

Machine-Learning Aided First-Principles Prediction of Earth-Abundant Pnictogen Chalcogenide Solid Solutions for Solar-Cell Devices

Cibrán López, Ivan Caño, David Rovira, Pol Benítez, José Miguel Asensi, Zacharie Jehl, Josep-Lluís Tamarit, Edgardo Saucedo, and Claudio Cazorla*

Discovering novel families of materials composed of earth-abundant elements and characterized by non-toxicity, high thermodynamic stability, and simple low-temperature synthesis processes, is paramount for the advancement of urgently needed energy storage and conversion technologies. Pnictogen chalcogenides, represented by the general formula ABC ($A = \text{Bi, Sb}$; $B = \text{S, Se}$; $C = \text{I, Br}$), emerge as a promising class of energy materials particularly well-suited for photovoltaic applications. However, the compositional landscape of $\text{Bi}_x\text{Sb}_{1-x}\text{S}_y\text{Se}_{1-y}\text{I}_z\text{Br}_{1-z}$ is vast and remains largely unexplored, with traditional experimental and theoretical exploration techniques facing limitations in covering the entire solid-solution range due to their labor-intensive and time-consuming nature. Here, an integrated *bottom-up* approach that combines first-principles calculations, machine learning models, experiments, and device optimizations is introduced to provide a comprehensive fundamental understanding of pnictogen chalcogenides with arbitrary composition and to expedite the design of high-performance multi-junction solar cells. The synergistic investigations unveil a broad and continuous spectrum of bandgaps and optical absorption coefficients ranging from 1.2 to 2.1 eV and from $2.5 \cdot 10^5$ to $6.6 \cdot 10^5 \text{ cm}^{-1}$, respectively, across a wide variety of thermodynamically stable compounds. Additionally, a tandem BiSBr–BiSeI device is identified as an optimal multi-junction solar cell, exhibiting a maximum short-circuit current density of 18.65 mA cm^{-2} under intensity-matching conditions. The introduced *bottom-up* materials design approach may facilitate an unprecedented and rapid translation of basic knowledge into the most demanded solar cell applications.

1. Introduction

Given the substantial influx of sunlight energy reaching the Earth's surface, solar power emerges as a highly promising low-carbon alternative to fossil fuels.^[1,2] With the ongoing decline in the cost of solar devices, photovoltaics present tremendous potential among renewable energy resources for effectively curbing greenhouse gas emissions and, consequently, mitigating the impacts of climate change.

Single-junction solar cells demonstrate impressive solar conversion efficiencies, reaching values as high as 28% for silicon (Si), 29% for gallium arsenide (GaAs), 24% for copper indium gallium selenide (CIGS), and 26% for lead hybrid halide perovskites.^[3,4] However, these efficiencies are nearing the theoretical Shockley–Queisser maximum efficiency limit of $\approx 33\%$.^[5] Advancements beyond single-junction photovoltaic technologies, therefore, are imperative to drive new developments in solar energy conversion and sustainability.

Multi-junction solar cells consist of various semiconducting layers, each generating an electric current in response to distinct light wavelengths. Utilizing multiple

C. López, P. Benítez, J.-L. Tamarit, C. Cazorla
Group of Characterization of Materials
Departament de Física
Universitat Politècnica de Catalunya
Campus Diagonal-Besòs
Av. Eduard Maristany 10–14, Barcelona 08019, Spain
E-mail: claudio.cazorla@upc.edu

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/adfm.202406678>

© 2024 The Author(s). Advanced Functional Materials published by Wiley-VCH GmbH. This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/adfm.202406678

C. López, I. Caño, D. Rovira, P. Benítez, Z. Jehl, J.-L. Tamarit, E. Saucedo, C. Cazorla
Research Center in Multiscale Science and Engineering
Universitat Politècnica de Catalunya
Campus Diagonal-Besòs
Av. Eduard Maristany 10–14, Barcelona 08019, Spain
I. Caño, D. Rovira, Z. Jehl, E. Saucedo
Department of Electronic Engineering
Universitat Politècnica de Catalunya
Barcelona 08034, Spain
J. M. Asensi
Departament de Física Aplicada
Universitat de Barcelona
Barcelona 08007, Spain

semiconducting materials enables the absorption of light across a wider spectrum of wavelengths compared to single junctions, thereby enhancing the overall conversion efficiency of sunlight into electrical energy. To quantify this improvement, an infinite array of junctions would theoretically achieve a maximum efficiency limit of 86.8% under full concentrated sunlight.^[6,7]

Earth-abundant pnictogen chalcogenides with the general formula ABC ($A = \text{Bi, Sb}; B = \text{S, Se}; C = \text{I, Br}$), commonly referred to as “MChX,” have recently garnered considerable interest as a burgeoning class of solar absorbers. This interest stems from their non-toxicity, low synthesis temperatures ($<300\text{ C}^\circ$),^[8,9] bandgaps ranging from 1.0 to 2.0 eV,^[10,11] exceptional thermodynamic stability,^[12,13] and inherent tolerance to defects.^[14] The electron affinities and ionization potentials of MChX materials also seem to align well with various established charge transport layer materials.^[15,16] Additionally, at low temperatures, these materials are archetypal photoferroics in which a second-order Jahn–Teller distortion is stabilized due to the presence of s^2 lone pair electrons in the pnictogen atoms.^[17–19] Given these interesting properties, earth-abundant pnictogen chalcogenides present promising opportunities for the realization of multi-junction solar cells.

However, the integration of MChX solid solutions into advanced solar cell devices continues to lag behind due to several factors. First, the complete landscape of $\text{Bi}_x\text{Sb}_{1-x}\text{S}_y\text{Se}_{1-y}\text{I}_z\text{Br}_{1-z}$ compositions is vast and remains largely unexplored. Consequently, there is a painstaking lack of fundamental knowledge regarding the global optoelectronic properties of MChX materials. Both traditional experimental and computational exploration techniques may fall short in covering the entire range of possible MChX solid solutions due to their labor-intensive and time-consuming nature. Second, a significant gap persists between the fundamental understanding of the optoelectronic properties of materials and the actual design of solar cell devices. This disconnect stems from the substantial disparity in length scales, concepts, and techniques employed in both expertise areas. As a result, valuable *bottom-up* solar cell design approaches rarely are accomplished, impeding a rapid translation of basic knowledge into practical applications.

In this study, we propose a potential solution to the MChX solar-cell design challenge described above. Leveraging high-quality quantum density functional theory (DFT) calculations, we initially curated a comprehensive database of $\text{Bi}_x\text{Sb}_{1-x}\text{S}_y\text{Se}_{1-y}\text{I}_z\text{Br}_{1-z}$ compounds. Subsequently, we trained machine learning convolutional neural network (ML-CNN) models on the DFT data to efficiently predict the thermodynamic stability, energy bandgap and optical absorption properties of MChX solid solutions across their entire chemical spectrum. Finally, we conducted a preliminary evaluation of multi-junction devices from an optical viewpoint with macroscopic radiative transfer theory (RTT), utilizing the calculated complex refractive indices for MChX absorbers and a realistic solar cell architecture.

Our combined ML-aided DFT investigation unveiled a broad spectrum of bandgaps and light absorption coefficients spanning from 1.2 to 2.1 eV and $25\text{--}66\ \mu\text{m}^{-1}$, respectively, for a great variety of stoichiometries yielding thermodynamically stable compounds. These findings were elucidated based on electronic band structure characteristics and overarching chemical

trends. A $\text{Bi}_{0.3}\text{Sb}_{0.7}\text{SeI}$ solid solution was experimentally synthesized through a physical route and characterized at the structural and optical levels, finding consistent agreement with the ML-aided DFT predictions. Furthermore, our macroscopic RTT modeling identified a tandem $\text{BiSBr}\text{--}\text{BiSeI}$ device as an optimal multi-junction solar cell, yielding a maximum short-circuit current density of $18.65\ \text{mA cm}^{-2}$ under intensity-matching conditions. Thus, this study not only advances the fundamental understanding of MChX solid solutions but also establishes an innovative and universally applicable synergistic simulation approach for the *bottom-up* design of solar cells.

2. Results and Discussion

2.1. Integrated *Bottom-up* Approach

The holistic *bottom-up* materials design approach introduced in this study consists of four different synergistic steps (**Figure 1**): 1) *DFT characterization and database curation*: Initial characterization of the eight ABC ($A = \text{Bi, Sb}; B = \text{S, Se}; C = \text{I, Br}$) parent compounds along with selected solid solutions using density functional theory (DFT); this step involved studying the electronic structure and optical properties of specific MChX to create an extensive database. 2) *Machine learning models generation*: Development of reliable machine learning convolutional neural network (ML-CNN) models to comprehensively characterize the complete compositional space of MChX solid solutions; these models aim to predict the thermodynamic stability and optoelectronic properties of any possible $\text{Bi}_x\text{Sb}_{1-x}\text{S}_y\text{Se}_{1-y}\text{I}_z\text{Br}_{1-z}$ compound. 3) *Experimental verification*: Validation of the computational predictions through materials synthesis and experimental analysis; this step involved synthesizing MChX solid solutions in the laboratory and measuring their structural and optical properties to compare against the predictions from DFT and ML models. 4) *Simulation and optimization of multi-junction solar cells*: The validated computational models were used to simulate and optimize multi-junction MChX solar cell devices; this task involved designing and optimizing device architectures to maximize performance metrics such as the short-circuit current density.

This integrated *bottom-up* approach (i.e., *bottom* for the initial step 1 and *up* for the final step 4 described above) facilitates a comprehensive understanding of MChX solid solutions and enables the design of high-performance solar cells through a combination of theoretical predictions, experimental validation, and device optimization. Details of this general materials design strategy are provided next.

2.1.1. Curated First-Principles MChX Database

An extensive database of first-principles calculations covering $\text{Bi}_x\text{Sb}_{1-x}\text{S}_y\text{Se}_{1-y}\text{I}_z\text{Br}_{1-z}$ solid solutions with compositions $x, y, z = \{0, 0.25, 0.5, 0.75, 1\}$ was created, totaling 125 different compounds.^[22,23] The crystal structure of all the investigated materials was assumed to be the centrosymmetric orthorhombic $Pnma$ phase experimentally observed in the eight parent compounds under ordinary room-temperature conditions (**Figure 2a, b**)^[10,11] (alternative ferroelectric phases possibly

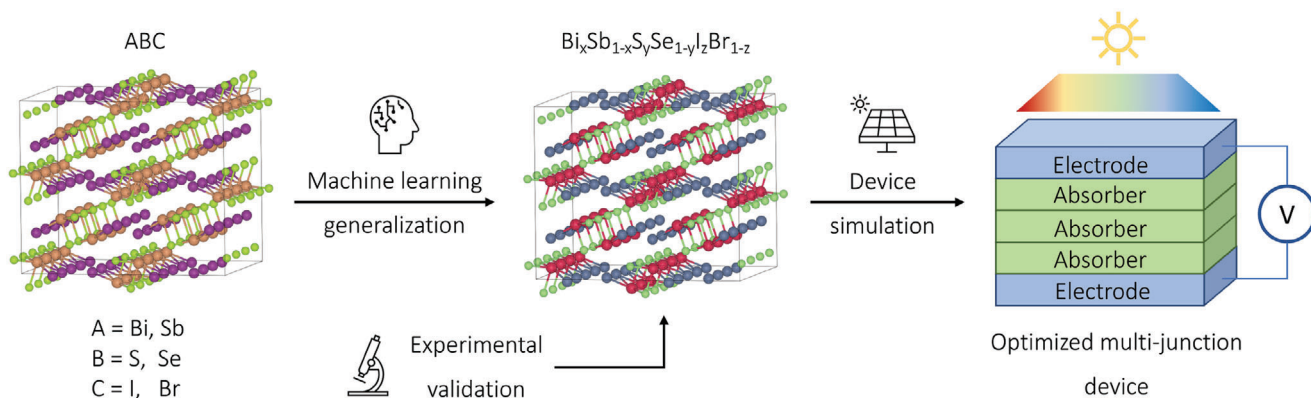


Figure 1. Integrated *bottom-up* approach proposed to expand knowledge on MChX solid solutions and design multi-junction solar cell devices. From left to right: Curation of a comprehensive DFT database of MChX solid solutions, generation of machine learning models^[20,21] for prediction of energy and optoelectronic properties, experimental synthesis of MChX solid solutions for machine learning validation, and simulation and optimization of multi-junction MChX solar cells.

stabilized at lower temperatures^[17–19] have not been considered in this study). The physical quantities estimated for each material using advanced first-principles methods included the lattice parameters, total energy, convex hull energy (Figure 2c), electronic bandgap and frequency-dependent complex dielectric function.

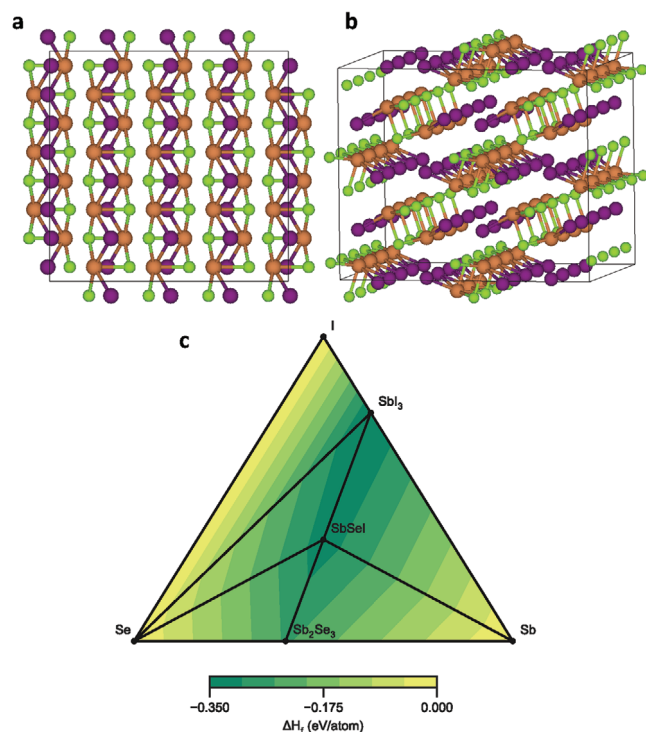


Figure 2. General structural and thermodynamic stability properties of MChX compounds. Characteristic orthorhombic *Pnma* phase a) along the [001] direction and b) in a 3D view particularized for the parent compound SbSeI. Sb, Se, and I atoms are represented with brown, green, and purple spheres, respectively. c) Convex-hull surface of SbSeI calculated with DFT methods. SbSeI is predicted to be thermodynamically stable against segregation into secondary phases because its formation enthalpy is negative relative to the convex-hull surface.

From the latter quantity, the optical refractive index, extinction coefficient, absorption coefficient, energy-loss function and reflectivity can be straightforwardly deduced.^[24]

First-principles calculations based on density functional theory (DFT)^[25] were carried out by employing the Perdew–Burke–Ernzerhof exchange–correlation energy functional revised for solids (PBEsol)^[26] as the baseline (Methods). The atomic structures were optimized at this level of theory. To estimate optoelectronic properties such as bandgaps and optical absorption coefficients, the range-separated hybrid functional HSEsol^[27,28] was used to mitigate the well-known shortcomings of semi-local functionals (e.g., self-interaction energy errors). Additionally, spin-orbit coupling effects (SOC), which may be particularly significant for Bi-containing compounds, were also taken into consideration (HSEsol+SOC).

Modeling of solid solutions and alloys poses technical challenges due to the requirement of multiple and large simulation supercells to accurately replicate arbitrary stoichiometries and chemical disorder.^[29,30] Simulation techniques like the special quasirandom structure (SQS) approach^[31] have significantly simplified the modeling of chemically disordered crystals by introducing radial correlations in site occupations. However, the size of SQS supercells can still be prohibitively large for HSEsol+SOC calculations, as pursued in this study.

Alternatively, we employed here the virtual crystal approximation (VCA),^[32] which assumes the presence of virtual atoms on potentially disordered sites, interpolating between the electronic traits of the actual components. VCA supercells match the size of the primitive cell, making them computationally manageable, and usually provide reliable optoelectronic results for isoelectronic atom substitutions.^[33]

Several tests were conducted to demonstrate the suitability of the VCA method in simulating MChX solid solutions. Specifically, we assessed the reliability of the VCA method in predicting the energy bandgap of Bi_{0.5}Sb_{0.5}SeI, a compound that, in terms of computational affordability, is one of the most suitable for undertaking DFT calculations. The energy bandgap was evaluated at the HSEsol+SOC level of theory using 1) the DFT-VCA method, 2) DFT and a combinatorial approach considering the 12-atoms unit cell, and 3) DFT and a combinatorial approach considering

a $2 \times 1 \times 1$ supercell. Our findings revealed only slight disparities, approximately within 0.1 eV, among the provided results. In particular, we obtained 1.87 eV from (1), 1.67 eV from (2), and 1.81 eV from (3). This variance is consistent with the typical accuracy of DFT calculations, thus demonstrating the suitability of the DFT–VCA method for addressing the problem of interest in this study.

2.1.2. Machine Learning Models

Machine learning (ML) methods have shown significant promise in predicting and characterizing chemically disordered materials,^[34] despite the inherent challenges associated with the modeling of solid solutions.^[35,36] In this study, we employed convolutional neural networks (CNN) implemented in the Scikit-learn package^[37] for predicting the thermodynamic stability and optoelectronic properties of MChX solid solutions (Methods). ML-CNN maintains high accuracy by directly using the stoichiometry of the chemically disordered crystal as input data. The performance of the generated ML-CNN models was evaluated using the mean absolute error (MAE) obtained from k -fold cross-validation tests ($k = 20$) conducted on both training and test sets.^[38] While alternative ML models based on random forest regressors were also explored, they were found to underperform CNN models in the prediction of MChX solid-solution properties (Figures S1–S3, Supporting Information).

ML-aided first-principles DFT predictions were made on a dense grid of $\text{Bi}_x\text{Sb}_{1-x}\text{S}_y\text{Se}_{1-y}\text{I}_z\text{Br}_{1-z}$ compositions taken at 0.05 intervals on x , y and z . This comprehensive approach enabled us to assess the thermodynamic stability and optoelectronic properties of a total of 9261 solid solutions. A repository containing our ML-CNN models, numerical tests and Python codes has been created and is freely accessible at ref. [39].

2.1.3. Experimental Synthesis

An MChX solid solution with a composition of $\text{Bi}_{0.3}\text{Sb}_{0.7}\text{SeI}$ was successfully synthesized through a physical route involving high-pressure treatment (Methods). The synthesis process consisted of two main steps: precursor synthesis and halogenation of the precursor to yield the pnictogen chalcogenide.^[8] Initially, a Bi_2Se_3 precursor was deposited on a molybdenum-coated glass substrate via co-evaporation. Subsequently, the MChX solid solution was synthesized through reactive annealing at moderate temperatures ($<300\text{ C}^\circ$) and pressures ($>1\text{ atm}$) in a halogen atmosphere composed of BiI_3 and SbI_3 (Methods). Although samples with different compositions were synthesized using the same experimental approach, their level of crystallinity was insufficient for subsequent accurate characterization, leading to their exclusion from further analysis.

2.1.4. Solar Cell Device Modeling

Macroscopic radiative transfer theory (RTT) serves as a robust modeling approach for determining the reflectance, transmittance, and specific absorbance of a material within a multilayered materials stack.^[40,41] It operates based on the formulation of

matrices that describe the propagation and boundary conditions of electromagnetic waves as they traverse through a medium and undergo transmission or reflection at interfaces. Each layer within the stack is characterized by parameters such as thickness, refractive index, and extinction coefficient. The propagation matrix incorporates factors accounting for phase changes as light propagates through each layer, while the interface matrix considers changes in amplitude at interfaces according to Fresnel equations. RTT offers high accuracy, provided reliable input parameters are used, and captures valuable information including interferences within the stack. Additionally, RTT is computationally efficient, yielding results within seconds, and relies on fully analytical methods.

However, RTT also has its limitations. First, the method assumes coherent light propagation, which may not hold true for very thick layers or layers with significant absorption. Second, scattering effects are not inherently accounted for in RTT. Nonetheless, there are workarounds for these limitations. For instance, effective medium theories can be employed to address scattering effects,^[42,43] and incoherent propagation can be simulated through an averaging process over the angular distribution while considering the intensity obtained from Fresnel equations. In the context of multi-junctions, well-defined specular interfaces are crucial to maintain the conformal growth of each subcell. Consequently, considering scattering effects in such cases might be unnecessary as it would deviate from the optimal scenario. Thus, RTT remains well-suited for providing quantitative optical results in these scenarios.

In this study, due to the lack of comprehensive experimental data on the optical properties of pnictogen chalcogenide absorbers, we employed a pragmatic approach. We utilized a set of partner materials with well-characterized optical constants to evaluate the optical performance of the MChX absorbers within a multi-junction design framework. The optical indices calculated using DFT methods served as inputs for this evaluation. RTT was employed to establish an upper bound for the photovoltaic current generated by each subcell. Additionally, it was feasible to impose a constraint on the intensity by aligning the current of each subcell, particularly in a 2-terminal multi-junction design where all subcells operate in series. Consequently, an optically optimal configuration was determined by adjusting the thickness and bandgap of each subcell to achieve the desired current.

2.2. MChX Parent Compounds

Table 1 provides both experimental and theoretical information regarding the optoelectronic and structural properties of the eight MChX parent compounds. According to our DFT calculations, the formation enthalpy of all these materials is below the corresponding convex hull surfaces, indicating their thermodynamic stability against phase separation (Figure 2c).

The agreement between our theoretical predictions and the experimental data is remarkably good concerning the energy bandgap, E_g , and lattice parameters. Specifically, for Bi-based compounds, the calculated bandgaps align closely with the experimental values, considering the typical numerical uncertainty of 0.1 eV. However, for Sb-based compounds, our DFT calculations slightly underestimate the measured E_g . It is worth noting that

Table 1. Experimental and calculated structural and optoelectronic properties of the eight MChX parent compounds. Theoretical DFT values of the energy bandgap, E_g , and valence band maximum relative to the vacuum level, VBM, were obtained at the HSEsol+SOC level (this work). Theoretical DFT values of the lattice parameters a , b , and c corresponding to the orthorhombic $Pnma$ phase were obtained at the PBEsol level (this work).

Compound	E_g^{Expt} [eV]	E_g^{DFT} [eV]	$(a, b, c)^{\text{Expt}}$ [Å]	$(a, b, c)^{\text{DFT}}$ [Å]	VBM ^{Expt} [eV]	VBM ^{DFT} [eV]
BiSI	1.6 [15]	1.5	4.12, 8.39, 10.11 [15]	4.14, 8.34, 10.02	-6.3 [15]	-7.2
BiSeI	1.3 [11]	1.3	4.22, 8.70, 10.58 [44]	4.19, 8.55, 10.34	-	-5.7
BiSBr	1.9 [9]	1.8	4.06, 8.14, 9.85 [9]	4.03, 8.06, 9.52	-5.9 [9]	-5.6
BiSeBr	1.3–1.7 [10]	1.3	-	4.08, 8.14, 10.12	-	-5.9
SbSI	1.8–2.2 [10]	1.7	-	4.04, 8.31, 9.97	-5.4 [10]	-6.7
SbSeI	1.7 [10]	1.5	-	4.10, 8.51, 10.27	-5.3 [10]	-6.4
SbSBr	1.8–2.2 [10]	1.9	-	3.90, 8.05, 9.58	-5.8 [10]	-6.0
SbSeBr	1.7 [10]	1.5	4.02, 8.32, 10.23 [44]	3.97, 8.14, 10.09	-	-5.5

the dispersion in the experimental values of SbSI and SbSBr is not negligible, contributing to the observed variance. Regarding the lattice parameters, the discrepancies between the experimental and theoretical data amount to only $\approx 2\%$, with the DFT values being consistently smaller. This regular underestimation of lattice parameters can be attributed to the neglect of thermal expansion effects in the calculations.

Regarding the valence band maximum relative to the vacuum level, VBM (Figure 3), the agreement between the DFT and experimental values is less satisfactory. Particularly, the theoretical values tend to underestimate this quantity by about 10% for Bi-based compounds and 20% for Sb-based compounds. The physical origins of these discrepancies are not clear to us.

Some general E_g and VBM trends can be deduced from the DFT data provided in Table 1, which systematically characterize the eight MChX parent compounds. Concerning E_g , it is observed that the larger the atomic radius of the element on the A, B, or C position, the smaller the resulting energy bandgap. Accordingly, the smallest E_g is obtained for BiSeI (1.3 eV), and the largest for SbSBr (1.9 eV). However, there is one exception to this behavior for Se-based compounds; BiSeI and BiSeBr, on one hand, and SbSeI and SbSeBr, on the other, exhibit identical energy bandgaps.

Regarding VBM, the results are less regular, making them more challenging to generalize in terms of simple atomic descriptors. Nevertheless, it is observed that in general when the

radius of the element on the C position is larger, the value of the VBM turns out to be larger in absolute value. Conversely, for Sb-based compounds, the larger the radius of the element on the B position, the smaller the value of the VBM is in absolute value.

In the following sections, we predict and analyze the thermodynamic stability and optoelectronic properties of arbitrary $\text{Bi}_x\text{Sb}_{1-x}\text{S}_y\text{Se}_{1-y}\text{I}_z\text{Br}_{1-z}$ solid solutions using the integral *bottom-up* approach described in Section 2.1. Machine learning predictions were made on a dense grid of compositions taken at 0.05 intervals on x , y , and z , implying that we assessed the properties of a total of 9261 solid solutions. We would like to emphasize that the obtained results do not simply conform to linear interpolations based on the values obtained for the eight MChX parent compounds described in this section.

2.3. Thermodynamic Stability of MChX Solid Solutions

The thermodynamic stability against segregation into secondary phases of all possible MChX solid solutions (≈ 9000) was assessed using convex-hull calculations based on ML-CNN models (Figure 2c; Figure S4, Supporting Information). Materials with formation enthalpies, ΔH_f , below the convex-hull surface are expected to be thermodynamically stable, while those with formation energies above the convex-hull surface are expected to

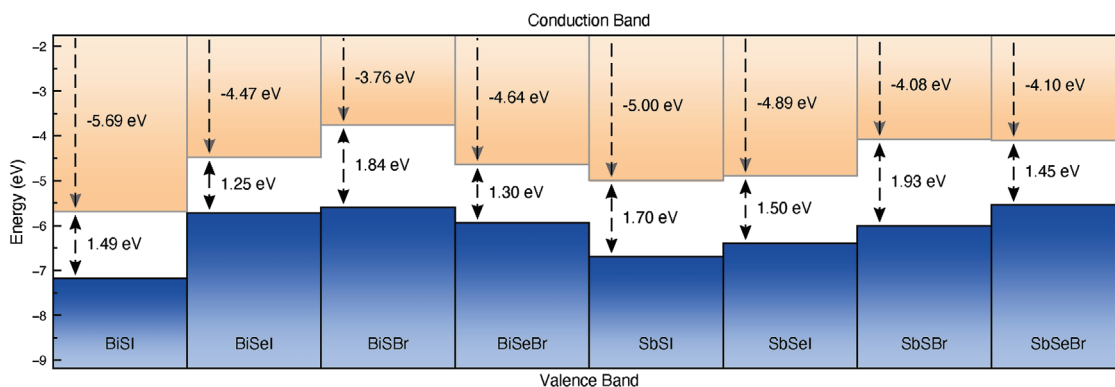


Figure 3. Band alignments of the eight MChX parent compounds estimated with DFT methods (HSEsol+SOC). The energy of the top of the valence band and the bottom of the conduction band are relative to the vacuum level. The results are not systematic, making them challenging to generalize in terms of simple atomic descriptors.

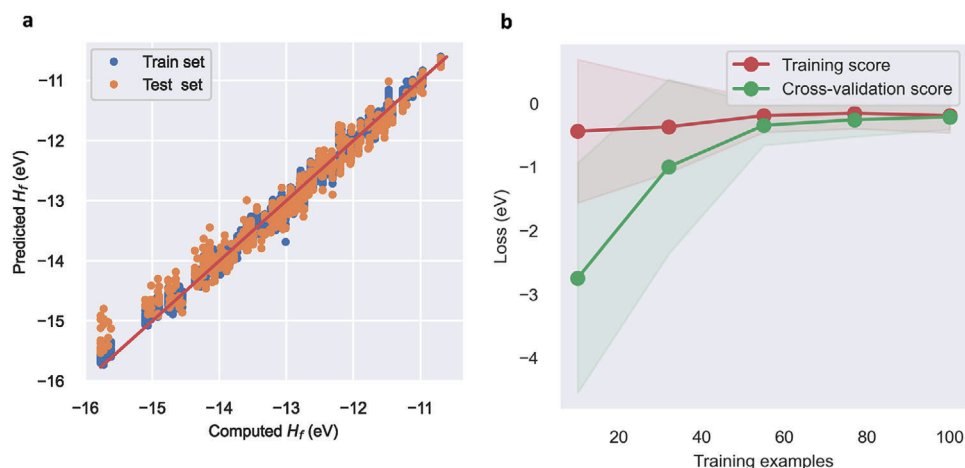


Figure 4. Evaluation of the performance of the ML-CNN model generated for the prediction of ground-state energies. a) The predicted values are compared with the computed DFT data. The solid red line represents the best ideal performance. b) Convergence of the ground-state energy model in terms of data samples. It is demonstrated that increasing the number of samples would not increase the accuracy of the model. For both the training and test sets, the negative mean squared error, that is, loss function, is represented against the number of training samples. A cross-validation strategy with 20 folds was implemented, obtaining mean absolute errors of 0.047 and 0.042 eV for the training and test sets, respectively.

decompose into other compounds. In this work, a reasonable, although somewhat arbitrary, numerical threshold of $\Delta H_f \geq 0.1$ eV per atom was adopted to distinguish thermodynamically stable compounds from unstable ones. This threshold accounts for the disregarded finite-temperature entropy effects in both the DFT calculations and the ML-CNN predictions, which are expected to favor the stabilization of chemically disordered systems in practice.^[29,30]

For the evaluation of ΔH_f , we trained ML-CNN models to accurately predict the DFT ground-state energy of MChX solid solutions and also used the pymatgen code.^[45] Figure 4a demonstrates the consistently good performance of the trained model, for which overfitting was efficiently avoided. Moreover, our DFT database is proven to be well-converged in terms of data samples, as depicted in Figure 4b. In particular, the training loss (i.e., negative mean square error) monotonically decreases with the number of training samples and converges to zero for both the training and test sets. This behavior implies that adding more data samples would not significantly improve the performance of the model. For both analyses, a k -fold cross-validation strategy with 20 validation runs was conducted, as this is common practice in the literature and provides statistically significant results.

Figure 5 presents our ΔH_f results, with compounds meeting the condition $\Delta H_f \leq 0.1$ eV per atom represented for clarity. A general domain resembling the contour of a “boxing ring” emerges. The systems contained in that compositional region are likely to be thermodynamically stable against segregation into secondary phases. In general, it is observed that compounds with mixed halogen content on the C position exhibit improved thermodynamic stability (i.e., smaller ΔH_f values). For instance, $\text{SbSeI}_{0.4}\text{Br}_{0.6}$ holds the minimum formation enthalpy relative to the convex hull of all MChX compounds, which amounts to -0.42 eV per atom. Similar ΔH_f values were also calculated for BiSe-, BiS-, and SbS-based solid solutions containing $\approx 50\%$ of each halide species.

Conversely, blending simultaneously pnictogen and chalcogen ions on the A and B positions induces a significant drop in the thermodynamic stability of the solid solution, as shown by the empty “ring” in Figure 5. Nonetheless, compounds with mixed

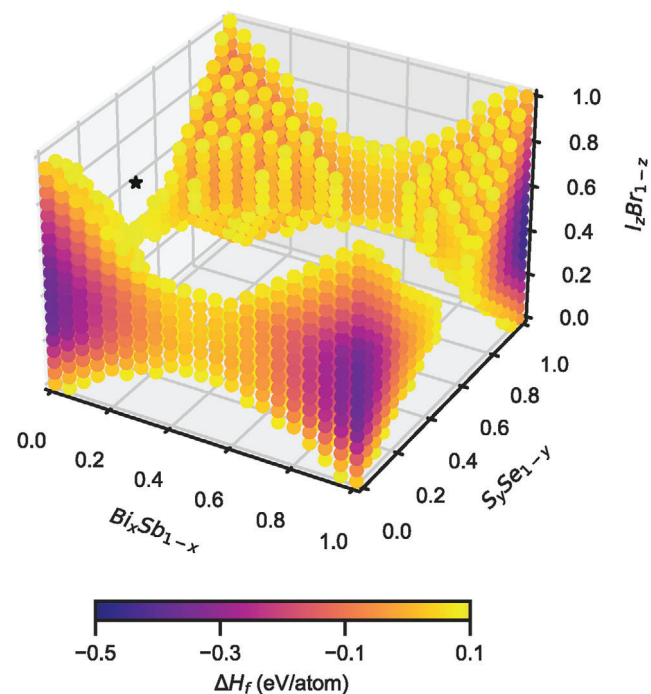


Figure 5. Thermodynamic stability against segregation into secondary phases of MChX solid solutions. Only regions of the configurational space that render thermodynamically stable compounds are represented. Those regions were determined with the condition $\Delta H_f \leq 0.1$ eV per atom, where the formation enthalpy is relative to the convex-hull surface. The black star represents the $\text{Bi}_{0.3}\text{Sb}_{0.7}\text{SeI}$ solid solution synthesized in this study.

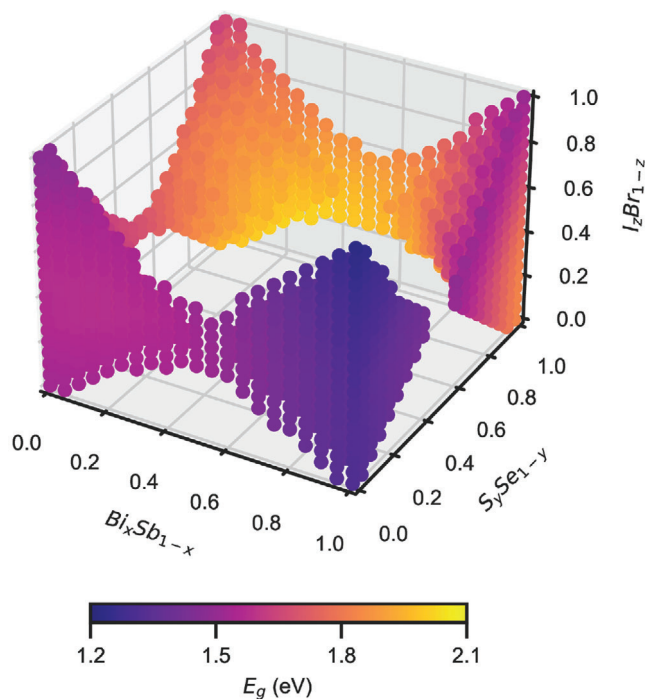


Figure 6. Predicted bandgaps of MChX solid solutions at the HSEsol+SOC level of theory. A continuous spectrum of bandgaps is obtained ranging from 1.2 to 2.1 eV. Results are presented only for compounds that according to our calculations are thermodynamically stable against segregation into secondary phases.

pnictogens on the A position or mixed chalcogens on the B position are also predicted to be synthesizable, as indicated by the sidelong regions that practically connect the four edges of the quadrilateral in Figure 5.

Regarding the somewhat arbitrary condition $\Delta H_f \leq 0.1$ eV per atom used here to distinguish thermodynamically stable from unstable compounds, it is noted that it appears to be physically well-justified. In this work, a $\text{Bi}_{0.3}\text{Sb}_{0.7}\text{SeI}$ solid solution was experimentally synthesized and observed to be thermodynamically stable at room-temperature conditions (Section 2.6). For this particular case, our ML-CNN model predicts a ΔH_f of 0.3 eV per atom, which is appreciably larger than 0.1 eV per atom. Therefore, our adopted formation enthalpy threshold appears to be quite conservative and highly reliable for our present purposes.

2.4. Bandgap of MChX Solid Solutions

We trained an ML-CNN model in our database of 125 pnictogen chalcogenides (Section 2.1.1) to accurately predict the energy bandgap of MChX solid solutions of any arbitrary composition at the HSEsol-SOC level of theory. The performance of the trained bandgap machine learning model is remarkably good, as demonstrated by its convergence in terms of data samples (Figure S5, Supporting Information). Our E_g results are shown in Figure 6, where we only include compounds that are thermodynamically stable against segregation into secondary phases (Figure 5).

Our ML-aided first-principles calculations reveal a continuous spectrum of energy bandgaps for synthesizable MChX solid

solutions that spans from 1.2 to 2.1 eV. Predominantly, energy bandgaps smaller than 1.8 eV are observed, as indicated by the overall blueish tone of the 3D plot in Figure 6. This range of possible E_g values, all achievable through compositional engineering, holds great promise for the design of bottom-cell absorbers in multi-junction photovoltaic devices^[6,7] (we will address this point in more detail in the next Section 2.7).

The smallest and largest E_g values were obtained for BiSeI and $\text{Bi}_{0.25}\text{Sb}_{0.75}\text{SI}_{0.25}\text{Br}_{0.75}$, which respectively amount to 1.25 and 2.05 eV. Additionally, it was found that the nature of the energy bandgap is typically indirect, although the secondary energy bandgap is usually direct and only around 0.1 eV larger than the primary one (Figure 7a).

Interestingly, we observed a significant “bowing” effect in E_g under pnictogen substitution on the A position, with a maximum occurring around $x = 0.25$. Similar effects have been previously observed in analogous solid-solution optoelectronic materials, such as chalcogenide anti-perovskites,^[47] albeit at different concentrations. This phenomenon elucidates the composition of the maximum in the energy bandgap. Regarding chalcogen substitutions on the B position, we did not observe a “bowing” effect; instead, we found a smooth and monotonic E_g behavior resembling a half parabola.

For halogen substitutions on the C position, we identified two distinct situations, akin to those observed in the parent MChX compounds. Specifically, for S-based materials, small variations in E_g were achieved under I/Br substitutions, accompanied by an also small “bowing” effect. However, for Se-based compounds, the size of the energy bandgap remained largely insensitive to the specific content of I and Br ions.

The explanation for the observed E_g invariance under I/Br substitution in Se-based MChX is elucidated by the partial density of electronic states (pDOS) depicted in Figure 7b. For both the SbSeBr and $\text{SbSeI}_{0.5}\text{Br}_{0.5}$ crystals, it can be observed that pnictogen orbitals predominantly occupy the bottom of the conduction band, while chalcogen orbitals are predominant at the top of the valence band. Halogen orbitals, on the other hand, are predominant in the low-energy region of the valence band. Therefore, since the contribution of halogen orbitals to the energy levels defining E_g is very limited, this quantity turns out to be insensitive to their relative content. For completeness, analogous pDOS plots are provided in Figure S6 (Supporting Information) for the compounds $\text{Bi}_{0.5}\text{Sb}_{0.5}\text{SeBr}$ and $\text{SbS}_{0.5}\text{Se}_{0.5}\text{Br}$.

2.5. Optical Absorption of MChX Solid Solutions

To evaluate the feasibility of MChX solid solutions as photoabsorbers, we calculated their optical absorption coefficient in the visible range, denoted as α , at the HSEsol-SOC level of theory. Additionally, we determined the energy loss, extinction coefficient, reflectivity, and refractive index for the 125 pnictogen chalcogenides comprising our database (not presented here).

We trained an ML-CNN model to accurately predict the α of MChX solid solutions of any arbitrary composition. The performance of the trained optical absorption coefficient machine learning model is remarkably good, as demonstrated by the high accuracy achieved in both the training and test sets (Figure S7, Supporting Information). Our α results are depicted in Figure 8,

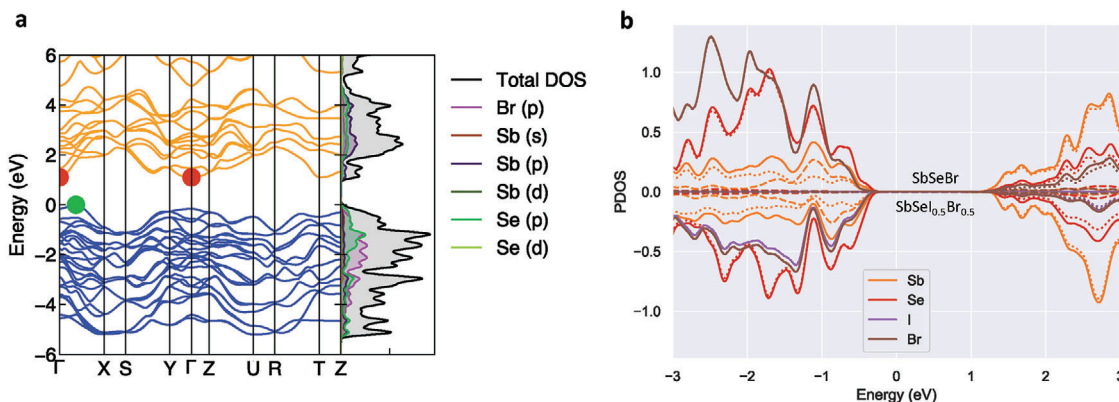


Figure 7. Electronic band structure properties of MChX solid solutions. a) Energy band structure of SbSeBr calculated along high-symmetry paths in reciprocal space at the PBEsol+SOC level of theory. Conduction and valence bands are represented in yellow and blue colors, respectively, and the contribution of each type of orbital is shown in the adjacent density of states. The Fermi energy level has been set to zero. Green and red dots represent the top of the valence band and the two lowest energy levels in the conduction band, respectively. The bandgap is indirect but the secondary bandgap is direct and very much similar in size to the primary one (by 0.1 eV). The figure was generated with the SUMO toolkit.^[46] b) Partial density of states (pDOS) of SbSeBr (top) and SbSe_{0.5}Br_{0.5} (bottom) calculated at the HSEsol+SOC level of theory. Solid, dotted, and dashed lines represent total, *p* and *s* pDOS, respectively.

where we only include compounds that are thermodynamically stable against segregation into secondary phases (Figure 5). It is worth noting that the optical absorption coefficient depends on the photon frequency. Therefore, for consistency, we considered

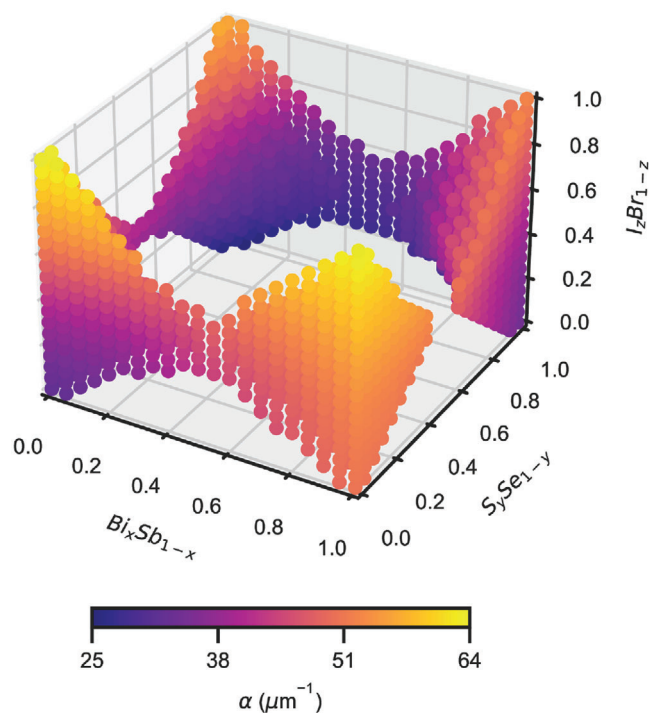


Figure 8. Optical absorption coefficient of MChX solid solutions in the visible spectrum. The optical absorption coefficient depends on the photon frequency; here we consider above the bandgap excitations corresponding to a photon frequency of 788 THz ($\lambda = 380$ nm, maximum in the visible spectrum). Results are presented only for compounds that according to our calculations are thermodynamically stable against segregation into secondary phases.

a fixed photon frequency of 788 THz (380 nm), corresponding to the maximum in the visible spectrum.

Our ML-aided first-principles calculations reveal a range of optical absorption coefficients for synthesizable MChX solid solutions that continuously span from 25 to 66 μm^{-1} (or equivalently, from $2.5 \cdot 10^5$ to $6.6 \cdot 10^5$ cm^{-1}). Notably, the smallest and largest α values were obtained for Bi_{0.05}Sb_{0.95}SBr (25.11 μm^{-1}) and BiSeI (65.25 μm^{-1}). These results are highly promising as they are comparable in size to the absorption coefficients measured for hybrid halide perovskites^[48] and other state-of-the-art photovoltaic materials.^[49]

The general α trends observed in Figure 8 are quite systematic. Essentially, it is found that as the average radius of the ions on the A, B and/or C positions increases, the absorption coefficient of the corresponding MChX solid solution also increases. The variation in α that is achievable through ion substitution is substantial. For instance, when considering ionic substitutions exclusively on the A or B positions, absorption coefficient variations of up to 60–70% relative to the limiting α value of 25 μm^{-1} can be attained. Remarkably, ionic substitutions performed on the C position, which were practically inconsequential for the energy bandgap, may induce giant absorption coefficient variations of >80%.

2.6. Experimental Validation

As a practical means of validating our ML-aided first-principles predictions on MChX solid solutions, we experimentally synthesized and characterized the material Bi_{0.3}Sb_{0.7}SeI (Methods). This compound exhibits a complex X-ray diffractogram (XRD) due to its low crystal symmetry (Figure 9a). From our XRD analysis, the following conclusions were drawn.

XRD peaks corresponding to the binary compounds Sb₂Se₃ and Bi₂Se₃ involved in the synthesis of the material were absent, indicating the successful completion of the halogenation reaction (i.e., secondary phases and decomposition products were

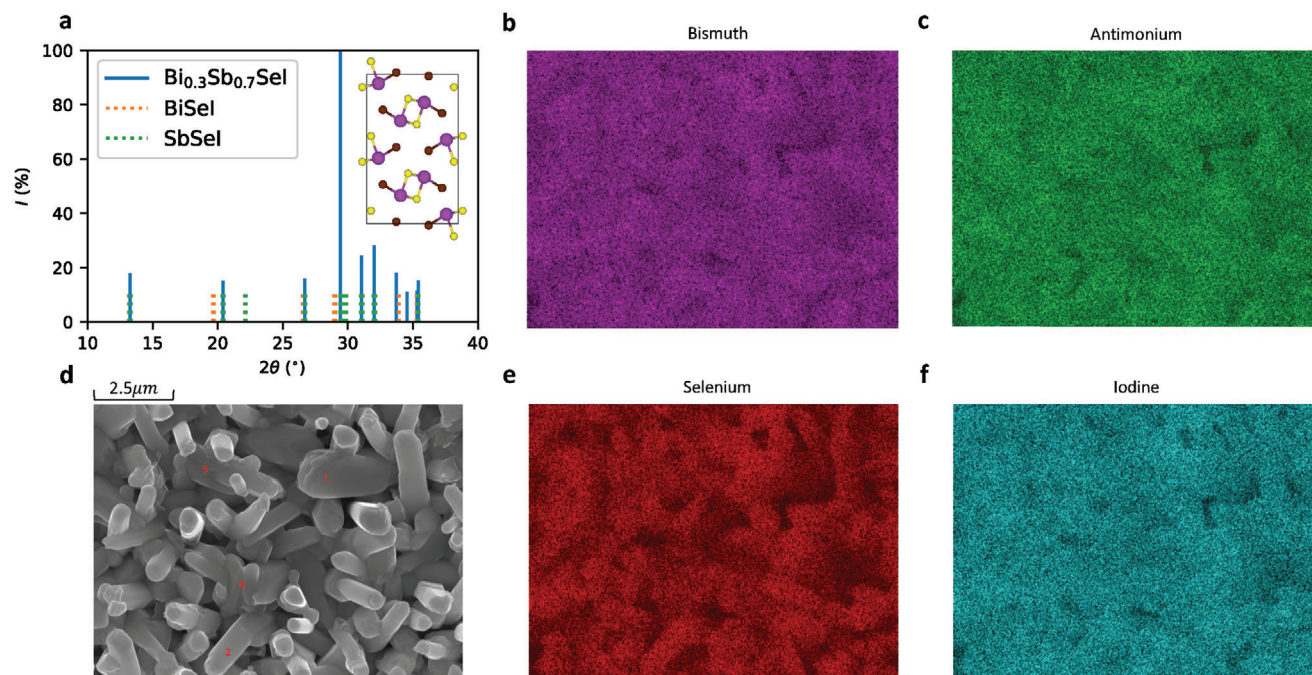


Figure 9. Experimental synthesis and characterization of the MChX solid solution $\text{Bi}_{0.3}\text{Sb}_{0.7}\text{SeI}$. a) XRD image of the crystal structure of the experimentally synthesized material. The XRD images of the parent compounds BiSeI and SbSeI are shown for comparison. Energy-dispersive X-ray spectroscopy maps showing very uniform concentration and distribution of b) bismuth, c) antimony, e) selenium, and f) iodine ions throughout the synthesized sample. d) SEM image showing columnar growth and from which four atomic locations were selected for further analysis of their composition (Table 2).

not identified). The solid solution was observed to crystallize in an orthorhombic $Pnma$ phase, consistent with the assumption made in the DFT calculations. The lattice parameters of the solid-solution crystal were determined through a LeBail refinement analysis performed with the FullProf software,^[50] yielding the values $a^{\text{Expt.}} = 4.13$, $b^{\text{Expt.}} = 8.70$ and $c^{\text{Expt.}} = 10.41$ Å. These results are in excellent agreement with the corresponding calculated values $a^{\text{DFT}} = 4.13$, $b^{\text{DFT}} = 8.60$, and $c^{\text{DFT}} = 10.59$ Å. It is worth noting that the theoretical lattice parameters were estimated at zero-temperature conditions, while the experimental values were obtained at room temperature. Therefore, the $\approx 1\%$ theoretical underestimation of the experimental lattice parameters is likely due to thermal lattice expansion effects.

As shown in Figure 9a, the 2θ reflections for BiSeI and SbSeI (represented by blue and red dashed lines, respectively), as extracted from a crystallographic data set (references 00–044–0162 and 01–076–1354, respectively), do not perfectly align with the peaks observed in our X-ray diffractogram for $\text{Bi}_{0.3}\text{Sb}_{0.7}\text{SeI}$. This discrepancy, more pronounced at larger angles, suggests that the structure of the solid solution exhibits some degree of cell-shape deformation compared to the structure of the parent MChX compounds BiSeI and SbSeI, as expected for chemically disordered systems. Specifically, for $\text{Bi}_{0.3}\text{Sb}_{0.7}\text{SeI}$ the peaks appear at larger 2θ values than for BiSeI (indicating a relatively contracted lattice) and smaller than for SbSeI (indicating a relatively expanded lattice). Additionally, the presence of double peaks in some reflections of the solid solution suggests a more complex interpretation. For example, it could indicate that the $\text{Bi}_{0.3}\text{Sb}_{0.7}\text{SeI}$ sample is not entirely homogeneous but rather exhibits some degree of compositional gradient.

In order to quantify the degree of compositional homogeneity in the synthesized solid solution, we conducted a series of scanning electron microscopy (SEM) energy-dispersive X-ray spectroscopy (EDX) measurements. The resulting compositional maps are presented in Figures 9b–f. The SEM image depicted in Figure 9d reveals a quasi 1D columnar morphology, similar to the morphology observed in the parent MChX compounds.^[10,11] From the SEM image, we selected four different atomic environments for which we performed the EDX analysis. The relative atomic abundances in these separate locations are reported in Table 2. The highly similar concentration of each atomic species observed in all the analyzed atomic environments indicates that the composition of the synthesized solid solution is highly homogeneous.

Regarding the optical properties, the transmittance, reflectance, and absorption coefficient of $\text{Bi}_{0.3}\text{Sb}_{0.7}\text{SeI}$ were experimentally measured (Figure S8, Supporting Information). How-

Table 2. Energy-dispersive X-ray (EDX) spectroscopy analysis of the synthesized $\text{Bi}_{0.3}\text{Sb}_{0.7}\text{SeI}$ sample. Detailed composition measured at the four specific columns depicted in Figure 9d. The highly similar concentration of each atomic species for all points indicates the synthesis of a very uniform solid solution.

Point	Bi [%]	Sb [%]	Se [%]	I [%]
1	11.55	23.90	32.63	31.91
2	12.85	23.92	34.91	29.02
3	13.34	22.95	31.63	32.07
4	12.78	22.70	31.47	33.05

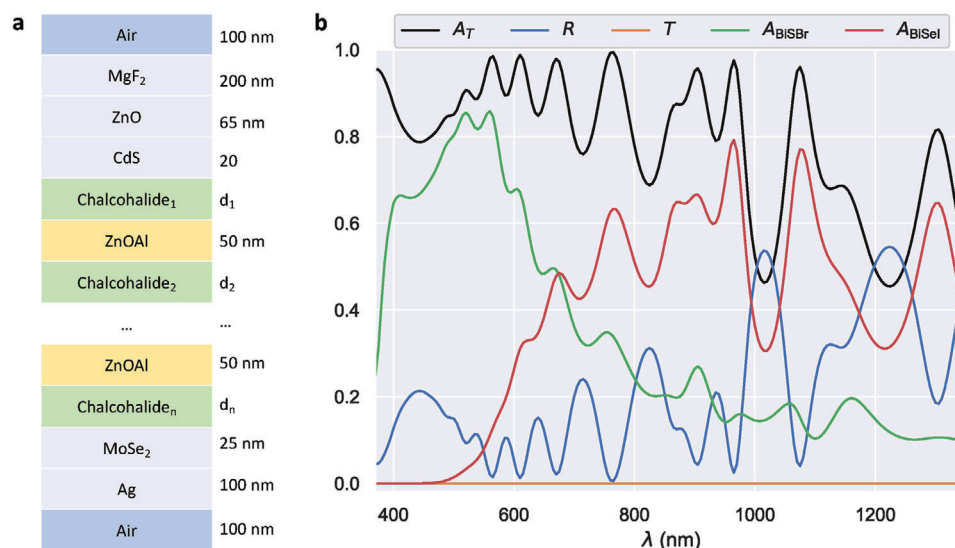


Figure 10. Modeling and optimization of multi-junction solar cell devices based on MChX compounds. a) General structure of the simulated multi-junction solar cell device considering a stacking configuration. The thickness of each chalcogenide layer (d_i) is obtained by an optimization process. The illumination direction goes from the top to bottom of the cell. b) Total absorptivity (A_T), reflectivity (R) and transmissivity (T) of the optimized multi-junction solar cell device, consisting of two layers of pnictogen chalcogenides: BiSBr (600 nm) and BiSeI (800 nm) as the top and bottom layers, respectively.

ever, the high roughness of the sample hindered the acquisition of reliable results, particularly for the absorption coefficient. Nonetheless, the energy bandgap of the synthesized solid solution was precisely determined using photo-thermal deflection spectroscopy (PDS). Specifically, we obtained $E_g^{\text{Expt.}} = 1.42$ eV (Methods and Figure S8, Supporting Information), which is in excellent agreement with the ML-aided first-principles prediction of $E_g^{\text{DFT}} 1.46$ eV.

Overall, based on the experimental information provided in this section, it can be concluded that the degree of reliability associated with our ML-aided first-principles results can be regarded as very high.

2.7. Multi-Junction MChX Solar Cells

The wide range of energy bandgaps exhibited by MChX solid solutions, coupled with their high optical absorption coefficients, positions them as exceptional candidates for the development of highly efficient multi-junction solar cell devices (Figure 10a). The rationale behind the enhanced optical absorption in multi-junction cells compared to single-junction solar cells lies, among other aspects, in the minimization of thermalization losses, a fundamental aspect in the derivation of the Shockley–Queisser limit.

Absorbers situated closer to the light source should possess larger energy bandgaps to absorb the majority of the incoming high-energy radiation, where valence-band electrons are directly promoted to the conduction band. Subsequent layers should feature progressively smaller energy bandgaps to capture the low-energy radiation that the preceding layers were unable to absorb due to their higher energy bandgaps. Moreover, these subsequent layers can capture photons generated through thermalization processes, where electrons initially excited with energies

higher than the absorber bandgap decay to the top of the conduction band, thus emitting low-frequency photons.

However, augmenting the number of layers to enhance overall light absorption leads to diminishing energy absorption in each layer and, due to the serial connection between layers, the current density generated by the entire device is dictated by that of the least efficient absorber (i.e., the intensity matching condition across absorbers). Hence, achieving optimal multi-junction solar cell devices necessitates a delicate balance between the number of considered absorber layers and their intrinsic optoelectronic properties.^[51,52]

Here, we employ macroscopic radiative transfer theory (RTT, Section 2.1.4 and Methods)^[40,41] to model and optimize multi-junction solar cell devices in stacking geometry (Figure 10a), considering MChX parent compounds as the optical absorber layers. The optimization process encompasses both the arrangement of the MChX absorbers in the device and the width of each layer (Methods). We confine the optimization to the parent MChX compounds, as their refractive indexes and extinction coefficients can be directly extracted from our HSEsol+SOC DFT calculations.

As essential components in a multi-junction solar cell, CdS and MoSe₂ were designated as the electron transport layer (ETL) and hole transport layer (HTL) of choice, respectively, facilitating the extraction of charged particles from the device (Figure 10a). Building on previous experimental knowledge, ZnOAl layers were interspersed among the absorber layers, serving as recombination layers (Figure 10a). This arrangement facilitates recombination between the electrons and holes generated in intermediate absorber layers, thereby maximizing the resulting voltage of the device.

Table 3 provides an overview of the results obtained from our multi-junction device optimizations, considering a varying number of absorber layers between 2 and 5. The best

Table 3. Optimized multi-junction solar cell devices as a function of the number of optical absorption layers. Short-circuit current densities estimated for the optimal multi-junction solar cell devices in stacking configuration (Figure 10a) considering from 2 to 5 optical absorption layers.

Layers	Disposition	Thickness [nm]	J_{SC} [mA cm ⁻²]
2	BiSBr	600	18.65
	BiSeI	800	
3	SbSBr	400	12.31
	BiSeBr	500	
	BiSeI	900	
4	SbSBr	400	7.99
	BiSBr	600	
	BiSeBr	800	
	BiSeI	1000	
5	SbSBr	200	6.33
	BiSBr	400	
	SbSeBr	600	
	BiSeBr	1000	
	BiSeI	1000	

overall performance, in terms of the short-circuit current density (J_{SC} , Methods), was achieved by a tandem device comprising a BiSBr (600 nm) and a BiSeI (800 nm) absorber layer. Specifically, based on the simulated absorptivity, reflectivity, and transmissivity (Figure 10b), we estimated an upper limit for J_{SC} of 18.65 mA cm⁻². The total absorptivity, reflectivity, and transmissivity of the remaining multi-junction solar cell devices listed in Table 3 can be found in Figure S9 (Supporting Information).

Several observations emerge from our results. First, it is evident that a considerable portion of the incident light is lost due to excessive reflectance across the spectral range, with an average value of 20% when transitioning from light wavelengths $\lambda = 400$ to 1300 nm. Although an MgF₂ layer is employed here to mitigate reflectance in the visible range, multi-junction designs necessitate, more than single-junction designs, the adoption of a broadband antireflective coating (ARC) to reduce the average reflectivity to below 5%. Therefore, while the calculated current density in the described configuration is 18.65 mA cm⁻², it is crucial to acknowledge that the actual upper limit for a broadband ARC is likely to be significantly higher, potentially exceeding 22 mA cm⁻². This range is commonly regarded as ideal for tandem devices, as recently reported for perovskite/silicon tandem solar cells (specifically, $J_{SC} = 20.99$ mA cm⁻²).^[53]

The alignment of the calculated current density value with the anticipated range, while utilizing film thickness within the typical range for real devices, instills confidence in the optical constants derived from the DFT calculations. However, a significant caveat remains. It is evident that sub-bandgap absorption occurs in both sub-cells (Figure 10b), a phenomenon not expected in actual devices unless a large concentration of defects is present (and here we have considered pristine samples). Such absorption is an artifact of the first-principles calculations. Specifically, DFT calculations typically employ partial smearing in the electronic Fermi-Dirac occupation distribution, as this greatly improves convergence in the electronic self-consistent loops. However, as a consequence, unphysical infra-bandgap absorption is

obtained, which does not occur under realistic ambient conditions. Therefore, while these optical modeling results provide a promising foundation and demonstrate how optoelectronic properties derived from DFT simulations can be utilized, they require cautious interpretation. They also underscore the importance of further experimental validation to refine and corroborate the optical model's predictions, which remains beyond the scope of the present work.

It is important to highlight that for obtaining more quantitative J_{SC} results, directly comparable with experiments, precise determination of the transport properties within the absorber layers is essential. To achieve this level of accuracy, knowledge of the defects energy profile is necessary for estimating recombination rates,^[14] along with key figures of merit such as mobilities to estimate carrier diffusion lengths. However, these investigations extend beyond the scope of the present fundamental analysis and are left for future work.

3. Conclusion

The integration of earth-abundant pnictogen chalcogenides (MChX) into advanced solar cell devices presents promising opportunities due to their high thermodynamic stability, favorable optoelectronic properties and low-temperature synthesis. However, challenges persist, including the vast unexplored landscape of MChX compositions and the disconnect between understanding optoelectronic properties and designing solar cell devices. To address these challenges, we have proposed a solution combining machine learning (ML) and quantum density functional theory (DFT), to efficiently predict the physico-chemical properties of MChX solid solutions, and DFT with macroscopic radiative transfer theory (RTT), to optimize the design of multi-junction MChX solar cell devices.

Our integrated *bottom-up* approach revealed a broad spectrum of energy bandgaps spanning from 1.2 to 2.1 eV and optical absorption coefficients ranging from 25 to 66 μm^{-1} for thermodynamically stable compounds across various stoichiometries. These results were rationalized in terms of electronic band structure features and general composition trends. Experimental synthesis and characterization of a Bi_{0.3}Sb_{0.7}SeI solid solution confirmed the accuracy of our ML-aided DFT predictions. Furthermore, macroscopic RTT modeling identified an optimal multi-junction solar cell design, with a tandem BiSBr–BiSeI device yielding a maximum short-circuit current density of 18.65 mA cm⁻² under intensity-matching conditions.

The findings presented in this study not only advance the fundamental understanding of MChX solid solutions but also establish an innovative and universally applicable *bottom-up* simulation approach for the design of solar cells, showcasing the potential of MChX materials in advancing solar energy conversion technologies.

4. Computational Section

4.1. First-Principles Calculations

Ab initio calculations based on density functional theory (DFT) were performed to analyze the physicochemical properties of

bulk pnictogen chalcogenides. We performed these calculations with the VASP code^[54] using the generalized gradient approximation to the exchange-correlation energy revised for solids due to Perdew et al.^[26] Quantum relativistic effects on the electronic bands were taken into consideration (i.e., spin-orbit coupling effects, SOC) for the calculation of optoelectronic properties along with hybrid functionals (HSEsol+SOC^[27,28]). The projector augmented-wave method was used to represent the ionic cores^[55] and the following electronic states were considered as valence: Bi's 6*p*, 5*d*, 6*s*; Sb's 5*p*, 4*d*, 5*s*; Se's 4*p*, 4*s*; S's 3*p*, 3*s*; I's 5*p*, 5*s*; Br's 4*p*, 4*s*.

Wave functions were represented in a plane-wave basis truncated at 850 eV for the semilocal structural relaxations and at 550 eV for the HSEsol+SOC calculations of optoelectronic properties. By using these parameters and dense *k*-point grids for integration over the irreducible Brillouin zone, the resulting zero-temperature energies were converged to within 1 meV per formula unit. In the geometry relaxations, a tolerance of 0.005 eV Å⁻¹ was imposed in the atomic forces. A primitive cell containing 12 atoms was employed for each compound and the corresponding Brillouin zone was sampled with a 10 × 5 × 4 Γ -centered mesh for the semilocal structural relaxations and 5 × 3 × 2 Γ -centered mesh for the HSEsol+SOC calculations of optoelectronic properties.

To estimate accurate band alignments, we followed the first-principles computational strategy employed in previous works.^[56,57] Briefly, both bulk and slab calculations were performed from which the alignment of the electrostatic potential within the dielectric material can be obtained relative to the vacuum level. From the slab calculations, the average electrostatic potential of the material relative to the vacuum level was obtained. From the bulk calculations, the valence band maximum and conduction band minimum relative to the average electrostatic potential were determined. These calculations involved the estimation of macroscopic and planar average potentials. The planar potential was computed by averaging potential values within a well-defined plane (for instance, perpendicular to the surface of the slab), and the macroscopic potential was obtained by taking averages of the planar potential over distances of one unit cell along the chosen direction. The slab systems should be thick enough to ensure that the electron density in the center of the slab is practically equal to that in the bulk material. We found that ≈15 Å thick slabs accompanied by similarly large portions of vacuum provided sufficiently well converged electrostatic potentials.

4.2. Machine Learning Models

Convolutional neural networks as implemented in the Scikit-learn package^[37] were used for the prediction of the energy and optoelectronic properties of MChX solid solutions. Specifically, a multilayer perceptron regressor with two layers of 32 and 64 neurons each was implemented. The input data comprised the relative abundance of each element in the solid solution and a targeted property. The input data was normalized using a standard scaler. Regarding the hyperparameters, the best performance was achieved with an Adam optimizer with a constant learning rate of 0.001, a rectified linear unit (ReLU) activation function, a mo-

mentum of 0.9, and $\alpha = 0.05$ strength for the L2 regularization term of the loss function. The performance of the models was evaluated through the mean absolute error (MAE) obtained from a *k*-fold cross-validation strategy applied on both the training and test datasets (with *k* = 20, as this is common practice in the literature and provides statistically significant results).

4.3. Experimental Synthesis and Characterization

Bi, Sb, and Se were deposited onto a molybdenum-coated glass substrate by co-evaporation, heating the crucibles at 680, 550, and 250 C°, respectively, while the substrate was kept at 180 C°. Later, the evaporated (Bi, Sb)₂Se₃ precursor was subjected to a high temperature and pressure reactive annealing under halogen atmosphere of BiI₃ and SbI₃, hence obtaining the pnictogen chalcogenide solid solution.^[8]

The optoelectronic properties of the Bi_{0.3}Sb_{0.7}SeI solid solution were experimentally characterized (Figure S8, Supporting Information). Urbach and Tauc models render the lower and upper part of the bandgap, respectively, from photothermal spectroscopy (PDS) measurements according to the formulas:

$$\begin{cases} \alpha d_{\text{Urbach}} = \alpha_0 \exp(E/E_u) & \text{if } E < E_1 \\ \alpha d_{\text{Tauc}} = \beta_0 \frac{\sqrt{E-E_g}}{E} & \text{if } E \geq E_1 \end{cases} \quad (1)$$

where optimal parameters $E_g = 1.40$ eV, $E_1 = 1.44$ eV, $E_u = 0.0868$ eV, $\alpha_0 = 1.17 \cdot 10^{-7}$, and $\beta_0 = 13.9$ were obtained for the synthesized MChX solid solution.

4.4. Optical Solar Cell Device Modeling

The macroscopic radiative transfer theory (RTT) serves as an efficient and systematic approach to simulate complex multi-junction light absorber structures, enabling performance analysis, design optimization, and prediction of system behavior, circumventing the need for computationally demanding simulations of individual layers, getting their absorptivity.^[40,41] In this approach, the short-circuit current density generated by the *n*-th absorber reads:^[41]

$$J_{SC} = q \int_0^\infty A_n(\lambda) \Phi(\lambda) d\lambda \quad (2)$$

where *q* is the electron charge, A_n the absorptivity of the *n*-th layer and Φ the incident solar spectrum. This equation assumes that every absorbed photon contributes to electrical current and generates one electron-hole pair (Shockley–Queisser limit). Consequently, this is an upper limit of the true defect-limited current density of the device, which neither considers saturation current from radiative recombination (which might be considered negligible here) nor non-radiative recombination.

The device optimization process follows a grid search strategy: every possible disposition of 2 to 5 parent pnictogen chalcogenides, ordered in descending energy bandgap, and layer widths ranging from 100 to 500 nm taken in steps of 10 nm. Despite more advanced search algorithms that could be used here, the little resources required for those simulations make grid search still

an easy option. Considering intensity matching, the optimized parameters of the device are selected as those maximizing the output current density of the device, which corresponds to the lowest current density generated by any of the absorbers within the device.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

C.C. acknowledges support from the Spanish Ministry of Science, Innovation and Universities under the fellowship RYC2018-024947-I and grants PID2020-112975GB-I00 and grant TED2021-130265B-C22. The authors also thankfully acknowledge technical support for the computational resources at MareNostrum4 provided by the Barcelona Supercomputing Center (FI-1-0006, FI-2022-2-0003, FI-2023-1-0002, FI-2023-2-0004, and FI-2023-3-0004). C.L. acknowledges support from the Spanish Ministry of Science, Innovation and Universities under an FPU grant. This publication was also supported by the predoctoral program AGAUR-FI ajuts (2024 FI-3 00065) Joan Oró, which is backed by the Secretariat of Universities and Research of the Department of Research and Universities of the Generalitat of Catalonia, as well as the European Social Plus Fund. This work is part of Maria de Maeztu Units of Excellence Programme CEX2023-001300-M funded by MCIN/AEI.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

C.C. and E.S. conceived the study and planned the research, which was discussed in-depth with the rest of the co-authors. C.L. and P.B. performed and analyzed the first-principles calculations. C.L. carried out the data analysis of the generated DFT computations as well as the training of the machine learning models. I.C., D.R., and E.S. synthesized the MChX solid solutions. I.C. and J.-L.T. performed the structural analysis of the MChX solid solutions. J.M.A., D.R., and I.C. measured and analyzed the optoelectronic properties of the MChX solid solutions. Z.J. and C.L. performed the simulations of multi-junction solar cell devices. The manuscript was written by C.C. and C.L. with substantial input from the rest of the co-authors.

Data Availability Statement

The data that support the findings of this study are openly available in GitHub at <https://github.com/IonRepo/ML-PCH>, reference number 0.

Keywords

density functional theory, machine learning, photovoltaics, pnictogen chalcogenides

Received: April 19, 2024

Revised: June 30, 2024

Published online:

- [1] W. Hermann, *Energy* **2006**, *31*, 1685.
- [2] L. M. Peter, *Philos. Trans. R. Soc. A* **2011**, *369*, 1840.
- [3] M. A. Green, E. Dunlop, G. Siefer, M. Yoshita, N. Kopidakis, K. Bothe, X. Hao, *Prog. Photovolt. Res. Appl.* **2023**, *31*, 3.
- [4] NREL, Best research-cell efficiencies, <https://www.nrel.gov/pv/cell-efficiency.html>, (accessed: May 2024).
- [5] W. Shockley, H. J. Queisser, *J. Appl. Phys.* **1961**, *32*, 510.
- [6] M. A. Green, *Third Generation Photovoltaics: Advanced Solar Energy Conversion*, Springer, Berlin **2023**.
- [7] M. Yamaguchi, F. Dimroth, J. F. Geisz, N. J. Ekins-Daukes, *J. Appl. Phys.* **2021**, *129*, 240901.
- [8] I. Caño, A. Navarro, E. Maggi, M. Barrio, J.-L. Tamarit, S. Svatek, E. Antolín, S. Yan, E. Barrena, B. Galiana, M. Placidi, J. Puigdollers, E. Saucedo, *J. Mat. Chem. A* **2023**, *11*, 17616.
- [9] S. Li, Z. Huang, Y. Ding, C. Zhang, J. Yu, Q. Feng, J. Feng, *Small* **2023**, *19*, 2306964.
- [10] U. V. Ghorpade, M. P. Suryawanshi, M. A. Green, T. Wu, X. Hao, K. M. Ryan, *Chem. Rev.* **2023**, *123*, 327.
- [11] J. He, X. Hu, Z. Liu, W. Chen, G. Longo, *Adv. Funct. Mater.* **2023**, *33*, 2306075.
- [12] R. Nie, J. Im, S. I. Seok, *Adv. Mater.* **2019**, *31*, 1808344.
- [13] Y. Li, S. Wang, J. Hong, N. Zhang, X. Wei, T. Zhu, Y. Zhang, Z. Xu, K. Liu, M. Jiang, H. Xu, *Small* **2023**, *19*, 2302623.
- [14] A. M. Ganose, S. Matsumoto, J. Buckridge, D. O. Scanlon, *Chem. Mater.* **2018**, *30*, 3827.
- [15] D. Tiwari, F. Cardoso-Delgado, D. Alibhai, M. Mombrú, D. J. Fermín, *ACS Appl. Energy Mater.* **2019**, *2*, 3878.
- [16] X. Guo, Y.-T. Huang, H. Lohan, J. Ye, Y. Lin, J. Lim, N. Gauriot, S. J. Zelewski, D. Darvill, H. Zhu, A. Rao, I. McCulloch, R. L. Z. Hoye, *J. Mater. Chem. A* **2023**, *11*, 22775.
- [17] E. Fatuzzo, G. Harbeke, W. J. Merz, R. Nitsche, H. Roetschi, N. Ruppert, *Phys. Rev.* **1963**, *127*, 2036.
- [18] K. T. Butler, J. M. Frost, A. Walsh, *Energy Environ. Sci.* **2015**, *8*, 838.
- [19] A. Walsh, D. J. Payne, R. G. Egdell, G. W. Watson, *Chem. Soc. Rev.* **2011**, *40*, 4455.
- [20] A. Mahmood, Y. Sandali, J.-L. Wang, *Phys. Chem. Chem. Phys.* **2023**, *25*, 10417.
- [21] A. Mahmood, A. Irfan, J.-L. Wang, *Chin. J. Polym. Sci.* **2022**, *40*, 870.
- [22] M. Scheidgen, L. Himanen, A. N. Ladines, D. Sikter, M. Nakhaee, Á. Fekete, T. Chang, A. Golparvar, J. A. Márquez, S. Brockhauser, S. Brückner, L. M. Ghiringhelli, F. Dietrich, D. Lehmborg, T. Denell, A. Albino, H. Näsström, S. Shabih, F. Dobener, M. Kühbach, R. Mozumder, J. F. Rudzinski, N. Daelman, J. M. Pizarro, M. Kuban, C. Salazar, P. Ondracka, H.-J. Bungartz, C. Draxl, et al., *J. Open Source Softw.* **2023**, *8*, 5388.
- [23] MChX-SS, <https://dx.doi.org/10.17172/NOMAD/2024.06.29-1> (accessed: May 2024).
- [24] V. Wang, N. Xu, J. C. Liu, G. Tang, W. T. Geng, *Comput. Phys. Commun.* **2021**, *267*, 108033.
- [25] C. Cazorla, J. Boronat, *Rev. Mod. Phys.* **2017**, *89*, 035003.
- [26] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, K. Burke, *Phys. Rev. Lett.* **2009**, *100*, 136406.
- [27] J. Heyd, G. E. Scuseria, M. Ernzerhof, *J. Chem. Phys.* **2003**, *118*, 8207.
- [28] L. Schimka, J. Harl, G. Kresse, *J. Chem. Phys.* **2011**, *134*, 024116.
- [29] J. Shenoy, J. N. Hart, R. Grau-Crespo, N. L. Allan, C. Cazorla, *Adv. Theory Simul.* **2019**, *2*, 1800146.
- [30] V. Tuli, P. A. Burr, A. Claisse, C. Cazorla, *Phys. Rev. Mater.* **2023**, *7*, 113607.
- [31] A. Zunger, S. H. Wei, L. G. Ferreira, J. E. Bernard, *Phys. Rev. Lett.* **1990**, *65*, 353.
- [32] L. Bellaiche, D. Vanderbilt, *Phys. Rev. B* **2000**, *61*, 7877.
- [33] C. Eckhardt, K. Hummer, G. Kresse, *Phys. Rev. B* **2014**, *89*, 165201.
- [34] G. R. Schleder, A. C. M. Padilha, C. M. Acosta, M. Costa, A. Fazzio, *J. Phys. Mater.* **2019**, *2*, 032001.

- [35] X. Liu, J. Zhang, Z. Pei, *Prog. Mater. Sci.* **2023**, *131*, 101018.
- [36] E. Kocer, J. K. Mason, H. Erturk, *J. Chem. Phys.* **2019**, *150*, 154102.
- [37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825.
- [38] C. López, A. Emperador, E. Saucedo, R. Rurali, C. Cazorla, *Mater. Horiz.* **2023**, *10*, 1757.
- [39] IonRepo/ML-PCH, <https://github.com/IonRepo/ML-PCH> (accessed: May 2024).
- [40] N. Dahan, Z. Jehl, J. F. Guillemoles, D. Lincot, N. Naghavi, J.-J. Greffet, *Opt. Express* **2013**, *21*, 2563.
- [41] S. Kim, J. A. Márquez, T. Unold, A. Walsh, *Energy Environ. Sci.* **2020**, *13*, 1481.
- [42] G. A. Niklasson, C. G. Granqvist, O. Hunderi, *Appl. Opt.* **1981**, *20*, 26.
- [43] P. Beckmann, *Prog. Optics* **1967**, *6*, 53.
- [44] V. A. Trifonov, A. V. Shevelkov, E. V. Dikarev, B. A. Popovkin, *Russ. J. Inorg. Chem.* **1999**, *44*, 5.
- [45] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, G. Ceder, *Comput. Mater. Sci.* **2013**, *68*, 314.
- [46] A. M. Ganose, A. J. Jackson, D. O. Scanlon, *J. Open Source Softw.* **2018**, *3*, 717.
- [47] I. Caño, J. W. Turnley, P. Benítez, C. López-Álvarez, J.-M. Asensi, D. Payno, J. Puigdollers, M. Placidi, C. Cazorla, R. Agrawal, E. Saucedo, *J. Mat. Chem. C* **2024**, *12*, 3154.
- [48] H. Fujiwara, M. Kato, M. Tamakoshi, T. Miyadera, M. Chikamatsu, *Phys. Status Solidi A* **2018**, *215*, 1700730.
- [49] A. Polman, M. Knight, E. C. Garnett, B. Ehrler, W. C. Sinke, *Science* **2016**, *352*, 4424.
- [50] J. Rodríguez-Carvajal, *Physica B* **1993**, *192*, 55.
- [51] A. Jimenez-Arguijo, A. Gon-Medaille, A. Navarro-Güell, M. Jimenez-Guerra, K. J. Tiwari, M. Placidi, M. S. Mkehlane, E. Iwuoha, A. Perez-Rodriguez, E. Saucedo, S. Giraldo, Z. J. Li-Kao, *Sol. Energy Mater. Sol. Cells* **2023**, *251*, 112109.
- [52] S. N. Hood, A. Walsh, C. Persson, K. Iordanidou, D. Huang, M. Kumar, Z. Jeh, M. Coure, J. Lauwaert, S. Lee, *J. Phys.:Energy* **2019**, *1*, 042004.
- [53] E. Aydin, E. Ugur, B. K. Yildirim, T. G. Allen, P. Dally, A. Razzaq, F. Cao, L. Xu, B. Vishal, A. Yazmaciyan, A. A. Said, S. Zhumagali, R. Azmi, M. Babics, A. Fell, C. Xiao, S. D. Wolfet, *Nature* **2023**, *623*, 732.
- [54] G. Kresse, J. Furthmüller, *Phys. Rev. B* **1996**, *54*, 11169.
- [55] P. E. Blöchl, *Phys. Rev. B* **1994**, *50*, 17953.
- [56] Z. Liu, B. Wang, D. Chu, C. Cazorla, *J. Mater. Chem. A* **2022**, *10*, 18132.
- [57] Z. Liu, C. Menéndez, J. N. Shenoy, J. N. Hart, C. C. Sorrell, C. Cazorla, *Nano Energy* **2020**, *72*, 104732.