Supporting Information

Designing Antiperovskite Derivatives via Atomic-Position Splitting for Photovoltaic Applications

Gang Tang[1,2,*], Xiaohan Liu[1], Shihao Wang[1], Tao Hu[4], Chunbao Feng[4], Cheng Zhu[1], Bonan Zhu[3], Jiawang Hong[2,3,#]

[1]Advanced Research Institute of Multidisciplinary Science, Beijing Institute of Technology, Beijing 100081, China

[2]Beijing Institute of Technology, Zhuhai Beijing Institute of Technology (BIT) Zhuhai 519088, P.R.China

[3]School of Aerospace Engineering, Beijing Institute of Technology, Beijing, 100081, China

[4]School of Science, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

E-mails:
*Gang Tang: gtang@bit.edu.cn or gangtang@outlook.com;
#Jiawang Hong: hongjw@bit.edu.cn.

S1 Ground-State Crystal Structure

For the cubic phase of antiperovskite derivatives, the presence of imaginary frequencies in the phonon dispersion curves corresponds to structural instabilities and indicates a lower symmetry ground state. To stabilize these unstable modes, we adopted the approach of freezing individual or pairs of unstable phonon modes[1,2]. Specifically, we displaced the atoms according to the force constant eigenvectors and created a pool of subgroup structures. Each of the obtained subgroup structures was then subjected to a full structural relaxation. Subsequently, the lowest energy structure was selected for phonon calculations to further check for additional imaginary modes. This process was repeated iteratively until phonon dispersion curves without imaginary frequencies were achieved.

S2 Machine Learning (ML) Procedures

S2.1 Compositional Descriptors and Feature Engineering

The descriptors used for training the ML model were automatically generated from the chemical compositions using a composition-based feature vector (CBFV) approach[3]. The CBFV is a widely adopted method for transforming chemical compositions into usable features and is generated from the descriptive statistics (such as maximum, minimum, composition-weighted average) of a compound's constituent element properties. Finally, a total of 264 descriptors were generated to construct the initial descriptor set (see Table S6) for ML.

To reduce the dimension of the descriptor set and avoid overfitting, feature engineering is a crucial step to improve the fitting accuracy and performance of the ML model. Initially, we employed the analysis of variance (ANOVA) method[4] from the scikit-learn package[5] to eliminate all features whose variance falls below a predefined threshold (threshold = 0.01). Subsequently, the Pearson correlation coefficients[6] (ρ) were calculated for each feature pair to identify and remove redundant raw features. Redundant descriptors were defined as those with $|\rho| \geq 0.90$. Lastly, we utilized the sequential forward selection (SFS) technique from the mlxtend package[7] to further obtain the optimal set of descriptors (see Figure S21a and c).

S2.2 ML Model Selection

We chose the eXtreme Gradient Boosting (XGBoost) algorithm[8] as the ML regression model in our work due to its accuracy, efficiency, and ease of use. Currently, XGBoost-based ML models have demonstrated excellent accuracy in predicting various material properties, such as electrocaloric temperature change[9], gas separation selectivity[10], perovskite catalytic properties[11], Debye temperature[12], and thermal conductivity[13]. The Bayesian optimization algorithm in the BayesianOptimization package[14] was used to optimize hyperparameters with five-fold cross-validation (see Table S7). The optimal parameters were determined based on the highest R^2 (the coefficient of determination) value achieved, with over 500 steps of Bayesian optimization performed. Additionally, the loss function of the XGBoost model for E_{hull} and E_{g}, showing the training and validation performance over iterations, is presented in Figure S21b and d.

To perform symbolic regression, we adopted the SISSO (Sure Independence Screening and Sparsifying Operator) algorithm[15], which combines sure independence screening (SIS) with the sparsifying operator (SO) to select a subspace of descriptors with the largest linear correlation with the targeted property. For establishing the feature spaces, we utilized the set of algebraic and functional operators given in the following:

$$\mathrm{H}^{(\mathrm{m})} \equiv \{+, -, \times, \div, \sqrt{}, \exp, ^{-1}, ^{2}, ^{3}, ^{4}\} \ (1)$$

S2.3 ML Model performance evaluation

The ML models from the five-fold cross-validation method were evaluated using two metrics: the root mean square error (RMSE) and the coefficient of determination (R^2). These metrics are defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_{i\_\text{pre}})^2} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} y_i - y_{i\_\text{pre}}}{\sum_{i=1}^{n} y_i - \bar{y}} \quad (3)$$

where y_i and y_{i_pre} are the real value and predicted result of the sample i, respectively, and $\bar{y}$ is the average of all the real values.

S2.4 Interpretation of the ML Model

To explain the output of the XGBoost ML model, we performed a SHAP (SHapley Additive exPlanations)[16] analysis, which is a game theoretic approach. This analysis combines feature importance with feature effects, showing the distribution of the Shapley values for each feature. Shapley values assign a value to each feature in a prediction, indicating how much each feature contributes to the difference between the model prediction and the expected prediction.

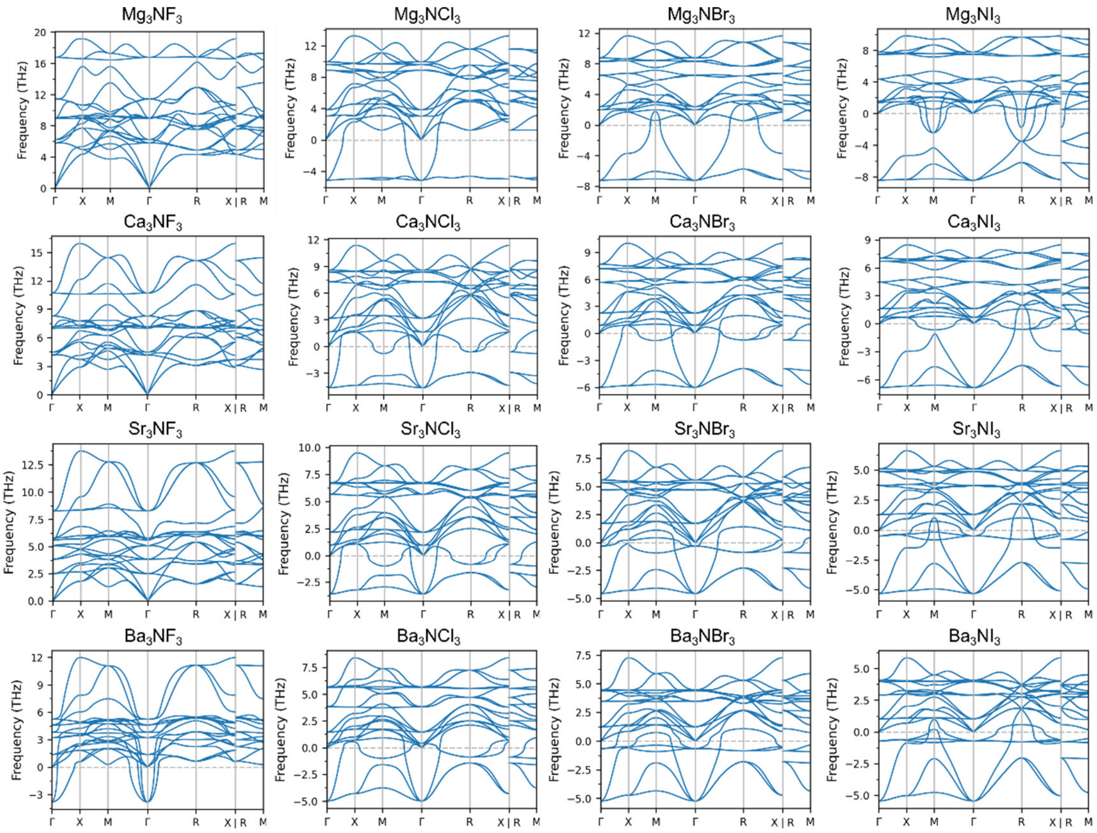
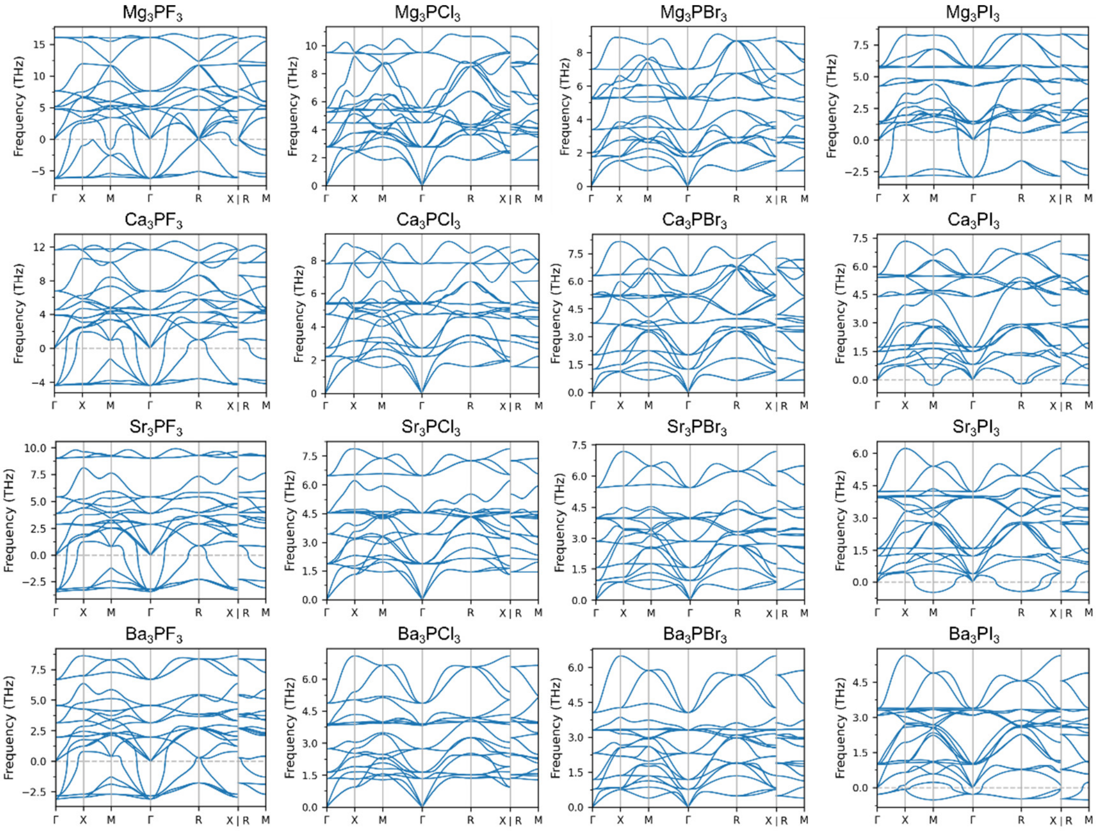


Figure S1. Phonon dispersion curves for the *Pm*-3*m* phase of $X_3NA'_3$ (X = Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+} and A′ = F^-, Cl^-, Br^-, I^-).



Mg_3PF_3 Mg_3PCl_3 Mg_3PBr_3 Mg_3PI_3

Ca_3PF_3 Ca_3PCl_3 Ca_3PBr_3 Ca_3PI_3

Sr_3PF_3 Sr_3PCl_3 Sr_3PBr_3 Sr_3PI_3

Ba_3PF_3 Ba_3PCl_3 Ba_3PBr_3 Ba_3PI_3

Figure S2. Phonon dispersion curves for the *Pm*-3*m* phase of $X_3PA'_3$ (X = Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+} and A′ = F^-, Cl^-, Br^-, I^-).

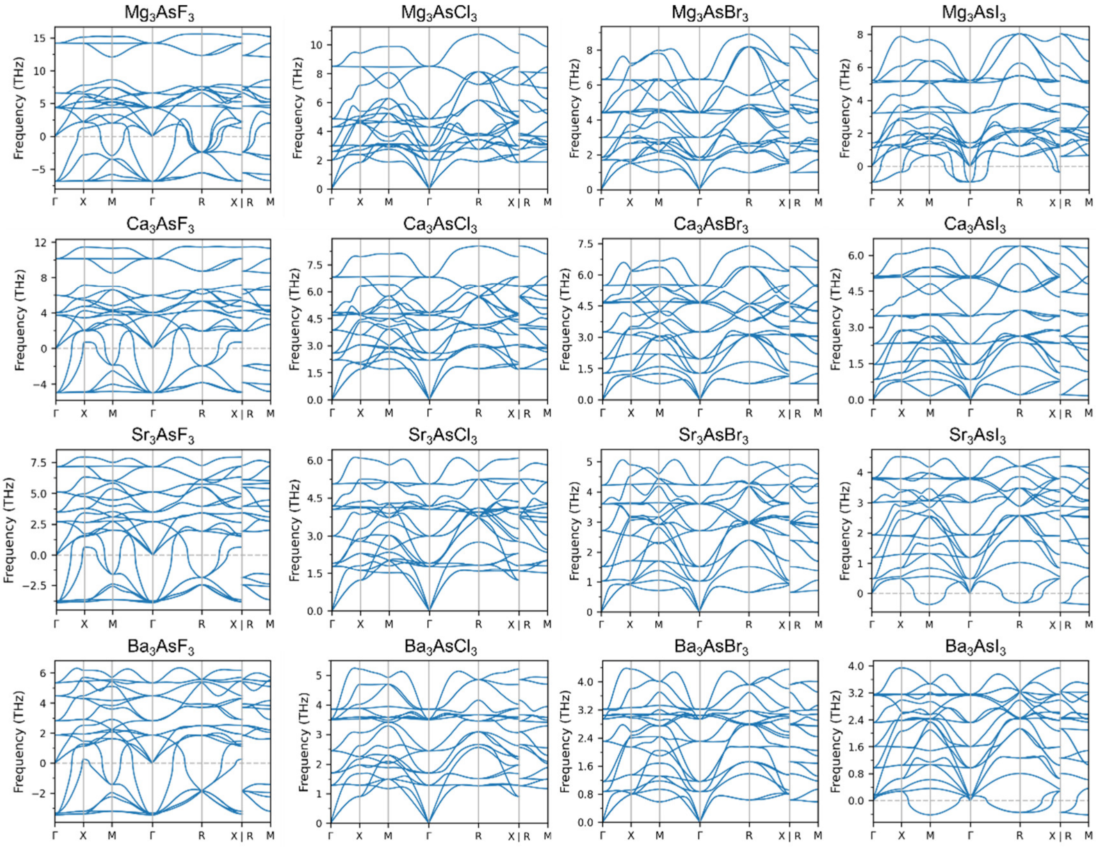
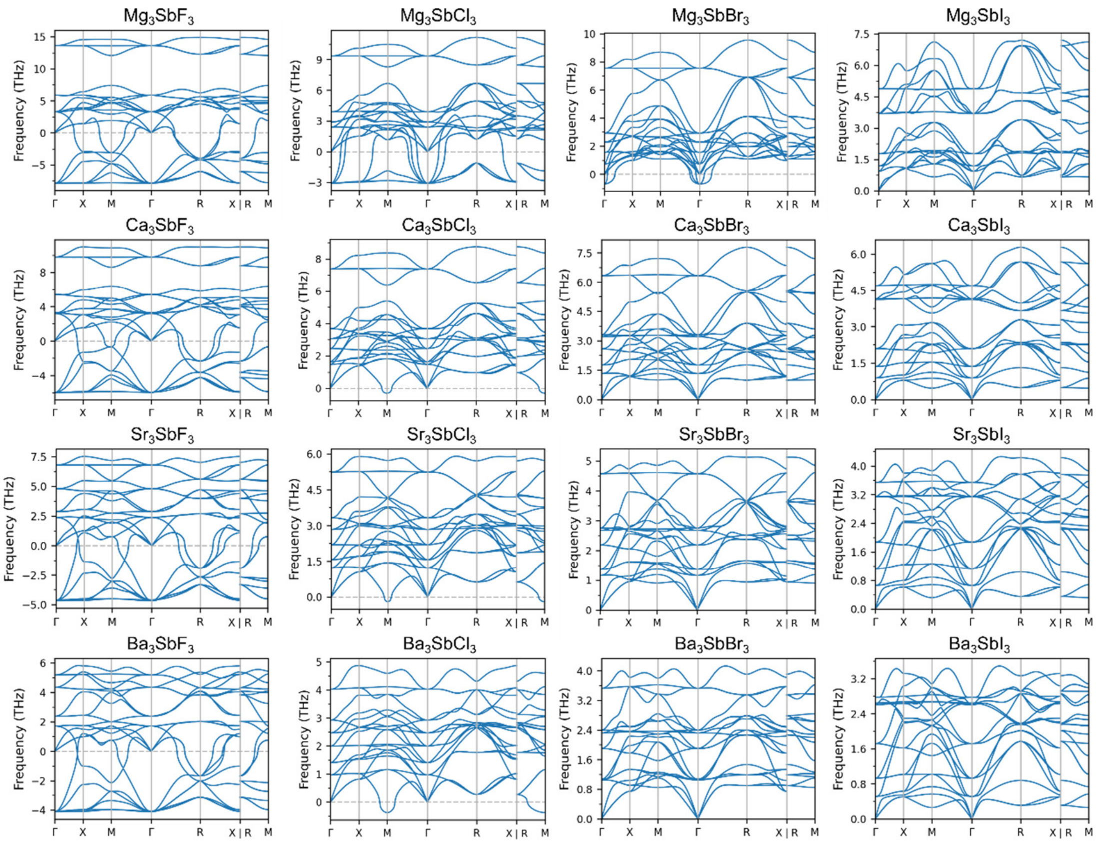


Figure S3. Phonon dispersion curves for the *Pm*-3*m* phase of X$_3$AsA′$_3$ (X = Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+} and A′ = F$^-$, Cl$^-$, Br$^-$, I$^-$).



Mg_3SbF_3 Mg_3SbCl_3 Mg_3SbBr_3 Mg_3SbI_3

Ca_3SbF_3 Ca_3SbCl_3 Ca_3SbBr_3 Ca_3SbI_3

Sr_3SbF_3 Sr_3SbCl_3 Sr_3SbBr_3 Sr_3SbI_3

Ba_3SbF_3 Ba_3SbCl_3 Ba_3SbBr_3 Ba_3SbI_3

Figure S4. Phonon dispersion curves for the *Pm*-3*m* phase of $X_3SbA'_3$ (X = Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+} and A′ = F^-, Cl^-, Br^-, I^-).

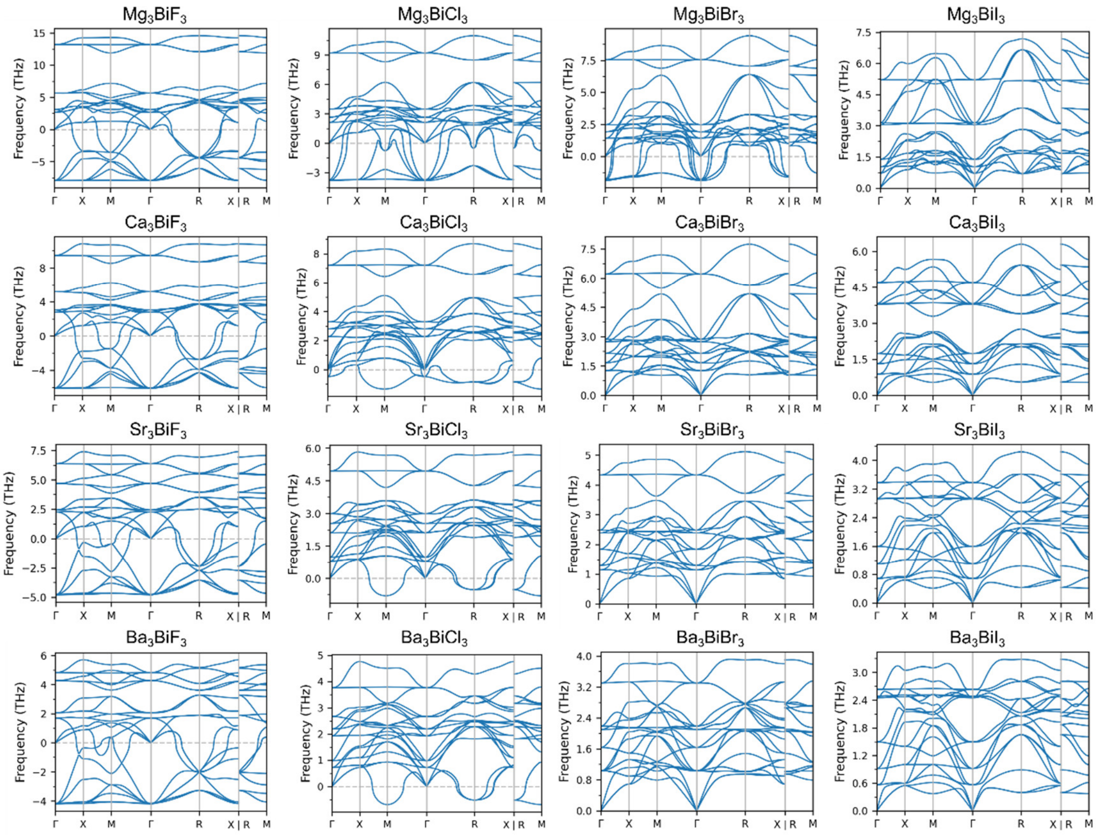


Figure S5. Phonon dispersion curves for the *Pm*-3*m* phase of $X_3BiA'_3$ (X = Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+} and A′ = F^-, Cl^-, Br^-, I^-).

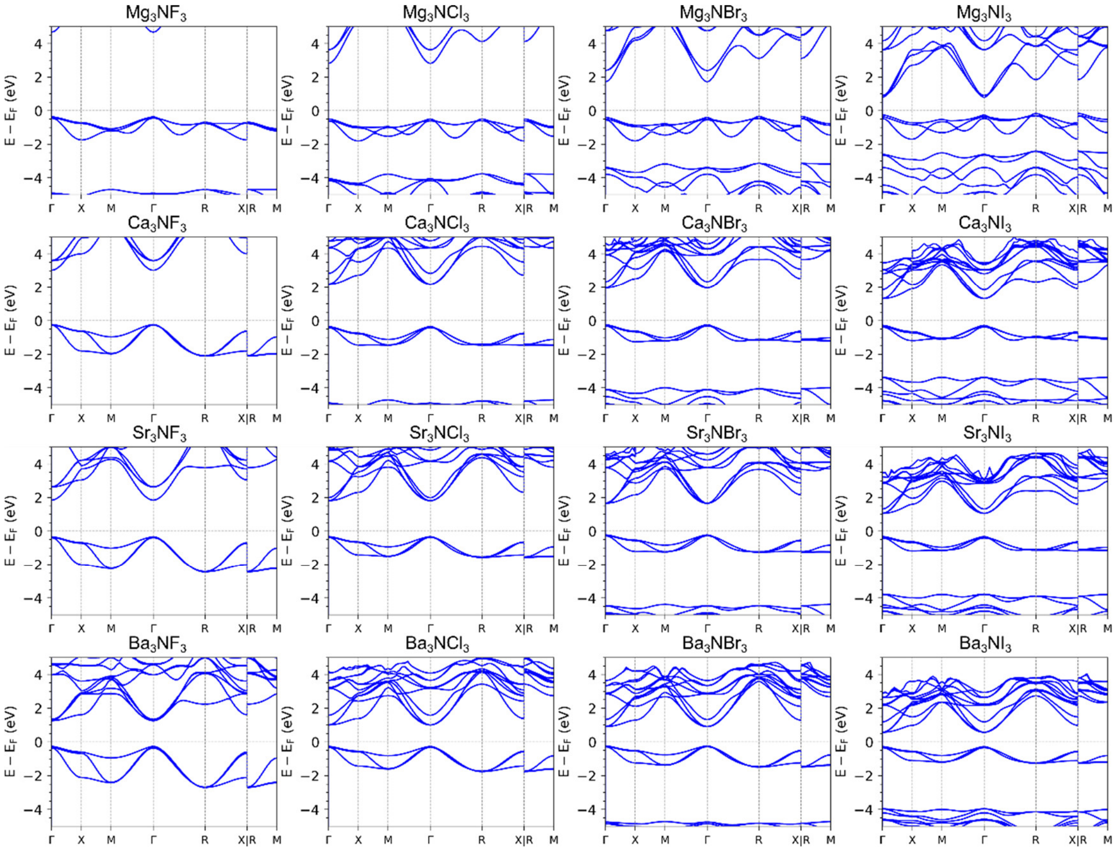


Figure S6. Calculated band structures for the *Pm*-3*m* phase of $X_3NA'_3$ (X = Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+} and A′ = F^-, Cl^-, Br^-, I^-), using the HSE06+SOC method.

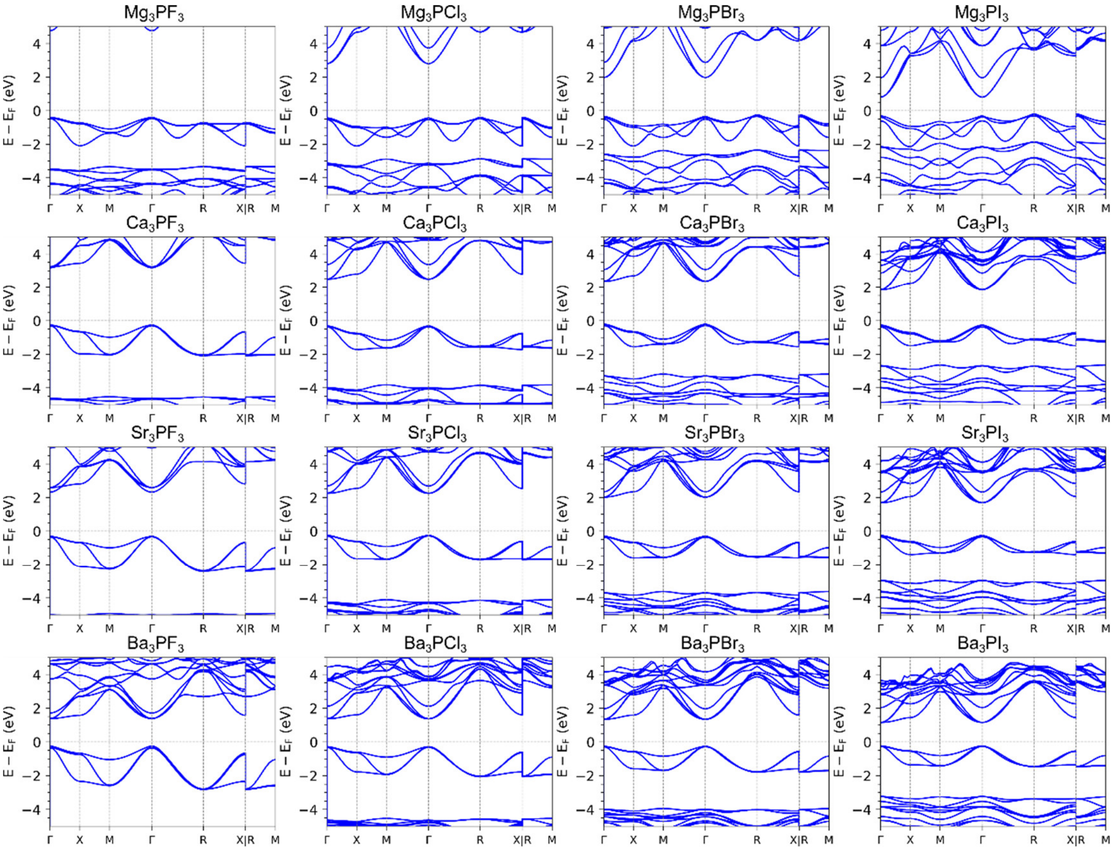


Figure S7. Calculated band structures for the *Pm*-3*m* phase of $X_3PA'_3$ (X = Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+} and A′ = F^-, Cl^-, Br^-, I^-), using the HSE06+SOC method.

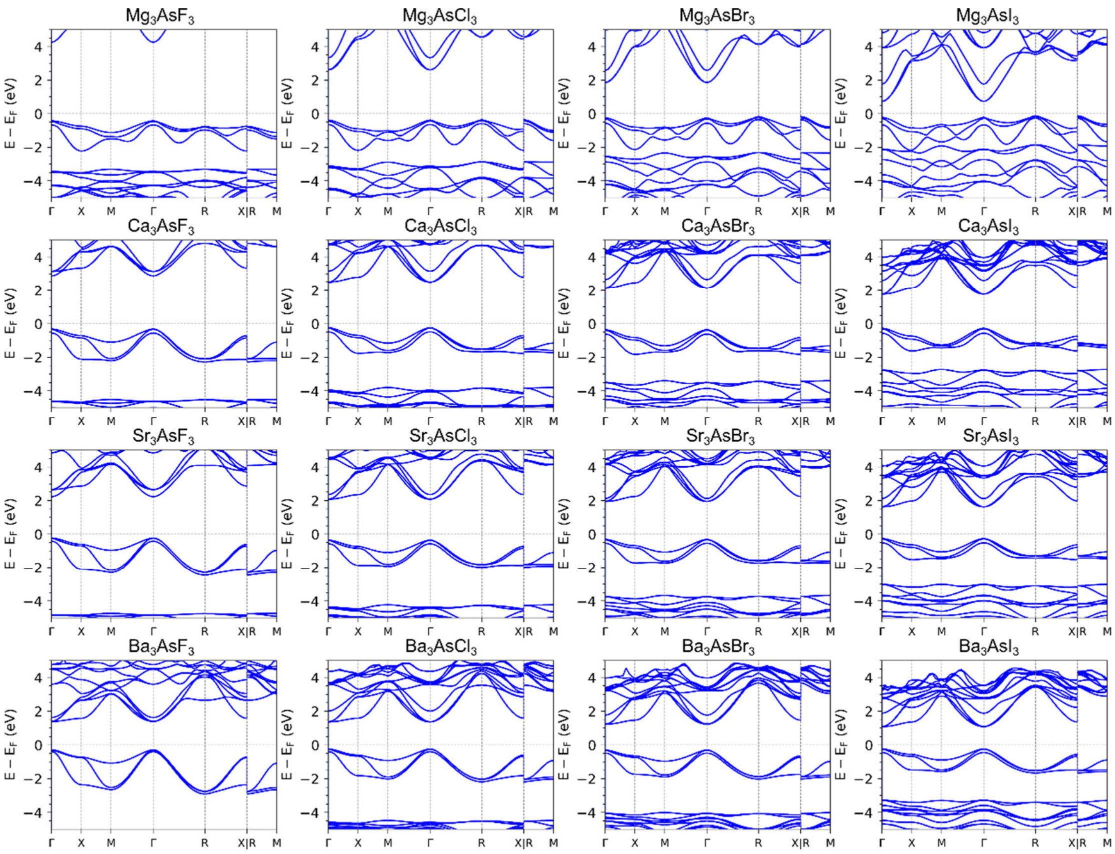


Figure S8. Calculated band structures for the *Pm*-3*m* phase of $X_3AsA'_3$ (X = Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+} and A′ = F^-, Cl^-, Br^-, I^-), using the HSE06+SOC method.

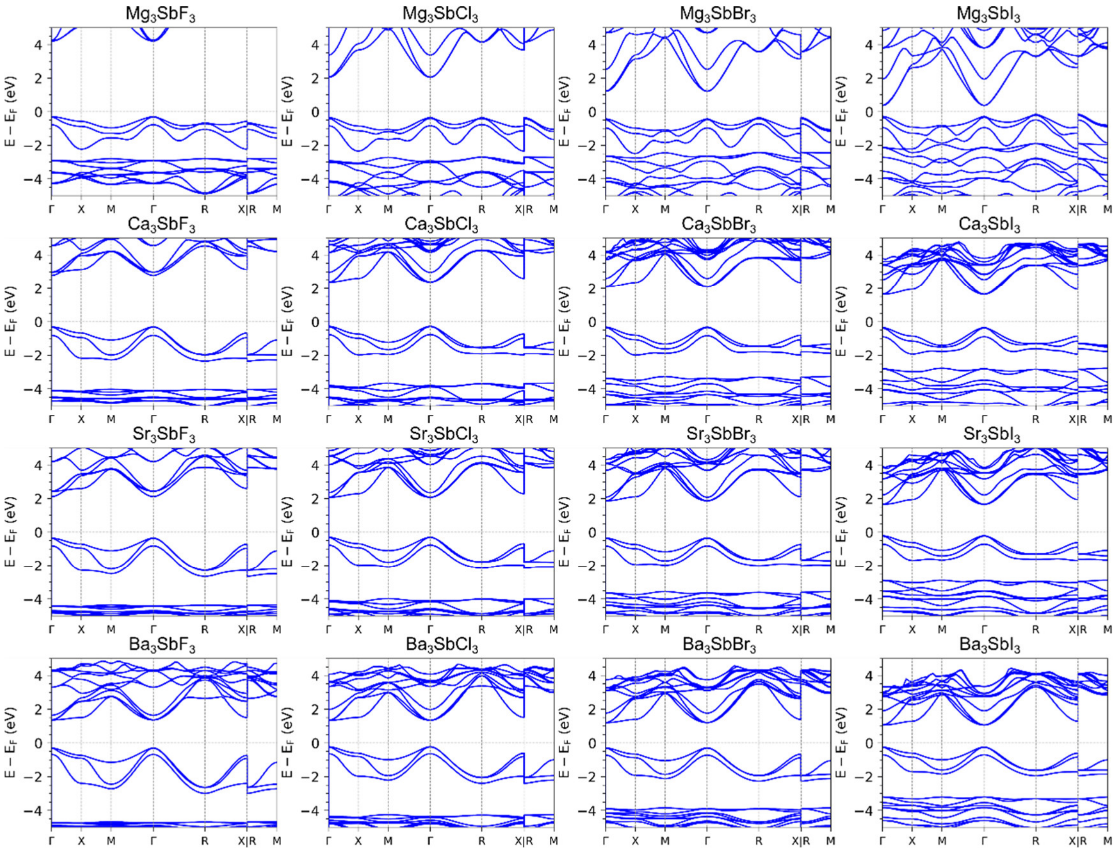


Figure S9. Calculated band structures for the *Pm*-3*m* phase of $X_3SbA'_3$ (X = Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+} and A′ = F^-, Cl^-, Br^-, I^-), using the HSE06+SOC method.

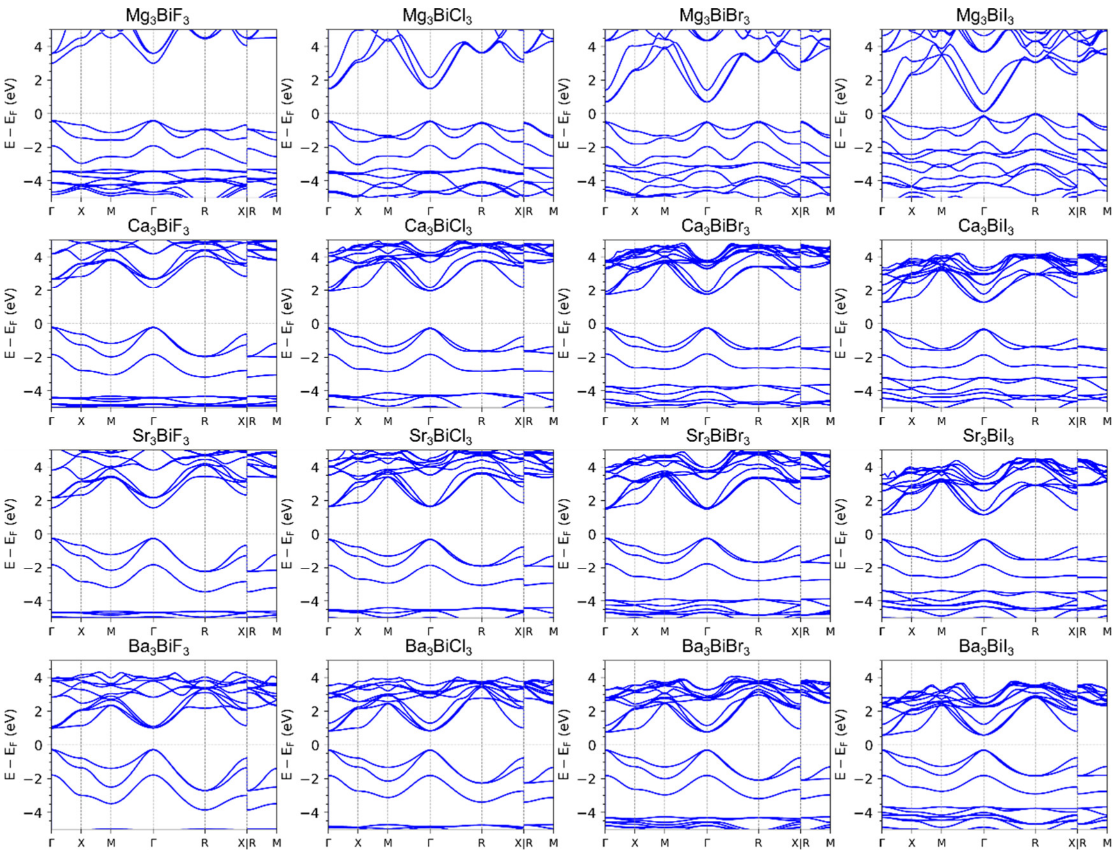


Figure S10. Calculated band structures for the *Pm*-3*m* phase of $X_3BiA'_3$ (X = Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+} and A′ = F^-, Cl^-, Br^-, I^-), using the HSE06+SOC method.

Table S1. The optimized lattice constants, the energy above the convex hull, and HSE06+SOC-calculated band gaps for 38 $X_3BA'_3$ antiperovskite derivatives (E_{hull}<80 meV/atom). The '☑' symbol indicates phonon stability, while the '☒' symbol indicates phonon instability.

Materials	Lattice Constants $a = b = c$ (Å)	E_{hull} (eV/atom)	Phonon Stability	Band Gap (eV)
Mg_3NF_3	4.262	0	☑	5.0409
Ca_3NF_3	4.832	0.034	☑	3.2542
Sr_3NF_3	5.196	0.064	☑	2.1783
Mg_3PCl_3	5.252	0.066	☑	3.1906
Ca_3PCl_3	5.715	0	☑	2.7968
Ca_3PBr_3	5.901	0	☑	2.5659
Sr_3PCl_3	6.056	0	☑	2.4904
Sr_3PBr_3	6.233	0	☑	2.3205
Sr_3PI_3	6.526	0	☒	1.9408
Ba_3PCl_3	6.439	0.037	☑	1.6387

Ba_3PBr_3	6.608	0	☑	1.587
Ba_3PI_3	6.888	0	☒	1.3769
Mg_3AsCl_3	5.330	0.074	☑	2.9758
Ca_3AsCl_3	5.786	0	☑	2.7272
Ca_3AsBr_3	5.964	0	☑	2.4896
Sr_3AsCl_3	6.125	0	☑	2.4261
Sr_3AsBr_3	6.295	0	☑	2.2495
Sr_3AsI_3	6.578	0	☒	1.8766
Ba_3AsCl_3	6.506	0	☑	1.5959
Ba_3AsBr_3	6.667	0	☑	1.5239
Ba_3AsI_3	6.937	0	☒	1.3122
Ca_3SbCl_3	5.989	0.044	☒	2.6324
Ca_3SbBr_3	6.142	0	☑	2.4065
Ca_3SbI_3	6.399	0.026	☑	1.9849
Sr_3SbCl_3	6.324	0.031	☒	2.3638
Sr_3SbBr_3	6.469	0	☑	2.1965
Sr_3SbI_3	6.717	0.018	☑	1.8567
Ba_3SbCl_3	6.696	0.065	☒	1.5519
Ba_3SbBr_3	6.834	0	☑	1.4835
Ba_3SbI_3	7.070	0.020	☑	1.2953
Ca_3BiCl_3	6.047	0.070	☒	2.2159
Ca_3BiBr_3	6.197	0	☑	1.9875
Ca_3BiI_3	6.449	0.001	☑	1.5732
Sr_3BiCl_3	6.382	0.073	☒	1.9222
Sr_3BiBr_3	6.523	0	☑	1.6986
Sr_3BiI_3	6.766	0	☑	1.4417
Ba_3BiBr_3	6.884	0.003	☑	1.053
Ba_3BiI_3	7.114	0	☑	0.8675

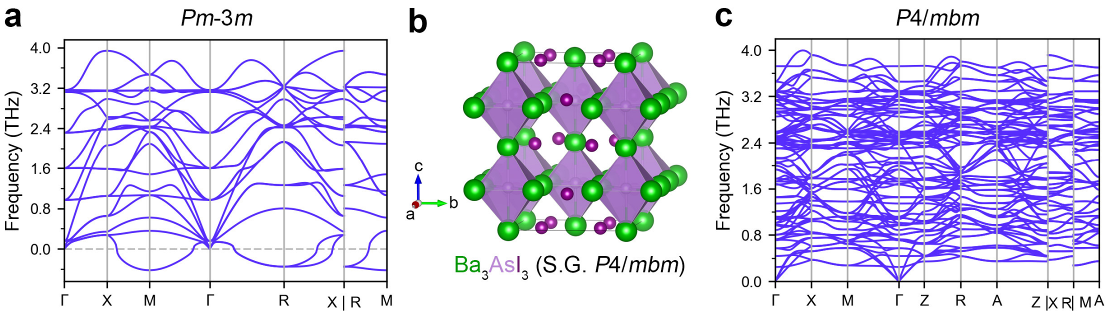


Figure S11. (a) Phonon spectra for the *Pm*-3*m* phase of Ba_3AsI_3, illustrating imaginary frequencies at the M and R points. (b) Crystal structure of the *P*4/*mbm* phase of Ba_3AsI_3. (c) Phonon spectra for Ba_3AsI_3 in the *P*4/*mbm* phase.

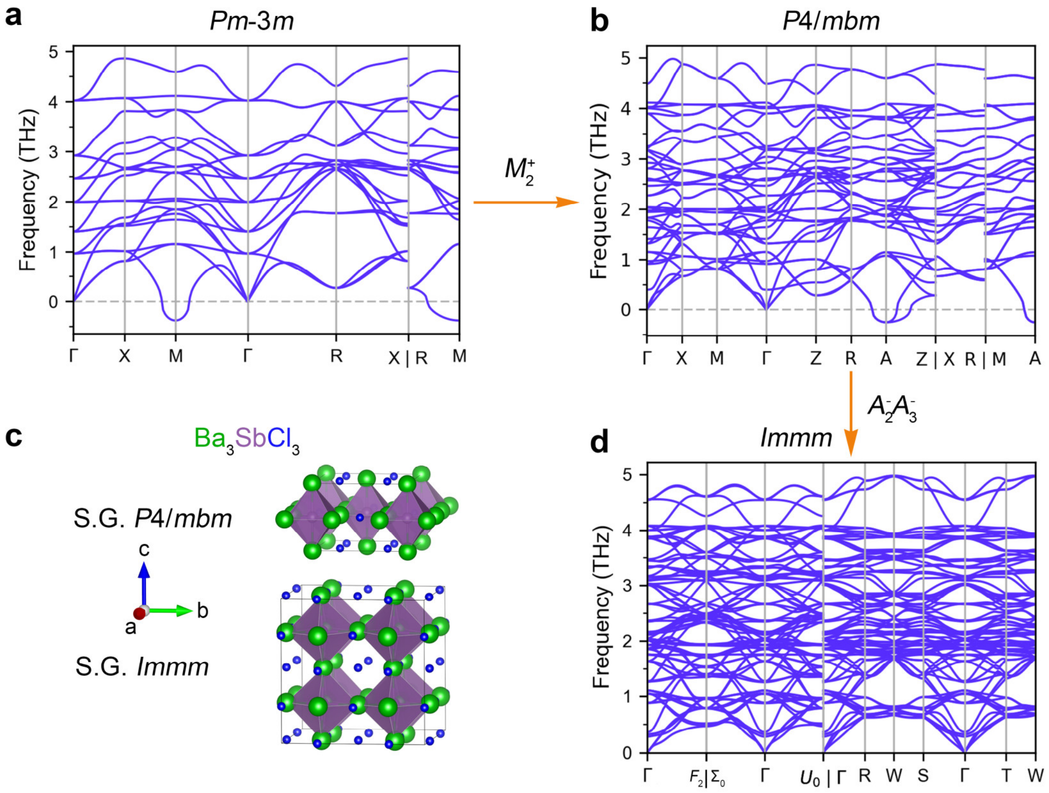


Figure S12. (a) Phonon spectra of the *Pm*-3*m* phase of Ba_3SbCl_3, showing imaginary frequencies at the M point. (b) Phonon spectra for the *P*4/*mbm* phase of Ba_3SbCl_3, displaying imaginary frequencies at the A point. (c) Crystal structure of Ba_3SbCl_3 in the *P*4/*mbm* and *Immm* phases. (d) Phonon spectra for Ba_3SbCl_3 in the *Immm* phase. Note that the arrows indicate that the structures of the *P*4/*mbm* and *Immm* phases are derived from the corresponding imaginary frequency modes.

Table S2. Possible low-energy structures of Ba_3PI_3 derived from the prototype *Pm*-3*m* phase based on unstable phonon modes. The total energies are presented relative to the energy of the *Pm*-3*m* phase.

Space group	Total energy (meV/f.u.)	irrep
*Amm*2 (No. 38)	-1.416	Γ_4^- $(a, a, 0)$
Cm (No. 8)	-0.919	Γ_4^- (a, a, b)
*P*1 (No. 1)	-1.681	Γ_4^- (a, b, c)
*P*4*mm* (No. 99)	-0.917	Γ_4^- $(a, 0, 0)$
Pm (No. 6)	-0.915	Γ_4^- $(a, b, 0)$
*R*3*m* (No. 160)	-1.641	Γ_4^- (a, a, a)
Fmmm (No. 69)	-25.472	R_3^+ (a, b)
*I*4/*mcm*(No. 140)	-25.673	R_3^+ $(0, a)$
*I*4/*mmm* (No. 139)	-25.274	R_3^+ $(a, 0)$
*P*4/*mbm* (No. 127)	-33.681	M_2^+ $(a; 0; 0)$
Pmma (No. 51)	-0.170	X_5^- $(a, 0; 0, 0; 0, 0)$
Cc (No. 9)	-25.572	$\Gamma_4^- \oplus R_3^+$ $(a, a, b—0, c)$
Cm (No. 8)	-25.556	$\Gamma_4^- \oplus R_3^+$ $(a, a, b—c, 0)$
*Fmm*2 (No. 42)	-25.350	$\Gamma_4^- \oplus R_3^+$ $(a, 0, 0—b, c)$
*Ima*2 (No. 46)	-25.675	$\Gamma_4^- \oplus R_3^+$ $(a, a, 0—0, b)$
*Imm*2 (No. 44)	-25.270	$\Gamma_4^- \oplus R_3^+$ $(a, a, b—0, c)$
*Amm*2 (No. 38)	-33.795	$\Gamma_4^- \oplus M_2^+$ $(a, 0, 0—b; 0; 0)$
Cm (No. 8)	-33.682	$\Gamma_4^- \oplus M_2^+$ $(a, b, 0—0; c; 0)$
*P*4*bm* (No. 100)	-33.681	$\Gamma_4^- \oplus M_2^+$ $(a, 0, 0—0; 0; \mathrm{b})$
Pc (No. 7)	-33.796	$\Gamma_4^- \oplus M_2^+$ $(a, a, b—c; 0; 0)$
$Pmc2_1$ (No. 26)	-33.796	$\Gamma_4^- \oplus M_2^+$ $(a, a, 0—b; 0; 0)$
Cmmm (No. 65)	-57.140	$M_2^+ \oplus R_3^+$ $(a; 0; 0—b, c)$
$P4_2/mnm$ (No. 136)	-33.759	$M_2^+ \oplus R_3^+$ $(a; 0; 0—b, 0)$
*P*4/*mbm*(No. 127)	-57.325	$M_2^+ \oplus R_3^+$ $(a; 0; 0—0, b)$

Table S3. Possible low-energy structures of Ba_3AsI_3 derived from the prototype *Pm*-3*m* phase based on unstable phonon modes. The total energies are presented relative to the energy of the *Pm*-3*m* phase.

Space group	Total energy (meV/f.u.)	irrep
Fmmm (No. 69)	-8.555	R_3^+ (a, b)
*I*4/*mcm* (No. 140)	-8.782	R_3^+ $(0, a)$
*I*4/*mmm* (No. 139)	-8.780	R_3^+ $(a, 0)$
*P*4/*mbm* (No. 127)	-13.675	M_2^+ $(a; 0; 0)$
Cmmm (No. 65)	-21.655	$M_2^+ \oplus R_3^+$ $(a; 0; 0—b, c)$
$P4_2/mnm$ (No. 136)	-13.652	$M_2^+ \oplus R_3^+$ $(a; 0; 0—b, 0)$
*P*4/*mbm* (No. 127)	-21.662	$M_2^+ \oplus R_3^+$ $(a; 0; 0—0, b)$

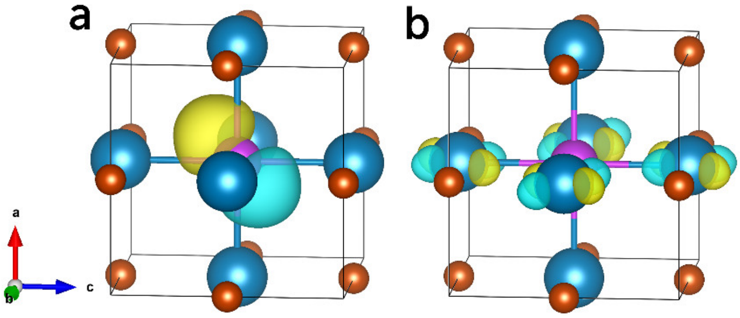


Figure S13. Calculated isosurfaces of wave functions in real space for (a) VBM and (b) CBM of Ba_3BiBr_3.

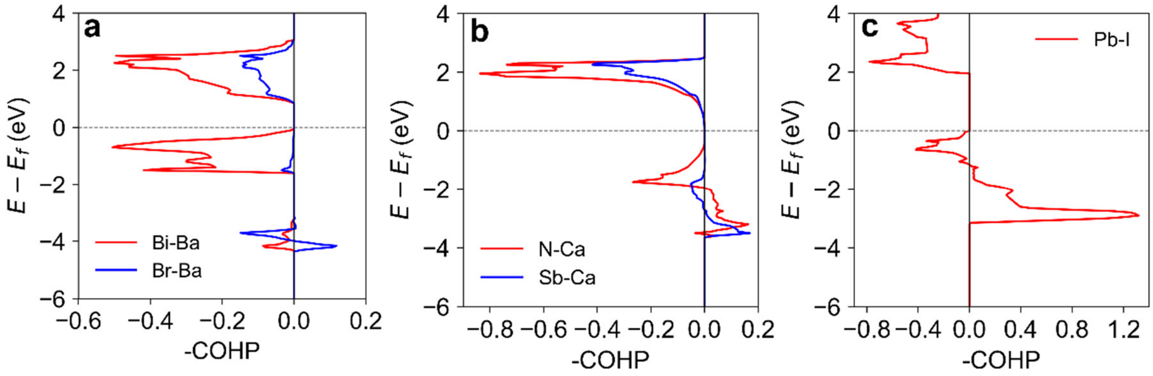


Figure S14. Crystal orbital Hamilton population (COHP) analysis for (a) Ba_3BiBr_3, (b) Ca_3NSb, and (c) $CsPbI_3$.

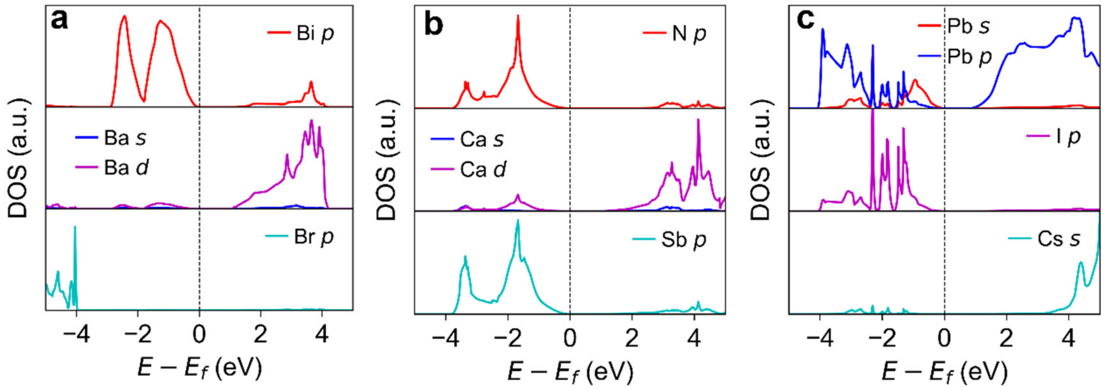


Figure S15. Orbital-projected electronic density of states (DOS) for (a) Ba_3BiBr_3, (b) Ca_3NSb, and (c) $CsPbI_3$, obtained using the HSE06+SOC method. The valence band maximum (VBM) is set to zero eV.

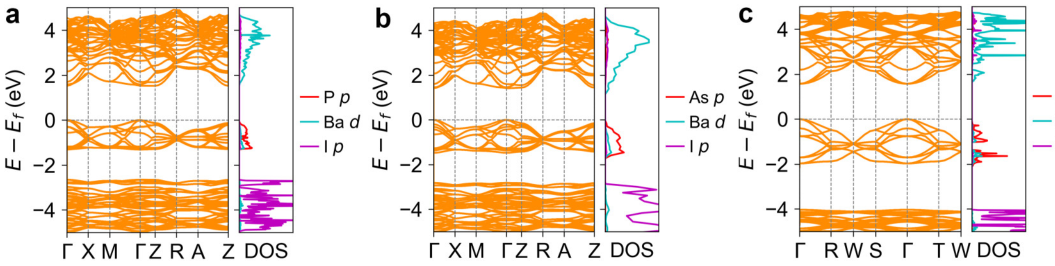


Figure S16. Electronic band structure and orbital-projected electronic density of states (DOS) for (a) tetragonal $P4/mbm$ Ba_3PI_3, (b) tetragonal $P4/mbm$ Ba_3AsI_3, and (c) orthorhombic $Immm$ Ba_3SbCl_3, calculated using the HSE06+SOC method. The valence band maximum (VBM) is set to zero eV.

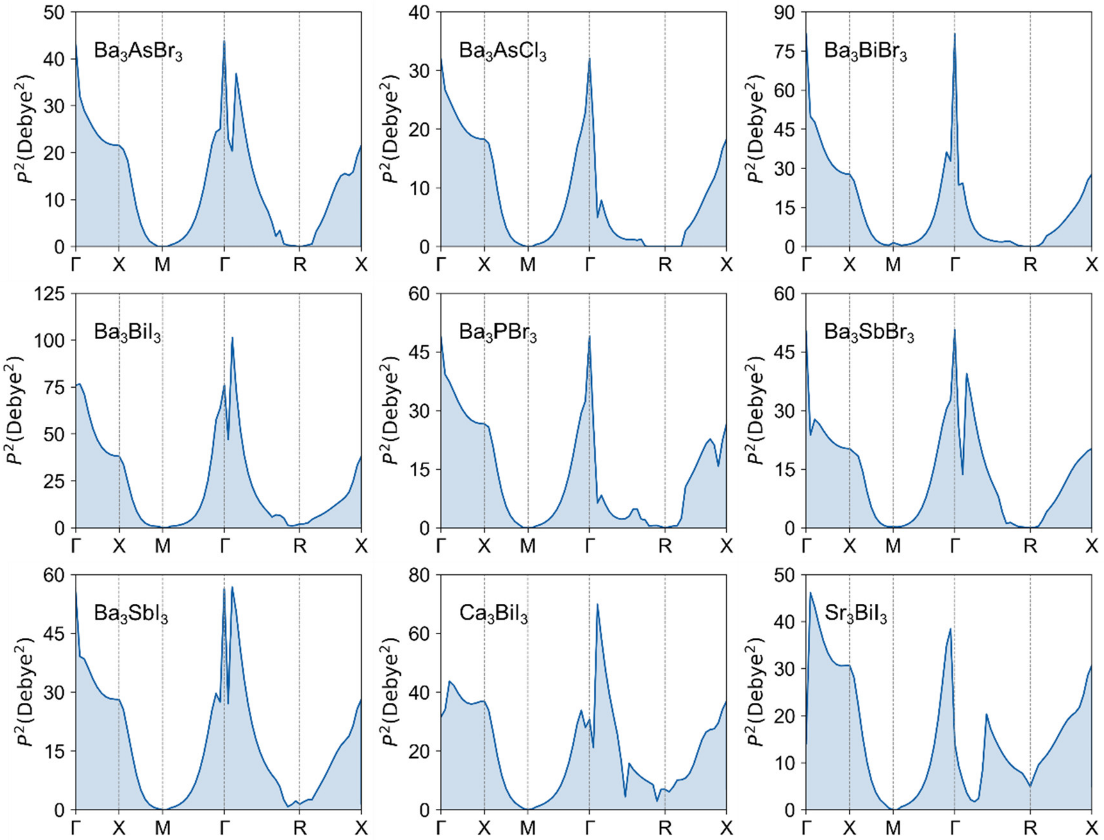


Figure S17. The squares of the transition dipole matrix elements (P^2) between the VBM and the CBM along different high-symmetry directions for the nine antiperovskite derivatives, calculated using the HSE06+SOC method.

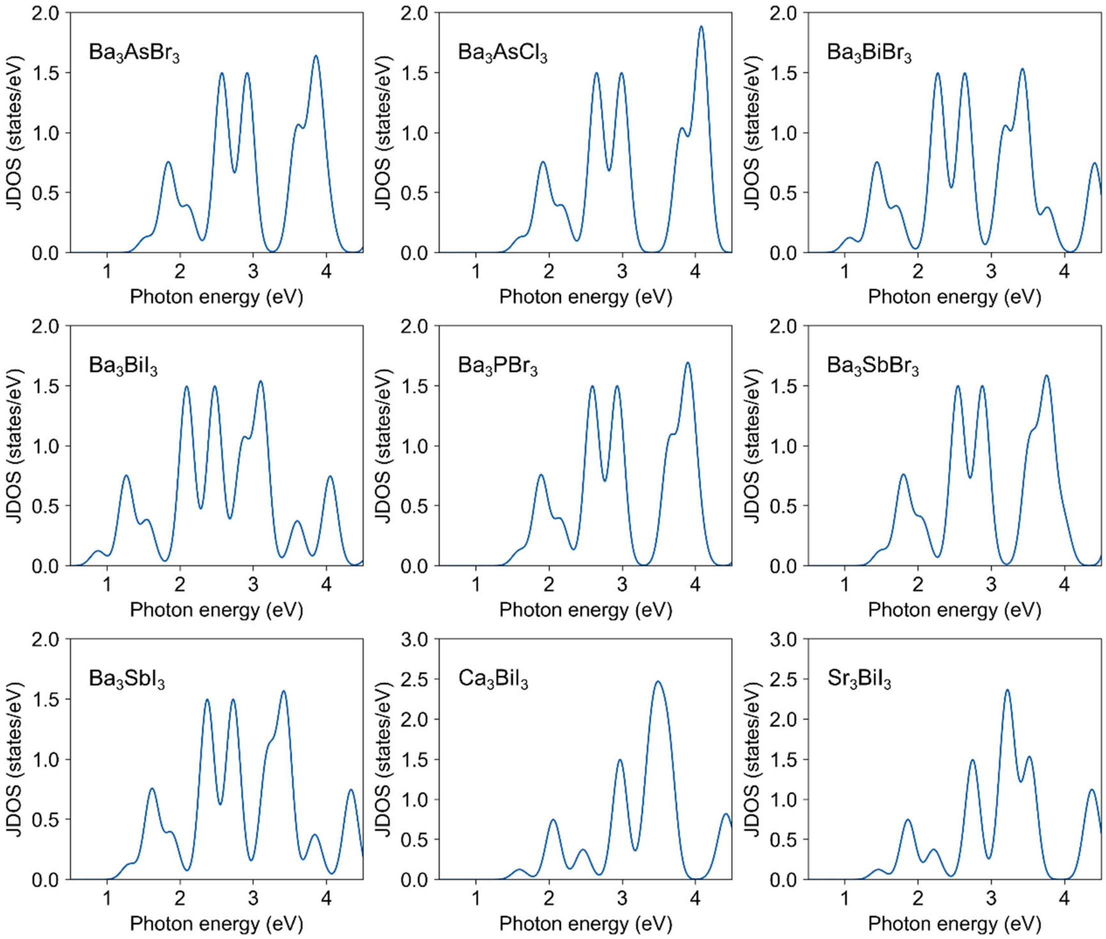


Figure S18. The calculated joint density of states (JDOS) for the nine antiperovskite derivatives, obtained through the HSE06+SOC method.

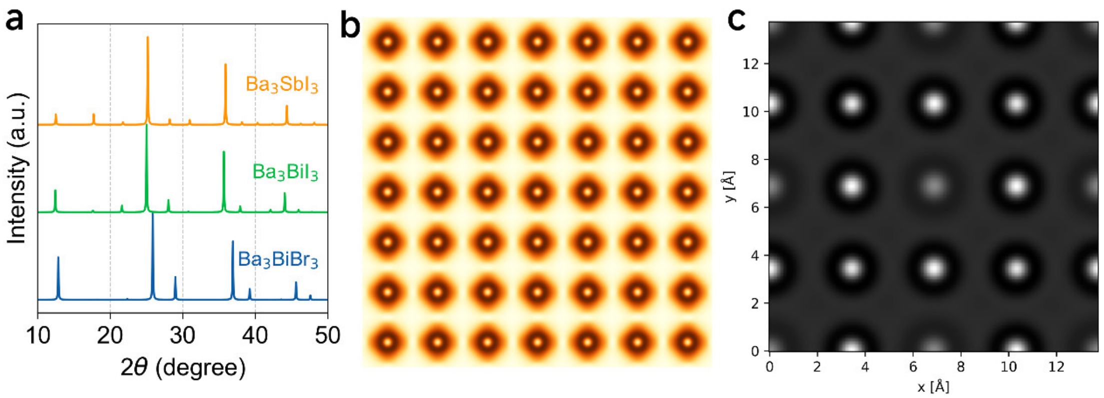


Figure S19. (a) Simulated XRD patterns of Ba_3BiBr_3, Ba_3BiI_3, and Ba_3SbI_3; (b) simulated STM image of Ba_3BiBr_3; and (c) simulated HRTEM image of Ba_3BiBr_3.

Table S4. Top 13 candidate materials predicted by XGBoost models with E_{hull} < 80 meV/atom and 0.8 eV < E_{g} < 1.6 eV.

Candidate Material	Predicted E_{hull} (meV/atom)	Predicted E_{g} (eV)
$Ba_3(P_{0.5}As_{0.5})Cl_3$	-19.44	1.58
$Ba_3(P_{0.5}As_{0.5})Br_3$	-10.26	1.41
$Ba_3(P_{0.5}As_{0.5})I_3$	-10.26	1.25
$Ba_3(As_{0.5}Bi_{0.5})Br_3$	2.84	1.08
$Ba_3(As_{0.5}Bi_{0.5})I_3$	2.84	0.84
$Sr_3(As_{0.5}Bi_{0.5})I_3$	5.29	1.43
$Ca_3(As_{0.5}Bi_{0.5})I_3$	9.87	1.53
$Ba_3(As_{0.5}Sb_{0.5})Br_3$	15.70	1.46
$Ba_3(As_{0.5}Sb_{0.5})I_3$	15.70	1.32
$Ba_3(P_{0.5}Bi_{0.5})Cl_3$	32.86	1.18
$Ba_3(P_{0.5}Sb_{0.5})Cl_3$	45.73	1.57
$Ba_3(P_{0.5}Bi_{0.5})Br_3$	47.09	0.94
$Ba_3(P_{0.5}Sb_{0.5})Br_3$	59.96	1.32
$Ba_3(P_{0.5}Sb_{0.5})I_3$	59.96	1.16

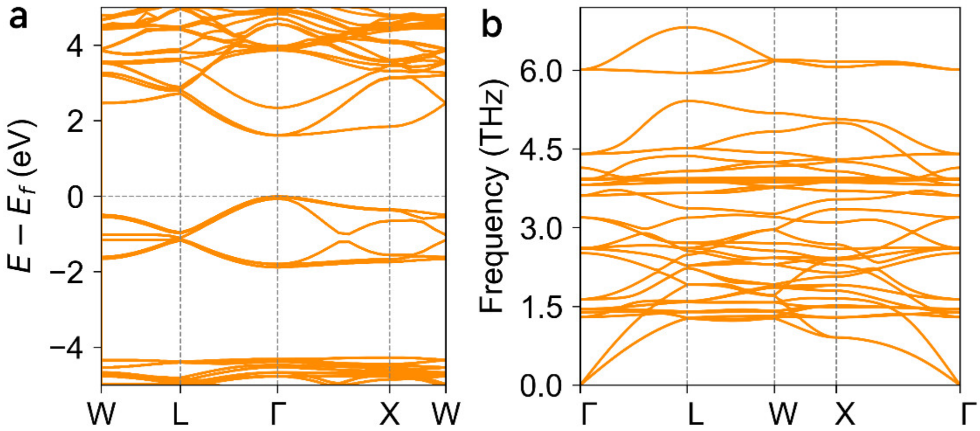


Figure S20. (a) Calculated band structures obtained through the HSE06+SOC method for $Ba_3(P_{0.5}As_{0.5})Cl_3$, and (b) phonon dispersion for $Ba_3(P_{0.5}As_{0.5})Cl_3$.

Table S5. Materials parameters used to compute transport properties. C is the elastic tensor in Voigt notation (unit: GPa); ε_s and ε_∞ are the static and high-frequency dielectric constants in ε_0; D_{vb} and D_{cb} are the absolute deformation potentials at the valence and conduction band edge, respectively; ω_{po} is the effective polar phonon frequency (unit: THz).

Material	C_{11}	C_{44}	C_{12}	$\varepsilon_{s,11}$	$\varepsilon_{s,22}$	$\varepsilon_{s,33}$	$\varepsilon_{\infty,11}$	$\varepsilon_{\infty,22}$	$\varepsilon_{\infty,33}$	$D_{vb,11}$	$D_{vb,22}$	$D_{vb,33}$	$D_{cb,11}$	$D_{cb,22}$	$D_{cb,33}$	ω_{po}
Ba_3BiBr_3	51.66	7.95	5.84	14.06	14.06	14.06	5.54	5.54	5.54	1.64	1.64	1.64	2.72	2.72	2.72	2.34
Ba_3BiI_3	48.15	6.69	5.12	13.18	13.18	13.18	6.34	6.34	6.34	1.18	1.18	1.18	3.10	3.10	3.10	2.14
Ba_3SbI_3	50.25	6.96	5.41	13.13	13.13	13.13	5.90	5.90	5.90	1.06	1.06	1.06	3.26	3.26	3.26	2.33

Table S6. Primary features for machine learning.

	Primary features	Abbreviation
1	Atomic number	Z
2	Atomic weight	A_r
3	Period number	P
4	Group number	G
5	Atomic radius	r^a
6	Covalent radius	r^c
7	Ionic radius	r^i
8	Crystal radius	r^m
9	Pauling electronegativity	χ^P
10	Allred-Rockow electronegativity	χ^{AR}
11	Mulliken electronegativity	χ^M
12	Martinov-Batsanov electronegativity	χ^{MB}
13	Gordy electronegativity	χ^G
14	Number of valence electrons	N_e
15	Density of element	ρ

Table S7. The optimized hyperparameters of XGBoost algorithm.

Model	Hyperparameter	Value
E_g	n_estimators	461
	learning_rate	0.435
	max_depth	2
	min_child_weight	1
E_{hull}	n_estimators	456
	learning_rate	0.500
	max_depth	4
	min_child_weight	10

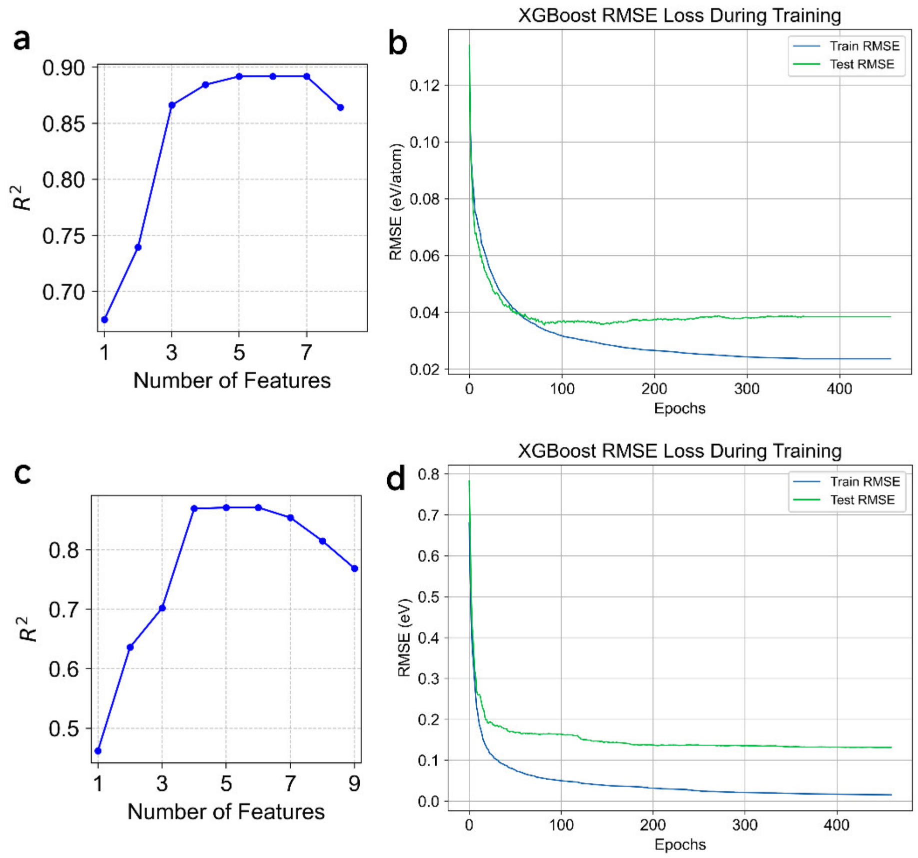


Figure S21. R^2 scores from 5-fold cross-validation of the sequential forward selection (SFS) results for the XGBoost model of (a) E_{hull} and (c) E_{g}. Loss function of the XGBoost model for (b) E_{hull} and (d) E_{g} across each epoch on the training and test datasets.

Reference:

1 W. Rahim, J. M. Skelton, C. N. Savory, I. R. Evans, J. S. O. Evans, A. Walsh and D. O. Scanlon, Polymorph exploration of bismuth stannate using first-principles phonon mode mapping, *Chem. Sci.*, 2020, **11**(30), 7904–7909.

2 A. Togo and I. Tanaka, Evolution of crystal structures in metallic elements, *Phys. Rev. B*, 2013, **87**(18), 184104.

3 L. Ward, A. Agrawal, A. Choudhary and C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, *npj Comput. Mater.*, 2016, **2**(1), 16028.

4 Y.-H. Wang, Y.-F. Zhang, Y. Zhang, Z.-F. Gu, Z.-Y. Zhang, H. Lin and K.-J. Deng, Identification of adaptor proteins using the ANOVA feature selection technique, *Methods*, 2022, **208**, 42–47.

5 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.

6 Y. Liu, Y. Mu, K. Chen, Y. Li and J. Guo, Daily activity feature selection in smart homes based on pearson correlation coefficient, *Neural Process Lett.*, 2020, **51**, 1771–1787.

7 S. Raschka, MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack, *J. open source softw.*, 2018, **3**(24), 638.

8 T. Chen and C. Guestrin, Xgboost: A scalable tree boosting system, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, 785–794. ACM, New York, NY USA.

9 J. Gong, S. Chu, R. K. Mehta and A. J. H. McGaughey, XGBoost model for electrocaloric temperature change prediction in ceramics, *npj Comput. Mater.*, 2022, **8**(1), 140.

10 E. Ren and F.-X. Coudert, Enhancing gas separation selectivity prediction through geometrical and chemical descriptors, *Chem. Mater.*, 2023, **35**(17), 6771–6781.

11 R. Jacobs, J. Liu, H. Abernathy and D. Morgan, Machine learning design of perovskite catalytic properties, *Adv. Energy Mater.*, 2024, **14**(12), 2303684.

12 A. D. Smith, S. B. Harris, R. P. Camata, D. Yan and C.-C. Chen, Machine learning the relationship between Debye temperature and superconducting transition temperature, *Phys. Rev. B*, 2023, **108**(17), 174514.

13 N. K. Barua, A. Golabek, A. O. Oliynyk and H. Kleinke, Experimentally validated machine learning predictions of ultralow thermal conductivity for SnSe materials, *J. Mater. Chem. C*, 2023, **11**(34), 11643–11652.

14 N. Fernando, Bayesian Optimization: Open source constrained global optimization tool for Python. https://github.com/bayesian-optimization/BayesianOptimization, 2014.

15 R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler and L. M. Ghiringhelli, SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates, *Phys. Rev. Mater.*, 2018, **2**(8), 083802.

16 H. W. Kuhn, A. W. Tucker, Contributions to the theory of games. Princeton University Press, 1953.