

ARTICLE

Open Access

Data-driven design of high-performance $\text{MASn}_x\text{Pb}_{1-x}\text{I}_3$ perovskite materials by machine learning and experimental realization

Xia Cai^{1,2,3}, Fengcai Liu^{1,3}, Anran Yu^{1,3}, Jiajun Qin⁴, Mohammad Hatamvand^{1,3}, Irfan Ahmed^{1,3}, Jiayan Luo^{1,3}, Yiming Zhang^{1,5}, Hao Zhang^{1,5,6} and Yiqiang Zhan^{1,3}

Abstract

The photovoltaic performance of perovskite solar cell is determined by multiple interrelated factors, such as perovskite compositions, electronic properties of each transport layer and fabrication parameters, which makes it rather challenging for optimization of device performances and discovery of underlying mechanisms. Here, we propose and realize a novel machine learning approach based on forward-reverse framework to establish the relationship between key parameters and photovoltaic performance in high-profile $\text{MASn}_x\text{Pb}_{1-x}\text{I}_3$ perovskite materials. The proposed method establishes the asymmetrically bowing relationship between band gap and Sn composition, which is precisely verified by our experiments. Based on the analysis of structural evolution and SHAP library, the rapid-change region and low-bandgap plateau region for small and large Sn composition are explained, respectively. By establishing the models for photovoltaic parameters of working photovoltaic devices, the deviation of short-circuit current and open-circuit voltage with band gap in defective-zone and low-bandgap-plateau regions from Shockley-Queisser theory is captured by our models, and the former is due to the deep-level traps formed by crystallographic distortion and the latter is due to the enhanced susceptibility by increased Sn^{4+} content. The more difficulty for hole extraction than electron is also concluded in the models and the prediction curve of power conversion efficiency is in a good agreement with Shockley-Queisser limit. With the help of search and optimization algorithms, an optimized Sn:Pb composition ratio near 0.6 is finally obtained for high-performance perovskite solar cells, then verified by our experiments. Our constructive method could also be applicable to other material optimization and efficient device development.

Introduction

Since the perovskite solar cells (PSCs) are proposed by Kojima et al. in 2009¹, they have been studied extensively with a rapid rise in power conversion efficiency (PCE)^{2–5} which exceeds 25.5% in single-junction PSC⁶. Organic metal halide perovskites (OMHPs) bestowed with outstanding optoelectronic properties are beneficial for high-

performance PSCs, including tunable band gap, low exciton binding energy, high light absorption coefficient and long carrier diffusion length^{7–10}. One of the most typical OMHPs is methylammonium lead tri-halide (MAPbI_3), which has attracted intensive interest and is one of the most promising materials for developing low-cost and solution-processed optoelectronic technology^{4,11}. Despite the rising efficiency records for PSC devices, the efficiency for single-junction MAPbI_3 -based device is limited by the band gap (≈ 1.6 eV), which is higher than the optimal range of Shockley-Queisser (S-Q) limit (≈ 1.35 eV)¹². To exceed the S-Q limit, recently all-perovskite tandem solar cells (PTSCs) composed of wide-bandgap subcells and low-bandgap

Correspondence: Hao Zhang (zhangh@fudan.edu.cn) or Yiqiang Zhan (yqzhan@fudan.edu.cn)

¹School of Information Science and Technology, Fudan University, Shanghai 200433, China

²College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, China

Full list of author information is available at the end of the article

© The Author(s) 2022



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

subcells have been proposed and proven to possibly further increase the efficiency, considering improved utilization of solar energy^{13,14}. Tin (Sn), as an environmentally-safer element in the same group of periodic table with Pb, can replace Pb in MAPbI₃ crystal partially or completely to form Sn-Pb alloying or Sn-based crystals, i.e. MASn_xPb_{1-x}I₃¹³, which can tune the band gap of Sn-Pb alloys between 1.1 eV and 1.6 eV by varying stoichiometric ratios of Sn to Pb^{15–18}. Therefore, MASn_xPb_{1-x}I₃ has been considered as the most promising candidate for high-efficiency single-junction PSCs or low-bandgap subcells in PTSC, responsible for absorbing low-energy photons^{13,19–21}. Much effort has been devoted and the efficiency for mixed Sn-Pb PSCs has increased to 18.6% in PTSCs with MA as A site²², and 23.3% in Sn-Pb PSC with MA-FA-Cs as A site²³, which is approaching the recorded efficiency achieved in single-junction Pb-based PSCs. However, for the goal of exceeding the efficiency of Pb-based PSCs and reaching S-Q limit in the future, it still needs a lot of optimization. The unsatisfactory performance for mixed Sn-Pb PSCs is mainly due to the poor morphology of fabricated Sn-based perovskites resulted from Sn vacancies formed by easy oxidation of Sn²⁺ to Sn⁴⁺ in ambient environment^{20,24}, and partially hindered by the insufficient understanding of the photovoltaic properties of mixed Sn-Pb perovskites, such as the underlying mechanisms for manipulations of band gap and photovoltaic-related parameters.

Current investigations of high-efficiency PSC devices require delicate control of chemical synthesis, laborious experimental steps, substantial resource input and a long research cycle to optimize the perovskite compositions, material of each transport layer, interfacial changes and other related parameters, with the purpose to establish the relationship between desirable PSC properties and fabrication parameters. However, due to the huge chemical space for mixed Sn-Pb alloys, the trial-error experiments are tedious as well as time and energy consuming, and sometimes it is beyond reach of providing a thorough investigation. Machine learning (ML) as a new tool learning from known data to solve intractable and complicated problems, can establish complex nonlinear relationship between input parameters and output property to make rapid prediction without prior knowledge²⁵. Moreover, with increasing amount of experimental data, the established model could be continuously optimized and its prediction ability could be further improved. At present, there are reports on the prediction of perovskite band gap using ML method^{26–30}, and some published studies have used ML to design organic solar cells (OPVs) and predict the performance of OPVs^{31–33}. Sahu et al. used 13 important microscopic properties of organic materials to build a ML model for predicting PCE of OPVs and constructed a dataset for 280 small molecule OPV systems³⁴. David et al. utilized a database consisting

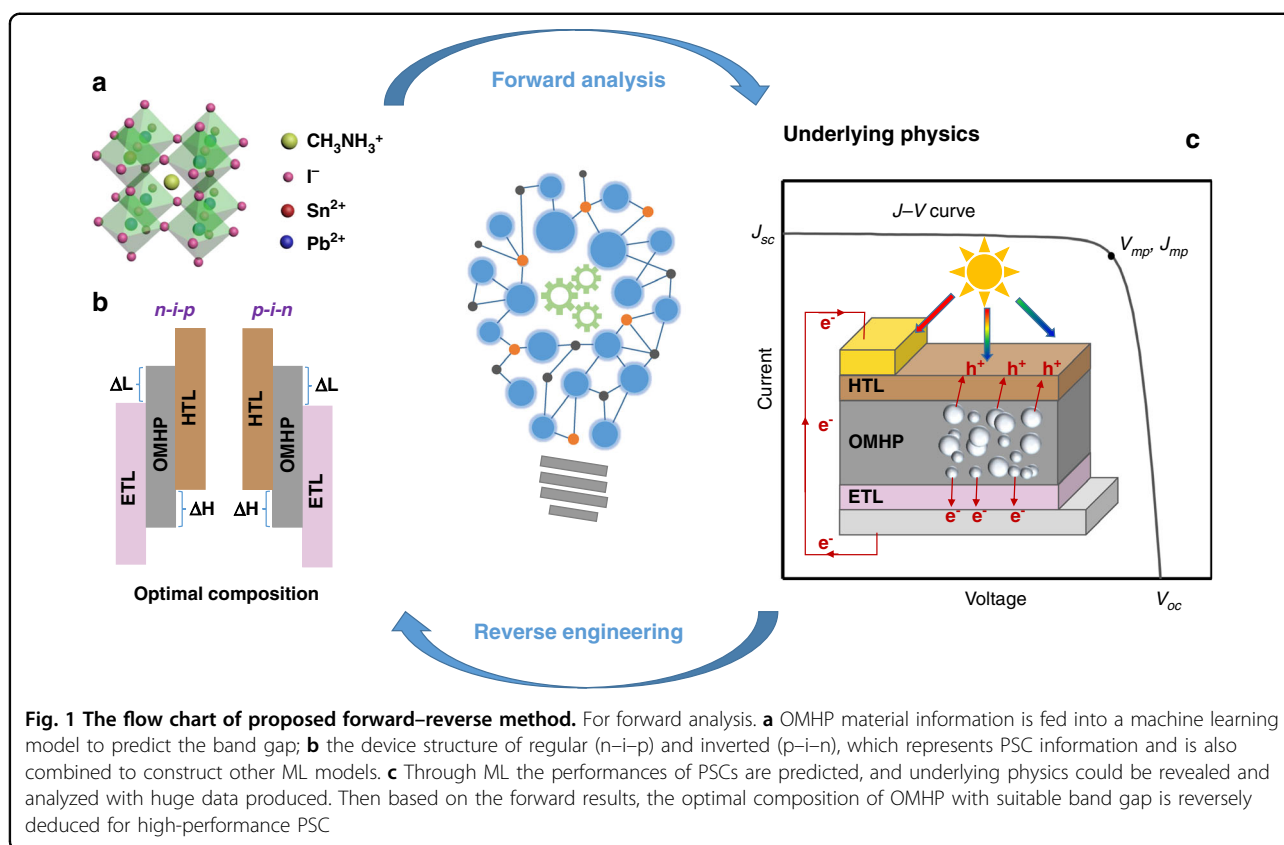
of 1850 entries of OPV characteristics, performance and stability data and employed a sequential minimal optimization regression model as means of determining the most influential factors governing the solar cell stability and PCE³⁵. However, few researches extend similar method to PSCs and ML is rarely used to find the relationship of PSC performance with material properties. Although Odabaşı et al. demonstrated the influencing factors of PSC performance using ML according to a large number of published papers³⁶, their work is mainly based on statistical analysis and a clear description about the underlying physics needs to be given. Li et al. predicted the band gap of perovskite materials and PCE of PSCs through ML³⁷ and the physical law learned by ML was explored. However, currently most work using ML is purely conducting simulations, and researches on PSCs combined simulations and experiments are still lacking.

In the present work, we propose a forward-reverse framework for the first time, to establish the relationship between key parameters and photovoltaic performances and realize the optimization of photovoltaic performances in mixed Sn-Pb PSCs. In the process of forward design, firstly, we establish a ML model that predicts the band gap (E_g) of OMHP materials with varied composition, and verify the E_g model by experiments. Then, considering the energy levels of OMHP and carrier transport layers, the models that predict PSC performance (short-circuit current J_{sc} , open-circuit voltage V_{oc} , fill factor FF and PCE) are designed and implemented without prior information. The mechanisms behind different performance models are analyzed in detail and several physical rules are concluded as well. Finally, with the goal of high PCE, three features (Sn-Pb ratio, E_g and the publication time of reported article) are used to build PCE model and directly retrodict the optimal OMHP composition. To the best of our knowledge, this process has not been reported on PSCs before. Based on the predicted optimal proportion for mixed Sn-Pb PSCs, we fabricate the samples and the predicted results are verified.

Results

Machine learning models

The OMHP with formula of ABX₃ is studied, where the cation A is methylammonium CH₃NH₃⁺ (MA⁺), B is the inorganic cation lead (Pb²⁺) or Tin (Sn²⁺), and X is iodide (I), and the basic device structure of PSC can be divided into two types according to the different contact materials with ITO: regular (n-i-p) and inverted (p-i-n)³⁸. The device performance is determined by the optical and electrical properties of OMHP, such as exciton binding energy, carrier mobility, the highest occupied molecular orbital (HOMO), the lowest unoccupied molecular orbital (LUMO), etc. The HOMO/LUMO level and interfacial properties of electron transport layer (ETL) and hole



transport layer (HTL) could also affect the performance of PSCs. The forward–reverse framework we propose, to study the mixed Sn–Pb perovskites, as shown in Fig. 1, includes two procedures: forward analysis and reverse engineering. The task of forward analysis procedure is to establish the E_g and device–performance models using ML, and that of reverse engineering procedure is to predict the optimization parameters of mixed Sn–Pb perovskites and experimental realization.

We apply five algorithms to different missions in both procedures, including linear regression (LR), support vector regression (SVR), k-nearest neighbor regression (KNR), random forest regression (RFR), gradient boosting regression (GBR) and neural network (NN), which can be implemented by Scikit-learn³⁹ and TensorFlow⁴⁰. In the reverse exploration, the established ML model is further inversely analyzed to carry out a search of maximum PCE value under some restrictions using genetic algorithm (GA) or Bayesian optimization algorithm (BO), which are implemented by Python packages (bayesian-optimization⁴¹ and deap⁴²). Related detailed explanations of these algorithms are provided in Supplementary notes.

Special care should be taken when preparing the data in ML modeling. During the preparation of ML dataset for E_g model, the repeated data points with the same Sn–Pb ratio and E_g is counted once, and if a given ratio

corresponds to different reported values of E_g , we keep all of those values to avoid data bias. For example, $\text{MASn}_{0.75}\text{Pb}_{0.25}\text{I}_3$ has four reported E_g values, i.e. two 1.18 eV, one 1.17 eV and one 1.27 eV. We keep one data point for 1.18 eV and two data points for 1.17 eV and 1.27 eV. To explore which elemental property under different Sn–Pb ratios is more dominant in determining the E_g of $\text{MASn}_x\text{Pb}_{1-x}\text{I}_3$, 14 physicochemical parameters⁴³ are used as descriptors of inputs for E_g regression model (see Table S1), and the SHAP (Shapley Additive exPlanations) method⁴⁴ is used to obtain the contribution of each feature to the E_g model.

When building the ML model to predict photovoltaic performance of PSCs including J_{sc} , V_{oc} , FF and PCE, three parameters are used as input parameters: i) the band gap (E_g) of OMHP material, ii) the energy difference (ΔH) between the HOMO of HTL and OMHP material, iii) the energy difference (ΔL) between the LUMO of ETL and OMHP material. The schematic energy band diagrams of n–i–p and p–i–n structures are presented in Fig. 1. For n–i–p or p–i–n, ΔH and ΔL are both calculated by $\Delta H = \text{HTL}_{\text{HOMO}} - \text{OMHP}_{\text{HOMO}}$ and $\Delta L = \text{OMHP}_{\text{LUMO}} - \text{ETL}_{\text{LUMO}}$.

Before reverse design, it is important to rebuild the PCE model for PSCs that involves features including Sn–Pb ratio, the E_g of OMHP and annual improvement of material qualities and optimization of processing (denoted

by publication date of research article) inspired from time series prediction. Since there is relatively uniform standard for PCE testing of PSCs, it is meaningful to use publication date as a feature to help the model explore the PCE progress trend of PSCs with passage of time. Due to the difference of preparation technologies, the performance of devices obtained by different research groups at the same ratio could also be different. So we use the maximum PCE result corresponding to each ratio in each year. Then, based on the mentioned model, the virtual design of OMHP materials is carried out by GA or BO.

It should be noted that, the data points are collected from published articles^{21,37} which are randomly divided into a training set and a test set in ratios of 90% and 10%. The dataset for ΔH and ΔL used to build the PCE model is shown in Figure S1 with 181 data points. To build the E_g prediction model, 43 data points are used from the performance model after discarding duplicate material composition data with the same reported value of E_g . In the model of reverse exploration, the data points with best PCE for a given preparation method in different publication times are chosen. For these models, we use 5-fold cross validation to optimize the hyper parameters of ML algorithm based on a Python library called hyperopt⁴⁵, which could effectively speed up the search of the values of multiple hyper parameters. The test subset is only used to test the quality of the built model. We evaluate the performance of constructed models by three indicators (the coefficient of determination R^2 , root mean square error RMSE and mean absolute error MAE).

Material design by ML: band gap

To utilize the solar energy to the maximum extent, the minimum E_g of mixed Sn-Pb perovskites used as the low-bandgap subcells to absorb low-energy photons, are critical to maximize the PCE of PTSCs. Moreover, manipulating the E_g by an appropriate ratio is also critical for the realization of high-efficiency and environmentally friendly single-junction PSCs based on mixed Sn-Pb perovskites. As previously reported, the valence bands (VBs) of MAPbI₃/MASnI₃ crystals are formed by the antibonding states of I-*p* and Pb/Sn-*s* atomic orbitals, and the conduction band (CBs) are formed by the antibonding states of I-*p* and Pb/Sn-*p* atomic orbitals. When the ratio of Sn increases in MASn_xPb_{1-x}I₃ alloys, the VB/CB levels undergo the shift from -5.45 eV/-3.90 eV ($x = 0$) to -5.47 eV/-4.17 eV ($x = 1$) respectively¹⁶ and Pb-*p* as CB and antibonding states of I-*p* and Sn-*p* as VB, which subsequently leads to the well-known bowing effect in the E_g of MASn_xPb_{1-x}I₃. However, the previously reported E_g values denoted as black dots in Fig. 2a are not directly linearly or parabolically dependent on the ratio x , since replacing of Pb by Sn will cause lattice compression and octahedral tilting in MASn_xPb_{1-x}I₃ structure, and increasing the former or enhancing the latter will increase or decrease the E_g ¹³. To reveal the underlying mechanism of the E_g in mixed Sn-Pb perovskites, we retrieve the contributions from elemental properties and structural information of mixed Sn-Pb perovskites by building the ML-bandgap model using 14 physicochemical parameters listed in Table S1 as inputs.

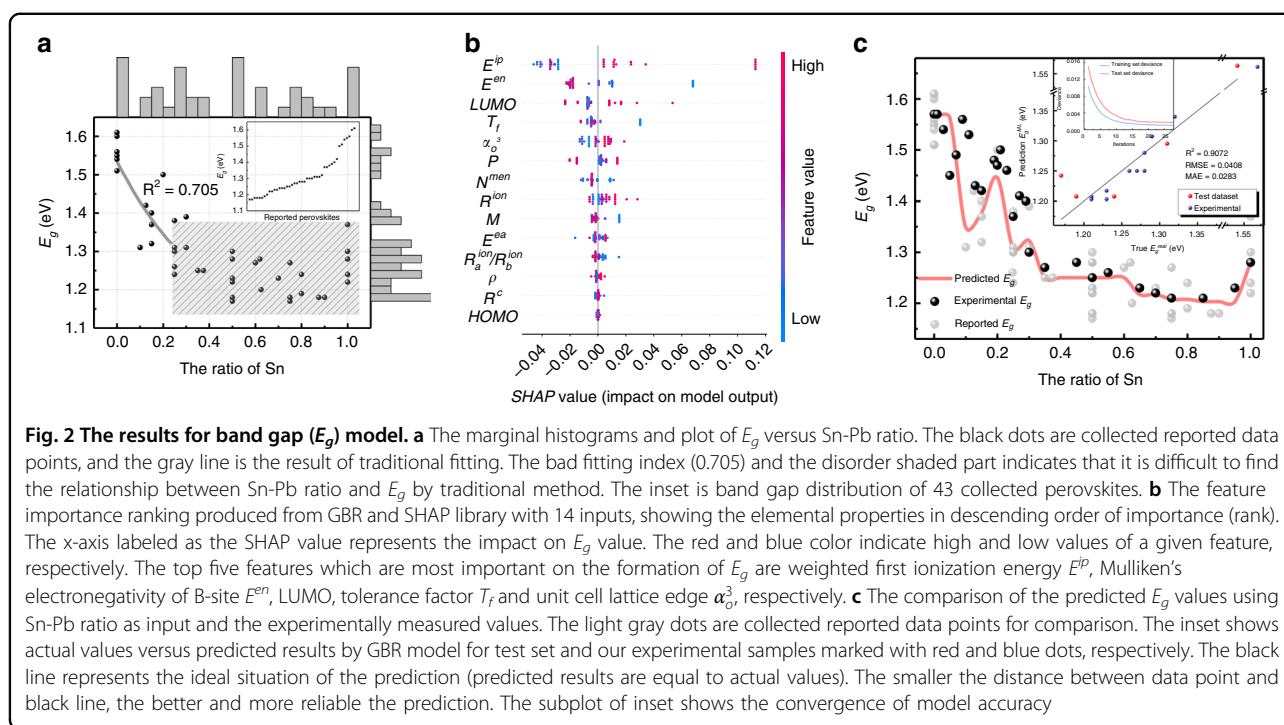


Table 1 The comparison of prediction performance of E_g models with different features and ML regression algorithms by three evaluation metrics (R^2 , RMSE and MAE). The best results are highlighted in bold.

Method	14 features			One feature		
	R^2	RMSE	MAE	R^2	RMSE	MAE
LR	0.6864	0.0751	0.0630	0.6064	0.0841	0.0707
SVR	0.8775	0.0469	0.0422	0.8221	0.0565	0.0516
KNR	0.8997	0.0425	0.0340	0.8833	0.0458	0.0365
RFR	0.9105	0.0401	0.0284	0.9065	0.0410	0.0344
GBR	0.9172	0.0386	0.0325	0.9072	0.0408	0.0283

Due to the limited available E_g data, we only use traditional ML method without NN. Five ML methods (LR, SVR, KNR, RFR and GBR) are used to build the E_g model with 14 features as input and the results are listed in Table 1. The GBR algorithm performs best with the corresponding values of R^2 , RMSE and MAE of 0.9172, 0.0386, and 0.0325, respectively. The SHAP method is used to further interpret the GBR-bandgap model⁴⁴. The calculated feature importance ranking produced from the GBR and SHAP library is shown in Fig. 2b, with x-axis labeled as the SHAP value representing the impact on E_g value, and the red and blue colors indicating high and low values of a given feature, respectively. The top five features which are most important on the formation of E_g are weighted first ionization energy E^{ip} , Mulliken's electronegativity of B-site E^{en} , LUMO, tolerance factor T_f and unit cell lattice edge α_o^3 , respectively. Obviously, LUMO plays a more important role than HOMO does, which is in good agreement with the large energy difference of -0.27 eV in CB and the small one of -0.02 eV in VB when Pb is replaced by Sn in $MASn_xPb_{1-x}I_3$ alloys mentioned above. In fact, when replacing Pb by Sn, E^{en} and T_f increase, and E^{ip} , LUMO and α_o^3 decrease. However, as shown in Fig. 2b, only the decreasing E^{ip} and increasing E^{en} clearly decrease the E_g . The decreasing LUMO when the ratio of Sn increases leads to the simultaneous increase and decrease of E_g , which is consistent with the bowing effects as shown in Fig. 2a. In a similar way, the replacing of big Pb atoms by small Sn atoms changes the structural characteristics, and subsequently decreases α_o^3 and increases T_f , which also leads to the simultaneous increase and decrease of E_g as shown in Fig. 2b.

To reduce the complexity and facilitate the PSC-device design, we further build a new E_g model based on one-dimensional feature (Sn-Pb ratio) as input. In this case, the GBR algorithm also performs best with the corresponding values of R^2 , RMSE and MAE of 0.9072, 0.0408, and 0.0283, respectively, as listed in Table 1. To verify the precision of our new ML-bandgap model, we experimentally fabricate a

series of new mixed Sn-Pb perovskite samples with various compositions. The measured and GBR-predicted E_g along with reported results is shown as inset in Fig. 2c, with blue dots denoting our experimental results and red dots denoting previously reported results. The experimental E_g values are deduced by Tauc plot (Fig. S2). It shows that our trained E_g model could not only predict the data in test set, but also accurately predict E_g of the new $MASn_xPb_{1-x}I_3$ samples. As previously mentioned, the replacing of Pb by Sn leads to significant changes in structure, e.g. lattice compression, octahedral tilting, and in electronic properties, e.g. VB/CB reconstructing, which thus results in a complicated dependence of E_g on Sn ratio in mixed Sn-Pb perovskites. To further reveal the dependence of E_g on Sn ratio, as shown in Fig. 2c, the ratio-dependent E_g for mixed Sn-Pb perovskite predicted by our new ML-bandgap model (red line), which match well with our fabricated ones (black dots), manifests itself with asymmetrically bowing shape and a minimum value of 1.198 eV of at the Sn ratio of 93.3%. The optimized Sn ratio for the S-Q limit at the E_g of 1.35 eV is predicted to be 10.0%, 12.2% and 23.3%.

Perovskite solar cell design: photovoltaic performance

As mentioned above, both n-i-p and p-i-n PSC devices are under study here as shown in Fig. 1b and are composed of a thin-layer photoactive OMHP absorbers sandwiched between two transport layers ETL and HTL, which are used to selectively transport electrons and holes to cathode and anode respectively. Ideally, incident photons are absorbed with nearly unity efficiency by OMHP film, and densities of photo-generated nonequilibrium free carriers in CB and VB can be evaluated by the quasi-Fermi level splitting (QFLS), i.e. $QFLS = E_F^e - E_F^h$ where $E_F^{e/h}$ is the quasi-Fermi level for electrons and holes respectively, which is also the open-circuit voltage (V_{oc}) of the PSC devices in the S-Q theory. The values of V_{oc} are deteriorated mainly by the unwanted nonradiative combinations occurring at the surface defects or grain boundaries in the OMHP film. Then the photo-generated electrons and holes are selectively extracted by the ETL and HTL to the cathode and anode respectively, and generate currents, e.g. J_{sc} at short-circuit condition. The extraction efficiency of carriers from OMHP film is mainly affected by the surface qualities and the energy differences between OMHP and HTL/ETL, i.e. ΔH (ΔL). The values of ΔH (ΔL) are generally larger than or equal to zero, since negative ones will introduce extraction barriers between OMHP film and transport layers, but they can not be too large to induce significant energy loss at the interface and form the transport barriers between transport layers and electrodes, which in turn may decrease J_{sc} . For example, in our case, the HOMO and LUMO of OMHP are -5.4 eV and -3.9 eV respectively, and the work functions of working anode (Au)/cathode (Al) are -5.1 eV and -4.3 eV, respectively. When ΔH (ΔL) between HTL (ETL) and OMHP film

Table 2 The prediction performance of different regression algorithms for four targets (FF, J_{sc} , V_{oc} and PCE) in designing PSCs device using experimental E_g , ΔH and ΔL as inputs

Method	FF			J_{sc}			V_{oc}			PCE		
	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE
LR	0.4750	14.3140	8.1743	-0.0676	5.3079	2.9698	0.6047	0.1345	0.0968	0.3422	4.0129	2.8447
SVR	0.2937	16.6026	7.4239	0.1519	4.7310	2.4441	0.4389	0.1602	0.1084	0.5518	3.3124	2.2708
KNR	0.4580	14.5433	6.9755	0.7913	2.3471	1.8209	0.8022	0.0951	0.0557	0.6229	3.0385	2.5294
RFR	0.5698	12.9566	6.9801	0.5262	3.5362	2.3123	0.6068	0.1341	0.0936	0.6598	2.8858	2.2669
GBR	0.5477	13.2863	7.1833	0.0239	5.0754	2.7716	0.6873	0.1196	0.0884	0.7651	2.3982	1.8816
NN	0.9096	4.5641	3.4786	0.8378	2.6438	1.6301	0.9024	0.0347	0.0289	0.9026	1.9626	1.5439

The best results are highlighted in bold and NN behaves the best among competitive algorithms due to higher R² and smaller RMSE and MAE

is higher than 0.3 (0.4) eV, HTL-anode (ETL-cathode) extraction potential barriers is generated. Generally, the open-circuit voltage V_{oc} relates to the short-circuit current J_{sc} by the expression,

$$V_{oc} = \frac{n_{id}k_B T}{q} \ln\left(\frac{J_{sc}}{j_0} + 1\right) \quad (1)$$

where n_{id} is the ideality factor and j_0 is the dark generation current. The electric power is zero at both open- and short-circuit conditions, and reaches a maximum power point at which the voltage V_{mp} and current J_{mp} give the fill factor FF of a PSC device, i.e. $FF = V_{mp}J_{mp}/J_{sc}V_{oc}$. Empirically FF can be written as⁴⁶,

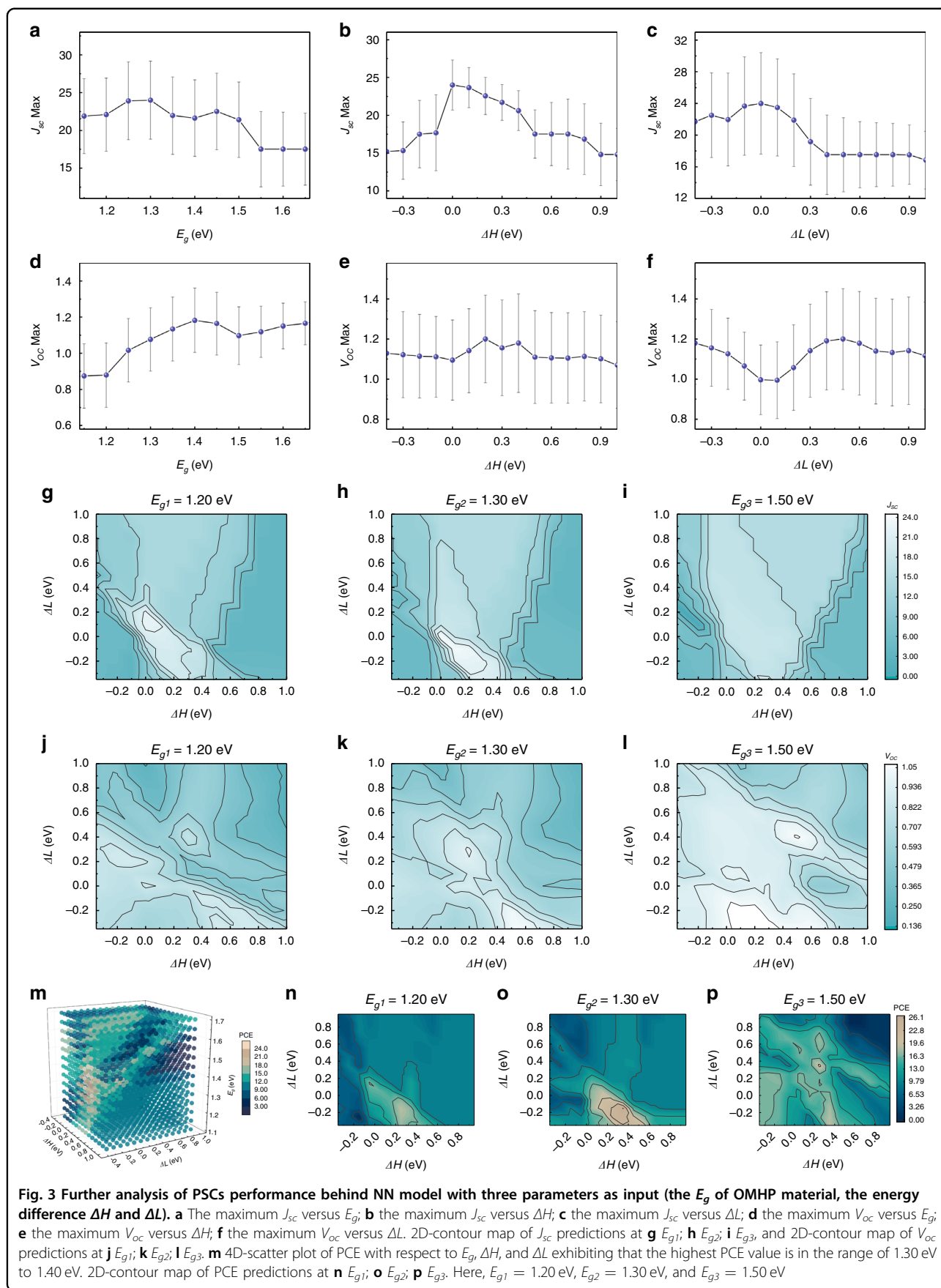
$$FF = \frac{v_m}{v_m + 1} \frac{v_{oc} - \ln(v_m + 1)}{v_{oc}(1 - e^{-v_{oc}})} \quad (2)$$

where $v_{ov} = V_{oc}/n_{id}k_B T$ and $v_m = v_{oc} - \ln(v_{oc} + 1 - \ln v_{oc})$. Subsequently PCE is given by $\eta_{PCE} = J_{sc} \times FF \times V_{oc}/P_{sun}$, where P_{sun} is the total incoming solar energy. For the mixed Sn-Pb perovskite as the OMHP film in PSC under study here, the increasing of Sn introduces more Sn vacancies and oxidized Sn⁴⁺ in the OHMP, which may enhance the non-radiative recombination via Shockley-Read-Hall and Auger recombination processes at the defects, decreasing V_{oc} and J_{sc} , even though E_g is tuned to the optimized value predicted by the S-Q theory. Therefore, to achieve the maximum PCE for MASn_xPb_{1-x}I₃-based PSC devices, a detailed optimization process regarding the influence on J_{sc} , V_{oc} and PCE from E_g , ΔH and ΔL is necessary.

Similar to the aforementioned process, the influences of E_g , ΔH and ΔL on photovoltaic performances of mixed Sn-Pb PSC devices, i.e. V_{oc} , J_{sc} , FF and PCE, are investigated using ML methods, and the performances of different ML algorithms are listed in Table 2, in which experimental E_g , ΔH and ΔL are used as inputs, as

comparison to those using predicted E_g as inputs listed in Table S2. In both simulations, NN behaves better than others, which confirms the superiority of NN in dealing with complex cases. By comparison, we find that, for J_{sc} using the predicted E_g as feature, the model metrics is improved, but for FF model, using the predicted E_g will worsen the model. For PCE and V_{oc} models, whether the predicted E_g is used, the model results change little. In addition, no matter what kind of situation and target, the values of R² of all models are significantly greater than 0.5, which means the built models can give relatively accurate prediction. The reason for large variation of metrics in FF with different models might be that the factors affecting FF are complex, and the dependence on involved features is beyond our collected data. The performances of NN in our dataset for four targets are shown in Figure S3, and the descent process of training/test set loss values during NN training is provided in Figure S4, both of which verify the convergence of our NN model. Herein, we set ΔH and ΔL from -0.35 eV to 1.00 eV and E_g from 1.15 eV to 1.65 eV in the prediction set.

The relationships of J_{sc} and V_{oc} with E_g are shown in Fig. 3, with the bar indicating standard deviation. As we know, in the S-Q theory, J_{sc} decreases and V_{oc} increases when E_g increase, where nonradiative recombination are totally neglected. However, as shown in Fig. 3a, in the E_g regions ranging roughly from 1.15 eV to around 1.25 eV and 1.40 eV to 1.50 eV, which are corresponding to low-bandgap-plateau (LBP) (rich-Tin, >50% Sn) and defective-zone (DZ) (poor-Tin, <20% Sn) regions respectively, J_{sc} increases when E_g increase, which are consistent with the unchanged or decreasing V_{oc} with increasing E_g as shown in Fig. 3d. The deviation observed in predicted J_{sc} and V_{oc} from the S-Q theory captured in our model, which was experimentally observed recently in ((HC(NH₂)₂)_{0.83}Cs_{0.17})(Pb_{1-y}Sn_y)I₃ family of perovskite materials as well⁴⁷, is due to the significant nonradiative recombination at the deep level traps induced by structural disorders (DZ) and



enhanced susceptibility of Sn^{2+} to oxidation (LBP). Similar to the situation in $((\text{HC}(\text{NH}_2)_2)_{0.83}\text{Cs}_{0.17})(\text{Pb}_{1-y}\text{Sn}_y)\text{I}_3$ perovskites, in the poor-Tin (DZ) region, the replacing of Sn in neat Pb perovskites induces local heterogeneity around Sn sites along with overall lattice compression and octahedral tilting, and such crystallographic distortion leads to energetic disorder near the band edges accommodating deeper traps formation, accompanied with a rapid change in E_g as shown in Fig. 2a. Therefore, the nonradiative J_{sc} are enhanced in the DZ region. However, in the rich-Tin (LBP) region, due to the cease of volume compression in the compositional perovskites, crystallographic and electronic stabilities are achieved, as shown in Fig. 2a as well. Since Sn content is beyond 50% in the LBP region, the subsequently enhanced electric susceptibility resulted from increased Sn^{4+} content also deteriorates the photovoltaic performances (J_{sc} and V_{oc}).

For ΔH and ΔL , the photo-generated carrier transport is not only affected by extraction barriers Δ , but also by the type of transport layer material, fabrication processing, etc. Thus it is complicated to obtain optimized J_{sc} and V_{oc} by manipulating ΔH and ΔL . As shown in Fig. 3b, c, the changes of J_{sc} with ΔH and ΔL are similar and it is not conducive to J_{sc} as ΔH (ΔL) is too large or too small. However, as shown in Fig. 3e, f, different from J_{sc} , the maximum value of V_{oc} does not appear in the region where ΔH (ΔL) is equal to zero. To further analyze the extraction barrier related to J_{sc} and V_{oc} with different fixed E_g values in detail, 2D-contour maps of maximum J_{sc} and V_{oc} prediction with different E_g are calculated and shown in Fig. 3g–l, where x-axis and y-axis represent ΔH and ΔL , respectively. The lighter color indicates the higher J_{sc} and V_{oc} . And for simplification, the values of E_g are chosen as 1.20 eV, 1.30 eV and 1.50 eV to conduct further analysis, which belong to the LBP, normal and DZ regions, respectively.

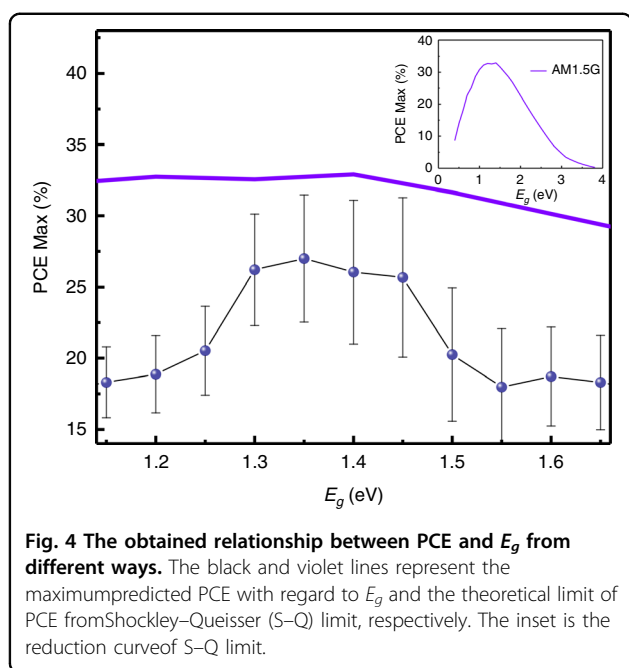
For J_{sc} as shown in Fig. 3g–i, as the E_g increase from 1.20 eV→1.30 eV→1.50 eV, the values of J_{sc} overallly decreases, which is consistent with the S-Q theory. When ΔH (ΔL) is much lower than zero, it means there is a big potential barrier between OMHP material and HTL (ETL), which will block the direct transfer of carriers and consequently decrease the current. When ΔH (ΔL) is much higher than zero, it could increase the speed of the carriers as they pass through transport layer, but this could also lead to induce energy loss at the interface and transport barrier between the transport layer and the electrode, and thus decreasing J_{sc} . Figure 3g–i also show that, when ΔL is negative, it is possible to manipulate ΔH to achieve the maximum J_{sc} , but not vice versa. Take Fig. 3h as an example. The maximum J_{sc} appears at $\Delta H = 0.2$ eV and $\Delta L = -0.2$ eV, i.e, there is 0.2 eV barrier between ETL and OMHP and 0.2 eV positive energy

difference between HTL and OMHP. In this condition, the maximum J_{sc} of 21.7 mA/cm² achieves. Conversely, for $\Delta H = -0.2$ eV and $\Delta L = 0.2$ eV, the value of J_{sc} drops (2.5 times) to 9.8 mA/cm², which should be due to the more difficult extraction for holes in HTL compared with electron extraction in ETL, since holes possess larger effective mass and smaller mobility compared to electrons⁴⁸. Therefore in the actual PSC design, employing excellent hole transport would lead to high photovoltaic performance^{49,50}.

For V_{oc} as shown in Fig. 3j–l, as the E_g increase from 1.20 eV→1.30 eV→1.50 eV, the values of V_{oc} overallly increase, which is consistent with the S-Q theory as well. The difficulty of hole extraction in HTL can also be observed. For example, as shown in Fig. 3l, the maximum V_{oc} appears in the region of $\Delta H = 0.2$ eV and $\Delta L = -0.2$ eV, which means 0.2 eV barrier between ETL and OMHP can promote the accumulation of electrons at the interface subsequently enlarging QFLS within OMHP and thus V_{oc} . Because the transport of holes in OMHP is more difficult than electrons, positive ΔH is required to facilitate the hole extraction, which could reduce the concentration of holes in OMHP thus reducing unwanted nonradiative recombination.

For the goal PCE, Fig. 3m shows a 4D-scatter plot of predicted PCE based on NN algorithm, where ΔH and ΔL are changed from -0.35 eV to 1.00 eV and E_g is changed from 1.15 eV to 1.65 eV, which reveals that, the highest PCE values are in the range of 1.30 eV to 1.40 eV. To further analyze the relationship between PCE and E_g , the maximum PCE of each E_g is extracted from Fig. 3m and plotted in Fig. 4, where the theoretical prediction by S-Q theory under AM1.5 G radiation is also shown. Both S-Q limit and our ML results show that the ideal E_g of 1.35 eV^{12,51,52}, and the trend of both lines is also consistent, which is an astonishing finding since only data from $\text{MASn}_x\text{Pb}_{1-x}\text{I}_3$ system are used in our model without providing any information about the solar spectrum. The peak at about 1.60 eV is probably due to the extensive use of this material by researchers and the continuous optimization achieved for the performance of PSCs on this material.

In addition, in order to obtain the relationship of PCE with ΔH and ΔL in the different values of E_g , 2D-contour maps are also drawn in Fig. 3n–p. The region of highest PCE shifting from smaller ΔH , ΔL to higher ΔH , ΔL with the increase of E_g of OMHP material, was also observed in perovskite materials of ABX_3 -type ($\text{A}=\text{MA}/\text{FA}/\text{Cs}^+$, $\text{B}=\text{Pb}/\text{Sn}^{2+}$, $\text{X}=\text{Br}/\text{Cl}/\text{I}$)³⁷, which manifests the extensibility of our ML model to general OMHP systems. Furthermore, it is interesting that as we compare Fig. 3n–p with Fig. 3g–l, the improvement of goal PCE at 1.30 eV (near optimal E_g) is mainly due to enhanced J_{sc} because both maximum region and contour overlap. For comparison to further check our model, according to Eq. (2),



the filling factor FF can be obtained with ML-predicted V_{oc} values and the ideality factor n_{id} of 2.5 is deduced according to Eq. (1), and finally PCE can be obtained. The calculated dependence of PCE on E_g , ΔH and ΔL is shown as Figure S5, which is reasonably consistent with the ML-predicted results as shown in Fig. 3n–p.

Reverse engineering

In order to further verify the effectiveness of ML, we have performed experimental tests. Through forward analysis, the process of OMHP material selection and PSC design are completed, but the relationship between Sn–Pb ratio and PCE of PSC cannot be directly given, that is, to determine which ratio could give the corresponding maximum predicted PCE. Therefore, in this section, a data-driven reverse engineering is proposed to explore potential property of $\text{MA}\text{Sn}_x\text{Pb}_{1-x}\text{I}_3$ material used in PSCs based on one-step spin coating process and inverted structure.

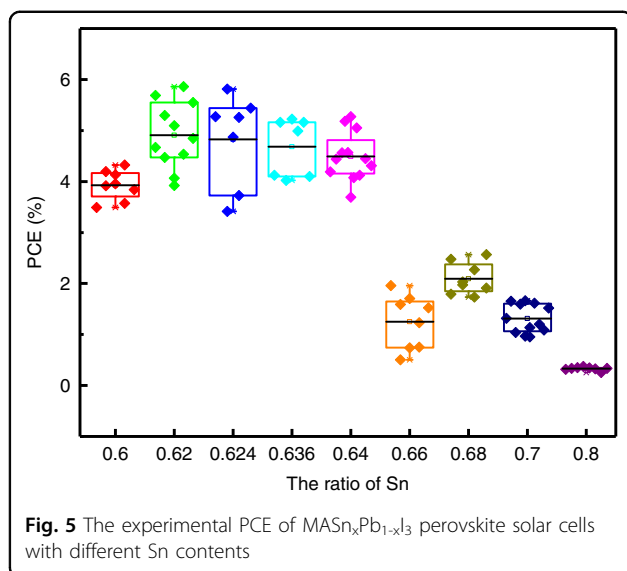
Due to the reduction of dataset resulting from the limit of spin coating process and device structure, as previously mentioned, three features of Sn–Pb ratio, E_g and the publication time of reported article are taken to build a new PCE model based on GBR algorithm. Then the established GBR model is further inversely analyzed to lead a search of maximum PCE using BO and GA. The search conditions of 10 initial points along with 20 iterations, and the population size of 100 along with 20 generations are used for BO and GA, respectively. The element Sn in the B-site has the doping ratio from 0.5 to 1.0 with step 0.001, and the second element Pb is given the remaining doping ratio, because the E_g obtained by

doping Sn in this range has the opportunity to obtain the maximum PCE according to previous forward analysis, while reducing the content of Pb as much as possible to reduce the toxicity of obtained OMHP. Figure S6 shows the fitness of PCE to the search round and shows how it changes with the optimization variable. GA presents a more stable and efficient exploration process than BO. However, both GA and BO give very close maximum PCE, i.e., 18.23% for $\text{MA}_{1.0}\text{Sn}_{0.624}\text{Pb}_{0.376}\text{I}_3$ by GA and 18.22% for $\text{MA}_{1.0}\text{Sn}_{0.636}\text{Pb}_{0.364}\text{I}_3$ by BO, respectively.

With the previous settings, we select a series of ratios from 0.5 to 1.0 for experimental validation under the set fabrication processing. As reported in previous works, the perovskite crystallization processes⁵³, grain boundary management⁵⁴, interface engineering⁵⁵ and charge transport layer selection⁵⁶ were proved to be critical aspects toward high-efficiency perovskite solar cells. Every precise tuning of above sections in each experimental condition is definitely difficult and time consuming. For now, we do not conduct special optimization for devices and choose a trade-off fabrication process to adapt all conditions to keep the consistency of the processing condition for different ratios to make the horizontal comparability more obvious. Actually, we focus on the curve trend and the obtained results are shown in Fig. 5, which reveals good consistency between experimental data and prediction trend in PCE distribution. The ratios corresponding to the best PCE obtained by GA and BO are near the region of highest experimental result, which illustrates the rationality and accuracy of our model. Moreover, in the range of experimental preparation, the PCE of device deviated from the optimal composition becomes very poor. The corresponding plots of J_{sc} , V_{oc} and FF for these devices are provided in Figure S7. Compared with the prediction of 18%, the efficiency corresponding to optimal ratio in our experiment still has much room for development that needs further experimental optimization. At present, the best efficiency obtained by Sadhanala et al. in set range of Sn content is 10% at 0.6 that ratio is close to our prediction. Specific data are provided in Table S3. With the accumulation of samples in the future, the optimization measures, such as considering additives, selecting better transport layer materials, etc., should be considered. In addition, our proposed ML approach in the forward–reverse framework can be applied to other perovskites materials with similar structures, such as ABX_3 -type perovskite materials ($A = \text{MA/FA/Cs}^+$, $B = \text{Pb/Sn}^{2+}$, $X = \text{Br/Cl/I}$).

Discussion

Combining with ML technology, we have proposed an efficient forward–inverse method to research $\text{MA}\text{Sn}_x\text{Pb}_{1-x}\text{I}_3$ material and explore high-performance PSCs. We use real experimental data aiming at getting



more practical results. During forward analysis, the E_g model of $\text{MASn}_x\text{Pb}_{1-x}\text{I}_3$ is first built with 14 physico-chemical parameters and Sn-Pb ratio as input respectively, and the asymmetrically bowing relationship between Sn-Pb ratio and E_g of OMHP is found, which match well with our fabricated ones. For the performance models of PSC, the established NN-based models exhibit satisfactory prediction for underlying data points and provide some guidance and new discoveries for PSC devices. The relationship of J_{sc} , V_{oc} and PCE with E_g of OMHP and energy level difference ($\Delta H/\Delta L$) are analyzed in detail respectively. The relationship of J_{sc} and V_{oc} with E_g matches the theoretical results well in most case. The deviation of J_{sc} and V_{oc} in LBZ and DZ region from the S-Q theory is also captured by our model, which is experimentally observed recently. The conclusion that hole extraction is more difficult is both obtained in J_{sc} and V_{oc} models. And the highest PCE of single-junction PSCs at about 1.35 eV is predicted without any prior information, which is quite consistent with the theoretical result of S-Q limit. Then in the reverse engineering, the optimal ratio of Sn in $\text{MASn}_x\text{Pb}_{1-x}\text{I}_3$ within a set range is directly obtained for high-performance PSCs and inference results fit well with experimental validation.

With the proposed target-driven approach, the physical laws obtained are reasonable and the predictions are verified experimentally in $\text{MASn}_x\text{Pb}_{1-x}\text{I}_3$ system. Our designed method could be expected to provide deeper understanding of physical phenomena as well as explore new functional materials and high-performance devices. For complex systems involving multiple variables, this method could give results significantly faster. With the accumulation of a database, it could also constantly learn with itself to obtain stronger predictive ability to give more helpful and accurate guidance.

Materials and methods

Materials

SnI_2 (99.999% purity) was purchased from Alfa Aesar. N, N-dimethylformamide (DMF), dimethyl sulfoxide (DMSO) and SnF_2 (99% purity) were purchased from Sigma-Aldrich. Methylammonium iodide (MAI) was purchased from Dyesol. Chlorobenzene was purchased from Thermo Fisher. PEDOT:PSS aqueous solution (AI 4083) was purchased from Heraeus Clevios. PbI_2 was purchased from TCL.

Mixed tin-lead perovskite film fabrication

A stock solution of 1.4 M MAPbI_3 solution (i.e. 0% Sn) was prepared by dissolving 1113 mg MAI and 3227 mg PbI_2 in 5 mL of a mixed solvent of 7:3 DMF:DMSO by volume. A stock solution of 1.4 M MASnI_3 (i.e. 100% Sn) with 10% molar excess of SnF_2 with respect to the SnI_2 content was prepared by dissolving 1113 mg MAI, 2604 mg SnI_2 , 219 mg SnF_2 in 5 mL of a mixed solvent of 7:3 DMF:DMSO by volume. Solutions with Sn content = 0, 15, 25, 30, 35, 45, 50, 55, 65, 70, 75, 85, 95, 100% were prepared by mixing the 0% and 100% Sn stock solutions in the appropriate ratio. Solutions with Sn content = 62, 62.4, 63.6, 64, 66, 68% were prepared through mixture of the 60% Sn solution with the 70% Sn solution. The precursor solution was stirred for at least 2 h at room temperature. Prior to spin-coating, the perovskite solution was filtered with a 0.22 μm PTFE syringe filter. The perovskite precursors with different Sn content were spin-coated onto the ITO/PEDOT:PSS substrates at 5000 rpm for 30 s. During the spin-coating, 750 μL of toluene was dropped onto the spinning substrates. Then perovskite films were annealed at 100 $^\circ\text{C}$ for 10 min to form mixed tin-lead perovskite films.

Mixed tin-lead perovskite solar cell fabrication

ITO-glass was cleaned using acetone, surfactant, deionized (DI) water, ethanol with ultrasonication for 30 min sequentially and dried with N_2 flow. ITO-glass was further cleaned by O_2 -plasma treatment for 15 min. PEDOT:PSS aqueous solution was then deposited on the ITO at 4000 rpm for 50 s followed by annealing on a hotplate at 175 $^\circ\text{C}$ for 60 min in ambient air. The perovskite films were prepared on the ITO/PEDOT:PSS using mixed tin-lead perovskite film fabrication method described above. After the deposition of the perovskite film, C60 (20 nm)/BCP (3 nm)/Ag (100 nm) were sequentially deposited by thermal evaporation.

Material characterization

The transmission spectra and absorption spectra were both measured using F20-UV thin-film analyzer (FILMETRICS) at wavelength range between 300 nm and 1100 nm. Tauc plots of absorption spectra to determine the impact of different Sn contents on perovskite band gaps was calculated by absorption spectra.

Device characterization

The J–V curves of PSCs were measured using a Keithley 2602B source in N₂ filled glove box at room temperature under AM 1.5 G condition at an intensity of 100 mW/cm², calibrated by a standard Si solar cell (PVM937, Newport). The light source was a 450 watt xenon lamp (Oriol solar simulator, 94023 A). The active area of PSCs was 0.107 cm², defined by the cross of patterned Ag and ITO electrode, and further calibrated by the microscope. The J–V curves were tested both at forward scan (from –0.2 V to 0.7 V, step 0.04 V) without any pre-conditioning before the test.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (grant numbers 61774046 and 11374063), and by Shanghai Municipal Natural Science Foundation under Grant Nos. 19ZR1402900.

Author details

¹School of Information Science and Technology, Fudan University, Shanghai 200433, China. ²College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, China. ³Center of Micro-Nano System, Fudan University, Shanghai 200433, China. ⁴Department of Physics, Chemistry and Biology, Linköping University, Linköping SE-58183, Sweden. ⁵Key Laboratory of Micro and Nano Photonic Structures and Department of Optical Science and Engineering, Fudan University, Shanghai 200433, China. ⁶Yiwu Research Institute of Fudan University, Chengbei Road, Yiwu City, Zhejiang 322000, China

Author contributions

Y.Z. and X.C. conceived the idea. X.C. did ML prediction and F.L. did experimental realization. X.C., F.L. and Y.Z. co-wrote the paper with all authors contributing to discussions and to finalizing the manuscript.

Competing interests

The authors declare no competing interests.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41377-022-00924-3>.

Received: 15 February 2022 Revised: 13 June 2022 Accepted: 30 June 2022
Published online: 26 July 2022

References

- Kojima, A. et al. Organometal halide perovskites as visible-light sensitizers for photovoltaic cells. *J. Am. Chem. Soc.* **131**, 6050–6051 (2009).
- Lee, M. M. et al. Efficient hybrid solar cells based on meso-superstructured organometal halide perovskites. *Science* **338**, 643–647 (2012).
- Jeon, N. J. et al. Solvent engineering for high-performance inorganic–organic hybrid perovskite solar cells. *Nat. Mater.* **13**, 897–903 (2014).
- Green, M. A., Ho-Baillie, A. & Snaith, H. J. The emergence of perovskite solar cells. *Nat. Photonics* **8**, 506–514 (2014).
- Yang, W. S. et al. Iodide management in formamidinium-lead-halide-based perovskite layers for efficient solar cells. *Science* **356**, 1376–1379 (2017).
- National Renewable Energy Laboratory (NREL). Best research-cell efficiency chart. (2022). <https://www.nrel.gov/pv/cell-efficiency.html>.
- Oga, H. et al. Improved understanding of the electronic and energetic landscapes of perovskite solar cells: high local charge carrier mobility, reduced recombination, and extremely shallow traps. *J. Am. Chem. Soc.* **136**, 13818–13825 (2014).
- Wehrenfennig, C. et al. High charge carrier mobilities and lifetimes in organolead trihalide perovskites. *Adv. Mater.* **26**, 1584–1589 (2014).
- Zhang, W. et al. Ultrasoft organic-inorganic perovskite thin-film formation and crystallization for efficient planar heterojunction solar cells. *Nat. Commun.* **6**, 6142 (2015).
- Jacobsson, T. J. et al. Exploration of the compositional space for mixed lead halogen perovskites for high efficiency solar cells. *Energy Environ. Sci.* **9**, 1706–1724 (2016).
- Liu, D. Y. & Kelly, T. L. Perovskite solar cells with a planar heterojunction structure prepared using room-temperature solution processing techniques. *Nat. Photonics* **8**, 133–138 (2014).
- Shockley, W. & Queisser, H. J. Detailed balance limit of efficiency of p–n junction solar cells. *J. Appl. Phys.* **32**, 510–519 (1961).
- Gu, S. et al. Tin and mixed lead–tin halide perovskite solar cells: progress and their application in tandem solar cells. *Adv. Mater.* **32**, 1907392 (2020).
- Eperon, G. E. et al. Perovskite–perovskite tandem photovoltaics with optimized band gaps. *Science* **354**, 861–865 (2016).
- Ogomi, Y. et al. CH₃NH₃Sn_xPb_{1-x}I₃ perovskite solar cells covering up to 1060 nm. *J. Phys. Chem. Lett.* **5**, 1004–1011 (2014).
- Hao, F. et al. Anomalous band gap behavior in mixed Sn and Pb perovskites enables broadening of absorption spectrum in solar cells. *J. Am. Chem. Soc.* **136**, 8094–8099 (2014).
- Tsai, C. M. et al. Role of tin chloride in tin-rich mixed-halide perovskites applied as mesoscopic solar cells with a carbon counter electrode. *ACS Energy Lett.* **1**, 1086–1093 (2016).
- Zhao, B. D. et al. High open-circuit voltages in tin-rich low-bandgap perovskite-based planar heterojunction photovoltaics. *Adv. Mater.* **29**, 1604744 (2017).
- Rajagopal, A. et al. Highly efficient perovskite–perovskite tandem solar cells reaching 80% of the theoretical limit in photovoltage. *Adv. Mater.* **29**, 1702140 (2017).
- Lin, R. X. et al. Monolithic all-perovskite tandem solar cells with 24.8% efficiency exploiting comproportionation to suppress Sn(II) oxidation in precursor ink. *Nat. Energy* **4**, 864–873 (2019).
- Wang, C. L. et al. Low-bandgap mixed tin-lead perovskites and their applications in all-perovskite tandem solar cells. *Adv. Funct. Mater.* **29**, 1808801 (2019).
- Chang, C. Y. et al. Solution-processed conductive interconnecting layer for highly-efficient and long-term stable monolithic perovskite tandem solar cells. *Nano Energy* **55**, 354–367 (2019).
- Kapil, G. et al. Tin-lead perovskite solar cells fabricated on hole selective monolayers. *ACS Energy Lett.* **7**, 966–974 (2022).
- Wei, M. Y. et al. Combining efficiency and stability in mixed tin–lead perovskite solar cells by capping grains with an ultrathin 2D layer. *Adv. Mater.* **32**, 1907058 (2020).
- Li, Z. Z. et al. Thermodynamic stability landscape of halide double perovskites via high-throughput computing and machine learning. *Adv. Funct. Mater.* **29**, 1807280 (2019).
- Ramprasad, R. et al. Machine learning in materials informatics: recent applications and prospects. *npj Computational Mater.* **3**, 54 (2017).
- Liu, Y. et al. Materials discovery and design using machine learning. *J. Mater. Res.* **3**, 159–177 (2017).
- Schleder, G. R. et al. From DFT to machine learning: recent approaches to materials science—a review. *J. Phys. Mater.* **2**, 032001 (2019).
- Liu, Z. et al. Computational functionality-driven design of semiconductors for optoelectronic applications. *InfoMat* **2**, 879–904 (2020).
- Zhao, X. G. et al. JAMIP: an artificial-intelligence aided data-driven infrastructure for computational materials informatics. *Sci. Bull.* **66**, 1973–1985 (2021).
- Lopez, S. A. et al. Design principles and top non-fullerene acceptor candidates for organic photovoltaics. *Joule* **1**, 857–870 (2017).
- Nagasawa, S., Al-Naamani, E. & Saeki, A. Computer-aided screening of conjugated polymers for organic solar cell: classification by random forest. *J. Phys. Chem. Lett.* **9**, 2639–2646 (2018).
- Sun, W. B. et al. The use of deep learning to fast evaluate organic photovoltaic materials. *Adv. Theory Simul.* **2**, 1800116 (2019).
- Sahu, H. et al. Toward predicting efficiency of organic solar cells via machine learning and improved descriptors. *Adv. Energy Mater.* **8**, 1801032 (2018).
- David, T. W. et al. Enhancing the stability of organic photovoltaics through machine learning. *Nano Energy* **78**, 105342 (2020).
- Odabasi, Ç. & Yildirim, R. Performance analysis of perovskite solar cells in 2013–2018 using machine-learning tools. *Nano Energy* **56**, 770–791 (2019).

37. Li, J. X. et al. Predictions and strategies learned from machine learning to develop high-performing perovskite solar cells. *Adv. Energy Mater.* **9**, 1901891 (2019).
38. Sani, F. et al. Advancement on lead-free organic-inorganic halide perovskite solar cells: a review. *Materials* **11**, 1008 (2018).
39. Pedregosa, F. et al. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
40. Abadi, M. et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. (2016). <https://arxiv.org/abs/1603.04467v1>.
41. Nogueira, F. Bayesian Optimization: open source constrained global optimization tool for Python. (2014). <https://github.com/fmfr/BayesianOptimization>.
42. Fortin, F. A. et al. DEAP: evolutionary algorithms made easy. *J. Mach. Learn. Res.* **13**, 2171–2175 (2012).
43. Shi, L. et al. Using data mining to search for perovskite materials with higher specific surface area. *J. Chem. Inf. Modeling* **58**, 2420–2427 (2018).
44. Lundberg, S. M. & Lee, S. I. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* 4765–4774 (Long Beach: MIT Press, 2017).
45. Bergstra, J., Yamins, D. & Cox, D. D. Hyperopt: a python library for optimizing the hyper parameters of machine learning algorithms. In *Proceedings of the 12th Python in Science Conference* 13–20 (Austin, Texas: SciPy Organizers, 2013).
46. Nayak, P. K. et al. Photovoltaic efficiency limits and material disorder. *Energy Environ. Sci.* **5**, 6022–6039 (2012).
47. Klug, M. T. et al. Metal composition influences optoelectronic quality in mixed-metal lead–tin triiodide perovskite solar absorbers. *Energy Environ. Sci.* **13**, 1776–1787 (2020).
48. Xing, G. C. et al. Long-range balanced electron- and hole-transport lengths in organic-inorganic $\text{CH}_3\text{NH}_3\text{PbI}_3$. *Science* **342**, 344–347 (2013).
49. Seo, J. Y. et al. Novel p-dopant toward highly efficient and stable perovskite solar cells. *Energy Environ. Sci.* **11**, 2985–2992 (2018).
50. Jeong, M. et al. Stable perovskite solar cells with efficiency exceeding 24.8% and 0.3-V voltage loss. *Science* **369**, 1615–1620 (2020).
51. Zong, Y. X. et al. Homogenous alloys of formamidinium lead triiodide and cesium tin triiodide for efficient ideal-bandgap perovskite solar cells. *Angew. Chem. Int. Ed.* **56**, 12658–12662 (2017).
52. Zong, Y. X. et al. Lewis-adduct mediated grain-boundary functionalization for efficient ideal-bandgap perovskite solar cells with superior stability. *Adv. Energy Mater.* **8**, 1800997 (2018).
53. Liu, H. et al. Modulated crystallization and reduced V_{OC} deficit of mixed lead-tin perovskite solar cells with antioxidant caffeic acid. *ACS Energy Lett.* **6**, 2907–2916 (2021).
54. Zhang, L. et al. Grain boundary passivation with dion-jacobson phase perovskites for high-performance Pb–Sn mixed narrow-bandgap perovskite solar cells. *Sol. RRL* **5**, 2000681 (2021).
55. Zhang, L. et al. Surface defect passivation of Pb–Sn-alloyed perovskite film by 1,3-propanediammonium iodide toward high-performance photovoltaic devices. *Sol. RRL* **5**, 2100299 (2021).
56. Gómez, P. et al. Pyrene-based small-molecular hole transport layers for efficient and stable narrow-bandgap perovskite solar cells. *Sol. RRL* **5**, 2100454 (2021).