. Legrain, J. Carrete, A. van Roekeghem, S. Curtarolo, and N. Mingo, *How Chemical Composition Alone Can Predict Vibrational Free Energies and Entropies of Solids*, Chem. Mater. **29**, 6220–6227 (2017).

[30] E. Gossett, C. Toher, C. Oses, O. Isayev, F. Legrain, F. Rose, E. Zurek, J. Carrete, N. Mingo, A. Tropsha, and S. Curtarolo, *AFLOW-ML: A RESTful API for machine-learning predictions of materials properties*, Comput. Mater. Sci. **in press**, arXiv:1711.10744 (2017).