PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2024, **26**, 23177

Screening of novel halide perovskites for photocatalytic water splitting using multi-fidelity machine learning†

Maitreyo Biswas, Rushik Desai and Arun Mannodi-Kanakkithodi *

Received 1st July 2024,
Accepted 12th August 2024

DOI: 10.1039/d4cp02330g

rsc.li/pccp

Photocatalytic water splitting is an efficient and sustainable technology to produce high-purity hydrogen gas for clean energy using solar energy. Despite the tremendous success of halide perovskites as absorbers in solar cells, their utility for water splitting applications has not been systematically explored. A band gap greater than 1.23 eV, high solar absorption coefficients, efficient separation of charge carriers, and adequate overpotentials for water redox reaction are crucial for a high solar to hydrogen (STH) efficiency. In this work, we present a data-driven approach to identify novel lead-free halide perovskites with high STH efficiency (η_{STH} > 20%), building upon our recently published computational data and machine learning (ML) models. Our multi-fidelity density functional theory (DFT) dataset comprises decomposition energies and band gaps of nearly 1000 pure and alloyed perovskite halides using both the GGA-PBE and HSE06 functionals. Using rigorously optimized composition-based ML regression models, we performed screening across a chemical space of 150 000+ halide perovskites to yield hundreds of stable compounds with suitable band gaps and edges for photocatalytic water splitting. A handful of the best candidates were investigated with in-depth DFT computations to validate their properties. This work presents a framework for accelerating the navigation of a massive chemical space of halide perovskite alloys and understanding their potential utility for water splitting and motivates future efforts towards the synthesis and characterization of the most promising materials.

1 Introduction

Over the past decade, tremendous efforts have been directed toward developing sustainable technologies and renewable energy sources. To address these critical energy challenges and to minimize fossil fuel dependence, hydrogen fuel can prove to be one of the most efficient alternatives. Hydrogen is much sought after because of its high calorific value and is known to produce water vapor as a byproduct of its combustion. However, harnessing pure hydrogen gas is still one of the key challenges to this technology, which has prevented its industrial adoption.

Photoelectrochemical (PEC) water splitting is one of the most efficient approaches to extract high-purity hydrogen. PEC water splitting aims to split the H_2O molecule into H_2 and O_2 *via* two half-reactions:

(i) Hydrogen evolution reaction (HER):

$$2H^+ + 2e^- \rightarrow 2H_2$$

(ii) Oxygen evolution reaction (OER):

$$2H_2O \rightarrow O_2 + 4H^+ + 4e^-$$

Materials suitable for water splitting should have a band gap greater than 1.23 eV to overcome the thermodynamic barrier of the endothermic water splitting reaction.[1] They should also have a straddling band alignment, *i.e.*, the conduction band minimum (CBM) should be above the reduction potential of H^+/H_2 and the valence band maximum (VBM) should be below the oxidation potential of O_2/H_2O to allow the HER and OER respectively to take place.[2] Incident photons excite electrons to the CB leaving behind holes in the VB, forming electron–hole pairs. The electrons in the CB facilitate the HER whereas holes take part in the OER.

TiO_2 is the most extensively studied photocatalyst because of its photo-chemical stability, corrosion resistance, abundance, and non-toxic nature.[3–8] But anatase and rutile TiO_2 have band gaps of 3.2 eV and 3.0 eV[9] respectively, limiting its photoactivity to the UV-range which is around 5% of the total irradiated solar energy. To narrow the band gap and to enhance the efficiency, numerous methods have been explored, such as

School of Materials Engineering, Purdue University, West Lafayette, IN 47907, USA. E-mail: biswasm@purdue.edu, amannodi@purdue.edu

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4cp02330g

the introduction of metals and non-metals as co-catalysts or dopants,[10–13] creating heterostructures,[14–16] and Z-scheme system construction.[17,18] In spite of these experimental and theoretical investigations, including large-scale synthesis of Z-scheme systems and heterostructures,[19] several challenges remain in TiO_2-based photocatalysis, such as the degradation of PEC efficiency due to the presence of point defects, dopants, and surface additives,[20] and heavy electron effective mass due to localized d-orbitals.[21]

Halide perovskites (HaPs) have been extensively studied for their high photovoltaic (PV) efficiency[22,23] and exciting optoelectronic applications.[24–26] Perovskites prove to be promising materials for photocatalytic water splitting because of their high solar absorption coefficient,[22,23] long electron and hole diffusion lengths,[27,28] long charge carrier lifetimes[28] and easily tunable band gaps for efficient absorption in the visible range of the solar spectrum.[28,29] Liu *et al.* reported a hydrogen evolution rate of 242.5 μmol g^{-1} h^{-1} by splitting H_2O using $CsPbI_3$ combined with graphitic carbon nitride (g-C_3N_4).[30] Fehr *et al.* reported a peak STH efficiency of 20.8% for integrated halide perovskite PEC cells using $Cs_{0.05}FA_{0.85}MA_{0.10}Pb(I_{0.95}Br_{0.05})_3$ and $FA_{0.97}MA_{0.03}PbI_3$ as the photocathode and photoanode respectively.[31] Karuturi *et al.*[32] reported an STH efficiency of over 17% for perovskite/Si dual-absorber tandem cells where a Si photocathode was paired with $Cs_{0.10}Rb_{0.05}FA_{0.75}MA_{0.15}PbI_{1.8}Br_{1.2}$ in tandem. Wang *et al.* implemented a data-driven approach to estimate the photocatalytic performance of lead-free $A_3B_2X_9$ perovskites and reported an STH efficiency of ~17% for $Cs_xRb_{3-x}Bi_yB'_{2-y}X_zX'_{9-z}$ compounds.[33] Thus, it can be well understood that composition engineering at cation or anion sites is an effective way to tune the band gap and enhance the photocatalytic efficiency of HaPs.

Despite these efforts, there are limitations in the detailed understanding of the effects of alloying on the photocatalytic performance of ABX_3 halide perovskites. The chemical space of ABX_3 perovskites comprises millions of possible alloying combinations at A, B, and X sites that would take decades to be screened experimentally. High-throughput DFT (HT-DFT) is one of the most effective ways to explore such combinatorial chemical spaces. HT-DFT combined with state-of-the-art ML models can be used for accelerated screening and discovery of novel stable perovskites with suitable band gaps and photocatalytic efficiencies. This kind of data-driven approach has been used previously by Pilania *et al.*,[34] Jin *et al.*,[35] and Wang *et al.*[33] to screen and identify suitable $AA'BB'O_6$ double perovskite oxides and $A_3B_2X_9$ halide perovskites for PEC water splitting.

In this work, we utilized our recently published multi-phase, multi-fidelity HaP alloy dataset,[36–38] containing 985 individual computations using the GGA-PBE and HSE06 functionals, on pure and alloyed inorganic and hybrid compounds, to screen promising candidates for photocatalytic water splitting. Each perovskite is represented by a 56-dimensional vector, used as the input to train ML predictive models for bulk stability and electronic band gaps and edges. Based on rigorously optimized regularized greedy forest (RGF)[39] models for the decomposition energy (ΔH) and band gap (E_g), prediction and screening were performed across a dataset of 150 000+ enumerated perovskite alloy compositions. For photocatalysis, the band gap and the position of band edges are crucial properties for determining the feasibility of HER and OER. Though the semi-local GGA-PBE[40] functional used for geometry optimization reproduces the lattice parameters and thermodynamic stability quite well, it severely underestimates the band gap.[37,41] Thus our RGF model was trained on a multi-fidelity dataset containing ΔH and E_g from both the GGA-PBE and the hybrid HSE06 functional (HSE),[42] such that the model learns the complex relationship between PBE and HSE band gaps for different perovskite systems. As reported in our recent work, learning from the PBE and HSE data together helps improve chemical space generalizability and prediction accuracy at the HSE-level.[36]

Screening is first performed based on predicted bulk stability and band edges empirically estimated using predicted band gaps and Mulliken electronegativities, following which the η_{STH} is calculated to determine the suitability for water splitting. We further examined the relationship between η_{STH} and material properties such as the band gap and electronegativity. It is found that alloying at the B-site plays a major role in enhancing the photocatalytic performance. Through this work, we present a list of promising HaP compositions for high-efficiency water splitting, including a few stable Pb-free perovskites to mitigate Pb toxicity issues. It is hoped that the insights and results from this computational screening effort will pave the way for future experimental synthesis of efficient halide perovskite-based photocatalysts. Fig. 1 shows an outline of this work, including perovskite descriptors, ML training, and screening across a massive space of possible compositions.

2 Computational methods

2.1 Multi-fidelity dataset for training

Multi-fidelity machine learning (MFML) leverages data from different sources with differing accuracies and computational costs to build efficient surrogate models for each fidelity. MFML models are more robust and generalizable for screening purposes due to their training on diverse data from multiple theoretical levels, enhancing their ability to identify correlations across different fidelities and improving efficiency and accuracy in predictive tasks. Such models exploit the inherent correlations between different data fidelities and are especially useful when high-fidelity data are lacking. As shown in our previous work,[36] training ML models on a combination of GGA and HSE data works better for HSE-level predictions than training on HSE alone, because the relationships between GGA and HSE help make better predictions where HSE data are missing. The multi-fidelity HaP dataset used in this work is compiled from our recently published works[36–38] and consists of 614 data points from PBE and 371 points from HSE. Computations are performed for HaPs in one of four prototype phases, namely cubic, tetragonal, orthorhombic, and hexagonal. The two main target properties are the decomposition

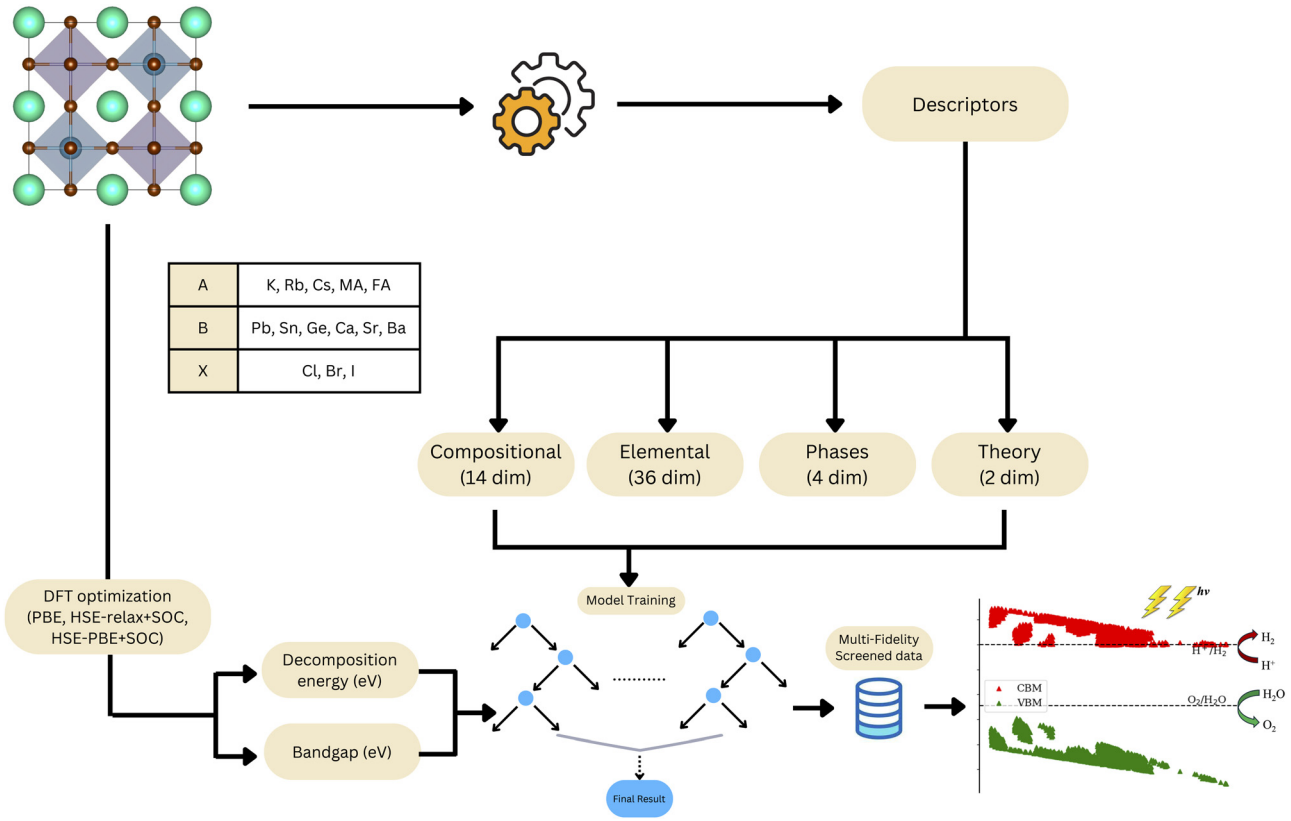

Fig. 1 DFT+ML workflow for multi-fidelity predictions of perovskite properties and screening for suitable photocatalysts.

energy (ΔH, defining the likelihood of ABX_3 decomposition to AX and BX_2 phases, including a configurational entropy term for alloys) and the band gap (E_g). The DFT data are restricted to HaP compositions within the chemical space defined by the A/B/X species pictured in Fig. 1, with mixing at any site only allowed in fractions of $n/8$ (n = 1, 2, 3, ..., 8). Supercells of mixed composition compounds were generated using the Special Quasi-Random Structures (SQS)[43,44] approach. To generate the random alloys, we implemented a simulated annealing[45] approach which iteratively improves the atom rearrangement to minimize the deviation from the perfect random alloy for the target composition. Every data point in the combined PBE + HSE dataset can be represented using a 56-dimensional vector that contains the following information:

(i) Compositional descriptors (14 dimensions). Encoding the HaP composition in terms of fractions (0.0, 0.125, 0.25, ..., 1.0) of the species present at the A/B/X sites.

(ii) Elemental descriptors (36 dimensions). Previously tabulated properties of A/B/X site species, such as the ionic radius and electron affinity. In the case of mixed ions, a weighted mean of the corresponding elemental or molecular properties is used.

(iii) Phase (4 dimensions). One-hot encoding of the perovskite phase (1: cubic, 2: tetragonal, 3: orthorhombic, 4: hexagonal).

(iv) Theory (2 dimensions). One-hot encoding of the level of theory used (1: PBE, 2: HSE) to facilitate multi-fidelity learning, based on the concept of multi-task learning.[46,47]

Table 1 Test RMSE and MAE for decomposition energy and band gap predictions using different functionals

Property	Functional	Test RMSE (eV)	Test MAE (eV)
Decomposition energy	PBE	0.03	0.02
	HSE	0.03	0.02
Band gap	PBE	0.10	0.07
	HSE	0.12	0.08

Further detailed analysis of this dataset can be found in our past publications (Table 1).[36–38]

2.2 ML model training

In this work, we chose regularized greedy forest (RGF) as the regression algorithm of choice to train predictive models for ΔH and E_g on the HT-DFT dataset. Surrogate models based on random forest regression (RFR), XGBoost, and gradient boosting decision trees (GBDT) all provide pretty accurate predictions of perovskite properties.[37,48,49] However, the accuracy of these ensemble-based models vastly depends on the size of the training dataset and is prone to overfitting issues due to the generation of overly complex decision trees.[50] The RGF model outperforms the ensemble models by incorporating tree-structured regularization into the learning formulation and by implementing the fully-corrective regularized greedy algorithm,[39] making it more generalizable. For rigorous training and

optimization of the surrogate models, we applied a 90–10 train–test split, 5-fold cross-validation, and hyperparameter optimization using GridSearchCV. Root mean square error (RMSE) was used as the metric to evaluate model performance. The ultimate goal is to use these surrogate models to screen across thousands of possible compositions to identify suitable materials for photocatalytic water splitting. To further eliminate any test–train split bias, we considered an ensemble of 4000 runs; *i.e.*, the RGF models were trained over a different test–train split in each iteration, and average test set predictions were obtained for each data point over the 4000 runs. All the code is available on GitHub (**https://github.com/maitreyo18/Multi-fidelity-screening-of-perovskite-photocatalysts**).

2.3 Enumerated dataset for prediction and screening

The DFT dataset consists of only a small subset of the combinatorial composition space. To perform a much more exhaustive screening of this space, we enumerated a "hypothetical" HaP dataset. We considered ABX_3 perovskites within the defined set of A/B/X species with mixing at any site in fractions of $n/8$. To keep the dataset tractable, we restricted the mixing to only one site at a time (*e.g.*, when we considered B-site mixing, the A and X sites were unalloyed). This leads to 37 785 unique A-site, B-site, and X-site mixed compositions based on the set of 5 unique A-site cations, 6 B-site cations, and 3 X-site anions shown in Fig. 1. Since each compound could exist in one of four prototype phases, this adds up to 151 140 total compounds. We extracted the 56-dimensional feature vectors for each of these compounds and ultimately fed them into the RGF models for predicting the ΔH and E_g, using averages over the 4000 individual runs as described above, also yielding the prediction uncertainty in terms of the standard deviation.

2.4 DFT details

All computations for validating the ML-screened compounds were performed using the Vienna ab initio simulation package (VASP),[51] employing projector augmented wave (PAW) pseudopotentials.[52,53] Geometry optimization was performed using the Perdew–Burke–Ernzerhof (PBE) functional within the generalized gradient approximation (GGA-PBE),[40] following which a

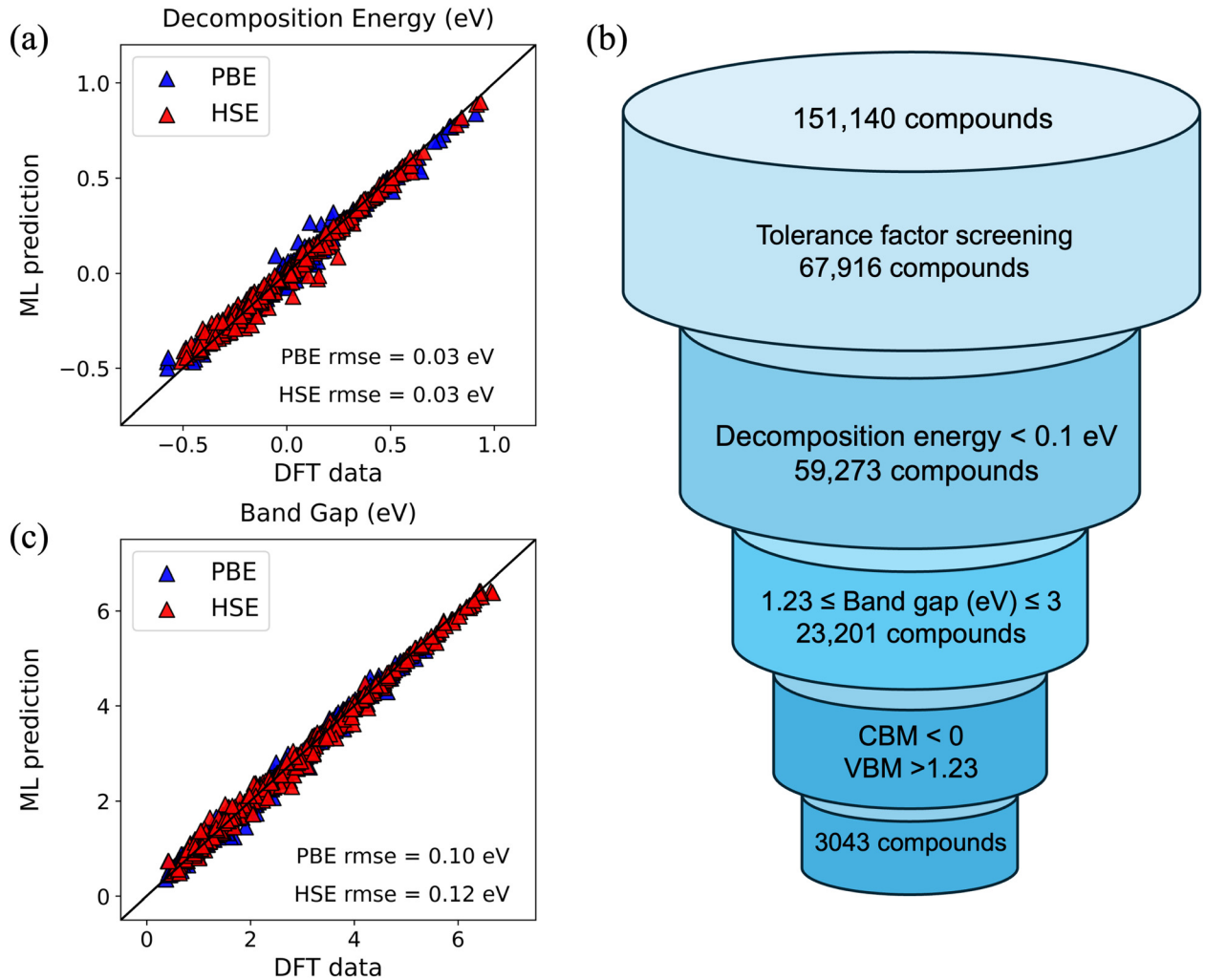

Fig. 2 Parity plots for RGF models showing effective test predictions over 4000 runs plotted against ground truth DFT values for (a) decomposition energy per formula unit and (c) band gap. The screening procedure for identifying suitable perovskites for water splitting is pictured in (b).

single-shot hybrid HSE06[42] (α = 0.25 and ω = 0.2) calculation was performed. A kinetic energy cutoff of 500 eV was used for the plain-wave basis set. For geometry optimization, the Brillouin zone was sampled using a 6 × 6 × 6 Monkhorst–Pack mesh for cubic unit cells, 3 × 3 × 3 mesh for the cubic supercells, and a 2 × 2 × 3 mesh for non-cubic supercells. During the optimization process, the atoms were allowed to fully relax to an energy convergence of 10^{-6} eV and a force convergence of −0.05 eV Å^{-1}. Spin–orbit coupling (SOC) was incorporated in the HSE calculations to capture the relativistic effects due to heavy elements,[54] using the LORBIT tag and the non-collinear magnetic version of VASP.[55] The frequency-dependent optical absorption coefficient $I(\omega)$ for each compound was calculated as:

$$I(\omega) = \frac{\sqrt{2}\omega}{c}\left(\sqrt{\varepsilon_1(\omega)^2 + \varepsilon_2(\omega)^2} - \varepsilon_1(\omega)\right)^{\frac{1}{2}} \quad (1)$$

from the complex dielectric function $\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega)$, using the LOPTICS tag. The VASP outputs were post-processed using VASPKIT.[56]

3 Results and discussion

3.1 Hierarchical screening

Parity plots in Fig. 2(a and b) show the RGF model performance against DFT ground truth, in terms of effective test predictions for all PBE and HSE data points. Our model shows test RMSE values of 0.03 eV and 0.10 eV for ΔH^{PBE} and E_g^{PBE} respectively and 0.03 eV and 0.12 eV for ΔH^{HSE} and E_g^{HSE} respectively. From the parity plots, it is clear that our model shows excellent predictions at both PBE and HSE levels and thus can be generalized to explore unknown compositions. These surrogate models were then used to predict ΔH^{HSE} and E_g^{HSE} for the 151 140 compounds in the enumerated dataset. We employed a hierarchical screening procedure on the enumerated dataset as shown in Fig. 2(b) to identify stable perovskites with suitable band gaps and band edges for water-splitting. To validate the formability of the ABX_3 perovskites, we first performed screening based on the well-known tolerance and octahedral factors which consider the ionic radii of the A, B, and X-site species. In addition to the Goldschmidt tolerance and octahedral factors, we also used a new tolerance factor proposed by Bartel *et al.*[57] The three stability factors are defined as follows:

Octahedral factor:

$$o = \frac{r_B}{r_X} \quad (2)$$

Tolerance factor:

$$t = \frac{r_A + r_X}{\sqrt{2}(r_B + r_X)} \quad (3)$$

Bartel tolerance factor:[57]

$$t_B = \frac{r_X}{r_B} - \left[1 - \frac{\frac{r_A}{r_B}}{\ln\left(\frac{r_A}{r_B}\right)}\right] \quad (4)$$

where r_A, r_B, and r_X represent the ionic radii of A, B, and X-site species respectively. In the case of alloying, the weighted average of the ionic radii is considered.

The accepted upper and lower bounds for the perovskite formability factors are as follows:[36–38] $o \in (0.442 - 0.895)$, $t \in (0.813 - 1.107)$, and $t_B < 4.18$; these conditions are satisfied by 67 916 of the 151 140 compounds. To assess the thermodynamic stability, we used a criterion where perovskites with decomposition energy $\Delta H^{HSE} < 0.1$ eV were accepted as likely being stable. This threshold accounts for potential errors in the machine learning (ML) predicted decomposition energies and includes more candidates. This step left us with 59 273 compounds. Next, to ensure that any compound is able to effectively absorb photons within the visible solar spectrum and to meet the threshold for minimum water electrolysis potential, we applied the condition of $1.23 \leq E_g^{HSE} \leq 3$ eV, reducing the number of compounds to 23 201. In Fig. 3, we visualize the ML-predicted E_g^{HSE} plotted against ΔH^{HSE} for the formable compounds; the shaded region shows where the 23 201 compounds lie.

Next, we must align the electronic band edges of the HaPs with respect to vacuum to determine whether they straddle the redox potentials of water. To do this, we adopted an empirical approach based on the Mulliken electronegativity and (ML-predicted HSE) band gap of the perovskites. The VBM and CBM are calculated as:

$$E_{CBM} = \chi(ABX_3) - E_e - \frac{1}{2}E_g$$
$$E_{VBM} = \chi(ABX_3) + E_g \quad (5)$$

where E_e is the energy of the free electron on the hydrogen scale (4.44 eV) and $\chi(ABX_3)$ is the geometric mean of the Mulliken

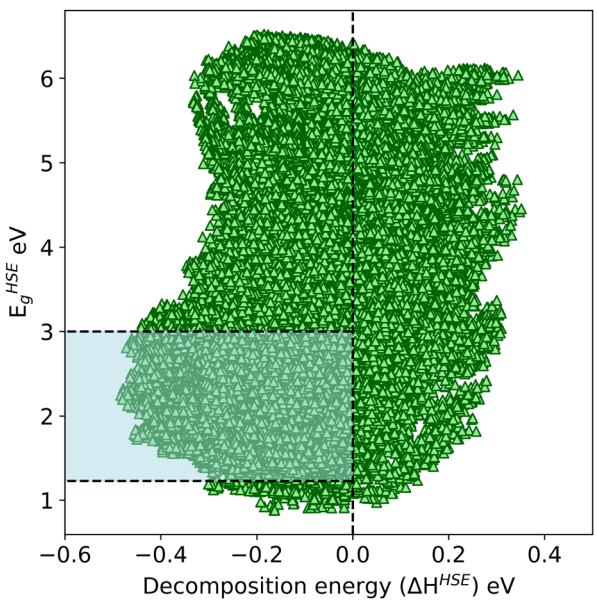

Fig. 3 Visualization of the ML-predicted HSE decomposition energies *vs.* band gaps for 23 201 compounds with desirable octahedral and tolerance factors.

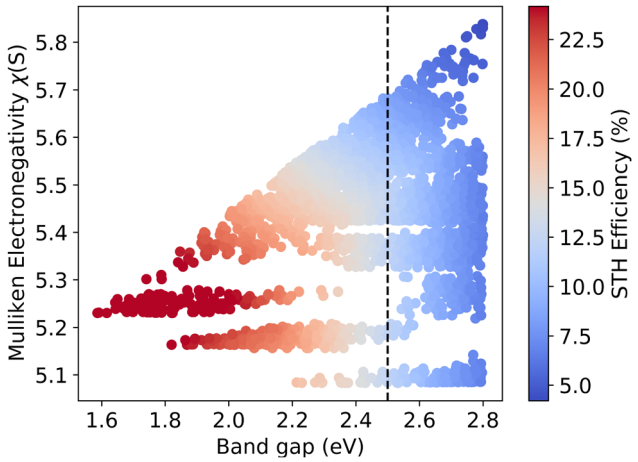

Fig. 4 Dependence of η_{STH} on the band gap and electronegativity of the perovskites. The dotted line shows $\eta_{STH} > 12\%$.

electronegativities of the A(χ_A), B(χ_B), and X-site(χ_X) species, calculated as:

$$\chi(S) = \sqrt[5]{\chi_A \chi_B (\chi_X)^3} \quad (6)$$

It should be noted that the electronegativities of all the A/B/X species used in this work are already tabulated and even used as part of the ML descriptors. This empirical approach has been successfully implemented previously,[33,34,58–61] and the estimated band edges have shown good agreement with experimentally measured VBMs and CBMs.[62] The band edges should have a straddling alignment to allow the HER and OER at the VBM and CBM respectively. Under the normal hydrogen electrode (NHE) standard, $E_{CBM} < 0$ and $E_{VBM} \geq 1.23$ should be satisfied for the necessary alignment. After this final round of band edge screening, 3043 perovskites were identified as suitable water splitting photocatalysts, which is only about 2% of the total number of enumerated compounds. In the next section, we provide further analysis of the screened compounds and DFT validation of a few selected perovskites.

3.2 Statistical analysis and STH efficiency(η_{STH})

Solar-to-hydrogen (STH) efficiency is the metric used to predict the performance of a photocatalyst for water splitting. Theoretical η_{STH} is calculated as:[33,63]

$$\eta_{STH} = \eta_{abs}\eta_{cu} \quad (7)$$

where η_{abs} is the efficiency of light absorption and η_{cu} is the efficiency of carrier utilization. η_{abs} is defined as:

$$\eta_{abs} = \frac{\int_{E_g}^{\infty} P(h\omega)\mathrm{d}(h\omega)}{\int_{0}^{\infty} P(h\omega)\mathrm{d}(h\omega)} \quad (8)$$

where E_g is the material band gap and $P(h\omega)$ is the AM1.5G solar energy flux at photon energy $h\omega$. η_{abs} is essentially the ratio of the power density absorbed by the material to the total power density of sunlight. The carrier utilization efficiency (η_{cu}) is defined as:

$$\eta_{cu} = \frac{\Delta G \int_{E}^{\infty} \frac{P(h\omega)}{h\omega}\mathrm{d}(h\omega)}{\int_{E_g}^{\infty} P(h\omega)\mathrm{d}(h\omega)} \quad (9)$$

where ΔG is the potential difference for the redox water splitting reaction and E is the actual photon energy utilized, which is calculated as:

$$E = \begin{cases} E_g, & (\chi(H_2) \geq 0.2,\ \chi(O_2) \geq 0.6) \\ E_g + 0.2 - \chi(H_2), & (\chi(H_2) < 0.2,\ \chi(O_2) \geq 0.6) \\ E_g + 0.6 - \chi(O_2), & (\chi(H_2) \geq 0.2,\ (\chi(O_2) < 0.6) \\ E_g + 0.8 - \chi(H_2) - \chi(O_2), & (\chi(H_2) < 0.2,\ (\chi(O_2) < 0.6) \end{cases} \quad (10)$$

$\chi(H_2)$ denotes the HER overpotential, *i.e.*, the potential difference between the CBM and the H^+/H_2 potential, and

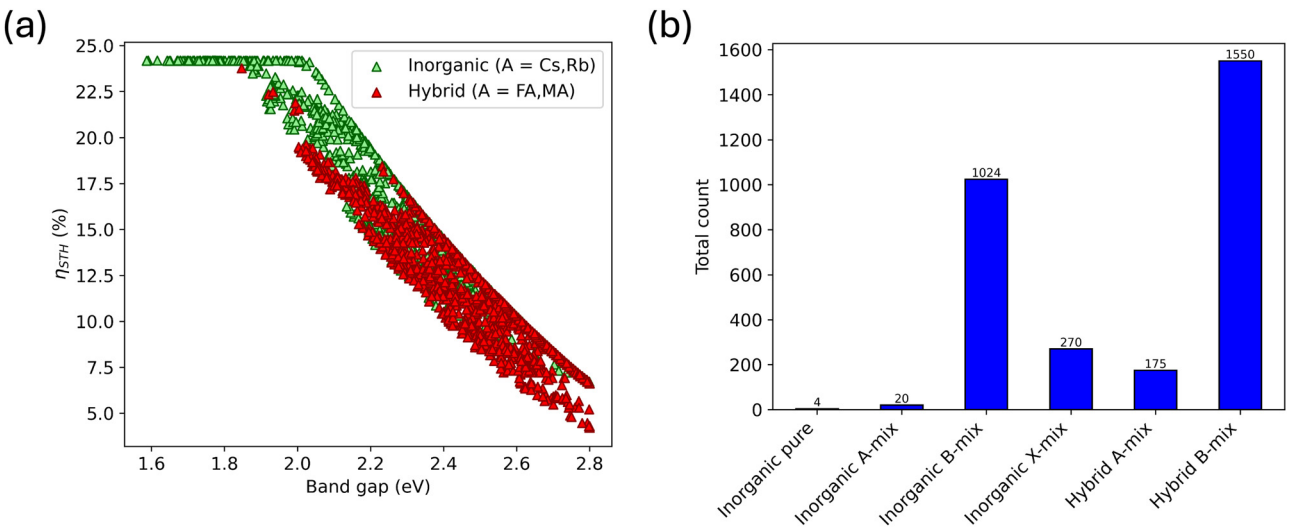

Fig. 5 (a) η_{STH} plotted as a function of band gap for the screened inorganic and hybrid organic–inorganic perovskites. (b) Different kinds of mixing present in the 3043 screened perovskites.

$\chi(O_2)$ denotes the OER overpotential which is the potential difference between the VBM and the O_2/H_2O potential.

Fig. 4 shows a visualization of the η_{STH} values (calculated in %) of the 3043 compounds post-screening, in terms of a plot between the Mulliken electronegativity and the HSE band gap. The truncated region represents perovskites with $\eta_{STH} > 12\%$ to the left. It can be seen that HaPs with band gaps in the range 1.6 eV $\leq E_g^{HSE} \leq$ 2.5 eV show high η_{STH}, clearly attributed to higher solar absorption in the visible spectrum which elevates η_{abs} and thus η_{STH}.

In general, η_{STH} seems to decrease as the electronegativity increases. Fig. 5(a) further shows a plot between η_{STH} and E_g^{HSE}, revealing something interesting: among these stable and formable HaPs with suitable band edges, the highest STH efficiencies are shown by purely inorganic compounds, and hybrid organic–inorganic perovskites (HOIPs) where the A-site contains some mix of MA and FA cations show lower efficiencies. This arises from the fact that in this band gap range, Cs-based inorganic perovskites are the most stable and lie on the lower E_g^{HSE} range thus showing $\eta_{STH} \approx 24\%$, whereas MA/FA-based compounds, which are largely far more stable than Cs/Rb/K-based compounds across the dataset,[36,37] lie in the larger E_g^{HSE} range and thus show $\eta_{STH} < 20\%$ for the majority of HOIPs. Decreasing the band gap in HOIPs below 2 eV also

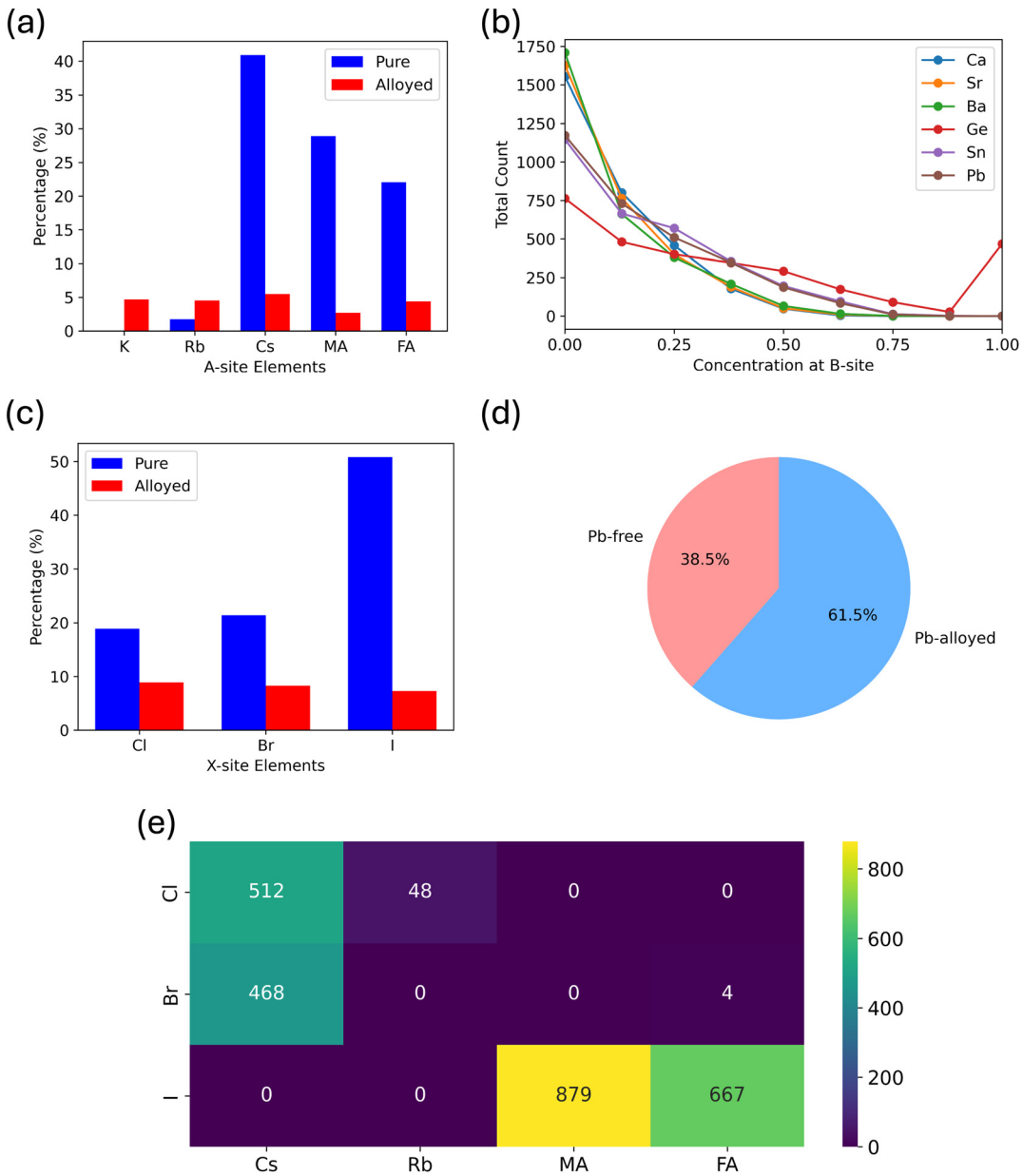

Fig. 6 Statistical analysis of the space of 3043 screened perovskites: (a) percentage of A-site species present. (b) Occurrence frequencies of different mixing fractions at the B-site. (c) Percentage of X-site species present. (d) Distribution of Pb *vs.* Pb-free compounds. (e) Number of perovskite compositions with pure compositions at A and X sites.

shifts the CBM downwards in energy below the H^+/H_2 redox potential, making it unfavorable for the HER.

Next, we discuss the general trends observed in tuning perovskite properties *via* composition engineering. As we go from Cs to Rb to K at the A-site, the cation size decreases, thereby strengthening p–p hybridization and consequently reducing the band gap.[64] The band gap decreases monotonically from Cl to Br to I at the X-site due to the decreasing electronegativity (Cl > Br > I).[65] It is known that B-site and/or X-site substitution are the most common ways to tune the band gap and band edge positions of HaPs, owing to the fact that the CBM and VBM majorly comprise the B-site s, p or d-orbitals and X-site p-orbitals, respectively.[65–67]

Fig. 5(b) shows the distribution of different types of mixing present in the 3043 screened perovskite list. These compounds predominantly involve B-site mixing (85%) in both inorganic HaPs and HOIPs, followed by scarce traces of X-site mixing (9%) and A-site mixing (6%), which corroborates the general trends as discussed. Fig. 6(a) shows that Cs is the A-site cation in a majority of the compounds followed by MA and FA, with only ~2% of the compounds containing Rb or K. The scarcity of A-site mixing in the screened list signifies that the stable perovskites tend to preserve pure compositions at the A-site. The lack of pure K-based or Rb-based perovskites can be attributed to their inherent instability and tendency to decompose.[68,69] Thus, K and Rb are only found as constituents in A-site mixed perovskites.

Fig. 6(b) further shows the prevalence of different mixing fractions of the B-site cations, revealing that mixing of several cations at once (thus forming high-entropy perovskite alloys) is indeed quite favorable, and each of the 6 cations is more likely to appear in smaller mixing fractions than larger quantities. At the X-site (Fig. 6(c)), about three-quarters of the compounds are iodides with the remaining compounds being nearly equally divided between bromides and chlorides. Interestingly, all the X-site mixed perovskites had pure Cs and Ge at the A and B sites respectively in different phases. No chlorides were

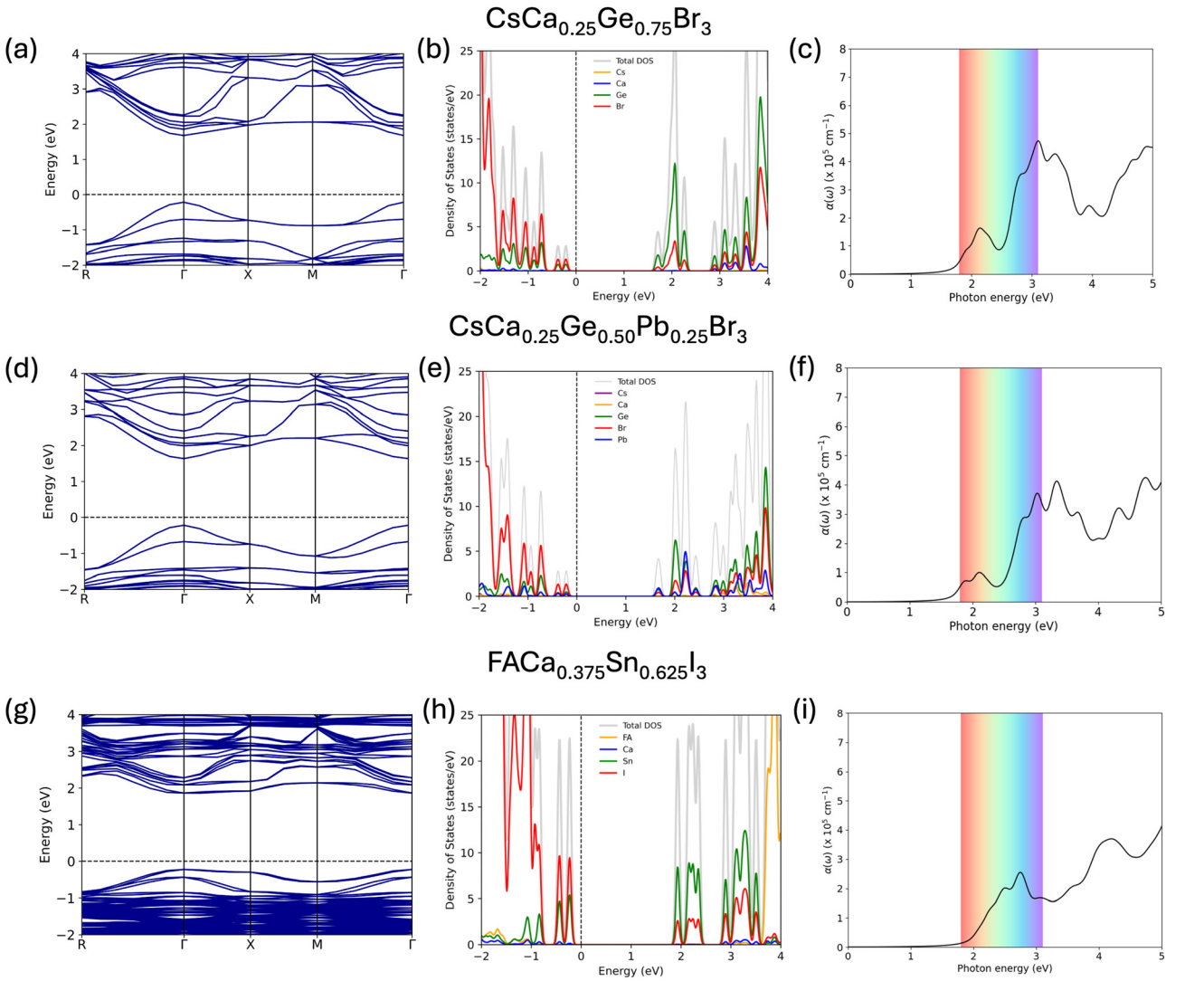

Fig. 7 HSE (HSE06+SOC) calculated band structures (a), (d) and (g), the projected density of states (PDOS) (b), (e) and (h), and optical absorption spectra (c), (f) and (i) for three selected compounds.

identified in combination with FA or MA. The incorporation of Cl in HOIPs either resulted in band gaps exceeding 3 eV or led to higher instability.

Since Pb-free perovskites are much sought after for mitigating Pb-toxicity issues, we performed a visualization of Pb-free *vs.* Pb-containing compounds in Fig. 6(d). We find that 1173 compounds out of 3043 do not contain any Pb at the B-site, constituting about 39% of the space, highlighting a significant exploration into alternative, environment friendly materials for water splitting. Fig. 6(e) shows that the HOIP space in the screened list comprises a majority of MA-I (879) and FA-I (667) compounds, and only four FA–Br compounds, whereas all the purely inorganic compounds are mostly Cs-based bromides (468) and chlorides (512) followed by Rb–Cl (48) compounds.

We find that the most suitable HOIPs with high η_{STH} are substitutional alloys of $FAPbI_3$, $FASnI_3$, and $MAPbI_3$ with alkaline earth metals Ca, Sr, or Ba at the B-site. The lower work function of the alkaline earth metals shifts the CBM which leads to band gap widening.[70] The most promising inorganic compounds are primarily alloys of $CsGeBr_3$ and $CsGeCl_3$ followed by alloys of $CsPbBr_3$ and $CsSnBr_3$. Similar to their hybrid counterparts, substitution with alkaline earth metals in inorganic perovskites widens the band gap and tunes the band alignment to be suitable for photocatalysis. The best STH efficiencies reported in the literature for perovskites lie in the ~20% range;[31,32] the best candidates identified here from the DFT-ML screening approach show efficiencies exceeding 24%, which represents a significant potential improvement in photocatalytic water splitting efficiency.

3.3 DFT validation

We selected five perovskites from the screened list of compounds and performed DFT calculations to validate the ML predictions. Fig. 7 shows the electronic band structure, projected density of states (PDOS) and optical absorption spectra of $CsCa_{0.25}Ge_{0.75}Br_3$, $CsCa_{0.25}Ge_{0.50}Pb_{0.25}Br_3$, and $FACa_{0.375}Sn_{0.625}I_3$, computed using the HSE (HSE06+SOC) functional, considering the cubic phase for all 3 compounds. All the band structures show a direct band gap with both band edges lying at the Γ point. The PDOS plots show expected trends, with a dominance of Ge and Br states in the CB and VB regions in $CsCa_{0.25}Ge_{0.75}Br_3$, states from a combination of multiple B cations and Br in $CsCa_{0.25}Ge_{0.50}Pb_{0.25}Br_3$, and primarily Sn and I states in $FACa_{0.375}Sn_{0.625}I_3$. The absorption spectra further show that all three compounds have large and rising absorption coefficients in the visible range.

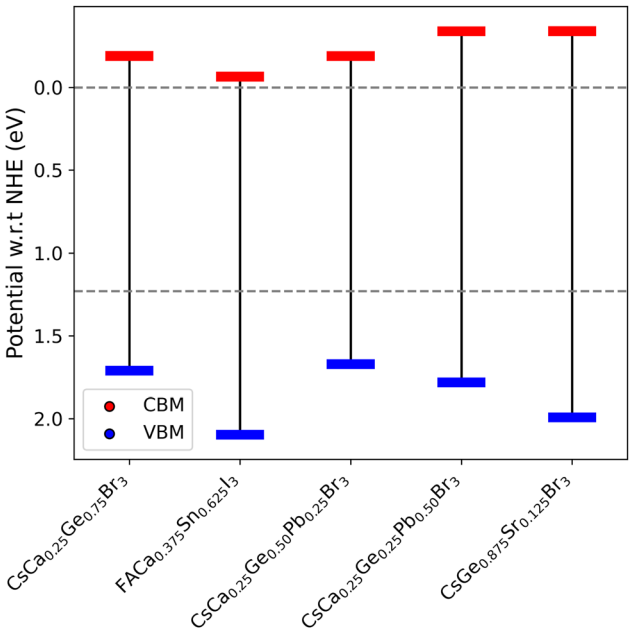

Fig. 8 Relative positions of band edges for 5 selected compounds, estimated empirically from HSE-computed band gaps.

Table 2 summarizes the DFT computed decomposition energies and band gaps and compares them against the ML predictions for the five selected compounds, in cubic and non-cubic phases (selected based on the ML-predicted lowest energy phase). ML predictions for ΔH^{HSE} and E_g^{HSE} are both in good agreement with the DFT values, validating the generalizability and reliability of our surrogate models for novel compositions. The negative (or close to zero) values for ΔH^{HSE} prove the stability of these novel compositions against decomposition into their respective binary AX and BX_2 phases. The band edge positions of these five perovskites relative to the redox potential of water, calculated using the HSE band gaps and eqn (8), are plotted in Fig. 8. Our DFT calculations verify the straddling band alignment of the chosen perovskites which is essential to facilitate the HER and OER processes. Direct band gap photocatalysts typically show higher solar absorption efficiency as compared to indirect band gap compounds because the interband transition of electrons from the VBM to the CBM does not require phonon transport.[35,71] All five perovskites reported in this work showed a direct band gap, which coupled with their high absorption coefficients bode well for efficient light-harvesting within the visible spectrum and subsequent OER

Table 2 DFT computed decomposition energy and band gap compared against the ML predictions for a few selected materials

		DFT calculations				ML predictions		
Compound	Phase	E_g^{HSE}	Gap-type	ΔH^{HSE}	ΔH^{PBE}	E_g^{HSE}	ΔH^{HSE}	ΔH^{PBE}
$CsCa_{0.25}Ge_{0.75}Br_3$	Cubic	1.90	Direct	−0.20	−0.23	2.22	−0.19	−0.23
$FACa_{0.375}Sn_{0.625}I_3$	Cubic (pseudo)	2.16	Direct	0.05	−0.05	2.40	0.05	0.07
$CsCa_{0.25}Ge_{0.50}Pb_{0.25}Br_3$	Cubic	1.86	Direct	−0.24	−0.19	2.14	−0.19	−0.22
$CsCa_{0.25}Ge_{0.25}Pb_{0.50}Br_3$	Tetragonal	2.12	Direct	−0.35	−0.22	2.12	−0.25	−0.21
$CsGe_{0.875}Sr_{0.125}Br_3$	Orthorhombic	2.33	Direct	−0.30	−0.29	2.17	−0.25	−0.26

Table 3 Structure and properties computed for the 5 compounds chosen for DFT validation: lowest energy phase, lattice parameters, electron and hole effective masses, and the STH efficiency

Compound	Phase	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	m_e^*/m_o	m_h^*/m_o	η_{STH} (%)
$CsCa_{0.25}Ge_{0.75}Br_3$	Cubic	11.32	11.32	11.32	90.00	90.00	90.00	0.209	0.528	24.18
$FACa_{0.375}Sn_{0.625}I_3$	Cubic (pseudo)	12.86	12.76	12.87	87.41	95.15	90.61	0.335	0.439	17.82
$CsCa_{0.25}Ge_{0.50}Pb_{0.25}Br_3$	Cubic	11.51	11.53	11.51	90.00	90.00	90.00	0.223	0.492	24.18
$CsCa_{0.25}Ge_{0.25}Pb_{0.50}Br_3$	Tetragonal	16.37	16.38	11.81	90.00	90.00	90.04	1.815	0.971	20.31
$CsGe_{0.875}Sr_{0.125}Br_3$	Orthorhombic	16.46	16.19	11.59	90.07	91.43	89.72	0.247	0.316	16.14

and HER productivity. We also note that three of these compounds are Pb-free perovskites and are thus of particular promise.

Another important aspect of efficient photocatalysis is a low electron effective mass (m_e^*) so as to achieve high charge carrier mobility,[28,35,71] long carrier lifetime,[28,35,71] and efficient electron transfer to facilitate the HER. We calculated m_e^* as well as the hole effective mass (m_h^*) by fitting a parabolic function to the dispersion relation at the CBM and VBM:

$$m_{e/h}^* = \pm\hbar^2\left(\frac{d^2E_k}{dk^2}\right)^{-1} \quad (11)$$

where E_k denotes the band edge eigenvalues and k is the wavevector. The calculated m_e^*, m_h^* and η_{STH} of the five compounds are listed in Table 3, alongside the optimized lattice parameters. The effective masses are primarily determined by the extent of orbital overlap between the B-site and X-site ions.[72] The abnormally high $m_e^* = 1.815m_0$ and $m_h^* = 0.971m_0$ of $CsCa_{0.25}Ge_{0.25}Pb_{0.50}Br_3$ in the tetragonal phase can be attributed to the increased disordering and octahedral tilting due to the mixing of three types of cations at the B-site. In general, in the tetragonal and orthorhombic phases, the orbital overlap between B and X ions is reduced as compared to the cubic phase, which in turn increases m_e^* and m_h^*. The increased disorder in $CsCa_{0.25}Ge_{0.25}Pb_{0.50}Br_3$ due to triple mixing at the B-site distorts the linearity of the B–X–B bonds, reducing the orbital overlap and increasing m_e^* and m_h^*. For the remaining compounds, our computed effective masses are in good general agreement with previously reported values for cubic HaPs.[64,72,73] Among the DFT-validated perovskites, $CsCa_{0.25}Ge_{0.75}Br_3$ and $CsCa_{0.25}Ge_{0.25}Pb_{0.50}Br_3$ show the highest $\eta_{STH} >$ 24%, which is substantially higher than the previously experimentally observed η_{STH} = 20.8%[31] for the Cs–FA–MA–Pb–I HOIP.

4 Conclusions

In this work, we applied a data-driven strategy to explore an alloyed perovskite space consisting of 150 000+ materials and discovered novel compounds for photocatalytic water splitting. This work is built upon a previously published high-throughput multi-fidelity halide perovskite DFT dataset and regularized greedy forest regression models trained on the data. We investigated the generalizability of our DFT-ML surrogate models and successfully validated the best predictions with DFT calculations. This work provides an analysis of the effects of alloying at the A/B/X sites on the thermodynamic landscape and optoelectronic properties of ABX_3 halide perovskites. For identifying suitable perovskites for water-splitting, we employed a hierarchical down-screening approach that filters out compositions based on their tolerance factors, decomposition energy, HSE band gaps, and empirically estimated electronic band edges. Through this approach, we identified 3043 promising materials, most of which are FA-based iodides or Cs-based bromides and contain multiple group II or group IV divalent cations mixed at the B-site.

We find that B-site alloying is the most ideal way to tune perovskite band gaps. Combined with low electron and hole effective masses and a high optical absorption coefficient ($>10^5$ cm^{-1}), these compounds show great promise as efficient photocatalysts. Among the screened perovskites, our DFT computations revealed $CsCa_{0.25}Ge_{0.75}Br_3$ and $CsCa_{0.25}Ge_{0.25}Pb_{0.50}Br_3$ to have a solar-to-hydrogen efficiency >24%, which is notably higher than the previously reported η_{STH} for perovskites both experimentally[31,32] and computationally.[33] The ML-predicted decomposition energies, band gaps and edges, and efficiencies are all made available. Our results also help identify several Pb-free perovskites that may be suitable for water splitting. We hope that this ML-accelerated hierarchical down-screening approach will inspire experimental efforts for validation in the near future. Our predictions and surrogate models are poised to enhance the exploration of this massive perovskite alloy space, enabling more informed and strategic research on perovskite based photocatalysts. As part of future work, the DFT dataset will be extended to more perovskite compositions and alternative ML algorithms will be explored for further improvement.

Author contributions

A. M. K. conceived and planned the research project. DFT computations and ML model training were performed by M. B. and R. D. For the manuscript, M. B. took the lead on writing while A. M. K. performed overall editing and quality control.

Data availability

The corresponding codes, .cif files and ML-predicted ΔH^{PBE}, ΔH^{HSE}, E_g^{PBE}, and E_g^{HSE} of 151 140 perovskites and the band edges and η_{STH} derived from the band gaps of all the 3043

screened perovskites can be found on Github: **https://github.com/maitreyo18/Multi-fidelity-screening-of-perovskite-photocatalysts**

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

A. M. K. acknowledges support from the School of Materials Engineering at Purdue University, as well as from Argonne National Laboratory under sub-contracts 21090590 and 22057223. This research used resources from the Laboratory Computing Resource Center (LCRC) and the Center for Nanoscale Materials (CNM) at Argonne National Laboratory, as well as the Rosen Center for Advanced Computing (RCAC) clusters at Purdue University. Work performed at the CNM, a U.S. Department of Energy Office of Science User Facility, was supported by the U.S. DOE, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

References

1 S. Thomas, N. Kalarikkal and A. R. Abraham, *Applications of Multifunctional Nanomaterials*, Elsevier, 2023.
2 O. Khaselev and J. A. Turner, *Science*, 1998, **280**, 425–427.
3 G. Wang, H. Wang, Y. Ling, Y. Tang, X. Yang, R. C. Fitzmorris, C. Wang, J. Z. Zhang and Y. Li, *Nano Lett.*, 2011, **11**, 3026–3033.
4 X. Zhang, S. Zhang, X. Cui, W. Zhou, W. Cao, D. Cheng and Y. Sun, *Chem. – Asian J.*, 2022, **17**, e202200668.
5 R. Dholam, N. Patel, M. Adami and A. Miotello, *Int. J. Hydrogen Energy*, 2008, **33**, 6896–6903.
6 E. P. Melián, O. G. Daz, A. O. Méndez, C. R. López, M. N. Suárez, J. D. Rodrguez, J. Navo, D. F. Hevia and J. P. Peña, *Int. J. Hydrogen Energy*, 2013, **38**, 2144–2155.
7 M. Ni, M. K. Leung, D. Y. Leung and K. Sumathy, *Renewable Sustainable Energy Rev.*, 2007, **11**, 401–425.
8 M. Matsuoka, M. Kitano, M. Takeuchi, M. Anpo and J. Thomas, *Top. Catal.*, 2005, **35**, 305–310.
9 G. Herman, Y. Gao, T. Tran and J. Osterwalder, *Surf. Sci.*, 2000, **447**, 201–211.
10 N. K. Bharti and B. Modak, *J. Phys. Chem. C*, 2022, **126**, 15080–15093.
11 M. Niu, D. Cheng and D. Cao, *Int. J. Hydrogen Energy*, 2013, **38**, 1251–1257.
12 M. A. Behnajady, B. Alizade and N. Modirshahla, *Photochem. Photobiol.*, 2011, **87**, 1308–1314.
13 D. M. Jang, I. H. Kwak, E. L. Kwon, C. S. Jung, H. S. Im, K. Park and J. Park, *J. Phys. Chem. C*, 2015, **119**, 1921–1927.
14 Y. Lin, Q. Wang, M. Ma, P. Li, V. Maheskumar, Z. Jiang and R. Zhang, *Int. J. Hydrogen Energy*, 2021, **46**, 9417–9432.
15 X. An, T. Li, B. Wen, J. Tang, Z. Hu, L.-M. Liu, J. Qu, C. Huang and H. Liu, *Adv. Energy Mater.*, 2016, **6**, 1502268.
16 W. Li, H. Zhang, M. Hong, L. Zhang, X. Feng, M. Shi, W. Hu and S. Mu, *Chem. Eng. J.*, 2022, **431**, 134072.
17 J. Yan, H. Wu, H. Chen, Y. Zhang, F. Zhang and S. F. Liu, *Appl. Catal., B*, 2016, **191**, 130–137.
18 T. Wei, Y.-N. Zhu, X. An, L.-M. Liu, X. Cao, H. Liu and J. Qu, *ACS Catal.*, 2019, **9**, 8346–8354.
19 H. Eidsvåg, S. Bentouba, P. Vajeeston, S. Yohi and D. Velauthapillai, *Molecules*, 2021, **26**, 1687.
20 Q. Guo, C. Zhou, Z. Ma and X. Yang, *Adv. Mater.*, 2019, **31**, 1901997.
21 W.-J. Yin, H. Tang, S.-H. Wei, M. M. Al-Jassim, J. Turner and Y. Yan, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2010, **82**, 045106.
22 Y. Fu, F. Meng, M. B. Rowley, B. J. Thompson, M. J. Shearer, D. Ma, R. J. Hamers, J. C. Wright and S. Jin, *J. Am. Chem. Soc.*, 2015, **137**, 5810–5818.
23 Z. Chen, Q. Dong, Y. Liu, C. Bao, Y. Fang, Y. Lin, S. Tang, Q. Wang, X. Xiao and Y. Bai, *et al.*, *Nat. Commun.*, 2017, **8**, 1890.
24 H. J. Snaith, *J. Phys. Chem. Lett.*, 2013, **4**, 3623–3630.
25 N.-G. Park, *J. Phys. Chem. Lett.*, 2013, **4**, 2423–2429.
26 Y. Cao, N. Wang, H. Tian, J. Guo, Y. Wei, H. Chen, Y. Miao, W. Zou, K. Pan and Y. He, *et al.*, *Nature*, 2018, **562**, 249–253.
27 A. A. Zhumekenov, M. I. Saidaminov, M. A. Haque, E. Alarousu, S. P. Sarmah, B. Murali, I. Dursun, X.-H. Miao, A. L. Abdelhady and T. Wu, *et al.*, *ACS Energy Lett.*, 2016, **1**, 32–37.
28 J. Chen, C. Dong, H. Idriss, O. F. Mohammed and O. M. Bakr, *Adv. Energy Mater.*, 2020, **10**, 1902433.
29 M. V. Kovalenko, L. Protesescu and M. I. Bodnarchuk, *Science*, 2017, **358**, 745–750.
30 Y. Liu and Z. Ma, *Colloids Surf., A*, 2021, **628**, 127310.
31 A. M. Fehr, A. Agrawal, F. Mandani, C. L. Conrad, Q. Jiang, S. Y. Park, O. Alley, B. Li, S. Sidhik and I. Metcalf, *et al.*, *Nat. Commun.*, 2023, **14**, 3797.
32 S. K. Karuturi, H. Shen, A. Sharma, F. J. Beck, P. Varadhan, T. Duong, P. R. Narangari, D. Zhang, Y. Wan and J.-H. He, *et al.*, *Adv. Energy Mater.*, 2020, **10**, 2000772.
33 T. Wang, S. Fan, H. Jin, Y. Yu and Y. Wei, *Phys. Chem. Chem. Phys.*, 2023, **25**, 12450–12457.
34 G. Pilania and A. Mannodi-Kanakkithodi, *J. Mater. Sci.*, 2017, **52**, 8518–8525.
35 H. Jin, H. Zhang, J. Li, T. Wang, L. Wan, H. Guo and Y. Wei, *J. Phys. Chem. Lett.*, 2019, **10**, 5211–5218.
36 J. Yang, P. Manganaris and A. Mannodi-Kanakkithodi, *J. Chem. Phys.*, 2024, **160**, 064114.
37 J. Yang, P. Manganaris and A. Mannodi-Kanakkithodi, *Digital Discovery*, 2023, **2**, 856–870.
38 J. Yang and A. Mannodi-Kanakkithodi, *arXiv*, 2023, preprint, arXiv:2309.16095, DOI: **10.48550/arXiv.2309.16095**.
39 R. Johnson and T. Zhang, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2013, **36**, 942–954.
40 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.
41 J. Yang and A. Mannodi-Kanakkithodi, *MRS Bull.*, 2022, **47**, 940–948.

42 J. Heyd, G. E. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2003, **118**, 8207–8215.
43 Z. Jiang, Y. Nahas, B. Xu, S. Prosandeev, D. Wang and L. Bellaiche, *J. Phys.: Condens. Matter*, 2016, **28**, 475901.
44 M. Ångqvist, W. A. Muñoz, J. M. Rahm, E. Fransson, C. Durniak, P. Rozyczko, T. H. Rod and P. Erhart, *Adv. Theory Simul.*, 2019, **2**, 1900015.
45 D. Bertsimas and J. Tsitsiklis, *Stat. Sci.*, 1993, **8**, 10–15.
46 C. W. Myung, A. Hajibabaei, J.-H. Cha, M. Ha, J. Kim and K. S. Kim, *Adv. Energy Mater.*, 2022, **12**, 2202279.
47 G. Pilania, J. E. Gubernatis and T. Lookman, *Comput. Mater. Sci.*, 2017, **129**, 156–163.
48 E. T. Chenebuah, M. Nganbe and A. B. Tchagang, *Mater. Today Commun.*, 2021, **27**, 102462.
49 S. Djeradi, T. Dahame, M. A. Fadla, B. Bentria, M. B. Kanoun and S. Goumri-Said, *Mach. Learn Knowl. Extr.*, 2024, **6**, 435–447.
50 T. Liu, S. Wang, Y. Shi, L. Wu, R. Zhu, Y. Wang, J. Zhou and W. C. Choy, *Sol. RRL*, 2023, **7**, 2300650.
51 G. Kresse and J. Furthmüller, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **54**, 11169.
52 G. Kresse and D. Joubert, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1999, **59**, 1758.
53 G. Kresse and J. Hafner, *J. Phys.: Condens. Matter*, 1994, **6**, 8245.
54 W. C. Ermler, R. B. Ross and P. A. Christiansen, *Advances in Quantum Chemistry*, Elsevier, 1988, vol. 19, pp. 139–182.
55 S. Steiner, S. Khmelevskyi, M. Marsmann and G. Kresse, *Phys. Rev. B*, 2016, **93**, 224425.
56 V. Wang, N. Xu, J.-C. Liu, G. Tang and W.-T. Geng, *Comput. Phys. Commun.*, 2021, **267**, 108033.
57 C. J. Bartel, C. Sutton, B. R. Goldsmith, R. Ouyang, C. B. Musgrave, L. M. Ghiringhelli and M. Scheffler, *Sci. Adv.*, 2019, **5**, eaav0693.
58 I. Hamideddine, H. Jebari, N. Tahiri, O. El Bounagui and H. Ez-Zahraouy, *Int. J. Energy Res.*, 2022, **46**, 20755–20765.
59 I. E. Castelli, T. Olsen, S. Datta, D. D. Landis, S. Dahl, K. S. Thygesen and K. W. Jacobsen, *Energy Environ. Sci.*, 2012, **5**, 5814–5819.
60 G. Wang, D. Cheng, T. He, Y. Hu, Q. Deng, Y. Mao and S. Wang, *J. Mater. Sci.: Mater. Electron.*, 2019, **30**, 10923–10933.
61 Y.-L. Liu, C.-L. Yang, M.-S. Wang, X.-G. Ma and Y.-G. Yi, *J. Mater. Sci.*, 2019, **54**, 4732–4741.
62 Y. Xu and M. A. Schoonen, *Am. Mineral.*, 2000, **85**, 543–556.
63 G. Wang, J. Chang, W. Tang, W. Xie and Y. S. Ang, *J. Phys. D: Appl. Phys.*, 2022, **55**, 293002.
64 D. Saikia, M. Alam, J. Bera, A. Betal, A. N. Gandi and S. Sahu, *Adv. Theory Simul.*, 2022, **5**, 2200511.
65 Q.-Y. Chen, Y. Huang, P.-R. Huang, T. Ma, C. Cao and Y. He, *Chin. Phys. B*, 2015, **25**, 027104.
66 D. H. Fabini, R. Seshadri and M. G. Kanatzidis, *MRS Bull.*, 2020, **45**, 467–477.
67 G. Tang, P. Ghosez and J. Hong, *J. Phys. Chem. Lett.*, 2021, **12**, 4227–4239.
68 A. Mannodi-Kanakkithodi and M. K. Chan, *Energy Environ. Sci.*, 2022, **15**, 1930–1949.
69 Y. Hu, M. F. Ayguler, M. L. Petrus, T. Bein and P. Docampo, *ACS Energy Lett.*, 2017, **2**, 2212–2218.
70 M. Pazoki, T. J. Jacobsson, A. Hagfeldt, G. Boschloo and T. Edvinsson, *Phys. Rev. B*, 2016, **93**, 144105.
71 Y. Wang, G. Brocks and S. Er, *ACS Catal.*, 2024, **14**, 1336–1350.
72 N. Ashari-Astani, S. Meloni, A. H. Salavati, G. Palermo, M. Gratzel and U. Rothlisberger, *J. Phys. Chem. C*, 2017, **121**, 23886–23895.
73 G. Giorgi, J.-I. Fujisawa, H. Segawa and K. Yamashita, *J. Phys. Chem. Lett.*, 2013, **4**, 4213–4216.

