ME 697R: Computational Methods for Nanoscale Energy Transport

Chapter 8: Machine learning techniques in nanoscale energy transport
Section 8.4: High throughput computation-driven material discovery

Xiulin Ruan
School of Mechanical Engineering
Purdue University
High Throughput Computation-driven Discovery

- **High throughput computations**: “use of many computing resources over long periods of time to accomplish a computational task”

- **High throughput screening approach**: Set up a large number of experiments/computations, perform these using available resources, and collate the obtained outputs/properties in a large organized database.

  - **Automated**: Should need *minimal human intervention*

  - **Output properties**: Instead of calculating only one target property, many physical properties can be calculated and correlations can then be extracted *a posteriori*.
Half Heusler (HH) Compounds

Example steps in high throughput (HT) screening:

- Identify low $\kappa_l$ Half Heusler (HH) compounds

---

Example steps in high throughput (HT) screening:

- Identify low $\kappa_l$ Half Heusler (HH) compounds

1. Screen for positive formation enthalpy

Obtained using high throughput DFT framework (AFLOW)

Example steps in high throughput (HT) screening:

- Identify low $\kappa_l$ Half Heusler (HH) compounds

1. Screen for positive formation enthalpy
2. Screen for mechanical stability

Second order force constants are characterized with full phonon dispersion curves

Example steps in high throughput (HT) screening:

- Identify low $\kappa_l$ Half Heusler (HH) compounds

1. Screen for positive formation enthalpy
2. Screen for mechanical stability

Options:

- Calculate $\kappa_l$ for 450 compounds
- Calculate $\kappa_l$ for few compounds and use ML correlations to predict $\kappa_l$ for the rest

Example steps in high throughput (HT) screening:

- Identify low $\kappa_l$ Half Heusler (HH) compounds

1. Screen for positive formation enthalpy
2. Screen for mechanical stability
3. Screen for thermodynamic stability

Thermodynamical information from AFLOWLIB used to check stability of all 450 compounds against more than 110,000 phases

Example steps in high throughput (HT) screening:

- Identify low $\kappa_l$ Half Heusler (HH) compounds

1. Screen for positive formation enthalpy
2. Screen for mechanical stability
3. Screen for thermodynamic stability

Properties including $\kappa_l$ of 75 HH Compounds

High throughput computations-driven discovery

- To perform the above steps, a **large number of ab-initio calculations** (or experiments) need to be carried out.

- An **automated process** is needed to set up the computations, perform them rapidly and collate the results.

- Should be able to handle **variability** in material structures, real and reciprocal lattices, $k$-point and $q$-point meshes etc. automatically.

- Also needs to **automatically respond to computational errors and failures** and restart failed computations after adjusting parameters.
High Throughput Materials Discovery

- AFLOW: An automatic framework for high-throughput materials discovery

AFLOW Procedure

AFLOW source code now available under GPL 3
materials.duke.edu/AFLOW/


AFLOW WORKSHOP: http://materials.duke.edu/AFLOW/AFLOW_WORKSHOP/
Use of such robust automated \textit{ab-initio} calculations framework generates large dataset on which ML can be applied (Bayesian optimization, random forest etc.).

- Suitable descriptors are chosen to form mapping between input structures and targeted outputs
- E.g. crystalline volume per atom ($V$) and density ($\rho$) were used to correlate to lattice thermal conductivity.

\begin{itemize}
\end{itemize}

- **Low $\kappa$ compounds using Bayesian Optimization:**
  - $\kappa$ of 101 materials was first evaluated by DFT calculations
  - Using the data, Bayesian Optimization was used to predict $\kappa$ for 54779 compounds
  - 221 materials with very low $\kappa$ was obtained from the high-throughput screening, which are potential thermoelectric materials

\[ Z(x^*) = \frac{f(x^*) - f_{\text{best}}}{\sqrt{v(x^*)}} \]

*FIG. 3 (color online). Dependence of the Z score on constituent elements for compounds in the MPD library. The magnitude of the Z score is shown by colors along with volume, $V$, and density, $\rho$, for each element.*

Low $\kappa$ compounds using Bayesian Optimization:

- Easily available properties (e.g. density) give less accurate measures of correlation.
- More expensive properties (e.g. phonon properties) give more accurate correlation but at computational cost.