A High-Throughput Computational Dataset of Halide Perovskite Alloys[†]

Jiaqi Yang[a], Panayotis Manganaris[a], and Arun Mannodi-Kanakkithodi[a]

Abstract

Novel halide Perovskites with improved stability and optoelectronic properties can be designed via composition engineering at cation and/or anion sites. Data-driven methods, especially high-throughput first principles computations and subsequent analysis based on unique materials descriptors, are key to achieving this goal. In this work, we report a Density Functional Theory (DFT) based dataset of 550 ABX_3 halide Perovskite compounds, with various atomic and molecular species considered at A, B and X sites, and different amounts of mixing considered at each site generated using the Special Quasirandom Structures (SQS) algorithm for alloys. We perform GGA-PBE calculations on pseudo-cubic Perovskite structures to determine their lattice constants, stability in terms of formation and decomposition energies, electronic band gaps, and properties extracted from optical absorption spectra. To elucidate the importance of the level of theory used, we further perform 300 calculations using the more expensive HSE06 functional and determine lattice constant, stability and band gap, and compare PBE and HSE06 properties with some experimentally measured results. Trends in the datasets are unraveled in terms of the effects of mixing at different sites, the composition in terms of specific atomic or molecular species, and averaged elemental properties of species at different sites. This work presents the most comprehensive DFT perovskite alloy dataset to date and the data, which is open-source, can be exploited to train predictive and optimization models for accelerating the design of completely new compositions that may yield large solar cell efficiencies and improved performance across many optoelectronic applications.

Introduction

The challenge of optimizing Perovskite performance is one with many facets. Almost every detail of a Perovskite crystal's structure and chemistry effects its performance as a semiconductor. The size of the unit cell effects its substrate affinity and in turn its carrier concentrations. The crystal phase effects many aspects of the electronic structure, including the band gap and optical response. Of course, the structure is ultimately dependent on the proportions and arrangement of the constituent elements.

Coverage of Chemical Space

Methodology

Building Perovskite Dataset

The dataset we report is based on standard cubic phase ABX_3 Perovskite structures obtained from public databases. Fourteen common Perovskite constituents are selected to form our Halide Perovskite composition space 1. Five constituents including Methylammonium and Formamidinium cations represent the possible A-site occupants. Six heavy metal elements represent the possible B-site occupants. Three halides represent the possible X-site occupants.

Each computational run is performed using a 2x2x2 supercell, this allows A and B site doping to be modeled in discrete $1/8^{th}$ fractions of the total site occupancy, and it allows X site doping to be modeled in $1/24^{th}$ fractions. At this level, it is appropriate to call these Perovskites alloys.

The pure (non-alloyed) possibilities are exhaustively sampled using 90 Perovskites. Based on these pure Perovskite structures, we mix candidates for A, B, and X sites systematically. The alloy space sees combinatorial scaling and must be sparsely sampled.

To create alloys, for example in B sites, we first select possible B sites mixing elements. Then SQS methods are applied to generate random alloys for given mixing concentrations. Thus, we constructed 126 A-site mixing samples, 151 B-site mixing samples and 127 X-site mixing samples. The resulting structures are optimized using a DFT variable-cell relaxation.

Calculation Details

Calculation DFT calculations are performed with Vienna Ab initio Simulation Package (VASP) version 6.1. The projector augmented wave (PAW) potentials were used The generalized gradient approximation (GGA) of Perdew, Burke and Ernzerhof (PBE) and (HSE) is used as exchange-correlation energy. The energy cutoff for the plane-wave basis is set to 500 eV. The Brillouin zone was sampled by Monkhorst-Pack k-point mesh, with a reciprocal mesh as

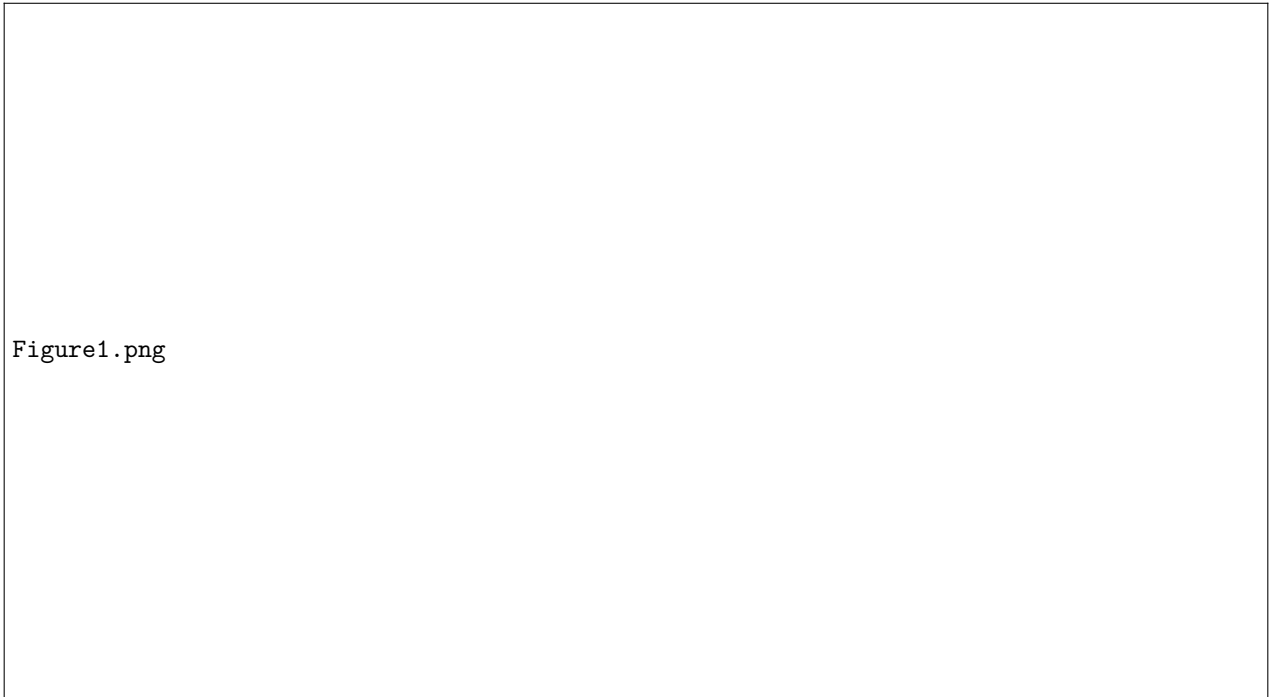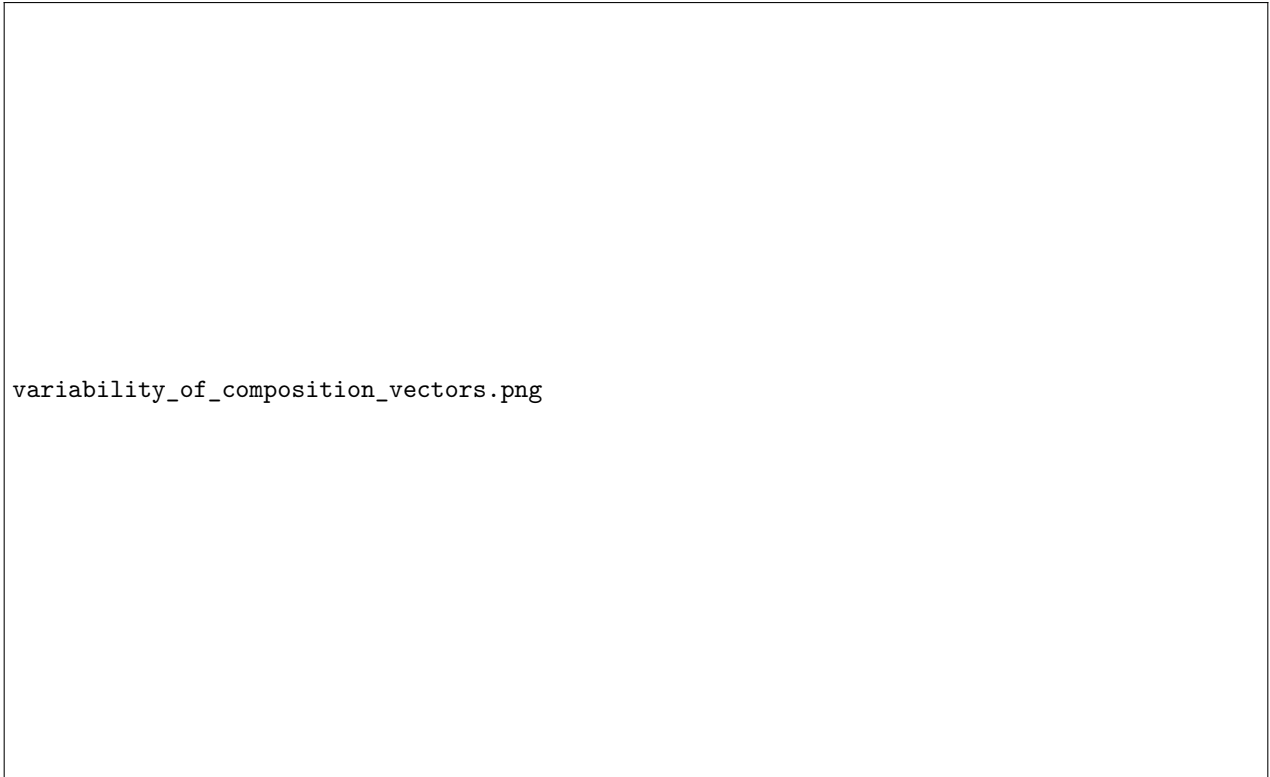[a]School of Materials Engineering, Purdue University, West Lafayette, IN 47907, USA; E-mail: amannodi@purdue.edu

Fig. 1 (a) Chemical space of ABX$_3$ perovskites. (b) Pie charts showing the amounts of various atomic and molecular species as well as types of mixing in the DFT dataset. (c) Detailed outline of this work.

Fig. 2 Plots showing number of Perovskites representing a constituent at a certain atomic fraction of a complete Perovskite.

4x4x4. The structural force convergence threshold is set to be 0.5 eV/Å.

LOPTICS Calculations and Absorption Spectrum

Analysis of DFT Computed Properties

Decomposition Energy

The decomposition energy indicates the stability of a compound. To calculate the decomposition energy for ABX3 perovskite, we assume it will decompose to two phases, AX and BX2. Using DFT calculations, we can get the optimized energy of a perovskite and that of its constituent phases.

The equation is presenting the details calculation for decomposition energy. For PBE decomposition energy, we use DFT energy results from PBE level calculation. For HSE decomposition energy, we use the energy from HSE level calculations.

Band Gap and Band Structure

SLME Package

The SLME metric developed by Yu and Zunger [1] is used as the primary criterion screening perovskites for their photovoltaic merits. The SLME value is computed for every Perovskite using Logan William's SL3ME [2].
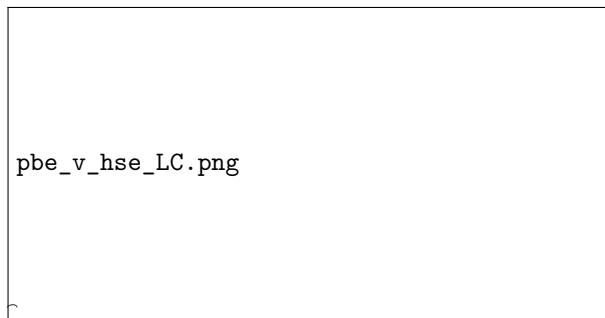
Results and discussion
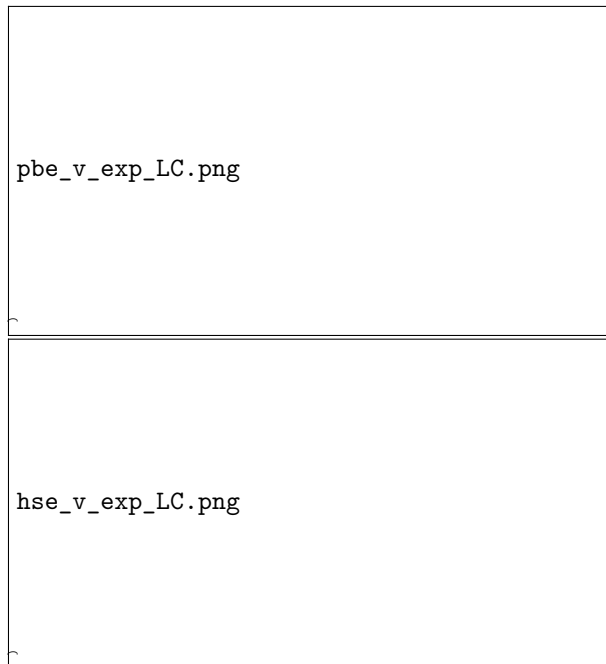
Visualization of DFT Data

Lattice constant PBE vs HSE

Fig ?? presents the lattice parameter comparison of PBE calculation and HSE calculations. Full relaxations at both levels of theory mostly agree on the lattice parameters. At least, any deviation appears to be very linearly explained. This suggests the accuracy of PBE relaxation is enough to optimize most Perovskite samples.

Note this also served as a reasonable validation for our results. A few samples did have significantly differing lattice parameters. This prompted checking the optimized structures. We found those Perovskite structures were substantially deformed and no longer had obvious octahedral structures. Thus, we exclude these outliers from any analysis concerned with the dominant pseudo cubic structures which are the focus of this report.

pbe_v_hse_LC.png

Comparing synthetic to physical data

pbe_v_exp_LC.png

hse_v_exp_LC.png

Decomposition energy vs band gap PBE and HSE

Fig 2(b) is showing the PBE band gap compared to the PBE decomposition energy. It presents the diversity of our perovskite dataset. The data sets cover a large range of band gap and decomposition energy. For example, we have perovskite with low decomposition energy (good stability) and suitable band gap value (between 1 eV to 2.5 eV for PBE calculations). And we can also find samples with low stability and large band gap. The distribution of band gap and decomposition energy shows a great diversity of all perovskite samples and indicates that our data set can statistically represent a sufficient perovskite space. Fig 2(c) shows the plot of HSE band gaps and HSE decomposition energy. Since we applied HSE calculations for part of the samples, it shows some grouping on low decomposition energy. Compared to PBE plot, more data should be added in high decomposition energy region and suitable band gap region.
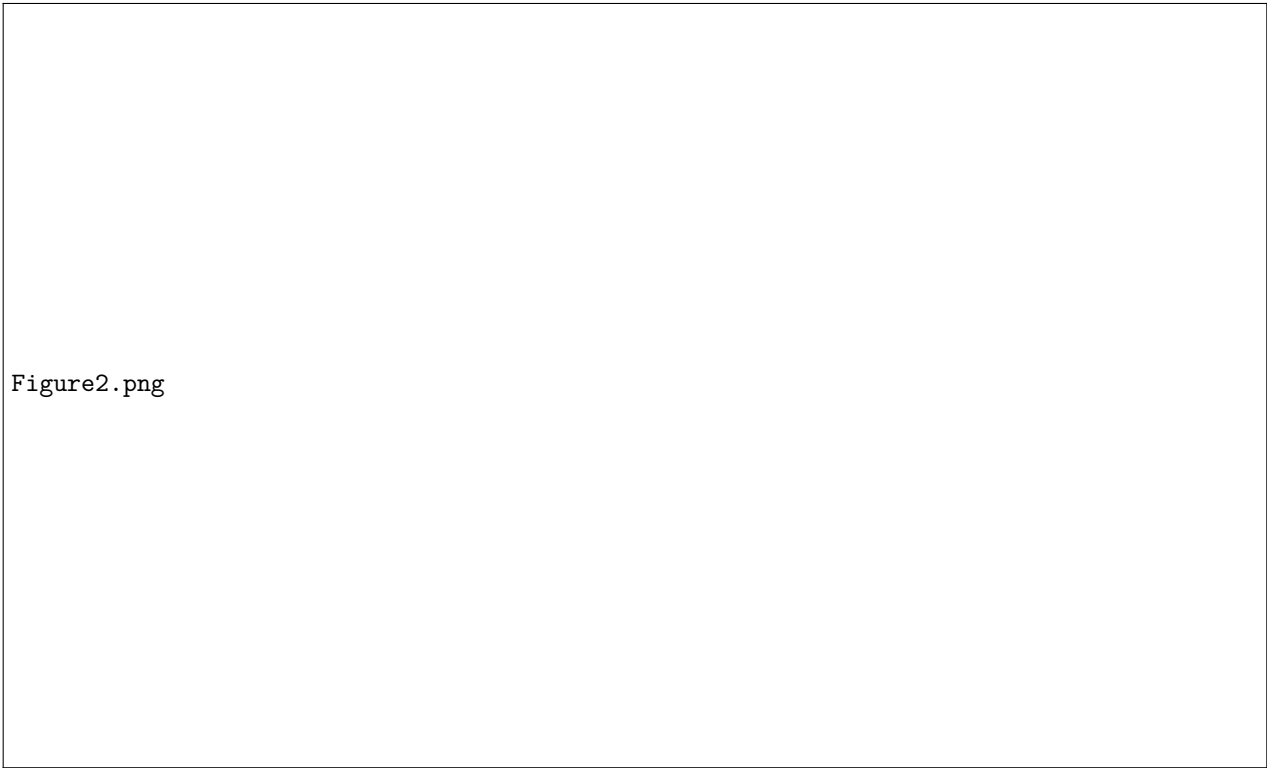
Fig. 3 DFT Results: PBE and HSE properties; lattice constants, decomposition energies, band gaps, and converter efficiencies.

Spectroscopic Limited Maximum Efficiency (SLME) vs PBE Band gap

Fig 2(d) presents the Spectroscopic Limited Maximum Efficiency (SLME) values related to the PBE band gap. Spectroscopic Limited Maximum Efficiency (SLME) is a very important properties for photovoltaic performance. SLME measures the absorption efficiency of light for the perovskite. As Fig 2(d) showing, a peak around 1.5 eV is obvious. The peak indicates that these samples with 1.5 eV PBE band gap will also have best absorption efficiency as photovoltaic materials. As the band gap increases, the SLME value decreases and eventually goes to zero due to the high band gap values.

Pearson Correlation Results

It is unlikely that any of the targets is fully explained by a single composition or composition derived axis. But there are helpful relations that aid in obtaining a physical understanding.

as in a Pearson correlation map is produced to check for strong relations. Those that exist, when plotted in detail show some trending, but always with extensive variability. Evidently, an accurate model will have to be formed on a multidimensional domain.

PCA

Principal Component Analysis is a method of projecting high dimensional data onto a plane defined by the two linear combinations of axes that explain as much of the variance as possible.

The method of PCA is the Singular Value Decomposition, a Unitary Transform which generalizes the familiar eigendecomposition. PCA will "rotate" the N data points in M-D space until their widest 2D cross section is displayed.

At this point it is readily apparent that this dataset is highly topological. The data exists on a mostly bounded domain in high dimensions, so there is some geometry the features constitute.

Our models will prefer to use this this geometric structure in their explanation for why perovskite properties vary, this can be useful for accuracy, it can also be a bias-inducing hindrance.

MDS / T-SNE / Isomap Results

Perspective and Future Work

The design of this dataset is uniquely suited to the exploration of alloying effects on Perovskite properties. The combinatorial space of possible alloys has been sparsely but systematically sampled along four primary alloy schemes. This sample space affords the opportunity for a QM/ML
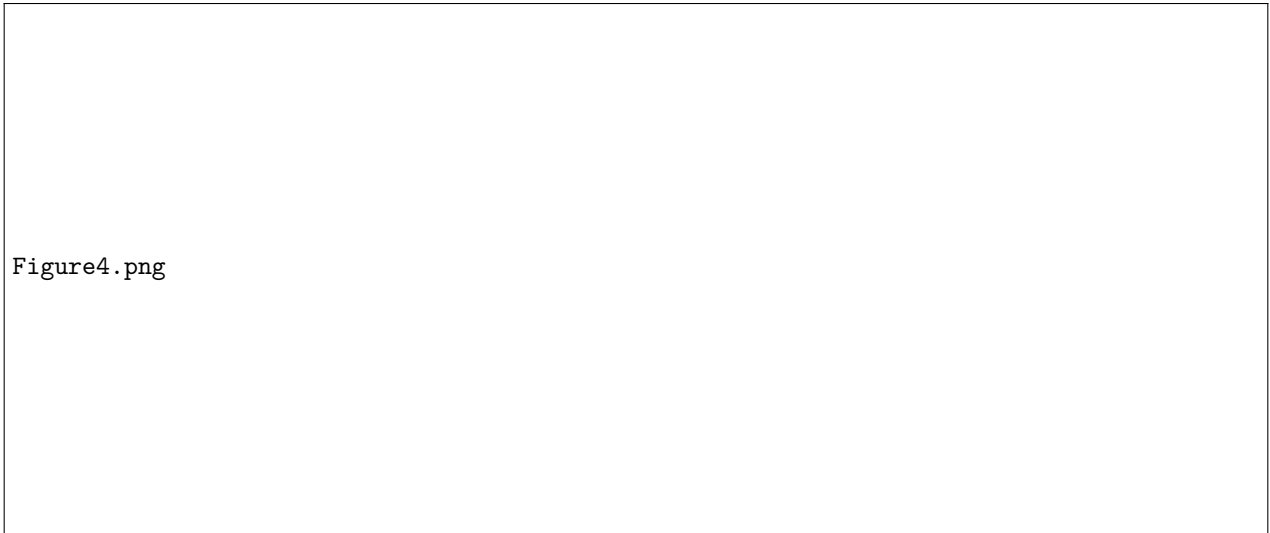
Fig. 4 Pearson linear correlation coefficients between 50 composition and elemental descriptors and (a) 6 PBE computed properties, and (b) 4 HSE computed properties.

Fig. 5 PCA

surrogate model to form the basis of an active learning strategy which can begin selecting potentially high performing multi-site alloy candidates based on the current sample set.

For the calculation of future data points obtained by such a scheme, we will follow a strategy of performing full structural optimization at a PBE level of theory unless circumstances demand otherwise. This is justified by ??.

We anticipate our principal challenge will be in extracting useful predictor variables from the composition information.

The basic feature sets examined here are highly correlated, but nonetheless show promise both as a basic screening criterion, and as robust classifier features under t-SNE transformation.

Modeling pipelines capable of predicting Perovskite decomposition energy will likely be very achievable using a transductive and invertible equivalent of the tSNE algorithm, potentially SONG.

For this reason, kernel learning methods appear to be particularly promising for high speed optimization of the space.

Conclusions

...

Conflicts of interest

There are no conflicts to declare.

acknowledgements

References

1 L. Yu and A. Zunger, Physical Review Letters, 2012, 108, year.

2 L. Williams, Sl3me – a Python3 Implementation of the Spectroscopic Limited Maximum Efficiency (SLME) Analysis of Solar Absorbers, `https://github.com/ldwillia/SL3ME`.

Supplemental Material