SUPPLEMENT

Supplementary Material for "Exploration of Double Perovskite Material Space via Machine Learning for Tandem Solar Cells"

Z.Q. Wang[a], Z.H. Xiong[a], W.J. Hu[b], J.J. Jiang[a], Z.B. Cheng[a], Y.M. Xue[a], L. Peng[a,*] and J. Lin[a]

[a]*Department of Physics, Shanghai University of Electric Power, Shanghai 200090, China*
[b]*College of Electrical Engineering, Shanghai University of Electric Power, Shanghai 200090, China*

Received: 18.03.2025 & Accepted: 09.06.2025

Doi: 10.12693/APhysPolA.147.488.S1 *e-mail: plpeng@shiep.edu.cn

1. Machine learning

We have established an eXtreme Gradient Boosting (XGBoost) model to predict the lattice constants of double perovskite structures and generate structural files for all perovskite materials. The resulting model-building process is as follows.

The cubic phase prototype structure file of $Cs_2InAgCl_6$ (see Fig. S1a) was obtained from the Crystallography Open Database [1–9]. The initial cubic phase structure of the material was obtained through element substitution. The corresponding tetragonal phase structure shown in Fig. S1b was derived from the Sr_2TaGaO_6 double perovskite. All the initial material structures of the tetragonal phase were generated by element substitution.

By comparing the structure files before and after structural relaxation, we observed that the most significant change was in the lattice constants. As shown in Fig. S1c (pre-relaxation cubic phase) and Fig. S1d (post-relaxation cubic phase), the relative atomic coordinates underwent only subtle changes. These changes were negligible compared to the lattice constant variations. Therefore, our prediction target is the crystal lattice structure.

We compiled the structure files after structural optimization. In the structure files of the cubic phase, the values of the first, second, and third basis vectors along the a, b, and c axes, respectively, are identical. Consequently, the prediction target for the cubic phase in the base learner only requires the precise value of one basis vector component. However, for the tetragonal phase, the

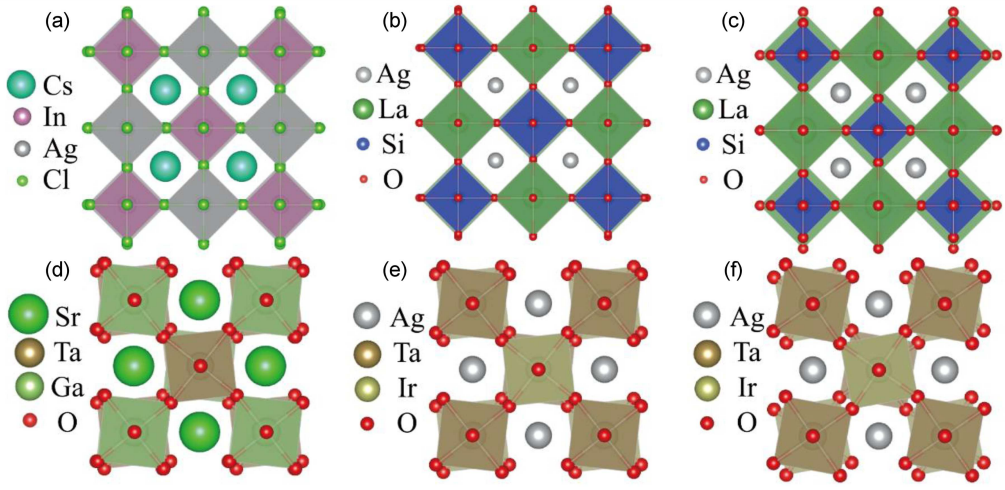


Fig. S1. (a) Prototypical cubic phase structure of $Cs_2InAgCl_6$. (b) Tetragonal phase structure of Sr_2TaGaO_6. (c) Cubic phase structure of Ag_2LaSiO_6 generated by element substitution. (d) Prototypical tetragonal phase structure of Ag_2TaIrO_6. (e) Cubic phase structure of Ag_2LaSiO_6 after structural optimization. (f) Tetragonal phase structure of Ag_2TaIrO_6 following structural optimization.

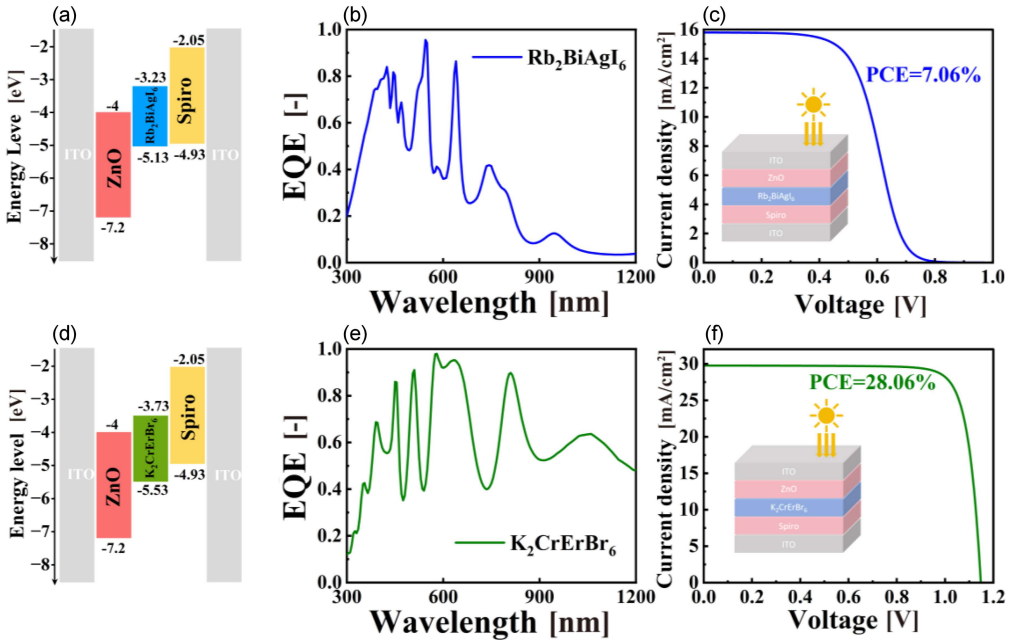


Fig. S2. Device simulation of single-junction Rb_2BiAgI_6 solar cell and the single-junction $K_2CrErBr_6$ solar cell. (a) Energy level diagram of different transmission layers of Rb_2BiAgI_6. (b) EQE and (c) PCE of Rb_2BiAgI_6 solar cell. (d) Energy level diagram of different transmission layers of $K_2CrErBr_6$. (e) EQE and (f) PCE of $K_2CrErBr_6$ solar cell.

TABLE SI

Hyperparameter adjustment of ALIGNN model.

Hyperparameter	Setting range	Selection value	Description
ALIGNN layers	uniform integer (0; 4)	3	A graph convolutional network (GNC) layer that composes an edge-gated graph convolution on the bond graph with an edge-gated graph convolution on the line graph.
GCN layers	uniform integer (0; 4)	3	A graph convolutional network layer that enhances the model's understanding of graph structure by introducing edge features.
edge input features	20, 40, 80, 100	40	The amount of feature information associated with edges in the GCN.
triplet input features	20, 40, 80, 160	40	The amount of feature information that represents the nodes in the graph and the relationships between them.
embedding features	16, 32, 64, 80	32	The number of representations that maps nodes, edges, or the entire graph in a graph to a low-dimensional vector space.
hidden features	64, 128, 256, 512	128	The width of the graph convolution layers.
batch size	16, 32, 64, 128, 256	64	The number of plots that are processed at the same time.
learning rate	$\log(10^{-4}; 10^{-1})$	0.001	The step size at which the parameters are updated during the training process.
epochs	100, 200, 500, 1000	100	The number of times the entire training set has been traversed once.

values of the second basis vector along the b axis and the third basis vector along the c axis are identical, differing from the value of the first basis vector along the a axis. Consequently, the prediction target for the tetragonal phase in the base learner requires the precise values of two basis vector components. To achieve accurate prediction of lattice constants across multiple space

TABLE SII

The key simulation parameters of tandem perovskite solar cells [11–13]. DOS is the density of states.

Parameter	Symbol [unit]	ZnO	Rb_2BiAgI_6	$K_2CrErBr_6$	Spiro-OMeTAD	CdS	CZTSSe
thickness	– [nm]	50	300	300	50	50	1000
electron affinity	χ [eV]	4	3.2285	3.7297	2.5	4.2	4.2
bandgap	E_g [eV]	3.3	1.9	1.8	2.7	2.4	1.13
relative permittivity	ε_r	9	8.67	4.11	3	9	13.6
eff. DOS for electron	N_C [cm^{-3}]	2.2×10^{18}	7.7×10^{21}	2.7×10^{21}	1.0×10^{19}	1.8×10^{19}	1.8×10^{19}
eff. DOS for hole	N_V [cm^{-3}]	5.7×10^{18}	1.9×10^{21}	3.9×10^{21}	1.0×10^{19}	2.2×10^{18}	2.2×10^{18}
mobility of electron	μ_e [cm^2 V/s]	100	81	11	0.0002	100	100
mobility of hole	μ_h [cm^2 V/s]	25	27	23	0.0002	25	25
SRH life time	$\tau_e = \tau_h$ [ns]	0.003	100	100	0.1	0.005	5.4

groups, we constructed XGBoost descriptors based on elemental properties. These include the atomic numbers, relative atomic masses, atomic volumes, mass density (mass per unit volume), atomic radii, covalent radii, and effective ionic radii at the A, B, and X sites. All these data come from the Database of Ionic Radii [10].

Table S1 summarizes the hyperparameter space and final selected values for the ALIGNN model. Seven key parameters were optimized on a predefined grid: the numbers of ALIGNN and GCN layers (3 each), edge and triplet input features (40 each), embedding dimension (32), hidden-layer width (128), batch size (64), learning rate (1×10^{-3}), and training epochs (100). These settings yield the best compromise between predictive accuracy and computational cost.

Table S2 summarizes the electrical parameters used in the COMSOL Multiphysics simulations for the tandem perovskite solar-cell stack. For each layer — ZnO, Rb_2BiAgI_6, $K_2CrErBr_6$, Spiro-OMeTAD, CdS, and CZTSSe — the table lists thickness, electron affinity (χ), band-gap energy (E_g), relative permittivity (ε_r), effective conduction- and valence-band density of states (N_C and N_V), electron and hole mobilities (μ_e and μ_h), and the common Shockley–Read–Hall (SRH) lifetime ($\tau_e = \tau_h$). These values from recent simulation studies [11–13] are directly imported into COMSOL's semiconductor module to define the material properties for all electrical simulations.

References

[1] A. Vaitkus, A. Merkys, T. Sander, M. Quirós, P.A. Thiessen, E.E. Bolton, S. Gražulis, *J. Cheminform.* **15**, 123 (2023).

[2] A. Merkys, A. Vaitkus, J. Butkus, M. Okulič-Kazarinas, V. Kairys, S. Gražulis, *J. Appl. Crystallogr.* **49**, 292 (2016).

[3] S. Gražulis, A. Merkys, A. Vaitkus, M. Okulič-Kazarinas, *J. Appl. Crystallogr.* **48**, 85 (2015).

[4] S. Gražulis, D. Chateigner, R.T. Downs, A.F.T. Yokochi, M. Quirós, L. Lutterotti, E. Manakova, J. Butkus, P. Moeck, A. Le Bail, *J. Appl. Crystallogr.* **42**, 726 (2009).

[5] S. Gražulis, A. Daškevič, A. Merkys, D. Chateigner, L. Lutterotti, M. Quirós, N.R. Serebryanaya, P. Moeck, R.T. Downs, A. Le Bail, *Nucl. Acids Res.* **40**, D420 (2012).

[6] A. Merkys, A. Vaitkus, A. Grybauskas, A. Konovalovas, M. Quirós, S. Gražulis, *J. Cheminform.* **15**, 25 (2023).

[7] A. Le Bail, *J. Appl. Crystallogr.* **38**, 389 (2005).

[8] M. Quirós, S. Gražulis, S. Girdzijauskaitė, A. Merkys, A. Vaitkus, *J. Cheminform.* **10**, 23 (2018).

[9] A. Vaitkus, A. Merkys, S. Gražulis, *J. Appl. Crystallogr.* **54**, 661 (2021).

[10] R.D. Shannon, *Acta Crystallogr. A* **32**, 751 (1976).

[11] A.W.Y. Ho-Baillie, J. Zheng, M.A. Mahmud, F.-J. Ma, D.R. McKenzie, M.A. Green, *Appl. Phys. Rev.* **8**, 041307 (2021).

[12] J.M. Siqueiros, R. Machorro, L.E. Regalado, *Appl. Opt.* **27**, 2549 (1988).

[13] L. Fanni, B. Delaup, B. Niesen, Y. Milstein, D. Shachal, M. Morales-Masis, S. Nicolay, C. Ballif, *Mater. Res. Express* **2**, 075006 (2015).