Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Digital Discovery*, 2026, **5**, 1215

Towards an understanding of photoluminescence in lead-free $Cs_2Ag_xNa_{1-x}Bi_yIn_{1-y}Cl_6$ double perovskites by machine learning prediction from density functional theory ground state properties

Marina S. Günthert, *[ab] Larry Lüer, [a] Oleksandr Stroyuk, [c] Oleksandra Raievska, [c] Christian Kupfer, [ac] Andres Osvet, [a] Bernd Meyer [b] and Christoph J. Brabec [acd]

Halide double perovskites are an emerging class of lead-free materials for optoelectronic and photovoltaic applications. Their properties can be tuned by changing ion ratios on different sublattices. An example is $Cs_2Ag_xNa_{1-x}Bi_yIn_{1-y}Cl_6$ (CANBIC), which shows impressive photoluminescence for a small range of compositions. In this work, we combine perfect experimental composition control in high-throughput synthesis with density functional theory (DFT) calculations and machine learning to identify the subspace of optimal ion ratios. For the example of CANBIC, we demonstrate that important excited state parameters determining photoluminescence can be successfully predicted by using only high-throughput DFT ground state data in a two-step machine learning algorithm. This approach reveals the relevant ground state features for the observed photoluminescence and is in accordance with the self-trapped exciton mechanism.

Received 8th December 2025
Accepted 23rd January 2026

DOI: 10.1039/d5dd00152h

rsc.li/digitaldiscovery

1 Introduction

Lead-halide perovskites have undergone extraordinary development over recent years in the field of optoelectronics and particularly photovoltaics (PV), with champion PV cells demonstrating power conversion efficiencies above 25%.[1,2] Nonetheless, they face two major problems: toxicity and stability. Therefore, the need to investigate lead-free alternatives has arisen, for example, through the substitution of Pb^{2+} in $CsPbX_3$ with two heterovalent ions: one monovalent and one trivalent in an alternating manner. This leads to a class of lead-free halide double perovskites with the general formula $A_2B(I)B(III)X_6$. The most popular representatives of this material class are Cs_2AgBiX_6,[3–6] Cs_2NaBiX_6,[7] Cs_2AgInX_6,[8–10] and Cs_2NaInX_6,[11] with X = Cl, Br, I.

Double perovskites have revealed unprecedented compositional flexibility due to the feasibility of independent variation of all four components as well as of alloying of two (and more) components on the B(I), B(III), and X sites.[12–14] The latter approach appeared to be especially fruitful for the design of new optoelectronic materials because alloying on B(I) and B(III) sites breaks the symmetry of the perovskite lattice, resulting in drastic changes in the absorption and emission of light as well as in the dynamics of photo-physical processes.[15]

A particular example of such compositional design is $Cs_2Ag_xNa_{1-x}Bi_yIn_{1-y}Cl_6$ (abbreviated as CANBIC by the first letters of constituent elements). Previous experiments have shown that the variation of the cation ratio at both the B(I) and B(III) sites in CANBIC results in a large variety of double perovskite compositions showing solid solution behavior rather than phase separation.[13,14] The most interesting feature of these materials is the existence of very strong photoluminescence under UV irradiation as reported previously.[13,16–18] The photoluminescence in CANBIC stems from a self-trapped exciton (STE). A STE is generally attributed to a strongly localized exciton (=trapped) in a lattice distortion. Therefore, it occurs in materials with a soft lattice.[19]

In the quest for high-performance, nontoxic PV materials, we need to maximize both energy efficiency (electrical output W_{out} per optical input W_{in}), which requires minimizing non-radiative losses, and external quantum efficiency (electrons out per photons in), which balances photon utilization and contributes to integral power efficiency. Achieving high external quantum efficiency requires a high absorption coefficient and efficient separation of primary photo-excited electron–hole pairs into mobile charge carriers.

[a]*Friedrich-Alexander-Universität Erlangen-Nürnberg, Materials for Electronics and Energy Technology (i-MEET), 91058 Erlangen, Germany*

[b]*Friedrich-Alexander-Universität Erlangen-Nürnberg, Interdisciplinary Center for Molecular Materials (ICMM), Computer Chemistry Center (CCC), 91052 Erlangen, Germany. E-mail: bernd.meyer@chemie.uni-erlangen.de*

[c]*Forschungszentrum Jülich GmbH, Helmholtz-Institut Erlangen-Nürnberg für Erneuerbare Energien (HI ERN), 91058 Erlangen, Germany*

[d]*Forschungszentrum Jülich GmbH, Institute of Energy Technologies (IET), Institute of Energy Materials and Devices (IMD), Photovoltaic (IMD-3), 52428 Jülich, Germany*

To do justice to the broad range of possible compositions, a high-throughput screening approach is needed for finding the best ion ratios in the two B-sublattices. Given the quasi-infinite compositional and processing space, computational pre-screening is mandatory to identify subspaces of high chance to encounter materials with the desired properties. However, the quantum-chemical prediction of the required excited-state properties, such as charge separation and recombination rates, is extremely resource consuming and therefore not apt for high-throughput screening.

For this reason, we explore in this work whether this gap can be bridged using a machine learning model, trained solely on easily accessible ground state data, which we collect from high-throughput density functional theory (DFT) calculations, to predict the excited-state properties of interest measured experimentally (see Fig. 1 for the overall workflow). We choose three target properties that are guaranteed to correlate with PV performance but can be obtained without actually building the devices, thereby strongly accelerating experimental screening of the subspaces identified by computational pre-screening: (i) the radiative recombination rate k_r is linked to the absorption coefficient *via* the Strickler/Berg relation and can be obtained by combining steady-state photoluminescence (PL) and transient photoluminescence (TRPL) measurements; (ii) from the same experiments, the non-radiative recombination rate k_{nr} is also available, controlling energy losses; (iii) from comparing steady-state PL and UV-Vis measurements, the Stokes shift E_s can be obtained.

These three target properties, for which high quality experimental data are available, are predicted by applying Gaussian Process Regression (GPR) as a machine learning method on a large set of ground state DFT data, spanning various structural, mechanical, electronic, and dynamical properties. To avoid overfitting, only such properties are included in the predictive model, which provide additional explanation of variance (mRMR – minimum Redundancy Maximum Relevance).[20] The combination of GPR and mRMR has been used successfully to identify the minimum feature set carrying the essential information to predict degradation in organic photovoltaics.[21] Combining the data driven machine learning (ML) approach with data sets based on experiments and DFT provides a holistic picture of the combinatorial possibilities in a material like CANBIC. We expect this algorithm to be of a general nature for STE-emissive materials and applicable to different compositions and classes of lead-free perovskites.

2 Methodology

2.1 Experimental

2.1.1 Synthesis. The synthesis was performed using the procedure developed by Stroyuk *et al.*[22] For a high-throughput synthesis, the routine was adjusted to a pipetting robot for automation. The synthesis is based on seven solutions: the first three (1–3) make up the first precursor and four more (4–7) make up the second precursor:

- Sol 1: 1 M $InCl_3$ in 4 M HCl.
- Sol 2: 1 M $BiCl_3$ in 4 M HCl.
- Sol 3: 5 : 1 2-propanol : 12 M HCl.
- Sol 4: 20 : 1 2-propanol : 13.4 M HCl.
- Sol 5: 1 M $AgNO_3$ in water.
- Sol 6: water.
- Sol 7: 7.5 : 2.5 : 6 water : 4 M NaAc : 4 M CsAc.

The 6 × 8 well plate geometry gave rise to the idea of the nominal x and y ratios described in Table 1. In the first step, two separate well plates were synthesized, one with varying x ratios for columns 1–8 and one with varying y ratios for rows A–F. The corresponding volumes of the individual solutions present in precursors I and II are shown in Table 1 as well.

In the second step, the two plates are mixed well by well under constant shaking and left overnight for crystallization. After crystallization, the residual solvent was decanted and the

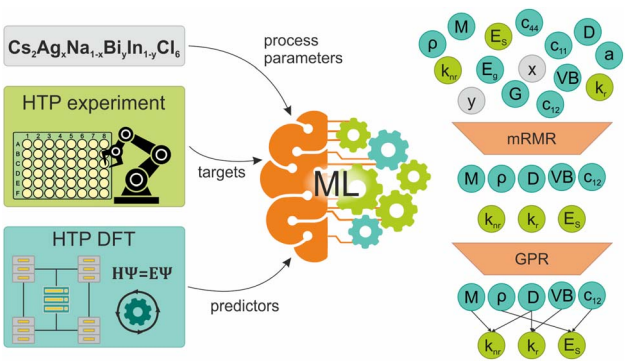

Fig. 1 Schematic workflow: parameters from various sources (gray: synthesis process parameters, green: high-throughput (HTP) experimental data, and turquoise: high-throughput DFT data) are fed into a two-step machine learning (ML) algorithm. The first step uses mRMR to choose the relevant DFT parameters for the experimental photoluminescence data (predictors: HTP DFT data and targets: HTP experimental PL data). Finally, the correlations between predictors and targets are evaluated by GPR.

Table 1 Synthesis amounts of individual solutions (in μl) and exchange ratios x and y on the experimental well plate produced *via* automated robotic synthesis

	Amounts and ratios							
	1	2	3	4	5	6	7	8
Sol 1	100	95	90	75	50	25	10	0
Sol 2	0	5	10	25	50	75	90	100
Sol 3	600	600	600	600	600	600	600	600
y	0.00	0.05	0.10	0.25	0.50	0.75	0.90	1.00

	Amounts and ratios					
	A	B	C	D	E	F
Sol 4	500	500	500	500	500	500
Sol 5	0	10	30	50	80	100
Sol 6	100	90	70	50	20	0
Sol 7	160	160	160	160	160	160
x	0.00	0.10	0.30	0.50	0.80	1.00

particles were resuspended in 1 ml of 2-propanol and centrifuged for 2 min at room temperature and 1000 rpm using a Hermle Z 32 HK. Afterwards, the residual solvent was removed again and the procedure was repeated another time.

The synthesis relies mainly on sodium salts, which results in an overall minimum percentage of sodium in all double perovskites of 15% even after purification (visible from EDX data; see Table S1 in the SI). After the synthesis, the samples for optical and structural analysis were produced by drop-casting of the fresh precipitates suspended in 0.20 ml of 2-propanol on glass (or adhesive carbon tape, respectively). After drying for 2–5 min, a dense and homogeneous layer forms.

2.1.2 Analysis. The EDX analysis was performed using a JEOL JSM-7610F Schottky field emission scanning electron microscope operating under 15–20 kV acceleration voltage equipped with a Deben Centaurus detector (for CL measurements) and an X-Max 80 mm^2 silicon drift detector (Oxford Instruments and AZtec Nanoanalysis software, for EDX measurements). The samples were prepared by drop-casting a suspension in 2-propanol on adhesive carbon tape placed on top of a single silicon crystal and dried under ambient conditions.

The XRD patterns were registered using a Panalytical X'Pert powder diffractometer with filtered Cu K_α radiation ($\lambda = 1.54178$ Å) and with an X'Celerator solid-state stripe detector in the Bragg–Brentano geometry with an angular range of $2\theta = 5$–$100°$ and a step rate of 0.05° per min. Analysis was performed with MAUD software to carry out Rietveld refinement. The samples were drop-cast as suspensions in 2-propanol on glass substrates and dried under ambient conditions.

After excitation with a 75 W xenon lamp from Thorlabs, the reflectance spectra were recorded using a Lack Comet spectrometer from StellarNet Inc. An optical Y-fiber probe with an identical geometry for the sample and the ultra-pure $BaSO_4$ reference from Alfa-Aesar was used for the registration of the spectra. Absorption spectra were then obtained by the division of the reflectance spectra of the sample and the reference, followed by subtracting the baseline.

The same spectrometer used for the reflectance spectra was used for the photoluminescence spectra in the range of 190–1000 nm with a UV LED from Thorlabs (360–370 nm) as the excitation source. For the kinetic curves of PL decay, a custom-designed setup based on a FluoTime300 luminescence spectrometer from PicoQuant GmbH was used. The setup is equipped with a 402 nm LDH-P-C-405B laser, which was used to excite the samples *via* an optical fiber. The PL signal was collected in the range of 420–800 nm with excitation and emission slits set to 4 nm.

By combining stationary measurements of the photoluminescence quantum yield (PLQY) from Stroyuk *et al.*[16] with the kinetic data on PL lifetime τ, the values of the rate constants of the radiative recombination k_r and the non-radiative recombination k_{nr} can be calculated using the following relationships:[16]

$$\tau = \frac{1}{k_r + k_{nr}} \quad \mathrm{PLQY} = \frac{k_r}{k_r + k_{nr}} \quad k_r = \frac{\mathrm{PLQY}}{\tau}. \qquad (1)$$

2.2 Density functional theory

DFT calculations were performed using the periodic plane-wave code PWscf of the Quantum Espresso software package (version 7.0).[23,24] The gradient-corrected exchange–correlation functional of Perdew, Burke and Ernzerhof revised for solids (PBEsol),[25] Vanderbilt ultrasoft pseudopotentials,[26] and a plane-wave basis set with an energy cutoff of 30 Ry were used. In the calculations of optical properties, the ultrasoft pseudopotentials were replaced with SG15 optimized norm-conserving Vanderbilt (ONCV) pseudopotentials and the plane-wave energy cutoff was increased to 60 Ry.[27,28] The RPA frequency-dependent complex dielectric function is calculated using the PBEsol eigenvalues and eigenstates.[23,24]

All parent compounds of CANBIC, *i.e.*, $Cs_2AgBiCl_6$ (CABC), $Cs_2NaBiCl_6$ (CNBC), $Cs_2AgInCl_6$ (CAIC), and $Cs_2NaInCl_6$ (CNIC), have a face-centered cubic (fcc) crystal structure belonging to the space group $Fm\bar{3}m$ (225).[29] The primitive and conventional double perovskite unit cells are shown in Fig. 2. For the calculations of various ground-state properties of CANBIC with different ion ratios x and y, we doubled the primitive fcc unit cell in all crystallographic directions to get an 80 atom ($2 \times 2 \times 2$) fcc supercell (see Fig. 2). This supercell has eight B(I) and eight B(III) sites, thus allowing us to sample altogether 81 different CANBIC compositions. Structure optimization was done by relaxing all atomic positions and the cubic lattice constant using a (2, 2, 2) Monkhorst–Pack k-point mesh for the supercell, *i.e.*, the crystal structure was constrained to the fcc Bravais lattice, but no constraints were imposed on the atomic positions. The atoms were allowed to locally break the cubic symmetry by, for example, random off-centering or octahedra tilts and rotations. The density of the k-point mesh was increased to (4, 4, 4) in the calculations of elastic and optical properties and to (6, 6, 6) in density-of-states (DOS) calculations.

2.3 Machine learning

The minimum Redundancy Maximum Relevance (mRMR) embedded Gaussian Process Regression (GPR) used in this

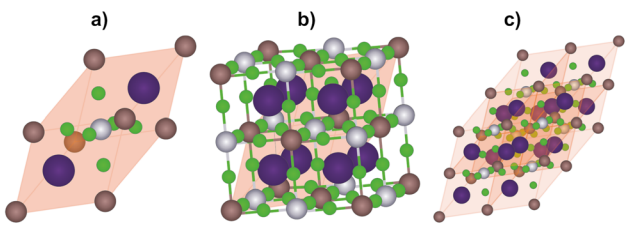

Fig. 2 Double perovskite structure (shown for the specific example of CAIC). (a) Primitive 10 atom fcc unit cell. (b) Conventional 40 atom cubic unit cell. (c) ($2 \times 2 \times 2$) 80 atom fcc supercell used for representing different ion ratios on the B(I) and B(III) sites. Cs: violet, Ag: gray, In: brown, and Cl: green.

work has previously been introduced by Liu *et al.*[21] It is capable of finding the essential predictors for a target quantity from a large number of features. For the evaluation of best predictors, the hierarchical mRMR algorithm was given the following ground state DFT properties to choose from as features:

- Structural features: lattice constant a [Å], unit cell volume V [Å^3], unit cell mass M [u], and density ρ [u Å^{-3}].
- Elastic constants: c_{11} [GPa], c_{12} [GPa], and c_{44} [GPa].
- Derived elastic properties: bulk modulus B [GPa], shear modulus G [GPa], Young's modulus Y [GPa], and Poisson ratio ν.
- Electronic properties: bandgap E_g [eV] and valence band shift VB_{shift} [eV].
- Dynamical properties: Debye temperature Θ_D [K].

The evaluated target properties are the three crucial parameters: Stokes shift E_s and radiative and non-radiative recombination rates k_r and k_{nr}, respectively.

The optimization is performed on the basis of known mRMR implementations,[30–32] but it does not use a dedicated redundancy measure but instead aims for the improvement of the root mean square error. This limits the number of possible predictors to a few hundred due to computational cost, which is still sufficient for the present study.

In addition to predictive uncertainty, model robustness is assessed using a bootstrap-like resampling procedure. For each GPR configuration, 50 random train/test splits (75/25) are performed, and only surrogate models with RMSE within 0.8–1.2 times the mean RMSE are retained, thereby excluding failed or degenerate fits. Hyperparameters are optimized using automatic relevance determination (ARD), with a lower bound imposed on standardized length scales to favor monotonic trends.

3 Results and discussion

For a good prediction of the experimentally measured photoluminescence properties by ML from the DFT ground state features, two criteria have to be fulfilled: first, the DFT results have to be in good accordance with experimental results, and second, data have to be continuous functions.

For the prediction of the photoluminescence criteria, the DFT calculations were performed to evaluate different ground-state properties of CANBIC perovskites, including structural properties, in particular the lattice parameter, elastic parameters (elastic constants, bulk, shear, and Young's moduli, and Poisson ratio), optical properties (dielectric function, reflectivity, energy loss function, absorption coefficient, refractive index, optical conductivity, and extinction coefficient), electronic properties (band structure and density-of-states), and dynamical properties (Debye temperature).

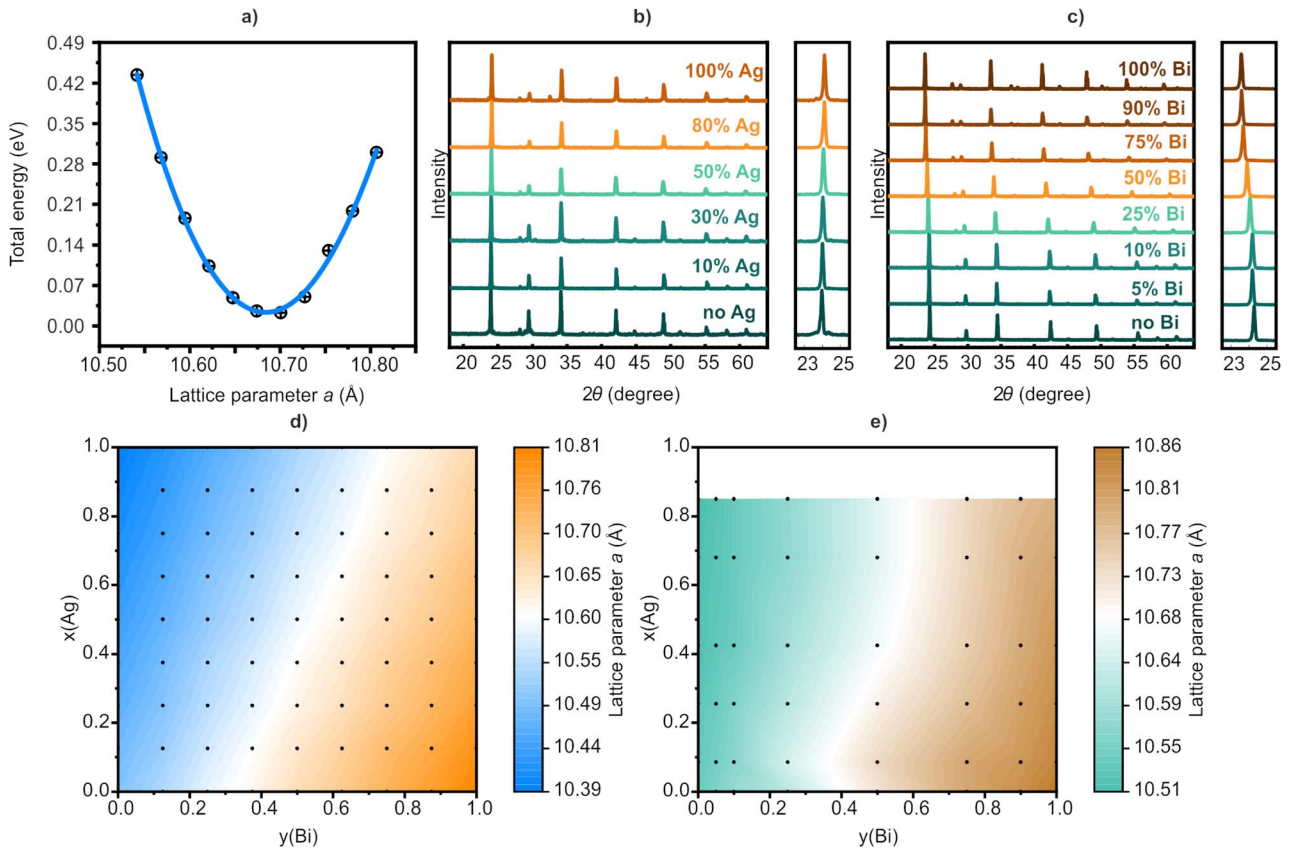

Fig. 3 Structural optimization of CANBIC compounds. (a) DFT total energy as a function of the cubic lattice parameter a for the example of $Cs_2AgBiCl_6$ (CABC). (b) XRD patterns for different Ag ratios; right: zoomed-in region of 23–25°. (c) XRD patterns for different Bi ratios; right: zoomed-in region of 23–25°. (d) 2D map of DFT-computed lattice parameters a. (e) 2D map of measured lattice parameters from XRD.

3.1 Density functional theory

The high-throughput DFT calculations were done with the 80 atom fcc supercell. This allows for eight B(I) and eight B(III) exchange positions and a total of 81 different compositions. For all compositions, the ground-state properties listed above were calculated and the results are presented in 2D colormaps with the B(I) ratio on one and the B(III) ratio on the other axes. Each dot in the map represents one distinct evaluated composition (the same goes for the presentation of the experimental data).

3.1.1 Structural properties. The lattice parameter is optimized by minimizing the total energy of the supercell, as shown in Fig. 3a for one particular composition. The four corners of the compositional space will be referred to as "parent compounds" due to those being pure double perovskites without interchangeable ion ratios. The 2D map in Fig. 3d shows a decline in lattice parameters going from $Cs_2NaBiCl_6$ (CNBC) to $Cs_2AgInCl_6$ (CAIC) with a slope of 0.14 Å for x and 0.28 Å for y. Both this trend and the absolute values are in accordance with experimental values obtained from XRD measurements (Fig. 3b and c) and corresponding Rietveld refinement. From the diffractograms, it is clear that all compositions with both varying Ag (Fig. 3b) and Bi (Fig. 3c) contents display the same pattern with only slight shifts in peak positions, as can be seen from the enlarged insets on the right side. This confirms the solid-solution behavior with changing lattice parameter for the varying compositions. The resulting lattice parameters displayed in the 2D map (Fig. 3e) overall fulfill the accordance and continuity criterion, the same as for the DFT counterpart in Fig. 3d.

3.1.2 Elastic and dynamical properties. Elastic properties describe the response of a material to mechanical stress. This is of particular interest for CANBIC considering the STE nature of the photoluminescence. STE is attributed to lattice distortions and therefore is expected to correlate with the elasticity of the material. Due to the cubic symmetry, the relevant elastic constants are c_{11}, c_{12}, and c_{44}. They were calculated by applying small strains to the equilibrium lattice, determining the resulting change in the total energy as described by Mehl *et al.*[33] From these three elastic constants, the derived parameters bulk modulus B, shear modulus G, Young's modulus Y, and Poisson ratio ν are calculated according to[34]

$$B = \frac{c_{11} + 2c_{12}}{3} \tag{2a}$$

$$G = \frac{c_{11} - c_{12} + 3c_{44}}{5} \tag{2b}$$

$$Y = \frac{9BG}{3B + G} \tag{2c}$$

$$\nu = \frac{3B - 2G}{2(3B + G)} \tag{2d}$$

The bulk and shear moduli together with the density ρ of the material determine the transversal and longitudinal velocities

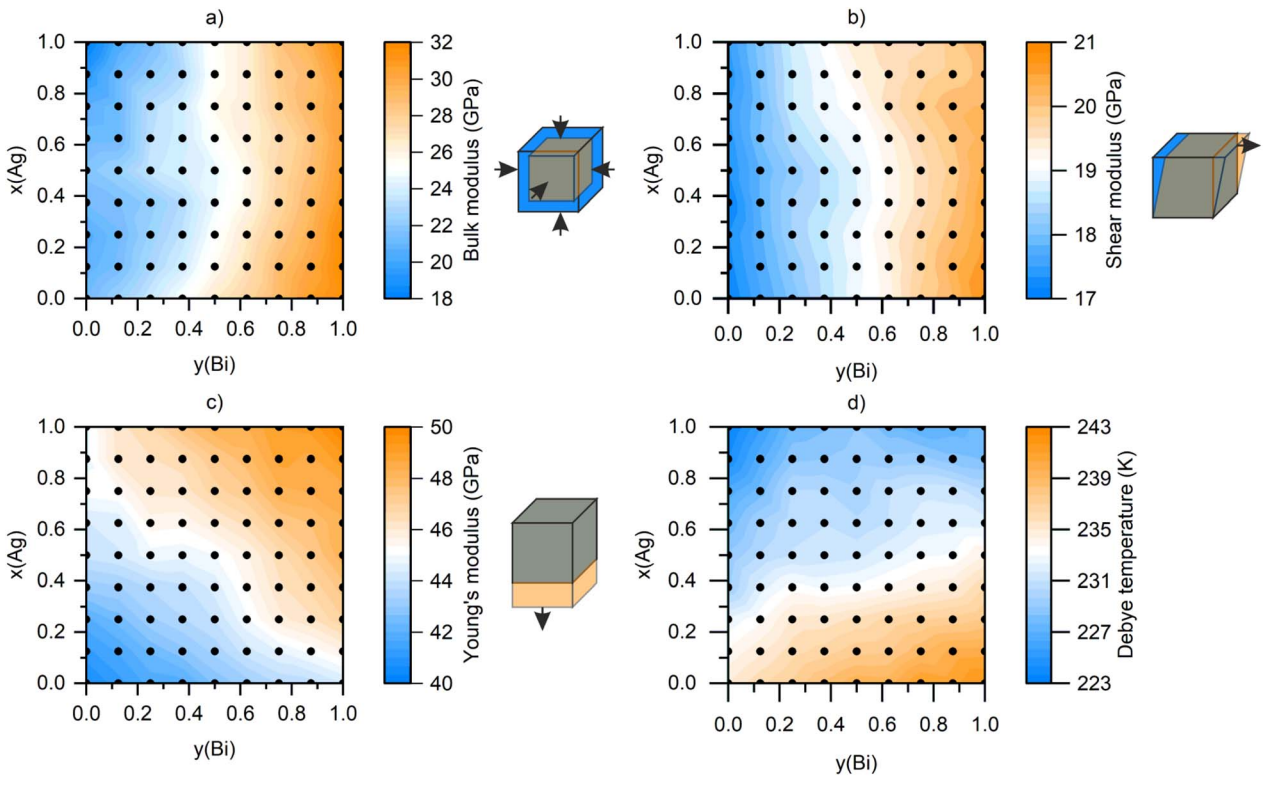

Fig. 4 2D maps of the DFT calculated elastic properties of CANBIC and schematic representation of the deformation. (a) Bulk modulus B, (b) shear modulus G, (c) Young's modulus Y, and (d) Debye temperature Θ_D.

of sound v_t and v_l, respectively, which can be used to derive the Debye temperature Θ_D using a simple Debye model:[34,35]

$$v_t = \sqrt{\frac{G}{\rho}} \quad v_l = \sqrt{\frac{B + \frac{4}{3}G}{\rho}} \quad v_m = \left[\frac{1}{3}\left(\frac{1}{{v_l}^3} + \frac{2}{{v_t}^3}\right)\right]^{-\frac{1}{3}} \tag{3a}$$

$$\Theta_D = \frac{h}{k}\left[\frac{3N_{at}}{4\pi}\left(\frac{\rho}{M}\right)\right]^{\frac{1}{3}} v_m. \tag{3b}$$

N_{at} is the number of atoms in the unit cell and M is their mass.

Fig. 4 depicts both the 2D colormaps for the bulk, shear, and Young's modulus with the corresponding distortion schemes and the 2D colormap for the Debye temperature. The derived elastic and dynamic constants (Fig. 4) show similarly smooth trends like the lattice parameter. The values obtained for the four parent compounds match with literature values using the same technique.[36–38] Both the bulk and shear moduli in Fig. 4a and b are almost invariant in terms of the Ag fraction. The opposite is visible for the Debye temperature, which is almost invariant to the Bi fraction (Fig. 4d). Interestingly, the Young's modulus displays a diagonal shift decreasing with almost the same slope on the x- and y-axes (Fig. 4c). Values for the individual elastic constants and the Poisson ratio can be found in Fig. S1 of the SI. All calculated elastic properties, both the elastic constants and the derived moduli, fulfill the continuity criterion.

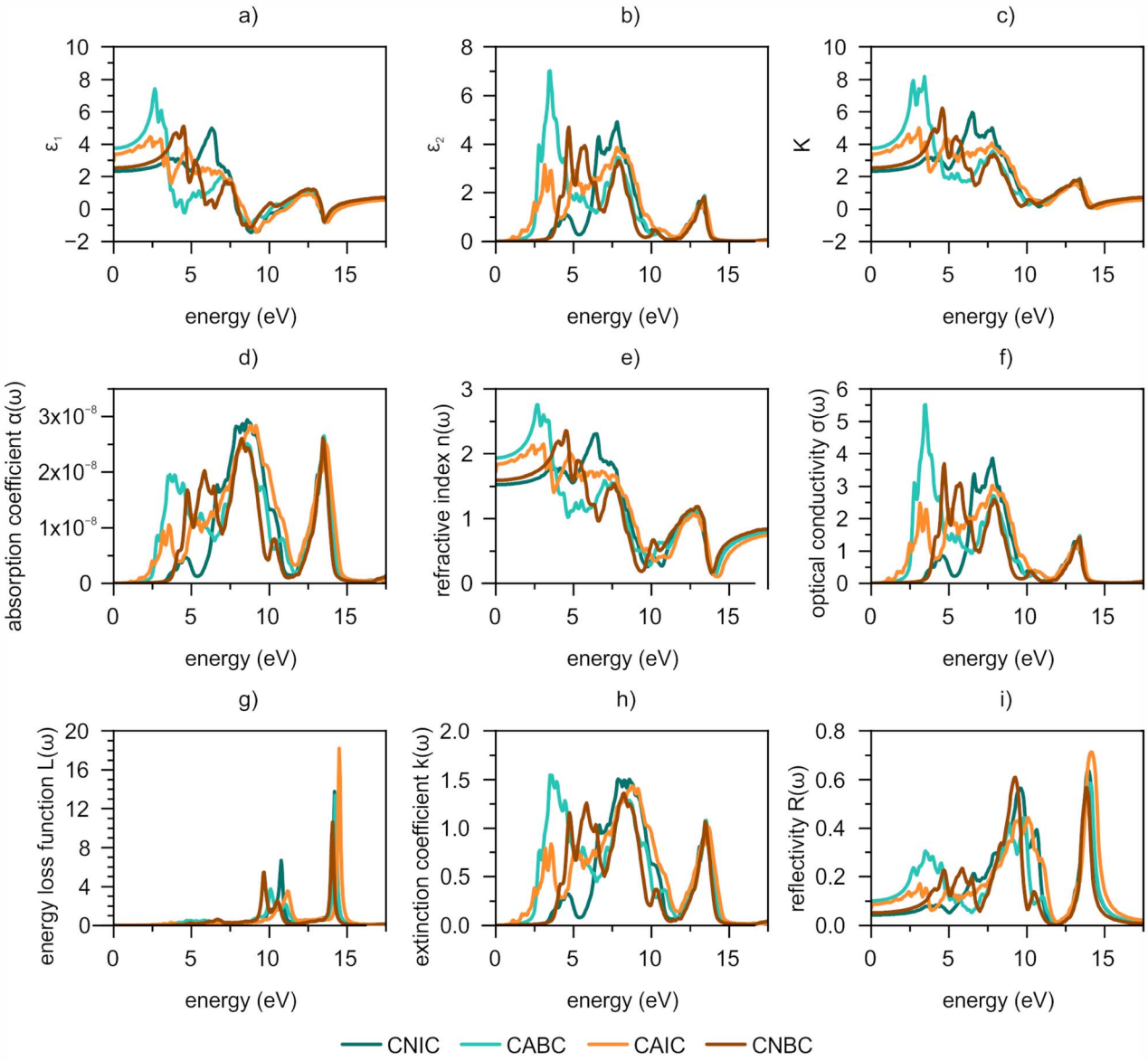

Fig. 5 Dielectric functions and derived properties of CANBIC parent compounds. (a) Real part of the dielectric function ε_1, (b) imaginary part of the dielectric function ε_2, (c) auxiliary factor K, (d) absorption coefficient α, (e) refractive index n, (f) optical conductivity σ, (g) energy loss function L, (h) extinction coefficient k, and (i) reflectivity R.

3.1.3 Optical properties. For the evaluation of an optoelectronic material, optical properties are crucial. The optical properties of CANBIC are obtained by first calculating the frequency-dependent complex dielectric function $\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega)$. From the real and imaginary parts of $\varepsilon(\omega)$, the following optical functions are derived: the absorption coefficient $\alpha(\omega)$, the refractive index $n(\omega)$, the optical conductivity $\sigma(\omega)$, the energy loss function $L(\omega)$, the extinction coefficient $k(\omega)$, and the reflectivity $R(\omega)$:

$$K(\omega) = \sqrt{\varepsilon_1{}^2(\omega) + \varepsilon_2{}^2(\omega)} \tag{4a}$$

$$\alpha(\omega) = \frac{\sqrt{2\omega}}{c}\sqrt{K - \varepsilon_1(\omega)} \tag{4b}$$

$$n(\omega) = \frac{1}{\sqrt{2}}\sqrt{K + \varepsilon_1(\omega)} \tag{4c}$$

$$\sigma(\omega) = \frac{\varepsilon_2(\omega)}{4\pi} \tag{4d}$$

$$L(\omega) = \frac{\varepsilon_2(\omega)}{K^2} \tag{4e}$$

$$k(\omega) = \frac{1}{\sqrt{2}}\sqrt{K - \varepsilon_1(\omega)} \tag{4f}$$

$$R(\omega) = \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2}. \tag{4g}$$

A full set of the derived optical functions for the CANBIC parent compounds is given in Fig. 5. Only the Ag-containing compounds CABC and CAIC show non-zero ε_2 values in the visible range (1.5–3.1 eV). Clearly $\varepsilon_1(0)$ increases from 2.33 to 3.74 eV for CNIC < CNBC < CAIC < CABC. For energies higher than 7.5 eV, all parent compounds show almost the same course for ε_1 and ε_2. The refractive index in the visible range is between 1.5 and 2.0 for all materials. The absorption coefficient shows a steep decrease at $\approx$10.0 eV for all compositions. Each of the compounds has at least two more absorption peaks (Table 2). From the absorption spectra, the bandgap can also be estimated by evaluating the absorption onset (Fig. 6c).

3.1.4 Electronic properties. The onset of the absorption spectra can be directly compared to the bandgap values obtained from the band structure calculations (Fig. 6d) and the experimental bandgap (Fig. 6e). The experimental bandgap was obtained using the measured absorption spectra and applying the indirect Tauc plot shown in Fig. 6f for varying Bi content at x(Ag) = 0.5, revealing a slight decrease in the bandgap with increasing Bi content. On the other hand, both DFT and experimental results show a strong decrease in the bandgap for increasing Ag fraction. Overall, DFT and experiment show the same trends. The main difference is the sample row with y(Bi) = 0, where the experimental values are higher than for y(Bi) > 0, which is not visible in the DFT results, neither from absorption nor from band structure calculations.

Along with the bandgap, additional information can be derived from the electronic structure calculations. To demonstrate the whole variance of the compositional space, two parent compounds with opposing compositions are shown in Fig. 6a and b: CAIC and CNBC. In particular, the data presented here show that the exchange of the B-site ions leads to a significant change in the valence bandwidth from almost 5.0 eV for CAIC to 2.5 eV for CNBC (Fig. 7a). This can be explained by analyzing the atom-resolved projected density-of-states (DOS) diagrams. For CAIC, the valence band is almost entirely composed of Ag- and Cl-related states. For CNBC, this Ag contribution obviously vanishes but a small Bi contribution very close to the valence band edge emerges. Another but less prominent feature is the composition of the conduction band. In both cases, the conduction band is made up of B(III) states, but for Bi there are significantly more unoccupied states than for In. More details on the elemental fractions of the conduction and valence bands over the whole material composition range can be found in Fig. S2 and S3 of the SI.

The position of the Cs-related peak was found to shift from $\approx$6.8 eV for CAIC to $\approx$6.2 eV for CNBC with respect to the valence band edge (Fig. 6a and b). The Cs peak originates from deep, almost core-like states, which are not affected by the composition. The observed shift can therefore directly be attributed to the change in the valence band edge due to the compositional variation. This gives us the opportunity to evaluate the valence band position in correlation to x and y (Fig. 7b). The calculated shifts of the valence band edge can be compared with experimental results from XPS measurements (Fig. 7c).[16] In the experiments, only a small number of compositions were evaluated, allowing a comparison only for a subset of the whole compositional space. However, this comparison shows a perfect alignment of the experimental (black) and theoretical (red) results. Therefore, the graph in Fig. 7c can be used as a calibration to determine the VB position with respect to the vacuum level from the calculated VB edge shifts for the missing experimental data points.

Table 2 Position of the main absorption peaks (in eV)

	CNIC	CABC	CAIC	CNBC
Peak 1	4.69	3.91	3.55	5.83
Peak 2	8.60	8.60	9.02	8.24
Peak 3	13.59	13.53	13.65	13.47

3.2 Photoluminescence data

The experimental photoluminescence data are summarized in Fig. 8. The luminescence is visible for compositions with non-zero Ag and Bi contents. Its bright luminescence under UV illumination is probably the most investigated feature of CANBIC in previous work.[12,16] The photoluminescence spectra (Fig. 8a) show a single broad emission peak that shifts to lower energies for increasing Bi content but is almost invariant for changing Ag content. The width of the emission peak is almost constant (Fig. 8b and c).

For a potential application of CANBIC, it is essential to understand how the luminescence features change with the composition. Therefore, three features, the Stokes shift E_s, the

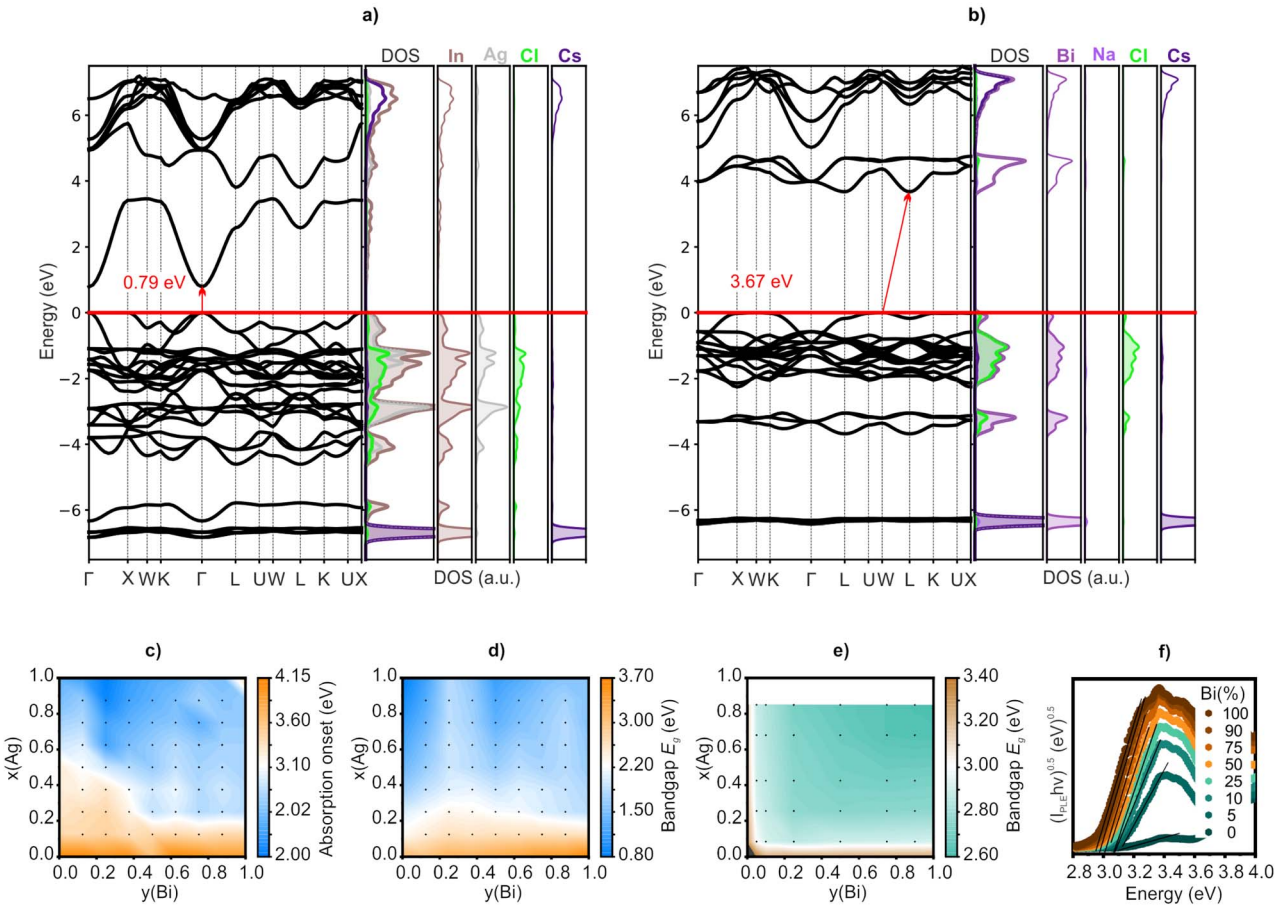

Fig. 6 Band structures of (a) CAIC and (b) CNBC with total and atom-resolved projected DOSs (calculated with the primitive 10 atom fcc unit cell). 2D maps of the electronic properties of CANBIC: (c) absorption onset from Fig. 5d, (d) bandgap from DFT band structure calculations, (e) experimental bandgap from the indirect Tauc plot. (f) Absorbance data of samples with varying Bi content at x(Ag) = 0.5 adapted for indirect Tauc plot analysis.

radiative recombination rate k_r and the non-radiative recombination rate k_{nr}, are investigated.

In Fig. 8g, the kinetic luminescence decay curves of different compositions with varying Bi content are shown. With increasing Bi content, the luminescence lifetime decreases from ≈1300 ns to ≈150 ns. The luminescence lifetime τ was evaluated by using a stretched exponential with exponent b to fit the time t dependent decay curves:[22,39,40]

$$f(t;\tau,b) = e^{-(t/\tau)^{1/b}}. \quad (5)$$

The results are shown in Fig. 8e. The radiative and non-radiative recombination rate constants k_r and k_{nr} are then determined by using eqn (1) (Fig. 8h and i).

Fig. 8d shows the combined PL and absorption spectra of samples with varying Bi content. All display a Stokes shift between 0.6 eV and 1.1 eV, which is quite stable for the

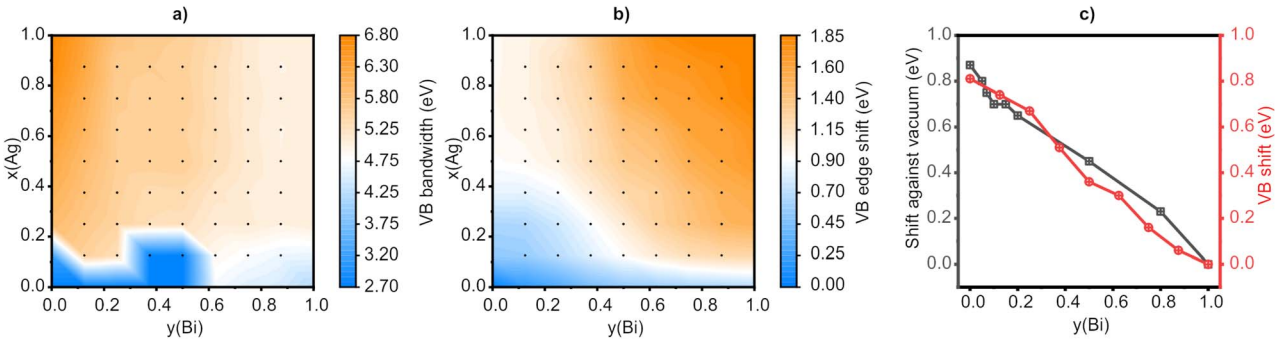

Fig. 7 (a) 2D map of the valence bandwidth. (b) 2D map of the valence band edge shift VB_{shift}. (c) Comparison between valence band shift against vacuum from XPS and DFT-calculated VB shift for varying Bi content at x(Ag) = 0.4 (experimental data from Stroyuk *et al.*[16]).

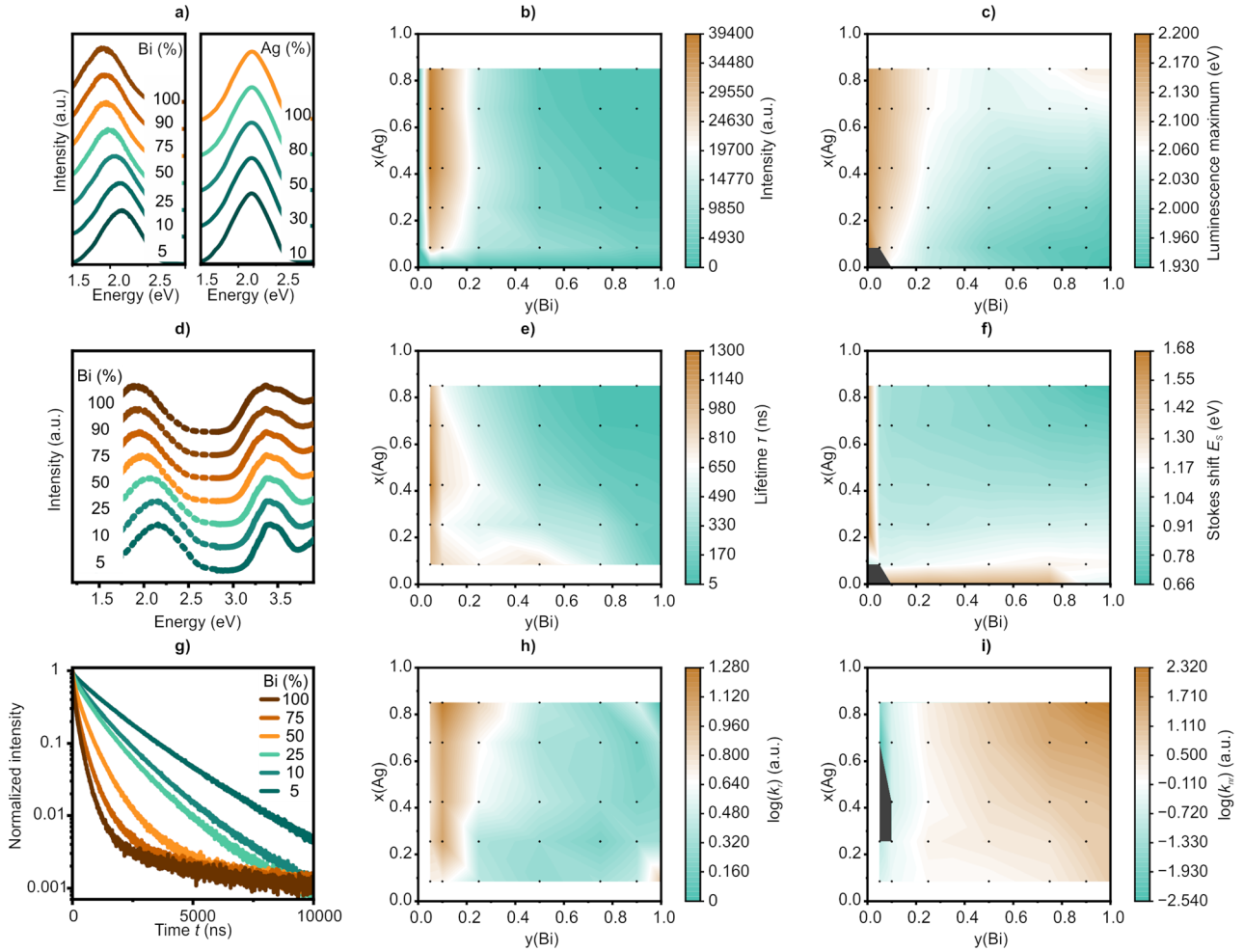

Fig. 8 (a) PL spectra of samples with varying Bi content at x(Ag) = 0.5 and with varying Ag content at y(Bi) = 0.5. (b) 2D map of the photoluminescence intensity. (c) 2D map of the photoluminescence maximum position. (d) Combined photoluminescence and absorption data of samples with varying Bi content at x(Ag) = 0.5 displaying the Stokes shift. (e) 2D map of the photoluminescence lifetime τ. (f) 2D map of the Stokes shift E_s. (g) Transient photoluminescence data of samples with varying Bi content at x(Ag) = 0.5. (h) 2D map of the radiative recombination rate k_r. (i) 2D map of the non-radiative recombination rate k_{nr}.

compositional exchange. Both the PL peak and the absorption edge move to lower energies for increasing Bi content, but the PL movement is more dominant.

CANBIC's highest luminescence intensity is found for $x = 0.4$ and $y = 0.05$ (Fig. 8b). The same composition also possesses the highest luminescence lifetime and radiative recombination rate (Fig. 8e and h). CANBIC is known to be a self-trapped exciton material, which is in accordance with the high PLQY and lifetime observed and explains the prominent Stokes shift.

In CANBIC, the ideal recombination rates (low non-radiative, high radiative) are found for compositions with high Ag (85%) and low (but nonzero!) Bi (10%) content. Unfortunately, at this composition, the Stokes shift is rather high, making exciton dissociation into mobile carriers an inefficient process. As shown in Fig. 8f, minimization of the Stokes shift would require maximizing the Bi content, which however has adverse effects on non-radiative recombination (see Fig. 8i). This means that there exists no composition at which CANBIC will be a good PV material. Using ML, we hope to identify orthogonal DFT predictors for the Stokes shift and the recombination properties, respectively; these would allow us to identify new double perovskites with promising excited state properties as PV materials by computational pre-screening.

3.3 Machine learning: mRMR and GPR

For the prediction of this multi-objective target, multiple steps are necessary. Primarily, the suitable and most relevant DFT-calculated identifiers for the photoluminescence parameters have to be found using a minimum redundancy maximum relevance (mRMR) algorithm. Afterwards, the predictors identified in this way are used to generate a predictive model *via* a regression algorithm, specifically Gaussian Process Regression (GPR).[20] The outcome of this predictive model is a possible indicator of the physics behind the photoluminescence.

To enhance the probability for generalization beyond the available dataset, we restricted the shape of the machine learned surrogate function to monotonous trends. A simple

Table 3 List of the best predictors from mRMR

Target	Best predictors
Stokes shift E_s	Density ρ Elastic constant c_{12}
Radiative rate k_r	Debye temperature Θ_D Valence band shift VB_{shift}
Non-radiative rate k_{nr}	Mass M Debye temperature Θ_D

example is displayed schematically in Fig. 8b: in our experimental dataset, PLQY shows a local maximum for a certain Ag/Na ratio. If we allow local maxima in the surrogate function along any given predictor, then mRMR will "happily" select the experimental Ag/Na ratio itself as a good predictor for PLQY. Even if we exclude experimental parameters from the predictor list, mRMR will still look for DFT properties that are strictly proportional to the Ag/Na ratio. Although this single predictor will correctly predict the Ag/Na ratio of maximum PLQY, it is improbable that it will generalize beyond the CANBIC class. However, if we disallow local extrema, then we force mRMR to identify the individual counter-players causing the local PLQY maximum, namely a single predictor that increases PLQY and another one that decreases it. These predictors have a higher chance to generalize beyond a specific material class.

Feature selection is performed by sequential addition of predictors to the GPR by a greedy algorithm, thereby allowing mRMR to be extended to non-linear correlations: first, GPR is run on all single predictors, and the one with the highest R^2 value is retained. Out of the remaining predictors, the one bringing the strongest increase in R^2 is added. The procedure ends when no further predictor increases R^2 significantly. To test the stability of the resulting surrogate function, each GPR is run 50 times in a bootstrap scheme.

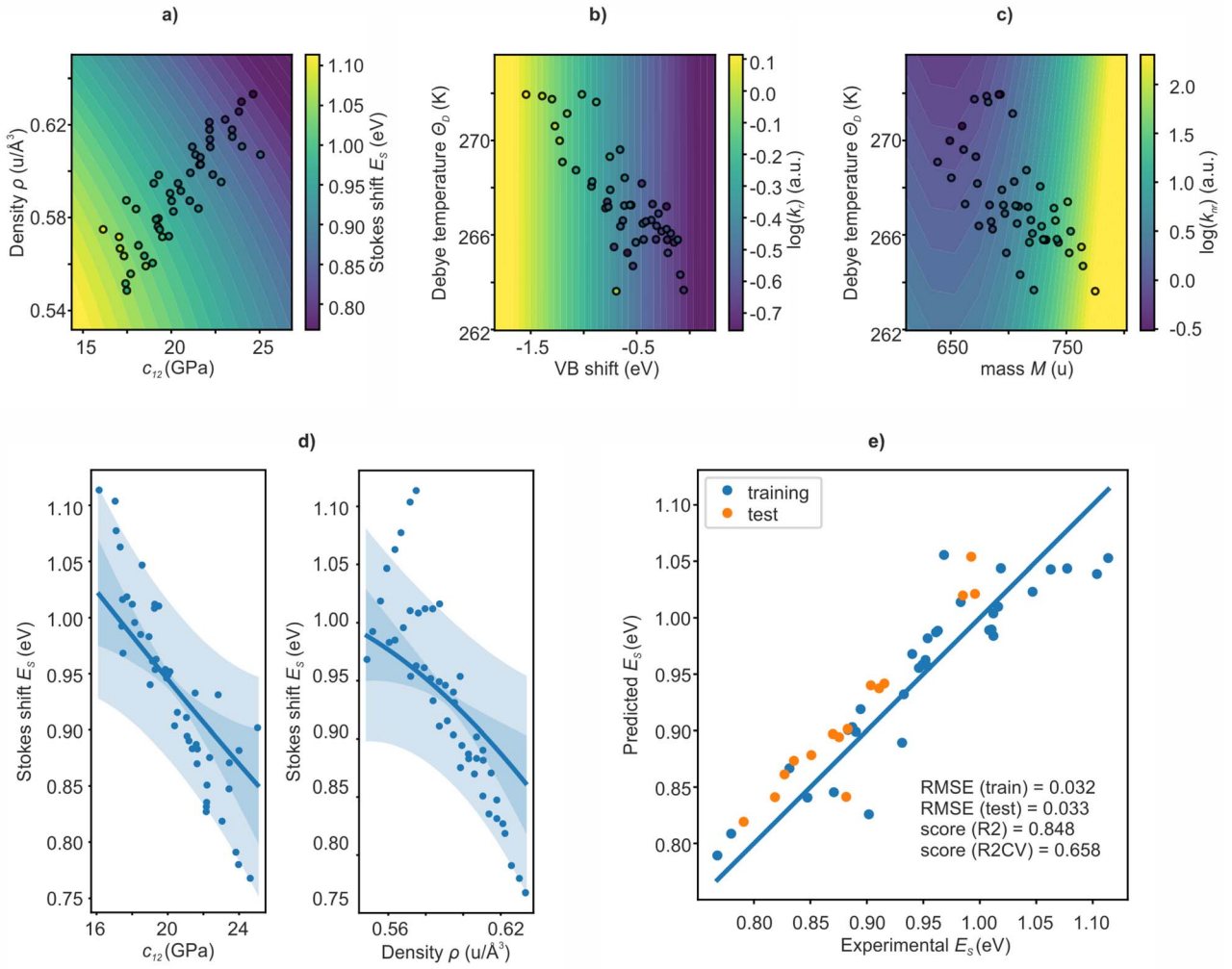

Fig. 9 2D maps of photoluminescence targets predicted from ground state DFT best predictors: (a) Stokes shift E_s (in eV), (b) decadic logarithm of the radiative recombination rate k_r (in a.u.), (c) decadic logarithm of the non-radiative recombination rate k_{nr} (in a.u.). The individual points represent the dataset; the color contour plot displays the prediction. Predictive model output for the prediction of the Stokes shift E_s from the elastic constant c_{12} and density ρ. (d) Single objective functions of individual predictors and targets. (e) Result plot with training and test data evaluation.

The list of best predictors for each target is given in Table 3. These predictors can now be used in a regression algorithm to predict the target properties. Fig. 9a–c show the corresponding surrogate models (colored isocontour) together with the experimental values (symbols). Color matching between symbols and isosurfaces shows the high quality of the prediction, as well as the absence of hidden parameters.

It is apparent that the dataset is not fully independent and isotropically distributed; specifically, there is some correlation between the elastic constant c_{12} and the density ρ in Fig. 9a and between the valence band shift VB_{shift} and the Debye temperature Θ_D in Fig. 9b. However, the spreading of the datapoints across the two-dimensional space is sufficient to quantify individual trends along both dimensions. This can be shown more clearly in one-dimensional intersections through the surrogate function. An example is displayed in Fig. 9d for the Stokes shift E_s; the corresponding plots for k_r and k_{nr} can be found in Fig. S5 of the SI. In Fig. 9d, the surrogate model is shown by a blue line. Light blue regions indicate the 95% confidence interval of a single prediction, as reported by GPR, while dark blue regions identify the 95% confidence interval for the trend. Fig. 9d shows that despite the correlation between c_{12} and ρ, the trends along both dimensions can be assessed with low uncertainty.

The set of decisive and sufficient features in Table 3 exhibits the desired orthogonality, allowing the optimization of one target without penalizing the other. However, with the current dataset, we cannot demonstrate the generalization of these predictors beyond CANBIC. For this reason, we apply domain knowledge to assess their potential for generalization.

3.4 Evaluation of the model

To gain physical insight from a numerical model, a closer look at the chosen best predictors is necessary.

For the Stokes shift in Fig. 9a, the most relevant predictors are the density and one of the elastic constants. Both correlations are monotonic with the elastic constant having a far stronger influence on the target variable. With CANBIC being a self-trapped exciton material and self-trapping being strongly correlated with localized lattice distortions, this correlation is expected but nevertheless gratifying.

The non-radiative recombination rate is almost purely dependent on the molecular mass of the system with a slight deviation depending on the Debye temperature. A smaller molecular mass in this prediction indicates a lower non-radiative recombination rate. On the other hand, the compounds with the smallest molecular weight of the synthesis procedure could not be evaluated in this prediction, as they are not luminescent. Therefore, no rate constants could be determined in this work.

The final target property (the radiative recombination rate) is described again by the Debye temperature and additionally the VB edge shift. This is of special interest, as for none of the targets the PBEsol calculated bandgap was chosen as a predictor by the algorithm, but instead another characteristic of the calculated electronic structure. The accuracy of this parameter has already been shown by the XPS data comparison (Fig. 7c). Again, similar to the non-radiative recombination rate, the Debye temperature is a weaker predictor.

A high radiative and low non-radiative recombination rate originates either from a low number of trap states or a high defect tolerance of the material, mostly attributed to shallow trap states. CANBIC is known to have PLQY $\geq$ 90%, which already indicates a high defect tolerance or shallow traps. This observation is in accordance with CANBIC being a self-trapped exciton material, which is in general highly defect tolerant.

4 Conclusions

Overall, we demonstrated the possibility to investigate lead-free halide double perovskites with interchangeable cation ratios by high-throughput DFT calculations of a broad range of structural, elastic, optical, electronic and dynamical ground state properties. The DFT results are in very good accordance with the available experimental results and at the same time broaden the knowledge with data that could not be obtained experimentally due to the synthesis procedure or the amount of material consumption.

For identifying subspaces of compositions with optimized ion ratios we developed a rapid pre-screening method of potential photovoltaic candidates for double perovskites. By using a GPR with embedded feature selection, we selected the relevant, non-redundant ground state DFT properties for the prediction of fluorescent properties. The model yields high-quality values for the Stokes shift and the radiative and non-radiative recombination rate without the use of sophisticated computational methods, which would be needed to determine excited-state properties.

Finally, we conclude that the Stokes shift is related to the lattice stiffness as expected for a self-trapped exciton material and therefore with a less stiff lattice the charge carrier mobility could be increased. Combining this information and the correlations for the recombination rates, a full set of proxies for the portfolio of requirements for a photovoltaic material is defined.

Author contributions

M. S. Günthert: investigation and writing – original draft; L. Lüer: investigation, methodology, and writing – review and editing; O. Stroyuk: conceptualization and writing – review and editing; O. Raievska: investigation and methodology; C. Kupfer: methodology; A. Osvet: writing – review and editing; B. Meyer: conceptualization, writing – review and editing, and project administration; C. J. Brabec: conceptualization, project administration, and funding acquisition.

Conflicts of interest

There are no conflicts to declare.

Data availability

The ML code as well as the corresponding dataset is available on GitHub at https://github.com/larryluer/mrmr-gpr (building predictive models in a materials science context) and on Zenodo at https://doi.org/10.5281/zenodo.18301315. The experimental data are archived on Zenodo at https://doi.org/10.5281/zenodo.6992560 and the DFT data are published on Zenodo at https://doi.org/10.5281/zenodo.17800198.

Supplementary information (SI): further characterization data of CANBIC, the ML training data, the mRMR training data, and the list of GPR predictions. See DOI: https://doi.org/10.1039/d5dd00152h.

Acknowledgements

The authors gratefully acknowledge financial support for this work by the Deutsche Forschungsgemeinschaft (DFG) under GRK 2495/2 (project number 399073171, projects E and L) and the German Federal Ministry for Economic Affairs and Climate Action (project Pero4PV, FKZ: 03EE1092A).

Notes and references

1 H.-S. Kim, C.-R. Lee, J.-H. Im, K.-B. Lee, T. Moehl, A. Marchioro, S.-J. Moon, R. Humphry-Baker, J.-H. Yum, J. E. Moser, M. Grätzel and N.-G. Park, *Sci. Rep.*, 2012, **2**, 591.
2 F. Qiu, J. Sun, Z. Zhang, T. Shen, H. Liu and J. Qi, *Mater. Today Energy*, 2021, **21**, 100837.
3 H. Absike, N. Baaalla, R. Lamouri, H. Labrim and H. Ez-zahraouy, *Int. J. Energy Res.*, 2022, **46**, 11053–11064.
4 H. C. Sansom, G. Longo, A. D. Wright, L. R. V. Buizza, S. Mahesh, B. Wenger, M. Zanella, M. Abdi-Jalebi, M. J. Pitcher, M. S. Dyer, T. D. Manning, R. H. Friend, L. M. Herz, H. J. Snaith, J. B. Claridge and M. J. Rosseinsky, *J. Am. Chem. Soc.*, 2021, **143**, 3983–3992.
5 D. Gill, P. Bhumla, M. Kumar and S. Bhattacharya, *J. Phys.: Mater.*, 2021, **4**, 025005.
6 F. Lv, T. Zhong, Y. Qin, H. Qin, W. Wang, F. Liu and W. Kong, *Nanomaterials*, 2021, **11**, 1361.
7 R.-I. Biega, Y. Chen, M. R. Filip and L. Leppert, *Nano Lett.*, 2023, **23**, 8155–8161.
8 M. Kumar, M. Jain, A. Singh and S. Bhattacharya, *Appl. Phys. Lett.*, 2021, **118**, 021901.
9 Y. Liu, I. J. Cleveland, M. N. Tran and E. S. Aydil, *J. Phys. Chem. Lett.*, 2023, **14**, 3000–3006.
10 M. Liang, W. Lin, Q. Zhao, X. Zou, Z. Lan, J. Meng, Q. Shi, I. E. Castelli, S. E. Canton, T. Pullerits and K. Zheng, *J. Phys. Chem. Lett.*, 2021, **12**, 4965–4971.
11 Y. Yu, W. Zhou, C. Li, P. Han, H. Li and K. Zhao, *Nanomaterials*, 2023, **13**, 549.
12 J. Luo, X. Wang, S. Li, J. Liu, Y. Guo, G. Niu, L. Yao, Y. Fu, L. Gao, Q. Dong, C. Zhao, M. Leng, F. Ma, W. Liang, L. Wang, S. Jin, J. Han, L. Zhang, J. Etheridge, J. Wang, Y. Yan, E. H. Sargent and J. Tang, *Nature*, 2018, **563**, 541–545.
13 K. Dave, W.-T. Huang, T. Leśniewski, A. Lazarowska, D. Jankowski, S. Mahlik and R.-S. Liu, *Dalton Trans.*, 2022, **51**, 2026–2032.
14 D. Manna, J. Kangsabanik, T. K. Das, D. Das, A. Alam and A. Yella, *J. Phys. Chem. Lett.*, 2020, **11**, 2113–2120.
15 W. Meng, X. Wang, Z. Xiao, J. Wang, D. B. Mitzi and Y. Yan, *J. Phys. Chem. Lett.*, 2017, **8**, 2999–3007.
16 O. Stroyuk, O. Raievska, A. Barabash, C. Kupfer, A. Osvet, V. Dzhagan, D. R. T. Zahn, J. Hauch and C. J. Brabec, *Mater. Adv.*, 2022, **3**, 7894–7903.
17 F. Locardi, E. Sartori, J. Buha, J. Zito, M. Prato, V. Pinchetti, M. L. Zaffalon, M. Ferretti, S. Brovelli, I. Infante, L. De Trizio and L. Manna, *ACS Energy Lett.*, 2019, **4**, 1976–1982.
18 Y. Gao, Y. Pan, F. Zhou, G. Niu and C. Yan, *J. Mater. Chem. A*, 2021, **9**, 11931–11943.
19 S. Li, J. Luo, J. Liu and J. Tang, *J. Phys. Chem. Lett.*, 2019, **10**, 1999–2007.
20 L.-Y. Bi, Y.-Q. Hu, M.-Q. Li, T.-L. Hu, H.-L. Zhang, X.-T. Yin, W.-X. Que, M. S. Lassoued and Y.-Z. Zheng, *J. Mater. Chem. A*, 2019, **7**, 19662–19667.
21 C. Liu, L. Lüer, V. M. L. Corre, K. Forberich, P. Weitz, T. Heumüller, X. Du, J. Wortmann, J. Zhang, J. Wagner, L. Ying, J. Hauch, N. Li and C. J. Brabec, *Adv. Mater.*, 2024, **36**, 2300259.
22 O. Stroyuk, O. Raievska, A. Barabash, M. Batentschuk, A. Osvet, S. Fiedler, U. Resch-Genger, J. Hauch and C. J. Brabec, *J. Mater. Chem. C*, 2022, **10**, 9938–9944.
23 P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal 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 and R. M. Wentzcovitch, *J. Phys.: Condens. Matter*, 2009, **21**, 395502.
24 P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. Buongiorno Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. Dal 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. Otero-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 and S. Baroni, *J. Phys.: Condens. Matter*, 2017, **29**, 465901.
25 J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou and K. Burke, *Phys. Rev. Lett.*, 2008, **100**, 136406.
26 D. Vanderbilt, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1990, **41**, 7892–7895.
27 D. R. Hamann, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **88**, 085117.

28 M. Schlipf and F. Gygi, *Comput. Phys. Commun.*, 2015, **196**, 36–44.
29 Y. Cai, W. Xie, Y. T. Teng, P. C. Harikesh, B. Ghosh, P. Huck, K. A. Persson, N. Mathews, S. G. Mhaisalkar, M. Sherburne and M. Asta, *Chem. Mater.*, 2019, **31**, 5392–5401.
30 Y. Zhu, Y. Zhu, X. Huang, J. Chen, Q. Li, J. He and P. Jiang, *Adv. Energy Mater.*, 2019, **9**, 1901826.
31 Y. Wei, Z. Chen, G. Lu, N. Yu, C. Li, J. Gao, X. Gu, X. Hao, G. Lu, Z. Tang, J. Zhang, Z. Wei, X. Zhang and H. Huang, *Adv. Mater.*, 2022, **34**, 2204718.
32 Z. Zhou, S. Xu, J. Song, Y. Jin, Q. Yue, Y. Qian, F. Liu, F. Zhang and X. Zhu, *Nat. Energy*, 2018, **3**, 952–959.
33 M. J. Mehl, J. E. Osburn, D. A. Papaconstantopoulos and B. M. Klein, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1990, **41**, 10311–10323.
34 M. Faghihnasiri, M. Izadifard and M. E. Ghazi, *J. Phys. Chem. C*, 2017, **121**, 27059–27070.
35 O. L. Anderson, *J. Phys. Chem. Solids*, 1963, **24**, 909–917.
36 M. N. Islam and J. Podder, *Heliyon*, 2022, **8**, e10032.
37 S. Choudhary, S. Tomar, D. Kumar, S. Kumar and A. S. Verma, *East Eur. J. Phys.*, 2021, 74–80.
38 A. A. Sholagberu, W. A. Yahya and A. A. Adewale, *Phys. Scr.*, 2022, **97**, 085824.
39 D. Zhu, J. Zito, V. Pinchetti, Z. Dang, A. Olivati, L. Pasquale, A. Tang, M. L. Zaffalon, F. Meinardi, I. Infante, L. De Trizio, L. Manna and S. Brovelli, *ACS Energy Lett.*, 2020, **5**, 1840–1847.
40 L. Zdražil, S. Kalytchuk, M. Langer, R. Ahmad, J. Pospíšil, O. Zmeškal, M. Altomare, A. Osvet, R. Zbořil, P. Schmuki, C. J. Brabec, M. Otyepka and Š. Kment, *ACS Appl. Energy Mater.*, 2021, **4**, 6445–6453.