EasyChair Preprint

№ 9440

Bandgap Prediction of Perovskite Solar Cell Using Multiple Regression Model Towards Higher Efficiency

Debmalya Sadhu, Devansh Dattatreya, Arjun Deo and Debasis De

December 11, 2022

Bandgap Prediction of Perovskite Solar Cell using Multiple Regression Model Towards Higher Efficiency

Debmalya Sadhu, Devansh Dattatreya, Arjun Deo, Debasis De*
Energy Institute, Bengaluru (Centre of Rajiv Gandhi Institute of Petroleum Technology)
Vidya Nagar, Bengaluru 562157, Karnataka
Corresponding author: debasisd@rgipt.ac.in

Abstract. Perovskite solar cells have been emerged as most promising third generation solar cell technology. In past few years, the efficiency of PSCs has increased drastically from 3.8% to 25.6% for lab scale devices in single junction architecture. The conventional way to develop materials is generally based on trial and error, continuous synthesis methods which are time consuming and costly. This motivates the use of autonomous experimentation toolkits like linear or multiple regression (MR) or various machine learning algorithms. A dataset containing 100 plus data points are collected from various published papers and analyzed using multiple regression algorithm in excel. The multiple regression (MR) model is applied for predicting the bandgap of the perovskite with the formula $Cs_aFA_bMA_{(1-a-b)}Pb(Cl_xBr_yI_{(1-x-y)})_3$ which takes into account the compositional engineering of cations and halide anions for predicting the optimum bandgap. The bandgap predicted by this model has a *R* square error of 0.96 which indicates the power of this model in prediction of the bandgap from their constituent material.

Keywords: Multiple Regression, Bandgap Optimization, Compositional Tuning, High Efficiency Perovskite Solar Cell

Introduction

Gustav Rose in 1839 had discovered perovskite for the first time in the Ural Mountains of Russia. Historically, oxide-based perovskite was most actively studied perovskite family as they exhibit superior ferroelectric, magnetic and superconductive properties. $CsPbX_3$ is the first halide-based perovskite structure identified by Moller in 1958. Organic Methylammonium (MA) appeared in halide perovskite family as cation, was first reported by Weber et al in 1978[1]. The solar-cells are categorized into three-generations where first-generation and second generation solar-cells are mainly based on wafer, thin film respectively, whereas the third-generation solar cells employ organic structures. Silicon based Tandem solar cells have achieved a large power conversion-efficiency (PCE) of 25–26% in recent few years. Nevertheless, this new variety of third generation solar-cells are termed as perovskite solar-cells which are an alternative for silicon based solar-cells having ability to exhibit PCE of 22.1%[2].

Hybrid organic-inorganic perovskite (HOIP) shows unique optical, electrical & optoelectronic performance in many applications. The properties like high efficiency, low temperature solution-based fabrication method, high absorption coefficient, high open circuit voltage and high diffusion length make the PSCs superior to other types of solar cells. The bandgap of perovskite materials plays an important role in light harvesting capability which governs the performance of different optoelectronic devices such as photodetectors, tandem solar cells, LEDs etc.[3,4].

The value of Tolerance factor (t) lies between 0.85 to 1.11 for halide perovskite and 0.89 to 1 for Oxide perovskites[5]. First 3D HOIPs discovered by replacing Cs in $CsPbX_3$ with MA cations by Weber. $CH_3NH_3PbI_3$ is most common materials for making high efficiency PSCs with a band gap of 1.55eV with absorption coefficient 10^4-10^5 cm^{-1}[6].

Compositional mapping of perovskites is required to extend the absorption to longer wavelengths without sacrificing α. Keeping α constant, we can use it under UV range or Visible range or Near IR range. The band gap of PSCs can be tuned from 1.55eV to 1.17eV by replacing Lead (Pb) to Tin (Sn)[7,8]. So, at 1.55eV bandgap the wavelength, the PSC can absorb is up to 800 nm. By changing the bandgap from 1.55 eV to 1.17 eV the absorption of incident wavelength can be varied up to near IR range(700nm) along with visible range(400-800nm).

To fabricate these types of HOIPs with the required bandgap the conventional way is trial and error-based experiments that requires high cost, lots of time, and materials. Meanwhile, though triple halide strategy with partial Br replaced by Cl is proved to be effective in achieving wide bandgap with reduced Br fraction, it faces challenges in doping Cl in the crystal lattice[9,10]. As Cl typically volatizes as MACl or FACl during annealing of the perovskite film[11]. Hence this limits the screening of the intrinsic elemental information of these materials.

To overcome this, Density Function Theory (DFT), Molecular Dynamics (MD) are introduced for exploring the relationship between the structural, elemental, and technical descriptors & performance parameters. Still, it would be a chronophagous and inefficient way to find stable and high-performance perovskites simply by performing experiments or DFT calculations. Based on many existing literature and computational data, modern algorithms like linear regression, multiple regression and artificial intelligence have gradually played a major role in Perovskite discovery[12,13,14].

Methods

Here we searched for the different literatures reporting the bandgap of the perovskites for building the dataset from web up to the end of 2021. Only Pb-based perovskites are considered, and Sn-based perovskites are excluded since Sn is not stable as it makes oxide on open atmosphere. Then the datasets are cleaned by removing the duplicate data points with same material composition and bandgap values. Finally, a dataset including 100 datapoints are extracted which cover Cl, Cl-Br mixed, Br, Br-I mixed, and I based MA, FA, and Cs. In the dataset $MAPbCl_3$ has the maximum bandgap of 3.16eV while the minimum value is 1.48eV for $FAPbI_3$[15]. The prepared dataset is shown in table 1 where 10 datapoints are shown.

Table 1. Dataset prepared from different experimental literatures.

Sample Id	MA	FA	Cs	Cl	Br	I	Bandgap (eV)	Ref.
01	1.00	0.00	0.00	0.00	0.00	1.00	1.60	[16]

02	0.70	0.20	0.10	0.00	0.00	1.00	1.56	[17]
03	0.50	0.40	0.10	0.00	0.00	1.00	1.54	[16]
04	0.40	0.60	0.00	0.00	1.00	0.00	2.20	[18]
05	0.25	0.75	0.00	0.00	0.00	1.00	1.55	[16]
06	0.20	0.70	0.10	0.00	0.00	1.00	1.49	[16]
07	0.17	0.83	0.00	0.00	0.00	1.00	1.53	[16]
08	0.00	0.80	0.20	0.00	0.18	0.82	1.66	[16]
09	0.00	0.83	0.17	0.00	0.17	0.83	1.63	[19]
10	0.00	0.85	0.15	0.00	0.15	0.85	1.53	[20]

After preparing the error free dataset a correlation between all the independent variables and dependent variable is established by calculating the Pearson Coefficient(r). Accordingly, from fig 1 it can be said MA, Cl and Br have the highest influence over the bandgap as the r values are 0.46, 0.88 and 0.44 respectively. More the r value near to 1 the parameters are more closely related. So other variables like FA, Cs and Br don't have much influence over bandgap.

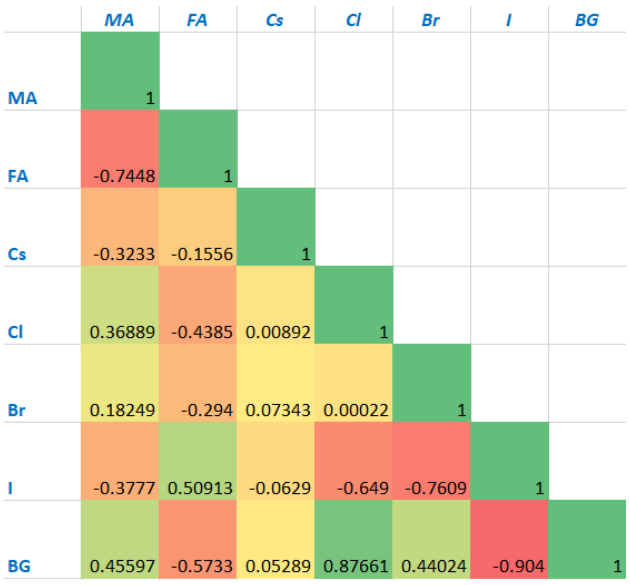

Fig 1. Relation between the influence of anions and cations over bandgap

90% of the total data points are used for training purpose and rest 10% for the testing. After applying the multiple regression algorithm in excel on the 90% training data, a fixed intercept (c) value of 6.42639 and the slope(m) for each composition of our perovskite solar cell e.g., 0.013 for MA or -0.084 for FA is obtained. So, the bandgap can be predicted by the formula

$$y = (m_1x_1+m_2x_2+m_3x_3+m_4x_4+m_5x_5+m_6x_6) + c \quad (1)$$

where, m_n and x_n represent the slope and percentage compositions of MA, FA, Cs, Cl, Br, I, respectively and y represent the bandgap.

Using this multiple regression algorithm, the bandgap of any perovskite having formula $Cs_aFA_bMA_{(1-a-b)}$ Pb $(Cl_xBr_yI_{(1-x-y)})_3$ can be predicted. A relative comparison is done between the actual bandgap and the predicted bandgap in fig 2. Some residual values are found for each bandgap reading, in some cases the value is high or sometimes it is negligible. This ununiform behavior of residuals is mainly due to the error of the algorithm itself and some noises present in the dataset. Noises refers to the data repetition, duplicate data, wrong data etc. To reduce this error, the residual plots shown in fig 3 are considered and accordingly refined the dataset.

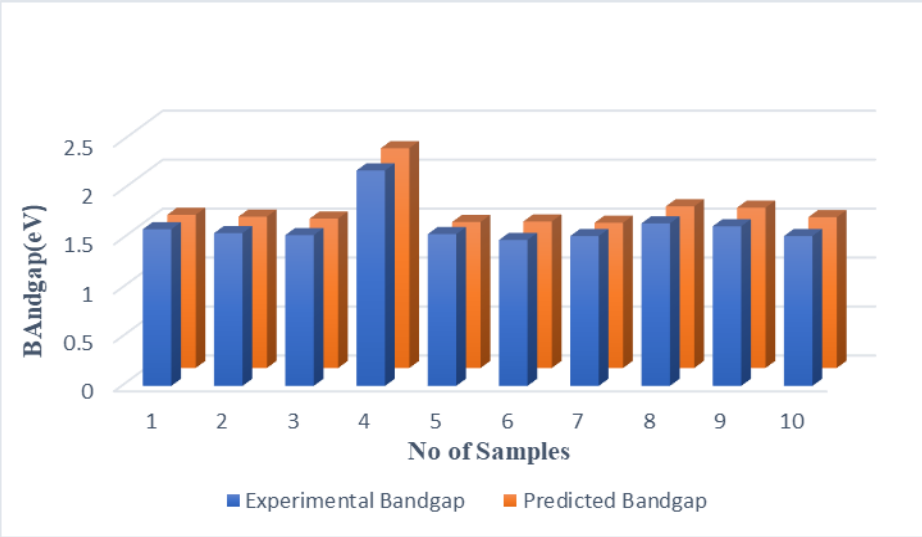

Fig. 2. Comparison between the predicted bandgap and the experimental bandgap from MR algorithm

Once it applied MR algorithm, the residual plots for each cationic and anionic compositions are obtained shown in fig 3. The percentage amount of every cationic and anionic compositions is extracted from the curves for which the residuals are minimum.

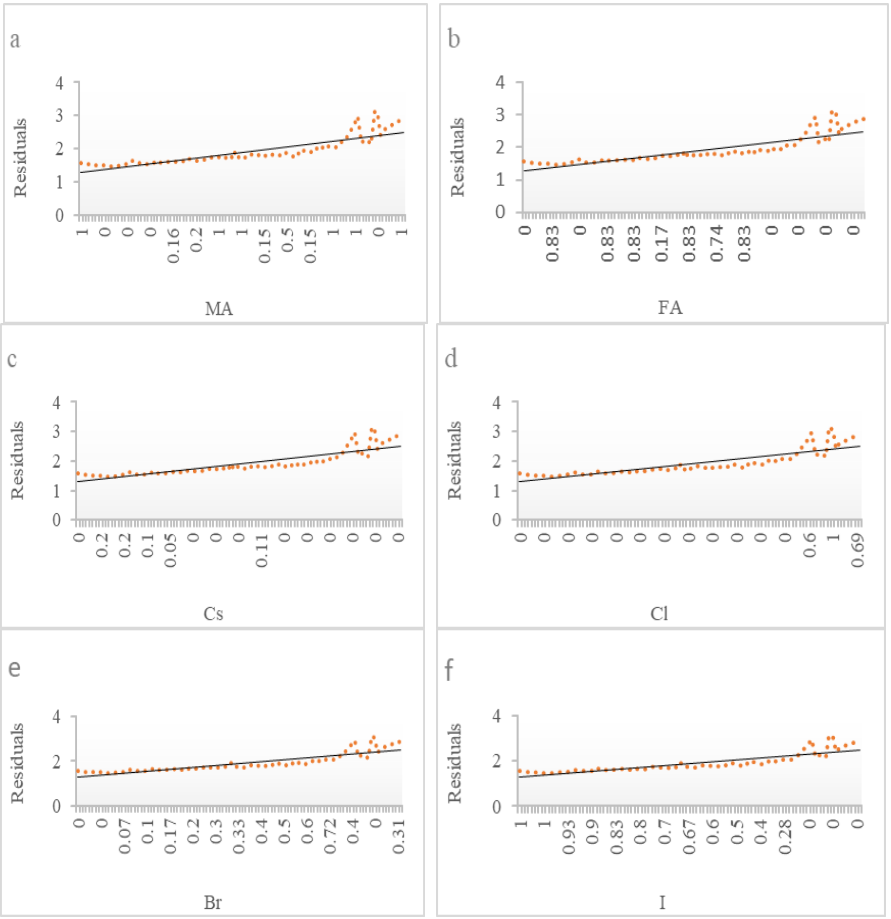

Fig. 3. Comparison of the residuals for each cationic and anionic compositions- (a) MA (b) FA (c) Cs (d) Cl (e) Br (f) I

A fresh dataset of compositions is prepared where the residuals are minimum. The newly generated dataset from the residual plots is shown in table 2. With the obtained equation (1) the bandgaps of newly generated datasets are predicted and then the optimum bandgap is finalized.

Table 2. Dataset from the residual plots of multiple regression algorithm.

Sample Id	MA	FA	Cs	Cl	Br	I	Predicted Bandgap (eV)
01	0.15	0.75	0.10	0.10	0.05	0.85	1.683
02	0.11	0.79	0.10	0.00	0.06	0.94	1.528
03	0.15	0.75	0.10	0.10	0.00	0.90	1.641
04	0.00	0.80	0.20	0.00	0.10	0.90	1.555
05	0.16	0.84	0.00	0.00	0.10	0.90	1.550
06	0.00	0.83	0.17	0.00	0.06	0.94	1.525

Table 3 indicates the best perovskites having bandgap 1.52 eV to 1.68 eV. These are generated from the residual plots of the MR algorithm. Among these the bandgap of 1.55 eV is considered as optimum bandgap as it is more focused on absorption of visible region wavelengths. So, PSCs like $CsFAMAPb(BrI)_3$, $FAMAPb(BrI)_3$ are selected as they have exact bandgap of 1.55eV.

Table 3. Obtained PSCs from residual plots with bandgap ranging from 1.52 eV to 1.68 eV

Sample ID	Obtained PSCs
01	$CsFAMAPb(ClBrI)_3$
02	$CsFAMAPb(BrI)_3$
03	$CsFAMAPb(ClI)_3$
04	$CsFAPb(BrI)_3$
05	$FAMAPb(BrI)_3$

The multiple regression model gives *R* square error of 0.959 ≈ 0.96. The *R* square value suggest the coefficient of determination of a model. These indicate the accuracy of the model is 96%. The Root Mean Square Error (RMSE) is calculated by the formula

$$rmse = \sqrt{[\sum_{i=1}^{n}[y(i) - \hat{y}(i)]]/n} \quad (2)$$

Where $y(i)$ is predicted value and $\hat{y}(i)$ is actual value which gives a value of 0.083 is shown in table 4. It determines the dataset prepared is with less error.

Table 4. Calculation of RMSE value from Actual vs Predicted Bandgap

Actual Bandgap	Predicted Bandgap	Residuals	Squared Error	RMSE
1.600	1.564	0.036	0.001277213	0.083988
1.560	1.545	0.015	0.000216058	
1.540	1.525	0.015	0.000200708	
1.520	1.561	-0.041	001.522E-05	
1.550	1.490	0.060	0.003450862	
1.490	1.496	-0.060	4.39632E-05	
1.530	1.483	0.047	0.002165165	

Conclusions

MR algorithm has reduced the time and material cost in experiment significantly to explore and design new HOIPs. It shows high precision in predicting the bandgap of perovskites from their elemental properties. Here we show how cationic or anionic compositional tuning lead to vary the bandgaps of PSCs. Along with that new PSCs having bandgap from 1.52 eV to 1.68 eV can be predicted which are not in the experimental literature. The MR model gives a *R* square error of 0.96 and RMSE value of 0.08 according to the dataset we fed to it. Also, the r value between experimental and predicted bandgap is calculated to be 0.979 which indicates a strong relation between them. These optimized bandgaps predicted by MR algorithm leads to the roads of high efficiency PSCs.

References

1. Suresh KumaraK.Chandra Babu Naidu. “A review on perovskite solar cells (PSCs), materials and applications.” Science Direct. Journal of Materiomics. 2021.7(5) 940-941.
2. Valerio Campanari, Faustino Martelli, Antonio Agresti, Sara Pescetelli, Narges Yaghoobi Nia, Francesco Di Giacomo, Daniele Catone, Patrick O’Keeffe, Stefano Turchini, Bowen Yang, Jiajia Suo, Anders Hagfeldt, and Aldo Di Carlo. “Re-evaluation of Photoluminescence Intensity as an Indicator of Efficiency in Perovskite Solar Cells” RRL Solar. 2022. 1-2.
3. Jost, L., Korte, L& Albrecht, S. Monolithic Perovskite tandem solar cells: a review of present status and advanced characterization methods towards 30% efficiency. Adv. Energy Mater. 2020,10, 1904102.
4. Tan, Z.-K. et al. Bright light-emitting diodes based on organometal halide perovskite. Nat. Nanotechnol. 9, 687-692 (2014).

5. Hua Ke. “The ideal Cubic Crystal structure of Perovskite materials often gets distorted or tilted into less symmetric tetragonal or orthorhombic structure.” Science Direct. Hybrid Perovskite Composite Materials. 2021.203-224.
6. Bowen, Yang Jiajia Suo, Anders Hagfeldt. “Interfacial Passivation Treatment towards High-efficiency and Operational Stable Perovskite Solar Cells.” nanoge.org. 2022. 1-2.
7. Jinhui Tong, Qi Jiang, Fei Zhang, Seok Beom Kang, Dong Hoe Kim, * and Kai Zhu. “Wide Bandgap Metal Halide Perovskites for Tandem Solar Cells”. ACS Energy Lett. 2021, 6, 1, 232–248.
8. Sneha a. Kulkarni et al. “Band-gap tuning of lead halide perovskites using a sequential deposition process.” In: Journal of Materials Chemistry A 2.24 2014, p. 9221.
9. M.kim, G-h. Kim T.K Lee et al, Methylammonium Chloride Induces Intermediate Phase Stabilization for Efficient Perovskite Solar Cells, joule, 2019,3,2179-2192.
10. C.Zuo, L.ding, An 80.11% FF record achieved for perovskite solar cells by using the NH_4Cl additive,Nanoscale, 2014, 6, 9935-9938
11. J.Jeong, H-B. Kim, H.Kim et al, Ternary Halide Perovskites for Highly Efficient Solution-Processed Hybrid Solar Cells, ACS energy Lett., 2016, 1, 712-718.
12. Qiuling Tao, Pengcheng Xu, Minjie Li, Wencong Lu. Machine Learning for Perovskite materials design and discovery. Npj Computational Materials (2021).
13. Meghna Srivastava, John M. Howard, Tao Gong, Mariama Rebello Sousa Dias, and Marina S. Leite. Machine Learning Roadmap for Perovskite Photovoltaics. J. Phys. Chem. Lett.2021, 12, 7886-7877.
14. B.Yilmaz, C. Odabasi and Ramazan Yildirim, Efficiency and Stability Analysis of 2D/3D Perovskite Solar Cells Using Machine Learning. Energy Technol. 2022, 10, 2100948.
15. D.Luo, R Su, W. Zhang, Q.Gong, and R.Zhu, Nat.Rev. Minimizing non-radiative recombination losses in perovskite solar cells, Nature Review Materials, 2020,5,44-60.
16. Herlina Arianita Dewi, Jia Li,a Enkhtur Erdenebileg, Hao Wang, Michele De Bastiani, Stefaan De Wolf, Nripan Mathews, Subodh Mhaisalkar and Annalisa Bruno. Efficient bandgap widening in co-evaporated MAPbI3 perovskite. RSC, Sustainable Energy & Fuels, 2022,issue 10.
17. Shaofu, Wang,Junjun Jin,Yuyang Qi,Pei Liu,Yu Xia,Yun Jiang,Rong-Xiang He,Bolei Chen,Yumin Liu,Xing-Zhong Zhao. δ-$CsPbI_3$ Intermediate Phase Growth Assisted Sequential Deposition Boosts Stable and High-Efficiency Triple Cation Perovskite Solar Cells. Advanced Functional Materials. Vol. 30, Feb 2020, 1908343.
18. B.Slimi, M Mollar, B. Mari and R. Choturou. Thin Film of Perovskite (Mixed-Cation of Lead Bromide $FA_{1-x}MA_xPbBr$) Obtained by One-Step Method. Journal of Electronic Materials. 48, 2019, 8014–8023.
19. Steffen Braunger, Laura E. Mundt, Christian M. Wolff, Mathias Mews, Carolin Rehermann, Marko Jošt, Alvaro Tejada, David Eisenhauer, Christiane Becker, Jorge Andres Guerra, Eva Unger, Lars Korte, Dieter Neher, Martin C. Schubert, Bernd Rech, and Steve Albrecht. $Cs_xFA_{(1-x)}Pb(I_{(1-y)}Br_y)_3$ Perovskite Compositions: The Appearance of Wrinkled Morphology and its Impact on Solar Cell Performance . J. Phys. Chem. C 2018, 122,17123–17135.
20. Shengquan Fu,Yueyue Xiao, Xinxin Yu, Tianxing Xiang, Fei Long, Junyan Xiao,a Zhiliang Ku, Jie Zhong,Wei Li,a Fuzhi Huang, Yong Peng and Yibing Cheng. Bandgap adjustment assisted preparation of >18% $Cs_yFA_{(1-y)}Pb(I_xBr_{(1-x)})_3$-based perovskite solar cells using a hybrid spraying process. RSC Advances. 2021, issue 29.