Computation of photovoltaic and stability properties of hybrid organic-inorganic perovskites via convolutional neural networks

Victor Alexander Aristizabal-Ferreira[a], José Manuel Guevara-Vela[a], Arturo Sauza-de la Vega[b], Ángel Martín Pendás[c], Gibran Fuentes-Pineda[d], Tomás Rocha-Rinza[a,*]

[a]*Instituto de Química, Universidad Nacional Autónoma de México, Circuito Exterior, C. U., Coyoacán, 04510, Ciudad de México, México.*

[b]*Department of Chemistry, Pritzker School of Molecular Engineering, James Franck Institute, Chicago Center for Theoretical Chemistry, University of Chicago, Chicago, Illinois 60637, United States*

[c]*Facultad de Química, Universidad de Oviedo, Av. Julián Clavería, 8, 33006 Oviedo, Asturias, España.*

[d]*Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Circuto Escolar 3000, C.U., Coyoacán, 04510, Ciudad de México, México.*

Abstract

Hybrid Organic Inorganic Perovskites (HOIPs) have gained considerable interest due to their potential applications as photovoltaic materials. Nevertheless, several issues have to be solved on this matter, such as the proper tuning of band gaps and those concerning stability, before these systems can realise their full potential. Here, we used deep learning techniques, more specifically Crystal Graph Neural Networks (Xie & Grossman, *Phys. Rev. Let.*, **2018**, 120), abbreviated as CGNN, to explore the chemical space of HOIPs and to address the above mentioned difficulties. We trained this CGNN with a data set comprised of $1,346$ density functional theory calculations and used it to compute band gaps, refractive indexes, atomisation energies, volumes of unit cells and volumetric densities of $3,840$ HOIPs. Our screening method permits a rapid selection of perovskites with suitable optoelectronic properties and only 7 have an adequate band gap to be used in photovoltaic technologies. The composition, ABX_3, of such perovskites is mainly of small molecular cations such as A = $[NH_4]^+$, $[NH_2NH_3]^+$ together with $[OHNH_3]^+$, B = In^{2+}, Zr^{2+} along with Sn^{2+}, and X = I^-. The consideration of further systems indicates that the occurrence of phosphorus and sulphur in the molecular cation diminishes strongly the band gap of the perovskite. We also considered the stability of the systems with optimal band gaps with

*To whom correspondence should be addressed: trocha@iquimica.com.mx

respect to their degradation in simple organic and inorganic salts. Overall, our investigation shows how deep learning techniques can be exploited to achieve a rapid screening of potential photovoltaic materials in terms of their electronic properties and stability.
Keywords:
Perovskites, Deep learning, Crystal Graph Neural Networks, Photovoltaic materials, Band gaps.

1. Introduction

Climate change is among the most important and urgent challenges that humanity is facing in this century. We require plenty of strategies to address this threat. According to the Paris agreement, [1] we need more efficient appliances, CO_2 capture technology, and the decarbonisation of our energy grid through the use of sustainable sources of energy, [2] in order to keep global warming below 2 °C. There is certainly a wide variety of renewable energies: geothermal, on- and offshore windmills, biomass, and solar cells. [3] The last-mentioned technology is particularly attractive because Earth receives a vast amount of energy from the Sun. Indeed, the luminous energy collected by Earth from the Sun in one hour equals the global energy requirements for a year[4, 5]. Furthermore, the use of photovoltaic cells to convert sunlight directly into electricity, is perhaps the renewable energy technology for which most progress has been made in the last decade. Due partly to (i) new developments in materials and (ii) economies of scale, the cost to produce solar energy has decreased in almost 90% since 2010. [6]

This remarkable progress has been boosted via investments not only in new manufacturing processes but also in basic science. There has recently been a very intense research activity dedicated to the discovery of new materials as light-harvesting compounds within solar cells that offer high efficiency at reasonable costs. Among the most promising systems that fulfil these requirements are perovskite-structured materials, i.e., compounds with the crystal structure of calcium titanium oxide $CaTiO_3$. More specifically, the perovskites which are considered for applications in photovoltaics are Hybrid Organic Inorganic Perovskites (HOIPs) with the ABX_3 formula where A is an organic cation, B a metal cation and X a halide anion. [7, 8] On top of being economically feasible and environmentally friendly materials, this type of perovskites are

susceptible to finely tune their band gap and other optoelectronic properties via changes of its composition. [9] This circumstance makes perovskites particularly attractive as potential photovoltaic materials. However, the enormity of the resulting chemical space renders impractical the investigation of all possible perovskites with traditional experimental methodologies.

In the last 10 years, the application of artificial intelligence has emerged as an alternative to solve very complex problems in science and engineering. Machine Learning (ML), a field of artificial intelligence, has found success in tackling diverse problems in physical chemistry such as the construction of better force fields, [10, 11], the computation of correlation energy [12, 13] along with the prediction of chemical reactions, [14–16] and catalytic activity [17, 18]. However, the most important application of ML in chemistry is perhaps the large-scale exploration of chemical space, which in turn allows for the discovery of new and exciting compounds and materials. [19–21] Hereof, one can find in the literature a number of investigations which exploit ML techniques for the study of perovskites. Some of these reports address the prediction of (i) HOIPs which can be feasibly synthesised [22, 23] (ii) low-dimension perovskites with iodine and lead [24], (iii) the formation energy of double perovskites with the oxide anion as a counterion [25], (iv) resistance to intrinsic dielectric breakdown [26, 27] and (v) cell parameters of double perovskites [28] among other applications. More specifically for this investigation, the band gap relates directly to the efficiency of opto-electronic devices. There are several papers which make use of ML techniques to predict band gaps [29–35]. Nevertheless, the investigations addressing HOIPs on this matter are scarce. For example, Shuaihua et al.[30] utilised ML to predict band gaps in HOIPs. These workers considered, however, metal centres whose oxidation states can hardly be +2, e.g., Si^{2+}, Sb^{2+} and Ti^{2+}. Other dissimilarities between our investigation and that in reference [30] is the use of Neural Networks (NN) as well as the consideration of elements in the third row in the molecular cations of the HOIPs and the stability degradation (*vide infra*). Likewise, Saidi et al.[35] used ML procedures for the computation of band gaps of HOIPs. An important difference of this work with this investigation is that we considered a far larger number of metallic centres and molecular cations which allows for a larger exploration of the chemical space of HOIPs.

Thus, we used machine learning techniques, in particular, artificial neural networks trained with results from density functional theory calculations [36] to explore the large chemical space of potential HOIPs. We considered a sizeable number of possible combinations between different ions to find the best candidate for its potential application in photovoltaic materials. We trained and tested a Crystal Graph Neural Network (CGNN) for the determination of properties of perovskites and then we used it in the study of $3,840$ HOIPs concerning its possible use as photovoltaics. Based on these results, we considered a further set of perovskites as prospective photovoltaic materials. We also addressed the stability of the HOIPs with optimal band gaps towards degradation reactions. This contribution shows how CGNNs can be exploited in the exploration of the chemical space for the acceleration of the discovery of perovskite-structured photovoltaics with tailored stability and electronic properties.

2. Methodology

Machine learning is an area of computer science and a branch of artificial intelligence whose objective is to develop techniques that allow to generalise behaviours from information supplied in the form of examples. Deep learning is a state-of-the-art subcategory of machine learning which deals with the use of NNs [37]. In this work, we use a convolutional neural network. This type of NN employs convolutions instead of general matrix multiplications in at least one of their layers, and it is specialised in the processing of data with topologies similar to those of a grid. Recently, Xie and Grossman [32] built a CGNN using graph representations that allow the use of crystal structures as input data. A graph in this context is a set of objects called vertices or nodes joined by links denoted as edges, which allow representing binary relationships between elements of a set. To create a crystal graph, the nodes are the atoms and the edges are the bonds between them. The whole system is represented by vectors and nodes of different lengths and characteristics, respectively.

These vectors are characterised by properties of the atoms such as group, period and block position in the periodic table, electronegativity, covalent radius, valence electrons, first ionisation energy, electronic affinity and atomic volume. Such feature vector is represented via one hot

encoding. Ditto for the interatomic distances taking a range between $0.7 - 5.2$ Å distributed in 10 categories. These procedures allow the consideration of constant-sized vectors as entries. Hereof, the pooling layer is used in the neural network to generate a fixed-size vector of characteristics that represents a crystal. The neural network is trained with each property separately, so that the optimised parameters are different for each case. This optimisation is achieved via stochastic gradient descent and backpropagation, to predict results comparable to those accurately calculated from Density Functional Theory (DFT). Indeed, the above mentioned CGNN yields errors which are comparable with those coming from electronic structure calculations for most of the desired properties of the studied perovskites while using training sets of manageable size for the training of the CGNN [32]. We address HOIPs with structures ABX_3 wherein A is a monovalent molecular cation, B is a divalent metallic cation and X is an halide anion (Figure 1 and Table 1). We optimised the structure of these perovskites, and computed band gaps, refractive indexes, atomisation energies, volumes of unit cells and volumetric densities with the above mentioned CGNN. The training of the exploited CGNN is detailed below.

Based on the predictions of the CGNN, we considered other systems with (i) A = $[PH_4]^+$, $[OHPH_3]^+$, $[NH_3PH_2]^+$, $[ClNH_3]^+$, $[ClCH_2NH_3]^+$, $[CH_3PH_3]^+$, $[C_2SH_5]^+$, $[C_3SH_7]^+$, $[C_4SH_9]^+$, $[(CH_3)_3S]^+$, $[CH_3SH_2]^+$, $[NH_3SH]^+$, $[OHSH_2]^+$, $[PH_3SH]^+$, $[SH_3C]^+$, along with $[SH_3]^+$ (ii) B = all the metallic cations reported in Table 1 and (iii) $X = I^-$ for which we carried a similar procedure, but we only determined band gaps. We also computed the energies for the decomposition of the perovskites with optimal band gaps according to the process,

$$ABX_{3(s)} \longrightarrow AX_{(s)} + BX_{2(s)},\ \Delta E_{\text{dissoc}}. \tag{1}$$

3. Computational details

3.1. Geometry optimisations

We optimised the crystal structure of 3, 840 different perovskites using DFT, in particular combining the PBE functional [38] with ad-hoc ultrasoft pseudopotentials, as implemented in

Table 1: Cations A and B in the structures of the perovskites shown in Figure 1.

32 molecular cations					
$[NH_4]^+$	$[CH_3C(NH_2)_2]^+$	$[NH_3OH]^+$	$[C_3H_4NS]^+$	$[CH_3NH_3]^+$	$[C_7H_6]^+$
$[NH_3NH_2]^+$	$[(CH)_4NH_2]^+$	$[(CH_2)_3NH_2]^+$	$[C(CH_3)_2CH_2NH_2]^+$	$[CH(NH_2)_2]^+$	$[(CH_3)_2(CH_2)_2NH_2]^+$
$[C_3N_2H_5]^+$	$[C(CH_5)NH_3]^+$	$[(CH_3)_2NH_2]^+$	$[C_2NH_6]^+$	$[NC_4H_8]^+$	$[C_2OH_5]^+$
$[CH_3CH_2NH_3]^+$	$[C_3OH_7]^+$	$([NH_2)_3C]^+$	$[C_4NH_{10}]^+$	$[(CH_3)_4N]^+$	$[C_4OH_9]^+$
$[(CH_3)_3NH]^+$	$[C_5NH_{12}]^+$	$[CH_3(CH_2)_2NH_3]^+$	$[C_5OH_{11}]^+$	$[(CH_3)_2CHNH_3]^+$	$[C_6NH_{14}]^+$
$[CH_3(CH_2)_3NH_3]^+$	$[C_6OH_{13}]^+$				

30 metallic centres											
Ag^{2+}	Os^{2+}	Ca^{2+}	Re^{2+}	Co^{2+}	Sc^{2+}	Cr^{2+}	Sn^{2+}	Cu^{2+}	Sr^{2+}	Fe^{2+}	Ta^{2+}
Ga^{2+}	Tc^{2+}	Ge^{2+}	Te^{2+}	Hf^{2+}	Ti^{2+}	In^{2+}	V^{2+}	Ir^{2+}	W^{2+}	Mg^{2+}	Y^{2+}
Mn^{2+}	Zn^{2+}	Mo^{2+}	Zr^{2+}	Nb^{2+}	Ni^{2+}						

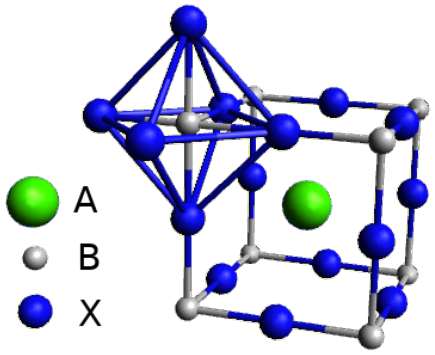

Figure 1: Structure of the unit cells of the hybrid organic-inorganic perovskites ABX_3 investigated herein. The green circle denotes a molecular cation A in the middle of the unit cell. The black circles represent metallic ions B in the vertices of the unit cell. The identity of the cations A and B are reported in Table 1. Finally, the purple circles indicate halide anions X^- (F^-, Cl^-, Br^- and I^-) in the midpoints of the edges of the unit cell.

the Quantum Espresso [39–41] suite of programs. The kinetic energy cutoff for wavefunction and charge density for these geometry optimisations were 16 and 200 Ry respectively. We took into account the combinations corresponding to 32 molecular cations, 30 metal cations and 4 halide anions for a total of 3, 840 data points as mentioned above. We considered the typical structure of HOIPs, ABX_3, in which A represents a monovalent molecular cation, B a divalent metallic ion and X a halide ion as shown in Figure 1 and Table 1. We started from initial orthorhombic crystal structures ($\alpha = \beta = \gamma = 90°$) but these angles were relaxed in geometry optimisations without restriction. We preprocessed the output data of Quantum Espresso with the aid of the software DiffPy[42–45] to get Cif files.

The structures for AX and BX_2 to be considered in Equation (1) were taken directly from the Materials Project website [46–49]. We performed an unrestricted geometry optimisation of AX and BX_2 in the same way that we did for ABX_3, to obtain the necessary electronic energies for the computation of ΔE_{dissoc}. Finally, we point out that the kinetic energy cutoffs for the wavefunction and charge dentsity for these electronic structure calculations in Quantum Espresso were 60 and 450 Ry respectively.

3.2. Training of the neural network

Xie and Grossman trained their CGNN by considering perovskites with the sites (i) A as a monovalent metallic cation, (ii) B as a divalent metallic cation and (iii) X as one of several elements from O, N, S, and F. We trained this CGNN with a database of perovskites wherein sites A and X are occupied by monovalent molecular cations and halide anions (F^-, Cl^-, Br^- or I^-) respectively. Our training employed 1, 346 HOIPs and their computed properties with density functional theory calculations.[36] This database comprised perovskites in different crystalline systems, formed by 16 organic cations in site A; the Ge, Pb and Sn divalent cations in position B and finally any of the halide anions in location X. As it is usual in machine learning protocols, we divided this database in three parts: training (60%), validation (20%) and testing (20%). The hyperparameters used in the CGNN are (i) the stochastic gradient descent to find the parameters that minimise the loss function; (ii) the learning rate (0.01); (iii) the number of hidden atom features in convolutional layers (64); (iv) the number of hidden features after pooling (128); (v) the number of convolutional layers (3); and (vi) the number of hidden layers after pooling (1).

4. Results and discussion

We present now the most important results of our investigation. We take first into account the training and the testing of the CGNN for the computation of the band gap, index of refraction, atomisation energy, volume of the unitary cell and volumetric density. Later, we consider the computation of these properties by the CGNN with the compositions indicated in Table 1. Finally, we present the examination of the stability for the perovskites with optimal band gaps as potential photovoltaic materials.

4.1. Training and testing of the crystal graph neural network

The data for the training and testing of the CGNN of Xie and Grossman were obtained via DFT calculations. The results for such training and testing for the computation of the above-mentioned properties are shown in Figures 2(a)–2(e).

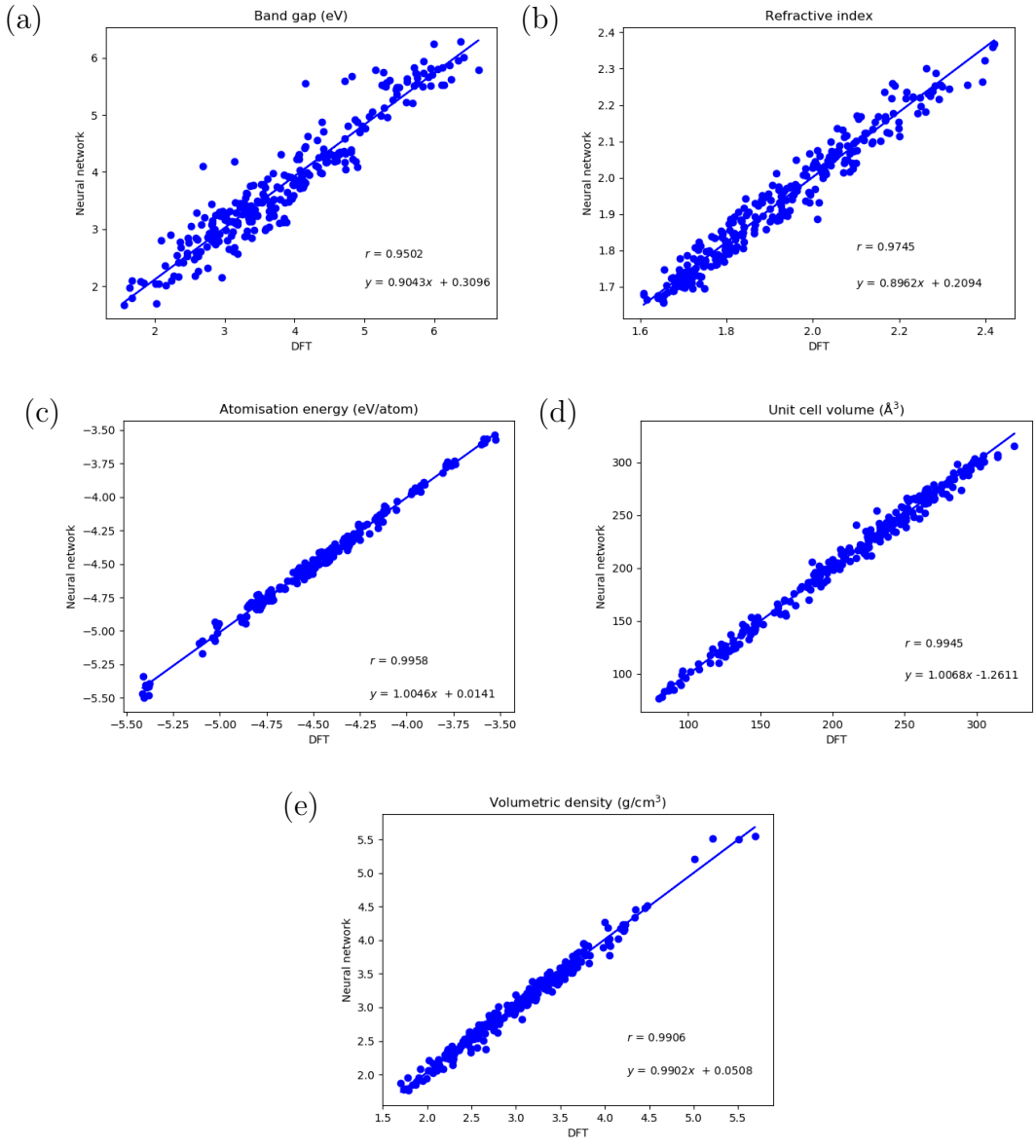

Figure 2: Correlation between the properties (a) band gap, (b) index of refraction, (c) atomisation energy, (d) volume of the unit cell and (e) volumetric density computed with (i) DFT and (ii) the crystal graph neural network exploited in this work.

The correlation coefficient, r, ideally equal to 1, provides information about the random error between the testing data and those computed by the CGNN. Besides, r is also a statistical measure of the reliability of the regression. Still, it is desirable to have other statistical parameters to assess different sources of error in the predictions of the CGNN. The systematic error can be appraised from the slope of the straight line which results from the correlation of the values computed with the employed CGNN and DFT. The optimum value of this slope is equal to 1. The intercept of this line should ideally equal zero and the corresponding deviations are indicators of a constant systematic error. Finally, the standard deviation around the line of regression, $S_{x/y}$, accounts for the dispersion of the data around such line and it gives an evaluation of the aleatory error in the method. Table 2 summarises the statistics of regression for the examined properties.

The parameters of Table 2 show that the CGNN performs well to predict the properties for which the training was performed, i.e., band gaps, indexes of refraction, atomisation energies, volumes of the unit cell and volumetric densities. In all cases, the correlation coefficient is larger than 0.95. This result indicates that the comparison of the data computed with DFT and with the CGNN are well-matched via a linear model and that they are highly correlated. Regarding the slope and the intercept, the obtained values are close to the ideal values, 1 and 0 respectively. These results indicate a good performance of the CGNN. Finally, in respect of the standard deviation, the value corresponding to the band gap is 8.6 % and those corresponding to the other properties are smaller than 3.0%, in all cases.

4.2. Computation of properties

The structures obtained as a result of the geometry optimisations of the unit cells with the program Quantum Espresso were used as input data for the CGNN trained with the properties shown in Figures 2(a)–2(e). The corresponding results are shown in Figures 3(a)–3(e) wherein one can observe the distribution of the data for each computed feature. Regarding the band gap (Figure 3(a)), the largest number of crystals are between 3.5 and 4.3 eV, in respect of the index of refraction (Figure 3(b)) in the range of 1.75–1.85, concerning the atomisation

Table 2: Statistical parameters for the regressions in Figures 2(a)–2(e).

Property	Correlation coefficient (r)	Slope	Intercept	Standard deviation ($S_{x/y}$)
Band gap	0.9502	0.9043	0.3096	0.326 (8.6%)
Index of refraction	0.9745	0.8962	0.2094	0.0365 (1.9%)
Atomisation energy	0.9958	1.0046	0.0141	0.0313 (0.7%)
Volume of the unit cell	0.9945	1.0068	−1.2611	6.2495 (3.0%)
Volumetric density	0.9906	0.9902	0.0508	0.0886 (2.9%)

energy (Figure 3(c)) in the interval −4.50 to −4.10 eV/atom. Finally, for the unit cell volume (Figure 3(d)) and the volumetric density (Figure 3(e)), most of the computed values fall within the 175–300 Å^3 and 2.6–3.5 g/cm^3 intervals, respectively.

These results are useful to examine the changes of every property when we modify one of the components of the crystal, either the organic moiety, the metallic centre or the halide anion, and we keep constant the other two constituents. For example, if we plot the band gap for the different metallic cations while maintaining unchanged the organic cation and the corresponding halide (Figure 4), we note that halides have different effects on the band gap, iodide/chloride anions generate the smallest/largest band gaps (Cl > F > Br >> I). We notice that fluorine and chlorine are exchanged in relation to their position in the periodic table, an observation which might warrant a posterior investigation. We also observe strong variations of this property when we change the metallic centre. We focused herein on the trends of the band gap values, due to its direct relationship with the performance of a photovoltaic device. As stated in Section 1, perovskites are promising materials to be used in solar cells due to their low production costs and the increase in efficiency that these materials have had in the last few years. The solar cells based on perovskites which represent the largest efficiency are of the tandem type, with two layers of photovoltaic materials with different band gaps.[50] For example, a tandem perovskite solar cell, including a sheet with a band gap equal to 1.75 eV can have an efficiency superior to a silicon-based photovoltaic device. This type of tandem photovoltaic cells require perovskites

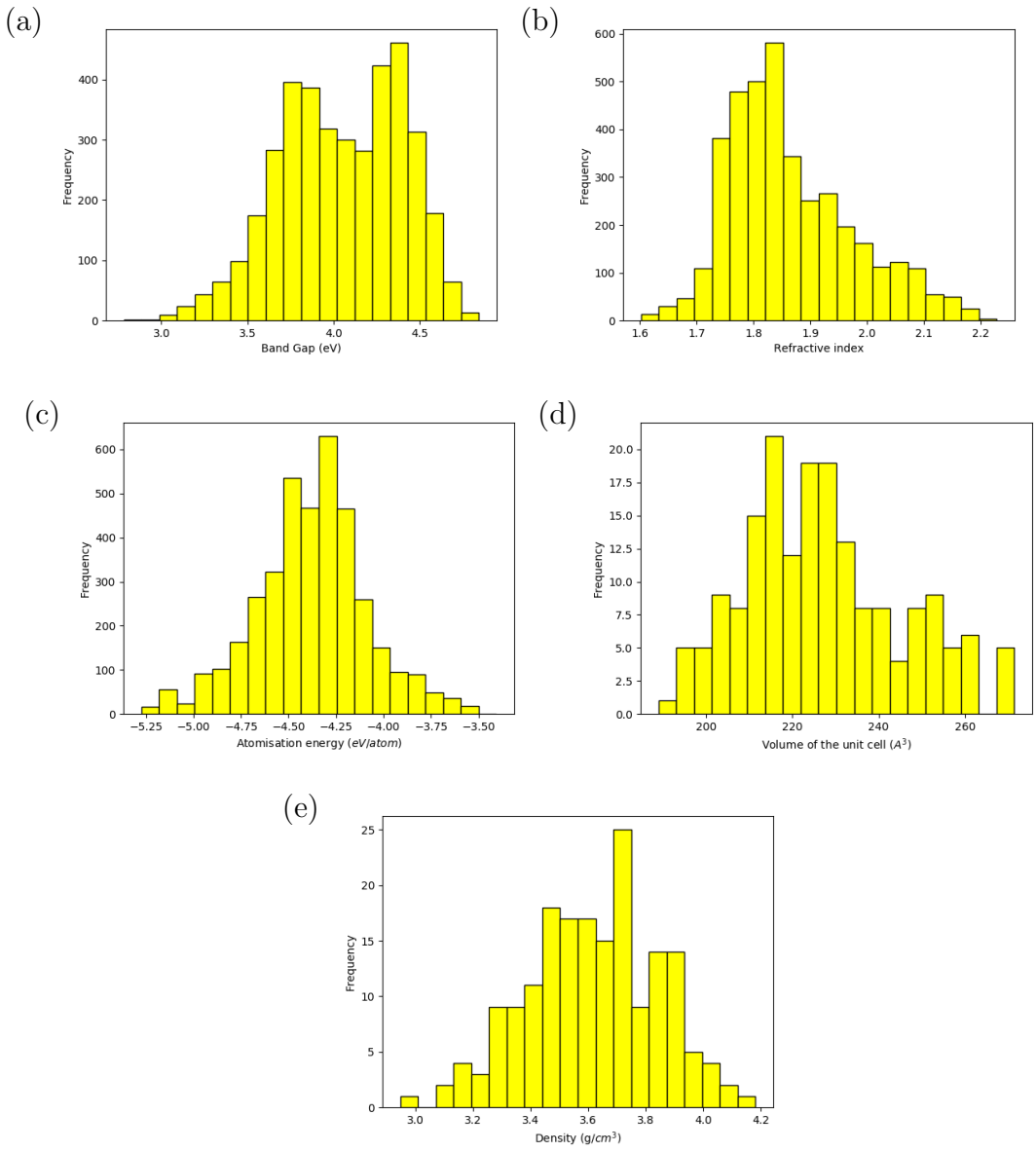

Figure 3: Distribution of the properties (a) band gap, (b) refractive index, (c) atomisation energy, (d) volume of the unit cells and (e) volumetric density for the perovskites schematised in Figure 1 and Table 1.

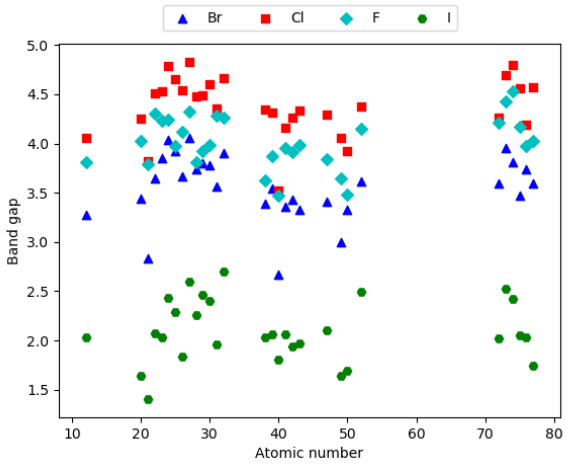

Figure 4: Band gap as a function of the atomic number of the metallic centre for the investigated hybrid organic inorganic perovskites with the molecular cation NH_4^+.

with band gaps between 0.9 and 1.2 eV and a frontal cell with a band gap between 1.7–1.9 eV.[50] Given the high values of the band gap of the perovskites computed in Figure 3(a), we are interested in determining the components which result in the smallest band gaps.

We noticed that the smallest molecular cations, e.g. NH_4^+, $[NH_2NH_3]^+$ and $[OHNH_3]^+$, are those for which the band gap is below 2.0 eV. Furthermore, Table 3 shows that the systems with the smallest band gaps have iodide as counterion as mentioned above,

Based on these results, we consider the small molecular cations A = $[PH_4]^+$, $[OHPH_3]^+$, $[NH_3PH_2]^+$, $[ClNH_3]^+$, $[ClCH_2NH_3]^+$, $[CH_3PH_3]^+$, $[C_2SH_5]^+$, $[C_3SH_7]^+$, $[C_4SH_9]^+$, $[(CH_3)_3S]^+$, $[CH_3SH_2]^+$, $[NH_3SH]^+$, $[OHSH_2]^+$, $[PH_3SH]^+$, $[SH_3C]^+$, along with $[SH_3]^+$ together with (i) all

Table 3: The 7 HOIPs based on ammonium cations with the band gaps within an optimum range to be used in solar cells.

Perovskite	Band gap (eV)	Perovskite	Band gap (eV)	Perovskite	Band gap (eV)
NH_4SnI_3	1.89	NH_4ScI_3	2.05	$OHNH_3SnI_3$	2.08
NH_4YI_3	1.90	NH_4CaI_3	2.06	$NH_2NH_3ZrI_3$	2.09
NH_4ZrI_3	2.02				

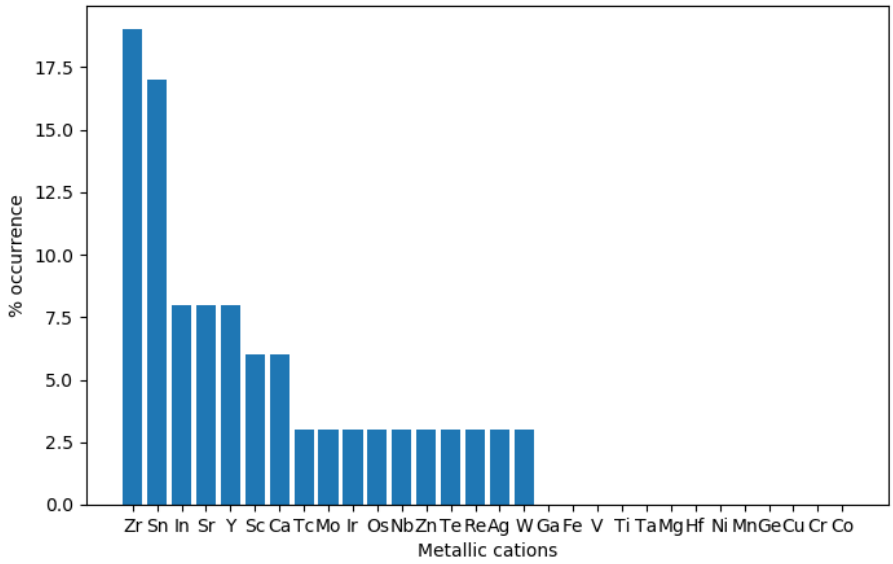

Figure 5: Percentage of occurrence of metallic cations in the examined perovskites with the band gaps in the suitable interval for their application as photovoltaic materials.

the metallic cations in Table 1 and (ii) $X^- = I^-$. We detected far more systems with the band gap in the required region for their application as photovoltaic materials, all of them containing phosphonium or sulphonium cations.

Finally, Figure 5 reports the percentage of occurrence of the investigated metallic cations in the perovskites whose band gap is below 2.0 eV. We note that the most frequent metal in such systems is zirconium (19%) and tin (17%) comes next. Indium, strontium and yttrium (8% each) are in third place. These five metals are included in sixty percent of the perovskites with suitable band gap intervals in their application as photovoltaics. The same figure shows metals which have the poorest performance in this regard, i.e., Ga, Fe, V, Ti, Ta, Mg, Hf, Ni, Mn, Ge, Cu, Cr, and Co. We acknowledge that among the systems with the determined smallest band gaps, some oxidation states will be easier to be synthesised such as Sn(II) as opposed to others like In(II). Nevertheless, we took into account the metallic cations for which the oxidation state II is still reasonable to be experimentally obtained.

Our results up to this point give directives to modulate the band gap of HOIPs for its use in photovoltaic devices. The band gap is reduced with (i) small molecular cations, (ii) the reduction of the electronegative character of the atom bearing the formal positive charge in the molecular cation and (iii) the occurrence of iodide as the halide anion. These results can partially be rationalised by considering the fact that ionicity generally increases the band gap of a material, a result which is well known in standard band theory. On the other hand, these observations also provoke questions regarding the relationship between chemical bonding and the band gap in the investigated systems. For example, the chloride anion is associated with the largest band gaps. We also observed a decrease of the band gap when there is a hard acid in the centre of the unit cell and a soft base in the corresponding edges. Nevertheless, we note an increase of the band gap when we have a softer molecular acid and a hard acid in the associated edges. We note how machine learning results can arise new scientific questions for which further investigation is required. Indeed, the studies which aim to relate band gap energies of HOIPs with their chemical bonding scenario are scarce. [51]

Table 4: The 29 HOIPs based on phosphonium and sulphonium cations with the band gaps within an optimum range to be used in solar cells.

Perovskite	Band gap (eV)	Perovskite	Band gap (eV)	Perovskite	Band gap (eV)
PH_4SnI_3	1.59	PH_4ScI_3	1.87	$OHPH_3SnI_3$	2.00
PH_4ReI_3	1.70	PH_4AgI_3	1.88	PH_4TcI_3	2.00
PH_4SrI_3	1.78	$OHPH_3ZrI_3$	1.91	PH_4OsI_3	2.02
PH_4InI_3	1.78	PH_4CaI_3	1.91	$NH_3PH_2ZnI_3$	2.04
$OHPH_3SrI_3$	1.80	PH_3SHInI_3	1.93	PH_3SHSnI_3	2.05
PH_4MoI_3	1.82	PH_4IrI_3	1.94	SH_3ZrI_3	2.06
PH_3SHZrI_3	1.82	$NH_3PH_2ZrI_3$	1.95	$OHPH_3YI_3$	2.07
PH_4ZrI_3	1.84	PH_4TeI_3	1.96	PH_4YI_3	2.09
PH_3SHSrI_3	1.85	SH_3SnI_3	1.97	$OHPH_3InI_3$	2.10
PH_4WI_3	1.86	PH_4NbI_3	1.97		

To conclude, we considered the stabilities of the investigated HOIPs towards its decompo-

sition in the associated ammonium/phosphonium and metallic halides as specified in equation (1). Table 5 reports those perovskites in Tables 3 and 4 for which $\Delta E_{\text{dissoc}} > 0$. We observe that most of the perovskites for the last-mentioned charts are energetically stable with respect to the degradation schematised in equation (1) to the extent that in some cases the values of ΔE_{dissoc} is in the order of tens of kcal/mol. We also noted that the most common molecular cations in Table 5 are based on phosphorus and sulphur, a fact which could give important directives for the production of more stable perovskites. We finally point out that to the best of our knowledge none of the systems in Tables 3–5 have been previously reported in the literature. We hope that those perovskites in Table 5 might be interesting prospective materials for HOIPs in photovoltaic devices.

5. Concluding remarks

We used deep learning techniques, more precisely a crystal graph convolutional neural network to explore the chemical space of hybrid organic-inorganic perovskites as potential photovoltaic materials. We used this neural network to compute distinct properties (band gaps, refraction indexes, atomisation energies, unit cell volumes and volumetric densities) of $3,840$ HOIPs. From these systems, our screening method yielded only 7 HOIPs that can be potentially used in photovoltaic materials. The exploration of further perovskites revealed that the use of phosphonium and sulphonium-based cations abates considerably the band gaps of the systems under study. We found that the systems with the most adequate band gaps have small molecular cations and

Table 5: Perovskites in Tables 3 and 4 for which their degradation into AX and BX_2 according to equation (1), is endothermic.

Perovskite	ΔE_{dissoc} (kcal/mol)	Perovskite	ΔE_{dissoc} (kcal/mol)	Perovskite	ΔE_{dissoc} (kcal/mol)
PH_4TeI_3	0.37	$OHPH_3InI_3$	6.50	SH_3SnI_3	9.97
PH_4SnI_3	0.79	$OHPH_3YI_3$	6.69	$OHPH_3ZrI_3$	11.72
$NH_3PH_2ZrI_3$	3.22	$OHPH_3SrI_3$	7.70	PH_3SHSrI_3	16.41
$OHPH_3SnI_3$	4.33	PH_3SHSrI_3	9.53	$OHNH_3SnI_3$	37.40
PH_3SHInI_3	5.56				

they contain iodide as counterion. Regarding the metallic cations, our results indicate that Zr, Sn, In, Sr and Y occur in sixty percent of the examined systems with suitable band gaps. We also assessed the energetic stability of the investigated perovskites and we found that phosphorus and sulphur are contained in the most recurrent molecular cations in the most stable perovskites towards degradation. Overall, our investigation illustrated how deep learning methods can be used to detect suitable HOIPs in terms of band gap and stability to be applied as photovoltaic materials and how these results can raise relevant questions about these perovskites.

6. Acknowledgements

We acknowledge financial support from CONACyT/Mexico (grant 596648/738983). We are also thankful to DGTIC-UNAM (grant LANCAT-UNAM-DGTIC-250) for computer time and AMP is grateful to Spanish MICINN for funding (grant PGC2018-095953-B-I00).

References

[1] M. G. Lawrence, S. Schäfer, H. Muri, V. Scott, A. Oschlies, N. E. Vaughan, O. Boucher, H. Schmidt, J. Haywood, J. Scheffran, *Nat. Commun.* **2018**, *9*, 3734.

[2] Y. Xu, V. Ramanathan, D. G. Victor, *Nature* **2018**, *564*, 30–32.

[3] L. C. Stokes, C. Warshaw, *Nat. Energy* **2017**, *2*, 17107.

[4] N. S. Lewis, D. G. Nocera, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 15729–15735.

[5] G. W. Crabtree, N. S. Lewis, *Phys. Today* **2007**, *60*, 37–42.

[6] Lazard Ltd - Financial advisory, asset management firm, *Levelized Cost Of Energy, Levelized Cost Of Storage, and Levelized Cost Of Hydrogen 2020*, Retrieved on Oct 18th, 2021. `https://www.lazard.com/perspective/levelized-cost-of-energy-levelized-cost-of-storage-and-levelized-cost-of-hydrogen-2020/`.

[7] R. Wang, M. Mujahid, Y. Duan, Z.-K. Wang, J. Xue, Y. Yang, *Adv. Funct. Mater.* **2019**, *29*, 1808843.

[8] M. Jošt, L. Kegelmann, L. Korte, S. Albrecht, *Adv. Energy Mater.* **2020**, *10*, 1904102.

[9] S. D. Wolf, J. Holovsky, S.-J. Moon, P. Löper, B. Niesen, M. Ledinsky, F.-J. Haug, J.-H. Yum, C. Ballif, *J. Phys. Chem. Lett.* **2014**, *5*, 1035–1039.

[10] P. L. A. Popelier, *Int. J. Quantum Chem.* **2015**, *115*, 1005–1011.

[11] V. Botu, R. Batra, J. Chapman, R. Ramprasad, *J. Phys. Chem. C* **2016**, *121*, 511–522.

[12] J. L. McDonagh, A. F. Silva, M. A. Vincent, P. L. A. Popelier, *J. Chem. Theory Comput.* **2017**, *14*, 216–224.

[13] J. T. Margraf, K. Reuter, *J. Phys. Chem. A* **2018**, *122*, 6343–6348.

[14] M. A. Kayala, C.-A. Azencott, J. H. Chen, P. Baldi, *J. Chem. Inf. Model.* **2011**, *51*, 2209–2222.

[15] M. A. Kayala, P. Baldi, *J. Chem. Inf. Model.* **2012**, *52*, 2526–2540.

[16] M. H. S. Segler, M. P. Waller, *Chem. Eur. J.* **2017**, *23*, 5966–5971.

[17] J. R. Kitchin, *Nat. Catal.* **2018**, *1*, 230–232.

[18] B. M. Bonk, J. W. Weis, B. Tidor, *J. Am. Chem. Soc.* **2019**, *141*, 4108–4118.

[19] B. Sánchez-Lengeling, A. Aspuru-Guzik, *Science* **2018**, *361*, 360–365.

[20] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Nature* **2018**, *559*, 547–555.

[21] K. T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, R. J. Maurer, *Nat. Commun.* **2019**, *10*, 5024.

[22] P. V. Balachandran, B. Kowalski, A. Sehirlioglu, T. Lookman, *Nat. Commun* **2018**, *9*, 1668.

[23] S. Zhang, T. Lu, P. Xu, Q. Tao, M. Li, W. Lu, *J. Phys. Chem. Lett.* **2021**, *12*, 7423–7430, PMID: 34337946.

[24] R. Lyu, C. E. Moore, T. Liu, Y. Yu, Y. Wu, *J. Am. Chem. Soc.* **2021**, *143*, 12766–12776, PMID: 34357756.

[25] J. Im, S. Lee, T.-W. Ko, H. W. Kim, Y. Hyon, H. Chang, *npj Comput. Mater.* **2019**, *5*, 37.

[26] V. Gladkikh, D. Y. Kim, A. Hajibabaei, A. Jana, C. W. Myung, K. S. Kim, *J. Phys. Chem. C* **2020**, *124*, 8905–8918.

[27] C. Kim, G. Pilania, R. Ramprasad, *J. Phys. Chem.* **2016**, *120*, 14575–14580.

[28] Y. Zhang, X. Xu, *Int. J. Quantum Chem.* **2021**, *121*, e26480.

[29] G. Pilania, A. Mannodi-Kanakkithodi, B. P. Uberuaga, R. Ramprasad, J. E. Gubernatis, T. Lookman, *Sci. Rep.* **2016**, *6*, 19375.

[30] S. Lu, Q. Zhou, Y. Ouyang, Y. Guo, Q. Li, J. Wang, *Nat. Commun.* **2018**, *9*, 3405.

[31] Y. Zhuo, A. Mansouri Tehrani, J. Brgoch, *J. Phys. Chem. Lett.* **2018**, *9*, 1668–1673.

[32] T. Xie, J. C. Grossman, *Phys. Rev. Lett.* **2018**, *120*, 145301.

[33] P. Dey, J. Bible, S. Datta, S. Broderick, J. Jasinski, M. Sunkara, M. Menon, K. Rajan, *Comput. Mater. Sci.* **2014**, *83*, 185–195.

[34] J. Lee, A. Seko, K. Shitara, K. Nakayama, I. Tanaka, *Phys. Rev. B* **2016**, *93*, 115104.

[35] W. A. Saidi, W. Shadid, I. E. Castelli, *npj Comput. Mater.* **2020**, *6*, 36.

[36] C. Kim, T. D. Huan, S. Krishnan, R. Ramprasad, *Sci. Data* **2017**, *4*, 170057.

[37] IBM Cloud Education, *Machine Learning*, **2020**, Retrieved on Oct 18th, 2021. `https://www.ibm.com/cloud/learn/machine-learning`.

[38] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[39] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, R. M. Wentzcovitch, *J. Phys. Condens. Matter* **2009**, *21*, 395502.

[40] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. B. Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. D. Corso, S. de Gironcoli, P. Delugas, R. A. DiStasio, A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J. Jia, M. Kawamura, H.-Y. Ko, A. Kokalj, E. Küçükbenli, M. Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H.-V. Nguyen, A. O. de-la Roza, L. Paulatto, S. Poncé, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A. P. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, X. Wu, S. Baroni, *J. Phys. Condens. Matter* **2017**, *29*, 465901.

[41] P. Giannozzi, O. Baseggio, P. Bonfà, D. Brunato, R. Car, I. Carnimeo, C. Cavazzoni, S. de Gironcoli, P. Delugas, F. F. Ruffino, A. Ferretti, N. Marzari, I. Timrov, A. Urru, S. Baroni, *J. Chem. Phys.* **2020**, *152*, 154105.

[42] P. Juhás, J. N. Louwen, L. van Eijck, E. T. C. Vogt, S. J. L. Billinge, *J. Appl. Cryst.* **2018**, *51*, 1492–1497.

[43] P. Juhás, C. L. Farrow, X. Yang, K. R. Knox, S. J. L. Billinge, *Acta Cryst.* **2015**, *71*, 562–568.

[44] L. Granlund, S. J. L. Billinge, P. M. Duxbury, *Acta Cryst.* **2015**, *71*, 392–409.

[45] P. Juhás, T. Davis, C. L. Farrow, S. J. L. Billinge, *J. Appl. Cryst.* **2013**, *46*, 560–566.

[46] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, *APL Materials* **2013**, *1*, 011002.

[47] M. de Jong, W. Chen, T. Angsten, A. Jain, R. Notestine, A. Gamst, M. Sluiter, C. Krishna Ande, S. van der Zwaag, J. J. Plata, C. Toher, S. Curtarolo, G. Ceder, K. A. Persson, M. Asta, *Scientific Data* **2015**, *2*.

[48] M. de Jong, W. Chen, H. Geerlings, M. Asta, K. A. Persson, *Scientific Data* **2015**, *2*.

[49] *materialsproject.org*, Retrieved on Oct 18th, 2021. `https://materialsproject.org`.

[50] G. E. Eperon, T. Leijtens, K. A. Bush, R. Prasanna, T. Green, J. T.-W. Wang, D. P. McMeekin, G. Volonakis, R. L. Milot, R. May, A. Palmstrom, D. J. Slotcavage, R. A. Belisle, J. B. Patel, E. S. Parrott, R. J. Sutton, W. Ma, F. Moghadam, B. Conings, A. Babayigit, H.-G. Boyen, S. Bent, F. Giustino, L. M. Herz, M. B. Johnston, M. D. McGehee, H. J. Snaith, *Science* **2016**, *354*, 861–865.

[51] M. A. Syzgantseva, O. A. Syzgantseva, *Theor. Chem. Acc.* **2019**, *138*, 52.