Laakso, Jarno; Todorovic, Milica; Li, Jingrui; Zhang, Guo-Xu; Rinke, Patrick

Compositional engineering of perovskites with machine learning

Published in:
Physical Review Materials

DOI:
10.1103/PhysRevMaterials.6.113801

Published: 07/11/2022

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Laakso, J., Todorovic, M., Li, J., Zhang, G.-X., & Rinke, P. (2022). Compositional engineering of perovskites with machine learning. *Physical Review Materials*, *6*(11), 1-10. Article 113801.
https://doi.org/10.1103/PhysRevMaterials.6.113801

PHYSICAL REVIEW MATERIALS **6**, 113801 (2022)

Compositional engineering of perovskites with machine learning

Jarno Laakso,[1] Milica Todorović,[2] Jingrui Li,[3] Guo-Xu Zhang,[4] and Patrick Rinke[1,*]
[1]*Department of Applied Physics, Aalto University, P.O. Box 11100, 00076 Aalto, Finland*
[2]*Department of Mechanical and Materials Engineering, University of Turku, 20014 Turku, Finland*
[3]*Electronic Materials Research Laboratory, Key Laboratory of the Ministry of Education & International Center for Dielectric Research, School of Electronic Science and Engineering, Xi'an Jiaotong University, 710049 Xi'an, China*
[4]*MIIT Key Laboratory of Critical Materials Technology for New Energy Conversion and Storage, School of Chemistry and Chemical Engineering, Harbin Institute of Technology, 150001 Harbin, China*

(Received 15 July 2022; accepted 11 October 2022; published 7 November 2022)

Perovskites are promising materials candidates for optoelectronics, but their commercialization is hindered by toxicity and materials instability. While compositional engineering can mitigate these problems by tuning perovskite properties, the enormous complexity of the perovskite materials space aggravates the search for an optimal optoelectronic material. We conducted compositional space exploration through Monte Carlo (MC) convex hull sampling, which we made tractable with machine learning (ML). The ML model learns from density functional theory calculations of perovskite atomic structures, and can be used for quick predictions of energies, atomic forces, and stresses. We employed it in structural relaxations combined with MC sampling to gain access to low-energy structures and compute the convex hull for $CsPb(Br_{1-x}Cl_x)_3$. The trained ML model achieves an energy prediction accuracy of 0.1 meV per atom. The resulting convex hull exhibits two stable mixing concentrations at 1/6 and 1/3 Cl contents. Our data-driven approach offers a pathway towards studies of more complex perovskites and other alloy materials with quantum mechanical accuracy.

DOI: 10.1103/PhysRevMaterials.6.113801

I. INTRODUCTION

Halide perovskite (ABX_3 with $X = Cl, Br, I$) materials have shown remarkable promise in emergent optoelectronic technologies. For example, perovskite solar cells (PSCs) have already achieved a record power-conversion efficiency of 25.5%, which is comparable to that of conventional crystalline silicon devices [1–3]. Notable advances have also been observed for perovskite-based light-emitting diodes (PeLEDs), which have achieved excellent monochromaticity, high external quantum efficiency, and high brightness [4–7]. The factors hindering the commercialization of halide perovskite-based materials are their instability against environmental stresses (such as heat, moisture, and oxygen) [8–11], and the toxicity of Pb as the most common B-site element [12–15].

The variety of candidates for the A, B, and X-site species makes it possible to tune the properties of perovskites via compositional engineering [16,17]. In most of today's high-performance PSCs both A and X sites are mixed: $(FA_{1-x-y}MA_xCs_y)Pb(I_{1-z}Br_z)_3$ where FA and MA stand for the two most common organic A-site cations $(H_2N)_2CH^+$ (formamidinium) and $CH_3NH_3^+$ (methylammonium), respectively, and x, y, and z indicate the mixing fractions of the elements [18–20]. This strategy is also widely utilized in PeLEDs, where halide mixing is the primary means of tuning the target emission wavelength [6,7,21], such as $MAPb(I_{1-x}Cl_x)_3$ for near-infrared [4,22], $MAPb(I_{1-x}Br_x)_3$ for red [4,23], and $MAPb(Br_{1-x}Cl_x)_3$ for blue emission [24,25].

Systematic screening of candidate materials for perovskite optoelectronics is still at its preliminary stage [17,20,26] due to the complexity of the multidimensional substitution space and the lack of fundamental understanding of the mechanisms that give rise to the property modifications. Atomistic first-principles computations can improve this understanding because they provide exact control over material nanostructure for property tuning studies, but employing them in compositional screening is challenging. Even for a specific composition such as $CsPb(Br_{1-x}Cl_x)_3$ with a fixed x value, the number of possible X-site (Cl and Br) configurations is infinite. On the one hand, large model systems are required to represent the (quasi)random Cl/Br configurations, and on the other hand, many possible structures with different configurations need to be explored. The intuitive ergodic approach to materials screening is thus intractable.

Within the past few years, machine learning (ML) has been applied to density functional theory (DFT) computational data sets of halide perovskites. Various ML-based attempts to identify possible stable perovskite structures have been reported, in which almost the entire periodic table was screened for candidate elements for A, B, and X sites [27–30]. Most of these works are based on DFT calculations of the smallest model systems. They also exclude perovskite alloys, leaving much of the materials space unexplored. Some ML studies of ionic mixing have emerged more recently: Park *et al.* predicted the possibility of forming mixed-ion perovskites based on ML similarity analysis performed on the DFT data set of "pure"

---

[*] patrick.rinke@aalto.fi

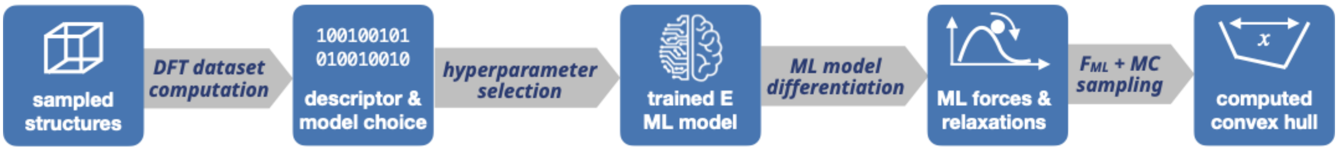

FIG. 1. The workflow of the DFT-based ML framework for configurational convex hull evaluation.

perovskites [31]. In the ML study by Lu *et al.*, a 5-atom unit cell model was used to represent mixed halide perovskites ABXX′X″ [32]. Limited compositions were included due to the perfect stoichiometry, and chemical disorder was ignored because the periodicity of the small unit cells did not allow it. Using the $2 \times 2 \times 2$ supercell, Mannodi-Kanakkithodi *et al.* studied the stability and properties of $\mathrm{MAPb_{0.875}B_{0.125}Br_3}$ with 55 candidate elements for B [33]. The supercell approach enabled a fine compositional study beyond stoichiometric systems, but substitution of only one Pb ion in the model system did not constitute realistic chemical disorder and the models are thus still rather simple.

In this paper, we present an ML framework for studying perovskite alloys based on DFT data. We demonstrate the approach for a cesium-lead-bromide-chloride alloy [$\mathrm{CsPb(Br_{1-x}Cl_x)_3}$]. Mixed bromide-chloride perovskites have recently generated interest as blue-light emitters in LEDs and near-ultraviolet light absorbers in photovoltaic devices [24,25]. As an inorganic binary alloy system, $\mathrm{CsPb(Br_{1-x}Cl_x)_3}$ serves as an appropriate starting point for future ML compositional engineering of more complex perovskite alloys. The chosen $2 \times 2 \times 2$ supercell model contains 24 halide anions so that the total number of possible configurations is $2^{24} \approx 1.68 \times 10^7$. This design space is already too large to traverse with DFT, even if structural symmetry reduces the number of configurations by one or two orders of magnitude. The traditional computational method for exploring such alloy spaces has been the cluster expansion (CE), which has also been applied in studying the stability of $\mathrm{CsPb(Br_{1-x}Cl_x)_3}$ [34,35]. Due to its rigid formalism, CE sometimes fails, for example, to accurately incorporate the effects of atomic relaxations [36]. Advanced data-driven research methods such as ML have recently emerged as a more flexible alternative to modeling alloy materials [37–39].

With the ultimate aim of discovering new materials, we design a multistep computational framework that is illustrated in Fig. 1. First, we generate a perovskite alloy data set with DFT. Different compositions, alignments, and structural distortions are considered to explore the space of structural variations. Based on these data, we develop and train an ML model for quick energy predictions. To map atomic structures to DFT total energies, our model combines two key components—the many-body tensor representation (MBTR) [40,41] for perovskite atomic structures and kernel-ridge regression (KRR) as the ML method. MBTR is a numerical descriptor for atomistic systems that provides detailed information of the local atomic structure and is compatible with kernel-based ML techniques [40]. KRR maps the MBTR vector of an atomic structure to its energy value. MBTR has been combined with KRR to successfully predict properties of molecules and solids [42–44]. In this paper, we extend previous research by deriving and incorporating the MBTR gradients. By differentiating our MBTR-KRR ML model, we can compute atomic forces and stress-tensor components and optimize the geometries of alloy configurations. Finally, we employ the ML geometry optimization to compute the convex hull for the binary alloy system, which is required for estimating which compositions form stable perovskites when synthesized: atomic configurations that are on the convex hull are stable while the rest are not. To this end, Monte Carlo (MC) sampling is utilized to explore the low-energy structures across all compositions.

Like our MBTR-KRR ML model, the CE model can be fitted to DFT calculations to provide fast energy predictions for atomic structures. However, structure enumeration in CE assumes ideal lattice positions and structural relaxation can only be incorporated through the training data by fitting to energies of relaxed geometries. Conversely, our approach provides the optimized atomic structures and lattice parameters in addition to the convex hull. Furthermore, with our approach a single ML model is capable of simulating different lattice types, which is not possible with the CE.

The outline for the remainder of this paper is as follows: In Sec. II, we briefly describe the methods involved in the development of our ML framework and its application in $\mathrm{CsPb(Br_{1-x}Cl_x)_3}$ stability computations. In Sec. III, we present the results of our study. In Sec. IV, we discuss the key findings, potential improvements, and future work. We conclude with a summary in Sec. V.

II. METHODS

A combination of methods is needed for our perovskite property-prediction ML framework. Here we describe how we (a) generated the $\mathrm{CsPb(Br_{1-x}Cl_x)_3}$ DFT data set, (b) developed and tested our ML model that learns from the DFT data and maps atomic structure to DFT energy, and (c) combined MC sampling with ML structure optimization to identify the low energy structures and compute the convex hull. The codes for all computational steps are available through a GitLab repository [45].

A. Data set computation

We generated a data set of atomic structures and DFT total energies for $\mathrm{CsPb(Br_{1-x}Cl_x)_3}$. We represented mixed-halide perovskites in a $2 \times 2 \times 2$ supercell containing 24 halide atoms. The data set consists of two types of structures. The first type are single point calculations featuring small deviations from the ideal atomic sites. Because our aim was to employ our ML model in geometry optimization, we also included structure snapshots from DFT relaxation simulations.

We started with the structures for single point calculations. We ensured the diversity of the data in terms of lattice

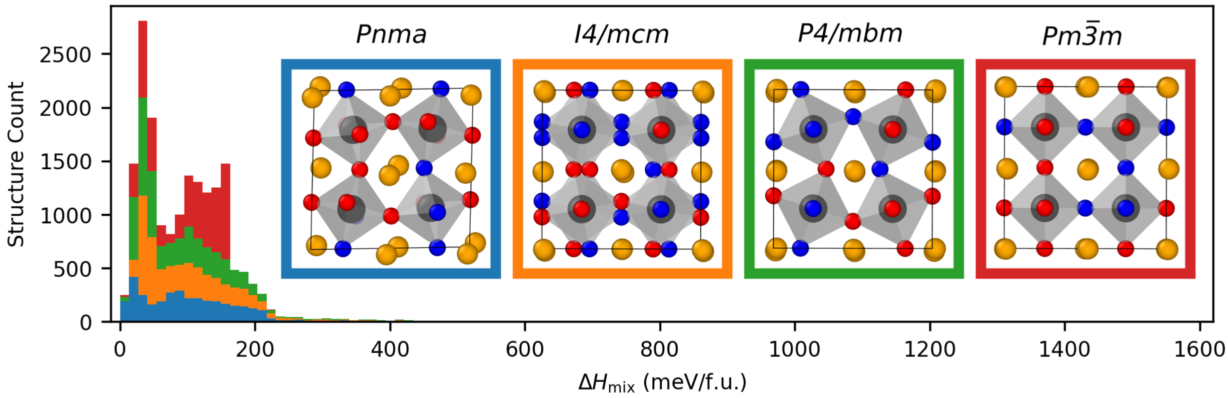

FIG. 2. Distribution of enthalpy of mixing (ΔH_{mix}) values per perovskite formula unit (f.u.) in the generated data set that consists of an equal number of structures from four different space groups ($Pm\bar{3}m$, $P4/mbm$, $I4/mcm$, and $Pnma$). The ΔH_{mix} values range from 0 to 1625 meV/f.u.

types (phases), compositions, configurations, lattice deformation, and deviations. First, we filled the substitution range uniformly by generating 100 structures per Cl concentration. This sums up to 10 000 structures, given the 25 Cl concentrations and four equally represented space groups ($Pm\bar{3}m$, $P4/mbm$, $I4/mcm$, and $Pnma$). To account for configurational diversity, the Cl/Br configuration was randomized in each alloy structure. We determined the lattice parameters of each alloy using Vegard's law [46]. Next, the atomic positions of each atomic structure were generated as follows: we initialized the atomic coordinates of each generated structure by linear interpolation between the relaxed $CsPbBr_3$ and $CsPbCl_3$ geometries. Then we randomly displaced Cs and halide ions to imitate structural fluctuations. Specifically, we displaced Cs positions uniformly within ±0.02 in terms of fractional coordinates along all three lattice vectors. The halide positions were displaced so that the octahedral tilting amplitude follows a $\sim(e^{-(x+1)^2} + e^{-(x-1)^2})$ distribution, which (a) is symmetric with respect to zero tilt, (b) has its maximum at the linearly-interpolated tilting structure, and (c) decays rapidly for larger tilting angles. We were thus able to capture disorder by generating many different structures, even for extremely low Cl or Br concentrations, with these random deviations. The total energies of the 10 000 structures were then computed with DFT.

Next, we generated the relaxation data. We used the same methodology as with the single point data to generate two more structures per Cl concentration for the four space groups, which sums up to 200 structures. The atomic positions in these structures were relaxed with DFT while keeping the lattice parameters fixed, which produced 8014 structure snapshots where both the total energies and the atomic forces are known. We then combined these structural snapshots with the 10 000 single-point structures to build the full data set of 18 014 structures. In the interest of open science [47], we made all relevant calculations for this paper available on the Novel Materials Laboratory (NOMAD) [48] and Zenodo [49].

Total DFT energies are difficult to learn directly due to their high dependence on the electron count in a structure. Instead, we calculated the enthalpy of mixing (ΔH_{mix}) for every structure in the data set,

$$\begin{aligned}\Delta H_{\mathrm{mix}} &= E(\mathrm{CsPb(Br}_{1-x}\mathrm{Cl}_x)_3) \\ &\quad - xE(\mathrm{CsPbCl_3}) - (1-x)E(\mathrm{CsPbBr_3}),\end{aligned} \tag{1}$$

where $E(\mathrm{CsPb(Br}_{1-x}\mathrm{Cl}_x)_3)$ is the DFT total energy per perovskite formula unit of a mixed structure. $E(CsPbCl_3)$ and $E(CsPbBr_3)$ are the DFT total energies per formula unit of the global minimum structures that we obtain for the pure $CsPbCl_3$ and $CsPbBr_3$ through DFT relaxation. Figure 2 shows the distribution of ΔH_{mix} values in the data set. We normalized the data so that the mean of the mixing enthalpies is 0 and the standard deviation is 1,

$$E_{\mathrm{label}} = \frac{\Delta H_{\mathrm{mix}} - \mu(\Delta H_{\mathrm{mix}})}{\sigma(\Delta H_{\mathrm{mix}})}. \tag{2}$$

Supervised ML requires different data pools for fitting the ML model and testing its performance, so we split the generated data into training and test sets. Our aim was to develop a model that performs well in structure optimization, which is reflected in our choice for the composition of the test set. We formed the test set by combining structure snapshots from 100 DFT relaxations. We selected these relaxations uniformly across the 25 concentration levels and four lattice types, ensuring that the materials space is well covered and the test results are not biased in terms of composition. The 10 000 single point structures and the remaining 100 DFT relaxations formed the training set. The resulting sizes for the training and test sets were 13 948 and 4 066 structures, respectively.

All DFT calculations were performed using the Perdew-Burke-Ernzerhof exchange-correlation functional for solids (PBEsol) [50] implemented in the all-electron numeric-atom-centered orbital code FHI-aims [51–53]. We chose the PBEsol functional because it predicts the lattice constants of halide perovskites well at a moderate computational cost [54,55]. Further computational settings included the standard FHI-aims tier-2 basis sets and "tight" grid settings, the zeroth-order regular approximation that accounts for scalar relativistic effects [56], and a Γ-centered 4 × 4 × 4 k-point mesh for Brillouin-zone integration. The DFT geometry optimizations had a convergence limit of 5 meV/Å for the maximum atomic

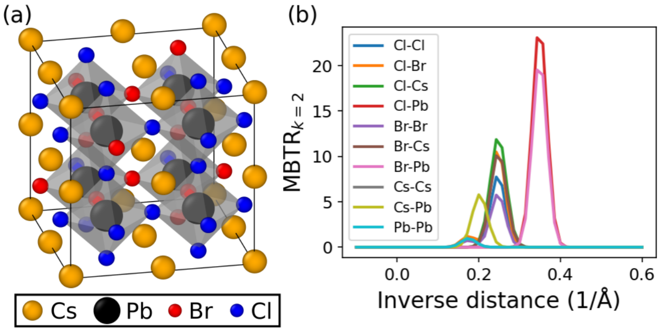

FIG. 3. (a) Example structure. (b) MBTR representation of (a).

force amplitude. Optimization of lattice constants was carried out using the analytical stress tensor [57].

B. Machine learning model

Model definition. We used supervised ML regression to map perovskite atomic structures to the corresponding energies. The ML model features two key aspects: representing the atomic structures in vector form with MBTR and mapping the MBTR vectors to E_{label} with KRR.

MBTR is a structural descriptor that considers structural motifs such as elemental contents ($k = 1$), interatomic distances ($k = 2$), and bond angles ($k = 3$) to form a vector representation of an atomic geometry [40,41]. A previous study [42] concluded that the $k = 1$ term did not improve the model accuracy when higher-order terms were included. The results of the same study, as well as our own tests, also indicate that including the $k = 3$ term produces a minimal improvement in accuracy while considerably increasing the computational cost of model predictions. In order to balance the accuracy and computational efficiency of our ML model, we employed only the $k = 2$ term in this paper.

The $k = 2$ term encodes each element pair (Z_1, Z_2) of a structure in a sum of Gaussians

$$\sum_{l\in Z_1}\sum_{m\in Z_2} w^{l,m}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-|\boldsymbol{R}_l-\boldsymbol{R}_m|^{-1})^2}{2\sigma^2}\right), \tag{3}$$

where the sum proceeds through all the atom pairs of elements Z_1 and Z_2 in a structure and $\boldsymbol{R}$ are the positions of these atoms. $w^{l,m}$ is a weighting function

$$w^{l,m} = \exp(-s|\boldsymbol{R}_l - \boldsymbol{R}_m|), \tag{4}$$

where s is a parameter that controls the magnitude of the weighting. The weighting function is required for the sum to converge in an infinite lattice of atoms. The parameter w_{cutoff} is a weight threshold for cutting off the sum, and with s it determines a cutoff distance r_{cutoff} for atom pairs that contribute to the representation. The MBTR functions are vectorized by evaluating them at N_{grid} discrete grid points spanning from x_{min} to x_{max}. The full vector representation $\boldsymbol{M}(\boldsymbol{s})$ of an atomic structure $\boldsymbol{s}$ is the concatenation of the element pair contributions. Figure 3(b) displays an example MBTR vector of a perovskite alloy structure.

We correlated the MBTR vectors with E_{label} using KRR. An energy prediction for a structure $\boldsymbol{s}$ is given by

$$E^{\mathrm{ML}}_{\mathrm{label}} = \sum_i^N \beta_i k(\boldsymbol{s}, \boldsymbol{s}_i), \tag{5}$$

where β_i are fitting coefficients, k is a kernel function, and $\boldsymbol{s}_i$ are a set of N reference structures. We chose the Gaussian kernel as it performs better than other common kernels when combined with MBTR according to an earlier study [42]. The Gaussian kernel function is

$$k(\boldsymbol{s}, \boldsymbol{s}') = \exp(-\gamma||\boldsymbol{M}(\boldsymbol{s}) - \boldsymbol{M}(\boldsymbol{s}')||_2^2), \tag{6}$$

where γ is a parameter that controls the width of the kernel distribution. The fitting coefficients β_i can be determined through

$$\boldsymbol{\beta} = (\boldsymbol{K} + \alpha\boldsymbol{I})^{-1}\boldsymbol{E}^{\mathrm{ref}}_{\mathrm{label}}, \tag{7}$$

where $\boldsymbol{K}$ is the kernel matrix $K_{i,j} := k(\boldsymbol{s}_i, \boldsymbol{s}_j)$, $\boldsymbol{E}^{\mathrm{ref}}_{\mathrm{label}}$ are the energy labels of the reference data set, and α is a regularization parameter.

Model selection. To achieve optimal ML model performance, we optimized its hyperparameters. The ML model has eight hyperparameters: six for MBTR (x_{min}, x_{max}, N_{grid}, s, w_{cutoff}, and σ) and two for KRR (α and γ). We set some of the hyperparameters based on the characteristics of the data and the implications that they have on the ML model efficiency, as described in the Supplemental Material (SM) [58]. The remaining hyperparameters we optimized with the Bayesian optimization code BOSS [59] following the procedure outlined in Ref. [60]. More details on the hyperparameter optimization and the hyperparameter values can be found in Sec. S2.A of the SM [58].

C. ML structure optimization and convex hull

To use our ML model for perovskite structure optimizations, the model needs to be capable of predicting atomic forces and stress tensor components, i.e., the energy derivatives. Instead of training the model to predict them directly, we differentiate the energy model predictions. The force on atom a relates to the energy gradient with respect to atomic positions

$$\boldsymbol{F}_a = -\nabla_a E. \tag{8}$$

Similarly, components of the stress tensor follow from derivatives with respect to strain

$$\sigma_{\mu\nu} = \frac{1}{V}\frac{\partial E}{\partial\epsilon_{\mu\nu}}, \tag{9}$$

where V is the volume of the simulation cell and $\epsilon_{\mu\nu}$ a component of the strain tensor. Inserting Eqs. (1) and (2) produces the expressions for the forces and stresses

$$\boldsymbol{F}^{\mathrm{ML}}_a = -8\sigma(\Delta H_{\mathrm{mix}})\nabla_a E^{\mathrm{ML}}_{\mathrm{label}}, \tag{10}$$

$$\sigma^{\mathrm{ML}}_{\mu\nu} = \frac{8\sigma(\Delta H_{\mathrm{mix}})}{V}\frac{\partial E^{\mathrm{ML}}_{\mathrm{label}}}{\partial\epsilon_{\mu\nu}}. \tag{11}$$

KRR and MBTR gradients are required for the derivatives of $E^{\mathrm{ML}}_{\mathrm{label}}$. The final expressions are derived in the SM [58].

Structure optimization with our ML model was carried out with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) minimizer implemented in the ASE package [61]. At each BFGS iteration, the forces and stresses were predicted with the ML model and both atomic positions and lattice parameters were updated accordingly until the maximum force component was smaller than 5 meV/Å.

Convex hull calculations. To map out the convex hull, we needed to find the structure with the lowest energy for each Cl concentration. This was a demanding task due to the large number of possible Cl/Br configurations. We made it tractable by applying our ML model.

We searched for the minimum configurations with Monte Carlo simulated annealing within the most stable *Pnma* phase. At each Cl concentration, we initialized the MC algorithm with a randomized Cl/Br configuration. With each subsequent iteration, the MC algorithm generated a new configuration by swapping a random pair of Cl and Br atoms. We then relaxed the crystal and atomic structures using the ML model. The energy of the relaxed structure was predicted with the ML model and compared to the energy before the atom pair swap. According to the Metropolis principle, if the energy decreased, the change was accepted and in its next step the algorithm continued from the modified Cl/Br configuration; if the energy increased, the change was accepted with the probability

$$P(\text{accept}) = \exp\left(-\frac{\Delta E^{\text{ML}}}{k_B T}\right), \tag{12}$$

where T is the simulated temperature.

We ran the MC algorithm for 200 steps, while decreasing T linearly from 10.0 K to 1.0×10^{-6} K. Due to the limited attempts, randomness of the initial Cl/Br configurations, and the choice of swapped atoms, we were not guaranteed to find the absolute minimum configuration every time. For a more accurate minimum search, we ran the MC sampling multiple times by initializing the algorithm with different configurations. We tested different restart counts, and found that repeating the sampling of 200 structures five times identified the minimum energy configuration with over 99% probability.

During the sampling, the MC algorithm might have constructed multiple symmetry-equivalent Cl/Br configurations. The relaxed energies of these configurations should be the same, but due to small uncertainties in the ML model, the predicted energies differed by up to 3 meV/f.u. We subsequently picked the median energy as representative of such a group of structures.

After using MC sampling to identify the minimum energy configurations for all concentrations, we calculated single point DFT energies for the corresponding ML optimized structures to make sure that the ML model is predicting their energies accurately. We then relaxed the minimum energy structures with DFT starting from the same initial geometry as the ML relaxation did. The energies obtained from the DFT relaxation are free from the effects of the limited ML model accuracy and thus constitute our most accurate estimates for the minimum configuration energies. We used these energies to determine the convex hull for $\text{CsPb}(\text{Br}_{1-x}\text{Cl}_x)_3$.

TABLE I. Prediction errors for our fully trained ML model for different phases. MAE_E is the energy prediction error. Initial MAE_F is the force prediction error on the initial structures (relaxation step 0) in the test set. MAE_F is the force prediction error over the whole test set. MAE_{rel} is the mean absolute difference between the ML predicted energies of ML optimized structures and DFT energies of the DFT optimized structures.

	MAE_E (meV/f.u.)	Initial MAE_F (meV/Å)	MAE_F (meV/Å)	MAE_{rel} (meV/f.u.)
$Pm\bar{3}m$	0.83	12.9	18.0	4.72
$P4/mbm$	0.50	13.7	16.5	1.98
$I4/mcm$	0.71	13.7	16.5	5.74
$Pnma$	0.70	15.3	16.7	1.32
Average	0.69	13.9	17.0	3.44

We also analyzed the relationship between the changes in energies and atomic structures. We defined a distance metric between two perovskite structures as the Euclidean distance between the MBTR vectors of the structures. To make interpretation easier, we used the same MBTR hyperparameters as for the predictive model, but normalized all the MBTR vectors to L2 norm of 1.

III. RESULTS

In this section we present results from model performance tests and $\text{CsPb}(\text{Br}_{1-x}\text{Cl}_x)_3$ convex hull computations. First, we assess the energy and force prediction accuracy of the ML model. Next, we analyze how the model performs in structure optimization. Finally, we present the results of the MC sampling and the convex hull.

A. Energy and force prediction

We trained our ML model with the optimized hyperparameters and predicted the energies of the test set structures. We then compared the predicted energies to the DFT values to obtain an estimate for the accuracy of the model,

$$\text{MAE}_E = \frac{1}{N}\sum_j |E^{\text{ML}}(\boldsymbol{s}_j) - E^{\text{DFT}}(\boldsymbol{s}_j)|, \tag{13}$$

where j runs over the test set structures $\boldsymbol{s}_j$ and N is the number of structures in the test set. We repeated the training procedure with different training set sizes ranging from 100 to ∼14 000 to generate the learning curves presented in Fig. 4(a). The prediction errors decrease rapidly with increasing training set size (irrespective of the space group) for about 1 000 training structures, and then begin to converge. The overall prediction error for the full training set is 0.69 meV/f.u. This corresponds to 0.14 meV per atom. Out of the four space groups, energies of $P4/mbm$ structures were the easiest to learn (see Table I).

The next step was to assess how well the ML model performs in force prediction tasks. We predicted the atomic forces for the test set structures and compared the results to the DFT

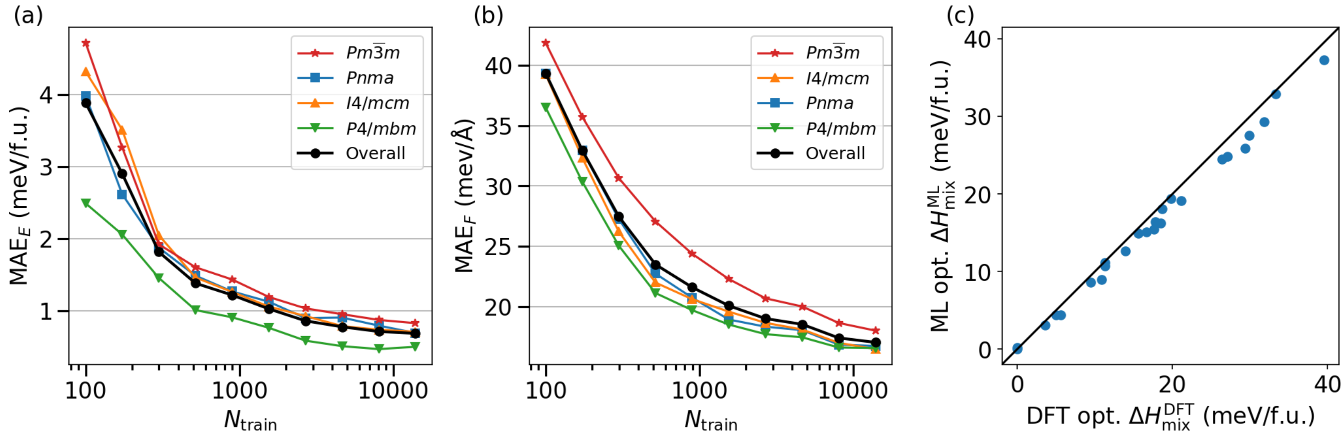

FIG. 4. (a) Energy learning curves. (b) Force learning curves. (c) ML optimized energies of the *Pnma* test structures compared to DFT optimization results.

values. The error metric that we used is

$$\mathrm{MAE}_F = \frac{1}{N}\sum_j \frac{1}{M}\sum_{a,i}\left|F_{a,i}^{\rm ML} - F_{a,i}^{\rm DFT}\right|, \tag{14}$$

where j runs over the test structures, a over the atoms in a structure, and i over the three components of a force vector. N corresponds to the number of test structures and M to the number of force components in a structure. We again repeated the test with different training set sizes to plot force learning curves [Fig. 4(b)]. The errors behave similarly to the energy prediction errors, decreasing quickly with the increasing training set size. The overall MAE$_F$ for the full training set is 17 meV/Å (see Table I). $Pm\bar{3}m$ stands out as the space group that was the most difficult to learn.

B. Structure optimization

Our aim was to employ the ML model in structure optimization for convex hull calculations. In the final accuracy check, we analyzed the performance of the ML model on the precomputed DFT relaxation trajectories by predicting the ML forces of relaxation snapshots and comparing them to the DFT forces of the same structures. We determined whether the ML model accuracy remains high throughout the entire relaxation trajectories by investigating how the force prediction error evolves with the number of relaxation steps.

Our test set consisted of snapshots from 100 DFT structure relaxations. We grouped these structures based on their relaxation step index, and calculated MAE$_F$ for each group using the ML model that was fitted on the full training set. During the data generation, the initial structures of the DFT relaxations (relaxation step 0) were generated in the same way as the single point structures that constitute the majority of the training set of the ML model, and thus the prediction accuracy on them provides an estimate of the optimal ML model performance. The initial MAE$_F$ differentiated by space group are shown in Table I. The mean deviation for all 100 initial structures is 13.9 meV/Å. Figure 5 shows how the force prediction error evolves as the relaxation proceeds for the 25 *Pnma* structures in the test set. The average error increases from the initial value but remains under 20 meV/Å. The maximum error remains under 40 meV/Å.

After analyzing the ML force predictions on DFT relaxations, we proceeded to test the structure optimization performance of the fully trained ML model. We performed ML optimizations starting from the initial structures of the DFT relaxations and compared the predicted energies to the DFT results. To make DFT and ML relaxations comparable in this test, we fixed the lattice parameters and only optimized the atomic positions. To assess the quality of the geometry optimization, we compared the ML energy of the final ML geometry to the DFT energy of the final DFT geometry.

All our ML structure optimizations converged. The MAE of optimized energies is 3.44 meV/f.u. overall, but depends significantly on the space group of the optimized structure (see Table I). The results for the 25 *Pnma* test structures are displayed in Fig. 4(c). The ML model systematically underestimates the optimized energy, but the MAE is only 1.32 meV/f.u.

C. Convex hull

The good accuracy of the ML structure optimizations makes them suitable for the convex hull computation for CsPb(Br$_{1-x}$Cl$_x$)$_3$. We employed MC sampling to access the low energy structures. The ML optimizations of all sampled structures converged. We obtained negative $\Delta H_{\rm mix}$ values over

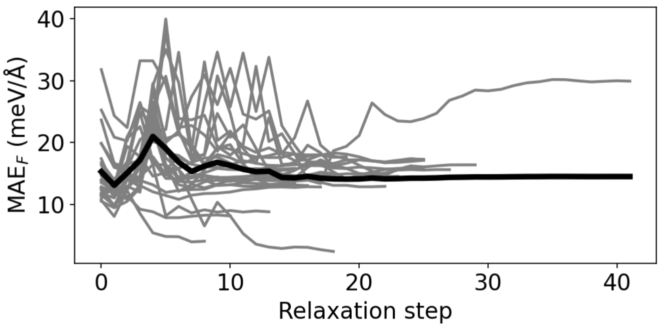

FIG. 5. Evolution of the force prediction error in the DFT relaxations of the *Pnma* test structures. The gray lines are individual relaxation trajectories. The mean of all relaxations is shown with a black line.

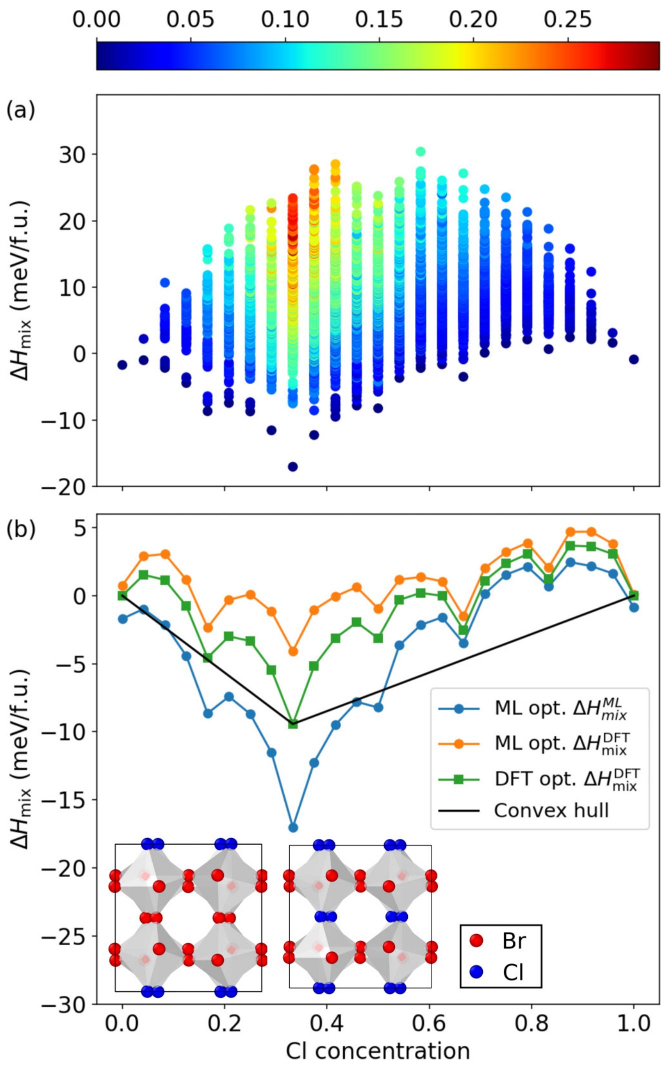

FIG. 6. (a) ML energies of the MC samples. Color coding is based on the MBTR distance between the sample and the minimum structure of the same Cl concentration. (b) ML and DFT relaxed energies for the minimum energy structures. The convex hull has been constructed based on the DFT results. Stable structures are found at 1/6 and 1/3 Cl concentrations.

a large part of the Cl concentration range with the lowest value at 1/3 Cl concentration, as illustrated in Fig. 6. We computed the energies of the ML-optimized minimum energy structures at each Cl concentration with DFT, which revealed that the ML model systematically underestimated their energies by 5.1 meV/f.u. on average. When we optimized the minimum structures with DFT starting from the same initial structures as the ML relaxations, the resulting energies were closer to the ML predictions but still underestimated by 3.2 meV/f.u. on average. We analyzed the difference between the DFT and ML relaxed structures and observed that the root-mean-square deviation (RMSD) of atomic positions between them was 0.03 Å on average, while the mean deviation of bond lengths was 0.02 Å.

We constructed a convex hull based on the energies of the DFT optimized structures. There are two structures on the hull: one at 1/6 and the other one at 1/3 Cl concentration.

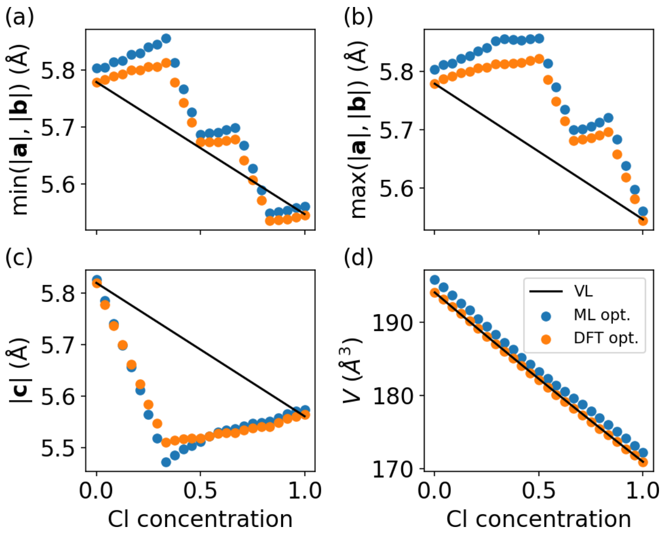

FIG. 7. Lattice vector lengths and unit cell volumes of $\mathrm{CsPb(Br_{1-x}Cl_x)_3}$ minimum energy structures after DFT and ML structure optimization compared to Vegards's law (VL): (a) length of the shorter of the two lattice vectors ***a*** and ***b***, (b) length of the longer of the two lattice vectors ***a*** and ***b***, (c) length of the lattice vector ***c***, and (d) unit cell volume.

Both of these structures are similar, with Cl and Br atoms ordered in layers. The preference for layered ordering is also shown in the periodicity of the minimum energies. Every four Cl concentration steps, the energy reaches a low point when a layer is filled. The other layered structures at 2/3 and 5/6 Cl content are nonetheless above the convex hull, which makes the hull unsymmetrical.

We analyzed the change in lattice parameters due to DFT and ML structure optimization by comparing how the relaxed lattice parameters of the minimum energy structures compare to Vegard's law. The results of the analysis are presented in Fig. 7. With both structure optimization methods, the $a-$ and $b-$ directional lattice vectors tend to elongate after the relaxation in contrast to Vegard's law values. At the same time, the relaxed structures are flatter in $c-$ direction. DFT relaxation preserved the overall volume of the perovskite simulation cells, whereas ML relaxation overestimated the volume systematically by 0.8% on average.

IV. DISCUSSION

Our ML model reached an energy prediction MAE of close to 0.1 meV per atom at a computational cost six orders of magnitude lower than DFT. This very low error value can be partly explained by data set redundancy: the atomic structures in our data set are very similar to each other, which makes it possible to reach very small prediction errors. The force component prediction MAE of the model is 17 meV/Å, which is over three times higher than our convergence criterion for the structure optimizations. Despite this, the ML model produced relaxations that were very similar to DFT results in terms of predicted energies, lattice parameters, and atomic positions in the relaxed structures.

When we employed the ML model in structure optimization, the accuracy depended significantly on the space group of the optimized structure. This is due to the fact that the number of iterations needed for an optimization to converge is different for all the space groups (Sec. S2.B of the SM [58]), and fewer iterations means less error propagation. On average, the *Pnma* optimizations are the shortest, which is reflected in the ML model reaching a low MAE of 1.3 meV/f.u. The ML structure optimizations are not only accurate but also very quick. Relaxing an atomic structure with the ML model is over four orders of magnitude faster than with DFT.

The convex hull that we computed for $CsPb(Br_{1-x}Cl_x)_3$ exhibits two stable mixing concentrations at $x = 1/6$ and $x = 1/3$. The stable structures are highly regular, with layered Cl/Br configurations. In the case of this material, the same results might have been obtained without the extensive MC exploration of structures, by systematically examining only the few regular configurations. In our approach, however, we made no assumptions about the regularity of the stable alloy configurations, and thus our approach for computing the convex hull would also work in the general case.

The shape of the convex hull that we obtained, as well as the observed preference for layered configuration of the Cl and Br atoms, agree with previous computational studies of the same material [34,35]. The main difference between our approach and the previous studies based on CE is that our ML relaxations provide direct access to the optimized geometries. Additionally, the earlier studies used a smaller supercell of 20 atoms to fit their CE models.

$CsPb(Br_{1-x}Cl_x)_3$ has been synthesized before in an experimental study that observed the material to be stable throughout the Cl concentration range [62]. In contrast, our computations predict only two stable concentration values. We suspect that the discrepancy between the experiments and our results is due to the fact that we modeled the system at 0 K and thus did not consider the effects of configurational entropy related to the Cl/Br mixing. In future work we will fabricate $CsPb(Br_{1-x}Cl_x)_3$ to verify the accuracy of our ML approach. In order to close the gap between computations and experiments, we plan to use advanced sampling methods and include configurational entropy in our computation of the convex hull.

We analyzed how the lattice parameters of relaxed geometries compare to the predictions from Vegard's law. Both the DFT and ML relaxations that we performed produce optimal lattice parameters that deviate significantly from Vegard's law. We repeated the MC sampling without optimizing the lattice parameters with the ML model (see Sec. S2.C of the SM [58]), and the resulting convex hull was very different. We conclude that in order to obtain the correct minimum energy Cl/Br configurations and capture the shape of the convex hull, optimizing the lattice parameters of each sampled structure is very important.

While generating the training data structures for the ML model, we introduced random variation to the octahedral tilting angles and thus to the simulation cell shape. This adds to the diversity of the training data, which helps the ML model retain its accuracy when the cell shape changes during the structure optimizations. The accuracy could likely be increased further by including snapshots from DFT cell optimization simulations into the training set.

In this paper, we fitted the KRR model purely based on energies. One potential way to increase the model's force prediction accuracy in the future would be to include atomic forces in the fitting process. One could define a KRR loss function that includes the forces, and refit the the model by finding the fitting coefficients that minimize this new loss function. The computational costs related to the model fitting would be larger with this approach, but the force predictions would likely be improved.

We applied the ML framework in the study of perovskites, but the same methods could be utilized also for other materials. This paper can benefit problems where the number of atomic systems that need to be considered is too large for DFT computations. This is often the case in the study of alloys, for which our approach can be applied by simply replacing the data set that is used for fitting the ML model.

V. CONCLUSIONS

In this study, we have developed and tested an ML framework for perovskite property prediction. We have demonstrated that our ML model that was trained on DFT calculations of $CsPb(Br_{1-x}Cl_x)_3$ atomic structures is capable of predicting energies with quantum mechanical precision. Our tests also show that the ML model can be used for accurate prediction of atomic forces as well as for structure optimization. The ML predictions are over four orders of magnitude faster than DFT.

We have employed the ML framework to study the stability of $CsPb(Br_{1-x}Cl_x)_3$. Our convex hull computations exhibit two stable mixing concentrations at $x = 1/6$ and $x = 1/3$. We observe that $CsPb(Br_{1-x}Cl_x)_3$ prefers layered Cl/Br alignments. Our approach to computing the convex hull is directly applicable to the study of different perovskite materials and other alloys.

ACKNOWLEDGMENTS

The authors wish to acknowledge Dr. A. Tiihonen for insightful discussions. This study was supported by the Academy of Finland through Project No. 334532. We further acknowledge CSC-IT Center for Science, Finland, the Aalto Science-IT project, and Xi'an Jiaotong University's HPC Platform for generous computational resources.

---

[1] H.-S. Kim, C.-R. Lee, J.-H. Im, K.-B. Lee, T. Moehl, A. Marchioro, S.-J. Moon, R. Humphry-Baker, J.-H. Yum, J. E. Moser *et al.*, Lead iodide perovskite sensitized all-solid-state submicron thin film mesoscopic solar cell with efficiency exceeding 9%, Sci. Rep. **2**, 591 (2012).

[2] D. P. McMeekin, S. Mahesh, N. K. Noel, M. T. Klug, J. Lim, J. H. Warby, J. M. Ball, L. M. Herz, M. B. Johnston, and H. J. Snaith, Solution-processed all-perovskite multi-junction solar cells, Joule **3**, 387 (2019).

[3] H. Min, D. Y. Lee, J. Kim, G. Kim, K. S. Lee, J. Kim, M. J. Paik, Y. K. Kim, K. S. Kim, M. G. Kim *et al.*, Perovskite solar cells with atomically coherent interlayers on SnO_2 electrodes, Nature (London) **598**, 444 (2021).

[4] Z.-K. Tan, R. S. Moghaddam, M. L. Lai, P. Docampo, R. Higler, F. Deschler, M. Price, A. Sadhanala, L. M. Pazos, D. Credgington *et al.*, Bright light-emitting diodes based on organometal halide perovskite, Nat. Nanotechnol. **9**, 687 (2014).

[5] S. D. Stranks and H. J. Snaith, Metal-halide perovskites for photovoltaic and light-emitting devices, Nat. Nanotechnol. **10**, 391 (2015).

[6] M. Lu, Y. Zhang, S. Wang, J. Guo, W. W. Yu, and A. L. Rogach, Metal halide perovskite light-emitting devices: Promising technology for next-generation displays, Adv. Funct. Mater. **29**, 1902008 (2019).

[7] X.-K. Liu, W. Xu, S. Bai, Y. Jin, J. Wang, R. H. Friend, and F. Gao, Metal halide perovskites for light-emitting diodes, Nat. Mater. **20**, 10 (2021).

[8] J. Huang, S. Tan, P. D. Lund, and H. Zhou, Impact of H_2O on organic–inorganic hybrid perovskite solar cells, Energy Environ. Sci. **10**, 2284 (2017).

[9] Y. Zhou and Y. Zhao, Chemical stability and instability of inorganic halide perovskites, Energy Environ. Sci. **12**, 1495 (2019).

[10] B.-w. Park and S. I. Seok, Intrinsic instability of inorganic-organic hybrid halide perovskite materials, Adv. Mater. **31**, 1805337 (2019).

[11] B. Chen, S. Wang, Y. Song, C. Li, and F. Hao, A critical review on the moisture stability of halide perovskite films and solar cells, Chem. Eng. J. **430**, 132701 (2022).

[12] F. Giustino and H. J. Snaith, Toward lead-free perovskite solar cells, ACS Energy Lett. **1**, 1233 (2016).

[13] M. Konstantakou and T. Stergiopoulos, A critical review on tin halide perovskite solar cells, J. Mater. Chem. A **5**, 11518 (2017).

[14] W. Ke and M. G. Kanatzidis, Prospects for low-toxicity lead-free perovskite solar cells, Nat. Commun. **10**, 965 (2019).

[15] Q. Tai, J. Cao, T. Wang, and F. Yan, Recent advances toward efficient and stable tin-based perovskite solar cells, EcoMat **1**, e12004 (2019).

[16] K. Wang, Z. Jin, L. Liang, H. Bian, D. Bai, H. Wang, J. Zhang, Q. Wang, and S. Liu, All-inorganic cesium lead iodide perovskite solar cells with stabilized efficiency beyond 15%, Nat. Commun. **9**, 4544 (2018).

[17] M. Saliba, Polyelemental, multicomponent perovskite semiconductor libraries through combinatorial screening, Adv. Energy Mater. **9**, 1803754 (2019).

[18] J.-P. Correa-Baena, A. Abate, M. Saliba, W. Tress, T. J. Jacobsson, M. Grätzel, and A. Hagfeldt, The rapid evolution of highly efficient perovskite solar cells, Energy Environ. Sci. **10**, 710 (2017).

[19] J.-P. Correa-Baena, Y. Luo, T. M. Brenner, J. Snaider, S. Sun, X. Li, M. A. Jensen, N. T. P. Hartono, L. Nienhaus, S. Wieghold *et al.*, Homogenized halides and alkali cation segregation in alloyed organic-inorganic perovskites, Science **363**, 627 (2019).

[20] S. Sun III, A. Tiihonen, F. Oviedo, Z. Liu, J. Thapa, Y. Zhao, N. T. P. Hartono, A. Goyal, T. Heumueller, C. Batali *et al.*, A data fusion approach to optimize compositional stability of halide perovskites, Matter **4**, 1305 (2021).

[21] S. Adjokatse, H.-H. Fang, and M. A. Loi, Broadly tunable metal halide perovskites for solid-state light-emission applications, Mater. Today **20**, 413 (2017).

[22] L. Zhao, K. M. Lee, K. Roh, S. U. Z. Khan, and B. P. Rand, Improved outcoupling efficiency and stability of perovskite light-emitting diodes using thin emitting layers, Adv. Mater. **31**, 1805836 (2019).

[23] T. Chiba, Y. Hayashi, H. Ebe, K. Hoshi, J. Sato, S. Sato, Y.-J. Pu, S. Ohisa, and J. Kido, Anion-exchange red perovskite quantum dots with ammonium iodine salts for highly efficient light-emitting devices, Nat. Photonics **12**, 681 (2018).

[24] M. Karlsson, Z. Yi, S. Reichert, X. Luo, W. Lin, Z. Zhang, C. Bao, R. Zhang, S. Bai, G. Zheng *et al.*, Mixed halide perovskites for spectrally stable and high-efficiency blue light-emitting diodes, Nat. Commun. **12**, 361 (2021).

[25] M. Luo, Y. Jiang, T. He, and M. Yuan, Metal halide perovskites for blue light emitting materials, APL Mater. **8**, 040907 (2020).

[26] J. C. Stanley, F. Mayr, and A. Gagliardi, Machine learning stability and bandgaps of lead-free perovskites for photovoltaics, Adv. Theory Simul. **3**, 1900178 (2020).

[27] J. Schmidt, J. Shi, P. Borlido, L. Chen, S. Botti, and M. A. L. Marques, Predicting the thermodynamic stability of solids combining density functional theory and machine learning, Chem. Mater. **29**, 5090 (2017).

[28] C. J. Bartel, C. Sutton, B. R. Goldsmith, R. Ouyang, C. B. Musgrave, L. M. Ghiringhelli, and M. Scheffler, New tolerance factor to predict the stability of perovskite oxides and halides, Sci. Adv. **5**, eaav0693 (2019).

[29] S. Lu, Q. Zhou, Y. Ouyang, Y. Guo, Q. Li, and J. Wang, Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning, Nat. Commun. **9**, 3405 (2018).

[30] Z. Li, Q. Xu, Q. Sun, Z. Hou, and W.-J. Yin, Thermodynamic stability landscape of halide double perovskites via high-throughput computing and machine learning, Adv. Funct. Mater. **29**, 1807280 (2019).

[31] H. Park, R. Mall, F. H. Alharbi, S. Sanvito, N. Tabet, H. Bensmail, and F. El-Mellouhi, Exploring new approaches towards the formability of mixed-ion perovskites by DFT and machine learning, Phys. Chem. Chem. Phys. **21**, 1078 (2019).

[32] S. Lu, Q. Zhou, L. Ma, Y. Guo, and J. Wang, Rapid discovery of ferroelectric photovoltaic perovskites and material descriptors via machine learning, Small Methods **3**, 1900360 (2019).

[33] A. Mannodi-Kanakkithodi, J.-S. Park, N. Jeon, D. H. Cao, D. J. Gosztola, A. B. F. Martinson, and M. K. Y. Chan, Comprehensive computational study of partial lead substitution in methylammonium lead bromide, Chem. Mater. **31**, 3599 (2019).

[34] W.-J. Yin, Y. Yan, and S.-H. Wei, Anomalous alloy properties in mixed halide perovskites, J. Phys. Chem. Lett. **5**, 3625 (2014).

[35] J. S. Bechtel and A. Van der Ven, First-principles thermodynamics study of phase stability in inorganic halide perovskite solid solutions, Phys. Rev. Mater. **2**, 045401 (2018).

[36] A. H. Nguyen, C. W. Rosenbrock, C. S. Reese, and G. L. W. Hart, Robustness of the cluster expansion: Assessing the roles of relaxation and numerical error, Phys. Rev. B **96**, 014107 (2017).

[37] A. Shapeev, Accurate representation of formation energies of crystalline alloys with many components, Comput. Mater. Sci. **139**, 26 (2017).

[38] A. R. Natarajan and A. Van der Ven, Machine-learning the configurational energy of multicomponent crystalline solids, npj Comput. Mater. **4**, 56 (2018).

[39] N. Artrith, A. Urban, and G. Ceder, Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species, Phys. Rev. B **96**, 014112 (2017).

[40] H. Huo and M. Rupp, Unified representation of molecules and crystals for machine learning, arXiv:1704.06439.

[41] L. Himanen, M. O. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke, and A. S. Foster, Dscribe: Library of descriptors for machine learning in materials science, Comput. Phys. Commun. **247**, 106949 (2020).

[42] A. Stuke, M. Todorović, M. Rupp, C. Kunkel, K. Ghosh, L. Himanen, and P. Rinke, Chemical diversity in molecular orbital energy predictions with kernel ridge regression, J. Chem. Phys. **150**, 204121 (2019).

[43] E. Lumiaro, M. Todorović, T. Kurten, H. Vehkamäki, and P. Rinke, Predicting gas–particle partitioning coefficients of atmospheric molecules with machine learning, Atmos. Chem. Phys. **21**, 13227 (2021).

[44] C. Nyshadham, M. Rupp, B. Bekker, A. V. Shapeev, T. Mueller, C. W. Rosenbrock, G. Csányi, D. W. Wingate, and G. L. W. Hart, Machine-learned multi-system surrogate models for materials prediction, npj Comput. Mater. **5**, 51 (2019).

[45] GitLab repository, https://gitlab.com/cest-group/learnsolar-cspbclbr

[46] A. R. Denton and N. W. Ashcroft, Vegard's law, Phys. Rev. A **43**, 3161 (1991).

[47] L. Himanen, A. Geurts, A. S. Foster, and P. Rinke, Data-driven materials science: Status, challenges, and perspectives, Adv. Sci. **6**, 1900808 (2019).

[48] NOMAD repository, https://doi.org/10.17172/NOMAD/2022.10.06-1.

[49] Zenodo repository, https://doi.org/10.5281/zenodo.7153019.

[50] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke, Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces, Phys. Rev. Lett. **100**, 136406 (2008).

[51] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, *Ab initio* molecular simulations with numeric atom-centered orbitals, Comput. Phys. Commun. **180**, 2175 (2009).

[52] V. Havu, V. Blum, P. Havu, and M. Scheffler, Efficient $O(N)$ integration for all-electron electronic structure calculation using numeric basis functions, J. Comput. Phys. **228**, 8367 (2009).

[53] S. V. Levchenko, X. Ren, J. Wieferink, R. Johanni, P. Rinke, V. Blum, and M. Scheffler, Hybrid functionals for large periodic systems in an all-electron, numeric atom-centered basis framework, Comput. Phys. Commun. **192**, 60 (2015).

[54] R. X. Yang, J. M. Skelton, E. L. da Silva, J. M. Frost, and A. Walsh, Spontaneous octahedral tilting in the cubic inorganic cesium halide perovskites $CsSnX_3$ and $CsPbX_3$ ($X = F$, Cl, Br, I), J. Phys. Chem. Lett. **8**, 4720 (2017).

[55] M. Bokdam, J. Lahnsteiner, B. Ramberger, T. Schäfer, and G. Kresse, Assessing Density Functionals Using Many Body Theory for Hybrid Perovskites, Phys. Rev. Lett. **119**, 145501 (2017).

[56] E. v. Lenthe, E. J. Baerends, and J. G. Snijders, Relativistic regular two-component Hamiltonians, J. Chem. Phys. **99**, 4597 (1993).

[57] F. Knuth, C. Carbogno, V. Atalla, V. Blum, and M. Scheffler, All-electron formalism for total energy strain derivatives and stress tensor components for numeric atom-centered orbitals, Comput. Phys. Commun. **190**, 33 (2015).

[58] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevMaterials.6.113801 for additional details on the machine learning model and its derivatives, hyperparameter optimization, and analysis of the computational results.

[59] M. Todorović, M. U. Gutmann, J. Corander, and P. Rinke, Bayesian inference of atomistic structure in functional materials, npj Comput. Mater. **5**, 35 (2019).

[60] A. Stuke, P. Rinke, and M. Todorović, Efficient hyperparameter tuning for kernel ridge regression with Bayesian optimization, Mach. Learn. Sci. Technol. **2**, 035022 (2021).

[61] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus *et al.*, The atomic simulation environment—A python library for working with atoms, J. Phys.: Condens. Matter **29**, 273002 (2017).

[62] T. G. Liashenko, E. D. Cherotchenko, A. P. Pushkarev, V. Pakštas, A. Naujokaitis, S. A. Khubezhov, R. G. Polozkov, K. B. Agapev, A. A. Zakhidov, I. A. Shelykh, and S. V. Makarov, Electronic structure of $CsPbBr_{3-x}Cl_x$ perovskites: Synthesis, experimental characterization, and DFT simulations, Phys. Chem. Chem. Phys. **21**, 18930 (2019).