

## ARTICLE OPEN



# Discovery of Pb-free hybrid organic–inorganic 2D perovskites using a stepwise optimization strategy

Byung Do Lee<sup>1,3</sup>, Jin-Woong Lee<sup>1,3</sup>, Minseuk Kim<sup>1</sup>, Woon Bae Park<sup>2</sup>✉ and Kee-Sun Sohn<sup>1</sup>✉

The current status of 2D organic–inorganic hybrid perovskites for use in photovoltaic (PV) and light-emitting diode (LED) applications lags far behind their 3D counterparts. Here, we propose a computational strategy for discovering novel perovskites with as few computing resources as possible. A tandem optimization algorithm consisting of an elitism-reinforced nondominated sorting genetic algorithm (NSGA-II) and a multiobjective Bayesian optimization (MOBO) algorithm was used for density functional theory (DFT) calculations. The DFT-calculated band gap and effective mass were taken as objective functions to be optimized, and the constituent molecules and elements of a Ruddlesden–Popper (RP) structure ( $n = 2$ ) were taken as decision variables. Fourteen previously unknown RP perovskite candidates for PV and LED applications were discovered as a result of the NSGA-II/MOBO algorithm. Thereafter, more accurate DFT calculations based on the HSE06 exchange correlation functional and ab initio molecular dynamics (AIMD) were conducted for the discovered 2D perovskites to ensure their validity.

*npj Computational Materials* (2022)8:83; <https://doi.org/10.1038/s41524-022-00781-z>

## INTRODUCTION

Pb-free 2D hybrid organic–inorganic perovskites (HOIPs) have attracted considerable attention due to several advantages over their 3D counterparts, including enhanced structural stability, improved moisture resistance, and a widened band gap<sup>1–7</sup>. In addition to typical photovoltaic (PV) and light-emitting diode (LED) applications, 2D perovskites have applications in other optoelectronic devices, such as photoelectric detectors, field-effect transistors, spintronics, waveguides, photocatalysis and nanolasers<sup>8–13</sup>. The present investigation focuses solely on the discovery of Pb-free 2D HOIPs for use in PV and LED applications. Although there are many reasons for the scarcity of promising Pb-free 2D HOIPs in the field, a primary concern is the lack of feasible preliminary candidates that at least fulfill band gap matching. Once a large pool of preliminary candidates has been secured, it will be much easier to identify suitable candidates that meet the remaining stricter conditions required for PV and LED applications. The search for such candidates, however, is not an easy task, as experimental high-throughput screening for HOIPs is still in its early stages<sup>14</sup>. One of the most efficient approaches in the search for appropriate candidates is employing density functional theory (DFT) calculations, which may lead to tangible discoveries. Rather than a pure DFT calculation-based approach, a data-driven machine learning (ML) approach in conjunction with DFT calculations has recently been used as an alternative, more efficient material discovery strategy<sup>15–21</sup> and has already been applied in the field of perovskite research<sup>22–34</sup>.

We propose a different type of data-driven strategy for the discovery of virtual Pb-free 2D HOIPs based on DFT calculations and nominate a number of virtual Pb-free 2D HOIPs suitable for PV and LED applications. The alternative data-driven strategy that we suggest is titled ‘stepwise optimization’, which does not use ML but rather involves metaheuristic<sup>35,36</sup> and Bayesian optimization (BO)<sup>37</sup>. Similar approaches, such as closed-loop optimization<sup>15</sup> or active learning<sup>16</sup>, have recently been reported, mainly based on

BO, despite the fact that ML was still in use. The underlying principle for stepwise optimization is similar to that of data-driven approaches<sup>15,16</sup>, but the stepwise optimization approach differs from other data-driven approaches in that we do not use a ML algorithm or parameter estimation. The stepwise optimization has already been utilized and allowed us to discover many inorganic functional materials over the last two decades<sup>38–42</sup>.

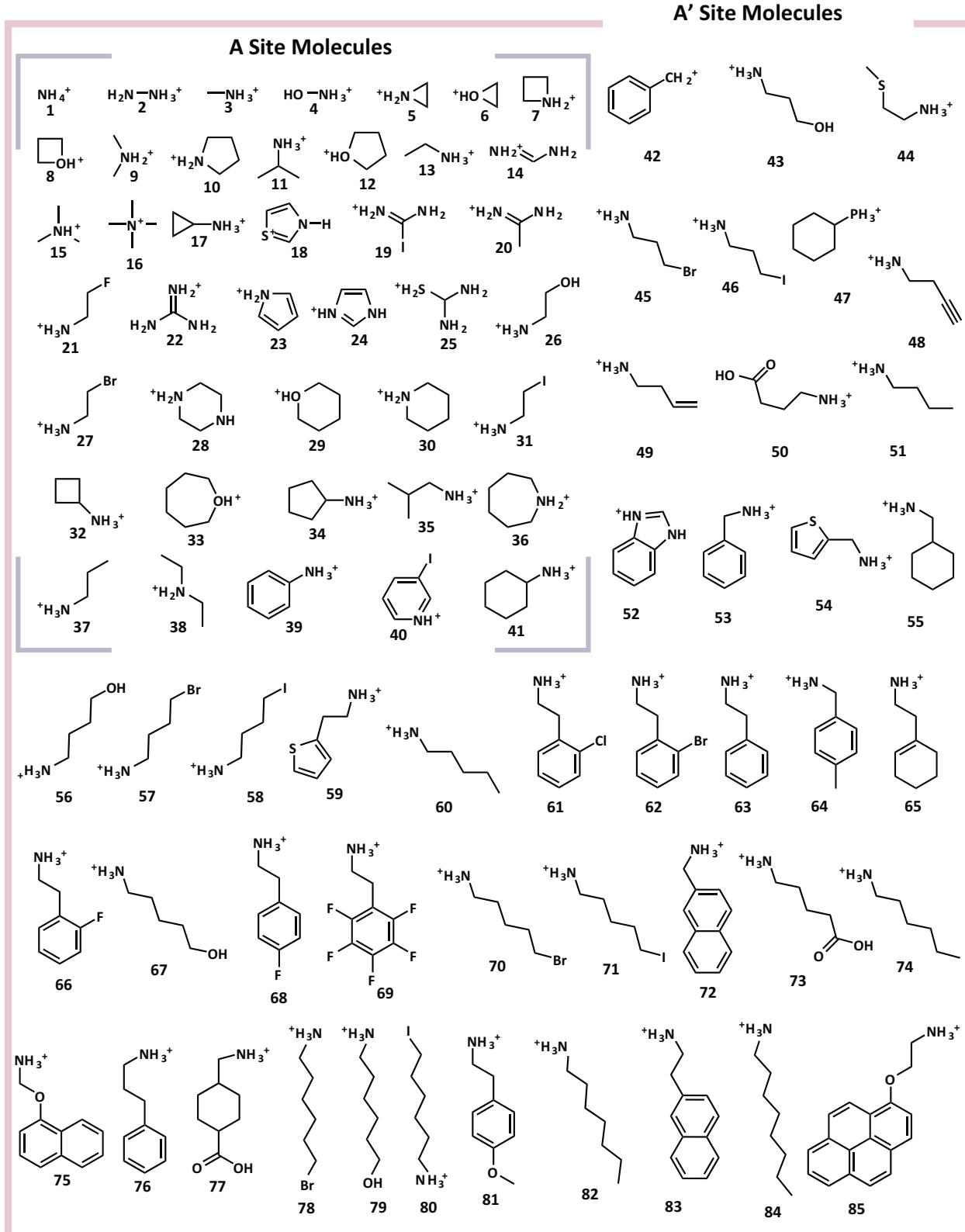
We use this strategy to discover novel Ruddlesden–Popper (RP) phase perovskites with a fixed number of inorganic octahedron layers ( $n = 2$ )<sup>43,44</sup>, which are regarded as representative 2D perovskites and are among the most commonly studied, although there are equally important Dion–Jacobson phases<sup>45–47</sup> and alternating cations in the interlayer space (ACI) phases<sup>48,49</sup>. There should be a tremendously large number of virtual RP phase perovskites by adopting many possible organic spacer molecules, smaller molecules residing in 3D octahedron slabs, octahedron-forming elements (except Pb) and several halide elements. There is no way to identify such a large number of candidates, even though we limited ourselves only to RP structures with  $n = 2$ . We need an optimization strategy to reduce the computational burden. We introduce a tandem optimization strategy that combines an elitism-reinforced nondominated sorting genetic algorithm (NSGA-II)<sup>50–53</sup> and a multiobjective Bayesian optimization (MOBO) algorithm<sup>54–56</sup>. Our major aim is to suggest hypothetical Pb-free HOIPs in a limited search space (RP with  $n = 2$ ). Other larger search space including other phases with other  $n$  values will be tracked down separately.

The key point of the proposed approach is the establishment of preset specific target properties (in this case, band gap and effective mass). The preset targets are achieved through a step-by-step improvement of the decision variables. A large amount of data is not required since no ML model training process is required in this approach. Only a small amount of data is required to instantaneously evaluate the objective function values in each step of the NSGA-II and MOBO iterations. The NSGA-II and MOBO

<sup>1</sup>Department of Nanotechnology and Advanced Materials Engineering, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul 143-747, South Korea. <sup>2</sup>Department of Advanced Components and Materials Engineering, Sunchon National University, Chonnam 57922, Republic of Korea. <sup>3</sup>These authors contributed equally: Byung Do Lee, Jin-Woong Lee.  
✉email: wbpark@schnu.ac.kr; kssohn@sejong.ac.kr

algorithms outperform well-known gradient-based optimization algorithms for small problems with at most a few decision variables. The NSGA-II and MOBO algorithms that we adopted for the HOIP discovery process considerably reduced the computational costs of DFT calculations. Although we have previously reported many similar NSGA-II-driven approaches for the

discovery of various materials<sup>38–42</sup>, a NSGA-II and MOBO tandem strategy has never been attempted. BO has been used for simple single-objective optimization problems in some cases<sup>57–60</sup>. However, it has never been used for a multiobjective optimization problem involving material discovery, despite the fact that MOBO has been well developed in other disciplines<sup>54–56</sup>.



**Fig. 1 Eighty-four molecule candidates to be inserted in the A'-site of the RP structure ( $n = 2$ ), among which 41 relatively smaller molecules (located on the upper-left side) were shared with the A-site.** 1. Ammonium, 2. Hydrazinium(HY), 3. Methylammonium(MA), 4. Hydroxylammonium, 5. Aziridinium, 6. Oxiran-1-i um, 7. Azetidinium(AZ), 8. Oxetan-1-i um, 9. Dimethylammonium(DiMA), 10. Pyrrolidinium(PY), 11. 2-Propanaminium, 12. Tetrahydrofuranium, 13. Ethylammonium(EA), 14. Formamidinium(FA), 15. Trimethylammonium(TriMA), 16. Tetramethylammonium(TetraMA), 17. Cyclopropylammonium, 18. Thiazolium, 19. Iodoformamidinium, 20. 1-Aminoethaniminium, 21. 1-Aminoethaniminium, 22. Guanidinium(GA), 23. 1H-Pyrrole-1-i um, 24. Imidazolium(IM), 25. Protonated thiourea cation, 26. Heteroatom substituted alkylammonium, 27. Ethylammonium bromide, 28. Piperazin-1-i um(PIP), 29. Tetrahydropyran-1-i um, 30. Piperidinium, 31. Ethylammonium iodide, 32. Cyclobutylammonium, 33. Oxepan-1-i um, 34. Cyclopentylammonium, 35. isobutyl-ammonium(IBA), 36. Azepanium, 37. propan-1-ammonium, 38. Diethylammonium(DiEA), 39. Phenyltrimethylammonium(PTA), 40. 3-iodopyridinium, 41. Cyclohexylammonium, 42. Tropylium, 43. Propylammonium hydroxide, 44. 2-(methylthio)ethylamine(MTEA), 45. Propylammonium bromide, 46. Propylammonium iodide, 47. Cyclohexylphosphonium, 48. But-3-yn-1-ammonium(BYA), 49. Allylammonium(ALA), 50. Ammonium 4-butyratric acid, 51. Butan-1-ammonium(BA), 52. Benzimidazolium, 53. Benzylammonium(BZA), 54. 2-thienylmethylammonium(ThMA), 55. Cyclohexylmethylammonium, 56. Butylammonium hydroxide, 57. Butylammonium bromide, 58. Butylammonium iodide, 59. 2-(2-thienyl)ethanaminium, 60. Penta-1-ammonium, 61. 2-chlorophenethylammonium, 62. 2-bromophenethylammonium, 63. Phenylethylammonium(PEA), 64. 4-methylbenzylammonium, 65. 2-(1-cyclohexenyl)ethylammonium, 66. 2-fluorophenethylammonium(FPEA), 67. Pentylammonium hydroxide, 68. 4-fluorophenylethylammonium(F-PEA), 69. Perfluorophenethylammonium(F5-PEA), 70. Pentylammonium bromide, 71. Pentylammonium iodide, 72. 1-(2-naphthyl)methanammonium(NMA), 73. 5-ammonium valeric acid(5-AVA), 74. Hexane-1-ammonium, 75. Naphthalene-O-ethylammonium, 76. Propyl phenyl ammonium(PPA), 77. (carboxy)Cyclohexylmethylammonium(TRA), 78. Hexylammonium bromide, 79. Hexylammonium hydroxide, 80. Hexylammonium iodide, 81. 4-methoxyphenethylammonium(MeO-PEA), 82. Heptan-1-ammonium, 83. 2-(2-naphthyl)ethanammonium(NEA), 84. Octan-1-ammonium, 85. Pyrene-O-ethylammonium.

## RESULTS AND DISCUSSION

### Problem setting

We independently implemented two closed-loop stepwise optimization processes: one for the discovery of HOIPs for LED applications and the other for PV applications. Different objective function settings with the same decision variable setting were used to investigate both targets, i.e., two different band gaps were set as targets and optimized independently in the same search space. Five generations (rounds or steps) were made for the NSGA-II/MOBO optimization process. Conventional NSGA-II (or MOBO) computations are iterated thousands of times until a complete optimization is achieved. In contrast, our NSGA-II/MOBO processes were completed after the fifth round due to the high-cost DFT calculation-based evaluation of the objective functions. However, the limited iterations up to at best the fifth round are acceptable by considering the fact that the most drastic improvement was always achieved during the initial stage of iterations, regardless of what metaheuristic-based optimization algorithms were used. The main goal of our NSGA-II/MOBO processes was not to achieve complete convergence on the global optimum but to provide a greater number of plausible HOIP candidates than other knowledge-based trial-and-error approaches.

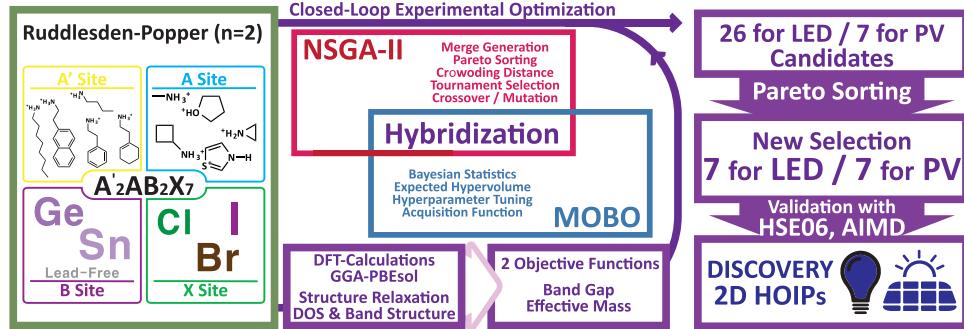
Objective function selection is critical for the optimization process. There is no single objective function that can incorporate all the properties of interest. We chose objective functions with two critical properties: band gap and effective mass. The formation energy, decomposition energy and energy above the convex hull, which represent thermodynamic stability, were omitted since complex molecule inclusion made it impossible to obtain proper information on the constituent compounds and their convex hull. It is inevitable to remove the thermodynamic stability from the NSGA-II/MOBO process since DFT calculations are unavailable for binary and ternary compounds that involve organic molecules. Almost none of the binary and ternary compounds involving A'- and A-site organic molecules have been discovered yet. It would be irrational to involve DFT calculations for the non-existent structures. Most plausible approach at the current status of development is to adopt hypothetical binary and ternary structures by introducing existing simple-metal-based binary and ternary compounds as a template structure for the non-existent organic-molecule-involved structures<sup>23,61–63</sup>. We also adopted this type of approach in our previous report<sup>34</sup>, wherein we were dealing with relatively small molecules only. In the present investigation, however, we are dealing with much bigger A' site molecules than ever before. It would be a complete

non-sense to use such a simple-metal-based binary or ternary compound as a template for such a complicated large-organic-molecule-based binary or ternary compound that could scarcely be actually synthesizable.

In addition, the DFT calculation-driven thermodynamic stability represented by formation energy, decomposition energy and energy above the convex hull, etc. would seem to play no significant role in judging if a hypothetical HOIP of concern is useful for the standard solar energy harvesting or light emitting applications, while both the band gap and effective mass play a crucial role. In fact, the decomposition energy had a much smaller variance than the band gap and effective mass in our previous report<sup>34</sup>. The variance for the min-max-normalized decomposition energy, band gap and effective mass for 266 hypothetical HOIPs were estimated to be 0.008, 0.032, and 0.014, respectively. The variance for the decomposition energy was much smaller in comparison to those for band gap and effective mass. This clearly confirmed that the decomposition energy would not play a key role as an objective function in the NSGA-II/MOBO process.

The decision variables that we adopted were A'-site molecules (85 molecules), A-site molecules (41 molecules), B-site atoms (2 atoms), and X-site atoms (3 atoms) in the RP structure ( $n = 2$ ), which yielded 20,910 possible HOIP candidates. Figure 1 shows the chosen candidates for A'- and A-site molecules. The careful selection of organic spacer materials, i.e., the A'-site molecule, plays a crucial role in determining the HOIP performance. We introduced all possible candidates as A'-site molecules, and some of the molecules were also selected as A-site molecules based on the size: 41 smaller molecules were simultaneously nominated as A'- and A-site molecule candidates. The B-site atoms were Sn and Ge (Pb was excluded), and the X-site atoms were Cl, Br, and I. It would be practically impossible to determine all the candidates via DFT and ab initio molecular dynamics (AIMD) calculations because the calculation time would be prohibitively long for HOIPs with relatively large cells involving many atoms per cell (>100) when using conventional lab-scale computational capacity. The partial mixing of molecules or atoms at their A', A, B and X sites was not allowed for the same reason.

Figure 2 shows a schematic of the entire NSGA-II/MOBO optimization process, elucidating the closed-loop optimization strategy that differs from conventional ML approaches. The objective functions and decision variables are schematically described in Fig. 2. NSGA-II was hybridized with the MOBO algorithm in such a way that every generation included two MOBO-selected entries. A brief description of the hybridized optimization process is also schematically provided in Fig. 2, and a more detailed description of the process is given in the



**Fig. 2 Schematic representation of the entire computational process involving the DFT-calculated-data-based NSGA-II/MOBO process.** The search space based on the RP structure ( $n = 2$ ) is schematically summarized on the leftmost side. The hybridized NSGA-II and MOBO algorithm are shown in the middle. All the operators for NSGA-II and MOBO are listed in the box. The NSGA-II/MOBO process led to 24 HOIP candidates for LEDs, which were reduced to seven candidates with stricter selection criterion, and in parallel, seven candidates for solar cell applications were directly selected from the NSGA-II/MOBO process.

Methods section. Two parallel optimizations were carried out independently: one for PV applications and one for LED applications. The band gap objective function for PVs ( $E_g$ \_PV) was set as the energy difference from 0.74 eV, which was determined by considering the GGA underestimation<sup>34,64–66</sup> of the Shockley–Queisser efficiency limit, which is 1.34 eV<sup>67</sup>. The band gap objective function for LEDs ( $E_g$ \_LED) was set as the energy difference from the GGA-based 1.98 eV, which corresponds to the GGA underestimation of a typical blue range peaking at 2.76 eV, since our goal is to pursue blue PeLEDs.

The effective mass objective function ( $m_{e/h}$ ) for both the PV- and LED-targeted NSGA-II/MOBO optimization processes was defined as the greater effective mass value of either an electron or a hole. In principle, a lower value should be preferred in PV and LED applications because it represents the charge transport potential. Nonetheless, the maximum value was chosen as an objective function because it was minimized in the NSGA-II/MOBO optimization process. It is logically more favorable to minimize the maximum rather than simply pursuing the minimum in an attempt to achieve minimization while simultaneously ensuring a balance between the electron and hole effective masses.

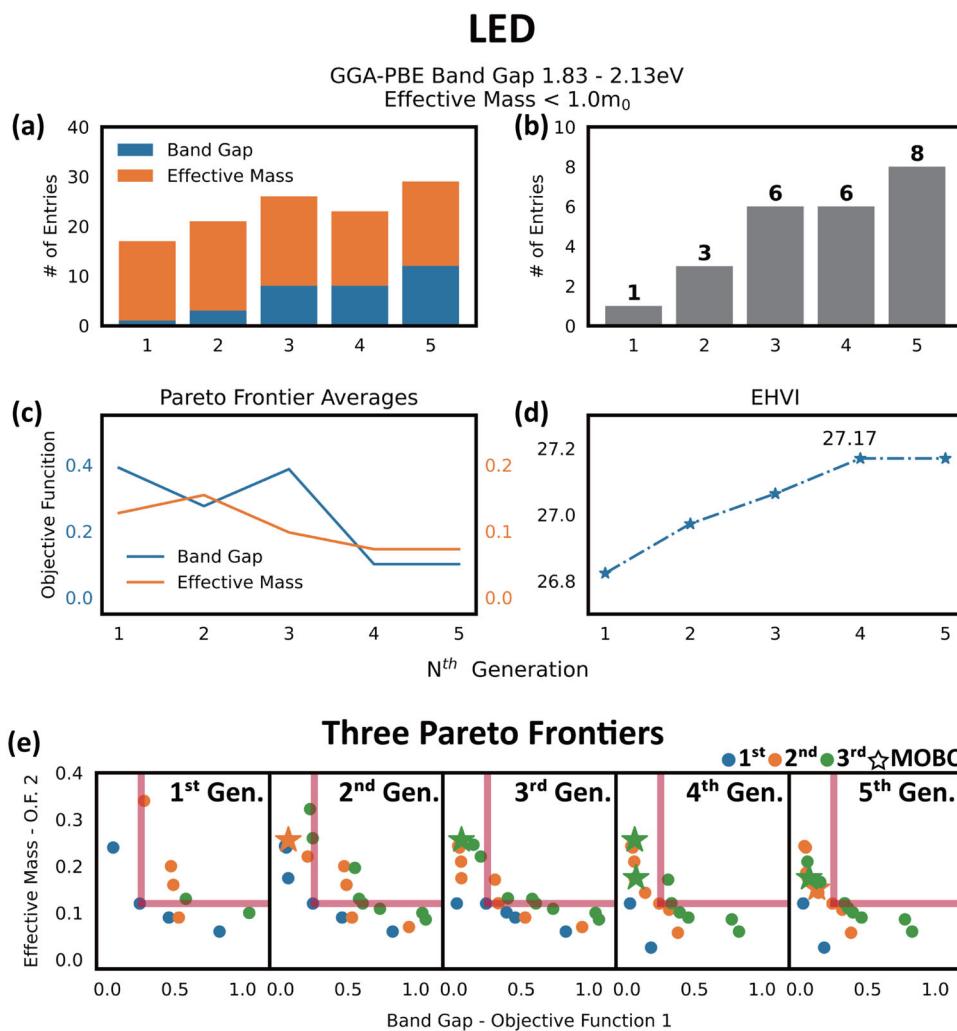
The population size for each generation was 20 for NSGA-II, and two entries recommended by the MOBO algorithm were inserted into each generation except for the randomly chosen first generation, yielding 22 HOIP entries for each generation after the first generation. This results in a total of 216 entries (108 for each of the PV and LED applications). The MOBO-recommended entries at each generation were derived from all previous generations, not just the immediately preceding generation. The NSGA-II/MOBO hybridized optimization process was implemented until the fifth round (generation). Although five NSGA-II generations, along with MOBO, may seem insufficient for complete optimization, it is not an issue because it is well known that the most significant improvement always occurs during the initial stage of optimization, as discussed above. In addition, it should be noted that the goal of the NSGA-II/MOBO hybridized optimization process is to obtain preliminary, rough guidance rather than complete optimization.

### Search for 2D HOIPs for use in LED applications

The current development status of blue perovskite light-emitting diodes (PeLEDs) lags far behind those of their red and green counterparts. A desirable blue light-emitting 2D HOIP has yet to be discovered, in contrast to well-established green and red light-emitting perovskites that achieve external quantum efficiencies beyond 20%<sup>68–70</sup>. Blue HOIPs can be fabricated by reducing the dimension of the perovskites, and phases with smaller  $n$  values typically have larger band gaps<sup>71</sup>. However, a strong exciton–phonon

coupling of 2D perovskites would limit the external quantum efficiency of 2D blue PeLEDs<sup>72</sup>. Although modifying the bromide/chloride composition is a well-known method for achieving blue emission, the practical implementation of this strategy has been challenging due to color instability, phase separation caused by electric fields in LED applications, and severe photoluminescence quenching<sup>73</sup>. Using mixed-dimensional perovskites or colloidal perovskite nanocrystals, bandgap engineering has been attempted<sup>74–76</sup>. However, the combination of 2D and 3D perovskites, known as quasi-2D perovskites, is not applicable for DFT calculations due to our computational capacity. Therefore, we changed the template structure to a 2D RP ( $n = 2$ ) structure and searched for novel blue light-emitting HOIPs within this structure. In this regard, the decision variables involve each constituent molecule/element at the A', A, B, and X sites in the fixed 2D RP ( $n = 2$ ) structure, with no co-occupancy allowed at any site.

Figure 3 shows the summarized result obtained from the NSGA-II/MOBO-hybridized optimization process. Evidence for the intended evolution was clearly detected. Specifically, the NSGA-II/MOBO process drove the HOIP entries to minimize both the objective functions,  $E_g$ \_LED and  $m_{e/h}$ . The bad gap objective function ( $E_g$ \_LED) differs from the band gap itself ( $E_g$ ). Accordingly, the band gap shifted toward the intended region (GGA-translated blue region) by minimizing  $E_g$ \_LED, and the effective mass continued to decrease as the NSGA-II/MOBO iteration progressed. Figure 3a shows the evidence for the evolution in terms of the number of promising entries with band gaps between 1.83 and 2.13 eV (GGA-translated blue region) and effective masses less than the rest mass of a free electron,  $m_0$ . Figure 3b also shows the number of promising entries that fulfill both requirements simultaneously. The number of promising entries increased as the NSGA-II/MOBO iteration continued until the fifth generation. The average objective function values for the first Pareto frontier are also plotted as a function of the generation number in Fig. 3c, with both  $E_g$ \_LED and  $m_{e/h}$  decreasing as the NSGA-II/MOBO iteration progresses. The expected hypervolume improvement (EHVI) is also plotted as a function of the generation number, with a clearly increasing trend shown in Fig. 3d. To confirm the evolution (improvement in both the band gap and effective mass) in more detail, the top three Pareto frontiers are plotted for 5 generations, as shown in Fig. 3e. It is worthwhile to focus on the promising area below and to the left of the red lines. The population in these areas becomes denser as the NSGA-II/MOBO iteration progresses. This is clear evidence that the NSGA-II/MOBO optimization algorithm was successful. In particular, the MOBO-nominated entries are specially marked as stars in Fig. 3e, indicating that relatively promising entries were nominated by the MOBO algorithm in later generations. It is clear that the NSGA-II/MOBO hybridization



**Fig. 3 The NSGA-II/MOBO screening result targeting HOIPs for LED applications.** **a** The numbers of HOIP candidates exhibiting  $1.83 < \text{GGA\_Eg} < 2.13 \text{ eV}$  and  $m_{e/h} < m_o$ . **b** The number of promising entries simultaneously fulfilling the above conditions. **c** The band gap and effective mass objective functions averaged from the first Pareto frontier are plotted as a function of the NSGA-II/MOBO round. Note that the band gap objective function does not represent the pristine band gap itself, and the effective mass objective function is the maximum effective mass out of electron and hole masses. **d** The EHVI is plotted as a function of the MOBO round. The increasing trend is indicative of a certain degree of intended evolution. **e** Pareto-sorted entries were collected from the top three Pareto frontiers of each generation for the NSGA-II/MOBO iteration (the MOBO-nominated entries are separately marked as pentacles).

was successful. All entries that appear during the NSGA-II/MOBO iteration are listed in Supplementary Table 1. The density of states (DOS), band structure, and relaxed structure schematics are presented in Supplementary Figs. 1 and 2.

We collected 24 entries that meet the predefined conditions of  $1.83 < E_g < 2.13 \text{ eV}$  and  $m_{e/h} < m_o$ , indicating that they could be promising HOIP candidates for LED applications. We finally identified seven candidates from the 24 entries based on the much stricter criterion of  $1.91 < E_g < 2.04 \text{ eV}$  and  $m_{e/h} < 0.24m_o$  to maintain a balance with the solar cell HOIP search case. These final candidates were validated using HSE06-based DOS and band structure, as well as AIMD. The results from the HSE06-driven DFT and the AIMD calculations are discussed in more detail in a later subsection.

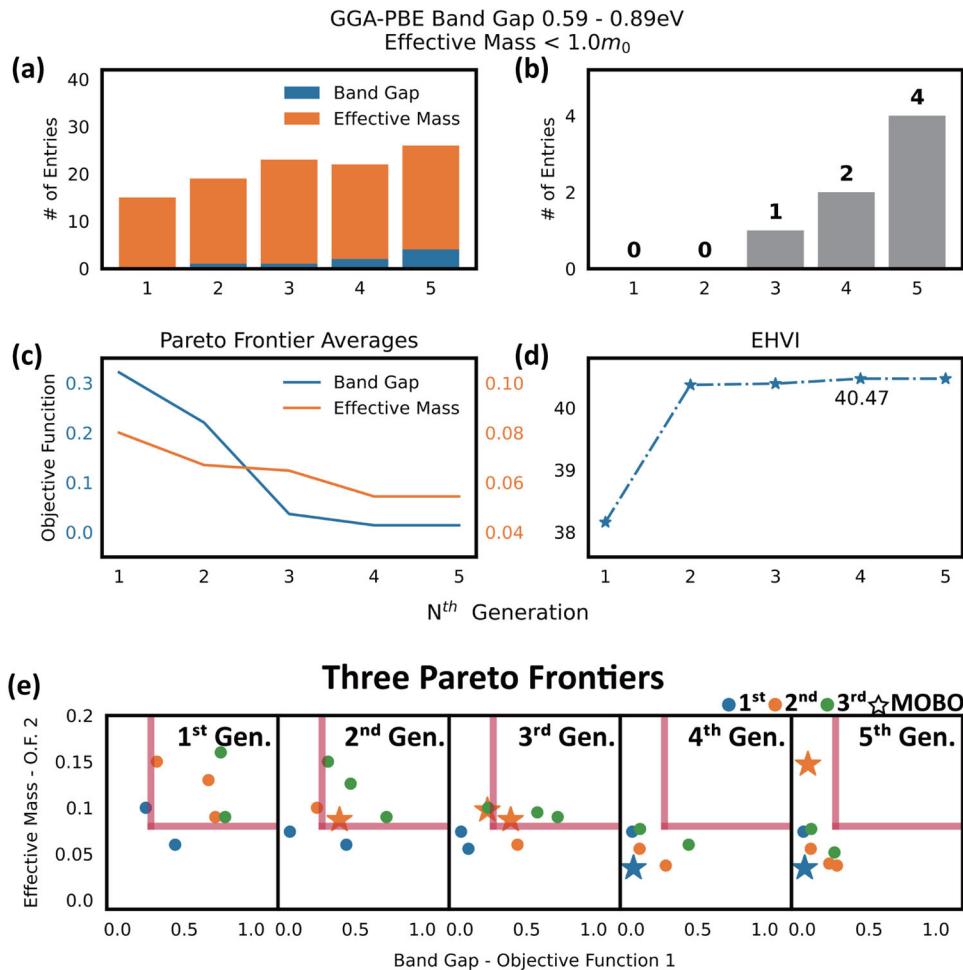
#### Search for 2D HOIPs for use in PV applications

It is generally known that 2D perovskites are not a good choice as absorbers for solar cell applications because of their wide optical bandgap and the limited charge transport associated with the 2D structure<sup>77</sup>. In this regard, a 3D/2D mixed-dimensionality

approach is recommended for PV applications<sup>6,78–82</sup>. Low-dimensional HOIPs and phases with smaller  $n$  values are known to have larger band gaps and exciton binding energies, so RP phase perovskites ( $n = 2$ ) may be more suitable for LED applications<sup>83</sup>. Therefore, the use of RP phase perovskites ( $n = 2$ ) as a template structure would appear to be a challenge if solar cell applications were the ultimate goal. Despite this challenge, however, we aimed to search the 2D RP structure pool for 2D perovskites suitable for solar cell applications. Although it would be difficult to find candidates with reduced band gaps and improved charge mobilities in the RP structure pool, the search effort was worthwhile because the NSGA-II/MOBO approach could facilitate the search process. It should be noted that the adoption of closed-loop optimization strategies based on NSGA-II and MOBO is motivated by this harsh and seemingly insurmountable challenge. This type of data-driven material discovery strategy could achieve unconventional results beyond the traditional knowledge-based approach.

Evidence for premeditated evolution was also clearly detected in solar cell application-targeted NSGA-II/MOBO implementation. Analogous to the previous case of LED application, NSGA-II/MOBO

## Solar Cell



**Fig. 4** NSGA-II/MOBO screening results targeting HOIPs for solar cell applications. **a** The numbers of HOIP candidates exhibiting  $0.61 < \text{GGA\_Eg} < 0.87 \text{ eV}$  and  $m_{e/h} < m_0$ . **b** The number of promising entries simultaneously fulfilling the above conditions. **c** The band gap and effective mass objective functions averaged from the first Pareto frontier are plotted as a function of the NSGA-II/MOBO round. Note that the band gap objective function does not represent the pristine band gap itself, and the effective mass objective function is the maximum effective mass out of electron and hole masses. **d** The EHVI is plotted as a function of the MOBO round. **e** Pareto-sorted entries were collected from the top three Pareto frontiers of each generation for the NSGA-II/MOBO iteration (the MOBO nominated entries are separately marked as pentacles).

drove the nominated HOIPs to minimize both the objective functions,  $E_g\text{-PV}$  and  $m_{e/h}$ . Accordingly, the band gap approached 0.74 eV (the GGA-translated Shockley–Queisser efficiency limit), and the effective mass continued to decrease as the NSGA-II/MOBO iteration progressed. Figure 4a shows the evidence for the evolution in terms of the number of promising entries with band gaps between 0.61 and 0.87 eV and effective masses less than  $m_0$ . The number of entries that meet the band gap requirement was lower than the number obtained from the LED-targeted HOIP search (Fig. 3a), although it showed an increasing trend as the NSGA-II/MOBO iteration progressed. Figure 4b also shows the number of promising entries that fulfill both requirements simultaneously. As a result, seven entries remain that fulfill  $0.61 < \text{GGA\_Eg} < 0.87 \text{ eV}$  and  $m_{e/h} < 0.24m_0$ . The number of promising entries discovered in the solar cell-targeted HOIP search was less than the number discovered in the LED-targeted HOIP search, where NSGA-II/MOBO nominated 24 promising HOIP entries (Fig. 3b). This reduced number supports the discipline-specific common knowledge that RP phase perovskites ( $n = 2$ ) are more suitable for LED applications than solar cell applications<sup>83</sup>. It should be noted, however, that the NSGA-II/MOBO strategy

allowed us to find seven candidates even in the RP phase perovskite pool, which is promising evidence that the NSGA-II/MOBO strategy was successful. However, the experimental realization of these hypothetical candidates has yet to be achieved, since the aim of the present investigation was limited to computational suggestions. The suggested HOIPs would be a good starting point for experimental investigators.

The average objective function values for the first Pareto frontier are also plotted as a function of the generation number in Fig. 4c, with both  $E_g\text{-PV}$  and  $m_{e/h}$  decreasing as the NSGA-II/MOBO iteration progresses. The EHVI is also plotted as a function of the generation number, with a clearly increasing shown in Fig. 4d. The top three Pareto frontiers are plotted for five generations, as shown in Fig. 4e. It is worthwhile to focus on the area below and to the left of the red lines. Analogous to the previous case of LED applications, the population in these areas became denser as the NSGA-II/MOBO iteration progressed. This is clear evidence that the NSGA-II/MOBO was successful for the solar-cell-targeted HOIP search. The MOBO-nominated entries are specially marked as stars in Fig. 4e, indicating that relatively promising entries were nominated by the MOBO algorithm in later generations. All entries that appear during the

**Table 1.** Details for the 14 hypothetical HOIPs.

Chemical formula	$E_g^{\text{GGA}}$ (eV)	$E_g^{\text{HSE06}}$ (eV)	$m_h^*$	$m_e^*$	D/I	Targeted application
(CH <sub>3</sub> (CH <sub>2</sub> ) <sub>4</sub> NH <sub>3</sub> ) <sub>2</sub> C <sub>4</sub> H <sub>7</sub> NH <sub>3</sub> Sn <sub>2</sub> Cl <sub>7</sub>	2.02	3.02	0.24	0.03	D	LED
(C <sub>6</sub> H <sub>5</sub> CH <sub>2</sub> NH <sub>3</sub> ) <sub>2</sub> C <sub>5</sub> NH <sub>12</sub> Sn <sub>2</sub> Cl <sub>7</sub>	1.95	2.8	0.24	0.04	D	LED
(CH <sub>3</sub> (CH <sub>2</sub> ) <sub>2</sub> NH <sub>3</sub> ) <sub>2</sub> NH <sub>3</sub> OHSn <sub>2</sub> Cl <sub>7</sub>	2.04	2.85	0.17	0.02	D	LED
(OH(CH <sub>2</sub> ) <sub>4</sub> NH <sub>3</sub> ) <sub>2</sub> Br(CH <sub>2</sub> ) <sub>2</sub> NH <sub>3</sub> Sn <sub>2</sub> Cl <sub>7</sub>	2.03	2.81	0.21	0.02	D	LED
(C <sub>7</sub> N <sub>2</sub> H <sub>7</sub> ) <sub>2</sub> C <sub>5</sub> OH <sub>11</sub> Sn <sub>2</sub> Br <sub>7</sub>	1.96	2.7	0.12	0.05	D	LED
(COOHC <sub>6</sub> H <sub>10</sub> CH <sub>2</sub> NH <sub>3</sub> ) <sub>2</sub> C <sub>3</sub> OH <sub>7</sub> Sn <sub>2</sub> Cl <sub>7</sub>	1.91	2.71	0.17	0.1	D	LED
(I(CH <sub>2</sub> ) <sub>4</sub> NH <sub>3</sub> ) <sub>2</sub> F(CH <sub>2</sub> ) <sub>2</sub> NH <sub>3</sub> Sn <sub>2</sub> Cl <sub>7</sub>	1.93	2.75	0.19	0.05	D	LED
(OH(CH <sub>2</sub> ) <sub>4</sub> NH <sub>3</sub> ) <sub>2</sub> C <sub>3</sub> N <sub>2</sub> H <sub>5</sub> Ge <sub>2</sub> I <sub>7</sub>	0.68	1.25	0.06	0.04	D	PV
(CH <sub>3</sub> CH <sub>2</sub> NH <sub>3</sub> ) <sub>2</sub> C <sub>4</sub> H <sub>7</sub> NH <sub>3</sub> Ge <sub>2</sub> I <sub>7</sub>	0.67	1.28	0.08	0.07	D	PV
((CH <sub>3</sub> ) <sub>2</sub> CHCH <sub>2</sub> NH <sub>3</sub> ) <sub>2</sub> OH(CH <sub>2</sub> ) <sub>2</sub> NH <sub>3</sub> Sn <sub>2</sub> I <sub>7</sub>	0.76	1.35	0.03	0.02	D	PV
(C <sub>6</sub> SH <sub>3</sub> CH <sub>2</sub> NH <sub>3</sub> ) <sub>2</sub> C <sub>3</sub> N <sub>2</sub> H <sub>5</sub> Sn <sub>2</sub> I <sub>7</sub>	0.87	1.44	0.08	0.02	D	PV
(C <sub>6</sub> H <sub>11</sub> CH <sub>2</sub> NH <sub>3</sub> ) <sub>2</sub> C <sub>5</sub> OH <sub>11</sub> Ge <sub>2</sub> I <sub>7</sub>	0.84	1.51	0.15	0.24	D	PV
((CH <sub>3</sub> ) <sub>2</sub> CHCH <sub>2</sub> NH <sub>3</sub> ) <sub>2</sub> C(NH <sub>2</sub> ) <sub>3</sub> Sn <sub>2</sub> I <sub>7</sub>	0.61	1.2	0.24	0.03	D	PV
(I(CH <sub>2</sub> ) <sub>2</sub> NH <sub>3</sub> ) <sub>2</sub> C <sub>5</sub> NH <sub>12</sub> Sn <sub>2</sub> I <sub>7</sub>	0.78	1.37	0.15	0.05	D	PV

Seven for LED applications and another seven for solar cell applications suggested by the NSGA-II/MOBO screening process.

NSGA-II/MOBO iteration are listed in Supplementary Table 2. The DOS, band structure, and relaxed structure schematics for all entries appearing during the NSGA-II/MOBO iteration are presented in Supplementary Figs. 3 and 4.

### NSGA-II/MOBO-nominated HOIPs

We identified 14 HOIP candidates from the solar cell- and LED-targeted HOIP search processes. Since 24 candidates were identified in the LED-targeted HOIP search, we reduced the number of candidates to seven candidates using a stricter criterion that the band gap range should be between 1.91 and 2.04 eV and the effective mass should be less than 0.24m<sub>o</sub>. As a result, 14 nominated HOIPs in the RP structure pool (seven for each of the PV and LED applications) are listed in Table 1. These nominated HOIP entries were recalculated using the HSE06 exchange correlation functional. Figure 5 shows the HSE06-based band structure, DOS, and schematic structure for the 14 HOIP candidates. The HSE06 band gap fell into realistic value range; 2.68 < E<sub>g</sub> < 2.89 eV for LED applications and 1.19 < E<sub>g</sub> < 1.49 eV for solar cell applications. We did not include SOC for HSE06 calculation. However, we examined the effect of SOC on the calculation. Supplementary Table 3 shows the effect of SOC on the HSE06 calculation result. The band gap values for HSE06 and HSE06-SOC calculations are similar. Although the effective mass differs between the HSE06 and HSE06-SOC calculations, it does not matter because the effective mass for HSE06-SOC is similar to that for the PBEsol-SOC that we used for the entire NSGA-II/MOBO process. It turned out that the SOC does not affect the band gap but greatly affect the effective mass.

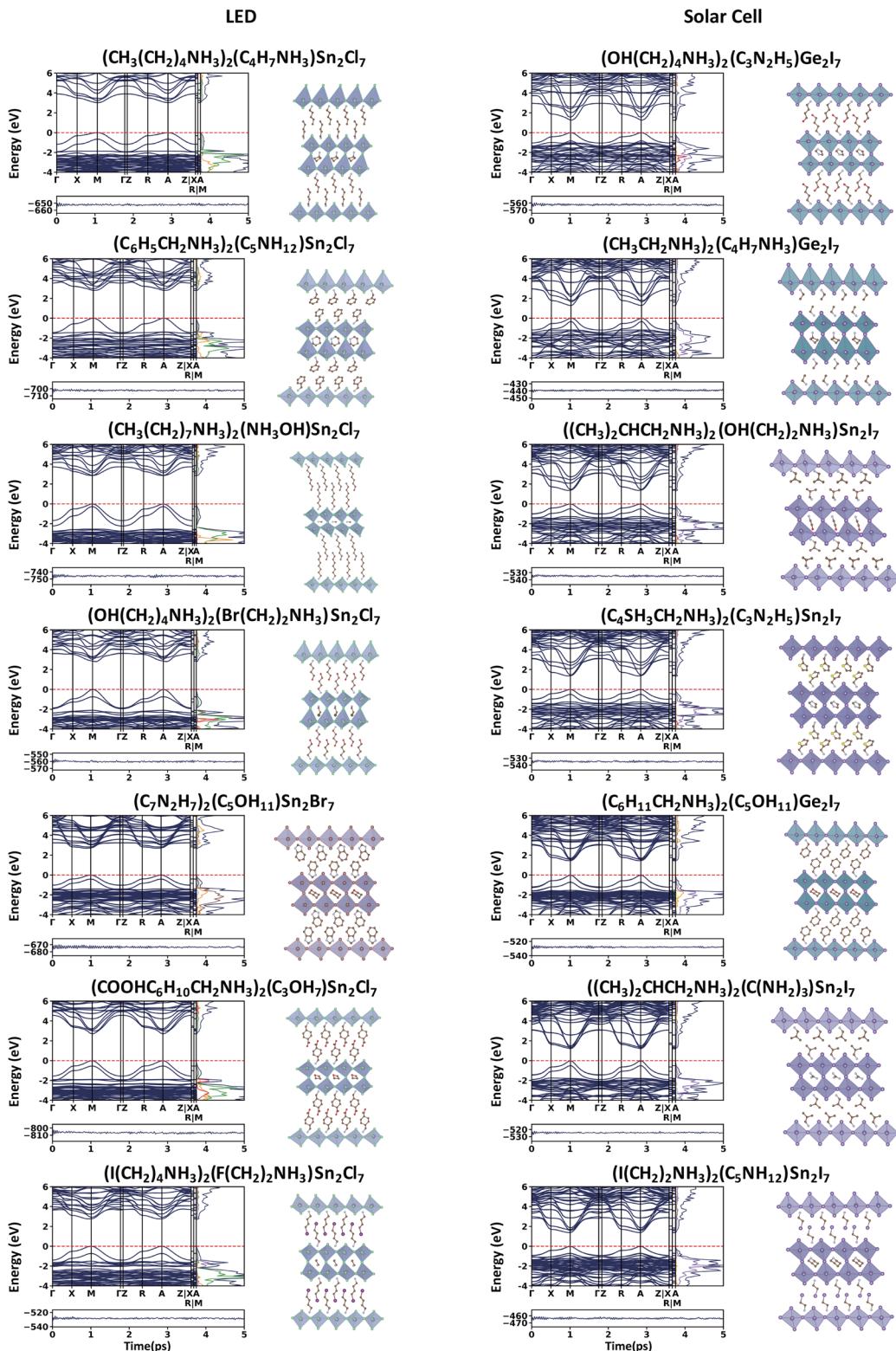
Additionally, AIMD was implemented to test the phase stability, and it was revealed that the AIMD total energy never changed significantly as a function of time for 5 ps for any of the nominated HOIP candidates. However, this does not guarantee the stability of the HOIP candidates. The AIMD result shown in Fig. 5 is a weak indicator of the stability of the selected HOIP candidates. Neither the AIMD total energy nor the DFT-driven thermodynamic energy preference can directly indicate the phase stability. The AIMD just shows the possibility of dynamic stability, and indicates neither formability nor synthesizability. However, these could be a rough indicator of the stability. The thermodynamic energy preference and synthesizability of the nominated candidates should be validated experimentally, which would be the future step in the discovery of novel HOIPs.

Although the GGA-band gap is significantly underestimated, it is not an issue to use it as an objective function since the NSGA-II/MOBO iteration operates based on the relative difference in the objective function rather than the absolute value of the objective function. As long as the relative differences among the GGA-based band gaps were consistent with those among the real values, the optimization would not be affected. More realistic band gap values can be easily obtained by referring to a correction curve. It is well known that the HSE06-based band gap is closer to the true value, and a linear relationship exists between the GGA- and HSE06-based band gaps. Thus, we constructed a correction curve using previous data<sup>23,34</sup>. Figure 6 shows the typical linear relationship between GGA- and HSE06-based band gaps for HOIPs. We superimposed the present 14 HOIP data points on the well-fitted plot. The new data never deviates from the plot.

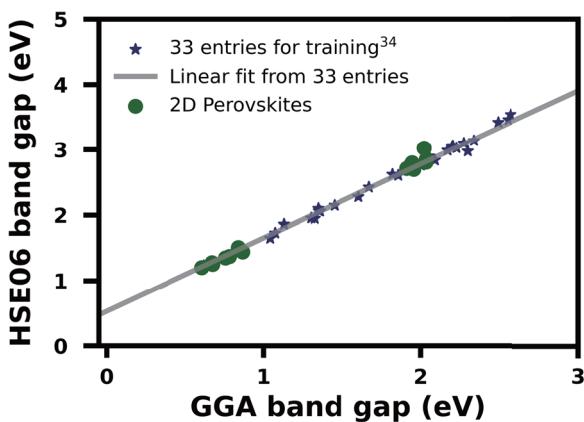
There is no overlap between the 14 NSGA-II/MOBO-nominated HOIP candidates and the well-known existing HOIPs listed in Table 2. Table 2 exhibits only a few HOIPs that meet the selection criterion adopted in the present investigation, as marked as LED and PV in the rightmost column. Six HOIPs are suitable for LED applications and only a HOIP is found useful in PV applications. The lack of PV-targeted HOIPs is rationalized by the commonsense that RP-based 2D perovskites are not suitable for the PV application. The reason for the lack of RP-based 2D perovskites for PV applications were already described above. It should be noted, however, that the seven hypothetical RP-based 2D perovskites for PV applications, which were nominated by the NSGA-II/MOBO strategy, are valuable discoveries in comparison to the scarcity in Table 2, although the synthesizability is yet to be accomplished. The six known HOIPs for use in LED applications are all based on Pb and well-known small molecules for both A' and A-sites. This means that the NSGA-II/MOBO-nominated, Pb-free HOIPs deserve attention.

In summary, we implemented both PV- and LED-targeted HOIP search procedures using a NSGA-II/MOBO hybridized optimization algorithm. The DFT-calculated band gap and effective mass were used as objective functions to be optimized by the NSGA-II/MOBO algorithm, while the choice of molecules and atoms in the RP structure ( $n = 2$ ) were taken as decision variables. We identified 14 promising, Pb-free HOIPs in the RP structure ( $n = 2$ ), seven candidates for each of the PV and LED applications. The nominated LED-targeted HOIPs have a HSE06-based band gap between 2.68–2.89 eV (GGA-translated 1.91–2.04 eV) and an effective mass below 0.24m<sub>o</sub>, while the nominated PV-targeted

## 14 hybrid 2D perovskites



**Fig. 5** Fourteen HOIPs chosen from the NSGA-II/MOBO screening results (seven for LED applications and seven for solar cell applications), including the HSE06-based band structure and DOS results; the band structure and DOS were drawn based on the original Bravais lattice. The AIMD total energy plot as a function of time for 5 ps is also presented on the bottom.



**Fig. 6 Plot of the GGA vs. HSE06 band gap for the 14 HOIP candidates (green circles) along with a linear regression fit from our previous investigation<sup>33</sup>.** The GGA-based band gap can be translated to an HSE06-based calculated version and vice versa by using this linear fit result.

HOIPs have a band gap between 1.19–1.49 eV (GGA-translated 0.61–0.87 eV) and an effective mass below  $0.24m_0$ . The phase stability of the nominated HOIPs was validated using AIMD calculations.

## METHODS

### NSGA-II and MOBO

The elitist-reinforced nondominated sorting genetic algorithm, known as NSGA-II<sup>50–53</sup>, has been widely used for multiobjective optimization problems in many fields. Multiobjective optimization problems deal with objectives that trade off with one another, meaning that while one objective increases, the other decreases. Of course, NSGA-II also works in non-trade-off problems. There is no unique global solution, but rather a set of solutions known as the Pareto frontier. The following mathematical forms represent a typical multiobjective optimization problem:

Minimize/Maximize :  $f_m(\mathbf{X})$ ,  $m = 1, 2, \dots, M$

A solution is a vector of  $n$  decision variables:

$$\mathbf{X} = (x_1, x_2, \dots, x_n)^T \text{ bounded such that } S : x_i^{(L)} \leq x_i \leq x_i^{(U)}, \\ i = 1, 2, \dots, n$$

In our specific problem, HOIP discovery was subjected to no constraints. The set of all feasible solutions belong to the feasible region, or  $S$ . The term “search space” is also used to refer to the feasible region. The objective space consists of the possible values of the  $M$  objective functions for all solutions in  $S$ .  $M$  is usually 2 in NSGA-II. If  $M > 2$ , then the more recently developed NSGA-III would work better<sup>34,84</sup>. NSGA-II enables us to obtain a nondominated set of the entire feasible search space  $S$ : the globally Pareto-optimal set, also called the Pareto frontier. NSGA-II drives the solutions to the Pareto frontier by employing two important principles, elitism and crowding distance, and shares all the other general GA principles, such as crossover, mutation, and tournament selection. NSGA-II uses an elitist principle on a large scale, in contrast to the limited elitism in normal GA, i.e., it performs nondominated sorting in the combined parent and offspring populations, i.e., the solutions are sorted based on the nondominance. NSGA-II also uses an explicit diversity-preserving mechanism (crowding distance) and thereby pursues evenly distributed nondominated solutions.

MOBO<sup>54–56</sup> has scarcely been used in the materials research domain, although single-objective BO has been used for materials research<sup>57–60</sup>. MOBO is still in its infancy, and its application in the materials research domain is rare. We hybridized MOBO and NSGA-II for the purpose of discovering virtual HOIPs in the present investigation and thereby were able to reduce the risk from insecure MOBO. We adopted a state-of-the-art MOBO algorithm that incorporated a parallel q-expected hypervolume improvement acquisition function (qEHVI), enabling efficient and effective gradient-based optimization methods<sup>60</sup>. BO is powerful for optimization tasks that require time-consuming evaluations of objective functions.

**Table 2.** Details for well-known existing HOIPs based on an RP structure with  $n = 2$ .

Formula	$E_g^{\text{Cal.}} (\text{eV})$	$E_g^{\text{Exp.}} (\text{eV})$	Targeted application
(BA) <sub>2</sub> (MA)Pb <sub>2</sub> I <sub>7</sub>	1.98 <sup>a,3</sup> , 1.58 <sup>a,B,5</sup> , 1.25 <sup>a,M,5</sup> , 1.88 <sup>a,14</sup> , 2.48 <sup>b,17</sup>	1.9–2.17 <sup>1–20</sup>	
(BA) <sub>2</sub> (MA)Sn <sub>2</sub> I <sub>7</sub>	0.795 <sup>a,24</sup> , 1.5 <sup>b,23</sup>	1.64 <sup>21–24</sup>	
(BA) <sub>2</sub> (MA)Sn <sub>1.6–8.6</sub> Cl <sub>0.1–14</sub>	0.821 <sup>a,24</sup>	–	
(PEA) <sub>2</sub> (MA)Pb <sub>2</sub> I <sub>7</sub>	–	1.95–2.17 <sup>19,25–35</sup>	
(PEA) <sub>2</sub> (MA)Sn <sub>2</sub> I <sub>7</sub>	–	1.8 <sup>36</sup>	
(IPA) <sub>2</sub> Sn <sub>2</sub> I <sub>7</sub>	–	2.1 <sup>22</sup>	
(AVA) <sub>2</sub> (MA)Pb <sub>2</sub> I <sub>7</sub>	2.05 <sup>b,37</sup>	2.36 <sup>37</sup>	
(BA) <sub>2</sub> (MA)Pb <sub>2</sub> Br <sub>7</sub>	–	2.85 <sup>38–40</sup>	LED
(BA) <sub>2</sub> (FA)Pb <sub>2</sub> Br <sub>7</sub>	–	2.86 <sup>41</sup>	LED
(BA) <sub>2</sub> (FA)Pb <sub>2</sub> I <sub>7</sub>	2.54 <sup>b,17</sup>	2.14–2.16 <sup>17,41</sup>	
(BA) <sub>2</sub> (MA)Sn <sub>2</sub> Br <sub>7</sub>	–	2.19 <sup>41</sup>	
(BA) <sub>2</sub> (MA)Sn <sub>2</sub> I <sub>7</sub>	–	1.64–1.83 <sup>21,41</sup>	
(BA) <sub>2</sub> (FA)Sn <sub>2</sub> I <sub>7</sub>	–	1.78–1.84 <sup>21,43</sup>	
(BA) <sub>2</sub> (MA)Pb <sub>2</sub> Cl <sub>7</sub>	–	3.57 <sup>42</sup>	
(BA) <sub>2</sub> (FA)Pb <sub>2</sub> Cl <sub>7</sub>	–	3.5 <sup>42</sup>	
(BA) <sub>2</sub> (GA)Pb <sub>2</sub> Cl <sub>7</sub>	2.55 <sup>b,17</sup>	2.17 <sup>17</sup>	
(BA) <sub>2</sub> CsPb <sub>2</sub> Br <sub>7</sub>	1.43 <sup>a,44</sup>	2.61–2.89 <sup>42,44</sup>	LED
(PA) <sub>2</sub> (MA)Pb <sub>2</sub> I <sub>7</sub>	–	2.12–2.18 <sup>45–48</sup>	
(HA) <sub>2</sub> (MA)Pb <sub>2</sub> I <sub>7</sub>	1.01 <sup>a,40</sup>	2.1–2.16 <sup>40,46,48–50</sup>	
(HA) <sub>2</sub> CsPb <sub>2</sub> I <sub>7</sub>	–	2.22 <sup>50</sup>	
(HA) <sub>2</sub> (AA)Pb <sub>2</sub> I <sub>7</sub>	–	2.2 <sup>50</sup>	
(HA) <sub>2</sub> (FA)Pb <sub>2</sub> I <sub>7</sub>	–	2.13 <sup>50</sup>	
(HA) <sub>2</sub> (DMA)Pb <sub>2</sub> I <sub>7</sub>	–	2.18 <sup>50</sup>	
(BA) <sub>2</sub> (DMA)Pb <sub>2</sub> I <sub>7</sub>	2.60 <sup>b,17</sup>	2.18 <sup>17</sup>	
(HA) <sub>2</sub> (EA)Pb <sub>2</sub> I <sub>7</sub>	–	2.19 <sup>50</sup>	
(HA) <sub>2</sub> (GA)Pb <sub>2</sub> I <sub>7</sub>	–	2.19 <sup>50</sup>	
(PBA) <sub>2</sub> CsPb <sub>2</sub> I <sub>7</sub>	– <sup>51</sup>	– <sup>51</sup>	
(NEA) <sub>2</sub> (FA)Pb <sub>2</sub> Br <sub>7</sub>	–	3.03 <sup>52</sup>	
(EA) <sub>2</sub> (MA)Pb <sub>2</sub> I <sub>7</sub>	–	1.89 <sup>19,47</sup>	
(BA) <sub>2</sub> (MA)Ge <sub>2</sub> I <sub>7</sub>	–	1.83 <sup>53</sup>	
PEA(RbxCs <sub>1-x</sub> )Pb <sub>2</sub> Br <sub>7</sub>	– <sup>54</sup>	– <sup>54</sup>	
(BA) <sub>2</sub> CsAgBiBr <sub>7</sub>	2.38 <sup>a,55</sup>	2.38 <sup>55</sup>	
(HPA) <sub>2</sub> (MA)Pb <sub>2</sub> Br <sub>7</sub>	1.46 <sup>a,40</sup>	2.7 <sup>40</sup>	LED
(OA) <sub>2</sub> (MA)Pb <sub>2</sub> Br <sub>7</sub>	1.5 <sup>a,40</sup> , 2.02 <sup>a,56</sup>	2.05–2.71 <sup>40,56</sup>	LED
(IBA) <sub>2</sub> (MA)Pb <sub>2</sub> I <sub>7</sub>	–	1.98–2.35 <sup>57,58</sup>	
(PEA) <sub>2</sub> (FA)Sn <sub>2</sub> I <sub>7</sub>	– <sup>59</sup>	– <sup>59</sup>	
(PPA) <sub>2</sub> (MA <sub>0.5</sub> FA <sub>0.5</sub> )Pb <sub>2</sub> I <sub>7</sub>	–	2.2 <sup>60</sup>	
(TPMA) <sub>2</sub> (MA)Pb <sub>2</sub> I <sub>7</sub>	–	1.99 <sup>31</sup>	
(IAA) <sub>2</sub> (MA)Pb <sub>2</sub> I <sub>7</sub>	–	2.4 <sup>58</sup>	
(OA) <sub>2</sub> (FA)Sn <sub>2</sub> I <sub>7</sub>	–	1.43 <sup>43</sup>	PV
(DA) <sub>2</sub> (FA)Sn <sub>2</sub> I <sub>7</sub>	–	1.54 <sup>43</sup>	
(ADA) <sub>2</sub> (FA)Pb <sub>2</sub> I <sub>7</sub>	2.42 <sup>b,61</sup>	2.06 <sup>61</sup>	
(ADAM) <sub>2</sub> (FA)Pb <sub>2</sub> I <sub>7</sub>	2.21 <sup>b,61</sup>	2.21 <sup>61</sup>	
(CMA) <sub>2</sub> (MA)Pb <sub>2</sub> I <sub>7</sub>	–	2.13 <sup>62</sup>	
(BZA) <sub>2</sub> (MA)Pb <sub>2</sub> I <sub>7</sub>	–	1.92 <sup>19</sup>	
(IM) <sub>2</sub> (MA)Pb <sub>2</sub> I <sub>7</sub>	–	1.99 <sup>19</sup>	
(OA) <sub>2</sub> (MA)Pb <sub>2</sub> I <sub>7</sub>	0.99 <sup>a,40</sup>	2.16 <sup>40</sup>	
(HA) <sub>2</sub> (MA)Pb <sub>2</sub> Br <sub>7</sub>	1.6 <sup>a,40</sup>	2.74 <sup>40</sup>	LED

The superscript attached to the band gap values stands for supplementary references that are available in the Supplementary Information. Note that the reference numbering differs from that for references in the manuscript. The superscript numbers with no band gap data represent HOIPs that are synthesized with no band gap value reported. The last column indicates entries that fulfill the selection criterion that we adopted for LED and PV applications in the NSGA-II/MOBO screening process.

<sup>a</sup>GGA.

<sup>b</sup>Hybrid functional.

<sup>c</sup>Optical band gap.

<sup>d</sup>Bulk structure.

<sup>e</sup>Monolayer structure.

For instance, our DFT-calculation-driven HOIP discovery task took the band gap and effective mass as objective functions to be optimized. Therefore, at least several hours (sometimes more than a day) were required for the band gap and effective mass calculations for a single HOIP candidate under our computational capacity. The hybridization of MOBO with NSGA-II would be a useful option to reduce the computational burdens for the current optimization task. Rather than a hard-to-handle objective function, an acquisition function that was easy-to-handle, cheap-to-evaluate, and easily optimizable (with exact gradients computed) was used to determine the next samples to be evaluated. It was sufficient to evaluate the objective function of only a limited number of samples.

Although many blindly trust the capabilities of BO, we do not. It is our opinion that BO capability has been slightly exaggerated due to its outstanding merit, namely low-cost optimization with a relatively small number of samples to be evaluated. Even a single-objective BO execution does not always yield an acceptable result when using a simple benchmark function as a single objective function. Moreover, when compared with the benchmark results for single-objective BO execution, a worse performance can be expected for MOBO. Hybridization with the well-known NSGA-II would be a good option to compensate for the shortcomings of MOBO. It should be noted that our hybridization is not a typical use of the NSGA-II or NSGA-III algorithms for acquisition function optimization since qEHVI is differentiable. We introduced  $q = 2$  for the MOBO process. Two independent samples were created for each step by the MOBO algorithm using all the existing samples previously introduced by both NSGA-II and MOBO. We monitored those generated by MOBO separately for comparison with the NSGA-II-recommended samples.

## DFT and AIMD calculations

First-principle calculations were performed using conventional lab-scale computational inventories based on the Vienna Ab initio Simulation Package (VASP5.4)<sup>85–89</sup>. The generalized gradient approximation (GGA) parameterized by Perdew, Burke, and Ernzerhof<sup>85</sup>, along with the van der Waals (vdW) functional (DFT-D3)<sup>90</sup>, was employed for the NSGA-II iterations. Projector-augmented wave potentials<sup>91,92</sup>, a cutoff energy of 500 eV and a k-mesh interval maintained below 0.3 eV were adopted using the Monkhorst-Pack scheme for all input model structures, which ensured that total energy converged at less than  $10^{-5}$  eV. We also incorporated spin-orbit coupling<sup>93</sup> for all entries in the NSGA-II/MOBO iterations. All structural aspects (atomic position, lattice size, and shape) were allowed to relax until the final force on all relaxed atoms was less than 0.01 eV/Å. Hybrid DFT calculations within the Heyd-Scuseria-Ernzerhof (HSE) formalism with a 25% Hartree-Fock exchange<sup>94,95</sup> were employed for some promising candidate compounds elicited from the NSGA-II/MOBO iteration results to perform linear regression fitting for a GGA band gap vs. an HSE06 band gap. To examine the room-temperature structural stability of the selected candidates, AIMD<sup>96,97</sup> calculations were implemented at 300 K within the vdW correction using the DFT-D3 method. The AIMD simulation lasted 5 ps, with a time step of 1 fs, and was based on the canonical ensemble (NVT) and Nosé-Hoover thermostat algorithm.

## DATA AVAILABILITY

The data used for this study are available in Supplementary Information.

## CODE AVAILABILITY

The code used for this study is available at <https://github.com/socoobblue/optimization>.

Received: 8 December 2021; Accepted: 5 April 2022;

Published online: 26 April 2022

## REFERENCES

1. Tsai, H. et al. High-efficiency two-dimensional ruddlesden–popper perovskite solar cells. *Nature* **536**, 312–316 (2016).
2. Hu, J. et al. Synthetic control over orientational degeneracy of spacer cations enhances solar cell efficiency in two-dimensional perovskites. *Nat. Commun.* **10**, 1276 (2019).
3. Gao, P., Bin Mohd Yusoff, A. R. & Nazeeruddin, M. K. Dimensionality engineering of hybrid halide perovskite light absorbers. *Nat. Commun.* **9**, 5028 (2018).
4. Quan, L. N. et al. Ligand-stabilized reduced-dimensionality perovskites. *J. Am. Chem. Soc.* **138**, 2649–2655 (2016).
5. Cao, D. H., Stoumpos, C. C., Farha, O. K., Hupp, J. T. & Kanatzidis, M. G. 2D homologous perovskites as light-absorbing materials for solar cell applications. *J. Am. Chem. Soc.* **137**, 7843–7850 (2015).
6. Grancini, G. et al. One-year stable perovskite solar cells by 2D/3D interface engineering. *Nat. Commun.* **8**, 15684 (2017).
7. Smith, I. C., Hoke, E. T., Solis-Ibarra, D., McGehee, M. D. & Karunadasa, H. I. A layered hybrid perovskite solar-cell absorber with enhanced moisture stability. *Angew. Chem., Int. Ed.* **53**, 11232–11235 (2014).
8. Straus, D. B. & Kagan, C. R. Electrons, excitons, and phonons in two-dimensional hybrid perovskites: connecting structural, optical, and electronic properties. *J. Phys. Chem. Lett.* **9**, 1434–1447 (2018).
9. Mao, L., Stoumpos, C. C. & Kanatzidis, M. G. Two-dimensional hybrid halide perovskites: principles and promises. *J. Am. Chem. Soc.* **141**, 1171–1190 (2019).
10. Chen, Y. et al. 2D Ruddlesden–popper perovskites for optoelectronics. *Adv. Mater.* **30**, 1703487 (2018).
11. Maculan, G. et al.  $\text{CH}_3\text{NH}_3\text{PbCl}_3$  single crystals: inverse temperature crystallization and visible-blind UV-photodetector. *J. Phys. Chem. Lett.* **6**, 3781–3786 (2015).
12. Saouma, F. O., Park, D. Y., Kim, S. H., Jeong, M. S. & Jang, J. I. Multiphoton absorption coefficients of organic–inorganic lead halide perovskites  $\text{CH}_3\text{NH}_3\text{PbX}_3$  ( $X=\text{Cl}, \text{Br}, \text{I}$ ) single crystals. *Chem. Mater.* **29**, 6876–6882 (2017).
13. Zhang, Z. et al. Bandgap narrowing in Bi-doped  $\text{CH}_3\text{NH}_3\text{PbCl}_3$  perovskite single crystals and thin films. *J. Phys. Chem. C* **121**, 17436–17441 (2017).
14. Li, Z. et al. Robot-accelerated perovskite investigation and discovery. *Chem. Mater.* **32**, 5650–5663 (2020).
15. Zhong, M. et al. Accelerated discovery of  $\text{CO}_2$  electrocatalysts using active machine learning. *Nature* **581**, 178–183 (2020).
16. Attia, P. M. et al. Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature* **578**, 397–402 (2020).
17. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
18. Jain, A. et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
19. Pilania, G. et al. Machine learning bandgaps of double perovskites. *Sci. Rep.* **6**, 19375 (2016).
20. Pilania, G., Gubernatis, J. E. & Lookman, T. Multi-fidelity machine learning models for accurate bandgap predictions of solids. *Comput. Mater. Sci.* **129**, 156–163 (2017).
21. Chen, C. et al. A critical review of machine learning of energy materials. *Adv. Energy Mater.* **10**, 1903242 (2020).
22. Nakajima, T. & Sawada, K. Discovery of Pb-free perovskite solar cells via high-throughput simulation on the K computer. *J. Phys. Chem. Lett.* **8**, 4826–4831 (2017).
23. Li, Y. & Yang, K. High-throughput computational design of organic–inorganic hybrid halide semiconductors beyond perovskites for optoelectronics. *Energy Environ. Sci.* **12**, 2233–2243 (2019).
24. Lu, S. et al. Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nat. Commun.* **9**, 3405 (2018).
25. Li, Z., Xu, Q., Sun, Q., Hou, Z. & Yin, W.-J. Thermodynamic stability landscape of halide double perovskites via high-throughput computing and machine learning. *Adv. Funct. Mater.* **29**, 1807280 (2019).
26. Schmidt, J. et al. Predicting the thermodynamic stability of solids combining density functional theory and machine learning. *Chem. Mater.* **29**, 5090–5103 (2017).
27. Im, J. et al. Identifying Pb-free perovskites for solar cells by machine learning. *npj Comput. Mater.* **5**, 37 (2019).
28. Jain, D., Chaube, S., Khullar, P., Srinivasan, S. G. & Rai, B. Bulk and surface DFT investigations of inorganic halide perovskites screened using machine learning and materials property databases. *Phys. Chem. Chem. Phys.* **21**, 19423–19436 (2019).
29. Agiorgousis, M. L., Sun, Y.-Y., Choe, D.-H., West, D. & Zhang, S. Machine learning augmented discovery of chalcogenide double perovskites for photovoltaics. *Adv. Theory Simul.* **2**, 1800173 (2019).
30. Balachandran, P. V. et al. Predictions of new  $\text{ABO}_3$  perovskite compounds by combining machine learning and density functional theory. *Phys. Rev. Mater.* **2**, 043802 (2018).
31. Balachandran, P. V., Kowalski, B., Sehirlioglu, A. & Lookman, T. Experimental search for high-temperature ferroelectric perovskites guided by two-step machine learning. *Nat. Commun.* **9**, 1668 (2018).
32. Bartel, C. J. et al. New tolerance factor to predict the stability of perovskite oxides and halides. *Sci. Adv.* **5**, 2 (2019).
33. Li, W., Jacobs, R. & Morgan, D. Predicting the thermodynamic stability of perovskite oxides using machine learning models. *Comput. Mater. Sci.* **150**, 454–463 (2018).

34. Lee, B. D. et al. Discovery of lead-free hybrid organic/inorganic perovskites using metaheuristic-driven DFT calculations. *Chem. Mater.* **33**, 782–798 (2021).
35. Bianchi, L., Dorigo, M., Gambardella, L. M. & Gutjahr, W. J. A survey on metaheuristics for stochastic combinatorial optimization. *Nat. Comput.* **8**, 239–287 (2009).
36. Blum, C. & Roli, A. Metaheuristics in combinatorial optimization: overview and conceptual comparison. *ACM Comput. Surv.* **35**, 268–308 (2003).
37. Mockus, J. *Bayesian Approach to Global Optimization: Theory and Applications* (Kluwer Academic Publishers, 2013).
38. Sohn, K.-S., Lee, J. M. & Shin, N. A search for new red phosphors using a computational evolutionary optimization process. *Adv. Mater.* **15**, 2081–2084 (2003).
39. Sharma, A. K., Son, K. H., Han, B. Y. & Sohn, K.-S. Simultaneous optimization of luminance and color chromaticity of phosphors using a nondominated sorting genetic algorithm. *Adv. Funct. Mater.* **20**, 1750–1755 (2010).
40. Park, W. B., Shin, N., Hong, K.-P., Pyo, M. & Sohn, K.-S. A new paradigm for materials discovery: heuristics-assisted combinatorial chemistry involving parameterization of material novelty. *Adv. Funct. Mater.* **22**, 2258–2266 (2012).
41. Park, W. B., Singh, S. P. & Sohn, K.-S. Discovery of a phosphor for light emitting diode applications and its structural determination,  $\text{Ba}(\text{Si},\text{Al})_5(\text{O},\text{N})_8:\text{Eu}^{2+}$ . *J. Am. Chem. Soc.* **136**, 2363–2373 (2014).
42. Jung, Y. H., Park, W. B., Pyo, M., Sohn, K.-S. & Ahn, D. A multi-element doping design for a high-performance  $\text{LiMnPO}_4$  cathode via metaheuristic computation. *J. Mater. Chem. A* **5**, 8939–8945 (2017).
43. Lee, H.-D. et al. Efficient ruddlesden–popper perovskite light-emitting diodes with randomly oriented nanocrystals. *Adv. Funct. Mater.* **29**, 1901225 (2019).
44. Zheng, Y. et al. Unique characteristics of 2D ruddlesden–popper (2DRP) perovskite for future photovoltaic application. *J. Mater. Chem. A* **7**, 13860–13872 (2019).
45. Yu, S. et al. Nonconfinement structure revealed in Dion–Jacobson type quasi-2D perovskite expedites interlayer charge transport. *Small* **15**, 1905081 (2019).
46. Mao, L. et al. Hybrid Dion–Jacobson 2D lead iodide perovskites. *J. Am. Chem. Soc.* **140**, 3775–3783 (2018).
47. Li, Y. et al. Bifunctional organic spacers for formamidinium-based hybrid Dion–Jacobson two-dimensional perovskite solar cells. *Nano Lett.* **19**, 150–157 (2019).
48. Zhang, Y. et al. Dynamical transformation of two-dimensional perovskites with alternating cations in the interlayer space for high-performance photovoltaics. *J. Am. Chem. Soc.* **141**, 2684–2694 (2019).
49. Soe, C. M. M. et al. New type of 2D perovskites with alternating cations in the interlayer space,  $(\text{C}(\text{NH}_2)_3)(\text{CH}_3\text{NH}_3)_n\text{Pb}_n\text{I}_{3n+1}$ : structure, properties, and photovoltaic performance. *J. Am. Chem. Soc.* **139**, 16297–16309 (2017).
50. Deb, K. & Jain, H. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach. part I: solving problems with box constraints. *IEEE Trans. Evol. Comput.* **18**, 577–601 (2014).
51. Jain, H. & Deb, K. An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part II: handling constraints and extending to an adaptive approach. *IEEE Trans. Evol. Comput.* **18**, 602–622 (2014).
52. Srinivas, N. & Deb, K. Multi-objective function optimization using non-dominated sorting genetic algorithms. *Evol. Comput.* **2**, 221–248 (1994).
53. Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6**, 182–197 (2002).
54. Hernández-Lobato, D., Hernández-Lobato, J. M., Shah, A. & Adams R. P. Predictive entropy search for multi-objective Bayesian optimization. Preprint at <https://arxiv.org/abs/1511.05467> (2016).
55. Wada, T. & Hino H. Bayesian optimization for multi-objective optimization and multi-point search. Preprint at <https://arxiv.org/abs/1905.02370> (2019).
56. Daulton, S., Balandat, M. & Bakshy, E. Differentiable expected hypervolume improvement for parallel multi-objective Bayesian optimization. Preprint at <https://arxiv.org/abs/2006.05078> (2020).
57. Herbol, H. C., Hu, W., Frazier, P., Clancy, P. & Poloczek, M. Efficient search of compositional space for hybrid organic–inorganic perovskites via Bayesian optimization. *npj Comput. Mater.* **4**, 51 (2018).
58. Kusne, A. G. et al. On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat. Commun.* **11**, 5966 (2020).
59. Ulissi, Z. W., Medford, A. J., Bligaard, T. & Nørskov, J. K. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. *Nat. Commun.* **8**, 14621 (2017).
60. Shuaihua, J. H. & Persson, K. A. A high-throughput framework for determining adsorption energies on solid surfaces. *npj Comput. Mater.* **3**, 14 (2017).
61. Xia, C. et al. Discovery of lead-free perovskites for high-performance solar cells via machine learning: ultrabroadband absorption, low radiative combination, and enhanced thermal conductivities. *Adv. Sci.* **9**, 2103648 (2021).
62. Tianmin, W. & Jian, W. Global discovery of stable and non-toxic hybrid organic–inorganic perovskites for photovoltaic systems by combining machine learning method with first principle calculations. *Nano Energy* **66**, 104070 (2019).
63. Shuaihua, L. et al. Accelerated discovery of stable lead-free hybrid organic–inorganic perovskites via machine learning. *Nat. Commun.* **9**, 3405 (2018).
64. Perdew, J. P. & Levy, M. Physical content of the exact Kohn-Sham orbital energies: band gaps and derivative discontinuities. *Phys. Rev. Lett.* **51**, 1884–1887 (1983).
65. Sham, L. J. & Schlüter, M. Density-functional theory of the energy gap. *Phys. Rev. Lett.* **51**, 1888–1891 (1983).
66. Crowley, J. M., Tahir-Kheli, J. & Goddard, W. A. Resolution of the band gap prediction problem for materials design. *J. Phys. Chem. Lett.* **7**, 1198–1203 (2016).
67. Shockley, W. & Queisser, H. J. Detailed balance limit of efficiency of p-n junction solar cells. *J. Appl. Phys.* **32**, 510–519 (1961).
68. Zhao, B. et al. High-efficiency perovskite-polymer bulk heterostructure light-emitting diodes. *Nat. Photonics* **12**, 783–789 (2018).
69. Cao, Y. et al. Perovskite light-emitting diodes based on spontaneously formed submicrometre-scale structures. *Nature* **562**, 249–253 (2018).
70. Lin, K. et al. Perovskite light-emitting diodes with external quantum efficiency exceeding 20 per cent. *Nature* **562**, 245–248 (2018).
71. Chen, Z. et al. High-performance color-tunable perovskite light emitting devices through structural modulation from bulk to layered film. *Adv. Mater.* **29**, 1603157 (2017).
72. Ni, L. et al. Real-time observation of exciton-phonon coupling dynamics in self-assembled hybrid perovskite quantum wells. *ACS Nano* **11**, 10834–10843 (2017).
73. Karlsson, M. et al. Mixed halide perovskites for spectrally stable and high-efficiency blue light-emitting diodes. *Nat. Commun.* **12**, 361 (2021).
74. Yuan, S. et al. Optimization of low-dimensional components of quasi-2D perovskite films for deep-blue light-emitting diodes. *Adv. Mater.* **31**, 1904319 (2019).
75. Lu, M. et al. Metal halide perovskite light-emitting devices: promising technology for next-generation displays. *Adv. Funct. Mater.* **29**, 1902008 (2019).
76. Liu, X.-K. et al. Metal halide perovskites for light-emitting diodes. *Nat. Mater.* **20**, 10–21 (2021).
77. Zhang, F. et al. Advances in two-dimensional organic–inorganic hybrid perovskites. *Energy Environ. Sci.* **13**, 1154–1186 (2020).
78. Jung, E. H. et al. Efficient, stable and scalable perovskite solar cells using poly(3-hexylthiophene). *Nature* **567**, 511–515 (2019).
79. Kim, D. H. et al. Bimolecular additives improve wide-band-gap perovskites for efficient tandem solar cells with CIGS. *Joule* **3**, 1734–1745 (2019).
80. Chen, P. et al. Progress and perspective in low-dimensional metal halide perovskites for optoelectronic applications. *Sol. RRL* **2**, 1700186 (2018).
81. Tong, J. et al. Carrier lifetimes of  $>1 \mu\text{s}$  in Sn-Pb perovskites enable efficient all-perovskite tandem solar cells. *Science* **364**, 475–479 (2019).
82. Wang, Z. et al. Efficient ambient-air-stable solar cells with 2D–3D heterostructured butylammonium-caesium-formamidinium lead halide perovskites. *Nat. Energy* **2**, 17135 (2017).
83. Muljarov, E. A., Tikhodeev, S. G., Gippius, N. A. & Ishihara, T. Excitons in self-organized semiconductor/insulator superlattices: Pbl-based perovskite compounds. *Phys. Rev. B* **51**, 14370–14378 (1995).
84. Han, W. G., Park, W. B., Singh, S. P., Pyo, M. & Sohn, K.-S. Determination of possible configurations for  $\text{Li}_{0.5}\text{CoO}_2$  delithiated Li-ion battery cathodes via DFT calculations coupled with a multi-objective non-dominated sorting genetic algorithm (NSGA-III). *Phys. Chem. Chem. Phys.* **20**, 26405–26413 (2018).
85. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
86. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **47**, 558–561 (1993).
87. Kresse, G. & Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous–semiconductor transition in germanium. *Phys. Rev. B: Condens. Matter Mater. Phys.* **49**, 14251–14269 (1994).
88. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
89. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B: Condens. Matter Mater. Phys.* **54**, 11169–11186 (1996).
90. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **132**, 154104 (2012).
91. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **50**, 17953–17979 (1994).
92. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **59**, 1758–1775 (1999).
93. Even, J., Pedesseau, L., Jancu, J.-M. & Katan, C. Importance of spin-orbit coupling in hybrid organic/inorganic perovskites for photovoltaic applications. *J. Phys. Chem. Lett.* **4**, 2999–3005 (2013).
94. Heyd, J., Scuseria, G. E. & Ernzerhof, M. Hybrid functionals based on a screened coulomb potential. *J. Chem. Phys.* **118**, 8207–8215 (2003).

95. Krukau, A. V., Vydrov, O. A., Izmaylov, A. F. & Scuseria, G. E. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *J. Chem. Phys.* **125**, 224106 (2006).
96. Li, Z., Kermode, J. R. & De Vita, A. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.* **114**, 096405 (2015).
97. Kresse, G. & Hafner, J. Ab initio molecular dynamics for open-shell transition metals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **48**, 13115–13118 (1993).

## ACKNOWLEDGEMENTS

This research was supported by the Creative Materials Discovery Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT, and Future Planning (2021R1A2C1009144), (2021R1A2C1011642) and (2021M3A7C2089778).

## AUTHOR CONTRIBUTIONS

K.-S.S. conceived the concept for the entire process and directed the computational process. W.B.P. and B.D.L. performed DFT and AIMD calculations. B.D.L. and J.W.L. took part in coding for NSGA-II/MOBO algorithms. W.B.P., M.K., and B.D.L. prepared figures and tables. K.-S.S. and W.B.P. wrote the paper. All authors discussed the results and commented on the manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41524-022-00781-z>.

**Correspondence** and requests for materials should be addressed to Woon Bae Park or Kee-Sun Sohn.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022