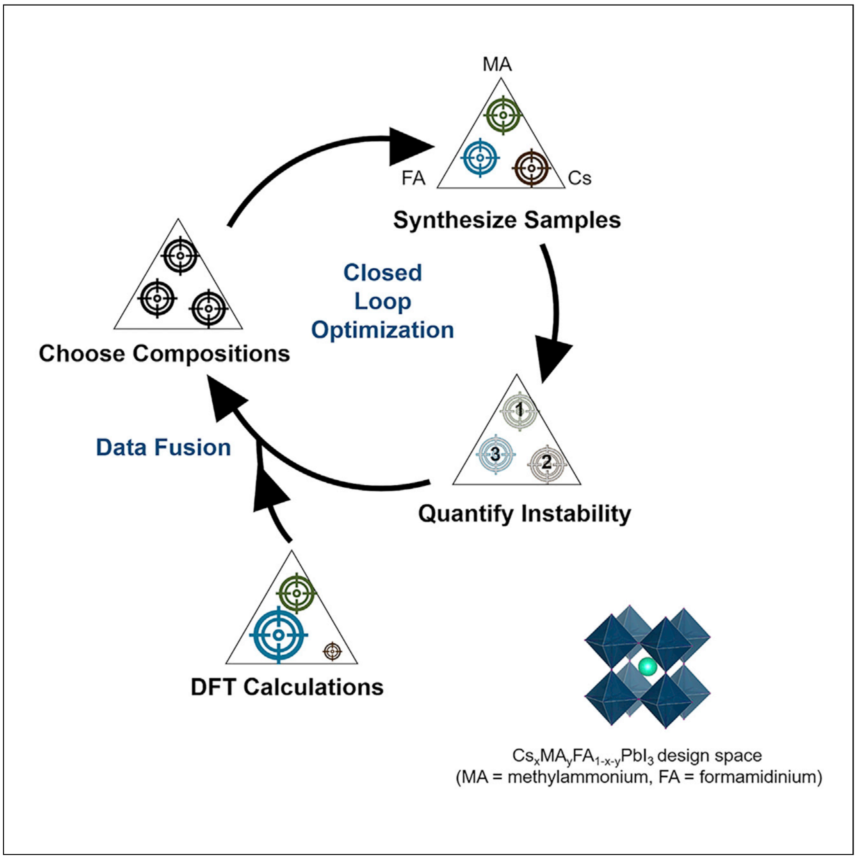
Matter

CellPress

Article

A data fusion approach to optimize compositional stability of halide perovskites



Data fusion combines first-principle calculations and high-throughput experimentation into an end-to-end closed-loop optimization framework, allowing an accelerated search of alloyed halide perovskites in a combinatorial space without human intervention.

Shijing Sun, Armi Tiihonen, Felipe Oviedo, ..., Vladan Stevanovic, John Fisher III, Tonio Buonassisi

shijings@mit.edu (S.S.)
buonassisi@mit.edu (T.B.)

HIGHLIGHTS

Physics-informed machine learning enables accelerated search of stable perovskites

Closed-loop Bayesian optimization takes the human out of the decision-making loop

>17-fold higher stability achieved within a combinatorial space of $Cs_xMA_yFA_{1-x-y}PbI_3$

Improved perovskite thin-film stability translates into enhanced solar cell reliability

4

Demonstrate

Proof-of-concept of performance with intended application/response

Sun et al., Matter 4, 1305–1322
April 7, 2021
https://doi.org/10.1016/j.matt.2021.01.008

Matter

CellPress

Article

A data fusion approach to optimize compositional stability of halide perovskites

Shijing Sun,[1,7,*] Armi Tiihonen,[1,7] Felipe Oviedo,[1] Zhe Liu,[1] Janak Thapa,[1] Yicheng Zhao,[2,6] Noor Titan P. Hartono,[1] Anuj Goyal,[3] Thomas Heumueller,[2,6] Clio Batali,[1] Alex Encinas,[1] Jason J. Yoo,[1] Ruipeng Li,[4] Zekun Ren,[5] I. Marius Peters,[2] Christoph J. Brabec,[2,6] Moungi G. Bawendi,[1] Vladan Stevanovic,[3] John Fisher III,[1] and Tonio Buonassisi[1,5,8,*]

SUMMARY

Search for resource-efficient materials in vast compositional spaces is an outstanding challenge in creating environmentally stable perovskite semiconductors. We demonstrate a physics-constrained sequential learning framework to subsequently identify the most stable alloyed organic-inorganic perovskites. We fuse data from high-throughput degradation tests and first-principle calculations of phase thermodynamics into an end-to-end Bayesian optimization algorithm using probabilistic constraints. By sampling just 1.8% of the discretized $Cs_xMA_yFA_{1-x-y}PbI_3$ (MA, methylammonium; FA, formamidinium) compositional space, perovskites centered at $Cs_{0.17}MA_{0.03}FA_{0.80}PbI_3$ show minimal optical change under increased temperature, moisture, and illumination with >17-fold stability improvement over $MAPbI_3$. The thin films have 3-fold improved stability compared with state-of-the-art multi-halide $Cs_{0.05}(MA_{0.17}FA_{0.83})_{0.95}Pb(I_{0.83}Br_{0.17})_3$, translating into enhanced solar cell stability without compromising conversion efficiency. Synchrotron-based X-ray scattering validates the suppression of chemical decomposition and minority phase formation achieved using fewer elements and a maximum of 8% MA. We anticipate that this data fusion approach can be extended to guide materials discovery for a wide range of multinary systems.

INTRODUCTION

The environmental instability of organic-inorganic halide perovskite materials limits their usage in optoelectronics, such as in solar cells, light emitters, lasers, and photodetectors.[1] Compositional engineering is, to date, one of the most effective methods to improve the stability of perovskites in the presence of heat, humidity, and light without sacrificing optoelectronic performance.[2] This fact has led to intensive research within combinatorial spaces, such as $A_xB_yC_{1-x-y}Pb(I_zBr_{1-z})_3$.[3] However, only a small fraction of this compositional space has been experimentally explored, in part due to the prohibitively expensive brute force synthesis. The paucity of resulting degradation data inhibits generalization of mechanisms across this diverse chemical and structural space, requiring each compositional search to begin with *ab initio* experimental investigations.[4] This challenge is similar to those faced by other materials communities, including the search for heterogeneous catalysts, alloyed battery electrodes, and high-entropy metal alloys for structural and magnetic materials.[5–7] The halide perovskite field and several others require new tools to experimentally navigate these vast spaces efficiently to locate optima and to extract generalizable scientific insights.[8–14]

Progress and potential

Despite recent intensive efforts to improve the environmental stability of halide perovskite materials for energy harvesting and conversion, traditional trial-and-error explorations face bottlenecks in the navigation of vast chemical and compositional spaces. We develop a closed-loop optimization framework that seamlessly marries data from first-principle calculations and high-throughput experimentation into a single machine learning algorithm. This framework enables us to achieve rapid optimization of compositional stability for $Cs_xMA_yFA_{1-x-y}PbI_3$ perovskites while taking the human out of the decision-making loop. We envision that this data fusion approach is generalizable to directly tackle challenges in designing multinary materials, and we hope that our successful showcase on perovskites will encourage researchers in other fields to incorporate knowledge of physics into the search algorithms, applying hybrid machine learning models to guide discovery of materials in high-dimensional spaces.

Machine learning-based sequential learning approaches (e.g., Bayesian optimization [BO]) have emerged as efficient materials search tools that explore vast variable spaces in a "closed-loop" fashion, whereby the outcome of one experimental round informs the next without human intervention. BO, which has attracted increasing attention in the recent developments of self-driving laboratories in various fields of materials science, recently successfully directed experimentation in the search of organic hole-transport materials,[15] piezoelectric oxides,[16] and organic photocatalysts.[10] Within the field of perovskite solar cells, machine learning has been combined with robotic liquid synthesis for microcrystal crystallization.[17–20] However, such a model-free statistical approach shows limitations without principled guidance from domain expertise, because it has to learn everything from scratch. Recent *in situ* experiments and first-principle calculations independently revealed insights into the fundamental composition-dependent instability in organic-inorganic perovskites and their alloys; however, merging computational and experimental insights on selective compositions into a generalizable optimization policy over the entire chemical space remains a challenge.[3] State-of-the-art two-step approaches of directly applying theoretical screening as a hard constraint before shortlisted synthesis are limited by inefficiencies arising from: (1) high-performing theoretical calculations for organic-inorganic systems are often too sparse to guide experimentation, and (2) the discrepancies between the calculation assumptions and the experiments at non-thermodynamic equilibria decreases search accuracy.[21,22] The lack of physics-informed and iterative materials search hinders the ultimate goal of designing perovskite compositions for enhanced environmental stability.

Here, we introduce a data fusion approach to incorporate both Gibbs free energy of mixing (ΔG_{mix}) from density functional theory (DFT) calculation[23] and experimentally quantified degradation from accelerated aging tests to every decision that the BO algorithm is making. We apply this closed-loop machine learning framework to optimize lead iodide perovskites that suffer from severe heat and moisture-induced degradation within the five-element space of $Cs_xMA_yFA_{1-x-y}PbI_3$. Under multiplex environmental stress tests with increased temperature, humidity, and illumination in air, we identified compositions overperforming the $MAPbI_3$ starting point by 17-fold and our state-of-the-art reference composition of ($Cs_{0.05}(MA_{0.17}FA_{0.83})_{0.95}Pb(I_{0.83}Br_{0.17})_3$) by 3-fold within three optimization rounds, and the results are found transferable to device stability. DFT here serves as principled guidance within the decision-making algorithm to constrain the search space to not only chemically, but also the structurally stable α-perovskite alloys.

RESULTS AND DISCUSSION

Closed-loop experimentation platform driven by physics-informed BO

To efficiently guide the compositional search, we constructed a physics-informed batch BO framework (Figure 1). In BO, promising compositions for the next experimental round are suggested by an acquisition function, such as expected improvement, $EI(\Theta)$, which balances the exploitation of the most stable regions and the exploration of high-uncertainty regions within the compositional space. As a key algorithm contribution, we fuse ΔG_{mix} as a probabilistic constraint of the BO acquisition function in the "composition selection" step, providing additional information on phase stability to effectively identify multi-cation perovskites that are thermodynamically stable relative to their single-cation counterparts (Figures 1A and 1B). We define "instability index" (I_c), a figure of merit for optimizing stability. The goal of each optimization round, which consists of three steps of composition selection, "film synthesis," and "instability quantification," is to minimize this value. Our batch BO algorithm makes use of a surrogate ML model, Gaussian process (GP)

[1]Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]Helmholtz-Institute Erlangen-Nürnberg (HI-ERN), Erlangen 91058, Germany
[3]Colorado School of Mines, Golden, CO 80401, USA
[4]Brookhaven National Laboratory, Upton, NY 11970, USA
[5]Singapore-MIT Alliance for Research and Technology, Singapore 138602, Singapore
[6]Institute of Materials for Electronics and Energy Technology (i-MEET), Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen 91058, Germany
[7]These authors contributed equally
[8]Lead contact
*Correspondence: shijings@mit.edu (S.S.), buonassisi@mit.edu (T.B.)

https://doi.org/10.1016/j.matt.2021.01.008

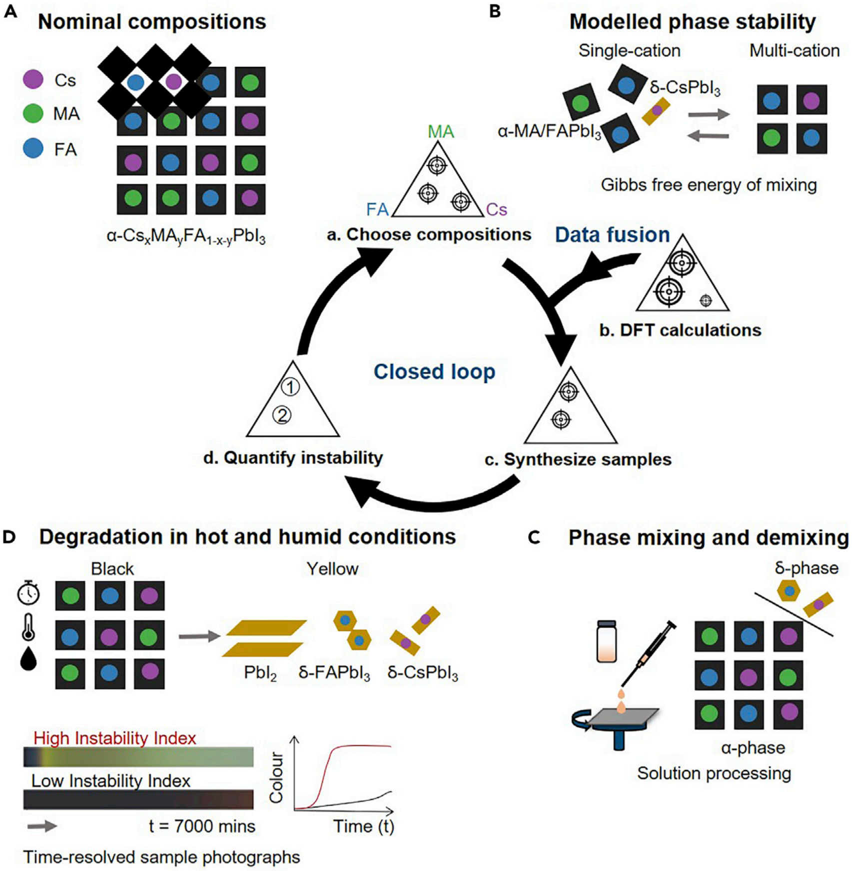


Figure 1. Each optimization round consists of three steps of "composition selection," "film synthesis," and "instability quantification," and a fourth step of theoretical incorporation into the closed-loop workflow using a data fusion approach

(A) Composition selection: the selection of nominal compositions of multi-cation perovskites is driven by machine learning algorithms.

(B) Data fusion: DFT-modeled ΔG_{mix} is incorporated in the optimization algorithm as a constraint.

(C) Film synthesis: tuning A-site cations in lead iodide perovskites forms $Cs_xMA_yFA_{1-x-y}PbI_3$, a space of 5,151 compositions (estimated experimental resolution 1%). Thin-film samples are spin coated in series using precursor solutions of nominal compositions.

(D) Instability quantification: we perform accelerated high-throughput degradation tests with *in situ* optical monitoring, enabling 28 thin-film samples to be degraded in parallel. Near-black photoactive perovskite films turn yellow over time due to the emergence of high-band-gap degradation products. Quantified optical changes over time, recorded in R (red), G (green), and B (blue) channels, are used as a proxy to evaluate the chemical instability of the samples under 85% relative humidity (RH), 85°C sample temperature, and 0.15 Sun visible only illumination. Two representative sets of sample photographs and curves of the total (R + G + B) area-averaged value as a function of time are illustrated in (D).

regression,[24] to estimate the value and uncertainty of I_c in non-explored regions of the compositional space (see the Experimental procedures).

Within each optimization round (one batch in BO), 28 spin-coated thin-film samples (Figure 1C) are examined *in situ* in parallel using an environmental chamber under 85% relative humidity (RH) and 85°C in the air (Figure S1). 0.15 Sun visible only illumination is applied to enable automatic image capture every 5 min using an RGB camera (~200 μm resolution). Photoactive α-perovskite phases within $Cs_xMA_{y-}FA_{1-x-y}PbI_3$ exhibit a band gap of ~1.5 eV, whereas their main degradation products under hot and humid conditions, PbI_2 (2.27 eV)[25] δ-$CsPbI_3$ (2.82 eV),[26] or δ-$FAPbI_3$ (2.43 eV)[27] show deteriorated photophysical properties (Figure S2). As shown in Figure 1D, we hence used a color-based metric as a proxy to capture the macroscopic evolution of the high-band-gap, non-perovskite phases (see Videos S1 and S2).[28–30] We define the instability index (I_c) as the integrated color change of an unencapsulated perovskite film over accelerated degradation test duration T. Complementary direct band-gap measurements before and after the degradation tests using UV-vis spectroscopy are listed in Figure S12.

$$I_c\left(\Theta\right) = \sum_{c=\{R,G,B\}} \int_{0min}^{T} \left|c(t,\ \Theta) - c(0,\Theta)\right| dt, \qquad \text{(Equation 1)}$$

where composition $\Theta = (x, y, 1 - x - y)$, t is time, and c are area-averaged, color-calibrated red, green, and blue pixel values of the sample. The cutoff time was set to T = 7,000 min based on the observed divergence between the most- and least-stable compositions (Figure S3). Our closed-loop and iterative workflow enable the systematic optimization of multi-cation perovskites against degradation by varying the nominal compositions, Θ, within $Cs_xMA_yFA_{1-x-y}PbI_3$ (x, y limit to two decimal places) (Tables S1 and S2).

Data fusion approach: incorporation of phase thermodynamics into automated composition selection

Due to their polymorphic nature, identical perovskite compositions crystallized into different phases can exhibit diverse degradation behaviors, making it essential to evaluate phase stabilities in any perovskite composition optimization.[1] The end members of the compositional space in this study consist of the cubic α-FA/$MAPbI_3$ perovskites and the non-perovskite δ-$CsPbI_3$ at the synthesis temperature.[31] Phase de-mixing during synthesis leads to minority phases within thin-film samples before degradation tests and are, therefore, not captured in I_c. Nevertheless, phase de-mixing during film formation or soon after is not desirable because it causes deterioration of the electronic properties of the perovskite.[32] Schelhas et al.[23] recently demonstrated the use of DFT calculations to predict the phase de-mixing tendency between α-$Cs_xMA_yA_{1-x-y}PbI_3$ (G_{mix}) and their single-cation perovskite polymorphs $APbI_3$ (A = Cs, MA, or FA) (G_0) at a given temperature. Here, we fuse the composition-dependent change in Gibbs free energy of mixing, ΔG_{mix} as a constraint into the experimental optimization loop (Figure 2A). This approach allows the α- and δ-phase relative stability in the non-degraded perovskite samples to be considered in the composition selection, thus enabling us to reduce sampling in regions with high probability of minority phase formation.

Data fusion refers to a set of techniques where ML is used to map two or more datasets coming from related but distinct distributions. In our case, we relate the theoretical $\Delta G_{mix}(\Theta)$ and the experimental $I_c\ (\Theta)$. The two data streams account for distinct mechanisms of modeled thermodynamic phase instability and measured macroscopic thermal-moisture instability, respectively. Hence, it is inadequate to combine both datasets as equivalent or include DFT directly as a prior following state-of-the-art model-free BO.[33,34] Here, we define a data-fused probabilistic constraint approach according to Equation 2:

$$P(\Delta G_{mix}(\Theta), \beta_{DFT}) = \frac{1}{1 + e^{-\Delta G_{mix}(\Theta)/\beta_{DFT}}}, \qquad \text{(Equation 2)}$$

where $P(\Delta G_{mix}(\Theta), \beta_{DFT})$ is a logistic cumulative distribution function modeling the phase mixing probability and β_{DFT} is a data fusion parameter calibrated according to ΔG_{mix} calculations to control the smoothness of the boundaries from stable to

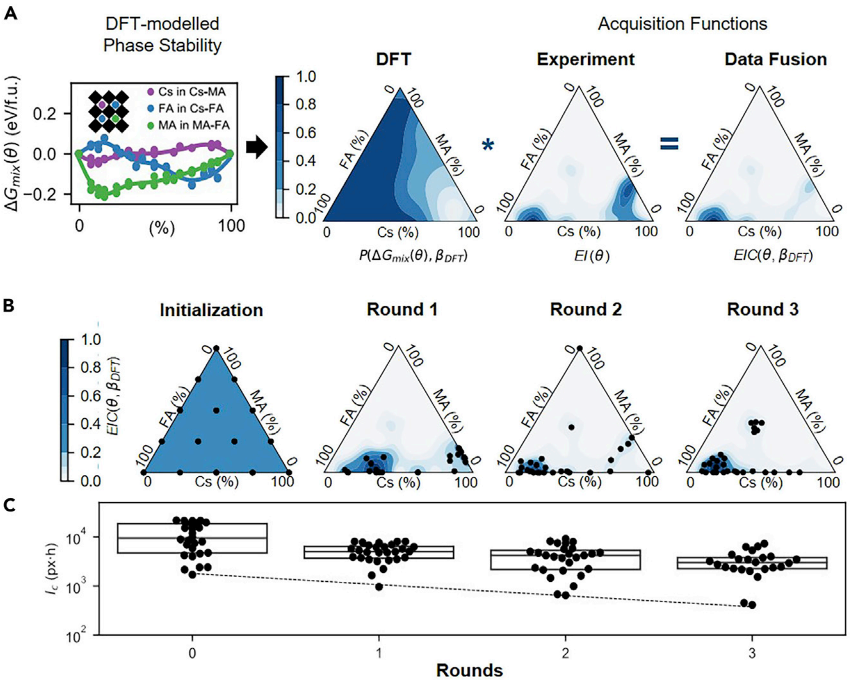


Figure 2. Fusion of DFT and experiments guides the optimization of compositional stability

(A) DFT-modeled Gibbs free energy of mixing, ΔG_{mix}, of 47 binary compositions of CsMA, CsFA, and MAFA α-perovskites relative to end members of δ-$CsPbI_3$, α-$MAPbI_3$, and α-$FAPbI_3$. We fit a Gibbs free energy model, $\Delta G_{mix}(\Theta)$, where Θ is a composition in the ternary space, to the phase thermodynamics data using Gaussian process regression. $\Delta G_{mix}(\Theta)$ is transformed into a probabilistic constraint, $P(\Delta G_{mix}(\Theta), \beta_{DFT})$ that models the cumulative probability of phase mixing at above 300 K. Multiplying P with the acquisition function of the Bayesian optimization algorithm, $EI(\Theta)$, gives a DFT-weighted acquisition function, $EIC(\Theta)$.

(B) Starting from 15 equally spaced compositions in the initialization round, for each optimization round, 28 sample films (black markers) are synthesized and undergo degradation tests. Compositions are chosen by a Bayesian optimization algorithm that suggests them using $EIC(\Theta)$ (blue surface color).

(C) Experimentally measured instability indices (I_c) (pixels × hours), of 112 samples over one initialization and three optimization rounds. The black boxes indicate the mean and standard deviation of each round. The dashed line indicating the most stable compositions in each experimental round is for eye guidance only.

unstable compositions, forming a soft compositional boundary as shown in Figure 2A (see the Experimental procedures for algorithm details).

Given the computational cost and complexity of DFT calculations on organic-inorganic hybrid systems, we first regress 85 DFT-modeled ΔG_{mix} values on 47 single-cation and binary alloyed compositions (29 MAFA and CsFA compositions from Schelhas et al.[23] and 12 CsMA compositions computed for the present work using the same methods) over the quasi-ternary $Cs_xMA_yA_{1-x-y}PbI_3$ phase space using an auxiliary GP model that defines $\Delta G_{mix}(\Theta)$. Figure 2A visualizes the probability of phase mixing $P(\Delta G_{mix}(\Theta), \beta_{DFT}) \in [0, 1]$, where low values suggest phase instability ($\Delta G_{mix} >> 0$) and high values suggest phase stability ($\Delta G_{mix} << 0$).

Our work is inspired by the unknown constraint BO proposed by Gelbart et al.[35] By developing a probabilistic constraint model $P(\Delta G_{mix}(\Theta), \beta_{DFT})$ instead of applying a hard constraint boundary, we are able to discount regions predicted by DFT to go through phase de-mixing rather than completely exclude any unfavorable regions.

This approach accounts for the inherent uncertainty in DFT predictions, chemical accuracy, and data scarcity through the use of the soft compositional boundary to model the stability threshold (see the Experimental procedures for β_{DFT} calibration). The proposed algorithm allows us to seamlessly adapt DFT into the experimental optimizations loop, thereby achieving a physics-informed and sample-efficient search without being limited by the unknown exact phase boundaries across a vast compositional space (Figures S4 and S5).

To integrate the probabilistic constraint into the BO formulation, we weigh the acquisition function with the value of $P(\Delta G_{mix}(\Theta), \beta_{DFT})$ and obtain a DFT-weighted BO acquisition function, $EIC(\Theta)$, as illustrated in Figure 2A. Traditional $EI(\Theta)$ utilizes the I_c results of our first experimental round without DFT and indicates two potential optima in Cs-poor and Cs-rich regions, respectively. The DFT-weighted $EIC(\Theta)$ effectively reduces sampling in energetically unfavorable Cs-rich regions despite low I_c: the subsequent optimization rounds converge to stable nominal compositions with a high probability of stable α-perovskite films among Cs-poor regions (Figures 2A, S6, and S7). Comparisons of optimization with and without DFT weighting using a teacher-student model are shown in Figures S8 and S9, which validates that, without data fusion, the model-free BO algorithm continues to suggest sampling in Cs-rich regions despite their phase instability.

Figure 2B demonstrates that batch BO sequentially identifies the most stable regions over one initialization and three optimization rounds of synthesis and degradation tests. Iterative evolution of the landscape (posterior mean of I_c, $I_c(\Theta)$, with uncertainty) is presented in Figures S3 and S4. Figure 2C reveals a rapid decrease in experimentally quantified I_c from Rounds 0–3. The search converges after three optimization rounds (see Figure S5 for convergence conditions) to an optimal composition region centered at $Cs_{0.17}MA_{0.03}FA_{0.80}PbI_3$ and bounded by 8%–29% Cs, <14% MA, and 68%–92% FA. The identification of the global optimum lying within an FA-rich, and Cs- and MA-poor region is consistent with the reports that FA-rich perovskites show superior environmental stability compared with their MA-rich counterparts and the less volatile Cs is expected to enhance the heat and moisture resistance.[36] Interestingly, we found a local optimum near $Cs_{0.26}MA_{0.36}FA_{0.38}PbI_3$, which emerged in Round 1. We sampled four additional compositions in Round 3 and validated that the non-intuitive local optima suggested by the algorithm is reproducible. The ability to rapidly identify non-intuitive regions of success is a major advantage of using an automated closed-loop optimization algorithm over materials search strategies leveraging human intuition alone. Further experimental validation and mechanisms study of the identified compositional regions of interest are discussed in the next subsection. We define the compositional space as the discretized quasi-ternary-phase space subdivided by the minimum achievable experimental resolution (1% composition). This yields 5,151 possible singular, binary, and ternary cation compositions, 1.8% of which were sampled experimentally while converging to the optimal region (i.e., 94 unique compositions and 112 samples within Round 0–3, see the Supplemental information for more details). Three additional degradation rounds of seven representative compositions were performed to validate the instability trend, with structural and optical characterization shown in Table S3 and Figures S10–S13.

Composition-, phase-, and time-dependent instability landscapes

We find the overall stability landscape within the $Cs_xMA_yFA_{1-x-y}PbI_3$ compositional space to be non-linear. To quantify the divergence in degradation profiles, Figure 3A presents the composition-dependent instability landscape, $I_c(\Theta)$ for

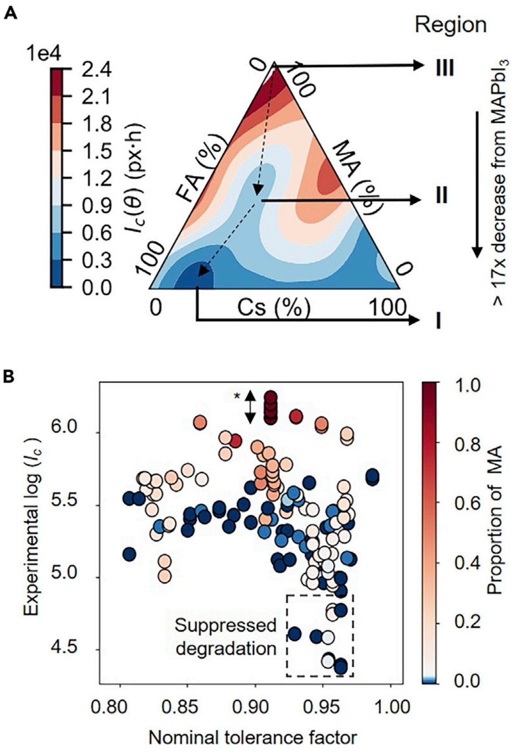


Figure 3. Composition-dependent instability landscape

(A) The instability landscape $I_c(\Theta)$ mapped over the $Cs_xMA_yFA_{1-x-y}PbI_3$ compositional space evolves in time. The posterior mean of the Bayesian optimization surrogate model estimating I_c is shown after 6 h at the end of the aging test (using the experimentally measured samples from Rounds 0–3). Three distinct compositional regions are identified. Regions I–III are labeled following ascending order of measured I_c.

(B) Roles of cations in the realized degradation routes. (A) I_c of all experimentally measured samples as a function of Goldschmidt's tolerance factor, co-visualizing the proportion of MA. *Experimental uncertainty of measured I_c across batches in the control composition, $MAPbI_3$ (see Data S1 for a full list of samples).

$Cs_xMA_yFA_{1-x-y}PbI_3$ after three experimental optimization rounds, where three distinct compositional regions (Regions III → I) with descending I_c are clearly identified. The evolution of $I_c(\Theta)$ as a function of degradation time, as shown in Figure S3, further reveals that fast degradations of MA-rich compositions are evident after 6 h of degradation tests (Region III), while two additional regions, representing compositions in local optima (Region II) and the global optima (Region I), are sequentially distinctive after 100 h of degradation tests. Experimentally measured I_c data reveal a >17-fold reduction from the $MAPbI_3$ endpoint in Region III to the ML-optimum, at $Cs_{0.17}MA_{0.03}FA_{0.80}PbI_3$, in Region I. Interestingly, several MA-containing compositions show comparable I_c with their CsFA binary cation counterparts as shown in Figure 3B. Up to 8% MA, the least chemically stable cation in the design space, can be added into the perovskite structure before environmental stability is significantly compromised.

Figure 3B visualizes experimentally measured I_c as a function of Goldschmidt's tolerance factor (TF) calculated using the average ionic radius of A-site cations in nominal compositions. TF is empirical guidance that has been widely applied to estimate the intrinsic structural stability of hybrid perovskites.[37,38] We find that TF optimization is necessary but not sufficient criteria for achieving high environmental stability. During optimization Rounds 1–3, an increasing number of compositions within a TF of 0.93–0.97 are suggested by ML, indicating high stability of compositions with a TF of around 0.95. This value is lower than TF = 1 of an ideal cubic structure, attributing to the incorporation of small-radius and non-volatile Cs into the α lattice to improve moisture and heat resistance (Figure S11).

To validate the scientific relevance of the data fusion approach, which effectively discounts experimental sampling in the regions with high probabilities of minority phase formation, we sought to determine the impact of thermodynamics-driven

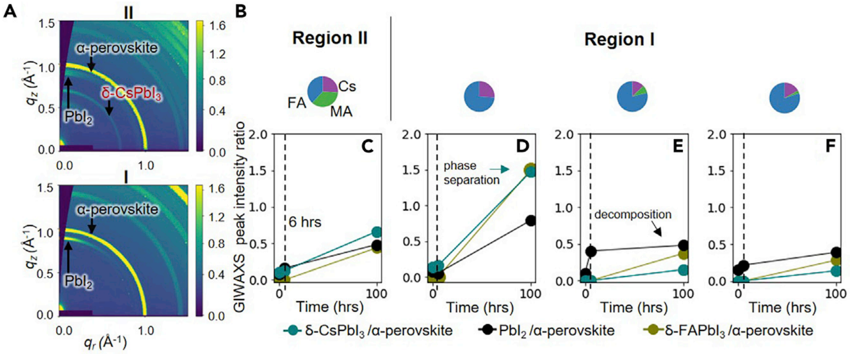


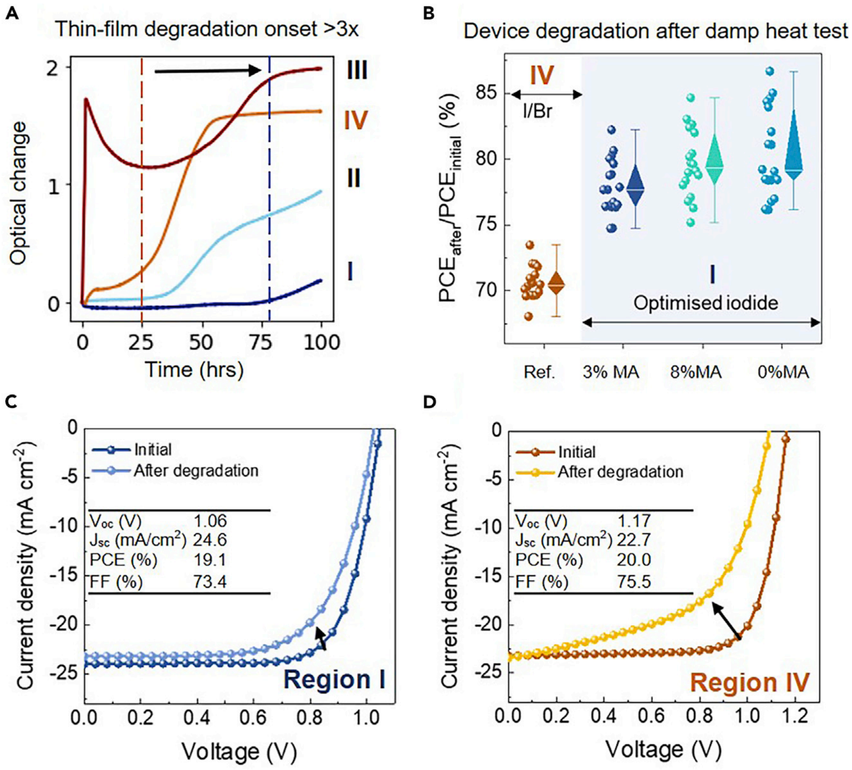
Figure 4. Degradation mechanisms in optimized perovskites

(A) GIWAXS images of the as-synthesized thin films of $Cs_{0.26}MA_{0.36}FA_{0.38}PbI_3$ in Region II and the ML-optimum composition, $Cs_{0.17}MA_{0.03}FA_{0.80}PbI_3$, in Region I. Over-stoichiometric precursors with excess PbI_2 were added in all samples following the high-efficiency perovskite solar cell recipe in Saliba et al.[36]

(B) GIWAXS peak intensity ratios of the non-perovskite phases and PbI_2 relative to the perovskite phase as an estimate for the extent of degradation for representative compositions in Regions I and II, including (C) $Cs_{0.26}MA_{0.36}FA_{0.38}PbI_3$ in Region II, (D) $Cs_{0.26}FA_{0.74}PbI_3$ near the boundary of Regions I and II with high Cs, (E) $Cs_{0.13}MA_{0.08}FA_{0.79}PbI_3$ near the boundary of Regions I and II with high MA, and (F) $Cs_{0.17}MA_{0.03}FA_{0.8}PbI_3$ at the center of Region I.

minority phases on degradation dynamics. Within the $Cs_xMA_yFA_{1-x-y}PbI_3$, we examined the structural evolution after 0-, 6-, and 100-h degradation tests, respectively, using *ex situ* synchrotron-based grazing incidence wide-angle scattering (GIWAXS) measurements (Figures S14 and S15). Comparing the ML-local optimum, (ii) $Cs_{0.26}MA_{0.36}FA_{0.38}PbI_3$ in Region II, and the ML-optimum, (i) $Cs_{0.17}MA_{0.03}FA_{0.80}PbI_3$ in Region I (Figure 4A), we observe that the two compositions (i) and (ii) exhibit comparable lattice parameters of α-perovskites; however, local optimum (ii) contains an additional δ-$CsPbI_3$ minority phase before degradation tests. Interestingly, only a slightly larger increase in PbI_2 is observed in (ii) than in global optimum (i) after a 6-h degradation run. After 100 h, (ii) exhibits a significantly greater loss of intensity of α-perovskites and crystallinity (Figure S14).

To understand the distinctive roles of MA, Cs, and minority phases that govern the divergence in degradation profiles behind the optimization results, we further quantified the peak intensities of PbI_2 (001), δ-$CsPbI_3$ (002), $FAPbI_3$ (001), and α-perovskite (001) of four representative compositions in Regions I and II (Figures 4C–4F). Two composition-dependent degradation mechanisms were observed (Figures S16 and S17; Tables S5 and S6). While all four samples show increased PbI_2 content (evidence of chemical decomposition to precursors), the emergence of δ-$CsPbI_3$ and δ-$FAPbI_3$ minority phases in Region I films with low Cs and MA content (Figures 4E and 4F) indicates additional mechanisms of phase separation during degradation tests. We find that MA plays a competing role as it accelerates chemical decomposition while suppressing phase separation (additional X-ray diffraction results are shown in Figure S10). Reducing MA content from 8% to 3% (Figures 4E and 4F) shows effective suppression of chemical decomposition within the first 6 h. If we remove MA completely, we observe a faster phase separation emerging between 6 and 100 h of degradation (Figures 4C and 4D). Phase separation as a degradation mechanism is observed to take place in a longer timescale compared with chemical decomposition. Overall, the stability optimization pathway from Region II to Region I to achieve kinetically suppressed degradation can be summarized as follows: (1) reducing MA to suppress chemical decomposition, (2) reducing Cs to limit



Unencapsulated under 85% RH/85 °C in air for 50 hours under dark

Figure 5. Suppressed degradation in thin-film and photovoltaic devices

(A) Optical changes as a function of degradation time, showing the onset of degradation for representative compositions in Regions I–IV, $Cs_{0.17}MA_{0.03}FA_{0.80}PbI_3$, $Cs_{0.26}MA_{0.36}FA_{0.38}PbI_3$, $Cs_{0.05}(MA_{0.17}FA_{0.83})_{0.95}Pb(I_{0.83}Br_{0.17})_3$ (I/Br), and $MAPbI_3$, respectively.

(B) The percentage ratios of the solar cell efficiency after 50 h of full damp heat degradation tests for unencapsulated devices over the initial efficiencies based on I/Br and three ML-informed compositions in Region I, $Cs_{0.17}MA_{0.03}FA_{0.80}PbI_3$, $Cs_{0.13}MA_{0.08}FA_{0.79}PbI_3$, and $Cs_{0.13}FA_{0.87}PbI_3$, respectively.

(C) Initial device efficiency based on the ML-optimum composition in Region I, and the their current-voltage curves before and after 50 hours accelerated degradation tests under 85% RH/85°C unencapsulated in air in dark.

(D) Initial device efficiency based on the I/Br reference composition in Region IV, and their current-voltage curves before and after the same degradation tests as in (C).

thermodynamics-driven minority phase formation, and (3) balancing MA, FA, and Cs for restraining additional minority phase formation in hot and humid conditions.

Insights into the effects of compositional complexity on thin-film and device stability

To determine the impact of ML-informed compositions in the field of perovskite optoelectronics, we first demonstrated the improvement of thin-film stability in this study against the state-of-the-art. Figure 5A illustrates the quantitative optical change analysis for three representative thin-film compositions from Region I, including the ML-optimum $Cs_{0.17}MA_{0.03}FA_{0.80}PbI_3$ (i), the ML-local optimum $Cs_{0.26}MA_{0.36}FA_{0.38}PbI_3$ (ii), and $MAPbI_3$ (iii). We further compared the optimized five-element iodide perovskites (i) with a six-element iodide-bromide reference composition, $Cs_{0.05}(MA_{0.17}FA_{0.83})_{0.95}Pb(I_{0.83}Br_{0.17})_3$ (iv) (I/Br). I/Br is outside the design space of this study (referred to as Region IV), but is among the most widely used compositions in high-efficiency perovskite solar devices.[36,39,40] We found that (i) yields a 3.5-fold lower I_c than (iv). In addition to suppressed total degradation in optimized iodide perovskites, the degradation onset is also postponed. This is reflected by the >3-fold delay in the onset of sharp optical change as in Figure 5A. The ML-optimum within the iodide perovskite space overperforms the more complex I/Br mix, which contradicts a long-standing assumption in the perovskite field that increasing compositional complexity entropically stabilizes the absorber. The overall environmental stability based on thin-film stability is Region I > Region II > Region IV > Region III.

We then demonstrated the enhanced stability of full photovoltaic devices without compromising conversion efficiency. To ensure rigorous environmental tests that match the reliability requirement for perovskite solar cell commercialization, we used ambient air, 85% RH, and 85°C temperature full damp heat testing on unencapsulated devices (close to the conditions of accelerated film degradation tests), which is a much harsher stress condition than most current laboratory testing for perovskite devices. Figure 5B reveals that devices of the three ML-informed compositions from Region I all overperform the state-of-the-art I/Br reference, leading to an increased efficiency (after aging/before aging) ratio from 73% to 87% after a 50-h accelerated degradation test. While devices of both the ML-optimum and the reference composition show initial efficiencies of >19% (as shown in Figures 5C and 5D), more severe photocurrent decay is observed in the reference. An average of the 77% of the performance is maintained in the ML-optimum in comparison with 71% in I/Br (Figure 5B, over 18 devices) after degradation tests. In addition to the ML-optimum (3% MA), two other Region I iodide perovskites, $Cs_{0.13}MA_{0.08}FA_{0.79}PbI_3$ (8% MA) and $Cs_{0.13}FA_{0.87}PbI_3$ (0% MA) also overperform I/Br in both film and device stability (Figure S18). We find that compositions with fewer elements lead to improved thermo-moisture stability, where the MA-free, iodide-only composition in this comparative study achieved the highest device stability.

We note that, in a solar cell, other layers than the absorber within a device may also accelerate the degradation depending on the device architectures. Unencapsulated full devices are often found to undergo faster degradations than bare films, in particular in the first several hours, attributing to interface-related degradations.[4,23] In this study we only focus on the correlations between film and device stability from an angle of perovskite layer degradation without device architecture optimization. To confirm if the device stability improvement holds under additional stress of 1 Sun illumination, we further compared the photostability of devices under 1 Sun at 65°C in N_2 (different from the conditions applied for thin-film optimization in this study) and the results show a photo-thermostability of 8% MA > ML-optimum (3% MA) > I/Br > 0% MA (Figure S19). The higher device photostability of MA-containing compositions suggest a beneficial role of MA in suppressing illumination-induced degradation in multi-cation iodide perovskites.

Our device stability results highlight the non-intuitive conclusion that simplifying the perovskite formulation yields a higher device stability in hot and humid environments. This finding emphasizes the importance of achieving a holistic understanding of a compositional space to effectively identify optima. The physics-constrained sequential learning approach developed in this study can be extended to experimentally navigate higher-dimensional spaces under operational conditions, such as to identify the most environmentally stable I/Br-chloride perovskite alloys in the growing chemical space of $A_xB_yC_zD_{1-x-y-z}Pb(I_pBr_qCl_{1-p-q})_3$, to further improve the efficiency and reliability of perovskite solar cells.

Conclusions

In this study we develop a closed-loop optimization strategy for $Cs_xMA_yFA_{1-x-y}PbI_3$ multi-cation perovskites against heat-, moisture-, and light-induced degradation by introducing a physics-constrained BO framework. We identify an FA-rich and Cs-poor region centered at $Cs_{0.17}MA_{0.03}FA_{0.8}PbI_3$ with >17-fold stability optimization from $MAPbI_3$ while sampling only 1.8% of the discretized compositional space, achieving superior search efficiency and scientific relevance to brute force screening and state-of-the-art model-free BO, respectively. The study demonstrates the power of data fusion to allow material search over vast and sparsely sampled compositional spaces, where the DFT-modeled phase mixing serves as a probabilistic constraint and provides principled guidance to ML-directed experimentation.

We apply this physics-informed optimization framework to achieve a holistic understanding of the fundamental composition-, phase-, and time-dependent behavior of organic-inorganic perovskites. As a consequence of competing roles of cations in different degradation mechanisms, a composition window of up to 8% addition of the least chemically stable cation, MA, contributes to kinetically suppressed degradation, whereas the most chemically stable cation in this design space, Cs, contributes to accelerated degradation through phase separation even in the most macroscopically stable candidates found in the $Cs_xMA_yFA_{1-x-y}PbI_3$ compositional space, which potentially limits the benefits of Cs as a perovskite stabilizing agent. These findings highlight the detrimental effects of minority phase formation as a degradation pathway, which occurs over a longer time frame than chemical decomposition, and hence are easily underestimated in the initial compositional engineering. We further suggest several optimized iodide perovskites, including $Cs_{0.17}MA_{0.03}FA_{0.8}PbI_3$ and $Cs_{0.13}MA_{0.08}FA_{0.79}PbI_3$, that show superior photo-, thermo-, and moisture device-stability to the state-of-the-art I/Br mixed perovskite ($Cs_{0.05}(MA_{0.17}FA_{0.83})_{0.95}Pb(I_{0.83}Br_{0.17})_3$), providing insights into simplifying perovskite compositions for solar cell reliability.

This data fusion approach combines multiple data sources into a single search algorithm and can be utilized to include other experimental or theoretical constraints with non-negligible uncertainty into the materials design strategy. The method can be generalized to direct experimentation in other material systems, merging complementary experiments and theory to collectively inform synthesis in a closed-loop fashion across a vast chemical and structural space.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead contact, Tonio Buonassisi (buonassisi@mit.edu).

Materials availability

This study did not generate new reagents. Reagents used were purchased from Sigma-Aldrich as described in subsection Materials.

Data and code availability

Details of materials, experimental methods, machine learning framework, and camera-based *in situ* degradation monitoring platform details are available in the Supplemental information. The codes and the datasets used for Bayesian optimization are available in GitHub repository https://github.com/PV-Lab/SPProC. Lists of samples and raw data of XRD characterization, and thin-film degradation results, are shown in Data S1.

Materials

Perovskite precursor solutions were spin coated on UV-ozone glass substrates. Glass microscope slides (VWR) cut into square pieces were cleaned with sonication in 2% Hellmanex-DI water mix, DI water, and IPA, respectively. We performed perovskite synthesis with over-stoichiometric PbI_2 in a molar ratio of 1.09 (PbI_2) to 1 (halide salt of CsI, MAI, and FAI). Lead(II) iodide stock solution was prepared in 9:1 *N*,*N*-dimethylformamide (Sigma-Aldrich) to dimethyl sulfoxide (Sigma-Aldrich) solvent. The perovskite precursor solution was prepared by mixing individual stock solutions following the ratios of Cs, FA, and MA suggested by the machine learning algorithm. The films were annealed at 403 K for 20 min using the central part of a hot plate in the glovebox. The spin-coating program follows a two-step approach: 1,000 rpm for 10 s and acceleration of 200 rpm/s, with a subsequent 6,000 rpm for 30 s and acceleration of 2,000 rpm/s. Chlorobenzene (Sigma-Aldrich) antisolvent (150 μL) was dropped 5 s at the beginning of the second step of spin coating. The annealed samples were cooled to room temperature before being transferred to the degradation test. $Cs_{0.05}(FA_{0.83}MA_{0.17})_{0.95}Pb(I_{0.87}Br_{0.13})_3$ was synthesized following Correa-Baena et al.[41]

Accelerated degradation tests

Humidity, temperature, and visible light illumination level were controlled at 85°C ± 2°C, 85% ± 5%, and 0.15 ± 0.01 Sun, respectively, using an in-house built environmental chamber. Samples were photographed automatically every 5 min during the aging test and, additionally, the humidity and temperature of the aging chamber were tracked automatically. That the illumination conditions remained stable during the aging tests was confirmed by following a reference color chart that had been placed into the picture area and the collected data were color-calibrated (see the Supplemental experimental procedures). Accelerated degradation, automation, parallel degradation of 28 samples, and a fault-resistant measurement method using photography enabled high-throughput aging testing compared with traditional methods. After the aging test, the samples were stored in a glovebox until further characterization.

Physics-informed BO framework

In the BO setting, a surrogate machine learning model, GP regression, is used to approximate the mean and uncertainty of $I_c(\Theta)$ in non-sampled regions of the compositional space. Once the model is fitted, an acquisition function is used for suggesting locations in the compositional space with a high chance of leading to an optimum. Our chosen base acquisition function that we modify by data fusion principle is expected improvement $\mathrm{EI}(\Theta)$,

$$EI(\Theta) = (\mu_n(\Theta) - \tau)\Phi\left(\frac{\mu_n(\Theta) - \tau}{\sigma_n(\Theta)}\right) + \sigma_n(\Theta)\varphi\left(\frac{\mu_n(\Theta) - \tau}{\sigma_n(\Theta)}\right),$$

where Φ is the standard normal cumulative distribution, $\mu_n(\Theta)$ is the mean of the surrogate model's (here GP regression) posterior, τ is an incumbent best point, σ_n is the variance of the GP model's posterior, φ is the standard normal probability distribution, and n is degradation round. By maximizing $EI(\Theta)$, the most promising location Θ^* (either due to a low expected instability index or a high uncertainty of the estimate) for the next experimental round is determined. To suggest more than one promising composition Θ per round, we use the local penalization algorithm to resample $EI(\Theta)$, as described by González et al.[42] To include the physical constraints in the BO algorithm, the acquisition function is weighted according to the probabilistic model $P(\Delta G_{\mathrm{mix}}(\Theta), \beta_{\mathrm{DFT}})$ following the method of Gelbart et al.[35]:

$$EIC(\Theta) = EI(\Theta)\ P(\Delta G_{\mathrm{mix}}(\Theta), \beta_{\mathrm{DFT}}).$$

Once ΔG_{mix} is modeled (see the Supplemental experimental procedures), the probabilistic model $P(\Delta G_{\mathrm{mix}}(\Theta), \beta_{\mathrm{DFT}})$ can be computed. The inherent bias and precision of DFT calculations justifies the probabilistic treatment of the constraint. Referring to the constraint value as $\Phi = P(\Delta G_{\mathrm{mix}}(\Theta),\ \beta_{\mathrm{DFT}})$, we formulate the data fusion process as maximizing the likelihood $\mathcal{L}(\beta_{\mathrm{DFT}}; \Phi,\ \Delta G_{\mathrm{mix}})$ such that:

$$\beta^{*}_{\mathrm{DFT}} = \mathrm{argmax}_{\beta_{\mathrm{DFT}}} \mathcal{L}(\beta_{\mathrm{DFT}}; \Phi,\ \Delta G_{\mathrm{mix}}).$$

Using a Bernoulli likelihood, this definition is equivalent to fitting a logistic regression model via maximum likelihood, with ΔG_{mix} as the independent variable and the probability of phase de-mixing as the dependent variable. In this sense, one could estimate a certain critical energy above which the crystalline structure is unstable and will decompose into its constituent phases. In the context of convex hull stability calculations, this value is often considered to be around -0.025 eV/f.u. We hence choose β_{DFT} so that it produces a cumulative probability of $P(-0.025\mathrm{eV}/f.u, \beta^{*}_{\mathrm{DFT}}) = 0.7$ and $P(-0.05eV/f.u, \beta^{*}_{\mathrm{DFT}}) = 0.9$. This assumption defines a smooth gradual boundary for phase mixing in the compositional space, considering the inherent uncertainty of first-principle calculations. Our choice of probabilistic model is common in machine learning literature, due to the simplicity and expressivity of logistic models.[43]

DFT calculations

The methodology used here for alloyed halide perovskites has been discussed in detail in our recently published work (Schelhas et al.[23] and Goyal et al.[44]). In this paper we performed additional DFT calculations for the binary (Cs-MA) and ternary (Cs-MA-FA) alloys, as well as using our previously published DFT calculations on binaries (Cs-FA, FA-MA) (Schelhas et al.[23]), to generate adequate initial DFT data in the three-dimensional-phase space that feeds into the optimization model. DFT calculations were performed within the projected augmented wave method[45] as implemented in the VASP code.[46] The Perdew Burke Ernzerhof exchange correlation function[47] was used with GGA, and spin-orbit coupling was included in the total energy calculations. A plane wave cutoff of 340 eV and a Monkhorst-Pack k-point sampling scheme[48] were used. Alloy structures were created using the pseudo-cubic as the starting structure for the pure compositions (obtained from Stoumpos et al.[49]) with random substitution at the A-site. A special quasi-random structure method,[50] as implemented in the ATAT package,[51],[52] was used to obtain structures for various A-site alloy compositions. Two different supercell sizes (96 and 144 atoms) were used and for each A-site composition with multiple (two or three) structures, variation in the orientation of the MA and FA molecules was considered. For both pure and alloy phase calculations all degrees of freedom (cell shape, volume, and ionic positions) were relaxed in DFT. Following the relaxations, A-site alloy structures at various compositions were found to retain the overall cubic symmetry. As benchmarked in Lejaeghere et al.,[53] the reproducibility and precision in our DFT total energy calculations was very high.

Thermodynamic modeling: to model the thermodynamic phase stability of mixed A-site halide perovskites, we compute the Gibbs free energy of mixing ($\Delta G_{\mathrm{mix}} = \Delta H_{\mathrm{mix}} - T\Delta S_{\mathrm{mix}}$) of these materials as a function of the A-site composition. The modeled ΔG_{mix} has two components: (1) the enthalpy of mixing (ΔH_{mix}) and (2) the entropy of mixing (ΔS_{mix}). The enthalpy of mixing is calculated from DFT by taking the difference between the total energy of the mixed A-site halide perovskite with respect to the total energy of the constituent, or pure, phase. The temperature dependence ($T\Delta S_{mix}$) to Gibbs free energy is incorporated by considering the entropic contributions associated with the configuration and rotations degrees of freedom. Further details of thermodynamic modeling can be found in Schelhas et al.[23] The variability in the computed value of Gibbs free energy at a specific composition is between 5 and 20 meV/unit, and this is a result of the varying orientation of the FA and MA molecules between the multiple structures considered in our simulations.

X-ray diffraction

Grazing incidence X-ray diffraction (incident angle of 1°) was performed using Rigaku SmartLab with Cu-$K\alpha$ sources on the as-synthesized thin films to investigate the crystal structures and to examine minority phases.

UV-visible spectroscopy

The absorbance of the films was calculated based on transmission and reflection measurements taken using a PerkinElmer Lambda 950 UV/Vis Spectrophotometer. Band gaps were calculated using Tauc methods assuming direct band gaps.

Scanning electron microscopy

The film morphology was investigated using a Zeiss Ultra-55 field-emission scanning electron microscope (FESEM, Zeiss), with an in-lens detector and a 3.00 kV EHT gun. The grain sizes were counted using ImageJ within an area of ~0.72 μm^2. Various compositions show different distributions of grain sizes. The grain sizes of the composition with high Cs ($Cs_{0.26}FA_{0.74}PbI_3$), with Cs > 20%, are mostly between 300 and 500 nm and some of them reach 1,200 nm, indicating the presence of the δ-phase. The grain sizes of $MAPbI_3$ are mostly between 200 and 400 nm, and reach 1,000 nm. The grain sizes of the rest of the compositions, which have low Cs (<20%), are mostly between 200 and 400 nm.

GIWAXS measurements

GIWAXS measurements were taken at beamline 11-BM (CMS) at the National Synchrotron Light Source II (NSLS-II) of Brookhaven National Laboratory. An X-ray beam with an energy of 13.5 keV was shone on the perovskite films in grazing incident geometry. The data presented in the study were taken at an incident angle $\theta = 0.2$, which probes the bulk structure of the films. The scattering spectra were collected using an exposure time of 30 s with an area detector (DECTRIS Pilatus 800K) placed 257 mm away from the sample. Data analysis was performed by using custom-made software (SciAnalysis).[54]

Device fabrication

Unless stated otherwise, all materials were purchased from Sigma-Aldrich or Merck and used as received. MAI and FAI were purchased from Xi'an P-OLED. PbI_2 was purchased from Lumtec. The Ta-WO_x colloidal solution was purchased from Avantama. The SnO_2-PEIE solution was prepared by mixing 15 wt % SnO_2 aqueous solution (300 μL) with 1.8 mL of isopropanol and H_2O (1/1, v/v) and 20 μL of PEIE. First, ITO substrates were sonicated in acetone/isopropanol for 10/5 min, respectively. Before spin coating the SnO_2-PEIE solution (80 μL) at 3,500 rpm for 30 s, the ITO substrates were treated with UV-ozone for 10 min in ambient air. After annealing at 150°C/10 min in ambient air, 80 μL of PCBM:PMMA solution was spin coated on a SnO_2/PEIE layer at 2,000 rpm for 30 s and then annealed at 150°C for 10 min. A 1.2 M PbI_2 and FAI solution was prepared first in DMF and DMSO (4:1, v/v), and 1.2 M MAI/CsI solution was prepared in DMSO. The $MA_xCs_yFA_{1-x-y}PbI_3$ precursors were prepared by mixing the mother solutions in the target ratio. The perovskite precursor solution was spin coated on the PCBM substrate using the following parameters: 200 rpm for 2 s, 2,000 rpm for 2 s, and 5,000 rpm for 40 s (a = 3 s). Then, 180 μL of chlorobenzene was dropped on the film at 20 s, followed by annealing at 110°C for 10 min and 150°C for 5 min. PDCBT for the hole transporting layer was spin coated at 2,000 rpm for 40 s and annealed at 90°C for 5 min. Finally, 100 μL of Ta-WO_x was coated on the PDCBT at 2,000 rpm for 30 s and annealed at 75°C in ambient air. A 100-nm-thick Au electrode was deposited through a shadow mask via thermal evaporation. For the devices used in the stability tests, a 200-nm Au layer was deposited.

Device characterization

The *J*–*V* curves of the solar cells were obtained using a Keithley source under 100 mW cm^{-2} AM1.5G illumination. The *J*–*V* characteristics were measured from −0.1 to 1.2 V (forward scan) at a scan rate of 20 mV/s. No hysteresis was observed in the devices. For the thermal stability test at 85°C/85% RH, the devices were stored in a sample box in a climate chamber in the dark without any encapsulation. The devices were tested before and after storing in the climate chamber for 50 h. For the photostability test at 85°C, the devices were stored in a sealed chamber with N_2 flow under metal halide light illumination (100 mW cm^{-2}). The devices were continuously measured with forward scan at a scan rate of 20 mV/s.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.matt.2021.01.008.

ACKNOWLEDGMENTS

The authors thank members of the MIT Photovoltaics Research Laboratory, in particularly Jim Serdy for setting up and maintaining the environmental chamber for accelerated aging tests; Antonio Buscemi and Isaac Metcalf for assistance in XRD measurements, Isaac Siyu Tian for initial image processing, and Shreyaa Raghavan for the development of the user-friendly Python package. S.S. thanks Dr. Masafumi Fukuto and Dr. Kevin Yager from Brookhaven National Laboratory for beamtime assistance. High-throughput GIWAXS images were collected at BM-11, NSLS-II. The authors thank Prof. Joshua Schrier (Fordham University) for the discussion on perovskite material discovery using machine learning techniques. S.S. further thanks Felice Frankel for valuable insights on graphic visualization, and Mariela Castillo for early-stage project management.

This study is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under contract no. HR001118C0036. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA. S.S., N.T.P.H., and T.B. thank TOTAL S.A., the US National Science Foundation (grant CBET-1605547), and the Skoltech NGP program. The research is also supported by Singapore Massachusetts Institute of Technology (MIT) Alliance for Research and Technology's Low Energy Electronic Systems research program. A.T. was supported by the Alfred Kordelin Foundation and Svenska Tekniska Vetenskaps-akademien i Finland. F.O. was supported by the U.S. Department of Energy under Photovoltaic Research and Development program under award DE-EE0007535. This research used computational resources sponsored by the DOE Office of Energy Efficiency and Renewable Energy and located at National Renewable Energy Laboratory. Y.Z., T.H., and

C.J.B. gratefully acknowledge financial support through the "Aufbruch Bayern" initiative of the state of Bavaria (EnCN and "Solar Factory of the Future"), the Bavarian Initiative "Solar Technologies go Hybrid" (SolTech), "ELF-PV Design and development of solution processed functional materials for the next generations of PV technologies" by the Bavarian State Government (no. 44-6521a/20/4), and the DFG, project nos. 182849149–SFB 953, IGK2495, and INST 90/1093/1. J.J.Y. and M.B. were supported by the Institute for Soldier Nanotechnology (ISN) (grant W911NF-13-D-0001), the National Aeronautics and Space Administration (NASA) (grant NNX16AM70H), and the Eni-MIT Alliance Solar Frontiers Center.

AUTHOR CONTRIBUTIONS

S.S. and T.B. conceived the project. S.S. designed the experiments and led the data collection and interpretation. A.T. developed color and chamber environment calibration. A.T. and Z.L. processed the camera data. F.O. and J.F. designed the Bayesian optimization algorithms and instability index, and F.O., A.T., and Z.L. wrote the Bayesian optimization implementation. J.T. synthesized all thin films samples supervised by S.S. Y.Z. fabricated the devices. J.T. and N.T.P.H. performed laboratory structural and morphological characterization. J.T., A.T., and N.T.P.H. performed the film degradation tests. A.G. performed DFT calculations. T.H., Y.Z., and I.M.P. performed device degradation tests. F.O., A.T., Z.L., and J.F. developed the incorporation of DFT into the Bayesian optimization loop, and A.T. implemented the Gaussian process model and probability distribution on phase stability. A.E. built the automatic humidity control supervised by S.S. and A.T. C.B. and A.T. developed the camera control program and visualized degradation data. C.B. performed optical characterization supervised by S.S. S.S., J.T., N.T.P.H., and J.J.Y. performed structural characterization using synchrotron radiation. R.L. assisted in the measurements at beamline BM-11, NSLS-II. Z.R. performed the teacher-student model for DFT and non-DFT optimization comparison. C.J.B., M.G.B., V.S., J.F., and T.B. supervised the project. S.S., A.T., F.O., and T.B. wrote the manuscript with contributions from all co-authors. All authors discussed the results and inferred the implications of the project.

DECLARATION OF INTERESTS

MIT Photovotlaics Research Laboratory has filed IP pertaining to various aspects of materials development.

Received: September 8, 2020
Revised: November 4, 2020
Accepted: January 6, 2021
Published: February 1, 2021

REFERENCES

1. Boyd, C.C., Cheacharoen, R., Leijtens, T., and McGehee, M.D. (2019). Understanding degradation mechanisms and improving stability of perovskite photovoltaics. Chem. Rev. *119*, 3418–3451, https://doi.org/10.1021/acs.chemrev.8b00336.
2. Jeon, N.J., Noh, J.H., Yang, W.S., Kim, Y.C., Ryu, S., Seo, J., and Seok, S.I. (2015). Compositional engineering of perovskite materials for high-performance solar cells. Nature *517*, 476–480, https://doi.org/10.1038/nature14133.
3. Saliba, M. (2019). Polyelemental, multicomponent perovskite semiconductor libraries through combinatorial screening. Adv. Energy Mater. *9*, 1803754, https://doi.org/10.1002/aenm.201803754.
4. Khenkin, M.V., Katz, E.A., Abate, A., Bardizza, G., Berry, J.J., Brabec, C., Brunetti, F., Bulović, V., Burlingame, Q., Di Carlo, A., et al. (2020). Consensus statement for stability assessment and reporting for perovskite photovoltaics based on ISOS procedures. Nat. Energy *5*, 35–49, https://doi.org/10.1038/s41560-019-0529-5.
5. Rohr, B., Stein, H.S., Guevarra, D., Wang, Y., Haber, J.A., Aykol, M., Suram, S.K., and Gregoire, J.M. (2020). Benchmarking the acceleration of materials discovery by sequential learning. Chem. Sci. *11*, 2696–2706, https://doi.org/10.1039/c9sc05999g.
6. Zhang, W., Mao, J., Li, S., Chen, Z., and Guo, Z. (2017). Phosphorus-based alloy materials for advanced potassium-ion battery anode. J. Am. Chem. Soc. *139*, 3316–3319, https://doi.org/10.1021/jacs.6b12185.
7. George, E.P., Raabe, D., and Ritchie, R.O. (2019). High-entropy alloys. Nat. Rev. Mater. *4*,

515–534, https://doi.org/10.1038/s41578-019-0121-4.

8. Zhong, M., Tran, K., Min, Y., Wang, C., Wang, Z., Dinh, C.T., De Luna, P., Yu, Z., Rasouli, A.S., Brodersen, P., et al. (2020). Accelerated discovery of CO_2 electrocatalysts using active machine learning. Nature *581*, 178–183, https://doi.org/10.1038/s41586-020-2242-8.

9. Kumar, G., Bossert, H., McDonald, D., Chatzidimitriou, A., Ardagh, M.A., Pang, Y., Lee, C., Tsapatsis, M., Abdelrahman, O.A., and Dauenhauer, P.J. (2020). Catalysis-in-a-Box: robotic screening of catalytic materials in the time of COVID-19 and beyond. Matter *3*, 805–823, https://doi.org/10.1016/j.matt.2020.06.025.

10. Burger, B., Maffettone, P.M., Gusev, V.V., Aitchison, C.M., Bai, Y., Wang, X., Li, X., Alston, B.M., Li, B., Clowes, R., et al. (2020). A mobile robotic chemist. Nature *583*, 237–241, https://doi.org/10.1038/s41586-020-2442-2.

11. Cole, J.M. (2020). A design-to-device pipeline for data-driven materials discovery. Acc. Chem. Res. *53*, 599–610, https://doi.org/10.1021/acs.accounts.9b00470.

12. Gómez-Bombarelli, R., Aguilera-Iparraguirre, J., Hirzel, T.D., Duvenaud, D., Maclaurin, D., Blood-Forsythe, M.A., Chae, H.S., Einzinger, M., Ha, D.G., Wu, T., et al. (2016). Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. Nat. Mater. *15*, 1120–1127, https://doi.org/10.1038/nmat4717.

13. Kirman, J., Johnston, A., Kuntz, D.A., Askerka, M., Gao, Y., Todorović, P., Ma, D., Privé, G.G., and Sargent, E.H. (2020). Machine-learning-accelerated perovskite crystallization. Matter *2*, 938–947, https://doi.org/10.1016/j.matt.2020.02.012.

14. Li, Z., Najeeb, M.A., Alves, L., Sherman, A., Parrilla, P.C., Pendleton, I.M., Zeller, M., Schrier, J., Norquist, A.J., and Chan, E. (2019). Robot-accelerated perovskite investigation and discovery (RAPID): 1. Inverse Temperature Crystallization. ChemRxiv. https://doi.org/10.26434/CHEMRXIV.10013090.V1.

15. MacLeod, B.P., Parlane, F.G.L., Morrissey, T.D., Häse, F., Roch, L.M., Dettelbach, K.E., Moreira, R., Yunker, L.P.E., Rooney, M.B., Deeth, J.R., et al. (2020). Self-driving laboratory for accelerated discovery of thin-film materials. Sci. Adv. *6*, eaaz8867, https://doi.org/10.1126/sciadv.aaz8867.

16. Xue, D., Balachandran, P.V., Yuan, R., Hu, T., Qian, X., Dougherty, E.R., and Lookman, T. (2016). Accelerated search for $BaTiO_3$-based piezoelectrics with vertical morphotropic phase boundary using Bayesian learning. Proc. Natl. Acad. Sci. U S A *113*, 13301–13306, https://doi.org/10.1073/pnas.1607412113.

17. Epps, R.W., Bowen, M.S., Volk, A.A., Abdel-Latif, K., Han, S., Reyes, K.G., Amassian, A., and Abolhasani, M. (2020). Artificial chemist: an autonomous quantum dot synthesis bot. Adv. Mater. *32*, 2001626, https://doi.org/10.1002/adma.202001626.

18. Gu, E., Tang, X., Langner, S., Duchstein, P., Zhao, Y., Levchuk, I., Kalancha, V., Stubhan, T., Hauch, J., Egelhaaf, H.J., et al. (2020). Robot-based high-throughput screening of antisolvents for lead halide perovskites. Joule *4*, 1806–1822, https://doi.org/10.1016/j.joule.2020.06.013.

19. Higgins, K., Valleti, S.M., Ziatdinov, M., Kalinin, S., and Ahmadi, M. (2020). Chemical robotics enabled exploration of stability and photoluminescent behavior in multicomponent hybrid perovskites via machine learning. ChemRxiv. https://doi.org/10.26434/CHEMRXIV.12436079.V1.

20. Choubisa, H., Askerka, M., Ryczko, K., Voznyy, O., Mills, K., Tamblyn, I., and Sargent, E.H. (2020). Crystal site feature embedding enables exploration of large chemical spaces. Matter *3*, 433–448, https://doi.org/10.1016/j.matt.2020.04.016.

21. Curtarolo, S., Hart, G.L., Nardelli, M.B., Mingo, N., Sanvito, S., and Levy, O. (2013). The high-throughput highway to computational materials design. Nat. Mater. *12*, 191–201, https://doi.org/10.1038/nmat3568.

22. Jesper Jacobsson, T., Correa-Baena, J.-P., Pazoki, M., Saliba, M., Schenk, K., Grätzel, M., and Hagfeldt, A. (2016). Exploration of the compositional space for mixed lead halogen perovskites for high efficiency solar cells. Energy Environ. Sci. *9*, 1706–1724, https://doi.org/10.1039/C6EE00030D.

23. Schelhas, L.T., Li, Z., Christians, J.A., Goyal, A., Kairys, P., Harvey, S.P., Kim, D.H., Stone, K.H., Luther, J.M., Zhu, K., et al. (2019). Insights into operational stability and processing of halide perovskite active layers. Energy Environ. Sci. *12* (4), 1341–1348, https://doi.org/10.1039/c8ee03051k.

24. SheffieldML/GPyOpt: Gaussian Process Optimization Using GPy n.d.. https://github.com/SheffieldML/GPyOpt.

25. Zhu, X.H., Wei, Z.R., Jin, Y.R., and Xiang, A.P. (2007). Growth and characterization of a PbI_2 single crystal used for gamma ray detectors. Cryst. Res. Technol. *42*, 456–459, https://doi.org/10.1002/crat.200610847.

26. Hu, Y., Bai, F., Liu, X., Ji, Q., Miao, X., Qiu, T., and Zhang, S. (2017). Bismuth incorporation stabilized α-$CsPbI_3$ for fully inorganic perovskite solar cells. ACS Energy Lett. *2*, 2219–2227, https://doi.org/10.1021/acsenergylett.7b00508.

27. Masi, S., Echeverría-Arrondo, C., Salim, K.M.M., Ngo, T.T., Mendez, P.F., López-Fraguas, E., Macias-Pinilla, D.F., Planelles, J., Climente, J.I., and Mora-Sero, I. (2020). Chemi-structural stabilization of formamidinium lead iodide perovskite by using embedded quantum dots. ACS Energy Lett. 418–427, https://doi.org/10.1021/acsenergylett.9b02450.

28. Hashmi, S.G., Tiihonen, A., Martineau, D., Ozkan, M., Vivo, P., Kaunisto, K., Ulla, V., Zakeeruddin, S.M., and Grätzel, M. (2017). Long term stability of air processed inkjet infiltrated carbon-based printed perovskite solar cells under intense ultra-violet light soaking. J. Mater. Chem. A *5*, 4797–4802, https://doi.org/10.1039/c6ta10605f.

29. Hashmi, S.G., Martineau, D., Li, X., Ozkan, M., Tiihonen, A., Dar, M.I., Sarikka, T., Zakeeruddin, S.M., Paltakari, J., Lund, P.D., and Grätzel, M. (2017). Air processed inkjet infiltrated carbon based printed perovskite solar cells with high stability and reproducibility. Adv. Mater. Technol. *2*, 1600183, https://doi.org/10.1002/admt.201600183.

30. Stoddard, R.J., Dunlap-Shohl, W.A., Qiao, H., Meng, Y., Kau, W.F., and Hillhouse, H.W. (2020). Forecasting the decay of hybrid perovskite performance using optical transmittance or reflected dark-field imaging. ACS Energy Lett. *5*, 946–954, https://doi.org/10.1021/acsenergylett.0c00164.

31. Kim, S., Eom, T., Ha, Y.-S., Hong, K.-H., and Kim, H. (2020). Thermodynamics of multicomponent perovskites: a guide to highly efficient and stable solar cell materials. Chem. Mater. *32*, 4265–4272, https://doi.org/10.1021/acs.chemmater.0c00893.

32. Knight, A.J., and Herz, L.M. (2020). Preventing phase segregation in mixed-halide perovskites: a perspective. Energy Environ. Sci. *13*, 2024, https://doi.org/10.1039/d0ee00788a.

33. Herbol, H.C., Poloczek, M., and Clancy, P. (2020). Cost-effective materials discovery: Bayesian optimization across multiple information sources. Mater. Horiz. *7*, 2113, https://doi.org/10.1039/D0MH00062K.

34. Doan, H.A., Agarwal, G., Qian, H., Counihan, M.J., Rodríguez-López, J., Moore, J.S., and Assary, R.S. (2020). Quantum chemistry-informed active learning to accelerate the design and discovery of sustainable energy storage materials. Chem. Mater. *32*, 6338, https://doi.org/10.1021/acs.chemmater.0c00768.

35. Gelbart, M.A., Snoek, J., and Adams, R.P. (2014). Bayesian optimization with unknown constraints. Uncertain. Artif. Intell. Proc. 30th Conf. UAI *2014*, 250–259.

36. Saliba, M., Matsui, T., Seo, J.Y., Domanski, K., Correa-Baena, J.P., Nazeeruddin, M.K., Zakeeruddin, S.M., Tress, W., Abate, A., Hagfeldt, A., and Grätzel, M. (2016). Cesium-containing triple cation perovskite solar cells: improved stability, reproducibility and high efficiency. Energy Environ. Sci. *9*, 1989–1997, https://doi.org/10.1039/C5EE03874J.

37. Kieslich, G., Sun, S., and Cheetham, A.K. (2014). Solid-state principles applied to organic–inorganic perovskites: new tricks for an old dog. Chem. Sci. *5*, 4712–4715, https://doi.org/10.1039/C4SC02211D.

38. Li, Z., Yang, M., Park, J.-S., Wei, S.-H., Berry, J.J., and Zhu, K. (2016). Stabilizing perovskite structures by tuning tolerance factor: formation of formamidinium and cesium lead iodide solid-state alloys. Chem. Mater. *28*, 284–292, https://doi.org/10.1021/acs.chemmater.5b04107.

39. Correa-Baena, J.P., Saliba, M., Buonassisi, T., Grätzel, M., Abate, A., Tress, W., and Hagfeldt, A. (2017). Promises and challenges of perovskite solar cells. Science *358*, 739–744, https://doi.org/10.1126/science.aam6323.

40. Saidaminov, M.I., Williams, K., Wei, M., Johnston, A., Quintero-Bermudez, R., Vafaie, M., Pina, J.M., Proppe, A.H., Hou, Y., Walters, G., et al. (2020). Multi-cation perovskites prevent carrier reflection from grain surfaces.

Nat. Mater. *19*, 412–418, https://doi.org/10.1038/s41563-019-0602-2.

41. Correa-Baena, J.P., Luo, Y., Brenner, T.M., Snaider, J., Sun, S., Li, X., Jensen, M.A., Hartono, N.T.P., Nienhaus, L., Wieghold, S., et al. (2019). Homogenized halides and alkali cation segregation in alloyed organic-inorganic perovskites. Science *363*, 627–631, https://doi.org/10.1126/science.aah5065.
42. González, J., Dai, Z., Hennig, P., and Lawrence(n.d.)., N. (2016). Batch Bayesian Optimization via Local Penalization. Artificial intelligence and statistics, 648–657.
43. Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep Learning (MIT press Cambridge).
44. Goyal, A., McKechnie, S., Pashov, D., Tumas, W., van Schilfgaarde, M., and Stevanović, V. (2018). Origin of pronounced nonlinear band gap behavior in lead–tin hybrid perovskite alloys. Chem. Mater. *30*, 3920–3928, https://doi.org/10.1021/acs.chemmater.8b01695.
45. Blöchl, P.E. (1994). Projector augmented-wave method. Phys. Rev. B Condens. Matter *50*, 17953–17979, https://doi.org/10.1103/PhysRevB.50.17953.
46. Kresse, G., and Furthmüller, J. (1996). Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput. Mater. Sci. *6*, 15–50, https://doi.org/10.1016/0927-0256(96)00008-0.
47. Perdew, J.P., Burke, K., and Ernzerhof, M. (1996). Generalized gradient approximation made simple. Phys. Rev. Lett. *77*, 3865–3868, https://doi.org/10.1103/PhysRevLett.77.3865.
48. Monkhorst, H.J., and Pack, J.D. (1976). Special points for Brillouin-zone integrations. Phys. Rev. B *13*, 5188–5192, https://doi.org/10.1103/PhysRevB.16.1748.
49. Stoumpos, C.C., Malliakas, C.D., and Kanatzidis, M.G. (2013). Semiconducting tin and lead iodide perovskites with organic cations: phase transitions, high mobilities, and near-infrared photoluminescent properties. Inorg. Chem. *52*, 9019–9038, https://doi.org/10.1021/ic401215x.
50. Zunger, A., Wei, S., Ferreira, L.G., and Bernard, J.E. (1990). Special quasirandom structures. Phys. Rev. Lett. *65*, 353–356, https://doi.org/10.1103/PhysRevLett.65.353.
51. Van de Walle, A., Asta, M., and Ceder, G. (2002). The alloy theoretic automated toolkit: a user guide. Calphad *26*, 539–553, https://doi.org/10.1016/S0364-5916(02)80006-2.
52. Van De Walle, A., Tiwary, P., De Jong, M., Olmsted, D.L., Asta, M., Dick, A., Shin, D., Wang, Y., Chen, L.-Q., and Liu, Z.-K. (2013). Efficient stochastic generation of special quasirandom structures. Calphad *42*, 13–18, https://doi.org/10.1016/j.calphad.2013.06.006.
53. Lejaeghere, K., Bihlmayer, G., Björkman, T., Blaha, P., Blügel, S., Blum, V., Caliste, D., Castelli, I.E., Clark, S.J., Dal Corso, A., et al. (2016). Reproducibility in density functional theory calculations of solids. Science *351*, aad3000, https://doi.org/10.1126/science.aad3000.
54. Yager, K.G. (2021). https://github.com/CFN-softbio/Sci/Analysis. [Accessed 21 January 2021].