Journal of Materials Chemistry C

ROYAL SOCIETY OF CHEMISTRY

COMMUNICATION

View Article Online
View Journal | View Issue

Cite this: J. Mater. Chem. C, 2023, 11, 5601

Received 1st January 2023,
Accepted 19th March 2023

DOI: 10.1039/d3tc00001j

rsc.li/materials-c

Investigating magnetic van der Waals materials using data-driven approaches†

Romakanta Bhattarai, Peter Minch and Trevor David Rhone *

In this work, we investigate magnetic monolayers of the form $A^{i}A^{ii}B_4X_8$ based on the well-known intrinsic topological magnetic van der Waals (vdW) material $MnBi_2Te_4$ (MBT) using first-principles calculations and machine learning techniques. We select an initial subset of structures to calculate the thermodynamic properties, electronic properties, such as the band gap, and magnetic properties, such as the magnetic moment and magnetic order using density functional theory (DFT). Data analytics approaches are used to gain insight into the microscopic origin of materials' properties. The dependence of materials' properties on chemical composition is also explored. For example, we find that the formation energy and magnetic moment depend largely on A and B sites whereas the band gap depends on all three sites. Finally, we employ machine learning tools to accelerate the search for novel vdW magnets in the MBT family with optimized properties. This study creates avenues for rapidly predicting novel materials with desirable properties that could enable applications in spintronics, optoelectronics, and quantum computing.

Introduction

Two-dimensional (2D) materials represent an important family of materials that have been of central interest since the discovery of graphene in 2004.[1] A large number of 2D materials have been investigated using computational and experimental approaches. The tunability of their properties down to a single layer makes them ideal candidates for a variety of applications including nanoelectronics, optoelectronics,[2–4] sensing,[5] memory devices,[6] spintronics,[7] and quantum computing.[8,9] This implies that 2D materials hold great potential for driving future industrial innovation. Materials with intrinsic 2D magnetism are of particular interest because these materials can exhibit phenomena such as ferromagnetism, and the quantum anomalous hall effect (QAHE). Although existing materials databases, such as C2DB,[10] and 2DMatPedia,[11] contain a lot of 2D magnetic materials that are predicted *via* computational methods, there are only a handful of such materials that are experimentally demonstrated to show intrinsic 2D magnetism. Examples of these include $Cr_2Ge_2Te_6$,[12] CrI_3,[13] VSe_2,[14] Fe_3GeTe_2,[15] and $FePS_3$.[16]

$MnBi_2Te_4$ is one of the most studied 2D magnetic materials[17–28] whose bulk phase was experimentally studied in 2013.[29] It consists of blocks of a septuple layer (SL) of Mn, Bi, and Te in the following order: Te–Bi–Te–Mn–Te–Bi–Te. Each SL can be exfoliated as a single layer that exhibits intralayer ferromagnetism (FM). Interestingly, any two adjacent SL blocks can be stacked to give rise to an anti-ferromagnetic (AFM) order. Bulk MBT has an AFM order and is a topological insulator (TI).[17,30] Meanwhile, the MBT monolayer is a direct band gap semiconductor with a band gap of 0.70 eV.[6] In addition, MBT also hosts several exotic phases that depend on the layer number. These include the quantum anomalous hall insulator,[26] Weyl semimetal,[6] and axion insulator.[31] The alternating magnetic ordering and rich topological properties make MBT an exciting 2D topological quantum magnetic material that can have potential applications in topological quantum computing, nanoelectronics, spintronics, and data storage.

Discovering novel materials with desirable properties in a large materials space *via* experiments or first-principles quantum calculations is prohibitive as both methods are expensive in terms of time and cost. Access to supercomputers has accelerated the exploration of materials resulting in an increasing number of novel materials in recent years. The rise of materials databases[10,11,32,33] and access to efficient ML algorithms facilitate the growing area of materials informatics, *i.e.* applying machine learning to materials science. Materials informatics enables the accelerated discovery of materials with desirable characteristics through the efficient exploration of a large set of materials.[34–45] ML lowers the computational cost of materials discovery due to its ability to learn from a small dataset and make predictions on a much larger dataset. As a result, ML techniques have been employed in various areas of materials science including 2D

Department of Physics, Applied Physics, and Astronomy, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA. E-mail: rhonet@rpi.edu

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3tc00001j

magnetic materials,[37,46] Janus materials,[47] energy materials,[48,49] MXenes,[50,51] catalysts[52] and other functional 2D materials[38] to accelerate materials discovery.

In this paper, we use first-principles calculations and machine learning techniques to investigate monolayers of the form $A^iA^{ii}B_4X_8$ based on the well-known intrinsic topological magnetic layered material $MnBi_2Te_4$. We consider a very large number of candidate materials ($\sim 10^4$) formed by making chemical substitutions at A, B, and X sites. We select an initial subset of 240 structures and study their thermodynamic (*i.e.* formation energy), electronic (*i.e.* band gap), and magnetic properties (*i.e.* magnetic moment, magnetic order) using DFT calculations. Data analytics methods are employed to gain insight into the microscopic origin of materials' properties. Finally, we use ML tools to predict novel materials from the MBT family having desirable properties, such as materials with large magnetic moments and high chemical stability. This study creates avenues for the design of novel materials that can enable technological innovation.

Methods

First, we generate a database of candidate structures of the type $A^iA^{ii}B_4X_8$ based on the parent material $MnBi_2Te_4$ with different elements occupying A, B, and X sites. In our study, we consider 20 transition metals at the A site, 4 elements from group IVA–VA at the B sites, and 3 chalcogens at the X site, yielding 12,600 $A^iA^{ii}B_4X_8$ candidate materials that comprise our chemical space. Initially, we consider a small subset of 240 structures for DFT calculations. The unit cell of $MnBi_2Te_4$ consists of 7 atoms. We use a $2 \times 1 \times 1$ supercell in the calculation to ensure that a minimum number of sites are available for making composite materials from the parent structure. The chemical formula of composite materials thus becomes $A^iA^{ii}B_4X_8$. All the structures considered in DFT calculations have an Mn atom at one of the two A sites of $A^iA^{ii}B_4X_8$ (eg. $(MnCr)Sb_4Te_8$). Geometrical optimization is performed with spin-polarization and spin–orbit interactions. We use VASP[53] to perform first-principles calculations with projected augmented wave pseudopotential[54] along with the GGA-PBE type[55] of exchange and correlation functional. The Hubbard U parameter ($U = 4$ eV) is introduced to consider the localized 3d-states of all the transition metals.[56] A plane wave basis set with a kinetic energy cut-off of 450 eV is used. The electronic and force convergence criteria are set to be 10^{-6} eV and 10^{-2} eV Å^{-1} between any two successive SCF steps respectively. The Gaussian smearing method with a width of 0.03 eV is used in the calculations. Gamma-centered k-points mesh of $5 \times 10 \times 1$ is used for the integration of the Brillouin zone. A vacuum region of more than 30 Å is chosen to avoid the interaction between any two adjacent layers. Using DFT, we calculate the total ground state energy, electronic band gap, magnetic moment, and magnetic order of the initial subset of structures. We also calculate the formation energy and use it as a proxy for the chemical stability of a material. It is the difference in energy between the chemical compound and individual elemental phases constituting that compound, which, for $A^iA^{ii}B_4X_8$ structures, is given as,

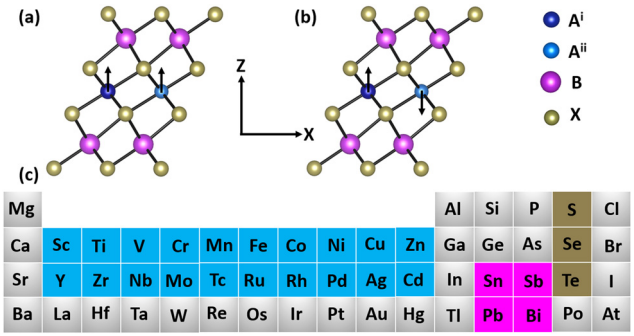

Fig. 1 The crystal structure of the $A^iA^{ii}B_4X_8$ (a) FM and (b) AFM state. (c) Elements that are considered for the substitution on A (blue), B (magenta), and X (tan) sites.

$$E_{form} = E(A^iA^{ii}B_4X_8) - E(A^i) - E(A^{ii}) - 4 \times E(B) - 8 \times E(X)$$

The corresponding energies are obtained from the DFT calculations. It should be noted that the materials with the lowest energy spin configurations are considered when calculating the formation energy.

To train the ML models on the DFT-calculated data, we construct a set of materials descriptors using atomic properties obtained from the Mendeleev python package.[57] We use a total of 55 descriptors built from 11 different atomic properties including dipole polarizability, ionization energy, electron affinity, covalent radius, number of valence electrons, and number of unpaired electrons (see ESI† for detailed information). These descriptors are the inputs of the ML models and DFT-calculated formation energy, magnetic moment, and band gap are the target variables of the ML models.

To train the machine learning model, we randomly split our dataset into a training set, a validation set, and a test set in the ratio of 0.6 : 0.2 : 0.2 respectively. Different types of supervised ML models are then employed to predict the materials' properties. For instance, random forest regression and extra trees regression are used for predicting the formation energy, whereas random forest regression is used for making predictions on the magnetic moment, and band gap.[58] The random seed is not kept fixed in the random forest regression models. Also, we use K-fold cross-validation with $K = 10$ while tuning the hyperparameters. See ESI† for details on the hyperparameters used for each ML model. These models are useful for making physical interpretations of the predictions because they allow us to extract the relative importance of the features used to make predictions (Fig. 1).[58,59]

Results and discussion

Workflow overview

The workflow of data-driven virtual screening of the magnetic vdW materials is presented in Fig. 2. We start with a monolayer of MBT as the archetypal material. Then we make chemical substitutions at the A, B, and X sites of $MnBi_2Te_4$, with one of the two A sites (*i.e.* A^i) unchanged, to get a subset of the chemical space for DFT calculations. 20 transition metals are chosen for the A^{ii} sites (making MnA^{ii}, *e.g.* MnCr, MnNi), 4

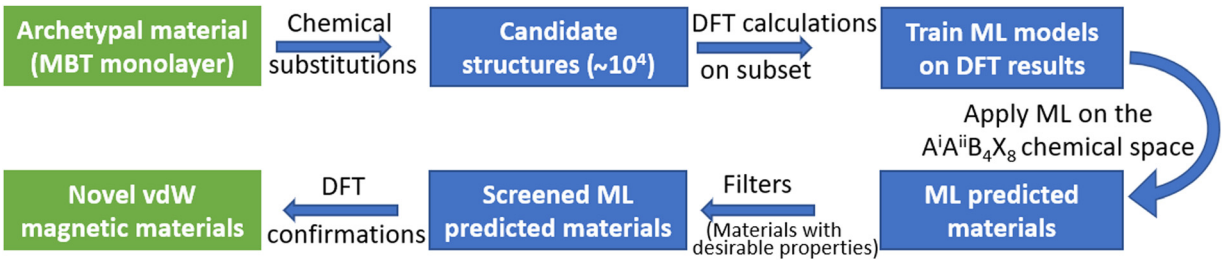

Fig. 2 The workflow of data-driven virtual screening of magnetic vdW materials.

group IVA–VA elements for the B sites (*i.e.* Bi_2, Sb_2, Sn_2, and Pb_2) and 3 chalcogens at the X sites (*i.e.* S_4, Se_4, and Te_4) are used for the substitutions. This generates a total of 240 candidate structures. Next, we perform DFT calculations on those structures and apply data-analytics techniques to analyze the DFT results. Our data-analytics method focuses on the following materials' properties: formation energy, magnetic moment, magnetic states, and band gap. We train ML models and then use them to predict the properties of the entire materials space of $A^iA^{ii}B_4X_8$ structures consisting of 12,360 candidate materials. Next, we apply three screening criteria on the ML predicted materials properties: formation energy (E_f) < 0 eV, (ii) magnetic moment (μ) > 4.55 μ_B, and (iii) band gap (E_g) > 0 eV. These filters will search for stable magnetic insulators from the materials space of $A^iA^{ii}B_4X_8$. Finally, the candidate materials that are passed through the filters are tested *via* DFT calculations. The dynamic stability tests are performed on the structures that are confirmed by DFT calculations.

Formation energy

Calculating formation energy is the very first step in screening the materials in our work. It is used as a proxy for the chemical stability of a material. A plot of the formation energies of 240 structures considered in the initial calculations is shown in Fig. 3(a). In the calculations, one of the two A sites (*i.e.* A^i) is always fixed to Mn. It is evident that the cases where the elements Sc, Zn, Y, and Cd are on the A^{ii} site have the lowest formation energies than any other substitutions. This behavior also resembles that of $Cr_2Ge_2Te_6$ in Ref. 37. The structures comprising certain transition metals such as Mo, Tc, Ru, Rh on the A^{ii} site tend to have larger formation energy implying less stability. The number of electrons present in their d-orbitals is associated with this behavior (see ESI†). Also, a trend of slightly decreasing formation energy is seen as we go up the column from Te to Se or S at X sites. For any given A and X sites, the structures with Bi on the B site tend to have lower energies. There is a trend of increasing formation energy for the B site as follows: Bi < Sn < Pb < Sb.

Magnetic moment

Calculating magnetic moments is another important step in our materials search. We aim to find materials that have ferromagnetic spin configurations as well as large magnetic moments. Fig. 3(b) shows a heatmap of the DFT-calculated

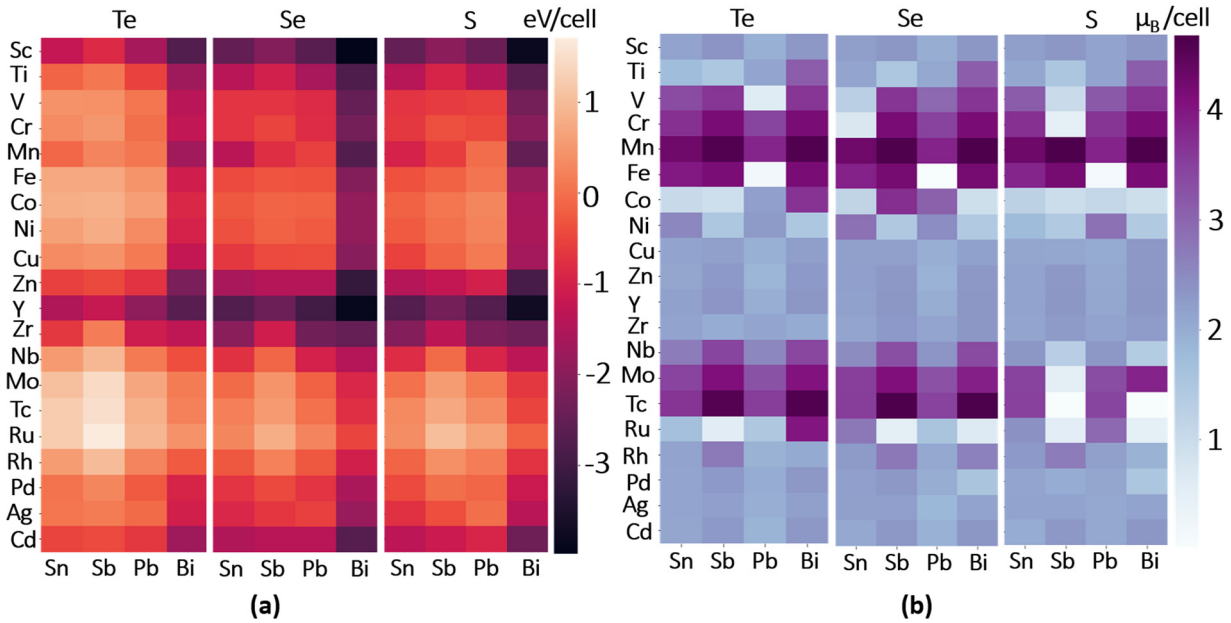

Fig. 3 (a) Formation energy per unit cell (in eV) and (b) magnetic moment per unit cell (in μ_B) of $A^iA^{ii}B_4X_8$ structures at the lowest energy states.

magnetic moment per unit cell for the 240 candidate structures in their lowest energy spin configurations. Since only the A site (A^{i} and A^{ii}) of MBT contains magnetic elements, the magnetic moments depend mainly on the A^{ii} site (A^{i} is fixed to Mn). A pattern in Fig. 3(b) emerges: as we go from left to right of the first two rows of transition metals in the periodic table, the structures comprising the elements such as V, Cr, Mn, Fe have larger (or nearly zero) magnetic moments if there is FM (AFM) ordering whereas the elements such as Sc, Ti, Cu, Zn have significantly smaller values. This is consistent with the total number of unpaired electrons in the atoms: the elements with the higher (lower) number of unpaired electrons have larger (smaller) magnetic moments (see ESI†). Also, elements like Sc, Y, Zr, Cu, Zn, and Ag are non-magnetic, substituting these in the A^{ii} site will reduce the magnetic moment to around 2.2 μ_B per unit cell. The white blocks represent the AFM structures having magnetic moments close to zero. In addition, some of the structures exhibit ferrimagnetism with the magnetic moment between 0 and 2 μ_B such as $(MnNi)Sb_4Te_8$, $(MnCo)Sn_4Se_8$, and $(MnV)Sb_4S_8$.

Interestingly, the elements on the B sites play a role in determining the magnetic moment, and magnetic order of the structures although their contribution is not highly significant. Relatively larger magnetic moments are found in the cases where Bi and Sb are on the B site (*i.e.* $B_2 = Bi_2, Sb_2$). We expect that structures with both Bi and Sb at the B site also have similar magnetic moments (*i.e.* B_2 = BiSb). On the other hand, Sn and Pb at the B site tend to lower each atom's local magnetic moments, resulting in smaller magnetic moments. This emphasizes the role of additional 5p (unpaired) electrons in group VA elements that are absent in group IVA. The magnetic moment of pure MBT monolayer is 4.62 μ_B per unit cell. We find that the following four additional candidate materials have slightly larger magnetic moments than pure MBT monolayer: $Mn_2Bi_4S_8$ (4.66 μ_B), $Mn_2Bi_4Se_8$ (4.65 μ_B), $Mn_2Sb_4S_8$ (4.65 μ_B), and $Mn_2Sb_4Se_8$ (4.64 μ_B).

Magnetic spin configurations

A plot of magnetic spin configurations of 240 candidate structures of $A^{i}A^{ii}B_4X_8$ is shown in Fig. 4(a). The blue and yellow colors represent the respective FM and AFM spin ordering in their ground states. Most of the structures with Co, Ni, Ru on the A^{ii} site strongly prefer AFM configurations irrespective of B and X sites whereas those comprising Mn, and Tc prefer FM configurations. A total of 182 out of 240 structures have FM and the remaining 58 have AFM spin configurations. Our analysis of magnetic ordering gives an overall scenario of how the magnetic spins are aligned in the candidate materials depending upon the atomic substitutions at different sites. This helps us to search for structures in the chemical space of $A^{i}A^{ii}B_4X_8$ with specific magnetic configurations.

Band gap

The electronic properties of the candidate structures are also investigated. The electronic band gaps are calculated using the pymatgen package.[60] The calculated electronic band gaps are shown in Fig. 4(b). Analysis of electronic band gaps is an important step to search for candidate materials in the family of MBT monolayers that exhibit topological properties. Since standard DFT is known to underestimate the band gap,[61] we use the DFT+U method to better approximate the band gaps. We expect that the results will also be useful for the comparative study of the family of MBT materials. The exact band gap calculation requires computationally very expensive methods such as hybrid-DFT (HSE),[62,63] many-body green function (GW),[64–66] and Bethe–Salpeter equations (BSE) approaches,[67,68] which are not suitable for high-throughput calculations. All the structures comprising the elements from group IVA, namely, Sn and Pb, at

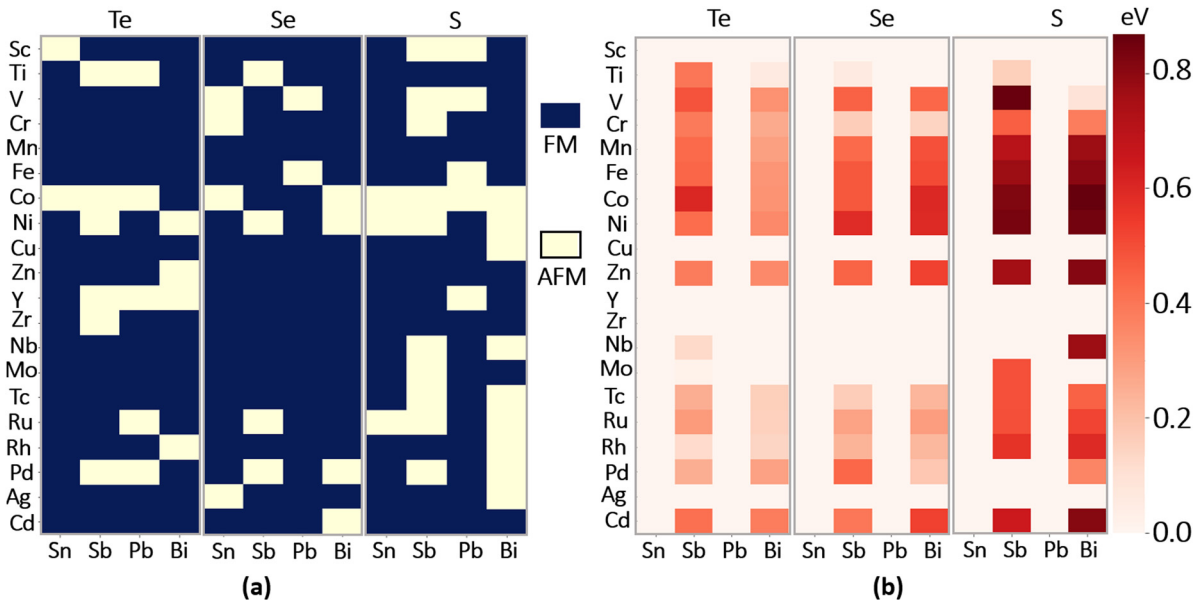

Fig. 4 (a) Magnetic spin states (FM/AFM) and (b) electronic band gap (in eV) of $A^{i}A^{ii}B_4X_8$ structures at the lowest energy states.

the B site show zero band gap hence the metallic behavior irrespective of A and X sites. But those with Sb and Bi at the B site are mostly semiconducting and show an increasing trend of the band gap as we go up from Te to Se or S at X sites. This implies that the band gap is sensitive to the occupancy of A, B, and X sites.

Next, we apply the following three filters on the materials in the DFT dataset: (i) formation energy (E_f) < 0 eV, (ii) magnetic moment (μ) > 4.55 μ_B, and (iii) band gap (E_g) > 0 eV. The first filter, E_f, is a proxy for chemical stability. It is a necessary but not sufficient condition for chemical stability. The second filter, μ, screens for the materials with magnetic moments greater than or equal to that of monolayer $MnBi_2Te_4$. This also ensures that the screened materials are ferromagnetic with large magnetic moments. The third filter is applied to exclude the candidate materials that are metallic. Since the topologically nontrivial insulators tend to have non-zero band gaps, we expect that the last filter may help to identify novel materials in the $MnBi_2Te_4$ family that have topological properties. We find that five candidate structures satisfy these criteria including one with Tc on A^{ii} site. Since Tc is a radioactive element, we do not consider the structure comprising Tc as a viable material for further investigation. These structures are $Mn_2Bi_4S_8$, $Mn_2Bi_4Se_8$, $Mn_2Sb_4S_8$, and $Mn_2Sb_4Se_8$. Their corresponding values are presented in Table 1.

It should be noted that the monolayer $Mn_2Sb_4S_8$ and few-layer $Mn_2Bi_4Se_8$ have been successfully synthesized in experiments.[69–73] While the monolayers of $Mn_2Bi_4Se_8$, and $Mn_2Sb_4Se_8$ are predicted using first-principles calculations in the past,[74] the bulk phase of $Mn_2Bi_4S_8$, and a quasi-one-dimensional phase of $Mn_2Sb_4Se_8$ are also experimentally synthesized.[75,76] This supports the ability of DFT methods to discover novel materials.

ML models for high-throughput screening

We train ML models on the DFT data aiming to facilitate a better understanding of the structure–property relationships so that this relationship can be exploited for quantitative predictions. ML will then be used to screen the entire space of $A^iA^{ii}B_4X_8$ materials. We use two non-linear regression models to predict the formation energy of the materials. Fig. 5(a) and (b) represent the performance of extra trees regression and random forest regression on predicting the formation energy of $A^iA^{ii}B_4X_8$ structures respectively. The models' performance is measured in terms of R^2 and mean absolute error (MAE). It is evident that both models work very well for predicting the formation energy, indicated by the high (R^2) test scores. We also examine the performance of random forest regression on the test data with the size of the training set in Fig. 5(c). The model performance increases with the training set size at first, then reaches almost a plateau region after the training set size exceeds 96 data points (40% of the DFT-generated dataset). The corresponding highest test score (R^2) is 0.92 with an MAE of 0.27 eV.

The random forest descriptor importances for the top descriptors for formation energy is shown in Fig. 5(d). This shows that certain descriptors such as atomic volume, electron affinity, electronegativity, dipole polarizability, and covalent radius are the most important for predicting the formation energy. The larger the difference in atomic volume between B and X sites, the more negative the formation energy (indicating higher stability). Similarly, the greater the difference in electronegativity between two A sites, the more positive the formation energy is (indicating lower stability. The structures with higher electronegativity, a tendency to make chemical bonds, are more chemically stable. This means the structures comprising two completely different elements on A sites (in terms of electronegativity) are less chemically stable.

The random forest model performance for the magnetic moment of $A^iA^{ii}B_4X_8$ structures is shown in Fig. 6(a). The model has a test score (R^2) of 0.59 and MAE of 0.35 μ_B. The model performance here is lower than that of the formation energy. This is likely due to the fact that the magnetic moment is a complex property whose behavior is not fully captured within the descriptors used. For instance, the magnetic moment is sensitive to the position of atoms within the structures because of the direct and indirect exchange interactions. However, the descriptors used do not directly include the atomic positions. Also, the model performance is impacted by the distribution of the data. For instance, a large portion of the data has μ near

Table 1 Predicted candidate structures. DFT and ML represent the structures that are predicted from DFT calculations and machine learning, respectively

Chemical formula	Predicted from	Magnetic state	Formation energy (eV per unit cell)	Magnetic moment (μ_B per unit cell)	Band gap (eV)
$Mn_2Bi_4S_8$[a76]	DFT	FM	−2.511	4.655	0.732
$Mn_2Bi_4Se_8$[a69,70,74]	DFT	FM	−2.751	4.651	0.493
$Mn_2Sb_4S_8$[a71–73]	DFT	FM	−0.601	4.652	0.707
$Mn_2Sb_4Se_8$[a74,75]	DFT	FM	−0.783	4.639	0.431
$Mn_2Bi_4S_4Te_4$[b84,85]	ML	FM	−1.522	4.613	0.169
$Mn_2Bi_4Se_4Te_4$[b84,85]	ML	FM	−1.928	4.619	0.178
$Mn_2Sb_2Bi_2Te_8$[b86]	ML	FM	−0.807	4.641	0.278
$Mn_2Sb_2Bi_2S_4Se_4$	ML	FM	−1.558	4.641	0.465
$Mn_2Sb_2Bi_2S_4Te_4$	ML	FM	−0.627	4.611	0.252
$Mn_2Sb_2Bi_2Se_4Te_4$	ML	FM	−1.024	4.618	0.249
$Mn_2Sb_2Bi_2Se_8$	ML	FM	−1.856	4.644	0.335
$Mn_2Sb_4Se_4Te_4$	ML	FM	−0.137	4.616	0.493
$Mn_2Bi_4S_4Se_4$	ML	FM	−2.439	4.645	0.533

[a] Structures that are synthesized in the experiment. [b] Structures predicted from DFT calculations.

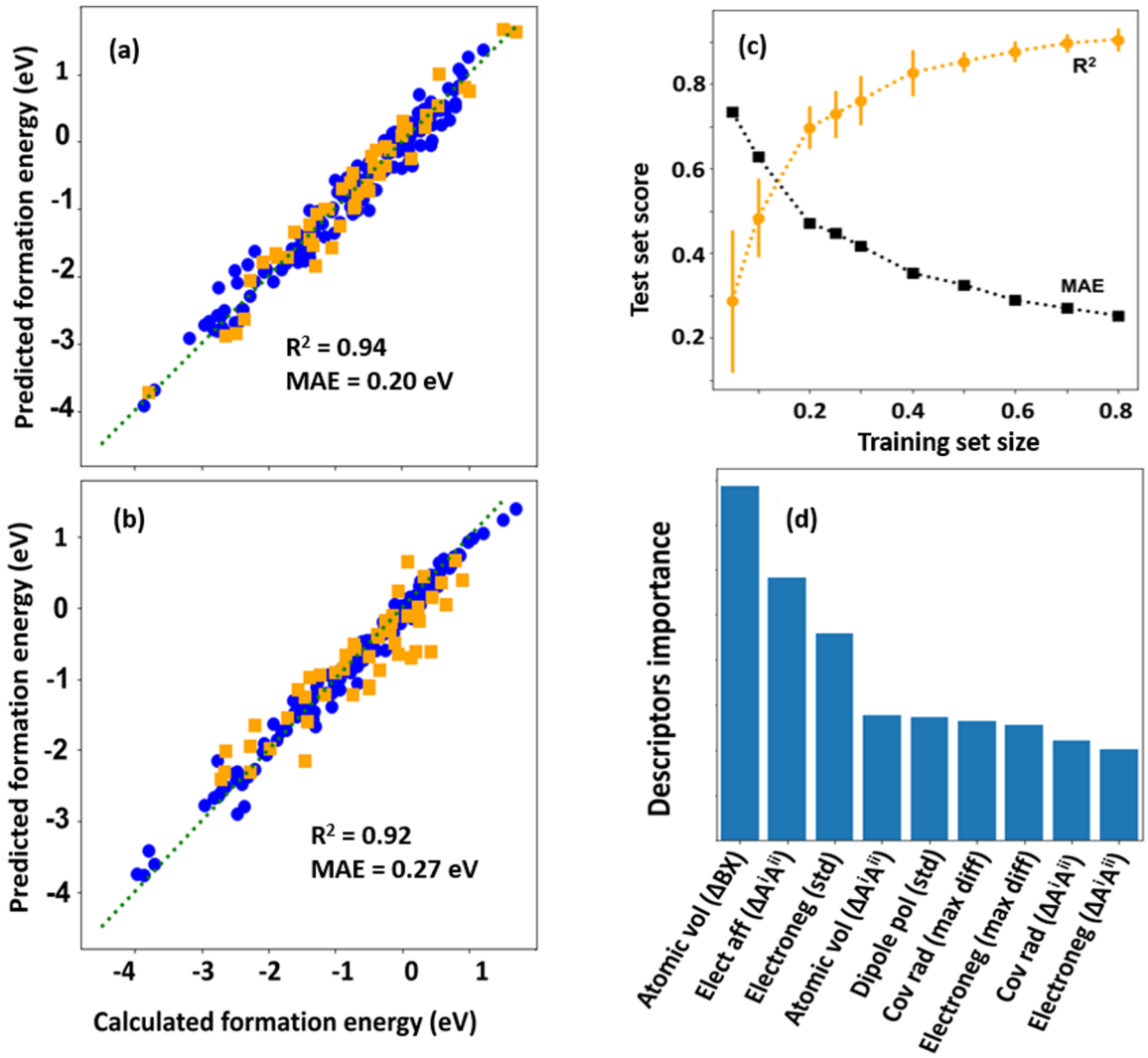

Fig. 5 ML prediction of formation energy (in eV) of $A^iA^{ii}B_4X_8$ structures (a) extra trees regression (b) random forest regression. Training and test data are displayed in blue circles and orange squares respectively. (c) Performance of random forest regression on the test data with the size of the training set (d) Top 9 descriptors for the formation energy prediction by random forest regression.

2 μ_B (see ESI† for more details). To more accurately predict μ, we need a sophisticated set of descriptors. Examples include, CGCNN,[77] and SOAP.[78] However, these models are more complex and less physically interpretable. So, we choose a less sophisticated interpretable model for our analysis. We are able to employ this model for making reasonably accurate predictions on magnetic moments of the new structures, which is discussed below. The top descriptors are the average number of unpaired electrons, average electron affinity, and standard deviation of valence electrons. Since the total magnetic moment comprises the local magnetic dipole moments of individual spins, a larger (smaller) number of spin-up electrons will yield a higher (lower) magnetic moment. Among the elements on A sites, Mn and Tc have the highest number of unpaired electrons. So we expect the structures comprising these elements to have the largest magnetic moments (see ESI† for the most important descriptors).

Fig. 6(b) shows the performance of random forest regression on band gap prediction. The model performs well with a test score (R^2) of 0.70 and MAE of 0.08 eV. The performance is not as good as formation energy but better than the magnetic moment. The average number of unpaired electrons, average dipole polarizability, average number of valence electrons, and average ionization energy are some of the most important descriptors. For example, a higher ionization energy implies that the electrons are more tightly bound to the nucleus and thus are not easily liberated from the atom. This results in a larger separation of the valence band and conduction band in solids thereby increasing the band gap. Also, the materials with larger (smaller) dipole polarizabilities have smaller (larger) band gaps. However, the reverse trend is observed in the case of unpaired electrons (see ESI† for additional details). Several studies have already shown that data-driven approaches can be used to efficiently predict the band gap of a variety of materials.[79–83]

After training the ML models, they are then used to make predictions on the entire space of $A^iA^{ii}B_4X_8$ materials containing 12,360 structures that are not included in the original dataset of

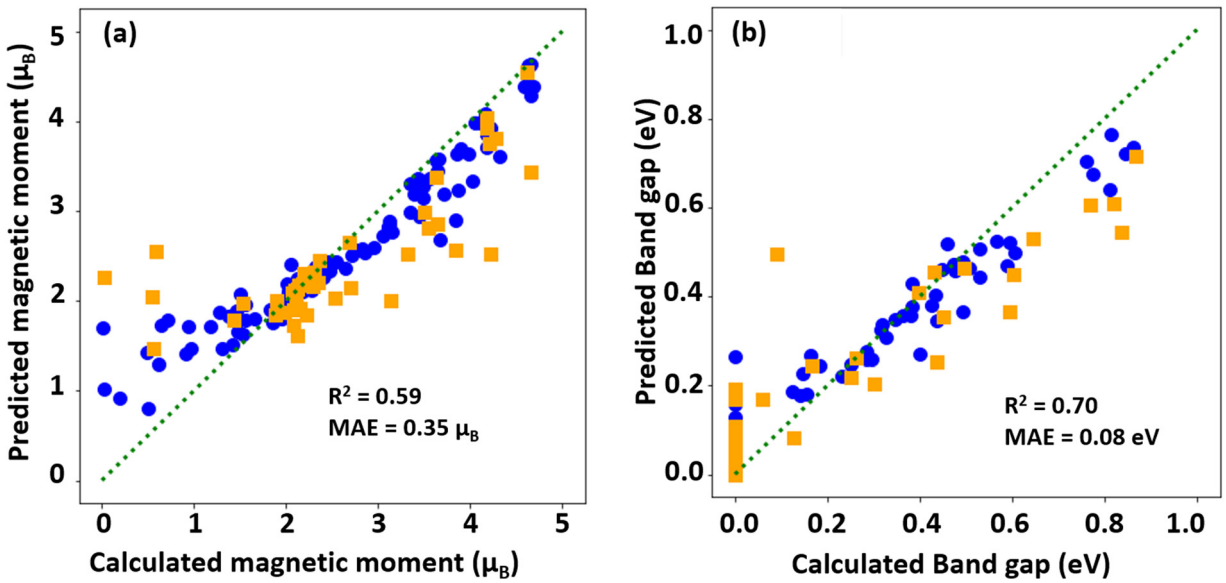

Fig. 6 ML prediction of $A^iA^{ii}B_4X_8$ structures (a) random forest regression on the magnetic moment (in μ_B) (b) random forest regression on band gap (in eV). Training and test data are displayed in blue circles and orange squares and red circles respectively. The dashed line (green) is a guide to the eye and represents the perfect prediction line.

240 structures used for DFT calculations. Next, we apply the following three filters on the materials in the resulting dataset: (i) formation energy (E_f) < 0 eV, (ii) magnetic moment (μ) > 4.55 μ_B, and (iii) band gap (E_g) > 0 eV. After applying the screening criteria to the ML predictions, we find 25 promising materials. Three of them contain technetium (Tc). Being a radioactive element, we exclude the structures comprising Tc during the final DFT simulations. We then check these 22 predictions with further DFT calculations and find that 9 of them satisfy all the criteria. Interestingly, 3 of these 9 structures are already reported as ferromagnetic vdW Janus materials. They are $Mn_2Bi_4S_4Te_4$,[84,85] $Mn_2Bi_4Se_4Te_4$,[84,85] and $Mn_2Sb_2Bi_2Te_8$.[86] This validates our data-driven method for discovering novel materials. All the predicted candidate structures (9 ML-predicted and 4 DFT-calculated) are presented in Table 1.

We find that our data-driven approach is successful in identifying the candidate materials in the MBT system that have not yet been reported in the literature. The ML model performance can be determined by the R^2 scores and the MAE. Using a random forest model we find an R^2 score (MAE) of 0.92 (0.27 eV) for the formation energy, 0.59 (0.35 μ_B) for the magnetic moment, and 0.70 (0.08 eV) for the band gap. These scores indicate that the models are performing well. We report a materials screening success rate of 41% (9 out of 22) that includes three criteria - formation energy, magnetic moment, and band gap. A 100% success rate would imply that ML predictions exactly match DFT calculations. Our result does not imply that the models' performance is poor. The metric for success of the screening is determined by a combination of the strict (*i.e.* well defined cutoff) screening criteria, model performance error (small in our case), and the number of screening criteria used.

To further test the stability of our best candidates, we test the dynamical stability by calculating the phonon spectra using the Phonopy package.[87] The stability is also verified by *ab initio* molecular dynamics simulations at 500 K (see ESI† for detailed information). We find that $Mn_2Bi_4S_8$, $Mn_2Sb_2Bi_2S_4Se_4$, $Mn_2Sb_2Bi_2Se_4Te_4$, $Mn_2Sb_2Bi_2Se_8$, $Mn_2Sb_4Se_4Te_4$, and $Mn_2Bi_4S_4Se_4$ are dynamically stable. In addition to the chemical stability tests performed in our study, we note that calculating the energy above the convex hull of the candidate materials could provide further evidence for assessing their synthesizability. Although, satisfying the convex hull test does not guarantee that the materials can be synthesized. Convex hull calculations involve exploring all the possible competing phases in the same materials system and comparing their energies. However, when dealing with complex materials like MBT from ternary, quaternary, and higher-order space, finding the competing phases is complicated because the number of possible phases increases tremendously with the number of constituent elements.[88] For this reason, we believe that identifying these phases is outside the scope of this work. Nevertheless, investigating competing phases is a worthwhile endeavor that we will plan to explore in a future publication. Furthermore, these results will provide the impetus for other researchers to carry out convex hull calculations, which are important metrics for chemical stability.

All the predicted $A^iA^{ii}B_4X_8$ structures have Mn at both A^i and A^{ii} sites. This is the result of the magnetic filter that searches for structures with the highest magnetic moments. Mn has the highest number of spin-up electrons, which is the major deciding factor for magnetic moment. Although the elements occupying B and X sites are non-magnetic, they possess minimal magnetic moment values resulting from the indirect exchange interaction with the magnetic elements at A sites. We find that the local magnetic moments at X sites are always negative for X = Te. However, this doesn't hold true when S and Se are substituted at X sites. In fact, they possess smaller (in

magnitude) positive magnetic moments as opposed to the former case. That's why the structures (Table 1) comprising Te at the X sites have relatively smaller magnetic moments as compared to the ones without Te. We expect that this behavior is associated with the large-spin–orbit coupling in Te as compared to S and Se. Also, all of the predicted structures are ferromagnetic.

$Mn_2Sb_2Bi_2S_4Se_4$, $Mn_2Sb_2Bi_2S_4Te_4$, $Mn_2Sb_2Bi_2Se_4Te_4$, $Mn_2Sb_2Bi_2Se_8$ are Janus 2D materials. They are a special class of materials showing exotic physical phenomena such as varying electronic band gap,[89,90] the Rashba effect,[91,92] and piezoelectricity,[86,89,90] which makes them promising candidates for numerous potential applications including photo-catalytic water splitting,[89,93] hydrogen-evolution reaction,[94,95] and sensing devices.[96,97] We believe these candidate materials will spark novel and emergent phenomena in 2D Janus materials. In addition, MBT-type materials are well-known for their rich topological properties.[6,17,26,30,31] Thus, our work creates avenues for the future study of the topological properties of the most promising candidate structures identified in this study. However, this is beyond the scope of the present study.

Conclusions

First-principles calculations combined with machine learning techniques are used to investigate monolayers of the form $A^iA^{ii}B_4X_8$ based on the intrinsic topological magnetic van der Waals material $MnBi_2Te_4$. DFT is used to investigate a small set of structures. The formation energy, band gap, magnetic moment, and magnetic order are calculated. Data analytics approaches are used to gain insight into the microscopic origin of materials' properties. We find that the formation energy and magnetic moment depend largely on A and B sites whereas the band gap depends on all three sites. Finally, we employ machine learning tools to search for novel materials within the family of MBT structures with desirable properties. This study creates avenues for predicting novel materials with desirable properties that have applications in spintronics, optoelectronics, and quantum computing.

Author contributions

T. R. conceived the idea of the project and assisted in writing the manuscript. R. B. performed the calculations, conducted the data analysis, and wrote the manuscript. P. M. assisted with the data analysis.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), supported by National Science Foundation grant number ACI-1548562. In addition, we used the resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. This material is based upon work supported by the NSF CAREER award under grant number 2044842.

References

1 K. S. Novoselov, A. K. Geim, S. V. Morozov, D.-E. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva and A. A. Firsov, *Science*, 2004, **306**, 666–669.
2 J. A. Wilson and A. Yoffe, *Adv. Phys.*, 1969, **18**, 193–335.
3 Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman and M. S. Strano, *Nat. Nanotechnol.*, 2012, **7**, 699–712.
4 M. Buscema, J. O. Island, D. J. Groenendijk, S. I. Blanter, G. A. Steele, H. S. van der Zant and A. Castellanos-Gomez, *Chem. Soc. Rev.*, 2015, **44**, 3691–3718.
5 C. Anichini, W. Czepa, D. Pakulski, A. Aliprandi, A. Ciesielski and P. Samor, *Chem. Soc. Rev.*, 2018, **47**, 4860–4908.
6 J. Li, Y. Li, S. Du, Z. Wang, B.-L. Gu, S.-C. Zhang, K. He, W. Duan and Y. Xu, *Sci. Adv.*, 2019, **5**, eaaw5685.
7 D. Zhong, K. L. Seyler, X. Linpeng, R. Cheng, N. Sivadas, B. Huang, E. Schmidgall, T. Taniguchi, K. Watanabe and M. A. McGuire, *et al.*, *Sci. Adv.*, 2017, **3**, e1603113.
8 M. Z. Hasan and C. L. Kane, *Rev. Mod. Phys.*, 2010, **82**, 3045.
9 X.-L. Qi and S.-C. Zhang, *Rev. Mod. Phys.*, 2011, **83**, 1057.
10 S. Haastrup, M. Strange, M. Pandey, T. Deilmann, P. S. Schmidt, N. F. Hinsche, M. N. Gjerding, D. Torelli, P. M. Larsen and A. C. Riis-Jensen, *et al.*, *2D Mater.*, 2018, **5**, 042002.
11 J. Zhou, L. Shen, M. D. Costa, K. A. Persson, S. P. Ong, P. Huck, Y. Lu, X. Ma, Y. Chen and H. Tang, *et al.*, *Sci. Data*, 2019, **6**, 1–10.
12 C. Gong, L. Li, Z. Li, H. Ji, A. Stern, Y. Xia, T. Cao, W. Bao, C. Wang and Y. Wang, *et al.*, *Nature*, 2017, **546**, 265–269.
13 B. Huang, G. Clark, E. Navarro-Moratalla, D. R. Klein, R. Cheng, K. L. Seyler, D. Zhong, E. Schmidgall, M. A. McGuire and D. H. Cobden, *et al.*, *Nature*, 2017, **546**, 270–273.
14 M. Bonilla, S. Kolekar, Y. Ma, H. C. Diaz, V. Kalappattil, R. Das, T. Eggers, H. R. Gutierrez, M.-H. Phan and M. Batzill, *Nat. Nanotechnol.*, 2018, **13**, 289–293.
15 Z. Fei, B. Huang, P. Malinowski, W. Wang, T. Song, J. Sanchez, W. Yao, D. Xiao, X. Zhu and A. F. May, *et al.*, *Nat. Mater.*, 2018, **17**, 778–782.
16 J.-U. Lee, S. Lee, J. H. Ryoo, S. Kang, T. Y. Kim, P. Kim, C.-H. Park, J.-G. Park and H. Cheong, *Nano Lett.*, 2016, **16**, 7433–7438.
17 Y. Gong, J. Guo, J. Li, K. Zhu, M. Liao, X. Liu, Q. Zhang, L. Gu, L. Tang and X. Feng, *et al.*, *Chin. Phys. Lett.*, 2019, **36**, 076801.
18 M. M. Otrokov, I. I. Klimovskikh, H. Bentmann, D. Estyunin, A. Zeugner, Z. S. Aliev, S. Gaß, A. Wolter, A. Koroleva and A. M. Shikin, *et al.*, *Nature*, 2019, **576**, 416–422.
19 E. D. Rienks, S. Wimmer, J. Sánchez-Barriga, O. Caha, P. S. Mandal, J. Růžicka, A. Ney, H. Steiner, V. V. Volobuev and H. Groiß, *et al.*, *Nature*, 2019, **576**, 423–428.
20 S. H. Lee, Y. Zhu, Y. Wang, L. Miao, T. Pillsbury, H. Yi, S. Kempinger, J. Hu, C. A. Heikes and P. Quarterman, *et al.*, *Phys. Rev. Res.*, 2019, **1**, 012011.

21 J. Wu, F. Liu, M. Sasase, K. Ienaga, Y. Obata, R. Yukawa, K. Horiba, H. Kumigashira, S. Okuma and T. Inoshita, *et al.*, *Sci. Adv.*, 2019, **5**, eaax9989.
22 B. Chen, F. Fei, D. Zhang, B. Zhang, W. Liu, S. Zhang, P. Wang, B. Wei, Y. Zhang and Z. Zuo, *et al.*, *Nat. Commun.*, 2019, **10**, 1–8.
23 J.-Q. Yan, Q. Zhang, T. Heitmann, Z. Huang, K. Chen, J.-G. Cheng, W. Wu, D. Vaknin, B. C. Sales and R. J. McQueeney, *Phys. Rev. Mater.*, 2019, **3**, 064202.
24 H. Li, S. Liu, C. Liu, J. Zhang, Y. Xu, R. Yu, Y. Wu, Y. Zhang and S. Fan, *Phys. Chem. Chem. Phys.*, 2020, **22**, 556–563.
25 P. M. Sass, W. Ge, J. Yan, D. Obeysekera, J. Yang and W. Wu, *Nano Lett.*, 2020, **20**, 2609–2614.
26 Y. Deng, Y. Yu, M. Z. Shi, Z. Guo, Z. Xu, J. Wang, X. H. Chen and Y. Zhang, *Science*, 2020, **367**, 895–900.
27 C. Liu, Y. Wang, H. Li, Y. Wu, Y. Li, J. Li, K. He, Y. Xu, J. Zhang and Y. Wang, *Nat. Mater.*, 2020, **19**, 522–527.
28 J. Ge, Y. Liu, J. Li, H. Li, T. Luo, Y. Wu, Y. Xu and J. Wang, *Natl. Sci. Rev.*, 2020, **7**, 1280–1287.
29 D. S. Lee, T.-H. Kim, C.-H. Park, C.-Y. Chung, Y. S. Lim, W.-S. Seo and H.-H. Park, *CrystEngComm*, 2013, **15**, 5532–5538.
30 M. Otrokov, I. P. Rusinov, M. Blanco-Rey, M. Hoffmann, A. Y. Vyazovskaya, S. Eremeev, A. Ernst, P. M. Echenique, A. Arnau and E. V. Chulkov, *Phys. Rev. Lett.*, 2019, **122**, 107202.
31 D. Zhang, M. Shi, T. Zhu, D. Xing, H. Zhang and J. Wang, *Phys. Rev. Lett.*, 2019, **122**, 206401.
32 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner and G. Ceder, *et al.*, *APL Mater.*, 2013, **1**, 011002.
33 R. T. Downs and M. Hall-Wallace, *Am. Mineral.*, 2003, **88**, 247–250.
34 T. Ueno, T. D. Rhone, Z. Hou, T. Mizoguchi and K. Tsuda, *Mater. Discovery*, 2016, **4**, 18–21.
35 Y. Zhu, X. Kong, T. D. Rhone and H. Guo, *Phys. Rev. Mater.*, 2018, **2**, 081001.
36 E. O. Pyzer-Knapp, J. W. Pitera, P. W. Staar, S. Takeda, T. Laino, D. P. Sanders, J. Sexton, J. R. Smith and A. Curioni, *npj Comput. Mater.*, 2022, **8**, 1–9.
37 T. D. Rhone, W. Chen, S. Desai, S. B. Torrisi, D. T. Larson, A. Yacoby and E. Kaxiras, *Sci. Rep.*, 2020, **10**, 1–11.
38 M. C. Sorkun, S. Astruc, J. Koelman and S. Er, *npj Comput. Mater.*, 2020, **6**, 1–10.
39 K. Takahashi, L. Takahashi, I. Miyazato and Y. Tanaka, *ACS Photonics*, 2018, **5**, 771–775.
40 R. K. Vasudevan, K. Choudhary, A. Mehta, R. Smith, G. Kusne, F. Tavazza, L. Vlcek, M. Ziatdinov, S. V. Kalinin and J. Hattrick-Simpers, *MRS commun.*, 2019, **9**, 821–838.
41 D. Torelli, K. S. Thygesen and T. Olsen, *2D Mater.*, 2019, **6**, 045018.
42 D. Torelli, H. Moustafa, K. W. Jacobsen and T. Olsen, *npj Comput. Mater.*, 2020, **6**, 158.
43 N. Mounet, M. Gibertini, P. Schwaller, D. Campi, A. Merkys, A. Marrazzo, T. Sohier, I. E. Castelli, A. Cepellotti and G. Pizzi, *et al.*, *Nat. Nanotechnol.*, 2018, **13**, 246–252.
44 Z.-X. Shen, C. Su and L. He, *npj Comput. Mater.*, 2022, **8**, 132.
45 A. Kabiraj, M. Kumar and S. Mahapatra, *npj Comput. Mater.*, 2020, **6**, 35.
46 Y. Xie, G. A. Tritsaris, O. Granas and T. D. Rhone, *J. Phys. Chem. Lett.*, 2021, **12**, 12048–12054.
47 A.-Y. Lu, H. Zhu, J. Xiao, C.-P. Chuu, Y. Han, M.-H. Chiu, C.-C. Cheng, C.-W. Yang, K.-H. Wei and Y. Yang, *et al.*, *Nat. Nanotechnol.*, 2017, **12**, 744–749.
48 G. H. Gu, J. Noh, I. Kim and Y. Jung, *J. Mater. Chem. A*, 2019, **7**, 17096–17117.
49 S. K. Kauwe, T. D. Rhone and T. D. Sparks, *Crystals*, 2019, **9**, 54.
50 S. Zhao, W. Kang and J. Xue, *Appl. Phys. Lett.*, 2014, **104**, 133106.
51 M. Khazaei, M. Arai, T. Sasaki, C.-Y. Chung, N. S. Venkataramanan, M. Estili, Y. Sakka and Y. Kawazoe, *Adv. Funct. Mater.*, 2013, **23**, 2185–2192.
52 B. R. Goldsmith, J. Esterhuizen, J.-X. Liu, C. J. Bartel and C. Sutton, 2018.
53 G. Kresse and J. Furthmüller, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **54**, 11169.
54 P. E. Blöchl, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1994, **50**, 17953.
55 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.
56 S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. Humphreys and A. P. Sutton, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1998, **57**, 1505.
57 *mendeleev - A Python resource for properties of chemical elements, ions and isotopes, ver. 0.11.0*, **https://github.com/lmmentel/mendeleev**, 2014.
58 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss and V. Dubourg, *et al.*, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.
59 G. James, D. Witten, T. Hastie and R. Tibshirani, *An introduction to statistical learning*, Springer, 2013, vol. 112.
60 A. Jain, G. Hautier, C. J. Moore, S. P. Ong, C. C. Fischer, T. Mueller, K. A. Persson and G. Ceder, *Comput. Mater. Sci.*, 2011, **50**, 2295–2310.
61 C. Wang and W. Pickett, *Phys. Rev. Lett.*, 1983, **51**, 597.
62 J. Heyd, G. E. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2003, **118**, 8207–8215.
63 J. Heyd and G. E. Scuseria, *J. Chem. Phys.*, 2004, **121**, 1187–1192.
64 L. Hedin, *Phys. Rev.*, 1965, **139**, A796.
65 M. S. Hybertsen and S. G. Louie, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1986, **34**, 5390.
66 G. Onida, L. Reining and A. Rubio, *Rev. Mod. Phys.*, 2002, **74**, 601.
67 S. Albrecht, L. Reining, R. Del Sole and G. Onida, *Phys. Rev. Lett.*, 1998, **80**, 4510.
68 M. Rohlfing and S. G. Louie, *Phys. Rev. Lett.*, 1998, **81**, 2312.
69 T. Zhu, A. J. Bishop, T. Zhou, M. Zhu, D. J. O'Hara, A. A. Baker, S. Cheng, R. C. Walko, J. J. Repicky and T. Liu, *et al.*, *Nano Lett.*, 2021, **21**, 5083–5090.
70 R. C. Walko, T. Zhu, A. J. Bishop, R. K. Kawakami and J. A. Gupta, *Phys. E*, 2022, **143**, 115391.

71 Z. Zhang, Y. Zhang, Y. Li, J. Lin, D. G. Truhlar and S. Huang, *Chem. Mater.*, 2018, **30**, 3208–3214.
72 X. Jiao, X. Hu, G. Xi, G. Li, L. Qiu, Y. Zou and X. Zhang, *Electrochim. Acta*, 2021, **366**, 137317.
73 Z. Lin, S. Ning, Z. Yang, Z. Zhang, S. Huang, J. Long, H. Lin and X. Wang, *Chem. Commun.*, 2016, **52**, 13381–13384.
74 H. Zhang, W. Yang, Y. Wang and X. Xu, *Phys. Rev. B*, 2021, **103**, 094433.
75 R. Kumar and A. Sundaresan, *Mater. Res. Bull.*, 2022, **145**, 111569.
76 C. L. Ciobanu, J. Brugger, N. J. Cook, S. J. Mills, P. Elliott, G. Damian and F. Damian, *Am. Mineral.*, 2014, **99**, 1163–1170.
77 T. Xie and J. C. Grossman, *Phys. Rev. Lett.*, 2018, **120**, 145301.
78 A. P. Bartók, R. Kondor and G. Csányi, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **87**, 184115.
79 B. Olsthoorn, R. M. Geilhufe, S. S. Borysov and A. V. Balatsky, *Adv. Quantum Technol.*, 2019, **2**, 1900023.
80 A. Saeki and K. Kranthiraja, *Jpn. J. Appl. Phys.*, 2019, **59**, SD0801.
81 J. R. Moreno, J. Flick and A. Georges, *Phys. Rev. Mater.*, 2021, **5**, 083802.
82 R. X. Yang, C. A. McCandler, O. Andriuc, M. Siron, R. Woods-Robinson, M. K. Horton and K. A. Persson, *ACS Nano*, 2022, **16**(12), 19873–19891.
83 C. Park, J.-W. Lee, M. Kim, B. Do Lee, S. P. Singh, W. B. Park and K.-S. Sohn, *Inorg. Chem. Front.*, 2021, **8**, 4610–4624.
84 Y. Li, S. Xu, J. Wang, C. Wang, B. Yang, H. Lin, W. Duan and B. Huang, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, **119**, e2122952119.
85 J.-Y. You, X.-J. Dong, B. Gu and G. Su, *Phys. Rev. B*, 2021, **103**, 104403.
86 S.-D. Guo and M.-X. Wang, *Phys. Chem. Chem. Phys.*, 2021, **23**, 22443–22450.
87 A. Togo and I. Tanaka, *Scr. Mater.*, 2015, **108**, 1–5.
88 C. J. Bartel, *J. Mater. Sci.*, 2022, **57**, 10475–10498.
89 W. Chen, X. Hou, X. Shi and H. Pan, *ACS Appl. Mater. Interfaces*, 2018, **10**, 35289–35295.
90 Y. Guo, S. Zhou, Y. Bai and J. Zhao, *Appl. Phys. Lett.*, 2017, **110**, 163102.
91 Y. Cheng, Z. Zhu, M. Tahir and U. Schwingenschlögl, *EPL*, 2013, **102**, 57001.
92 Y. Wang, W. Wei, H. Wang, N. Mao, F. Li, B. Huang and Y. Dai, *J. Phys. Chem. Lett.*, 2019, **10**, 7426–7432.
93 R. da Silva, R. Barbosa, R. R. Mancano, N. Duraes, R. B. Pontes, R. Miwa, A. Fazzio and J. E. Padilha, *ACS Appl. Nano Mater.*, 2019, **2**, 890–897.
94 Y. Hao, L.-C. Xu, Z. Yang, R.-P. Liu, X.-Y. Li, J. Pu and L.-F. Huang, *Electrochim. Acta*, 2019, **310**, 153–161.
95 W. Shi, G. Li and Z. Wang, *J. Phys Chem. C*, 2019, **123**, 12261–12267.
96 L. Zhang, Z. Yang, T. Gong, R. Pan, H. Wang, Z. Guo, H. Zhang and X. Fu, *J. Mater. Chem. A*, 2020, **8**, 8813–8830.
97 C. Jin, X. Tang, X. Tan, S. C. Smith, Y. Dai and L. Kou, *J. Mater. Chem. A*, 2019, **7**, 1099–1106.

