Journal of Materials Chemistry A

Machine Learning-Empowered Study of Metastable γ-CsPbI3 under Pressure and Strain

Journal:	*Journal of Materials Chemistry A*
Manuscript ID	TA-ART-01-2024-000174.R1
Article Type:	Paper
Date Submitted by the Author:	08-Mar-2024
Complete List of Authors:	Han, Minkyung ; Stanford University, Earth and Planetary Sciences; Stanford University, Materials Science and Engineering Peng, Cheng; SLAC National Accelerator Laboratory, Stanford Institute for Materials and Energy Sciences Song, Ruyi; SLAC National Accelerator Laboratory, Stanford Institute for Materials and Energy Sciences Ke, Feng; SLAC National Accelerator Laboratory, Stanford Institute for Materials and Energy Sciences Nashed, Youssef; SLAC National Accelerator Laboratory Mao, Wendy; Stanford University, Earth and Planetary Sciences; SLAC National Accelerator Laboratory, Stanford Institute for Materials and Energy Sciences Jiang, Chunjing; University of Florida, Physics Lin, Yu; SLAC National Accelerator Laboratory, Stanford Institute for Materials and Energy Sciences

Machine Learning-Empowered Study of Metastable γ-$CsPbI_3$ under Pressure and Strain

Minkyung Han[1,2,*], Cheng Peng[3], Ruyi Song[3], Feng Ke[3,#], Youssef S. G. Nashed[4], Wendy L. Mao[1,3], Chunjing Jia[5,*], and Yu Lin[3,*]

[1]Department of Earth and Planetary Sciences, Stanford University, Stanford CA, 94305, United States
[2]Department of Materials Science and Engineering, Stanford University, Stanford CA, 94305, United States
[3]Stanford Institute for Materials and Energy Sciences, SLAC National Accelerator Laboratory, Menlo Park CA, 94025, United States
[4]Machine Learning Initiative, SLAC National Accelerator Laboratory, Menlo Park CA, 94025, United States
[5]Department of Physics, University of Florida, Gainesville FL, 32611, United States
[#]Present Address: State Key Laboratory of Metastable Materials Science and Technology, Yanshan University, Hebei, 066104, China.
[*]mhan8@stanford.edu, chunjing@phys.ufl.edu, lyforest@slac.stanford.edu

ABSTRACT

Metastable γ-$CsPbI_3$ is a promising solar cell material due to its suitable band gap and chemical stability. While this metastable perovskite structure can be achieved via introducing external pressure or strain, experimenting with this material is still challenging due to its phase instability. In this work, we present the first instance of exploiting various machine learning (ML) models to efficiently predict the band gap and enthalpy of metastable γ-$CsPbI_3$ under pressure or strain while identifying key structural features that determine these properties. ML models trained on experimentally benchmarked, first-principles calculation datasets exhibit excellent performance in predicting the behavior of tuned systems, comparable to predictions made for ambient material databases. In particular, graph neural networks (GNNs) that explicitly include a graph encoding the bond angle information outperform other ML models in most scenarios. The pressure-tuned system demonstrates a strong linear relationship between structural features and properties, effectively captured by global structural features using linear regression models. In contrast, the strain-tuned system shows a non-linear relationship, exhibiting superior prediction performance using GNNs trained on local environments. This study opens up opportunities to apply and develop ML models for understanding and designing materials at extreme conditions.

Introduction

Halide perovskites have garnered significant attention in the past decade due to their remarkable performance as solar cell absorber materials[1–4]. These materials offer several advantages, including efficient light absorption and photoluminescence within the suitable band gap range, high structural flexibility for chemical and physical tuning, and low-cost production through solution processing[5–8]. Among halide perovskites, those with three-dimensional structures possess an ABX_3 chemical formula, where B-site cations and surrounding X-site halogen anions form octahedra that encapsulate A-site cations. By substituting different elements or molecules at each site, the optoelectronic properties of halide perovskites can be modified. All-inorganic halide perovskites, such as the ones with Cs^+ occupying the A site, offer high optoelectronic performance and stability to heat and humidity[9, 10]. $CsPbI_3$ perovskite phases exhibit proper band gaps of 1.6 - 1.8 eV, positioning them as promising candidates for photovoltaic applications within the $CsPbX_3$ family (X = Cl, Br, I)[11–15].

Despite the high chemical stability, $CsPbI_3$ exhibits phase instability under ambient conditions, with the functional perovskite phases (α, β, and γ) spontaneously transforming into the thermodynamically stable non-perovskite δ-phase $CsPbI_3$[16–20]. Given that δ-$CsPbI_3$ possesses a larger band gap and is unsuitable as a solar cell absorber, considerable efforts have been devoted to (meta)stabilizing the functional perovskite phases at ambient conditions[13, 14, 21]. Among the three perovskite-phased polymorphs, the metastable orthorhombic γ-phase $CsPbI_3$ has the lowest formation energy due to its substantial octahedral tilts[18, 22–25]. Recently, we successfully preserved the metastable γ-$CsPbI_3$ to ambient conditions by manipulating the octahedral tilting angles through temperature and pressure engineering[26], thereby unlocking the potential of utilizing metastable halide perovskites in practical applications.

Pressure and strain are powerful techniques for effectively tuning the structures and properties of halide perovskites, owing to these materials' soft lattice structures that are highly susceptible to these tuning mechanisms. Various efforts have been made to stabilize metastable halide perovskites at room temperature and fine-tune their band gaps and charge carrier mobilities through precise control of the structure using pressure[26–28] or strain[29–31] engineering. Thus, gaining a deeper understanding of the behavior and structure-property relationships of these materials perturbed by external stimuli can provide

insights for effectively manipulating them to achieve desirable optoelectronic properties.

However, experimental access to metastable halide perovskite phases and their characterization is challenging, especially under perturbed conditions. To amend this challenge, computational simulations, particularly density functional theory (DFT) calculations, are often used to study these systems under pressure and strain, albeit DFT calculations can be computationally expensive[32]. To overcome the limitation of prohibitively expensive DFT calculations, machine learning (ML) techniques have been widely applied in materials science for predicting material properties, components, and structures based on first-principles calculation databases[33–36]. By achieving accuracies comparable to DFT calculations, ML shows promise in replacing expensive calculations while substantially reducing the required computation time[32]. ML techniques ranging from classical models[37–40] to neural networks[41–44] have been utilized to understand and predict various material properties in diverse material systems.

Existing ML work in materials science has mostly been trained on databases containing materials primarily under ambient-pressure conditions. ML models trained on such databases may potentially overlook metastable phases that become accessible under extreme conditions[32]. In this study, for the first time, we predict material properties of a metastable phase – γ-$CsPbI_3$ – by training ML models specifically on pressurized or strained systems. We employ both classical ML models, such as linear regression and random forest, as well as newly developed graph neural networks (GNNs) to identify the important features that determine the physical properties of the tuned γ-$CsPbI_3$ system under extreme conditions. We specifically investigate the impact of local environments between neighboring atoms to gain a deeper understanding of the intricate structure-property relationships. By utilizing the simulated structures from high-throughput DFT calculations under pressure and strain that are benchmarked by experiments, we aim to contribute to the development of ML studies for efficiently predicting material properties across a spectrum of tuned structures.

Results

Training database generation

To simulate the structure and property changes, we computationally applied hydrostatic pressure or plane strain to the experimental γ-phase $CsPbI_3$ structure at ambient conditions[20]. Each tuned structure was

optimized using DFT variable cell relax-calculation with generalized gradient approximations (GGA)[45] exchange-correlation functionals as implemented in Quantum Espresso[46]. We generated a total of 241 pressure-tuned structures by applying hydrostatic pressure from 0 to 2.4 GPa. Additionally, we produced 505 strain-tuned structures by applying tensile or compressive plane strain along the *ab*, *bc*, or *ac* directions, ranging from -3.0 to +3.0%. Note that positive and negative strains denote tensile and compressive strains, respectively. Figure 1(a) illustrates the tuning processes under pressure and strain, while detailed information on the structure tuning procedures and DFT settings can be found in the Methods section. The selection of these specific pressure and strain ranges was motivated by previous experimental results reporting significant property changes, such as band gap change and phase transitions, at these levels[26, 30]. Pressure-tuning compresses the structure hydrostatically while strain-tuning isobarically deforms the system along certain directions. In Supplementary Information, we presented the DFT-calculated database and evaluated the accuracy of DFT calculations (Figures S1-S3 and S11-S12).

Classical ML model predictions based on global structural features

We extracted eight main features from the DFT-calculated structures, including in-phase and out-of-phase octahedral tilts[47], three lattice parameters, and three Pb-I bond lengths (Figure 1(c)-(e)). These structural features were chosen because they play a crucial role in determining the stability and properties of a perovskite phase[26, 48, 49]. To assess the effectiveness of these structural features in representing the pressure and strain-tuned γ-$CsPbI_3$ systems, we trained classical ML models using these features as inputs and the corresponding band gap and enthalpy values as outputs. Linear regression and random forest models were employed to evaluate the linear and non-linear relationships between the structure and properties. Table 1 presents the prediction accuracies of each model, expressed as the MAD:MAE (MAD: mean absolute deviation, MAE: mean absolute error) ratio averaged from 5-fold cross-validation. The MAD:MAE ratio, a loss function accounting for different scales of each target property, indicates good model performance when exceeding 5[50]. The coefficient of determination (R^2) and its variance over cross-validation are shown in Table S1 and Figures S4-S5. For the pressure-tuned system, linear regression achieved high prediction accuracy with MAD:MAE ratios of 24.63 and 60.13 for band gap and enthalpy, respectively. In contrast, it exhibited poor performance in the strain-tuned system with MAD:MAE ratios of 2.28 for band gap and 4.32 for enthalpy. These results indicate a strong linear relationship between the eight structural features

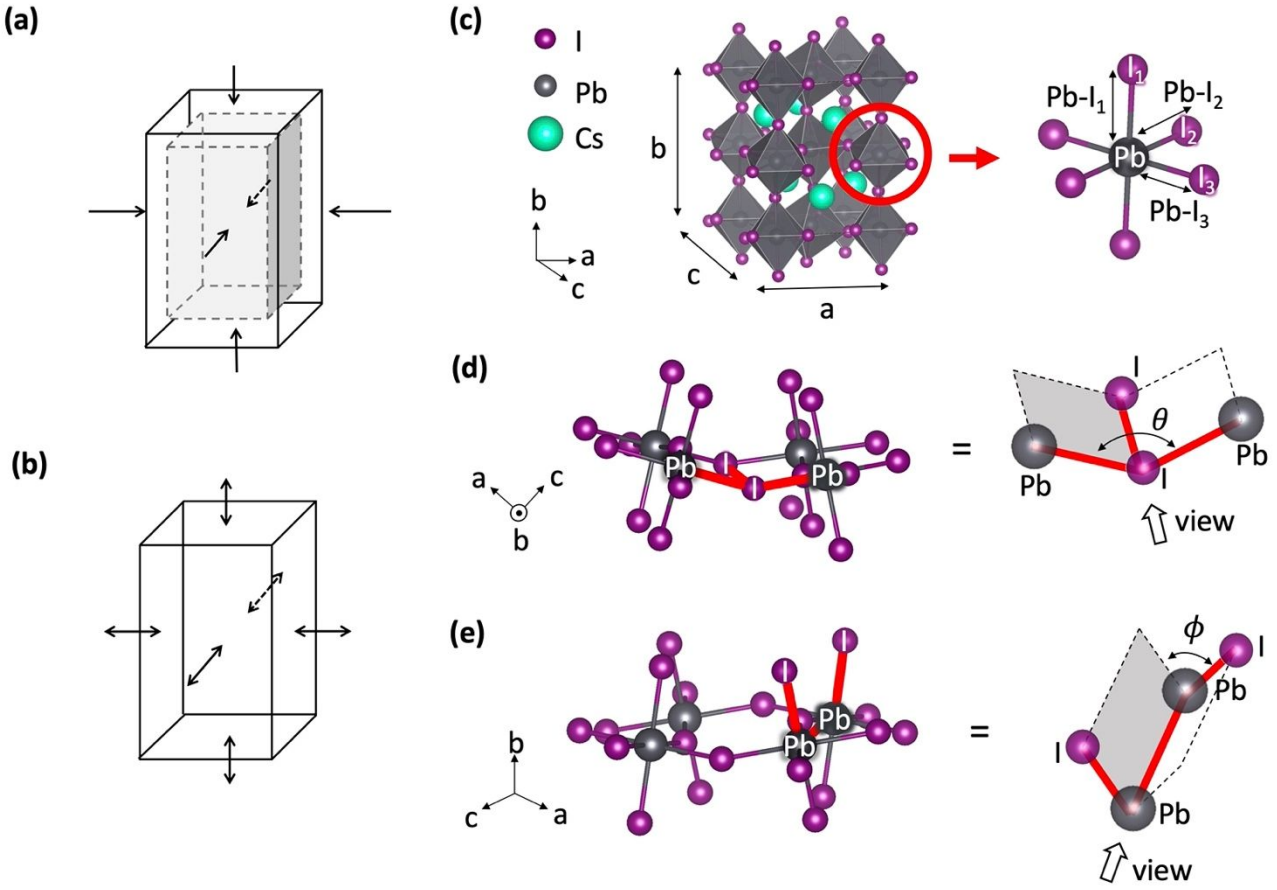

Figure 1. Illustrations of (a) pressure-tuning and (b) strain-tuning mechanisms. The hydrostatic pressure is uniformly applied to the unit cell from all directions, resulting in a reduction of the unit cell volume. The solid lines represent the original unit cell, while the grey dashed lines depict the deformed unit cell. On the other hand, in the strain-tuning, we deformed the unit cell along two lattice parameters (combinations of *ab*, *bc*, or *ac*), while adjusting the other lattice parameter (*c*, *a*, or *b*, respectively) to maintain a constant volume. Additionally, (c)-(e) illustrate the eight global structural features used to represent the γ-$CsPbI_3$ system for training classical ML models. Specifically, (c) displays the lattice parameters *a*, *b*, and *c* (left) and the Pb-I bond lengths Pb-I_1, Pb-I_2, and Pb-I_3 within the octahedron (right). (d) In-phase tilt ([180-θ(Pb-I-I-Pb)]/2) and (e) out-of-phase tilt (ϕ(I-Pb-Pb-I)/2) are dihedral angles that characterize the octahedral tilts. (d) and (e) were reproduced from reference[26] by Ke *et al.* with permission from *Nature Communications*, copyright 2021.

and output properties in the pressure-tuned system. Conversely, the strain-tuned system showed improved accuracies using random forest, with MAD:MAE ratios of 13.37 and 8.60 for band gap and enthalpy, respectively. However, the pressure-tuned system demonstrated slightly decreased performance with the random forest model, revealing a non-linear relationship between structural features and properties in

the strain-tuned γ-$CsPbI_3$ system. The non-linearity observed in the strain-tuned system is evident in the linear regression band gap prediction results, where several data points form distinct shapes, deviating from the ideal prediction line (Figure S4(c)). These distinct shapes were not observed in the random forest prediction results, which account for non-linearity in the data.

MAD:MAE		Linear Regression	Random Forest	CGCNN	ALIGNN
Band gap	Pressure	**24.63**	6.82	8.05	14.00
	Strain	2.28	13.37	9.01	**32.64**
Enthalpy	Pressure	60.13	56.90	51.85	**65.26**
	Strain	4.32	8.60	17.06	**35.16**

Table 1. Summary of prediction performance on test sets. MAD:MAE value presents the prediction accuracies of each model and indicates good model performance when exceeding 5. Each MAD:MAE value is averaged from the 5-fold cross-validation. The best prediction results for each target are indicated in bold.

Overall, the results indicate that the eight structural features effectively capture the structure-property relationships with high accuracy. In the process of identifying the primary structural features that influence the band gap and enthalpy predictions, it was observed that these features exhibit significant mutual correlation (Figure S8). The presence of strong multicollinearity among input features poses challenges in accurately evaluating and interpreting feature importance, which can lead to potentially misleading conclusions[51, 52]. We analyzed the impact of multicollinearity on evaluating feature importance in the Supplementary Information and assessed the performance of new models restructured based on the features identified through some of the feature selection techniques (Figures S8-S10). For the strain-tuned system, the new random forest models demonstrated similar prediction accuracy to the original models. However, the new linear regression models for the pressure-tuned system exhibited a notable decrease in prediction accuracy, particularly for the band gap. This decrease can be attributed to the high degree of correlation among the eight global features within the pressure-tuned system, including instances where octahedral tilts and Pb-I lengths are inherent components of lattice parameters. Thus, to enhance the prediction accuracy of the pressure-tuned system, it is essential to consider all eight features together as inputs.

Impact of local environments in GNN predictions

GNN models are based on a message-passing algorithm, which enables the inclusion of interactions between neighboring atoms. Unlike the classical models trained solely on global features, GNNs utilize

the coordinates of the entire unit cell structure and learn from local environments[42]. In this context, local environments encompass not only the interactions between the central atom and its neighboring atoms but also the interactions between neighboring atoms and their further neighbors. These intricate local details are captured and encoded in the crystal graph by updating the node corresponding to each atom. In this study, we trained and compared two GNNs: crystal graph convolutional neural network (CGCNN)[43] and atomistic line graph neural network (ALIGNN)[44]. Both models represent the crystal structure of interest by constructing a graph composed of nodes and edges, encoding atom and bond information, respectively. The key distinction between the two models is that ALIGNN incorporates a line graph that describes bond angle information among three atoms. We compare these models to examine the effect of considering angle information on the ML performance, particularly considering that the octahedral tilting angles were identified as significant structural features determining the bandgap and enthalpy of the γ-$CsPbI_3$ system[26]. Although the three-body bond angles considered in the ALIGNN model are distinct from the dihedral in-phase and out-of-phase tilting angles observed in the γ-$CsPbI_3$ system, they are closely related.

Table 1 summarizes the performance of CGCNN and ALIGNN for band gap and enthalpy predictions in both the pressure- and strain-tuned systems (Figure S6-S7). ALIGNN consistently outperforms CGCNN, indicating that the inclusion of bond angles as inputs during the learning process enhances the prediction accuracy of GNNs, in agreement with previous studies. Chaudhary and DeCost conducted a similar comparison on the Materials Project (MP) database[44, 53], demonstrating ALIGNN's superior performance, with approximately 1.8 times higher MAD:MAE ratio for band gap and formation energy predictions. In our study, ALIGNN exhibits a better performance than CGCNN, showing 3.62- and 2.06-times as high MAD:MAE ratios for the band gap and enthalpy predictions in the strain-tuned system. This indicates a more substantial improvement than the ambient MP database training results. Conversely, the pressure-tuned system shows 1.74- and 1.26-times as high MAD:MAE ratios for band gap and enthalpy predictions from ALIGNN compared to CGCNN, demonstrating comparable or slightly lower values than those obtained from the MP database training results. These findings suggest that the strain-tuned system is more responsive to the inclusion of local bond angle information than the ambient or pressure-tuned crystal structures on average.

Comparing the performance of classical models and GNNs as shown in Table 1, different trends emerge depending on the tuning methods. For pressure-tuning, classical models yield higher or comparable accuracies compared to GNNs, indicating that global features effectively capture the γ-$CsPbI_3$ system's behavior without considering local interactions. In contrast, the strain-tuned system exhibits the highest prediction accuracies with GNNs, particularly with ALIGNN (Figure 2). This suggests that the inclusion of local bond angle information benefits the prediction accuracy of the strain-tuned system.

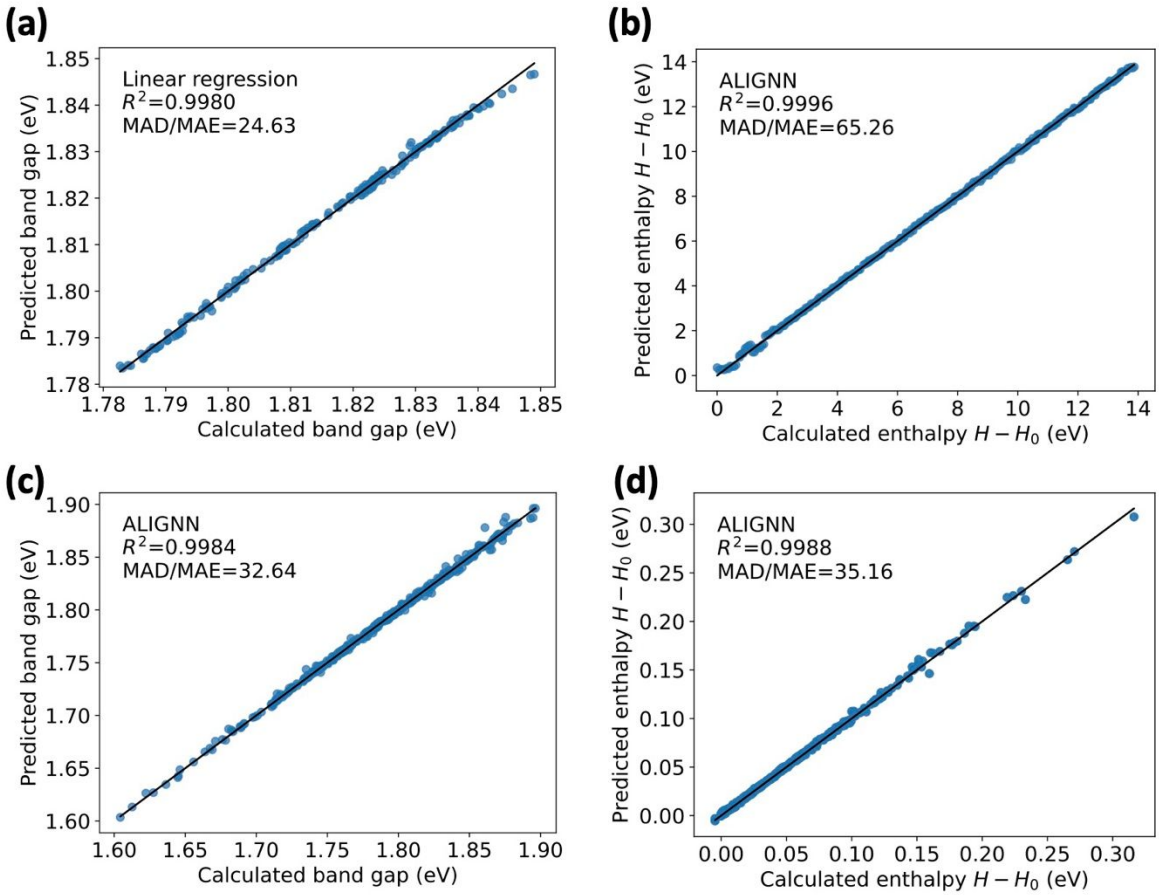

Figure 2. Prediction results from the best performing models for each target property. Pressure-tuned (a) band gap by linear regression and (b) enthalpy by ALIGNN. Strain-tuned (c) band gap and (d) enthalpy by ALIGNN. Solid lines are X=Y curves that represent the ideal predictions. In (a) and (d), the prediction results exhibit underestimation for higher band gap and enthalpy values. This discrepancy can be attributed to the data imbalance problem, where the training data had fewer input structures in these regions. As a result, the model underestimated the target properties for the corresponding structures.

Discussion

In this study, we explored the predictive capabilities of ML models for physical properties in pressure- and strain-tuned γ-$CsPbI_3$ systems, highlighting the influence of different tuning methods on model performance. Our findings indicate that ML models can effectively predict the perturbed systems as accurately as ambient-conditioned materials, thus demonstrating the potential of ML techniques for exploring materials under extreme conditions. In particular, classical models perform comparably well to GNNs in predicting properties in the pressure-tuned system, leveraging the representation of eight structural features and revealing a strong linear structure-property relationship. In contrast, the strain-tuned system exhibits improved prediction accuracy when incorporating local interactions and embracing more non-linearity using GNNs.

The observed discrepancy in model performance between pressure- and strain-tuned systems can be attributed to the distinct nature of their tuning mechanisms. Pressure tuning primarily entails progressive and hydrostatic modification of the structures, whereas strain tuning involves deformation along specific directions. This anisotropic distortion likely introduces non-linear structure-property effects into the dataset. Notably, the predictions from the pressure-tuned system primarily rely on global features, encompassing octahedral tilting angles, Pb-I bond lengths, and lattice parameters. In contrast, the strain-tuned system is better characterized by local environments, including local bond angles and chemical information.

The superior performance of classical models compared to GNNs in the pressure-tuned system highlights the limitation of GNN models in capturing global features, as previously noted in related research by Gong *et al.*[54] If GNNs could effectively consider all structural features during the learning process, they would offer broader applicability across various systems, including the pressure-tuned system like the one studied in this work. Therefore, further work is needed to combine GNNs with global structural feature considerations and establish a larger database encompassing accurate property values from pressure- and strain-tuned systems. This would facilitate the generalization of ML models for predicting properties in diverse perovskite systems, including those with lower-dimensional structures, and enable better comparison with experimental data.

Methods

Structure tuning

The γ-$CsPbI_3$ system was computationally tuned by applying hydrostatic pressure or introducing tensile or compressive plane strains to the experimental ambient structure[20]. The initial experimental structure was optimized at 0 GPa using DFT relax-calculation. From this zero-pressure optimized structure, a total of 241 pressure-tuned structures were obtained at pressures ranging from 0 GPa to 2.40 GPa with a 0.01 GPa increment. For strain tuning, 505 structures were created by deforming two of the lattice parameters from -3.0 to +3.0% with a step size of 0.5%, while maintaining a constant volume and the zero-pressure condition. Note that positive and negative strains indicate tensile and compressive strains, respectively. For instance, if the lengths of lattice parameters *a* and *b* were elongated by +3.0% each to simulate tensile strain, the *c* parameter was compressed to maintain the initial volume. Within the specified pressure and strain ranges, both the pressure- and strain-tuned systems showed potential phase transitions marked by discontinuous structural changes at certain pressure and strain levels. These findings are elaborately detailed in the Supplementary Information.

First-principles DFT calculations

The strain-tuned γ-$CsPbI_3$ structures were optimized using DFT relax-calculation, while the pressure-tuned structures were optimized using variable-cell relax-calculation with Quantum Espresso[46]. Ultrasoft pseudopotentials with PBE exchange-correlation functionals[55] were selected, and a Monkhorst-Pack K-point grid of 4x3x4 centered on the Γ point was used[56]. The kinetic energy cutoff was set at 75 Ry for wavefunctions and 500 Ry for the charge density. From the optimized structures, the band gap value at the Γ point and the enthalpy ($H = E + PV$) at the respective pressure or strain conditions were obtained. Supplementary Information provides a comparison between results obtained from different exchange-correlation functionals and band gap corrections.

Classical models

Multiple linear regression and random forest models were trained using structural features as inputs and the corresponding band gap and enthalpy as outputs. Classical models were employed initially to assess how well the global features could represent the tuned γ-$CsPbI_3$ system for band gap and enthalpy

predictions. Each model utilized eight structural parameters as inputs and predicted band gap or enthalpy for the corresponding pressure and strain levels, which included 241 and 505 structures, respectively. The random forest model consisted of 200 estimators (decision trees). The models underwent 5-fold cross-validation, and the loss functions from each fold were averaged. All the classical models in this study were implemented using the Scikit-learn library in Python[57].

GNN models

Two GNN models, CGCNN[43] and ALIGNN[44], were employed for band gap and enthalpy prediction. CGCNN and ALIGNN represent the crystal structure as a graph, with nodes representing atoms and edges encoding atomic bonds. CGCNN utilizes convolutional neural networks with convolutional and pooling layers on the crystal graph. ALIGNN, in addition to the crystal graph, incorporates a line graph that captures bond angle information. The optimized structures in Crystallographic Information Framework (CIF) format from the training dataset were used to train these models, which predicted target properties such as band gap and enthalpy. The dataset was divided into a training set (60%), a validation set (20%), and a testing set (20%), with training conducted over 200 epochs. To eliminate biases resulting from different property scales, the target properties were standardized to have a mean value of zero and a standard deviation of one.

Model evaluation

To evaluate the performance of the models, MAE, MAD, and coefficients of determination (R^2) were calculated using the ML-predicted and DFT-calculated target values in the testing dataset. Since different chemistry and properties have different scales, evaluating model prediction accuracy solely based on MAE values can be misleading. Therefore, the MAD:MAE ratios are presented in Table 1, which were obtained by dividing MAE by MAD, a metric used by Choudhary and DeCost[44]. The MAD:MAE ratios allow for an unbiased comparison of model performance across different properties. Models with MAD:MAE ratios beyond 5 are generally considered to be well-performing, with higher ratios indicating higher prediction accuracy[50].

Data availability

The complete input database for γ-$CsPbI_3$ utilized in this study can be accessed on GitHub at https://github.com/mhan8/Metastable_ML.

Code availability

The ML models trained using our database and their respective training configurations are available on GitHub at https://github.com/mhan8/Metastable_ML.

Acknowledgements

We thank Samuel Girdzis for his contribution to collecting experimental data for $CsPbBr_3$. This work is supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Division of Materials Sciences and Engineering, under contract No. DE-AC02-76SF00515. R.S. acknowledges support by the U.S. Department of Energy, Laboratory Directed Research and Development program at SLAC National Accelerator Laboratory, under contract No. DE-AC02-76SF00515. C.P. also acknowledges support by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Grant No. DE-SC0022216. Some of the computing for this project was performed on the Sherlock cluster. We would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results. Parts of the computation work used resources of the National Energy Research Scientific Computing Center, a Department of Energy Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Author contributions

Y.L. designed the project. C.J. and Y.L. supervised and guided the project direction. M.H. performed ML and DFT work. C.P., R.S., and C.J. evaluated the accuracy of DFT functionals. Y.S.G.N. helped evaluate ML models. F.K., W.L.M., and Y.L. provided experimental support. M.H. and Y.L. wrote the paper with all authors contributing to the discussion and revision of the paper.

References

1. Tao, Q., Xu, P., Li, M. & Lu, W. Machine learning for perovskite materials design and discovery. *npj Comput. Mater.* **7**, 23 (2021).

2. NREL. Best research-cell efficiency chart. *figshare* https://www.nrel.gov/pv/cell-efficiency.html (2023).

3. Tress, W. Maximum efficiency and open-circuit voltage of perovskite solar cells. *Organic-Inorganic Halide Perovskite Photovoltaics: From Fundamentals to Device Archit.* 53–77 (2016).

4. Park, N.-G., Miyasaka, T. & Grätzel, M. Organic-inorganic halide perovskite photovoltaics. *Cham, Switzerland: Springer* (2016).

5. Kojima, A., Teshima, K., Shirai, Y. & Miyasaka, T. Organometal halide perovskites as visible-light sensitizers for photovoltaic cells. *J. american chemical society* **131**, 6050–6051 (2009).

6. Deschler, F. *et al.* High photoluminescence efficiency and optically pumped lasing in solution-processed mixed halide perovskite semiconductors. *The journal physical chemistry letters* **5**, 1421–1426 (2014).

7. Ahmadi, M., Wu, T. & Hu, B. A review on organic–inorganic halide perovskite photodetectors: device engineering and fundamental physics. *Adv. Mater.* **29**, 1605242 (2017).

8. Arabpour Roghabadi, F., Ahmadi, V. & Oniy Aghmiuni, K. Organic–inorganic halide perovskite formation: in situ dissociation of cation halide and metal halide complexes during crystal formation. *The J. Phys. Chem. C* **121**, 13532–13538 (2017).

9. Beal, R. E. *et al.* Cesium lead halide perovskites with improved stability for tandem solar cells. *The journal physical chemistry letters* **7**, 746–751 (2016).

10. Yang, D. *et al.* All-inorganic cesium lead halide perovskite nanocrystals: synthesis, surface engineering and applications. *J. Mater. Chem. C* **7**, 757–789 (2019).

11. Eperon, G. E. *et al.* Inorganic caesium lead iodide perovskite solar cells. *J. Mater. Chem. A* **3**, 19688–19695 (2015).

12. Wang, J. *et al.* 21.15%-efficiency and stable γ-cspbi3 perovskite solar cells enabled by an acyloin ligand. *Adv. Mater.* 2210223 (2023).

13. Swarnkar, A. *et al.* Quantum dot–induced phase stabilization of α-cspbi$_3$ perovskite for high-efficiency photovoltaics. *Science* **354**, 92–95 (2016).

14. Wang, Y. *et al.* Thermodynamically stabilized β-cspbi$_3$ –based perovskite solar cells with efficiencies> 18%. *Science* **365**, 591–595 (2019).

15. Wang, Y., Chen, Y., Zhang, T., Wang, X. & Zhao, Y. Chemically stable black phase cspbi3 inorganic perovskites for high-efficiency photovoltaics. *Adv. Mater.* **32**, 2001025 (2020).

16. Dastidar, S. *et al.* Quantitative phase-change thermodynamics and metastability of perovskite-phase cesium lead iodide. *The journal physical chemistry letters* **8**, 1278–1282 (2017).

17. Yang, Z. *et al.* Impact of the halide cage on the electronic properties of fully inorganic cesium lead halide perovskites. *ACS Energy letters* **2**, 1621–1627 (2017).

18. Sutton, R. J. *et al.* Cubic or orthorhombic? revealing the crystal structure of metastable black-phase cspbi$_3$ by theory and experiment. *ACS Energy Lett.* **3**, 1787–1794 (2018).

19. Wang, B., Novendra, N. & Navrotsky, A. Energetics, structures, and phase transitions of cubic and orthorhombic cesium lead iodide (cspbi$_3$) polymorphs. *J. Am. Chem. Soc.* **141**, 14501–14504 (2019).

20. Straus, D. B., Guo, S. & Cava, R. J. Kinetically stable single crystals of perovskite-phase cspbi$_3$. *J. Am. Chem. Soc.* **141**, 11435–11439 (2019).

21. Sutton, R. J. *et al.* Bandgap-tunable cesium lead halide perovskites with high thermal stability for efficient solar cells. *Adv. Energy Mater.* **6**, 1502458 (2016).

22. Marronnier, A. *et al.* Anharmonicity and disorder in the black phases of cesium lead iodide used for stable inorganic perovskite solar cells. *ACS nano* **12**, 3477–3486 (2018).

23. Masi, S., Gualdrón-Reyes, A. F. & Mora-Sero, I. Stabilization of black perovskite phase in fapbi3 and cspbi$_3$. *ACS Energy Lett.* **5**, 1974–1985 (2020).

24. Woodward, P. M. Octahedral tilting in perovskites. ii. structure stabilizing forces. *Acta Crystallogr. Sect. B: Struct. Sci.* **53**, 44–66 (1997).

25. Li, Z. *et al.* Stabilizing perovskite structures by tuning tolerance factor: Formation of formamidinium and cesium lead iodide solid-state alloys. *Chem. Mater.* **28**, 284–292 (2016).

26. Ke, F. *et al.* Preserving a robust cspbi$_3$ perovskite phase via pressure-directed octahedral tilt. *Nat. communications* **12**, 461 (2021).

27. Beimborn, J. C., Hall, L. M., Tongying, P., Dukovic, G. & Weber, J. M. Pressure response of photoluminescence in cesium lead iodide perovskite nanocrystals. *The J. Phys. Chem. C* **122**, 11024–11030 (2018).

28. Ma, Z. *et al.* Pressure-induced emission of cesium lead halide perovskite nanocrystals. *Nat. Commun.* **9**, 4506 (2018).

29. Steele, J. A. *et al.* Thermal unequilibrium of strained black cspbi$_3$ thin films. *Science* **365**, 679–684 (2019).

30. Chen, Y. *et al.* Strain engineering and epitaxial stabilization of halide perovskites. *Nature* **577**, 209–215 (2020).

31. Zhu, C. *et al.* Strain engineering in perovskite solar cells and its impacts on carrier dynamics. *Nat. communications* **10**, 815 (2019).

32. Schmidt, J., Marques, M. R., Botti, S. & Marques, M. A. Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* **5**, 83 (2019).

33. Liu, Y., Zhao, T., Ju, W. & Shi, S. Materials discovery and design using machine learning. *J. Materiomics* **3**, 159–177 (2017).

34. Pilania, G. *et al.* Machine learning bandgaps of double perovskites. *Sci. reports* **6**, 19375 (2016).

35. Hsu, T. *et al.* Efficient and interpretable graph network representation for angle-dependent properties applied to optical spectroscopy. *npj Comput. Mater.* **8**, 151 (2022).

36. Lu, S. *et al.* Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nat. communications* **9**, 3405 (2018).

37. Rajan, A. C. *et al.* Machine-learning-assisted accurate band gap predictions of functionalized mxene. *Chem. Mater.* **30**, 4031–4038 (2018).

38. Lee, J., Seko, A., Shitara, K., Nakayama, K. & Tanaka, I. Prediction model of band gap for inorganic compounds by combination of density functional theory calculations and machine learning techniques. *Phys. Rev. B* **93**, 115104 (2016).

39. Zhuo, Y., Mansouri Tehrani, A. & Brgoch, J. Predicting the band gaps of inorganic solids by machine learning. *The journal physical chemistry letters* **9**, 1668–1673 (2018).

40. Legrain, F., Carrete, J., van Roekeghem, A., Curtarolo, S. & Mingo, N. How chemical composition alone can predict vibrational free energies and entropies of solids. *Chem. Mater.* **29**, 6220–6227 (2017).

41. Hong, Y., Hou, B., Jiang, H. & Zhang, J. Machine learning and artificial neural network accelerated computational discoveries in materials science. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **10**, e1450 (2020).

42. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for quantum chemistry. In *International conference on machine learning*, 1263–1272 (PMLR, 2017).

43. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. review letters* **120**, 145301 (2018).

44. Choudhary, K. & DeCost, B. Atomistic line graph neural network for improved materials property predictions. *npj Comput. Mater.* **7**, 185 (2021).

45. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. review letters* **77**, 3865 (1996).

46. Giannozzi, P. *et al.* Quantum espresso: a modular and open-source software project for quantum simulations of materials. *J. physics: Condens. matter* **21**, 395502 (2009).

47. Glazer, A. M. The classification of tilted octahedra in perovskites. *Acta Crystallogr. Sect. B: Struct. Crystallogr. Cryst. Chem.* **28**, 3384–3392 (1972).

48. Amat, A. *et al.* Cation-induced band-gap tuning in organohalide perovskites: interplay of spin–orbit coupling and octahedra tilting. *Nano letters* **14**, 3608–3616 (2014).

49. Garcia-Fernandez, P., Aramburu, J., Barriuso, M. & Moreno, M. Key role of covalent bonding in octahedral tilting in perovskites. *The J. Phys. Chem. Lett.* **1**, 647–651 (2010).

50. Ward, L., Agrawal, A., Choudhary, A. & Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. npj comput mater 2: 16028 (2016).

51. Biau, G. & Scornet, E. A random forest guided tour. *Test* **25**, 197–227 (2016).

52. Chan, J. Y.-L. *et al.* Mitigating the multicollinearity problem and its machine learning approach: a review. *Mathematics* **10**, 1283 (2022).

53. Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1** (2013).

54. Gong, S., Xie, T., Shao-Horn, Y., Gomez-Bombarelli, R. & Grossman, J. C. Examining graph neural networks for crystal structures: limitations and opportunities for capturing periodicity. *arXiv preprint arXiv:2208.05039* (2022).

55. Perdew, J. P., Burke, K. & Ernzerhof, M. Perdew, burke, and ernzerhof reply. *Phys. Rev. Lett.* **80**, 891 (1998).

56. Monkhorst, H. J. & Pack, J. D. Special points for brillouin-zone integrations. *Phys. review B* **13**, 5188 (1976).

57. Pedregosa, F. *et al.* Scikit-learn: Machine learning in python. *J. machine Learn. research* **12**, 2825–2830 (2011).

58. Hinuma, Y., Pizzi, G., Kumagai, Y., Oba, F. & Tanaka, I. Band structure diagram paths based on crystallography. *Comput. Mater. Sci.* **128**, 140–184 (2017).

59. Hubbard, C. R. & Calvert, L. D. The pearson symbol. *Bull. Alloy. Phase Diagrams* **2**, 153–157 (1981).