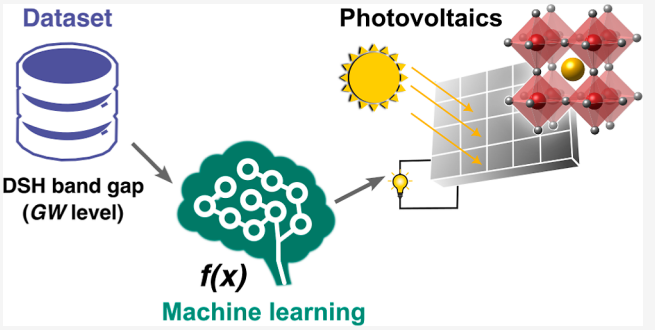
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

High-Quality Data Enabling Universality of Band Gap Descriptor and Discovery of Photovoltaic Perovskites

Haiyuan Wang,* Runhai Ouyang,* Wei Chen, and Alfredo Pasquarello

Cite This: J. Am. Chem. Soc. 2024, 146, 17636−17645
Read Online
ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Extensive machine-learning-assisted research has been dedicated to predicting band gaps for perovskites, driven by their immense potential in photovoltaics. Yet, the effectiveness is often hampered by the lack of high-quality band gap data sets, particularly for perovskites involving d orbitals. In this work, we consistently calculate a large data set of band gaps with a high level of accuracy, which is rigorously validated by experimental and state-of-the-art GW band gaps. Leveraging this achievement, our machine-learning-derived descriptor exhibits exceptional universality and robustness, proving effectiveness not only for single and double, halide and oxide perovskites regardless of the underlying atomic structures but also for hybrid organic−inorganic perovskites. With this approach, we comprehensively explore up to 15,659 materials, unveiling 14 unreported lead-free perovskites with suitable band gaps for photovoltaics. Notably, $MASnBr_3$, $FA_2SnGeBr_6$, $MA_2AuAuBr_6$, $FA_2AuAuBr_6$, $FA_2InBiCl_6$, $FA_2InBiBr_6$, and Ba_2InBiO_6 stand out with direct band gaps, small effective masses, low exciton binding energies, and high stabilities.



INTRODUCTION

Recently, perovskites have received tremendous attention in multifarious applications due to their high degree of synthetic tunability.[1−6] Their typical ABX_3 structure with corner-sharing BX_6 octahedra surrounding the cation A is shown in Figure 1a, where elements A and B can span the whole periodic table and X generally corresponds to a halogen or chalcogen element. The A site can also be occupied with an organic cation such as formamidinium (FA^+) and methylammonium (MA^+) to form hybrid organic−inorganic perovskites (Figure 1b), whereas the B site could contain two different elements leading to double perovskites described by the formula $A_2BB'X_6$ (Figure 1c). Perovskites can additionally exhibit several structural variations by going from the highly symmetric cubic phase to geometries of lower symmetry. Such a vast compositional and configurational variety allows one to tailor the properties of these materials, thereby making them highly versatile for various devices, such as photovoltaic solar cells,[1,2] light-emitting diodes,[4,7] and field-effect transistors.[5,6] One key issue required for developing such devices is achieving control over the band gap. For example, the Shockley−Queisser limit states that the power conversion efficiency of a single junction solar cell can reach 20% when the band gaps of the involved materials fall in the range 0.7−2.1 eV.[8] More specifically, when the band gaps are constrained to the range 0.93−1.61 eV, efficiencies may reach 30%.[8] Therefore, the capability of accurately and efficiently predicting band gaps is an important topic in this research area of material science.

Data-driven approaches[9−11] offer fast and powerful predictions for a given property of interest, such as the band gap. To achieve this, constructing a large and high-quality data set becomes a crucial step. Experimental approaches to obtain band gaps are generally bound by several time-consuming and expensive procedures. Furthermore, method inconsistencies when collecting data from different sources often lead to increased noise levels, making it difficult to establish high-quality data sets. Computationally, many-body perturbation theory in the GW approximation is widely recognized as the most accurate scheme for band gap calculations.[12,13] However, these calculations are highly demanding and not efficient for building large data sets. In the framework of density functional theory, (semi)local approximations like the Perdew−Burke−Ernzerhof (PBE) functional inherently underestimate the band gap by about 50%.[14] Meta-GGA functionals, such as GLLB-SC,[15] modified Becke−Johnson (mBJ),[16] and TASK,[17] have demonstrated improvements in predicting band gaps. However, the average error generally exceeds 0.5 eV.[18−20] Standard hybrid functionals, like PBE0[21,22] or range-separated HSE,[23,24] fail in consistently predicting band gaps of high accuracy when

Received: July 29, 2023
Revised: April 17, 2024
Accepted: April 18, 2024
Published: May 3, 2024

https://doi.org/10.1021/jacs.4c03507
J. Am. Chem. Soc. 2024, 146, 17636−17645

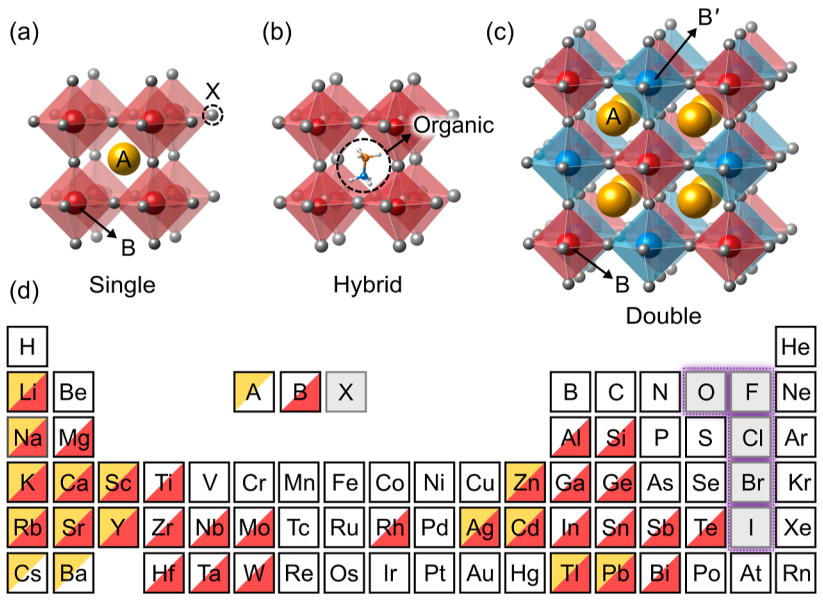


Figure 1. Representative perovskite structures and considered compositions. (a) Single perovskite ABX_3 in the cubic phase, where the A site is surrounded by corner-sharing BX_6 octahedra. (b) Hybrid organic−inorganic perovskite, where the A site is constituted by organic cations. (c) Double perovskites $A_2BB'X_6$ in the cubic phase, where the A atoms are surrounded by BX_6 and $B'X_6$ octahedra. (d) Elements considered in this work as distributed in the periodic table. Occupations of A, B, and X sites are indicated by the colors yellow, red, and gray, respectively.

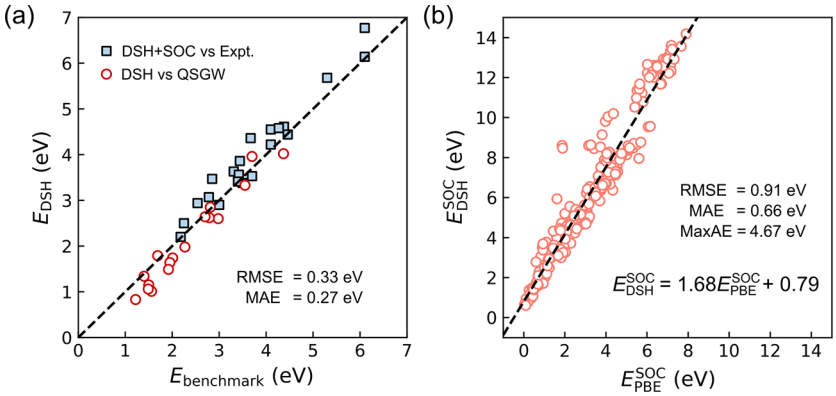


Figure 2. Validation of DSH band gaps. (a) DSH band gaps plotted against experiment (blue solid squares, cf. Table S1) and QSGW (red empty circles, cf. Table S2). In the comparison with respect to experiment, the calculated band gaps include SOC corrections. (b) DSH band gaps plotted vs PBE band gaps.

applied to a large variety of materials.[25,26] In contrast, the latest dielectric-dependent hybrid (DDH) functionals, such as doubly screened hybrid (DSH)[27] and dielectric-dependent range-separated hybrid functional based on the Coulomb attenuating method (DD-RSH-CAM),[28] have been shown to generally yield highly accurate band gaps, comparable to state-of-the-art GW calculations[27−29] but at a computational cost only slightly higher than conventional hybrid functionals due to the overhead of dielectric function calculations. In essence, the nonlocal part of the DDH functional is based on a range-separated Fock exchange expression parametrized by the dielectric function. By combining high accuracy and moderate computational cost, DDH functionals thus constitute an ideal theoretical tool for establishing a data set of band gaps.

High-throughput searches and machine learning (ML) have significantly advanced the research in material property prediction and computational screening across various materials, including perovskite[18,30−35] and nonperovskite materials.[9,36] Particularly, numerous studies have contributed to the band gap prediction of perovskites given their importance in photovoltaic applications, e.g., kernel ridge regression for the band gap of double perovskites,[30] convolutional neural networks for the band gap of metal halide perovskites,[34] and a comparison of different ML methods for hybrid organic−inorganic perovskites.[31] Despite these advancements, several issues remain to be addressed in order to achieve a more comprehensive and systematic approach. First, an optimal model should be trained on a consistent and high-quality data set. Second, it is unclear to what extent the previous ML models universally apply to various perovskite classes, such as single and double perovskites, halide- and oxide-based perovskites, and hybrid organic−inorganic perovskites. Finally, the previous ML models generally lack transparency, which hinders communication and understanding.

In this work, we systematically develop a high-quality band gap data set of 246 single and double perovskites based on highly accurate calculations with range-separated DDH functionals. Especially, we demonstrate the importance of semicore states for band gaps of halide-based double

https://doi.org/10.1021/jacs.4c03507
J. Am. Chem. Soc. 2024, 146, 17636−17645

perovskites. Next, we apply the sure independence screening and specifying operator (SISSO) method[37] to establish an accurate and insightful expression to describe and predict the band gap. We find that the resulting model works for both single and double, halide and oxide perovskites regardless of the underlying atomic structures, featuring root-mean-square errors (RMSEs) of 0.38 and 0.37 eV for the training and testing sets, respectively. This model is further found to hold for mixed perovskites and hybrid organic−inorganic perovskites, yielding RMSEs of 0.33 and 0.19 eV, respectively. Finally, we apply this model to identify promising perovskites in the huge material space encompassing 15,659 materials, finding 14 unreported Pb-free perovskites potentially suitable for photovoltaic applications. Among these perovskites, $MASnBr_3$, $FA_2SnGeBr_6$, $MA_2AuAuBr_6$, $FA_2AuAuBr_6$, $FA_2InBiCl_6$, $FA_2InBiBr_6$, and Ba_2InBiO_6 show particularly promising properties combining direct band gap, low effective masses, small exciton binding energy, and high stabilities.

RESULTS AND DISCUSSION

High-Quality Data Set Establishment. To establish a high-quality data set for band gaps, we evaluate DSH band gaps by conducting comparisons with both experiment and the higher level quasi-particle self-consistent GW (QSGW) theory. In the comparison with experiment, spin−orbit coupling (SOC) corrections are systematically included (cf. Table S1), whereas such corrections are not considered in the comparison with QSGW (cf. Table S2) to ensure consistency between the two theoretical results. The outcome of both these comparisons is illustrated in Figure 2a, where the DSH band gaps show low RMSE and mean absolute error (MAE) of 0.33 and 0.27 eV, respectively. The MAE value of 0.27 eV found here coincides with that determined previously for a set of single perovskites using the DSH functional.[29] The small size of the achieved error supports the use of the DSH functional supplemented with SOC corrections as an accurate tool for identifying suitable materials in large databases.

Thermal and zero-point motion effects resulting from the coupling to phonons affect the calculated band gap and need to be estimated when compared with the experiment. For single perovskites, we remark that a significant phonon-related band gap enhancement (∼0.7 eV[38]) has so far only been observed for the cubic phase, likely due to the significant local distortions occurring in this symmetry.[39] For noncubic single perovskites, Wang et al. demonstrated that the thermal and zero-point motion effects mostly lead to negative band gap renormalization, with an average value of about −0.15 eV.[29] To investigate the importance of such effects in the case of double perovskites, we conduct specific calculations for four representative halide-based cubic compounds (cf. Section II in the Supporting Information). Focusing on thermal effects, we find that these perovskites are subject to a band gap increase of less than 0.2 eV (cf. Figure S1). Taking into consideration the zero-phonon motions, we consistently find negative renormalizations, leading to reductions of the calculated band gaps, which do not exceed 0.1 eV. This indicates that phonon-related corrections are also minor for double perovskites. Hence, this analysis globally suggests that the combined effect of thermal and zero-point motions is generally weak compared to the intrinsic accuracy of the DSH functional. This justifies that these corrections can be initially disregarded in model training and in high-throughput search. In the case that cubic single halide perovskites would emerge from the search, it is understood that further attention should be devoted to phonon-related effects to verify whether the associated band gap enhancement would lead to the disqualification of identified candidates.

We remark that the band gaps in our data set correspond to fundamental band gaps. However, the photon absorption edge is determined by the optical band gap due to exciton formation. In this work, we estimate excitonic effects through the hydrogenic Wannier−Mott model (Table S3 and Figure S2). Typical exciton binding energies are generally small, on the order of 10 to 100 meV for single perovskites[40−43] and up to 210 meV for the double perovskites considered in our benchmark set (see Section II of the Supporting Information). We remark that excitonic effects always contribute to reducing the band gap. Thus, no material will escape detection by the upper limits set on the fundamental band gap. For strongly bound excitons, the ionization is hampered, resulting in unfavorable transport properties. At variance, for small exciton binding energies, the selection can be operated on the basis of the fundamental band gap. Therefore, band gap reductions due to excitonic effects are evaluated only for the materials emerging from the selection process.

Another key point related to data quality concerns the choice of pseudopotentials (PPs) to be used in the Quantum ESPRESSO suite of codes. We consider PPs of different accuracy denoted stringent and semicore. The stringent PPs are taken from the Pseudo Dojo set (http://www.pseudo-dojo.org) and correspond to those of highest accuracy made available. The semicore PPs include semicore states among the valence electrons and have purposely been generated for the present study. Focusing on the single perovskites given in Tables S1 and S2, we find that the stringent PPs exhibit an acceptable MAE of about 0.28 eV. However, the accuracy of these PPs breaks down when applied to double perovskites, as can be inferred from the large MAE of 0.68 eV found in Table S4. Using the semicore PPs on the same set of double perovskites, the MAE drops back to 0.32 eV (cf. Table S4). These results highlight the crucial role of semicore states in describing halide-based double perovskites. Thus, we use semicore PPs in the band gap calculations of all materials in this work. These PPs are provided in the Materials Cloud, and the link is shown in the section of Data Availability.

To further appreciate the accuracy of our scheme, we compare in Figure 2b the DSH band gaps of all 246 perovskites in our data set with their PBE band gaps, corresponding to the level of theory most widely used in ML studies to date. As expected, the PBE functional yields significantly smaller band gaps. Furthermore, when considering the optimal linear relationship to improve the PBE band gaps (cf. Figure 2b), the discrepancy with respect to the DSH band gaps still features RMSE and MAE values of 0.91 and 0.66 eV, respectively, with a maximum absolute error reaching 4.67 eV. The large size of these deviations motivates us to develop improved band gap models based on DSH data by making use of data-driven methods.

Feature Engineering and Band Gap Model Construction and Validation. To target the data set toward solar cells, we consider in our selection 173 materials with DSH band gaps up to 8 eV (data set I). By excluding materials with larger band gaps, we aim to improve the efficiency of data-driven models. This data set is then randomly split into a training set comprising about 80% of the materials (80 single and 62 double) and a testing set. Several atomic properties,

https://doi.org/10.1021/jacs.4c03507
J. Am. Chem. Soc. 2024, 146, 17636−17645

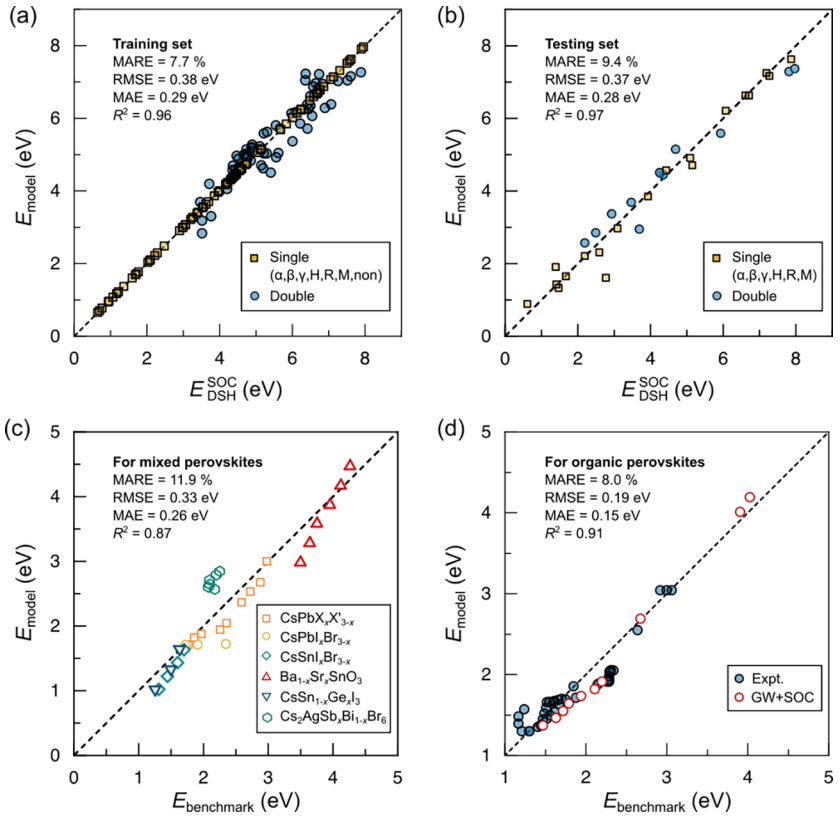


Figure 3. Performance of the developed model for band gap estimation. (a) Band gaps obtained through the SISSO-derived 3D@1C model benchmarked against DSH band gaps for the training set. (b) Same as in (a) but for the testing set. Various phases for both training and testing sets are considered, including cubic (α), tetragonal (β), orthorhombic (γ), hexagonal (H), rhombohedral (R), and monoclinic (M) phases, as well as ABX_3-based nonperovskites (non) and double perovskites (double). (c) Performance evaluation is extended to mixed-cation and mixed-anion perovskites taking benchmark band gaps from experiments in the literature (cf. Table S6). (d) Validity of the model is examined for hybrid organic−inorganic perovskites, with either experimental or GW band gap benchmarks from the literature (cf. Tables S7 and S8).

including the highest-occupied and lowest-unoccupied Kohn−Sham levels, the ionization potential (I), Pauling's electronegativity (χ), the ionic radius (r), and the oxidation number (n), as well as the PBE band gap are considered as input features. The energy levels of the atoms at the A sites lie far away from the valence band maximum (VBM) and the conduction band minimum (CBM) and thus do not influence the band gap directly.[29,44] In the following, we do not consider the properties of the atomic species at the A sites to keep the number of input features in ML to a minimum.

To identify the best model for the band gap, the RMSEs pertaining to models of varying dimensions (D) and feature complexities (C) (see Methods) are displayed in Figure S3a,b for the training and testing sets, respectively. It is evident from Figure S3b that the model with three dimensions and a feature complexity of 1 (denoted as 3D@1C) is already highly accurate. Other models with higher feature complexity showing smaller errors for the training set (Figure S3a) do not lead to any appreciable advantage when applied to the testing set (Figure S3b). The explicit expression of the band gap model 3D@1C is given by

$$E_{model} = 1.28(E^{SOC}_{PBE} - l_{B'}) + (9.05/n_{B'})l_X - 0.77 n_X l_{B'} + 2.50 \quad (1)$$

Here, E^{SOC}_{PBE} is the PBE band gap including the SOC correction, $l_{B'}$ and l_X are the energy levels of lowest-unoccupied molecular orbitals (LUMO) of the isolated atomic species B′ and X (from ref 30), respectively, as occurring in the formula $ABB'X_6$. $n_{B'}$ and n_X are the oxidation numbers of B′ and X, respectively. We draw attention to the fact that the $A_2BB'X_6$ formula reduces to that of a single perovskite when B = B′. The ordering of the B cations is defined by their oxidation state, i.e., $n_B \leq n_{B'}$. When $n_B = n_{B'}$, we further impose the requirement that $r_B > r_{B'}$. From this expression, we can see that the predictive accuracy of the band gap can be attained through the utilization of PBE band gaps with frontier orbitals of the X and B′ species, along with considering the oxidation state of B′ and X. As described in ref 45, these parameters, which are referred to as "materials genes", can effectively capture complex relationships but do not offer a comprehensive understanding of the underlying processes.[45] This is in part due to the fact that the chosen input features might be affected by correlated information, by which describing the intricate physical relationships becomes challenging. Nevertheless, the availability of an analytical expression provides a clear rule for tuning the band gaps through the LUMO and oxidation numbers of X and B′. The work by Slavney et al.[46] also emphasized the importance of frontier orbitals in determining the energy levels of the VBM and CBM.

Figure 3 illustrates the performance of our ML model in estimating band gaps. The model is free of overfitting, as it yields similar RMSEs for the training (Figure 3a) and testing (Figure 3b) sets. The predictive ability of the model applies to single and double perovskites of various structural phases. Concretely, for the testing set, the benchmark against the DSH

https://doi.org/10.1021/jacs.4c03507
J. Am. Chem. Soc. 2024, 146, 17636−17645

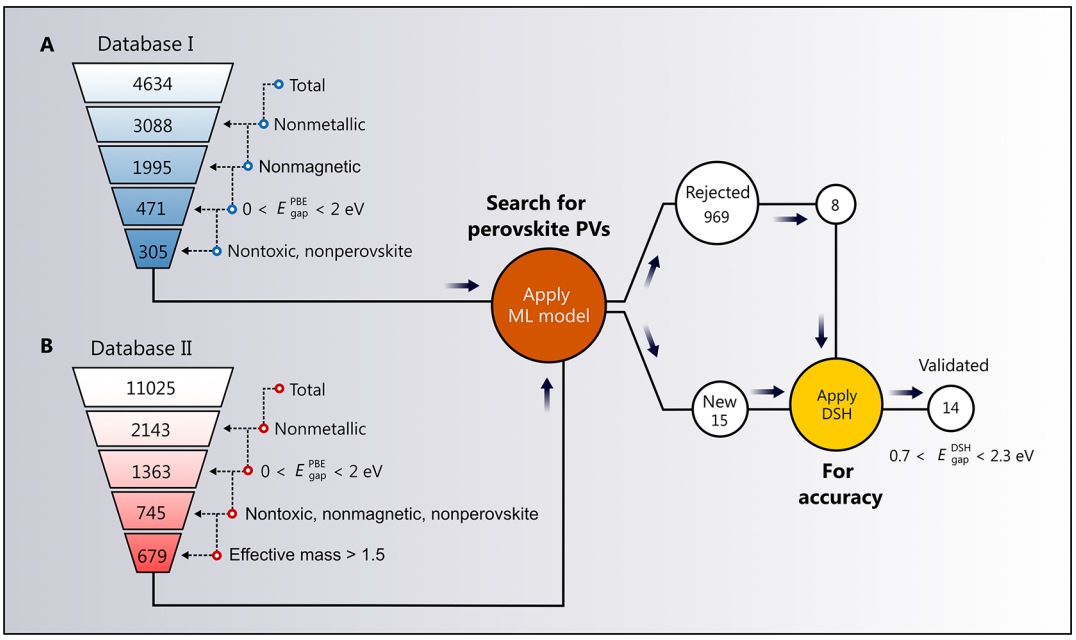


Figure 4. Screening workflow leading to data validation. The procedure for screening the databases from the Materials Project[48] (database I) and from ref 49 (database II) is illustrated. The databases are screened, based on metallicity, PBE band gap, toxicity, magnetism, nonperovskites, and effective mass. By applying the SISSO-derived ML model, the remaining data are further filtered to only retain suitable and so-far unidentified perovskites. To further increase the data accuracy, the identified materials are finally validated by DSH functional calculations. The order of the various selection steps is determined by the ease of operation, given the organization mode of the considered database.

band gaps results in a mean absolute relative error (MARE) of 9.4%, a RMSE of 0.37 eV, a MAE of 0.28 eV, and a R^2 of 0.97. To demonstrate the predictive power of the developed model, we further apply the model to mixed-cation or mixed-anion perovskites (Figure 3c) and to hybrid organic−inorganic perovskites (Figure 3d). The model shows outstanding performance with MAEs of 0.26 eV for the mixed perovskites and 0.15 eV for the organic−inorganic perovskites. These results demonstrate that eq 1 universally predicts accurate band gaps for single, double, mixed, and hybrid organic−inorganic perovskites. We note that, for the mixed perovskites, we use Vegard's law,[47] by which the band gap results from a linear combination of the band gaps of the parent materials weighted by their respective volume fractions. However, consideration of the bowing effect through the experimental bowing parameter only marginally affects the quality of the comparison with experiment in Figure 3c, resulting in a decrease of the MAE from 0.26 to 0.25 eV. Therefore, for the purpose of our benchmarking the bowing can be disregarded. Explicit consideration of the bowing effect for individual compounds is given in Tables S6 and S8.

To explore the potential of our methodology for addressing larger band gaps, we also consider a data set comprising 149 materials characterized by DSH band gaps exceeding 5 eV (data set II). Our investigation reveals that with the same input features mentioned previously, the resulting models are less satisfactory (cf. Figure S4a,b). For this data set, the incorporation of the high-frequency dielectric constant (ϵ_∞) among the input features has notable impact, resulting in a substantial decrease of the lowest RMSE for the testing set from 0.83 to 0.52 eV (cf. Figure S4b,d, and Table S5). The optimal model that includes ϵ_∞ (cf. eq S1) exhibits excellent predictive capabilities for various perovskites, as shown in Figure S5. We remark that this approach comes with both higher computational cost and practical limitations due to the extra computations required for the determination of ϵ_∞. Hence, when the focus is on materials with relatively smaller band gaps as is the case for solar cells, it is preferable to rely on models that do not require ϵ_∞ and still yield RMSEs as low as 0.37 eV (cf. Table S5). Consequently, in the present work, we use the band gap model generated on the basis of data set I without including ϵ_∞.

Discovery of New Perovskites for Photovoltaics. We employ eq 1 to perform high-throughput screening over two databases of perovskites. The first database is taken from the Materials Project[48] and contains 4634 double perovskites for which the PBE band gaps are provided. The second database consists of 11,025 organic and inorganic perovskites and is taken from ref 49. For both databases, a multistep and multistage screening procedure is carried out (cf. Figure 4). From the compounds collected from the Materials Project, we start by removing all metals (1546 compounds) and magnetic materials (1093) (see Methods). Next, to match the solar cell criterion, we additionally exclude 1524 compounds because of their PBE band gap exceeding 2 eV. Compounds including toxic elements (such as As, Tl, Pb, Hg, and Cd) and nonperovskites are further excluded, leading to 305 candidate materials for the first database.

For the second database, we first apply two screening criteria based on nonmetallicity and PBE band gap, which remove 8882 and 780 compounds, respectively. Additionally, 618 materials are excluded due to the their toxic, magnetic, and nonperovskite characteristics. The final database comprises 679 materials, achieved after further excluding compounds with an effective mass exceeding 1.5 eV, which would lead to large exciton binding energies, rendering them not ideal for solar cells. Information concerning effective masses is provided in the database II.[49]

Upon applying our model to the two databases, we further rule out 969 materials, thereby identifying 15 remaining

https://doi.org/10.1021/jacs.4c03507
J. Am. Chem. Soc. 2024, 146, 17636−17645

Table 1. Candidate Perovskites Identified for Photovoltaic Applications[a]

	E_{PBE}^{SOC}	D/I	ϵ_∞	E_{model}	E_{DSH}^{SOC}	ΔE	ΔH	T_f	O_f
				$\{B^{2+}B'^{2+}\}$					
$MASnBr_3$	0.88	D	5.03	1.95	1.81	0.14	55	0.89	0.53
$FAInBr_3$	0.52	I	7.01	1.47	1.25	0.22	44	1.07	0.41
$FA_2SnGeBr_6$	0.60	D	5.64	1.59	1.46	0.13	53	1.03	0.45
FA_2SnZnI_6	0.90	I	7.00	2.22	2.21	0.01	24	1.01	0.40
				$\{B^{+}B'^{3+}\}$					
$MA_2AuAuBr_6$	0.62	D	6.09	2.38	1.54	0.84	81	0.87	0.57
$MA_2AuBiBr_6$	0.34	I	5.80	1.99	1.64	0.35	33	0.84	0.61
$FA_2AuAuBr_6$	0.71	D	5.86	2.50	1.66	0.84	79	0.96	0.57
FA_2AgBiI_6	0.48	I	5.63	2.13	1.60	0.53	42	0.95	0.50
FA_2AgSbI_6	0.28	I	4.78	1.81	1.39	0.42	42	0.99	0.43
$FA_2CuBiBr_6$	0.53	I	8.16	2.23	1.99	0.24	36	1.03	0.46
$FA_2InBiCl_6$	0.35	D	4.25	2.29	1.45	0.84	61	1.04	0.51
$FA_2InBiBr_6$	0.13	D	5.69	1.72	0.87	0.85	50	1.02	0.47
$K_2CuInCl_6$	0.00	I	7.42	1.75	2.03	−0.28	36	0.94	0.43
Ba_2InBiO_6	0.23	D	6.89	2.16	1.26	0.90	273	0.98	0.56

[a]The identified halide-based perovskites are divided into two categories depending on the oxidation states of the atoms at the B site: $\{B^{2+}B'^{2+}\}$ and $\{B^{+}B'^{3+}\}$. E_{PBE}^{SOC} and E_{DSH}^{SOC} are the PBE and DSH band gaps, respectively, including SOC corrections. E_{model} is the band gap estimated with the model given in eq 1. $\Delta E = E_{model} - E_{DSH}^{SOC}$ stands for the energy difference between E_{model} and E_{DSH}^{SOC}. The direct (D) or indirect (I) nature of the band gap is indicated. The high-frequency dielectric constant ϵ_∞ has been calculated at the PBE level in this work. ΔH, T_f, and O_f are the decomposition enthalpy, tolerance factor, and octahedral factor, respectively. All energies are in eV, except ΔH, which is expressed in meV.

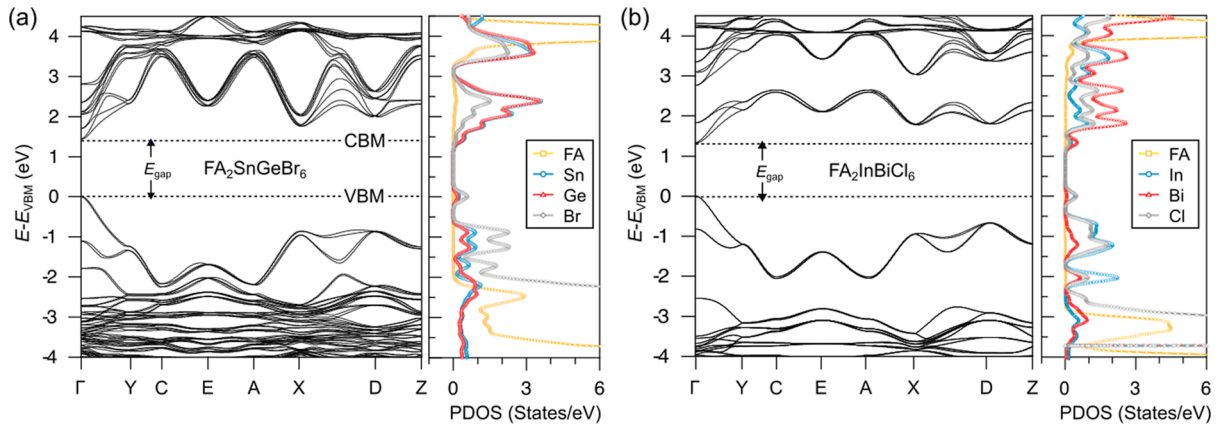


Figure 5. Electronic properties. Band structure and PDOS for (a) $FA_2SnGeBr_6$ and (b) $FA_2InBiCl_6$ calculated with the functional DSH + SOC. Yellow, blue, red, and gray lines are the PDOS on A, B, B′, and X, respectively.

candidates. The rejected materials encompass 869 perovskites having band gaps above 2.3 eV and 52 having been reported previously (the band gap estimated from our model and the corresponding references are given in Table S9). Furthermore, 48 materials have been identified as being unstable on the basis of both thermodynamic and crystallographic considerations (vide infra for the applied criteria). For the 15 materials that have successfully met our selection criteria, we determined ϵ_∞ and recalculated the band gap with the DSH functional to enhance the reliability of our predictions. This leads to a subset of 12 materials for which the DSH band gaps fall in the range from 0.7 to 2.3 eV. Similarly, suitable candidates might have been erroneously discarded due to band gaps overestimated by our model. Therefore, we revisit the set of 969 materials and perform selective DSH calculations for 8 materials meeting the following criteria: (i) $E_{model} < 3$ eV, (ii) direct band gap at the PBE level, (iii) stability, and (iv) being unreported so far. This additional step allows us to identify two more materials with band gaps falling within the desired range between 0.7 and 2.3 eV. From this analysis, it appears that the initial exclusion criteria of our model successfully filtered out most of the less promising candidates, but the application of the DSH functional remains essential for a finer identification. Among the identified materials, there are no single cubic halide perovskites. Thus, we expect that thermal and zero-point motion effects are smaller than 0.2 eV, as discussed above. Moreover, since the band gaps of these materials are generally less than 2 eV, these effects do not critically affect the identification. Hence, in total, this selection procedure leads to 14 suitable candidates, which are reported in Table 1.

From Table 1, we first notice that the most promising materials for solar-cell applications show high-frequency dielectric constants ϵ_∞ that are systematically larger than 5. Furthermore, we observe that a majority of identified materials are halide-based double perovskites. Generally, these materials can be characterized by a couple of $\{BB'\}$ replacing the divalent B atom of single perovskites. Two categories can be distinguished for these perovskites. The first category shows the isovalent couple $\{B^{2+}B'^{2+}\}$ with both B and B′ being divalent, while the second category corresponds to $\{B^{+}B'^{3+}\}$

https://doi.org/10.1021/jacs.4c03507
J. Am. Chem. Soc. 2024, 146, 17636−17645

couples, where B is monovalent and B′ is trivalent. The band structures and projected density of states (PDOS) of $FA_2SnGeBr_6$ and $FA_2InBiCl_6$, which we take as representatives of the two categories, are displayed in Figure 5. As is usual in such perovskites, the electronic states associated with the atoms at the A sites lie far away from the CBM and VBM, while the band-edge states show hybridization among the electronic states arising from the atoms at the B, B′, and X sites. The difference between the materials of the two categories lies in the fact that the atomic states from the B^{2+} and B'^{2+} atoms contribute equally to the density of states in the vicinity of the band gap (cf. Figure 5a), whereas the contributions coming from B^+ and B'^{3+} atoms clearly differ (cf. Figure 5b). Following this rationale, we inspect the PDOS of $MASnBr_3$, $FAInBr_3$, $MA_2AuAuBr_6$, and $FA_2AuAuBr_6$ (cf. Figure S6) and infer that $MASnBr_3$ and $FAInBr_3$ belong to the category $\{B^{2+}B'^{2+}\}$, whereas $MA_2AuAuBr_6$ and $FA_2AuAuBr_6$ are representative of the category $\{B^+B'^{3+}\}$.

The thermodynamic stability is evaluated by the decomposition enthalpy (ΔH) with respect to possible decomposing pathways, primarily characterized by the decomposition into corresponding binary materials.[50] As an example, for a double perovskite belonging to the category $\{B^+B'^{3+}\}$, the decomposition enthalpy is expressed as $\Delta H = 2E_{AX} + E_{BX} + E_{B'X_3} - E_{A_2BB'X_6}$. The decomposition pathways of the identified materials are listed in Table S10. As established by ref 50, positive values of ΔH exceeding 20 meV/atom indicate favorable stability. To evaluate the crystallographic stability, we also examine the tolerance factor, $T_f = (r_A + r_X)/[\sqrt{2}(r_{\bar{B}} + r_X)]$, and the octahedral factor, $O_f = r_{\bar{B}}/r_X$, where r_A and r_X are the ionic radii of the corresponding ions and $r_{\bar{B}}$ is the average ionic radius of B and B′.[31] Prior examinations of perovskites have demonstrated that these materials exhibit optimal potential for formation when $0.8 < T_f < 1.2$ and $0.4 < O_f < 0.7$.[31] The values of T_f and O_f calculated for the 14 selected materials are given in Table 1. The criteria of stability are met for all materials with values of T_f and O_f, which generally lie away from the limits, whereby the application of more sophisticated geometrical criteria[51] are not necessary. Moreover, we consider the energy above hull, recognizing that its value should be close to 0 eV/atom for stable semiconductors.[52] Taking $FA_2SnGeBr_6$ as an example, we find its energy above the hull to be 0.2 eV/atom. For three experimentally synthesized materials, namely $Cs_2AgBiBr_6$, $MA_2AgBiBr_6$, and MA_2AgBiI_6,[53] we determine energy-above-hull values of 0.01, 0.19, and 0.15 eV/atom, respectively. In view of this outcome, we refrain from conducting energy-above-hull investigations for organic−inorganic perovskites, as this criterion does not appear to be decisive for their formation. However, for the two inorganic perovskites identified in this work, i.e., $K_2CuInCl_6$ and Ba_2InBiO_6, we calculate energy-above-hull values of 0.02 and 0.00 eV/atom, respectively, which imply that they meet the stability criterion.

Among the discovered materials in Table 1, seven perovskites, viz., $MASnBr_3$, $FA_2SnGeBr_6$, $MA_2AuAuBr_6$, $FA_2AuAuBr_6$, $FA_2InBiCl_6$, $FA_2InBiBr_6$, and Ba_2InBiO_6 stand out as particularly promising candidates for light-harvesting in photovoltaic devices because of their direct band gap. In addition, these seven materials also carry features that advantageously affect their carrier transport properties.[54] Indeed, with hole and electron effective masses as low as $0.11m_0$ and $0.32m_0$ (cf. Table S3), respectively, where m_0 is the rest mass of the electron, these materials compare favorably with the perovskite $MAPbI_3$, which shows respective masses of $0.29m_0$ and $0.23m_0$.[55] Furthermore, the exciton binding energies in these materials can be estimated through the hydrogenic Wannier−Mott model[56] and are found to be at most 118 meV, corresponding to the case of Ba_2InBiO_6 (cf. Table S3). The exciton binding energies in $FA_2SnGeBr_6$ and $FA_2InBiBr_6$ are even lower than 40 meV (cf. Table S3), comparable to that of $FAPbI_3$ (35 meV[57]), the perovskite known for achieving the highest measured power efficiency so far (25.5%[2]).

It is important to note that the identified materials may be subject to limitations that require further experimental investigation and solution. For instance, perovskites are known to contain intrinsic point defects, such as vacancies, which could affect the performance of these materials in harvesting solar energy. However, unlike conventional semiconductors, metal halide perovskites appear to show exceptional defect tolerance, by which defect states are unable to deteriorate the convenient electronic properties of the pristine materials.[58,59] Since the identified materials mostly belong to this class of perovskites, it can reasonably be expected that the occurrence of typical defects would not be detrimental. A more critical aspect concerns the occurrence of the identified materials with unfavorable oxidation states. For instance, Ge^{2+} and Sn^{2+} are generally prone to oxidation, resulting in the +4 oxidation state. Similarly, In^+ tends to form In^{3+} under ambient conditions.[60] Indeed, the stability of unusual oxidation states of tin and germanium is currently a focal point for many research groups.[61−65] Of particular interest in this context, is the report by Kama et al.,[64] who found that Ge^{2+} is able to stabilize the concomitant occurrence of Sn^{2+}, thereby preventing the formation of higher oxidation states. Another interesting stabilization effect is observed for the here-identified Au-based materials, $MA_2AuAuBr_6$ and $FA_2AuAuBr_6$, which overcome the formation of the unfavorable +2 oxidation state by spontaneously giving rise to the formation of double perovskites based on the $\{Au^{1+}Au^{3+}\}$ couple, in line with the experimentally demonstrated synthesis of $Cs_2AuAuCl_6$.[66] Hence, while the demonstration of performance ultimately always rests on experiment, the list of materials in Table 1 provides precious guidelines for indicating research directions.

CONCLUSIONS

Through the use of DSH functionals and semicore PPs, we construct in this work high-quality data sets of band gaps, which carry an accuracy comparable to that of GW calculations. Combining DSH calculations and the SISSO machine-learning method, we further develop a band gap descriptor, which allows us to rapidly and accurately predict band gaps. To achieve this high level of description, it is sufficient to inform an analytical expression with the PBE band gap as well as the properties of the frontier orbitals and oxidation numbers of the isolated constituent atomic species. The resulting model is universally applicable to various classes of perovskites, including single and double, halide and oxide perovskites regardless of the underlying atomic structures, as well as hybrid organic−inorganic perovskites. The band gap model has also been found to successfully apply to mixed-cation or mixed-anion perovskites.

Through high-throughput screening of two databases containing 15,659 materials based on our model, we discover 14 Pb-free perovskites with suitable band gaps for solar-cell

https://doi.org/10.1021/jacs.4c03507
J. Am. Chem. Soc. 2024, 146, 17636−17645

applications, which have so far escaped detection. In particular, the perovskites $MASnBr_3$, $FA_2SnGeBr_6$, $MA_2AuAuBr_6$, $FA_2AuAuBr_6$, $FA_2InBiCl_6$, $FA_2InBiBr_6$, and Ba_2InBiO_6 stand out due to the direct nature of their band gap, low effective masses, low exciton binding energies, and favorable stability properties.

It is generally important to explore novel photovoltaic materials from large material spaces covering other types of perovskites, such as oxynitrides and oxysulfides and even other crystal prototypes. For this, more generalized ML models might be required. These will rely on the production of large amounts of additional high-quality and diverse data as well as on the engineering of structural features that are sensitive to crystal symmetry.

METHODS

Calculation Details. The DSH band gaps of 246 experimentally synthesized inorganic perovskites (126 single and 120 double nonmagnetic perovskites) are determined through the implementation of dielectric dependent hybrid functionals based on the Coulomb-attenuating method,[28] which is available in the Quantum ESPRESSO code.[67] The range-separated DSH has $\alpha_{SR} = 1$ in the short-range and $\alpha_{LR} = 1/\epsilon_\infty$ in the long-range.[27−29] The range-separation parameter $\mu = 2\tilde{k}_{TF}/3$, where $\tilde{k}_{TF}^2 = k_{TF}^2[(\epsilon_\infty - 1)^{-1} + 1]/\alpha$ with $\alpha = 1.563$. $k_{TF} = 2\sqrt[6]{(3n/\pi)}$ is the Thomas−Fermi screening length and depends on the valence charge density n.[27−29] In fact, the DSH functional shares the same functional form with the DD-RSH-CAM but does not require the computationally expensive dielectric function for the determination of the range-separation parameter.[28] The latter parameter is indeed given by an analytical expression that only depends on the high-frequency dielectric constant and on the valence charge density.[27,29] Fully relativistic PPs generated through the optimized norm-conserving Vanderbilt PP scheme[68] are used to account for SOC effects. The SOC correction ΔE_{SOC} is calculated with the PBE functional, which has been demonstrated to give an accurate description.[29] Namely, $\Delta E_{SOC} = E_{SOC} - E_{bare}$, where E_{SOC} and $E_{bare} = E_{CBM} - E_{VBM}$ are the Kohn−Sham band gaps at the PBE level with and without SOC effects, respectively. We then used ΔE_{SOC} to correct the bare DSH band gap. All DSH band gaps mentioned throughout this work include SOC corrections calculated in this way. By consequence, the band gaps estimated through the models implicitly include the SOC effect. To ensure a band gap convergence within 0.1 eV, all calculations use in-house-generated semicore-based PPs. A kinetic plane-wave cutoff of 100 Ry is used for all materials together with sufficiently dense k-point grids. In this work, we fix the lattice parameters from the experiment and fully relax atomic positions for the calculations of band gaps and high-frequency dielectric constants ϵ_∞. The value of ϵ_∞, which is used to fix the parameter α_{LR} in the DSH functional, is obtained at the PBE level through the application of finite electric fields.[69] A more detailed description can be found in previous studies.[27−29] All 246 structures have been downloaded from the inorganic crystal structure database (ICSD: https://icsd.fiz-karlsruhe.de/search/basic.xhtml) and Springer Materials database (https://materials.springer.com/). The atomic structures of the hybrid organic−inorganic compounds are taken from refs 49 and 57 where an accurate representation of the molecular orientations was obtained through a rigorous procedure involving molecular dynamics and the selection of various initial structures. Magnetic materials have not been considered in this work because the validity of the DSH functional for this class of materials has not yet sufficiently been ascertained. The assessment of energy above convex hull takes into consideration the PBE formation energies of all pertinent compounds available in the Materials Project[48] and is carried out with the pymatgen code.[70]

The computational cost to calculate the band gap with the DSH functional is the same as that for standard hybrid functionals, such as PBE0 or HSE06. Sufficiently accurate dielectric constants can be achieved at the PBE level and thus do not yield any significant overhead. Hybrid functional calculations imply a computational cost lower by at least 1 order of magnitude compared to GW calculations.[26]

Machine Learning Method-SISSO. SISSO is a data-driven method that combines symbolic regression and compressed sensing for explicit mathematical expressions. The initial feature space comprises 17 primary features $\Phi_0 = [E_{PBE}^{SOC}, h_B, h_{B'}, l_B, l_{B'}, h_X, l_X, I_X, n_B, n_{B'}, n_X, \chi_B, \chi_{B'}, \chi_X, r_B, r_{B'}, r_X]$, and the mathematical operations for feature construction are $\{+,-,\times,/,^{-1}, ^2, \sqrt[3]{}^3, |-|\}$. With these primary features and mathematical operations, a large number of expressions are generated. Details of the SISSO algorithm can be found in refs 37 and 71. The SISSO code is available at http://github.com/rouyang2017/SISSO. The values for Pauling's electronegativity and for the ionization potentials are taken from ref 72. For isolated atomic species, the highest occupied and the lowest unoccupied Kohn−Sham levels calculated at the PBE level and the ionic radii are taken from refs 30 and 73, respectively.

ASSOCIATED CONTENT

Data Availability Statement

The relevant data, including 246 geometry structures with corresponding DSH band gaps, semicore-based pseudopotentials, and database screening results, can be found in the Materials Cloud: https://archive.materialscloud.org/record/2024.61.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.4c03507.

Benchmark for DSH band gaps, DSH band gaps with SOC correction vs experimental values, DSH vs QSGW band gaps; band gap corrections, thermal effects and zero-point motion, exciton binding energies; validation: PP validation for double perovskite; effect of the high-frequency dielectric constant on the performance of the band gap model, models yielding the lowest RMSEs, effect of ϵ_∞ on the SISSO errors for band gaps larger than 5 eV, effect of ϵ_∞ on the SISSO errors for band gaps below 8 eV, model performance for band gaps below 8 eV; band gap model applied to mixed single perovskites; model band gaps for organic and inorganic perovskites compared to reference values; band gap model applied to mixed organic perovskites; materials passing the screening for photovoltaic applications but already reported; decomposition pathways and ionic radii; and identification of $\{B^{2+}B'^{2+}\}$ or $\{B^{+}B'^{3+}\}$ through the PDOS (PDF)

AUTHOR INFORMATION

Corresponding Authors

Haiyuan Wang − *Chaire de Simulation à l' Echelle Atomique (CSEA), Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland;* orcid.org/0000-0002-2411-1501; Email: haiyuan.wang@epfl.ch

Runhai Ouyang − *Materials Genome Institute, Shanghai University, 200444 Shanghai, China;* orcid.org/0000-0001-9020-0484; Email: rouyang@shu.edu.cn

Authors

Wei Chen − *Institute of Condensed Matter and Nanosciences, Université catholique de Louvain (UCLouvain), B-1348 Louvain-la-Neuve, Belgium;* orcid.org/0000-0002-7496-0341

Alfredo Pasquarello − *Chaire de Simulation à l' Echelle Atomique (CSEA), Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland;* orcid.org/0000-0002-9142-2799

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.4c03507

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank S. de Gironcoli, D. Park, A. Tal, and Y. W. Li for their useful discussions. We also thank G. Pilania for sharing the data of the HOMO and LUMO energy levels of the atomic elements calculated in ref 30. This work was supported by the Swiss National Science Foundation (SNSF) under grant no. 200020-172524. R.O. acknowledges the financial support by the National Natural Science Foundation of China (grant no. 22173058). The calculations have been carried out at the Swiss National Supercomputing Center (CSCS) (Project ID No.s1123) and at SCITAS-EPFL.

■ REFERENCES

(1) Kojima, A.; Teshima, K.; Shirai, Y.; Miyasaka, T. Organometal halide perovskites as visible-light sensitizers for photovoltaic cells. *J. Am. Chem. Soc.* **2009**, *131*, 6050−6051.
(2) Jeong, J.; Kim, M.; Seo, J.; Lu, H.; Ahlawat, P.; Mishra, A.; Yang, Y.; Hope, M. A.; Eickemeyer, F. T.; Kim, M.; et al. Pseudo-halide anion engineering for α-$FAPbI_3$ perovskite solar cells. *Nature* **2021**, *592*, 381−385.
(3) Tan, Z.-K.; Moghaddam, R. S.; Lai, M. L.; Docampo, P.; Higler, R.; Deschler, F.; Price, M.; Sadhanala, A.; Pazos, L. M.; Credgington, D.; Hanusch, F.; Bein, T.; Snaith, H. J.; Friend, R. H. Bright light-emitting diodes based on organometal halide perovskite. *Nat. Nanotechnol.* **2014**, *9*, 687−692.
(4) Lin, K.; Xing, J.; Quan, L. N.; de Arquer, F. P. G.; Gong, X.; Lu, J.; Xie, L.; Zhao, W.; Zhang, D.; Yan, C.; et al. Perovskite light-emitting diodes with external quantum efficiency exceeding 20 per cent. *Nature* **2018**, *562*, 245−248.
(5) Kagan, C. R.; Mitzi, D. B.; Dimitrakopoulos, C. D. Organic-inorganic hybrid materials as semiconducting channels in thin-film field-effect transistors. *Science* **1999**, *286*, 945−947.
(6) Liu, A.; Zhu, H.; Bai, S.; Reo, Y.; Zou, T.; Kim, M.-G.; Noh, Y.-Y. High-performance inorganic metal halide perovskite transistors. *Nat. Electron.* **2022**, *5*, 78−83.
(7) Liu, X.-K.; Xu, W.; Bai, S.; Jin, Y.; Wang, J.; Friend, R. H.; Gao, F. Metal halide perovskites for light-emitting diodes. *Nat. Mater.* **2021**, *20*, 10−21.
(8) Rühle, S. Tabulated values of the Shockley−Queisser limit for single junction solar cells. *Sol. Energy* **2016**, *130*, 139−147.
(9) Curtarolo, S.; Hart, G. L.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O. The high-throughput highway to computational materials design. *Nat. Mater.* **2013**, *12*, 191−201.
(10) Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M. Big Data of Materials Science: critical Role of the Descriptor. *Phys. Rev. Lett.* **2015**, *114*, 105503.
(11) Bartel, C. J.; Sutton, C.; Goldsmith, B. R.; Ouyang, R.; Musgrave, C. B.; Ghiringhelli, L. M.; Scheffler, M. New tolerance factor to predict the stability of perovskite oxides and halides. *Sci. Adv.* **2019**, *5*, No. eaav0693.
(12) Hedin, L. New method for calculating the one-particle Green's function with application to the electron-gas problem. *Phys. Rev.* **1965**, *139*, A796−A823.
(13) Tal, A.; Chen, W.; Pasquarello, A. Vertex function compliant with the Ward identity for quasiparticle self-consistent calculations beyond *GW*. *Phys. Rev. B* **2021**, *103*, L161104.
(14) Chen, W.; Pasquarello, A. Band-edge positions in G W: Effects of starting point and self-consistency. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *90*, 165133.
(15) Kuisma, M.; Ojanen, J.; Enkovaara, J.; Rantala, T. T. Kohn-Sham potential with discontinuity for band gap materials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *82*, 115106.
(16) Tran, F.; Blaha, P. Accurate band gaps of semiconductors and insulators with a semilocal exchange-correlation potential. *Phys. Rev. Lett.* **2009**, *102*, 226401.
(17) Aschebrock, T.; Kümmel, S. Ultranonlocality and accurate band gaps from a meta-generalized gradient approximation. *Phys. Rev. Res.* **2019**, *1*, 033082.
(18) Castelli, I. E.; Olsen, T.; Datta, S.; Landis, D. D.; Dahl, S.; Thygesen, K. S.; Jacobsen, K. W. Computational screening of perovskite metal oxides for optimal solar light capture. *Energy Environ. Sci.* **2012**, *5*, 5814−5819.
(19) Tran, F.; Ehsan, S.; Blaha, P. Assessment of the GLLB-SC potential for solid-state properties and attempts for improvement. *Phys. Rev. Mater.* **2018**, *2*, 023802.
(20) Borlido, P.; Schmidt, J.; Huran, A. W.; Tran, F.; Marques, M. A.; Botti, S. Exchange-correlation functionals for band gaps of solids: benchmark, reparametrization and machine learning. *npj Comput. Mater.* **2020**, *6*, 96.
(21) Perdew, J. P.; Ernzerhof, M.; Burke, K. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.* **1996**, *105*, 9982−9985.
(22) Burke, K.; Ernzerhof, M.; Perdew, J. P. The adiabatic connection method: a non-empirical hybrid. *Chem. Phys. Lett.* **1997**, *265*, 115−120.
(23) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **2003**, *118*, 8207−8215.
(24) Krukau, A. V.; Vydrov, O. A.; Izmaylov, A. F.; Scuseria, G. E. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *J. Chem. Phys.* **2006**, *125*, 224106.
(25) Chen, W.; Pasquarello, A. Band-edge levels in semiconductors and insulators: Hybrid density functional theory versus many-body perturbation theory. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2012**, *86*, 035134.
(26) Yang, J.; Falletta, S.; Pasquarello, A. One-Shot Approach for Enforcing Piecewise Linearity on Hybrid Functionals: Application to Band Gap Predictions. *J. Phys. Chem. Lett.* **2022**, *13*, 3066−3071.
(27) Cui, Z.-H.; Wang, Y.-C.; Zhang, M.-Y.; Xu, X.; Jiang, H. Doubly screened hybrid functional: an accurate first-principles approach for both narrow-and wide-gap semiconductors. *J. Phys. Chem. Lett.* **2018**, *9*, 2338−2345.
(28) Chen, W.; Miceli, G.; Rignanese, G.-M.; Pasquarello, A. Nonempirical dielectric-dependent hybrid functional with range separation for semiconductors and insulators. *Phys. Rev. Mater.* **2018**, *2*, 073803.
(29) Wang, H.; Tal, A.; Bischoff, T.; Gono, P.; Pasquarello, A. Accurate and efficient band-gap predictions for metal halide perovskites at finite temperature. *npj Comput. Mater.* **2022**, *8*, 237.
(30) Pilania, G.; Mannodi-Kanakkithodi, A.; Uberuaga, B.; Ramprasad, R.; Gubernatis, J. E.; Lookman, T. Machine learning bandgaps of double perovskites. *Sci. Rep.* **2016**, *6*, 19375.
(31) Lu, S.; Zhou, Q.; Ouyang, Y.; Guo, Y.; Li, Q.; Wang, J. Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nat. Commun.* **2018**, *9*, 3405.
(32) Im, J.; Lee, S.; Ko, T.-W.; Kim, H. W.; Hyon, Y.; Chang, H. Identifying Pb-free perovskites for solar cells by machine learning. *npj Comput. Mater.* **2019**, *5*, 37.
(33) Li, Z.; Xu, Q.; Sun, Q.; Hou, Z.; Yin, W.-J. Thermodynamic stability landscape of halide double perovskites via high-throughput computing and machine learning. *Adv. Funct. Mater.* **2019**, *29*, 1807280.
(34) Saidi, W. A.; Shadid, W.; Castelli, I. E. Machine-learning structural and electronic properties of metal halide perovskites using a

hierarchical convolutional neural network. *npj Comput. Mater.* **2020**, *6*, 36.
(35) Castelli, I. E.; Landis, D. D.; Thygesen, K. S.; Dahl, S.; Chorkendorff, I.; Jaramillo, T. F.; Jacobsen, K. W. New cubic perovskites for one-and two-photon water splitting using the computational materials repository. *Energy Environ. Sci.* **2012**, *5*, 9034−9043.
(36) Zhuo, Y.; MansouriTehrani, A.; Brgoch, J. Predicting the band gaps of inorganic solids by machine learning. *J. Phys. Chem. Lett.* **2018**, *9*, 1668−1673.
(37) Ouyang, R.; Curtarolo, S.; Ahmetcik, E.; Scheffler, M.; Ghiringhelli, L. M. SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Phys. Rev. Mater.* **2018**, *2*, 083802.
(38) Wiktor, J.; Rothlisberger, U.; Pasquarello, A. Predictive determination of band gaps of inorganic halide perovskites. *J. Phys. Chem. Lett.* **2017**, *8*, 5507−5512.
(39) Cannelli, O.; Wiktor, J.; Colonna, N.; Leroy, L.; Puppin, M.; Bacellar, C.; Sadykov, I.; Krieg, F.; Smolentsev, G.; Kovalenko, M. V.; et al. Atomic-Level Description of Thermal Fluctuations in Inorganic Lead Halide Perovskites. *J. Phys. Chem. Lett.* **2022**, *13*, 3382−3391.
(40) Filip, M. R.; Eperon, G. E.; Snaith, H. J.; Giustino, F. Steric engineering of metal-halide perovskites with tunable optical band gaps. *Nat. Commun.* **2014**, *5*, 5757.
(41) Chen, Z.; Yu, C.; Shum, K.; Wang, J. J.; Pfenninger, W.; Vockic, N.; Midgley, J.; Kenney, J. T. Photoluminescence study of polycrystalline $CsSnI_3$ thin films: Determination of exciton binding energy. *J. Lumin.* **2012**, *132*, 345−349.
(42) Baranowski, M.; Plochocka, P.; Su, R.; Legrand, L.; Barisien, T.; Bernardot, F.; Xiong, Q.; Testelin, C.; Chamarro, M. Exciton binding energy and effective mass of $CsPbCl_3$: a magneto-optical study. *Photonics Res.* **2020**, *8*, A50−A55.
(43) Dias, A. C.; Lima, M. P.; DaSilva, J. L. F. Role of Structural Phases and Octahedra Distortions in the Optoelectronic and Excitonic Properties of CsGeX3 (X = Cl, Br, I) Perovskites. *J. Phys. Chem. C* **2021**, *125*, 19142−19155.
(44) Tao, S.; Schmidt, I.; Brocks, G.; Jiang, J.; Tranca, I.; Meerholz, K.; Olthof, S. Absolute energy level positions in tin-and lead-based halide perovskites. *Nat. Commun.* **2019**, *10*, 2560.
(45) Foppa, L.; Rüther, F.; Geske, M.; Koch, G.; Girgsdies, F.; Kube, P.; Carey, S. J.; Hävecker, M.; Timpe, O.; Tarasov, A. V.; et al. Data-Centric Heterogeneous Catalysis: Identifying Rules and Materials Genes of Alkane Selective Oxidation. *J. Am. Chem. Soc.* **2023**, *145*, 3427−3442.
(46) Slavney, A. H.; Connor, B. A.; Leppert, L.; Karunadasa, H. I. A pencil-and-paper method for elucidating halide double perovskite band structures. *Chem. Sci.* **2019**, *10*, 11041−11053.
(47) Hill, R. Energy-gap variations in semiconductor alloys. *J. Phys. C: Solid State Phys.* **1974**, *7*, 521−526.
(48) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002.
(49) Nakajima, T.; Sawada, K. Discovery of Pb-free perovskite solar cells via high-throughput simulation on the K computer. *J. Phys. Chem. Lett.* **2017**, *8*, 4826−4831.
(50) Zhao, X.-G.; Yang, J.-H.; Fu, Y.; Yang, D.; Xu, Q.; Yu, L.; Wei, S.-H.; Zhang, L. Design of lead-free inorganic halide perovskites for solar cells via cation-transmutation. *J. Am. Chem. Soc.* **2017**, *139*, 2630−2638.
(51) Filip, M. R.; Giustino, F. The geometric blueprint of perovskites. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 5397−5402.
(52) Siriwardane, E. M. D.; Zhao, Y.; Perera, I.; Hu, J. Generative design of stable semiconductor materials using deep learning and density functional theory. *npj Comput. Mater.* **2022**, *8*, 164.
(53) Shadabroo, M. S.; Abdizadeh, H.; Golobostanfard, M. R. Elpasolite structures based on A_2AgBiX_6 (A: MA, Cs, X: I, Br): Application in double perovskite solar cells. *Mater. Sci. Semicond. Process.* **2021**, *125*, 105639.
(54) Liu, Y.; Yang, Z.; Liu, S. Recent progress in single-crystalline perovskite research including crystal preparation, property evaluation, and applications. *Adv. Sci.* **2018**, *5*, 1700471.
(55) Giorgi, G.; Fujisawa, J.-I.; Segawa, H.; Yamashita, K. Small photocarrier effective masses featuring ambipolar transport in methylammonium lead iodide perovskite: a density functional analysis. *J. Phys. Chem. Lett.* **2013**, *4*, 4213−4216.
(56) Tilchin, J.; Dirin, D. N.; Maikov, G. I.; Sashchiuk, A.; Kovalenko, M. V.; Lifshitz, E. Hydrogen-like Wannier−Mott excitons in single crystal of methylammonium lead bromide perovskite. *ACS Nano* **2016**, *10*, 6363−6371.
(57) Bokdam, M.; Sander, T.; Stroppa, A.; Picozzi, S.; Sarma, D.; Franchini, C.; Kresse, G. Role of polar phonons in the photo excited state of metal halide perovskites. *Sci. Rep.* **2016**, *6*, 28618.
(58) Kovalenko, M. V.; Protesescu, L.; Bodnarchuk, M. I. Properties and potential optoelectronic applications of lead halide perovskite nanocrystals. *Science* **2017**, *358*, 745−750.
(59) Ono, L. K.; Liu, S.; Qi, Y. Reducing detrimental defects for high-performance metal halide perovskite solar cells. *Angew. Chem., Int. Ed.* **2020**, *59*, 6676−6698.
(60) Xiao, Z.; Du, K.-Z.; Meng, W.; Wang, J.; Mitzi, D. B.; Yan, Y. Intrinsic instability of $Cs_2In(I)M(III)X_6$ (M= Bi, Sb; X= halogen) double perovskites: a combined density functional theory and experimental study. *J. Am. Chem. Soc.* **2017**, *139*, 6054−6057.
(61) Song, T.-B.; Yokoyama, T.; Stoumpos, C. C.; Logsdon, J.; Cao, D. H.; Wasielewski, M. R.; Aramaki, S.; Kanatzidis, M. G. Importance of reducing vapor atmosphere in the fabrication of tin-based perovskite solar cells. *J. Am. Chem. Soc.* **2017**, *139*, 836−842.
(62) Yao, H.; Zhou, F.; Li, Z.; Ci, Z.; Ding, L.; Jin, Z. Strategies for improving the stability of tin-based perovskite ($ASnX_3$) solar cells. *Adv. Sci.* **2020**, *7*, 1903540.
(63) Chiara, R.; Morana, M.; Malavasi, L. Germanium-based halide perovskites: materials, properties, and applications. *ChemPlusChem* **2021**, *86*, 879−888.
(64) Kama, A.; Tirosh, S.; Itzhak, A.; Ejgenberg, M.; Cahen, D. New Pb-Free Stable Sn−Ge Solid Solution Halide Perovskites Fabricated by Spray Deposition. *ACS Appl. Energy Mater.* **2022**, *5*, 3638−3646.
(65) Azizman, M. S. A.; Azhari, A. W.; Halin, D. S. C.; Ibrahim, N.; Sepeai, S.; Ludin, N. A.; Nor, M. N. M.; Ho, L. N. Progress in tin-germanium perovskite solar cells: A review. *Synth. Met.* **2023**, *299*, 117475.
(66) Elliott, N.; Pauling, L. The crystal structure of cesium aurous auric chloride, $Cs_2AuAuCl_6$, and cesium argentous auric chloride, $Cs_2AgAuCl_6$. *J. Am. Chem. Soc.* **1938**, *60*, 1846−1851.
(67) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens.Matter* **2009**, *21*, 395502.
(68) Hamann, D. Optimized norm-conserving Vanderbilt pseudopotentials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *88*, 085117.
(69) Umari, P.; Pasquarello, A. Ab initio molecular dynamics in a finite homogeneous electric field. *Phys. Rev. Lett.* **2002**, *89*, 157602.
(70) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **2013**, *68*, 314−319.
(71) Ouyang, R.; Ahmetcik, E.; Carbogno, C.; Scheffler, M.; Ghiringhelli, L. M. Simultaneous learning of several materials properties from incomplete databases with multi-task SISSO. *J. Phys. Mater.* **2019**, *2*, 024002.
(72) Dean, J. A.; Lange, N. A. *Lange's Handbook of Chemistry*; McGraw-Hill: New York, 1999.
(73) Ouyang, R. Exploiting ionic radii for rational design of halide perovskites. *Chem. Mater.* **2020**, *32*, 595−604.

https://doi.org/10.1021/jacs.4c03507
J. Am. Chem. Soc. 2024, 146, 17636−17645