Electronic Supplementary Material (ESI) for Materials Advances.

Machine learning the vibrational free energy of perovskites : Electronic supplementary information

Krishnaraj Kundavu[1], Suman Mondal[1], and Amrita Bhattacharya[1*]

[1]Ab-initio Computational Materials Science laboratory, Department of Metallurgical Engineering and Materials Science, Indian Institute of Technology Bombay, Maharashtra, 400076, India,

[*]Address correspondence to: b_amrita@iitb.ac.in

In this work, we build a machine learning model to predict the vibrational Helmholtz free energy of perovskite compounds. The datasets and codes used in our work are provided in github repository ML-FreeEnergyPerorskites

Link: https://github.com/krishnarajmayya/MLFreeEnergyPerorskites.git

1 Virtual design space

Perovskite compounds have very wast compositional space with a large percentage of elements in the periodic table eligible to take one of the three positions in ABX_3 stoichiometry. Figure 1 shows the design space used in this work to select the perovskite candidates for machine learning and prediction of stability. Many of the elements can occupy either A or B sites in ABX_3 structure (indicated in bi-color in the figure).

2 Feature correlation

Pearson correlation scores of the elemental features with each other are checked to remove redundancy by dropping the features showing a high correlation score ($> \pm 0.9$). The heat map showing the correlation of the descriptors for A and X site elements are given in Figure 2.

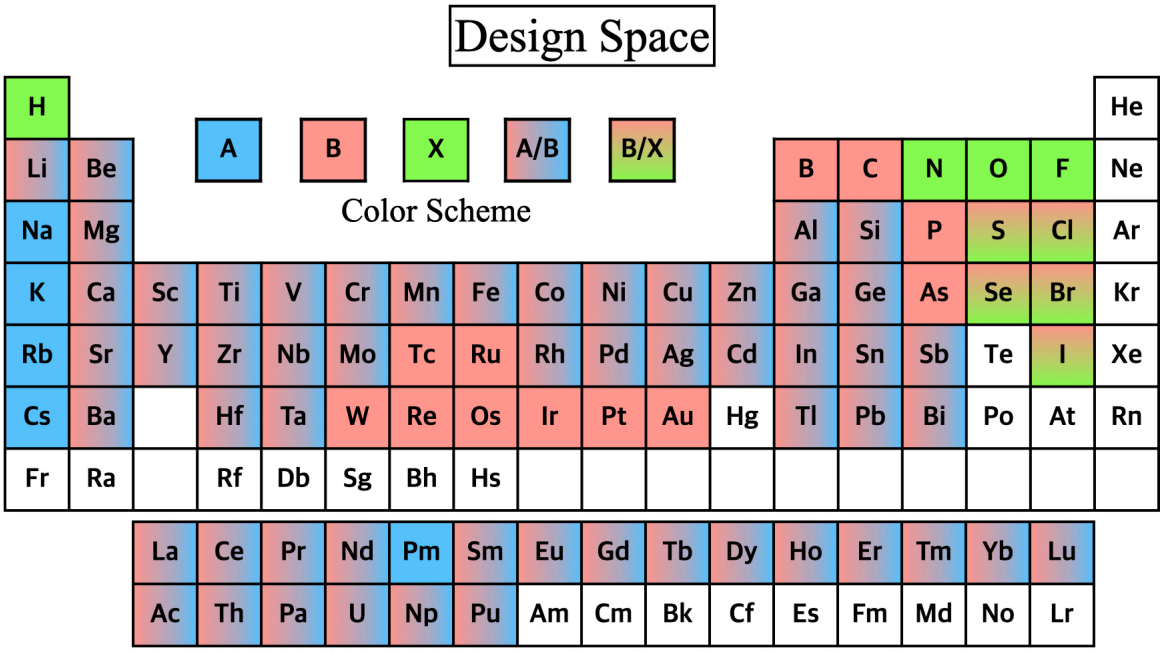

Figure 1: Design space of the perovskite compounds used in this study. The color scheme in given inside.

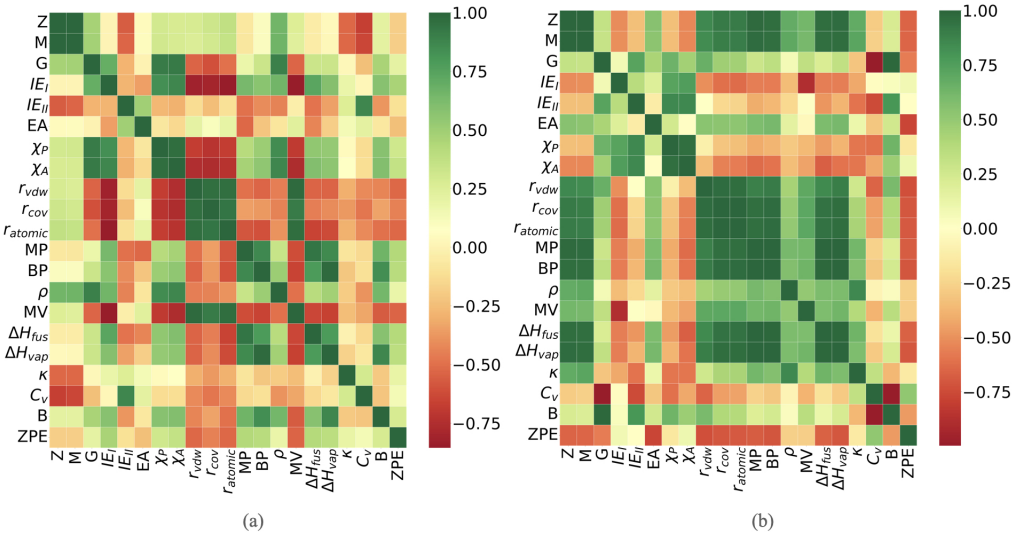

Figure 2: Correlation heat map plotted for all the elemental descriptors of the elements at A and X positions. Descriptors with a correlation score of $> \pm 0.9$ are concluded to be strongly correlated (symbols are in their usual meaning as given in Table 1 of the main paper).

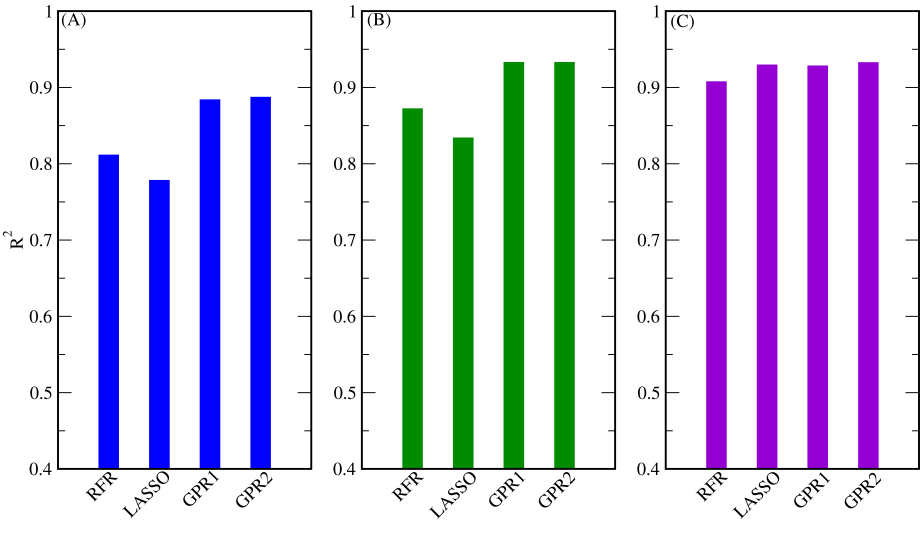

Figure 3: R^2 scores of different ML algorithms built using top 40 features for coefficient A, B, and C of free energy polynomial fit.

3 ML algorithms

To build the ML models, various algorithms implemented in the Scikit Learn package are used viz. Least Absolute Shrinkage and Selection Operator regression (LASSO), Random forest regression (RF), and Gaussian process regression (GPR) using the radial-basis function (RBF) along with white kernel and Rational quadratic kernel (henceforth referred to as GPR-1 and GPR-2 respectively). A comparison of the cross-validation R^2 scores obtained for all these models with the top 40 features is shown in Figure 3. For all three coefficients, GPR-2 model gives the best CV performance and hence are used for the final ML models.

4 Feature importance

Analyzing the important features of a ML model for the given target variable can give insights into the inherent relation between the target property and the properties of materials. Figure 4 shows the top 15 features for coefficients A, B, and C as predicted using select-K-best method. Bond lengths of the perovskite structures and properties of X anion appear to be very important in defining free energy variation. Elemental properties of A cation become dominant in predicting A while that of cation B become dominant for B.

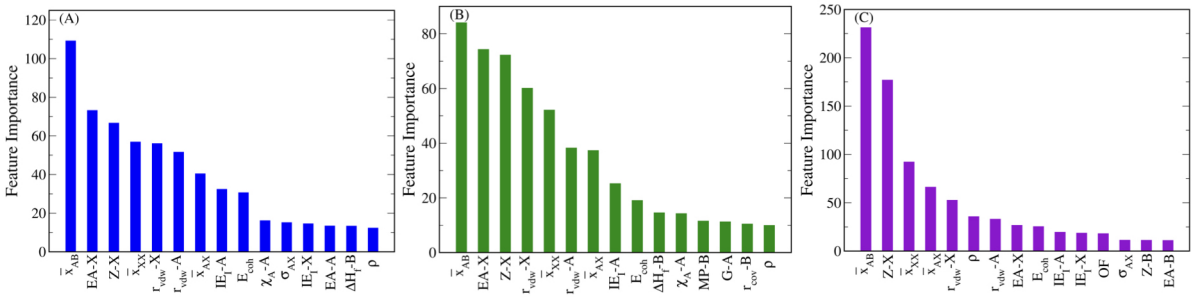

Figure 4: Feature importance of the top 15 descriptors as predicted using the select-K-best method for coefficient A, B, and C of free energy polynomial fit.

5 Validation of ML models

To check the accuracy and usefulness of ML models built in our study, F_{H} values are predicted for few compounds unseen in our dataset. Our models successfully predicted the phase changes in these compounds. Figure 5 shows the plot of total energy i.e. the sum of the DFT static energy (electronic and ionic) and free energy (predicted from the model) as a function of temperature (T) for $BaBiO_3$, $KCaCl_3$, $CsSrCl_3$ and $LaAlO_3$. Our results correctly predict stability of these compounds correctly. It can be noted that our plots indicate a long range of temperatures where the difference between the energies of competing phases are very close to each other. This manifests in the observation of gradual transition from one phase to another in experiments and also leads to co-existence of such phases for a range of temperatures.

6 Data analysis

Analyzing the trends and correlations in the dataset of perovskite compounds in our work gives useful insights into the relation between the composition, structure and vibrational free energy of these compounds. Figure 6 and 7 show the variation of Coefficient D (ZPE) and coefficient A of the perovskite compounds in our design space for different elements in A and B sites respectively. The elements are arranged in the ascending order of their atomic number. The plots hint at the periodic variation of free energy values as elements go from left to right in the periodic table. However the correlation is very subtle to gain high importance in the ML models. Figure 8 shows the range and curvature of F_{H} v/s T plots for different space groups.

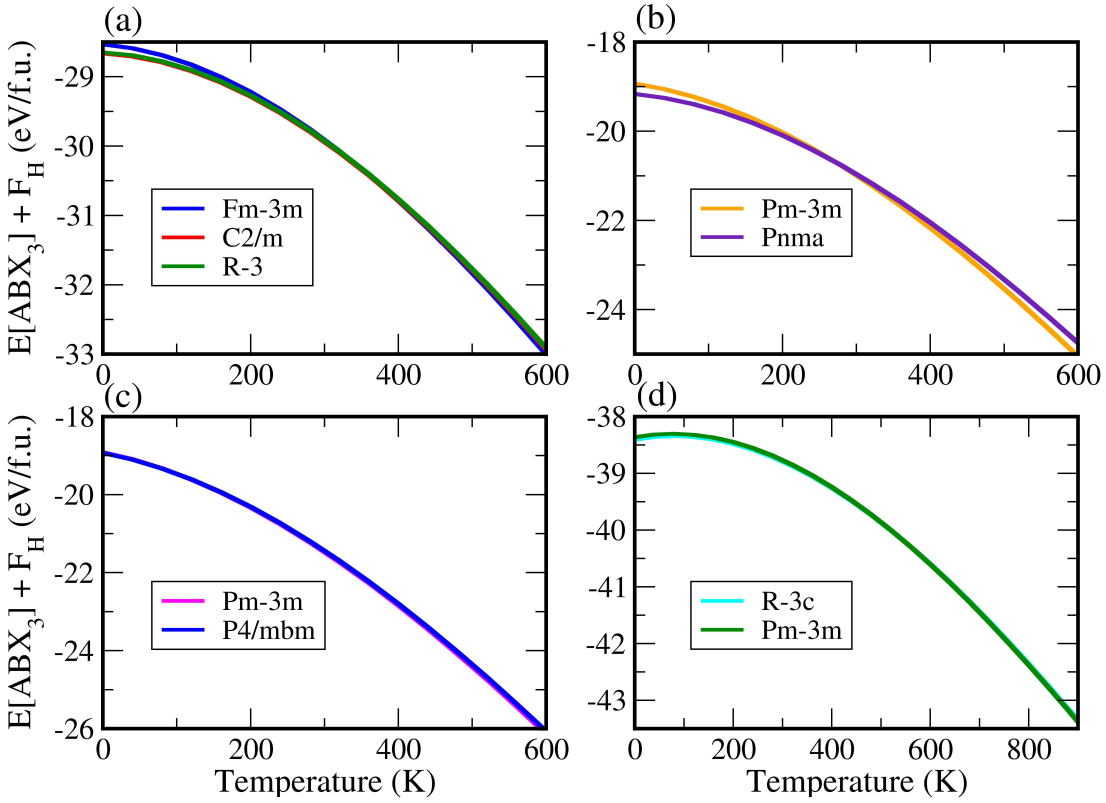

Figure 5: DFT total energy ($E[ABX_3]$) + Helmholtz free energy (F_H) plotted as a function of temperature (T) for competing phases of (a) $BaBiO_3$, (b) $KCaCl_3$, (c) $CsSrCl_3$ and (d) $LaAlO_3$ using the coefficients of polynomial fit - predicted by the ML models built in this study

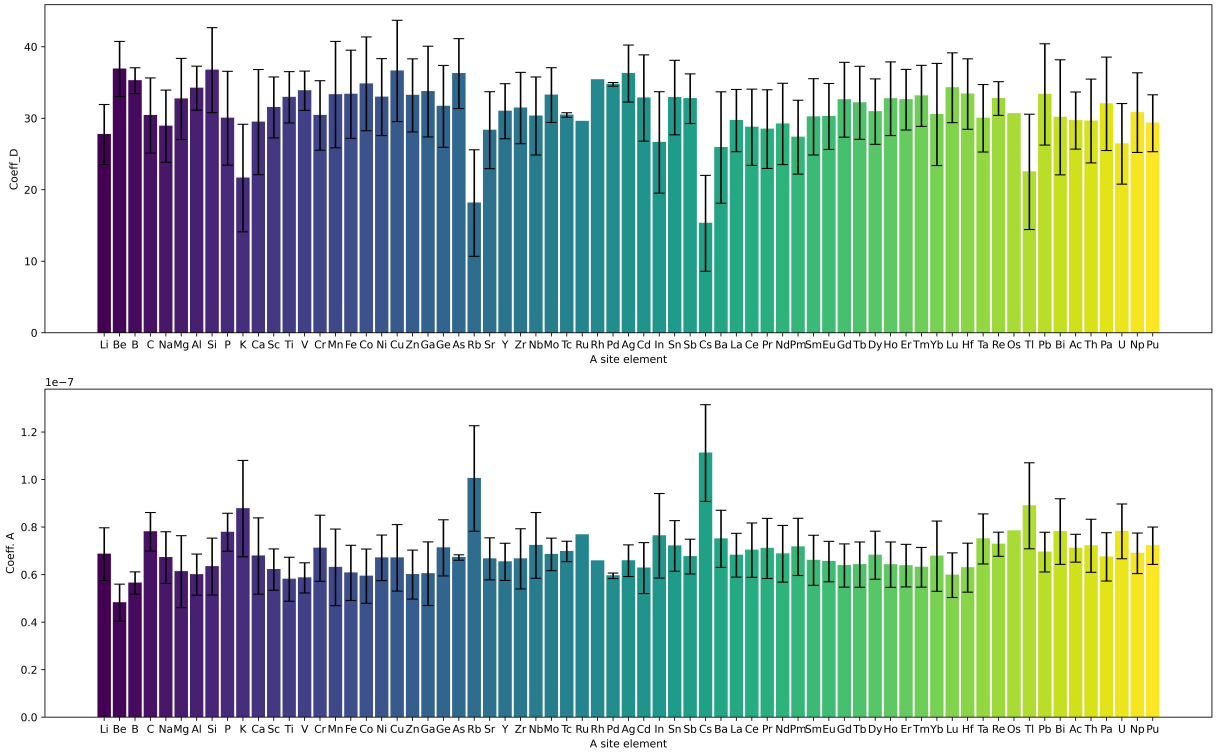

Figure 6: Variation of Coefficient D and coefficient A of polynomial fit respectively with A site element.

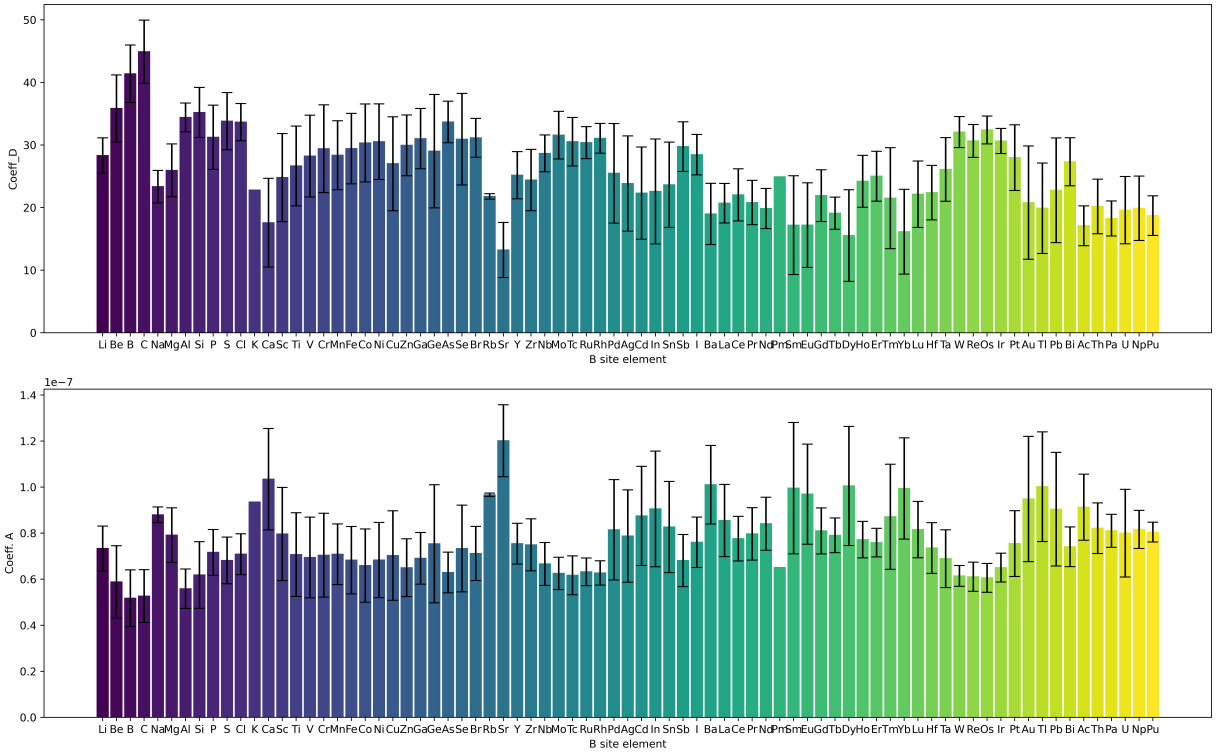

Figure 7: Variation of Coefficient D and coefficient A of polynomial fit respectively with B site element.

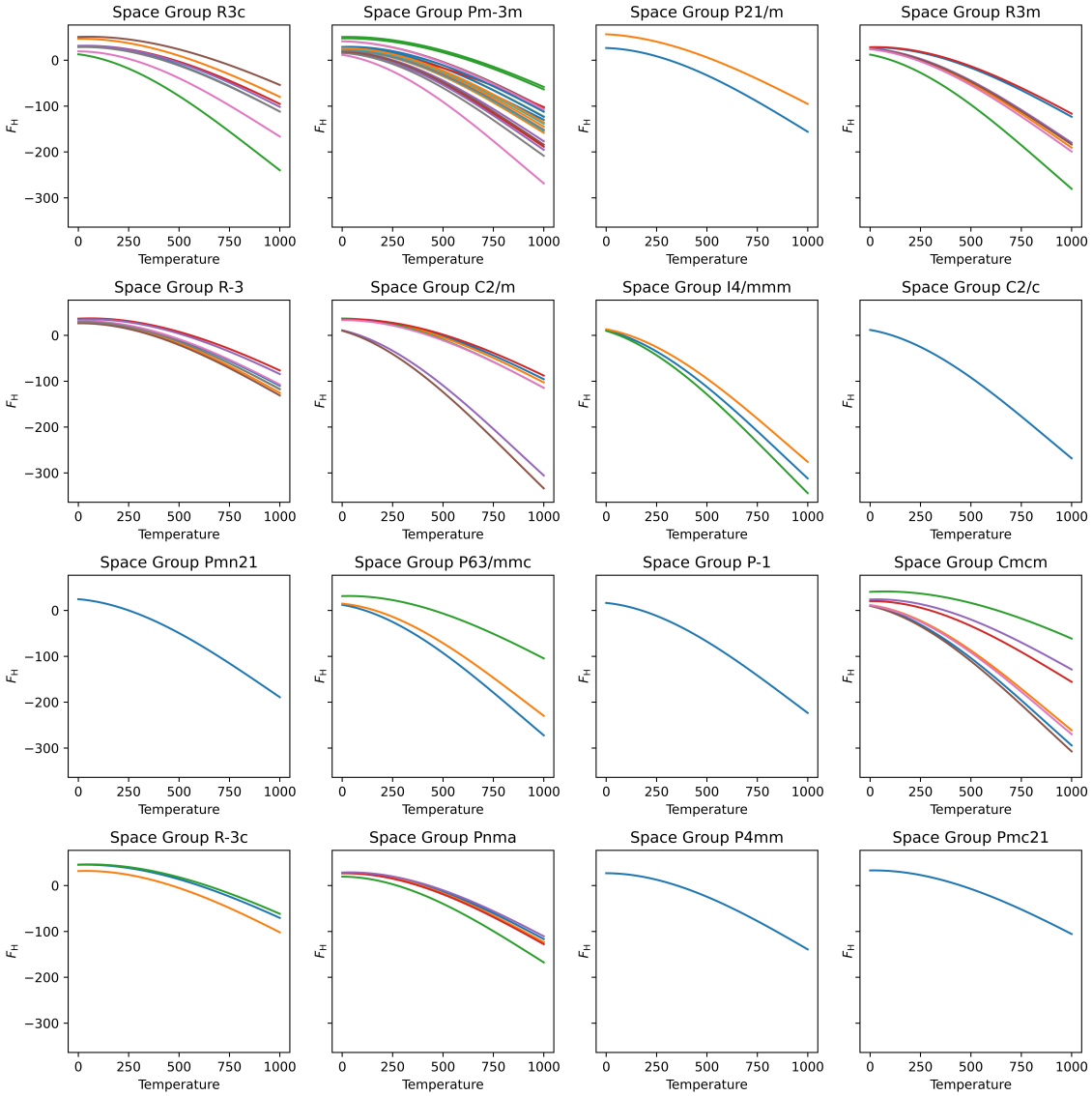

Figure 8: Variation of free energy with temperature for different space groups in our training dataset.