ARTICLE   OPEN

Check for updates

Generative design of stable semiconductor materials using deep learning and density functional theory

Edirisuriya M. Dilanga Siriwardane [1,2], Yong Zhao [1], Indika Perera[3] and Jianjun Hu [1] ✉

Semiconductor device technology has greatly developed in complexity since discovering the bipolar transistor. In this work, we developed a computational pipeline to discover stable semiconductors by combining generative adversarial networks (GAN), classifiers, and high-throughput first-principles calculations. We used CubicGAN, a GAN-based algorithm for generating cubic materials and developed a classifier to screen the semiconductors and studied their stability using first principles. We found 12 stable $AA'MH_6$ semiconductors in the F-43m space group including $BaNaRhH_6$, $BaSrZnH_6$, $BaCsAlH_6$, $SrTlIrH_6$, $KNaNiH_6$, $NaYRuH_6$, $CsKSiH_6$, $CaScMnH_6$, $YZnMnH_6$, $NaZrMnH_6$, $AgZrMnH_6$, and $ScZnMnH_6$. Previous research reported that five $AA'IrH6$ semiconductors with the same space group were synthesized. Our research shows that $AA'MnH_6$ and $NaYRuH_6$ semiconductors have considerably different properties compared to the rest of the $AA'MH_6$ semiconductors. Based on the accurate hybrid functional calculations, $AA'MH_6$ semiconductors are found to be wide-bandgap semiconductors. Moreover, $BaSrZnH_6$ and $KNaNiH_6$ are direct-bandgap semiconductors, whereas others exhibit indirect bandgaps.

npj Computational Materials (2022)8:164 ; https://doi.org/10.1038/s41524-022-00850-3

INTRODUCTION

Semiconductors are essential components of modern devices that use transistors, light-emitting diodes[1], integrated circuits[2], photovoltaic[3], solar cells[4], and so on[5–7]. Semiconductors exhibit variable resistance since electron flow can be controlled by light and heat. Therefore, these materials can be used for energy conversion, and digital switching[8]. The elemental semiconductors found from Group XIV in the periodic table, like Si and Ge, and the compounds of Ge are widely used in electronics, photovoltaic and optoelectronic devices. However, semiconductors with various properties are required for industrial applications[8,9]. For instance, good thermal conductivity and electric field breakdown strength, and also wide bandgap of SiC semiconductor make it a suitable material for high-temperature, high-power, high-frequency, and high-radiation conditions[10]. Thus, computational approaches for exploring semiconductors are essential to enhance future technologies. High-throughput screening with the aid of first-principles calculations was performed by several groups to discover optoelectronic semiconductors. Setyawan et al. and Ortiz et al. reported the high-throughput screening and data-mining frameworks to investigate bandgap materials for radiation detection[11–13]. High throughput material screening by Zhao et al. found that Cu-In-based Halide Perovskite as potential photovoltaic solar absorbers[13,14]. Based on 4507 hypothetical materials, Li et al. suggest 23 candidates for light-emitting applications, and 13 potential compounds for solar cell technologies[13,15]. Such examples indicate that high-throughput screening can now be used to explore promising semiconductor materials.

Generative adversarial networks (GANs) are a kind of generative models that learn patterns/distribution from input data[16]. GANs use two sub-models to train a generative model. The generator model generates fake data, and the discriminator model learns to tell fake data from real data. The two sub-models are trained simultaneously to achieve a Nash Equilibrium: the generator can generate data that the discriminator can recognize half the chance. Wasserstein distance[17] and gradient penalty[18] are introduced during training in order to overcome mode collapse and improve the training stability in original GANs[16]. There are a limited number of works that leverage GANs to generate crystal structures in material science. The reasons behind that are: 1) Crystal structures have so many formations, such as a different number of elements and number of atoms in a unit cell. It is hard to come up with a unified representation to make GANs learn from them like images or text; 2) GANs used in computer vision cannot generate crystal structures that satisfy physics or symmetric constraints. For instance, GANs easily generate materials that are not recognizable or that have crowd atoms in a unit cell. CrystalGAN[19] is believed to be the first work that uses GANs to generate materials. It applies CyClyGAN[20] to simple systems mapping ternary a hydride into another. In[21], Kim et al. use WGAN-GP[18] to train a generative model to generate Mg-Mn-O systems with atom coordinates as the input. All the works above only consider a simple or specific family of materials at a limited scale. CubicGAN proposed by Zhao et al.[22], however, is the first work that generates materials at a large scale.

In this research, we developed a binary classifier to filter the semiconductors/Insulators (nonmetals) from the dynamically stable quaternary Cubic materials discovered using the CubicGAN model, where high-throughput calculations were done with the assistance of a GAN model and density functional theory (DFT). We studied the most important elemental and electronic properties, which are helpful to distinguish the nonmetals and metals using the machine learning models. In addition, we carried out DFT calculations for those semiconductors to corroborate the thermodynamic stability and semiconductor properties. As a result, we find that 12 cubic semiconductors of a particular class of materials, which we label as $AA'MH_6$, are thermodynamically stable against their competing phases. We further performed the DFT calculations to study their structural, mechanical, thermodynamic, and electronic properties. Our results show that $AA'MnH_6$ and

[1]Department of Computer Science and Engineering, University of South Carolina, 550 Assembly St., Columbia 29201 SC, USA. [2]Department of Physics, University of Colombo, Colombo 3, Sri Lanka. [3]Department of Computer Science and Engineering, University of Moratuwa, Moratuwa, Sri Lanka. ✉email: jianjunh@cse.sc.edu

NaYRuH$_6$ have higher C_{ii} (i = 1, 2, 3) elastic constants, bulk modulus, shear modulus, and Young's modulus compared to the respective mechanical properties of the rest of the AA'MH$_6$ materials. At temperatures less than 200 K, AA'MnH$_6$ and NaYRuH$_6$ have lower specific thermal capacity (C_v) relative to other AA'MH$_6$ materials. The highest C_v at 300 K found in this work is from BaSrZnH$_6$ (127.96 JK^{-1}mol^{-1}). Moreover, hybrid functional calculations show that all AA'MH$_6$ materials are wide-bandgap semiconductors, which will be useful to develop optical and high-temperature power devices[23,24].

RESULTS AND DISCUSSION

Dataset of nonmetals and metals

As the CubicGAN model generates only ternary and quaternary materials, we first analyzed the number of nonmetals (semiconductors and insulators), and metals in the material project (MP) database[25], as shown in Table 1. We collected all the ternary and quaternary materials, where the bandgap details are available, using the Pymatgen code[26]. It could be found that ≈44 % of the ternary materials are nonzero bandgap materials while ≈56 % are metals. However, ≈73 % of the quaternary materials are semiconductors or insulators, whereas only ≈27 % of them are metals. This indicates that the probability of finding a stable quaternary material with a nonzero bandgap is higher compared to finding that in a ternary material set. We also compared the same details of the cubic materials. Interestingly, ≈80 % of the cubic ternary materials are metals, and only ≈20 % of them are nonmetals. On the contrary, the quaternary cubic materials have 30 % more nonzero bandgap materials than the number of metals. It shows that there is a low probability of discovering a nonzero bandgap cubic ternary compound. Instead, in this project, we mainly focused on the quaternary cubic materials for finding stable semiconductors. In this way, by reducing the search space of the materials, we can shorten the computational time taken by the DFT calculations.

Feature importance

Understanding which features are significant during the classification will be vital for discovering semiconductors. In Section 2.1, we could show that quaternary materials have a higher percentage of semiconductors compared to the ternary materials. Next, we analyzed which features have higher importance than others for classifying a quaternary material as metal or nonmetal. Feature importance (FI) of random forest algorithm is defined as the mean of the impurity decrease within each tree. This built-in feature of the random forest makes it convenient and a widely used method to calculate FI. Here, we trained our RFC model for the whole quaternary materials data set. The classification report of this model is in Supplementary Information. Even though both Avg. and the maximum difference of each atomic/electronic property were considered for the RFC model, only three features related to maximum difference have FI greater than 1 %. This indicates that Avg. value of the properties plays a significant role when classifying a material as metal or nonmetal. The top features of FI $\gtrsim$ 2.0% are mentioned in Fig. 1. Avg. Availability of metallic elements has the highest FI, while Avg. availability of nonmetal also has a FI of around 2 %. This indicates that having a metallic or

nonmetallic element is important for the material to be a metal or a semiconductor/insulator. It is generally accepted that metallic elements have a higher boiling point and higher density compared to that of nonmetals. It should be noted that the elemental properties like metallicity, being semiconductor/insulator, density, and boiling point are properties of the bulk material formed with a given element. Since the availability of metallic and nonmetallic elements plays a significant role, the boiling point and density of those elements also can become important features when classifying metals and nonmetals. It is also clear that electronic properties like Avg. number of unfilled orbitals, Avg. number of p-valence electrons, and Avg. availability of +2 and +3 oxidation states have high FI.

We also studied the descriptors to understand how the number of metals and nonmetals depends on the percentage availability of the metal (P_M), nonmetal (P_{NM}) and transition-metal (P_{TM}) elements in the chemical formulas. We use M, NM, and TM to indicate the type of elements to avoid confusion between material class (metal or nonmetal) and element type (metal, nonmetal, transition metal). Figure 2 shows the violin plots with all the 39024 quaternary materials against those three atomic properties. Here, $P_M = 100\%$, $P_{NM} = 100\%$, and $P_{TM} = 100\%$ for a given chemical formula when all the elements are M, NM, and TM, respectively. Figure 2(a) clearly evidences that nonmetals dominate until $P_M \approx 60$ %. The ratio between amounts of metals and nonmetals (metals : nonmetals) is around 1:3 at $P_M < 60$ %. This becomes approximately 5:1 after 60 %, showing the probability of finding a semiconductor/insulator decreases. On the contrary, Fig. 2(b) shows the opposite behavior of metals and nonmetals, while P_{NM} alters. Moreover, it is clear that semiconductors and insulators prefer a lower number of TM elements relative to the other element types. At $P_{TM} > 30$ %, number of metals become significant compared to that of nonmetals. When $P_{TM} \leq 5$ %, metals : nonmetals ratio is 1:6.
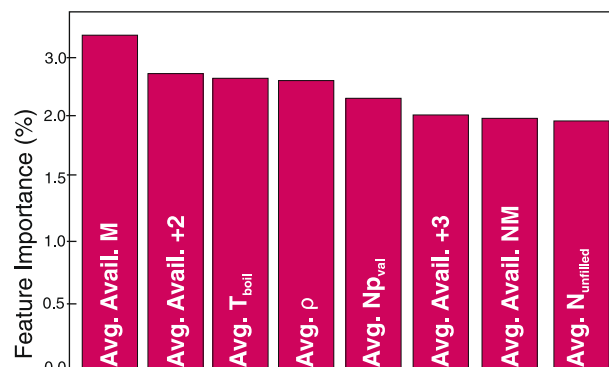

Fig. 1 Feature importance as a percentage ($\gtrsim$ 2.0%) from the RFC model. Labels on the x-axis: metal atom (M), +2 oxidation state (+2), boiling point (T$_{boil}$), atomic density (ρ), number of p-valence electrons (Np$_{val}$), +3 oxidation state (+3), nonmetal atom (NM), number of unfilled valence orbitals (N$_{unfilled}$). Avg. and Avail. stands for weighted average and availability, respectively.

Table 1. The total number (*N*) and percentage of ternary and quaternary nonmetals (NM) and metals (M) for all the types of Bravais lattice and also for Cubic structures (Cubic-NM and Cubic-M) in the Material project database[25].

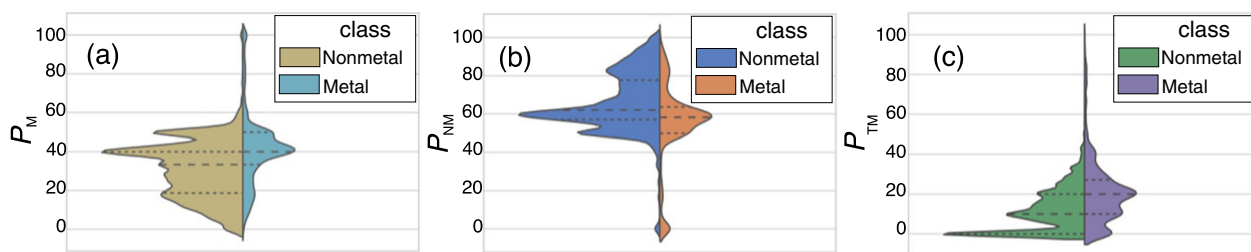| Material	N_{NM}	N_M	NM%	M%	$N_{Cubic-NM}$	$N_{Cubic-M}$	Cubic-NM%	Cubic-M%
Ternary	28102	35682	44.06	55.94	2094	8385	19.98	80.02
Quaternary	28527	10497	73.10	26.90	2578	1438	64.19	35.81

Fig. 2 Number of materials in each material class (metal or nonmetal) as a function of percentage availability (*P*) of element type in the chemical formula. Here, we considered (**a**) metal:M, **b** nonmetal:NM, and **c** transition-metal:TM elements.
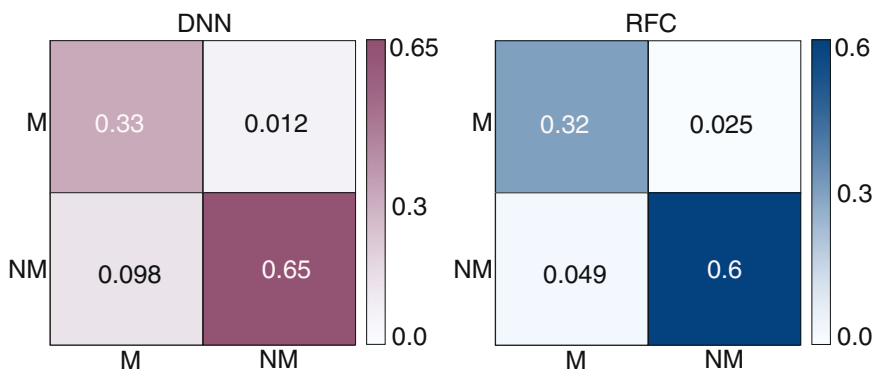

Fig. 3 Normalized confusion matrices of the classifiers for metals (M) and nonmetals (NM). The confusion matrix for deep neural network (DNN) is in purple, while that of random forest (RFC) is in blue.

Table 2. The classification report of the test set for the nonmetal/metal classifiers.

	DNN			RFC			
	Precision	Recall	F1_score	Precision	Recall	F1_score	Support
Metal	0.76	0.93	0.84	0.84	0.93	0.88	28
Nonmetal	0.96	0.85	0.90	0.96	0.91	0.93	53
Accuracy			0.88			0.91	81
Macro avg	0.86	0.89	0.87	0.90	0.92	0.91	81
Weighted Avg	0.89	0.88	0.88	0.92	0.91	0.91	81

Predicting Semiconductors

We further analyzed the error of the DNN and RFC models trained with quaternary cubic materials data. The 10-fold cross-validation accuracy results for each training step of the DNN model are 0.86, 0.92, 0.91, 0.97, 0.94, 0.94, 0.88, 0.94, 0.94, 0.86. Those of the RFC model are 0.86, 0.89, 0.88, 0.9, 0.87, 0.88, 0.90, 0.90, and 0.88. Thus, the mean accuracy was obtained for the DNN (RFC) model as 0.92 ± 0.034 (0.88 ± 0.013). Figure 3 shows the normalized confusion matrices for the classifiers. It is apparent that 33 (32) % of the instances were classified as true metals while 65 (60) % of the materials were listed as true nonmetals by the DNN (RFC) classifier. The percentages of false metals and false nonmetals from the DNN (RFC) model were 9.8 (4.9) % and 1.2 (2.5) %, respectively. The classification report for the model is shown in Table 2. It is clear that the DNN (RFC) classifier predicts whether a quaternary material is a metal or nonmetal with 0.88 (0.91) accuracy. Precision is the matrix that compares the number of true positive instances with the number of predicted positive instances. In our work, the DNN (RFC) model classifies a material as a nonmetal with 0.76 (0.96) and metal with 0.76 (0.84) precision. The recall is a measure of the number of correctly predicted positive cases compared to the total number of positive cases in the dataset. Table 2 shows

that there is 0.85 (0.91) recall for nonmetal, while there is 0.93 (0.93) recall for metals from the DNN (RFC) model. By combining precision and recall, F1-score can be calculated as 0.90 (0.93) for nonmetal and 0.84 (0.88) for metal classes. Therefore, the predictions of semiconductors/insulators from our DNN and RFC models can be expected to be highly accurate.

As seen in Table 2, the RFC model exhibits a slight improvement over the DNN model. To show the methodology of finding stable semiconductors based on generative adversarial networks, we applied our RFC classifier on CubicGAN predicted mechanically and dynamically stable quaternary materials. Out of 323 quaternary materials predicted by the CubicGAN model, 137 compounds were classified as nonmetals.

Structure and thermodynamic stability

We carried out our DFT calculations on those nonmetals to find thermodynamically stable semiconductors. We discovered that 12 semiconductors, which have chemical formulas in the form of $AA'MH_6$, exhibit zero energy-above-hull against the respective competing phases. Those are $BaNaRhH_6$, $BaSrZnH_6$, $BaCsAlH_6$, $SrTlIrH_6$, $KNaNiH_6$, $NaYRuH_6$, $CsKSiH_6$, $CaScMnH_6$, $YZnMnH_6$, $NaZrMnH_6$, $AgZrMnH_6$, and $ScZnMnH_6$. We also find that
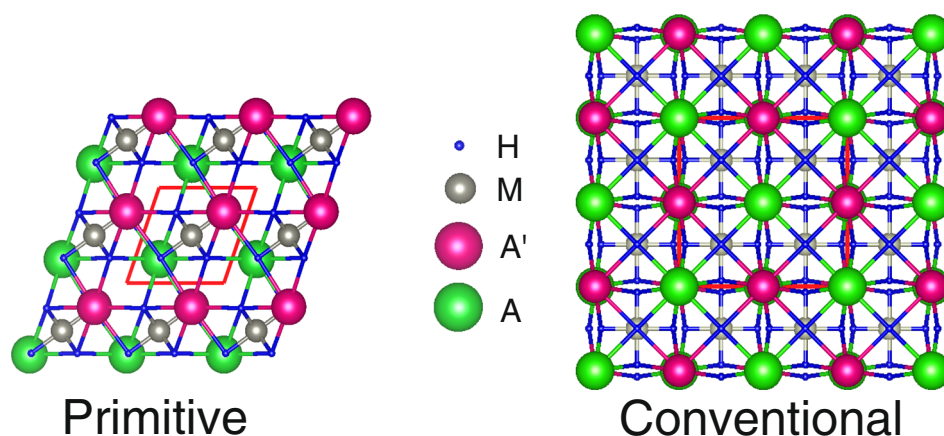
Fig. 4 Side view of the structure of AA′MH$_6$ materials with primitive and conventional unitcells, which are indicated by red lines. The green, purple, grey and blue spheres indicate the A, A′, M and H atoms, respectively.

Kadir et al. reported 5 different AA′MH$_6$ type semiconductors, where M = Ir[27]. They were able to synthesize NaCaIrH$_6$, NaSrIrH$_6$, NaBaIrH$_6$, KSrIrH$_6$, and KBaIrH$_6$ by direct combination of the alkali (Na and K), alkaline earth (Ca, Ba, and Sr) binary hydrides/ deuterides with Ir powder. Their X-ray and neutron powder diffraction studies confirm that those semiconductors have the space group symmetry F-43m. Furthermore, the open quantum materials database (OQMD)[28,29] contains the structural properties and band gaps of NaCaIrH$_6$, NaSrIrH$_6$, NaBaIrH$_6$ semiconductors and the MP database has those information on NaCaIrH$_6$, and NaBaIrH$_6$ semiconductors[25] (See Supplementary Information).

CubicGAN generates conventional structures with cubic Bravais lattice with F-43m (216) space group for AA′MH$_6$ materials, which have 36 atoms. On the contrary, the primitive unit cell with hexagonal Bravais lattice has only 9 atoms. Therefore, we considered the hexagonal unit cell to lower the computational time of the DFT calculations. In the primitive unit cells (see Fig. 4), green and red sites are symmetrically equivalent, while grey sites are located in the right middle of the hexagonal unit cell. Thus, we label the green and red sites as A and A′, while the middle site is M. Rest of the 6 sites are occupied by H atoms. In the research work of Kadir et al., they considered alkali atoms as A atoms, alkaline earth atoms as A′ atoms, and M atoms as Ir. In this research, our findings show that both A and A′ atoms can be alkali atoms (E.g., CsKSiH$_6$) or alkaline earth atoms (E.g., BaSrZnH$_6$). Moreover, the M atom can be a transition metal atom or even Al or Si. Therefore, our experiments show that those materials can have high chemical diversity.

The lattice parameters, A-H, M-H, A-M, and A-A′ bond lengths, are mentioned in Table 3. The primitive hexagonal unit cells have $a/c = 1$ lattice parameter ratio making $a = b = c$. As shown in Table 3, Mn-related AA′MH$_6$ and NaYRuH$_6$ structures have the shortest lattice parameters compared to the rest of the materials. They have lattice parameters less than 5.0 Å, while other materials have greater than 5.4 Å. All A, A′ and M elements make bonds with H atoms. A and A′ elements are bonded to twelve equivalent H atoms to form AH$_{12}$ and A′H$_{12}$ cuboctahedra. And also, M atoms make MH$_6$ octahedra by making bonds with 6 H atoms. An AH$_{12}$ (A′H$_{12}$) cuboctahedra shares corners with twelve equivalent AH$_{12}$ (A′H$_{12}$) cuboctahedra. Moreover, they share faces with four MH$_6$ octahedra[30]. Due to symmetry, A-H and A′-H bond lengths are equal. M-H bond lengths are the shortest compared to other bonds for a given compound. A-A′ of Mn-related AA′MH$_6$ and NaYRuH$_6$ structures are less than 3.4 Å, and A-M and A′-M distances are less than 3.1 Å. It can cause strong interactions between those atoms. A-A′ distance for the rest of the materials is greater than 3.8 Å, and A-M and A′-M distances are greater than 3.3 Å, indicating relatively weaker interactions.

The thermodynamic stability of the AA′MH$_6$ materials against their elements was studied using the formation energies, which were based on the following equation.

$$E_{form} = \frac{1}{N}\left(E_{tot} - \sum_i x_i E_i\right) \tag{1}$$

Here, E_{tot} is the total energy per unit formula of the material. x_i is the number of atoms of each element in the unit formula; i.e., 1 for A, A′, M atoms and 6 for H. $N = \sum x_i$; i.e., 9 for AA′MH$_6$. To find the atomic energies (E_i), we collected the most stable structures of each element using the Pymatgen code[26]. Same DFT settings were used to calculate the energy of each element. It is clear that all the six materials have negative formation energies, which confirms their stability. We also carried out spin-polarized calculations for the AA′MH$_6$ semiconductors with transition metal atoms to reveal whether they form magnetism. We observed that those materials do not have magnetic groundstates. Thus, all the AA′MH$_6$ semiconductors are nonmagnetic materials.

Mechanical properties and stability

Next, we studied the mechanical properties and stability of the AA′MH$_6$ materials by calculating the elastic constants using the DFPT method. To analyse the mechanical properties, we used the Vaspkit code[31], which computes the elastic constants by considering the AA′MH$_6$ cubic system. Since cubic unitcells has $a = b = c$ lattice lengths and $\alpha = \beta = \gamma = 90^0$ lattice angles, $C_{11} = C_{22} = C_{33}$, $C_{44} = C_{55} = C_{66}$, and $C_{12} = C_{13} = C_{23}$[32]. Therefore, we mention only the three independent elastic constants (C_{11}, C_{12} and C_{44}) in Table 4. It is clear that AA′MH$_6$ materials have relatively higher C_{11} for AA′MnH$_6$ and NaYRuH$_6$, compared to the other four materials in Table 4. As discussed before, the lattice constants and A-A′ bond lengths of AA′MnH$_6$ and NaYRuH$_6$ structures are considerably lower than that of the rest of the materials. As illustrated by Fig. 4, A-A′ bonds are aligned in **a**, **b** and **c** directions. C_{11}, C_{22}, and C_{33} are parallel to the **a**, **b** and **c** directions, respectively. Therefore, higher C_{ii} (i = 1, 2 and 3) can be mainly due to the strong interactions between the A and A′ atoms. Born stability criteria for the cubic systems are $C_{11} - C_{12} > 0$, $C_{11} + 2C_{12} > 0$ and $C_{44} > 0$[32]. It is clear from Table 4 that all the eight materials comply with the above requirements.

We also calculated the Bulk modulus (K), Young's modulus (Y), and isotropic Poisson's ratio (μ) based on the Hill approximation[33] as mentioned in Table 4. The smallest K values were found from CsKSiH$_6$ (16.615 GPa), while the largest value was calculated from AgZrMnH$_6$ (120.755 GPa). SrTlIrH$_6$ (21.915 GPa) provides the lowest Y, while NaZrMnH$_6$ (156.876 GPa) exhibits the maximum Y. It is clear
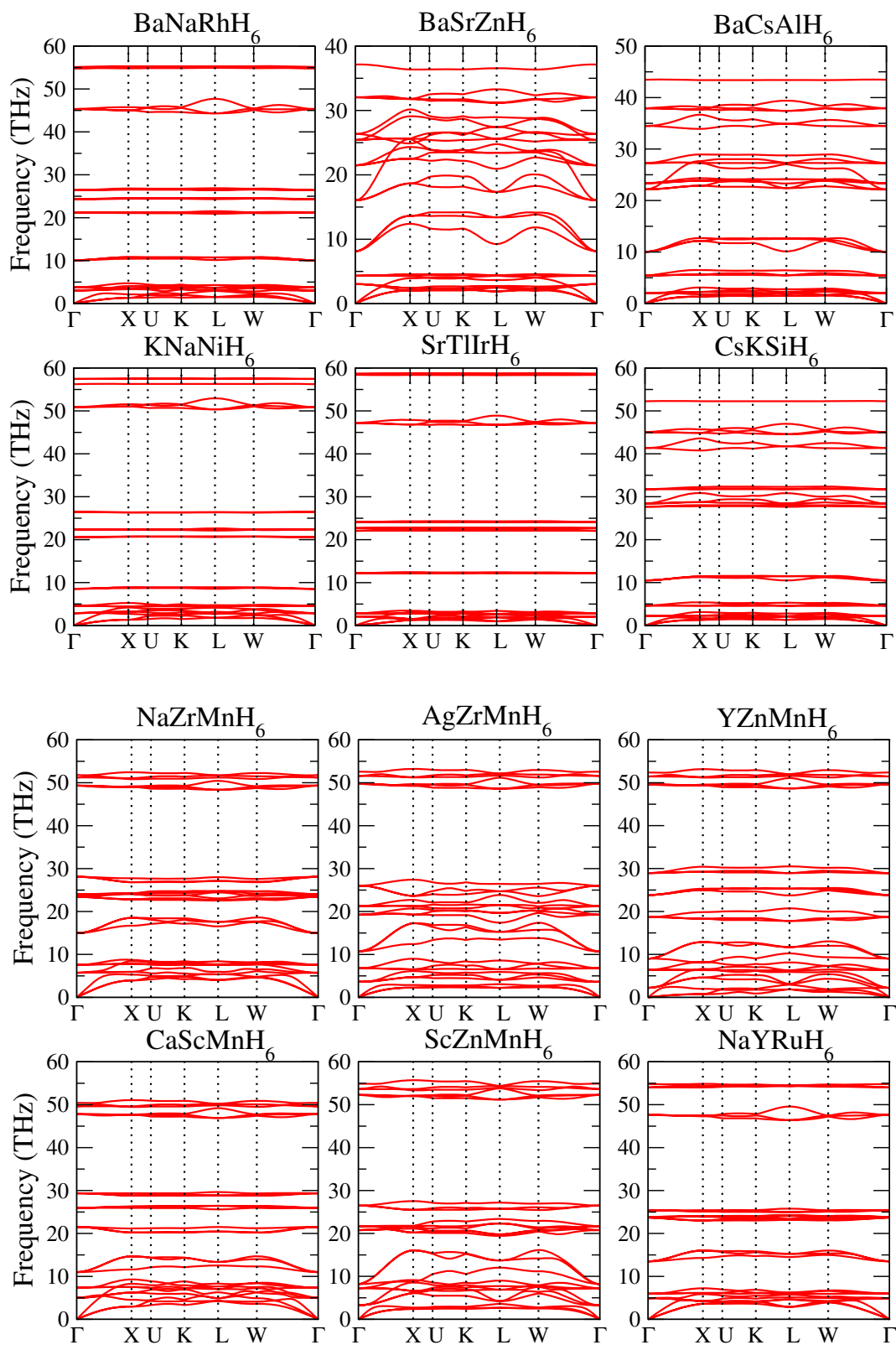
Fig. 5 Phonon dispersion of AA′MH₆ materials. The positive phonon frequencies in the figure indicate that the materials are dynamically stable.

that NaYRuH$_6$ and all the Mn-based materials have significantly larger K and Y values than that of the other six materials. This can be mainly because of high C_{ii} ($i = 1, 2,$ and 3) formed due to strong A-A′ bonds. Because of low Y, NaYRuH$_6$ and Mn-based AA′MH$_6$ materials can be considered stiffer materials relative to the other six semiconductors. And also, they exhibit more resistance to compression due to high K. All the μ values of the AA′MH$_6$ materials are between 0.2 and 0.4. maximum μ was found from SrTlIrH$_6$.

Table 3. The lattice parameter (*a*) and bond lengths in Å, and the formation energies in eV/atom for AA′MH$_6$ materials.

Material	*a*	A-H	M-H	A-M	A-A′	E_{form}
BaNaRhH$_6$	5.5105	2.76924	1.67023	3.37448	3.89651	−0.4678
BaSrZnH$_6$	5.47201	2.73801	1.83004	3.35091	3.8693	−0.3496
BaCsAlH$_6$	5.84781	2.93934	1.76669	3.58104	4.13503	−0.3159
SrTlIrH$_6$	5.56492	2.79622	1.69039	3.4078	3.93499	−0.3488
CsKSiH$_6$	5.86576	2.968	1.61863	3.59203	4.14772	−0.1817
KNaNiH$_6$	5.40252	2.73044	1.51196	3.30835	3.82016	−0.1715
NaYRuH$_6$	4.97064	2.48578	1.70949	3.04388	3.04388	−0.4999
CaScMnH$_6$	4.74883	2.37467	1.64383	2.90805	3.35793	−0.5261
YZnMnH$_6$	4.69455	2.34755	1.62362	2.87481	3.31955	−0.3099
NaZrMnH$_6$	4.64255	2.32129	1.63348	2.84297	3.28278	−0.3111
AgZrMnH$_6$	4.62246	2.31127	1.62077	2.83067	3.26857	−0.2070
ScZnMnH$_6$	4.50635	2.25322	1.60682	2.75957	3.18647	−0.3070

Thus, SrTlIrH$_6$ has considerably low *Y* and high *μ*. This indicates that SrTlIrH$_6$ semiconductor is less stiff due to small *Y* and more deformable elastically at small strains due to large *μ*.

Thermodynamic properties and dynamical stability

The temperature of the highest normal mode of a crystal is known as the Debye temperature θ_D. This can be obtained by employing Debye sound velocity (v_D) as explained by Eq. (2). Debye sound velocity can be calculated using the longitudinal and transverse sound velocities, which can be determined based on *K* and *G* as shown in Eq. (4)[34]. Here, *N*, V_0, and *ρ* are the number of atoms, volume, and density of the unicell, respectively. And also, *h* is Plank's constant, and k_B is Boltzmann's constant.

$$\theta_D = \frac{h}{k_B}\left(\frac{3N}{4\pi V_0}\right)^{\frac{1}{3}} v_D \qquad (2)$$

$$v_D = \left[\frac{1}{3}\left(\frac{2}{v_l^3} + \frac{1}{v_t^3}\right)\right]^{-\frac{1}{3}} \qquad (3)$$

$$v_l = \left(\frac{3K+4G}{3\rho}\right)^{\frac{1}{2}} \text{ and } v_t = \left(\frac{G}{\rho}\right)^{\frac{1}{2}} \qquad (4)$$

Table 5 shows the respective *ρ*, v_l, v_t, v_D and θ_D values for AA′MH$_6$ crystals. Debye temperature of NaYRuH$_6$ and Mn-based AA′MH$_6$ materials are significantly higher than that of other AA′MH$_6$ materials. As θ_D depends on *K* and *G* (see Eq. (4) and (2)), enhanced θ_D is due to the high *K* and *G* of those semiconductors.

We also plotted C_v as a function of temperature *T* using the Phonopy code[35]. C_v can be determined based on the following expression,

$$C_v = \sum_{\mathbf{q}j} k_B \left(\frac{\hbar\omega_{\mathbf{q}j}}{k_B T}\right)^2 \frac{\exp(\hbar\omega_{\mathbf{q}j}/k_B T)}{[\exp(\hbar\omega_{\mathbf{q}j}/k_B T) - 1]^2}, \qquad (5)$$

where $\omega_{\mathbf{q}j}$ is the phonon frequency for **q** wave vector at *j*th phonon band index and $\hbar$ is the reduced Plank's constant[35]. The phonon frequency for each K-point is plotted in Fig. 5. As can be seen in Fig. 6, the C_v of NaYRuH$_6$ and Mn-based AA′MH$_6$ materials are plotted with broken lines, and that of the rest of the materials is indicated by solid lines. It is clear that the C_v of NaYRuH$_6$ and Mn-based AA′MH$_6$ materials are smaller than that of the other materials at the low temperatures (0 to 150 K). At the low-temperature limit ($T \geq \theta_D$, $\theta_D/T << 1$), C_v is proportional to $(T/\theta_D)^3$. Since θ_D is higher compared to that of other materials, C_v is smaller at low temperatures for NaYRuH$_6$ and Mn-based AA′MH$_6$.

Electronic Properties

As can be seen in Table 6, A, A′ and M elements lose electrons (except in Ru, where it has small negative value), while H atoms gain electrons. Thus, we can expect an ionic character in A-H, A′-H, and M-H bonds. Even though A and A′ sites are symmetrically equivalent, the atoms at those sites can lose a different amount of electrons. This is mainly because atoms at those sites have different oxidation states. Based on Table 6, Na, K, and Cs alkali atoms have their usual oxidation state (+1), while alkaline earth atoms such as Ca, Sr, and Ba lose more than 1 electron as they can donate up to 2 electrons. Al, Si, and Tl exhibit their most common oxidation states, which are +3, +4, and +1, respectively. It is reported that first-principles computations provide only negligible changes in the local transition-metal charge for semiconducting crystals[36]. Therefore, we propose that we can consider MH$_6^{n-}$ complex as a single unit since the M-H bond lengths are very short compared to other H-related bonds. *n* can be found by computing $\Delta q_M + 6 \times \Delta q_H$, which is greater than 2 for all the M atoms except for Ni and Si. For those two atoms, $n \approx 1.6$. Therefore, we can expect MH$_6^{2-}$ for Si and Ni complexes, while MH$_6^{3-}$ for the rest of the complexes. Kadir et al. suggest that IrH$_6^{3-}$ complexes exist in AA′IrH$_6$ semiconductors[27]. Therefore, MH$_6^{3-}$ can be the common complex that exists in AA′MH$_6$ materials.

Figures 7 and 8 show the band structures and partial density of states (PDOS) of the AA′MH$_6$ materials. It is clear that all six AA′MH$_6$ materials are semiconductors. The bandgap for each material is mentioned in Table 7. The DFT calculations with PBE exchange-correlation functional underestimate the band gaps due to self-interaction error. It has been shown that the Heyd-Scuseria-Ernzerhof (HSE) screened Coulomb hybrid functional calculations provide reasonable estimation for the band gaps of semiconductors[37,38]. HSE06 uses $\frac{1}{4}$ of exact exchange and $\frac{3}{4}$ of PBE exchange. Based on our HSE06 computations, all the AA′MH$_6$ semiconductors can have bandgaps greater than 2.00 eV (see Supplementary Information). The bandgap range of wide-bandgap semiconductors is considered as the range above 2 eV[23]. Thus, we can identify that those materials are wide-bandgap semiconductors. As reported by Kadir et al., NaCaIrH$_6$, NaSrIrH$_6$, NaBaIrH$_6$, KSrIrH$_6$ and KBaIrH$_6$ have bandgaps between 2.91 and 3.33 eV[27] (see Supplementary Information). Wide-bandgap semiconductors are vital for manufacturing optical devices emitting green, red, and UV frequencies and also power devices functioning at higher temperatures[23,24].

Other than in BaCsAlH$_6$ and CsKSiH$_6$, all the AA′MH$_6$ materials have their conduction band minimum (CBM) at X high-symmetric K-point. The CBM of BaCsAlH$_6$ and CsKSiH$_6$ are at Γ points. The valence band maximum (VBM) of BaNaRhH$_6$, SrTlIrH$_6$, YMnZnH$_6$, NaYRuH$_6$, and AgZrMnH$_6$ exist in W K-point. BaSrZnH$_6$, KNaNiH$_6$, and BaSrZnH$_6$ have VBM at X, while that of CaScMnH$_6$ and AgZrMnH6 is at K high-symmetric point in the reciprocal space. Thus, both CBM and VBM of BaSrZnH$_6$ and KNaNiH$_6$ reside at X K-point, indicating those materials are direct bandgap semiconductors. Direct bandgap semiconductors are preferred for LED and laser devices over their indirect counterparts. Wide-bandgap semiconductors with direct bandgap are widely investigated for solar cells due to optical transparency[39]. BaNaRhH$_6$, KNaNiH$_6$, CaScMnH$_6$, and NaYRuH$_6$ materials have very flat bands near the Fermi level, which is indicated by zero energy. Relative to other materials, BaSrZnH$_6$ contains narrow (less flat) bands near the Fermi level. As a result, this can lower the effective mass of the carriers. Some research has shown that low effective mass will help developing efficient thermoelectric devices[40–42]. As shown by electronic band theory, the electron effective mass can be very high in the flat bands[43]. It is also shown that flat bands at the bottom of the conduction bands can provide high thermoelectric power[44]. YMnZnH$_6$ and ScZnMnH$_6$ materials also exhibit that the CBM are relatively flat. Moreover, as shown in Fig. 7, we can modulate the shape of the bands near the

Table 4. The mechanical properties of the AA′MH$_6$ materials.

Material	C_{11}	C_{12}	C_{44}	K	G	Y	μ
BaNaRhH$_6$	60.514	20.174	11.901	33.621	14.722	38.541	0.309
BaSrZnH$_6$	86.238	20.648	30.492	42.512	31.393	75.577	0.204
BaCsAlH$_6$	50.455	16.174	13.119	27.601	14.602	37.24	0.275
SrTlIrH$_6$	52.588	26.735	5.549	35.353	7.845	21.915	0.397
CsKSiH$_6$	28.939	10.454	7.203	16.615	7.960	20.592	0.293
KNaNiH$_6$	36.027	11.720	6.618	19.822	8.463	22.225	0.313
NaYRuH$_6$	131.450	36.248	44.149	67.982	45.499	111.600	0.226
CaScMnH$_6$	192.385	28.606	38.997	83.199	52.744	130.627	0.238
YZnMnH$_6$	171.063	50.92	23.563	90.968	34.649	137.167	0.331
NaZrMnH$_6$	196.714	47.003	57.275	96.907	63.761	156.876	0.23
AgZrMnH$_6$	190.957	85.654	50.274	120.755	51.211	134.606	0.314
ScZnMnH$_6$	191.749	60.602	60.537	104.317	62.504	156.295	0.250

The C_{11}, C_{12}, and C_{44} elastic constants, bulk modulus (K), Shear modulus (G) and Young's modulus (Y) were calculated in GPa. μ is the isotropic Poisson's ratio.

Table 5. The density (ρ), longitudinal (v_l), transverse (v_t) and average (v_D) sound velocity, Debye temperature (θ_D) and specific thermal capacity at 300 K ($C_v^{300\,K}$) for the AA′MH$_6$ materials.

Material	ρ (gcm^{-3})	v_l (ms^{-1})	v_t (ms^{-1})	v_D (ms^{-1})	θ_D (K)	$C_v^{300\,K}$ (JK^{-1}mol^{-1})
BaNaRhH$_6$	4.0401	3630.4733	6902.4643	2858.5852	337.0342	119.113
BaSrZnH$_6$	3.4683	4932.0885	12405.1302	3901.0279	463.1759	127.961
BaCsAlH$_6$	2.9078	4023.4038	7286.8499	3164.0609	351.5324	115.715
SrTlIrH$_6$	5.4546	2898.0880	4445.3015	2265.8564	264.5375	120.116
CsKSiH$_6$	1.9584	3728.7223	5212.2198	2901.6325	321.3897	109.520
KNaNiH$_6$	1.5423	4490.9338	5199.8403	3445.3146	414.3299	121.746
NaYRuH$_6$	3.4194	6133.7278	14188.8339	4846.6876	633.4998	112.140
CaScMnH$_6$	2.6144	7663.0364	14886.6869	6036.7538	825.9060	110.588
YZnMnH$_6$	3.9895	5863.5797	10231.3126	4606.2522	637.4814	116.661
NaZrMnH$_6$	3.3572	7361.3351	16649.9772	5814.9095	813.7671	106.507
AgZrMnH$_6$	5.0489	6118.9252	12770.7618	4827.2738	678.4886	106.507
ScZnMnH$_6$	3.5897	7230.2461	15811.8943	5708.4879	823.0192	115.952
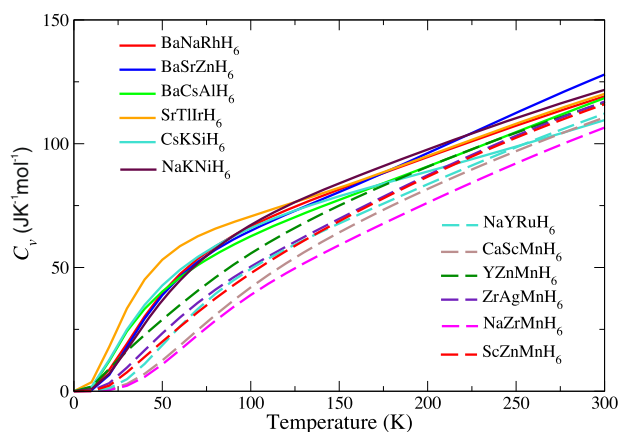
Fig. 6 The specific heat capacity (C_v) of AA′MH$_6$ materials as a function of temperature (T). Here, broken lines indicate C_v of AA′MnH$_6$ and NaYRuH$_6$ materials, while solid lines indicate that of the other materials.

Table 6. The Bader charge transfer (Δq) in electrons for each element of AA′MH$_6$ Materials.

| Material	Δq_A	$\Delta q_{A'}$	Δq_M	Δq_H
BaNaRhH$_6$	1.4256	0.8613	0.0127	−0.3841
BaSrZnH$_6$	1.3057	1.4588	0.6579	−0.5701
BaCsAlH$_6$	1.4853	0.7135	3.0000	−0.8581
SrTlIrH$_6$	1.5551	0.4674	0.2944	−0.3855
CsKSiH$_6$	0.7625	0.8483	4.0000	−0.9345
KNaNiH$_6$	0.7781	0.8734	0.5139	−0.3612
NaYRuH$_6$	0.8341	1.8845	−0.1099	−0.4356
CaScMnH$_6$	1.3626	1.6427	0.2943	−0.5509
YZnMnH$_6$	1.7862	0.7053	0.3301	−0.4717
NaZrMnH$_6$	0.8179	1.8115	0.3468	−0.4978
AgZrMnH$_6$	0.2876	1.8191	0.3904	−0.4168
ScZnMnH$_6$	1.6590	0.7511	0.2937	−0.4484
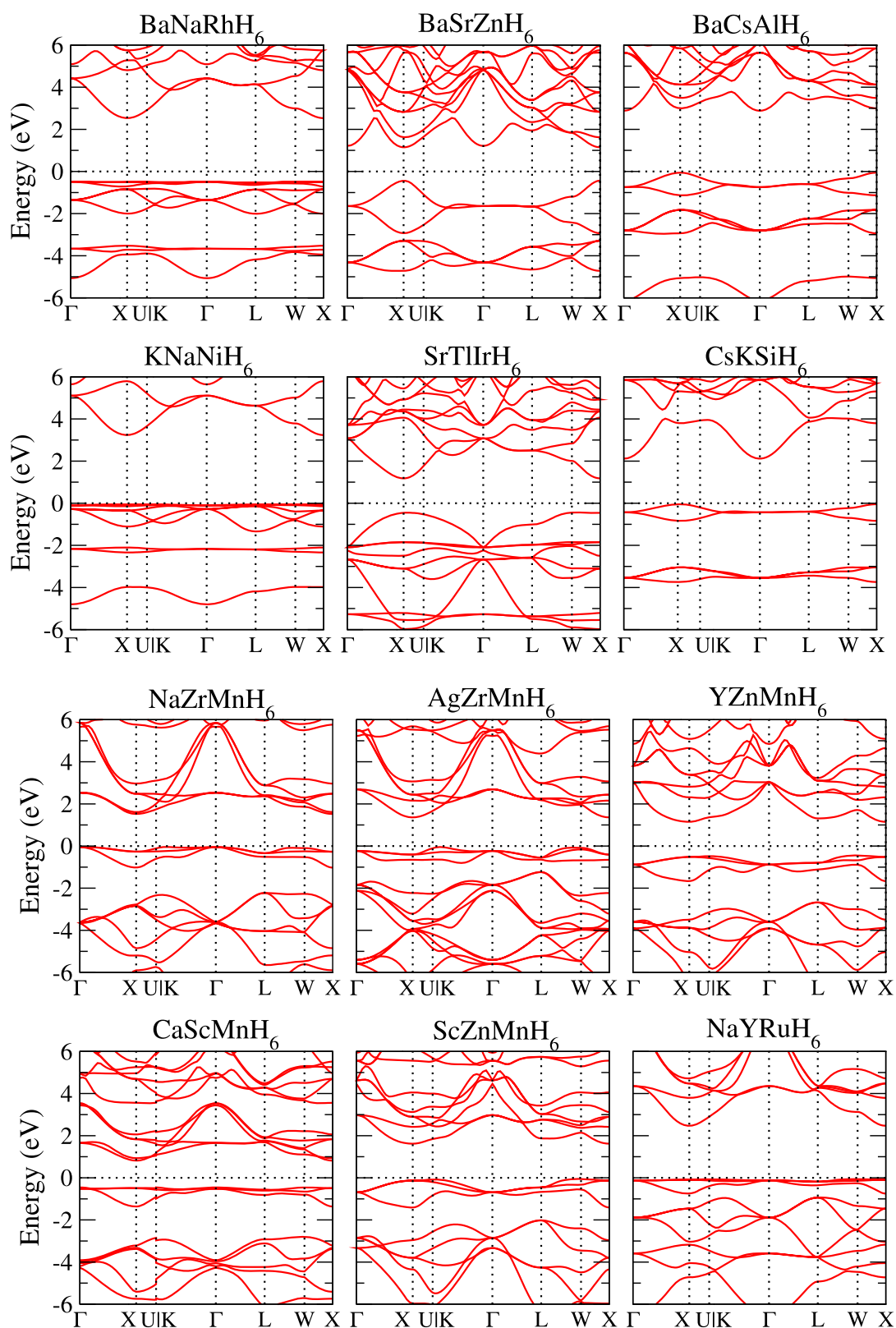
Fig. 7 Band structures of AA′MH₆ materials. Fermi energy marks zero energy.

Fermi level using the chemical formula. As a result, the thermo-electric properties can be tuned. Therefore, we propose that AA′MH₆ semiconductors should be investigated for thermoelectric applications. Our partial density of states (PDOS) studies reveal that d-orbitals of transition metal atoms reside at the M site dominate in the valence region near the Fermi level. Even though the transition metal atoms can be found at A and A′ sites, their pdos of d-orbitals are not significant near the Fermi level.
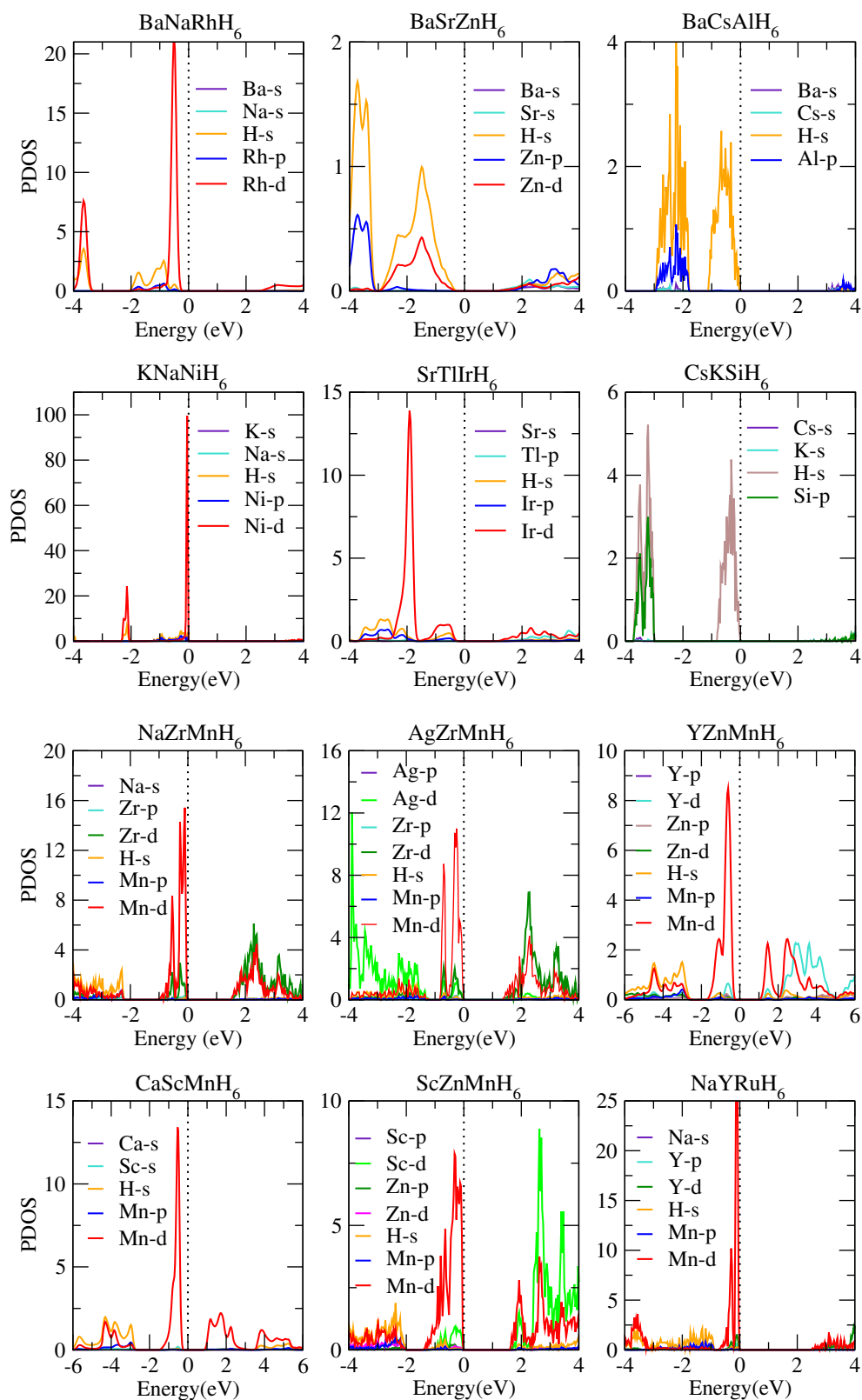
Fig. 8 Partial density of states (PDOS) of AA′MH₆ materials. Fermi energy marks zero energy.

METHOD

Generative adversarial network

The hypothetical materials used in our research are generated by our CubicGAN[22], a generative adversarial network (GAN) based model for generating cubic crystal structures in a high-throughput manner. Our GAN model consists of a generator network and a discriminator/critic network. The discriminator learns to tell real materials from fake materials generated by the generator. The generator learns how to generate samples with similar distribution as the training samples. After trained, we can sample

Table 7. The band gap, conduction band minimum (CBM), valence band maximum (VBM) and type of semiconductor for $AA'MH_6$ materials.

Material	Band Gap (eV)	CBM	VBM	Type
$BaNaRhH_6$	3.0181	X	W	Indirect
$BaSrZnH_6$	1.5864	X	X	Direct
$BaCsAlH_6$	2.9382	G	X	Indirect
$SrTlIrH_6$	1.6245	X	W	Indirect
$CsKSiH_6$	2.1647	G	X	Indirect
$KNaNiH_6$	3.3024	X	X	Direct
$NaYRuH_6$	2.5322	X	W	Indirect
$CaScMnH_6$	1.2716	X	K	Indirect
$YZnMnH_6$	1.6003	X	W	Indirect
$NaZrMnH_6$	1.5694	X	K	Indirect
$AgZrMnH_6$	1.4111	X	K	Indirect
$ScZnMnH_6$	1.6732	X	W	Indirect
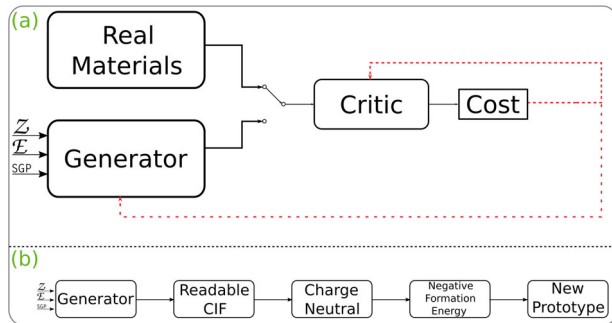
Fig. 9 The main framework of CubicGAN. a WGAN architecture and **b** postprocessing of generated samples. Z is the random noise. E is the element vector embedded by 23 element properties. SGP is the one hot embedding for the space group.

from the generator to generate nonexisting materials. In CubicGAN, we focused on generating ternary and quaternary materials with the space groups 221, 225, and 216. Moreover, to simplify the problem, CubicGAN uses special fractional coordinates, all in the set of {0.0, 0.25, 0.5, 0.75}. The CubicGAN is trained using material data from OQMD[45,46] and is evaluated on material data from Materials Project[47] and ICSD[48]. The main framework of CubicGAN and the post-processing for the generated materials are shown in Fig. 9. It is notoriously hard to train the original GAN model because the adversarial loss is not continuous in the generator, which causes vanishing gradients and saturation in the discriminator. We take advantage of the Wasserstein GAN with gradient penalty by penalizing the norm of gradients of the critic with respect to the inputs[18]. The critic takes real materials and fake materials generated by the generator and then outputs a score which can be interpreted as how real the input materials are. The score is used to update the parameters of the models of the generator and the critic. The adversarial loss is defined as:

$$\mathcal{L} = \mathbb{E}_{\tilde{\mathbf{x}} \sim \mathbb{P}_g}[D(\tilde{\mathbf{x}})] - \mathbb{E}_{\mathbf{x} \sim \mathbb{P}_r}[D(\mathbf{x})] + \lambda \mathop{E}_{\hat{\mathbf{x}} \sim \mathbb{P}_{\hat{\mathbf{x}}}}[(\|\nabla_{\hat{\mathbf{x}}} D(\hat{\mathbf{x}})\|_2 - 1)^2] \quad (6)$$

where D means the score function from the critic. $\hat{\mathbf{x}}$ is the linear interpolation between a real material $\mathbf{x}$ and the generated one $\hat{\mathbf{x}}$ and $\mathop{E}_{\hat{\mathbf{x}} \sim \mathbb{P}_{\hat{\mathbf{x}}}}[(\|\nabla_{\hat{\mathbf{x}}} D(\hat{\mathbf{x}})\|_2 - 1)^2]$ is the gradient penalty which enforces gradients with the norm at most 1 everywhere. λ is set 10 by default in this work.

Conditioning on random noise, three or four-element combinations, and space group, the generator not only generates materials with existing prototypes but also generates stable ones with nonexisting prototypes. When the CubicGAN generates 10 million materials, it can rediscover most of the cubic materials in Materials Project and ICSD. In CubicGAN, we only focus on the generated materials with prototypes, which are defined by the anonymous formula and the space group ID. In total, 24 and 1 nonexisting prototypes are found in 10 million generated ternary and

quaternary materials, respectively. Sub-figure (a) of Fig. 9 shows how to filter out the materials. On average, 90% of generated materials have readable CIFs, and we only select materials with neutral charge and negative formation energy predicted by CGCNN[49]. After filtering down materials with nonexisting prototypes, we performed DFT calculations, and 36847 candidate materials have been relaxed successfully. Further, 506 stable materials are verified by phonon dispersion.

Nonmetal - metal classifier

To develop a nonmetal - metal classifier, we first collected the pretty formulas, Bravais lattice type, and bandgap details of all the cubic quaternary materials from the MP database. There were 2578 nonzero bandgap materials (semiconductors and insulators) and 1,438 metals in the collected dataset. We considered 55 elemental and electronic structure attributes, such as the first ionization energy, atomic volume, electronegativity, total number of valence electrons, and number of valence electrons in s, p, d, and f orbitals, to develop the feature set (see Supplementary Information). The weighted average (Avg.) and a maximum difference of those properties for a given chemical formula were added to the feature set. The Avg. of a property S of a quaternary compound $A_\alpha B_\beta C_\gamma D_\delta$ was calculated based on the following expression,

$$S_{A_\alpha B_\beta C_\gamma D_\delta}^{Avg} = \frac{1}{\alpha + \beta + \gamma + \delta}(\alpha S_A + \beta S_B + \gamma S_C + \delta S_D), \quad (7)$$

where S_A, S_B, S_C and S_D are the property S of A, B, C, and D elements, respectively. Altogether, 119 features were considered for training the models.

We created the DNN classifier with two hidden layers using Keras[50] on top of TensorFlow[51]. The first and second hidden layers of DNN include 200, and 100 neurons, respectively. To include the nonlinearity in the system, we shifted the summed weighted inputs of each layer through the rectified linear unit (ReLu) activation function. We randomly dropped out 5% of the units of the hidden layers while training the models. This process is very important for limiting the overfitting of training data. Another useful approach to diminishing overfitting is weight regularization. We employed Ridge (L2) regularization method for adding penalties during updating weights. The adaptive moment estimation (Adam) optimizer with a 0.001 learning rate was considered with binary cross-entropy as the loss function and the metric during the calculations. The optimized number of epochs and batch size are 500 and 1500, respectively.

We developed a random forest classifier (RFC) as the second model, which uses an ensemble technique. Here, data is divided randomly, which is known as bagging and carries out training with multiple decision trees. The final prediction is given by averaging the output of all the decision trees. The hyperparameter optimization was performed using GridSearchCV algorithm as implemented in the Scikit-learn code[52]. The optimized number of decision trees, minimum samples split, minimum samples leaf, and maximum depth are 500, 10, 3, and 90, respectively. Furthermore, we used the RFC model to study the feature importance for whole quaternary materials data set. It will help discovering semiconductors in the future.

For both DNN and RFC models, the cubic quaternary materials dataset with 4016 materials was split randomly into 98 % and 2 % as the training and testing sets, respectively. The 10-fold cross-validation with accuracy as the scoring method was performed on the training set. Here, the training set was partitioned into 10 subsets, where 9 subsets were for training the model and the remaining subset was for validating.

Density functional theory (DFT) calculations

Density functional theory calculations were performed as implemented in the Vienna ab simulation package (VASP) code[53–56]. The electron wave functions were described using the PAW pseudopotentials[57,58]. The exchange-correlation interactions were treated based on the generalized gradient approximation (GGA) within the Perdew-Burke-Ernzerhof (PBE) formulation[59,60]. The energy threshold value of the plane-wave basis was set as 500 eV. In addition, the energy convergence criteria were set to 10^{-8} eV, and the force convergence criterion for the ionic steps is set to 10^{-2} eV/Å. The Brillouin zone integrations were performed using a dense k-point mesh within the Monkhorst-Pack scheme for the structure optimizations, band structure, density of states, mechanical properties, and phonon calculations. For instance, a $14 \times 14 \times 14$ K-mesh was used for $BaNaRhH_6$ with 5.5105 Å lattice constant. The $2 \times 2 \times 2$ supercells were employed for obtaining Phonon dispersions using the Phonopy code[35].

The elastic constants were calculated by employing density functional perturbation theory (DFPT) as implemented in VASP[61]. VASPKIT code[31] was used to obtain the bulk modulus (K), Shear modulus (G), Young's modulus (Y), and Poisson's ratio (μ) of the materials based on the Hill method[62].

DATA AVAILABILITY

CODE AVAILABILITY

REFERENCES

1. Chuang, R. W., Wu, R.-X., Lai, L.-W. & Lee, C.-T. Zno-on-gan heterojunction light-emitting diode grown by vapor cooling condensation technique. *Appl. Phys. Lett.* **91**, 231113 (2007).
2. Yu, L. et al. High-performance wse2 complementary metal oxide semiconductor technology and integrated circuits. *Nano Lett.* **15**, 4928–4934 (2015).
3. Green, M. A. & Bremner, S. P. Energy conversion approaches and materials for high-efficiency photovoltaics. *Nat. Mater.* **16**, 23–34 (2017).
4. Lin, Y. et al. Graphene/semiconductor heterojunction solar cells with modulated antireflection and graphene work function. *Energy Environ. Sci.* **6**, 108–115 (2013).
5. Oba, F. & Kumagai, Y. Design and exploration of semiconductors from first principles: A review of recent advances. *Appl. Phys. Express* **11**, 060101 (2018).
6. Rom, S., Ghosh, A., Halder, A. & Dasgupta, T. S. Machine learning classification of binary semiconductor heterostructures. *Phys. Rev. Materials* **5**, 043801 (2021).
7. Charles, H. & Sujan, G. In *Microelectronic packaging: Electrical interconnections* (Elsevier, 2016).
8. Rahman, M. A. A review on semiconductors including applications and temperature effects in semiconductors. *ASRJETS* **7**, 50–70 (2014).
9. Hinuma, Y. et al. Discovery of earth-abundant nitride semiconductors by computational screening and high-pressure synthesis. *Nat. Commun.* **7**, 11962 (2016).
10. Casady, J. & Johnson, R. Status of silicon carbide (sic) as a wide-bandgap semiconductor for high-temperature applications: A review. *Solid State Electron.* **39**, 1409–1422 (1996).
11. Ortiz, C., Eriksson, O. & Klintenberg, M. Data mining and accelerated electronic structure theory as a tool in the search for new functional materials. *Comput. Mater. Sci.* **44**, 1042–1049 (2008).
12. Setyawan, W., Gaume, R. M., Lam, S., Feigelson, R. S. & Curtarolo, S. High-throughput combinatorial database of electronic band structures for inorganic scintillator materials. *ACS Comb. Sci.* **13**, 382–390 (2011).
13. Luo, S., Li, T., Wang, X., Faizan, M. & Zhang, L. High-throughput computational materials screening and discovery of optoelectronic semiconductors. *WIRES Rev. Comput. Mol. Sci.* **11**, e1489 (2021).
14. Zhao, X.-G. et al. Cu-in halide perovskite solar absorbers. *J. Am. Chem. Soc.* **139**, 6718–6725 (2017).
15. 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).
16. Goodfellow, I. et al. Generative adversarial nets. *Advances in neural information processing systems* **27** (2014) .
17. Arjovsky, M., Chintala, S. & Bottou, L. *Wasserstein generative adversarial networks*, 214–223 (PMLR, 2017).
18. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. & Courville, A. Improved training of wasserstein gans. *arXiv preprint arXiv:1704.00028* (2017).
19. Nouira, A., Sokolovska, N. & Crivello, J.-C. Crystalgan: learning to discover crystallographic structures with generative adversarial networks. *arXiv preprint arXiv:1810.11203* (2018).
20. Zhu, J.-Y., Park, T., Isola, P. & Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks, 2223–2232 (2017).
21. Kim, S., Noh, J., Gu, G. H., Aspuru-Guzik, A. & Jung, Y. Generative adversarial networks for crystal structure prediction. *ACS Cent. Sci.* **6**, 1412–1420 (2020).
22. Zhao, Y. et al. High-throughput discovery of novel cubic crystal materials using deep generative neural networks. *Adv. Sci. Lett.* **8**, 2100566 (2021).
23. Takahashi, K., Yoshikawa, A. & Sandhu, A. *Wide bandgap semiconductors*. (Springer-Verlag Berlin Heidelberg. 239 2007).
24. Millan, J., Godignon, P., Perpiñà, X., Pérez-Tomás, A. & Rebollo, J. A survey of wide bandgap power semiconductor devices. *IEEE Trans. Power Electron.* **29**, 2155–2163 (2013).
25. Jain, A. et al. The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **1**, 011002 (2013).
26. Ong, S. P. et al. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* **68**, 314–319 (2013).
27. Kadir, K., Moser, D., Münzel, M. & Noréus, D. Investigation of counterion influence on an octahedral irh6-complex in the solid state hydrides aaeirh6 (a = na, k and ae = ca, sr, ba, and eu) with a new structure type. *Inorg. Chem.* **50**, 11890–11895 (2011).
28. Zolotariov, D. Development of the approximating functions method for problems in a planar waveguide with constant polarization. *Int. J. Math. Comput. Res.* **9**, 2515–2520 (2021).
29. Shimbaleva, I. *Hidden treasures of music* (Independently Published, 2021).
30. Project, T. M. Materials data on nacah6ir by materials project (2020).
31. Wang, V., Xu, N., Liu, J.-C., Tang, G. & Geng, W.-T. Vaspkit: A user-friendly interface facilitating high-throughput computing and analysis using vasp code. *Comput. Phys. Commun.* **267**, 108033 (2021).
32. Mouhat, F. & Coudert, F.-X. Necessary and sufficient elastic stability conditions in various crystal systems. *Phys. Rev. B* **90**, 224104 (2014).
33. Hill, R. The elastic behaviour of a crystalline aggregate. *Proc. Phys. Soc. A* **65**, 349 (1952).
34. Li, C. & Wang, Z. In *9 - computational modelling and ab initio calculations in max phases - I* (ed. Low, I.). *Advances in Science and Technology of Mn+1AXn Phases* 197–222 (Woodhead Publishing, 2012).
35. Togo, A. & Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **108**, 1–5 (2015).
36. Raebiger, H., Lany, S. & Zunger, A. Charge self-regulation upon changing the oxidation state of transition metals in insulators. *Nature* **453**, 763–766 (2008).
37. Henderson, T. M., Paier, J. & Scuseria, G. E. Accurate treatment of solids with the hse screened hybrid. *physica status solidi (b)* **248**, 767–774 (2011).
38. Heyd, J. & Scuseria, G. E. Efficient hybrid density functional calculations in solids: Assessment of the heyd-scuseria-ernzerhof screened coulomb hybrid functional. *J. Chem. Phys.* **121**, 1187–1192 (2004).
39. Woods-Robinson, R. et al. Wide band gap chalcogenide semiconductors. *Chem. Rev.* **120**, 4007–4055 (2020).
40. Witting, I. T. et al. The thermoelectric properties of bismuth telluride. *Adv. Electron. Mater.* **5**, 1800904 (2019).
41. Suwardi, A. et al. Inertial effective mass as an effective descriptor for thermoelectrics via data-driven evaluation. *J. Mater. Chem. A* **7**, 23762–23769 (2019).
42. Pei, Y., LaLonde, A. D., Wang, H. & Snyder, G. J. Low effective mass leading to high thermoelectric performance. *Energy Environ. Sci.* **5**, 7963–7969 (2012).
43. Zhong, C., Xie, Y., Chen, Y. & Zhang, S. Coexistence of flat bands and dirac bands in a carbon-kagome-lattice family. *Carbon* **99**, 65–70 (2016).
44. Yabuuchi, S., Okamoto, M., Nishide, A., Kurosaki, Y. & Hayakawa, J. Large seebeck coefficients of fe2tisn and fe2tisi: First-principles study. *Appl. Phys. Express.* **6**, 025504 (2013).
45. Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (oqmd). *Jom* **65**, 1501–1509 (2013).
46. Kirklin, S. et al. The open quantum materials database (oqmd): assessing the accuracy of dft formation energies. *Npj Comput. Mater.* **1**, 1–15 (2015).
47. Jain, A. et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
48. Bergerhoff, G., Brown, I. & Allen, F. et al. Crystallographic databases. *IUCr* **360**, 77–95 (1987).
49. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *PRL* **120**, 145301 (2018).
50. Chollet, F. et al. Keras. https://keras.io (2015).
51. Abadi, M. et al. TensorFlow: Large-scale machine learning on heterogeneous systems. https://www.tensorflow.org/ (2015). Software available from tensorflow.org.
52. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *JMLR* **12**, 2825–2830 (2011).
53. Kresse, G. & Hafner, J. ab initio. *Phys. Rev. B* **47**, 558–561 (1993).
54. Kresse, G. & Hafner, J. ab initio. *Phys. Rev. B* **49**, 14251–14269 (1994).
55. G. Kresse, J. F. Efficiency of ab initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).

56. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).

57. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).

58. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).

59. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

60. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple [phys. rev. lett. 77, 3865 (1996)]. *Phys. Rev. Lett.* **78**, 1396–1396 (1997).

61. Baroni, S., de Gironcoli, S., Dal Corso, A. & Giannozzi, P. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.* **73**, 515–562 (2001).

62. Hill, R. The elastic behaviour of a crystalline aggregate. *Proc. Phys. Soc. A* **65**, 349–354 (1952).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

Conceptualization, J.H. and E.S.; methodology, E.S., Y.Z.; software, J.H., Y.S.; resources, J.H., I.P.; writing–original draft preparation, E.S., Y.Z.; writing–review and editing, J.H., I.P., and E.S.; visualization, E.S. and Y.Z.; supervision, J.H.; funding acquisition, J.H.

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/s41524-022-00850-3.

Correspondence and requests for materials should be addressed to Jianjun Hu.

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.