

Supplementary Information for investigating magnetic van der Waals materials using data-driven approaches

Romkanta Bhattarai^a, Peter Minch^a, and Trevor David Rhone^{*a}

^aDepartment of Physics, Applied Physics, and Astronomy, Rensselaer Polytechnic
Institute, Troy, NY, 12180

*rhonet@rpi.edu

Distribution of E_f , μ , and E_g in $A^iA^{ii}B_4X_8$ structures

The distributions of DFT-calculated formation energy (E_f), magnetic moment (μ), and band gap (E_g) in $A^iA^{ii}B_4X_8$ structures are shown in Figure 1. In Figure 1 (a), the majority of structures have formation energy per unit cell between -1 and 1 eV. Also, there exist more X = Te structures with positive formation energies (towards the right) than X = Se and X = S.

The distribution of magnetic moment in Figure 1 (b) shows that the majority of $A^iA^{ii}B_4X_8$ structures have magnetic moment per unit cell close to $2 \mu_B$ irrespective of elements at the X sites. This is why our ML model has a better performance around this value. The accuracy of the ML models is proportional to the amount of training data. The performance deteriorates as we go far from the center of the distribution. Therefore, we expect that adding more structures with magnetic moments to the left or right of the center of the distribution will improve the overall ML performance. Additionally, in the region of higher magnetic moments, the distribution shows there are more structures with X = Te and X = Se as compared to X = S.

The distribution of band gaps (Figure 1 (c)) shows that most of the structures are metallic. Also, structures with X = S lie in the larger band gap region while structures with X = Se and X = Te lie in the smaller band gap region. It will be interesting to investigate the nature of the band gap in future studies. This gives insight into the topological properties of the materials.

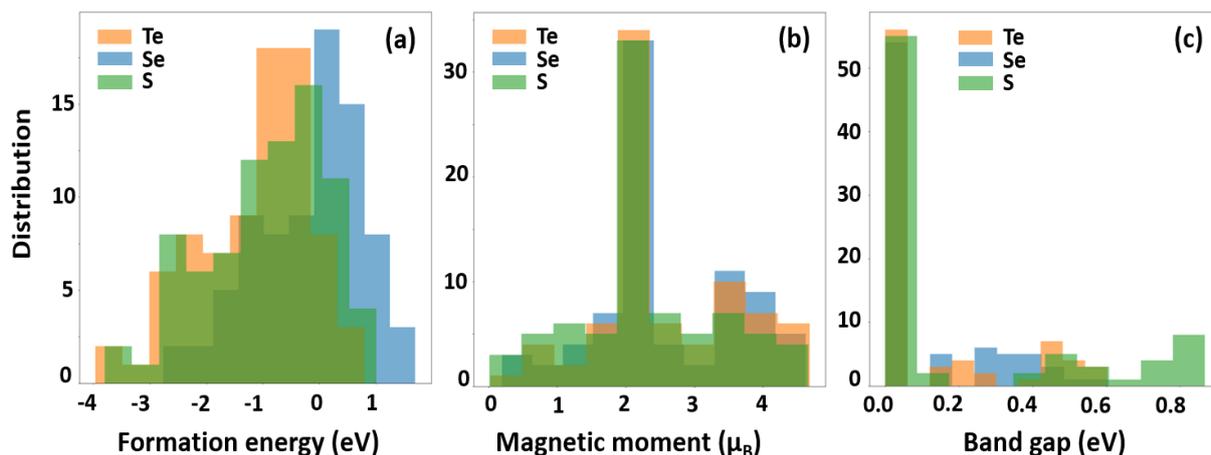


Figure 1: Distribution of (a) formation energy, (b) magnetic moment, and (c) band gap in $A^iA^{ii}B_4X_8$ structures

Descriptors

We construct the chemical descriptors from 11 different atomic properties. There are 5 distinct operations for constructing descriptors from each atomic property. The atomic properties are obtained from the Mendeleev python package¹. Thus, a total of 55 descriptors are considered in our ML models. Table 1 shows the 11 atomic properties and 5 distinct operations along with their naming conventions.

Table 1: Atomic properties and the mathematical transformations used to construct the descriptors

Atomic properties, P	Distinct descriptor operations
van der Waals radius (vdW rad)	combined average (avg)
covalent radius (covalent rad)	$(\frac{1}{n} \sum P)$
atomic radius (atomic rad)	difference between A^i and A^{ii} sites ($\Delta A^i A^{ii}$)
atomic volume (atomic vol)	$(P(A^i) - P(A^{ii}))$
dipole polarizability (dipole pol)	difference between B and X sites (ΔBX)
number of valence electrons (val elect)	$(P(B) - P(X))$
number of unpaired electrons (unp elect)	maximum difference among all sites (max diff)
ionization energies (ion energy)	$(\max(P(A^i) - P(A^{ii}) , P(B) - P(X) , \dots))$
total number of electrons (no elect)	combined standard deviation (std)
electronegativity (electroneg)	$(\sqrt{\frac{1}{n-1} \sum (P - \bar{P})^2})$
electron affinity (elect aff)	

Top descriptors for the formation energy prediction

The top 3 descriptors for predicting the formation energy are the difference in atomic volume between B and X sites, the difference in electron affinity between A^i and A^{ii} sites, and the standard deviation of electronegativity among all sites. Table 2 shows the top 10 descriptors.

Figure 2 shows the scatter plots of the top three descriptors with respect to E_f . It shows that the formation energy and the atomic volume are inversely related (Figure 2(a)). i.e. E_f decreases when the

Table 2: Top 10 descriptors for the formation energy prediction

Rank	Descriptors	Rank	Descriptors
1	atomic vol (ΔBX)	6	cov rad (max diff)
2	elect aff ($\Delta A^i A^{ii}$)	7	electroneg (max diff)
3	electroneg (std)	8	cov rad ($\Delta A^i A^{ii}$)
4	atomic vol ($\Delta A^i A^{ii}$)	9	electroneg ($\Delta A^i A^{ii}$)
5	dipol pol (std)	10	atomic vol (avg)

difference in atomic volume (ΔBX) increases. A similar pattern is observed in the case of electronegativity (Figure 2(c)). In Figure 2(b), E_f increases when the electron affinity ($\Delta A^i A^{ii}$) increases.

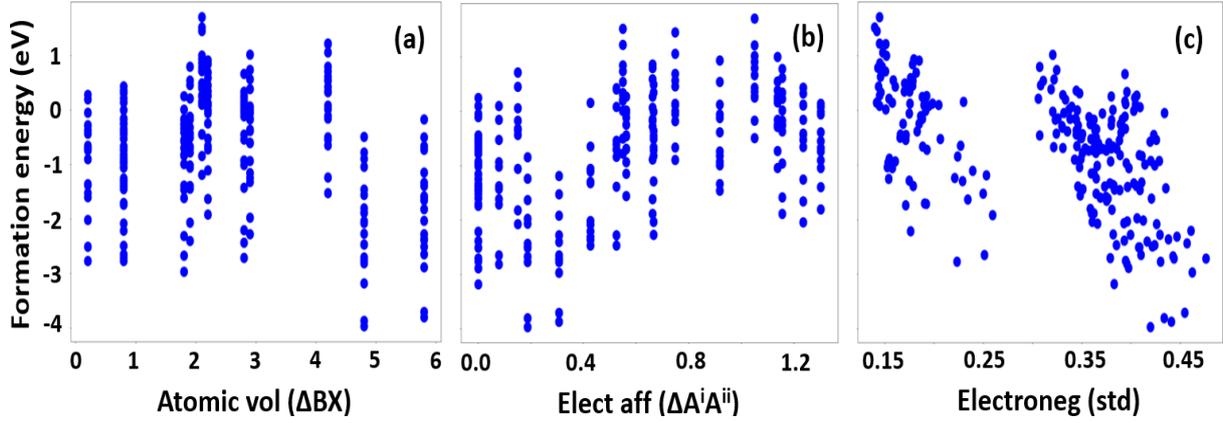


Figure 2: Scatter plots of top 3 descriptors for formation energy prediction

Top descriptors for the magnetic moment prediction

The top 3 descriptors for predicting the magnetic moment are the average number of unpaired electrons, the average electron affinity, and the standard deviation (std) of valence electrons. Table 3 shows the top 10 descriptors.

Table 3: Top 10 descriptors for the magnetic moment prediction

Rank	Descriptors	Rank	Descriptors
1	unp elect (avg)	6	val elect ($\Delta A^i A^{ii}$)
2	elect aff (avg)	7	unp elect (std)
3	val elect (std)	8	no elect ($\Delta A^i A^{ii}$)
4	val elect (max diff)	9	dipole pol (avg)
5	vdW rad ($\Delta A^i A^{ii}$)	10	unp elect ($\Delta A^i A^{ii}$)

Figure 3 shows the scatter plots of top three descriptors. It shows that the extrema of magnetic moments correspond to the higher number of unpaired electrons (Figure 3 (a)) and lower std of valence electrons (Figure 3 (c)). In addition, the average electron affinity is also determining factor of magnetic moment as shown in Figure 3 (b).

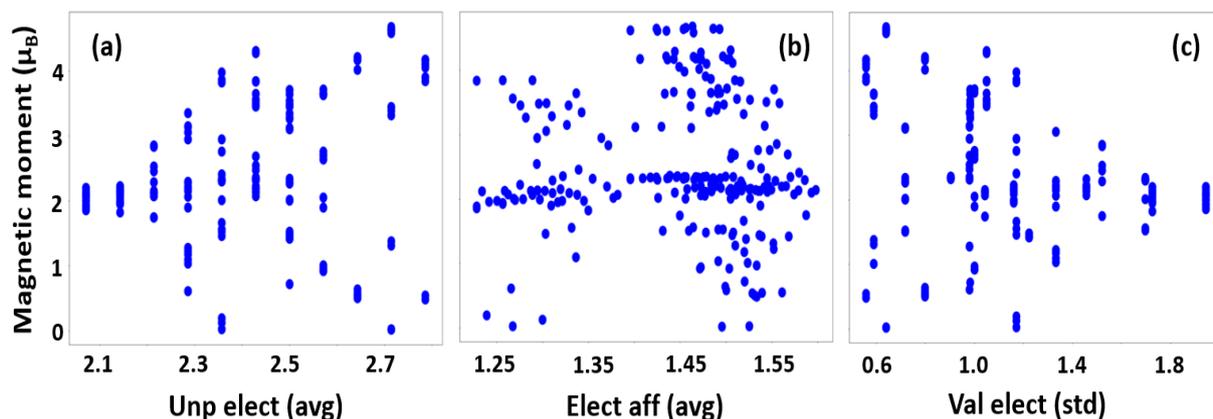


Figure 3: Scatter plots of top 3 descriptors for magnetic moment prediction

Top descriptors for the band gap prediction

Table 4 shows the 10 most important descriptors for predicting the band gap. Average unpaired electrons, average dipole polarizability, and average valence electrons are the first three important descriptors.

Table 4: Top 10 descriptors for the band gap prediction

Rank	Descriptors	Rank	Descriptors
1	unp elect (avg)	6	elect aff ($\Delta A^i A^{ii}$)
2	dipole pol (avg)	7	ion energy (avg)
3	val elect (avg)	8	vdW rad ($\Delta A^i A^{ii}$)
4	val elect (ΔBX)	9	vdW rad (avg)
5	unp elect (ΔBX)	10	electroneg (avg)

Figure 4 shows the scatter plots of the top three descriptors for the band gap. It shows that the band gap is larger in the region where the average number of unpaired electrons is higher (Figure 4(a)). Also, it increases with average valence electrons (Figure 4(c)). However, the relationship between band gap and with dipole polarizability is showing an opposite trend as shown in Figure 4(b).

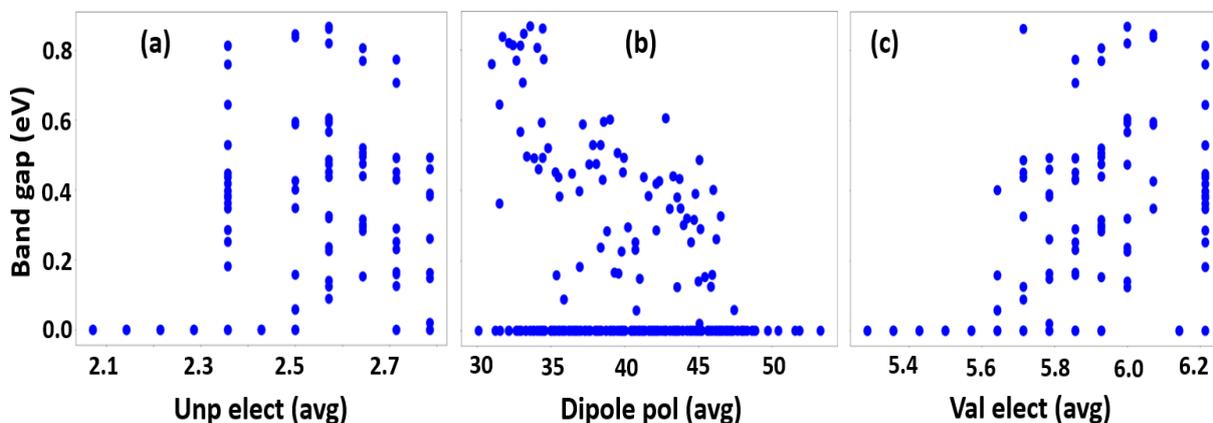


Figure 4: scatter plots of the top 3 descriptors for the band gap prediction

Hyperparameters for ML models

We tune the hyperparameters in all three random forest regression models to optimize the performance. The optimized hyperparameters for formation energy, magnetic moment, and band gap are presented in Table 5.

Table 5: Hyperparameters used to optimize the random forest model predictions for three distinct targets: formation energy, magnetic moment, and band gap

Hyperparameter	Formation energy	Magnetic moment	Band gap
max_depth	20	24	16
max_features	auto	auto	auto
min_samples_leaf	2	2	2
min_samples_split	4	4	2
n_estimators	30	10	20

Dynamic stability of predicted structures

We test the dynamic stability of the pure MBT monolayer and all the predicted structures by performing two calculations: (i) phonon spectra, and (ii) ab initio molecular dynamics (AIMD) simulations at high temperatures. The phonon calculations are performed on large supercells size ($4 \times 4 \times 1$) using the finite displacement method as implemented in the Phonopy package². From the calculations, we find that there are no significant imaginary phonon frequencies or soft modes in the phonon spectra of the $\text{Mn}_2\text{Bi}_4\text{S}_8$, $\text{Mn}_2\text{Sb}_2\text{Bi}_2\text{S}_4\text{Se}_4$, $\text{Mn}_2\text{Sb}_2\text{Bi}_2\text{Se}_4\text{Te}_4$, $\text{Mn}_2\text{Sb}_2\text{Bi}_2\text{Se}_8$, $\text{Mn}_2\text{Sb}_4\text{Se}_4\text{Te}_4$, and $\text{Mn}_2\text{Bi}_4\text{S}_4\text{Se}_4$ monolayer structures, thus implying their dynamical stability. A very small negative frequency near the Γ point in the Brillouin zone is due to a numerical artifact in the calculations. This can be removed by using much tighter convergence criteria or increasing the supercell size during the simulations. However, we see significant negative frequencies between Γ and M in the phonon spectrum of $\text{MnSbBiS}_2\text{Te}_2$. This indicates that the structure is dynamically metastable or unstable. We expect that there are stable competing phases of $\text{Mn}_2\text{Sb}_2\text{Bi}_2\text{S}_4\text{Te}_4$, the exploration of which is outside the scope of this work. The AIMD simulations are carried out on supercells size ($3 \times 3 \times 1$) at 500 K for 20 ps with a time step of 2 fs. The potential energy profiles are plotted as a function of time. The energy profile shows that the total energy fluctuates during the entire simulation period of 20 ps without any sudden drops in the energy. This indicates that the predicted structures are stable at high temperatures.

This stability implies that the predicted materials are promising candidates for further investigation, leading toward a set of promising candidates for synthesis³. We also note that an additional test that further bolsters synthesizability is calculating the energy above the convex hull. Nevertheless, this does not guarantee that the materials can be synthesized. There’s also the fact that metastable phases, although not the lowest energy phases, may still be synthesized in the experiment^{4;5;6;7;8}. The convex hull analysis will still provide useful information on the chemical stability and synthesizability of the predicted

materials. We hope that this present study will motivate other researchers to investigate the competing phases of the most promising materials candidates identified in this study.

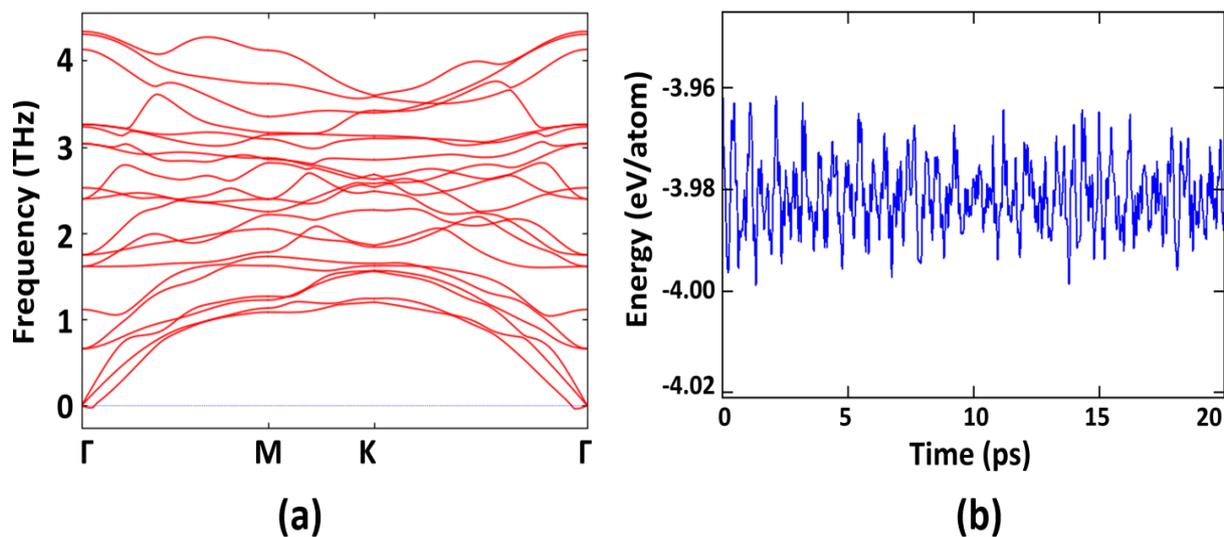


Figure 5: Phonon dispersion of the $\text{Mn}_2\text{Bi}_4\text{Te}_8$ (MBT) monolayer. (b) Energy profile of $\text{Mn}_2\text{Bi}_4\text{Te}_8$ monolayer during AIMD simulation at 500 K over 20 ps.

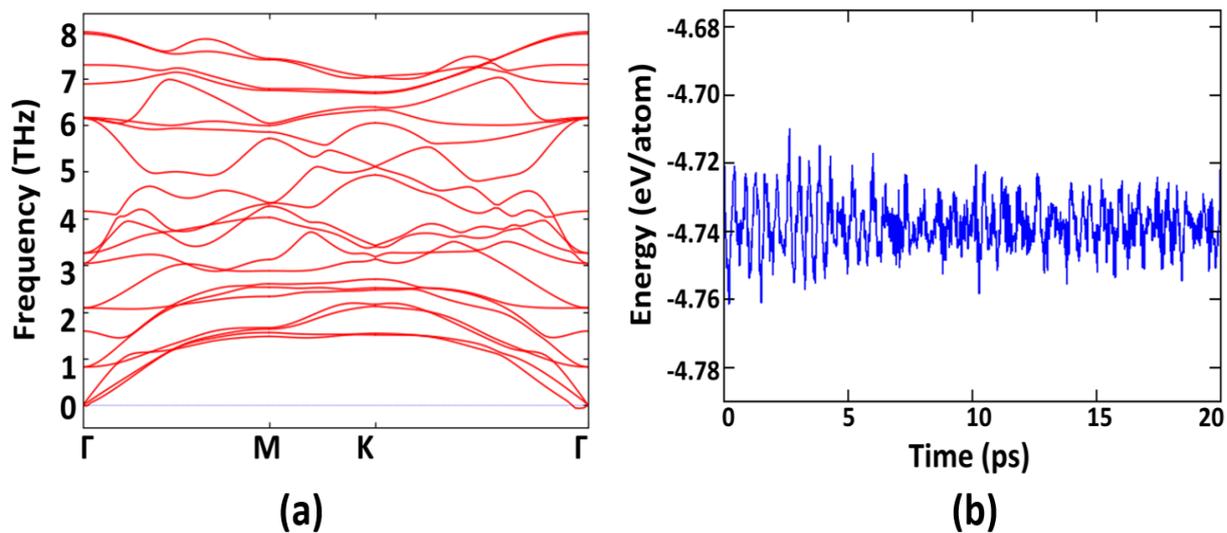
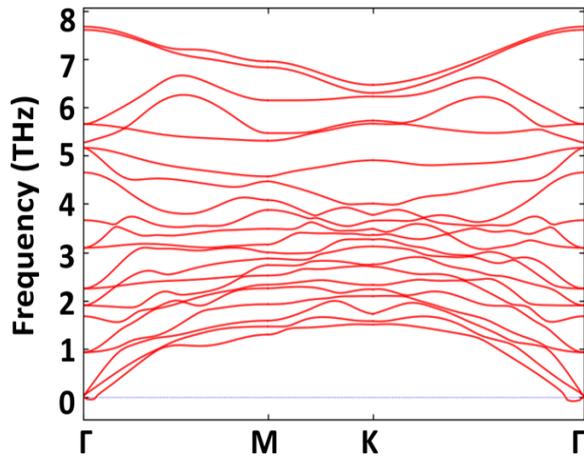
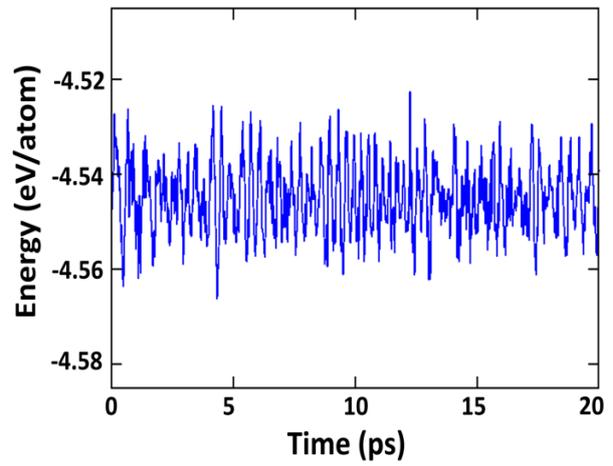


Figure 6: Phonon dispersion of the $\text{Mn}_2\text{Bi}_4\text{S}_8$ monolayer. (b) Energy profile of $\text{Mn}_2\text{Bi}_4\text{S}_8$ monolayer during AIMD simulation at 500 K over 20 ps.

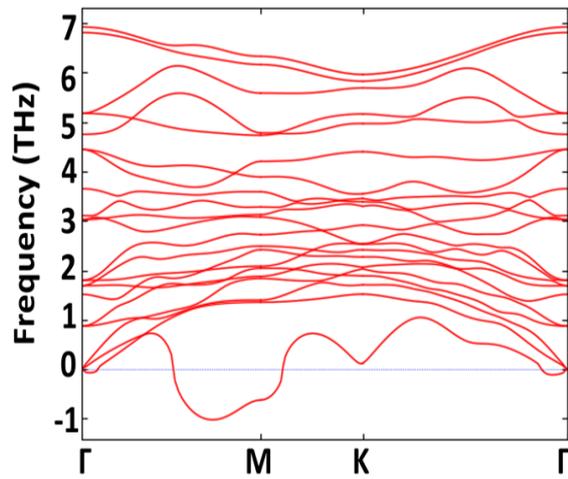


(a)

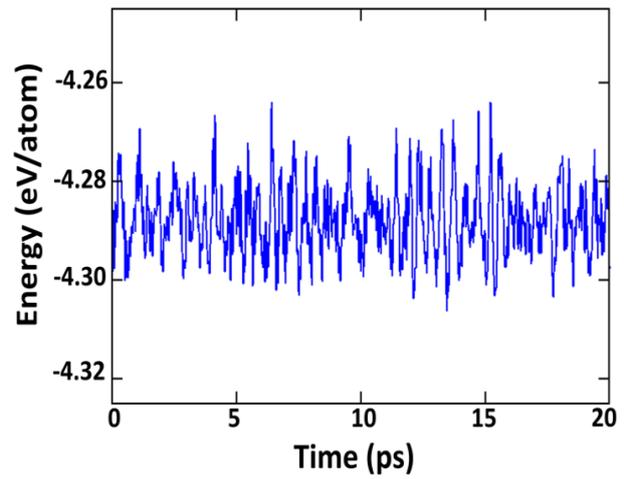


(b)

Figure 7: Phonon dispersion of the $\text{Mn}_2\text{Sb}_2\text{Bi}_2\text{S}_4\text{Se}_4$ monolayer. (b) Energy profile of $\text{Mn}_2\text{Sb}_2\text{Bi}_2\text{S}_4\text{Se}_4$ monolayer during AIMD simulation at 500 K over 20 ps.



(a)



(b)

Figure 8: Phonon dispersion of the $\text{Mn}_2\text{Sb}_2\text{Bi}_2\text{S}_4\text{Te}_4$ monolayer. (b) Energy profile of $\text{Mn}_2\text{Sb}_2\text{Bi}_2\text{S}_4\text{Te}_4$ monolayer during AIMD simulation at 500 K over 20 ps.

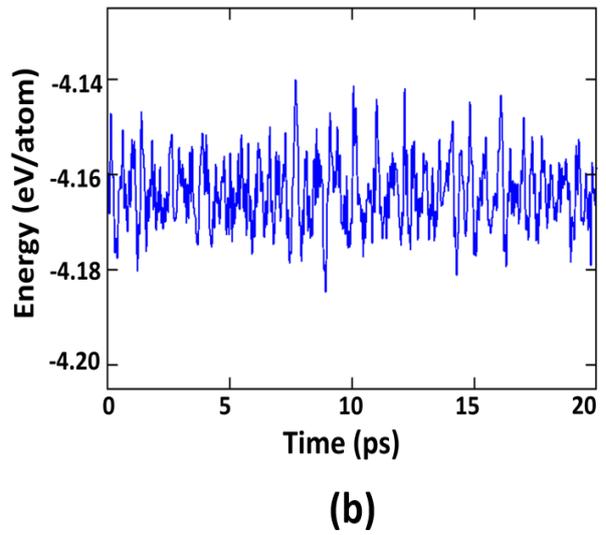
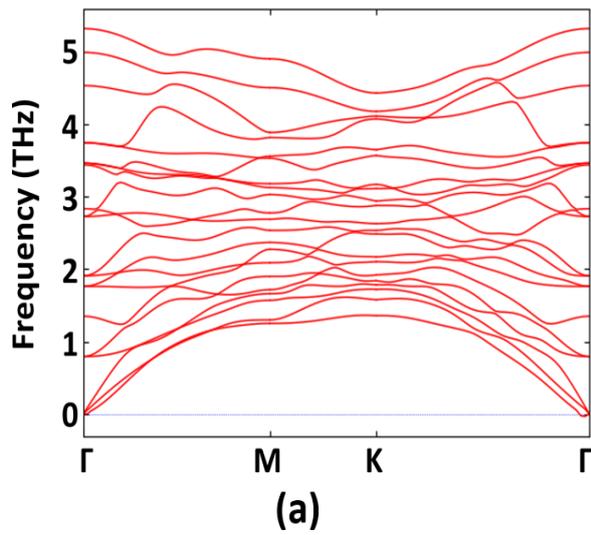


Figure 9: Phonon dispersion of the $\text{Mn}_2\text{Sb}_2\text{Bi}_2\text{Se}_4\text{Te}_4$ monolayer. (b) Energy profile of $\text{Mn}_2\text{Sb}_2\text{Bi}_2\text{Se}_4\text{Te}_4$ monolayer during AIMD simulation at 500 K over 20 ps.

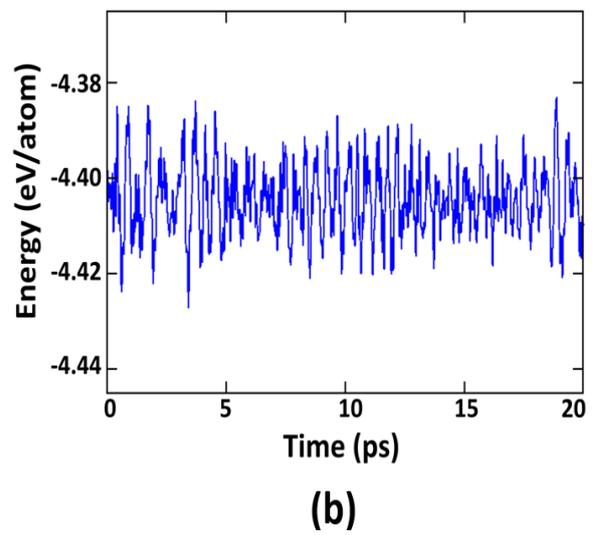
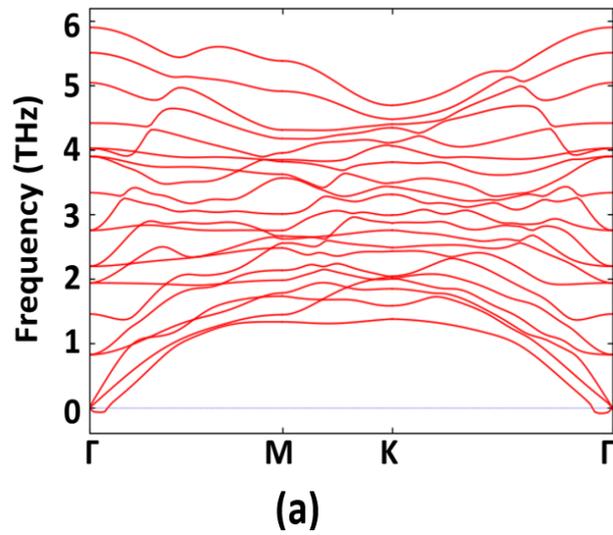


Figure 10: Phonon dispersion of the $\text{Mn}_2\text{Sb}_2\text{Bi}_2\text{Se}_8$ monolayer. (b) Energy profile of $\text{Mn}_2\text{Sb}_2\text{Bi}_2\text{Se}_8$ monolayer during AIMD simulation at 500 K over 20 ps.

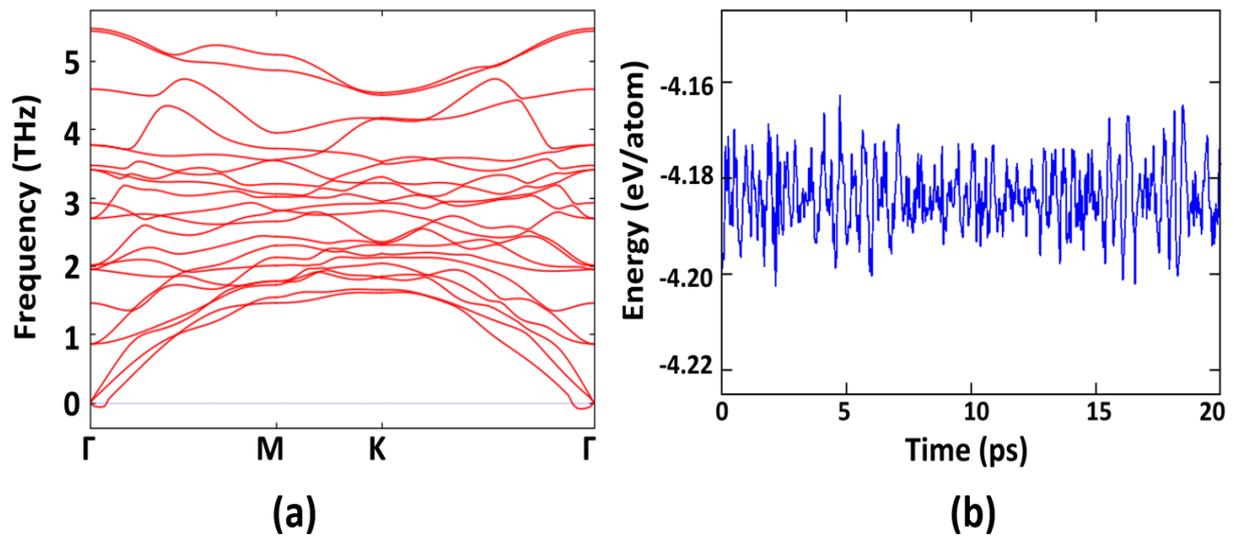


Figure 11: Phonon dispersion of the $\text{Mn}_2\text{Sb}_4\text{Se}_4\text{Te}_4$ monolayer. (b) Energy profile of the $\text{Mn}_2\text{Sb}_4\text{Se}_4\text{Te}_4$ monolayer during AIMD simulation at 500 K over 20 ps.

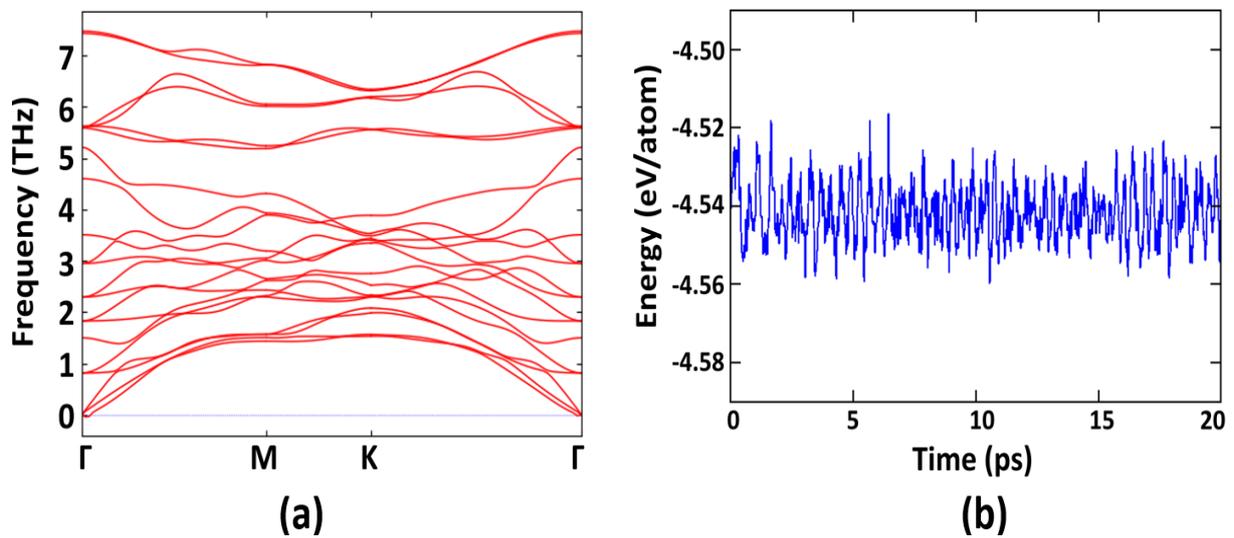


Figure 12: Phonon dispersion of the $\text{Mn}_2\text{Bi}_4\text{S}_4\text{Se}_4$ monolayer. (b) Energy profile of $\text{Mn}_2\text{Bi}_4\text{S}_4\text{Se}_4$ monolayer during AIMD simulation at 500 K over 20 ps.

Data availability

The dataset (generated from the density functional theory calculations) that supports the finding of the study will be made available upon reasonable request.

Notes and references

- [1] *mendeleev* – A Python resource for properties of chemical elements, ions and isotopes, ver. 0.11.0, <https://github.com/lmmentel/mendeleev>, 2014–.
- [2] A. Togo and I. Tanaka, *Scripta Materialia*, 2015, **108**, 1–5.
- [3] O. I. Malyi, K. V. Sopiha and C. Persson, *ACS applied materials & interfaces*, 2019, **11**, 24876–24884.
- [4] M. S. Sokolikova and C. Mattevi, *Chemical Society Reviews*, 2020, **49**, 3952–3980.
- [5] H. Jin, T. Song, U. Paik and S.-Z. Qiao, *Accounts of Materials Research*, 2021, **2**, 559–573.
- [6] J. Yang, Q. Xu, Y. Zheng, Z. Tian, Y. Shi, C. Ma, G. Liu, B. Peng, Z. Wang and W. Zheng, *ACS nano*, 2022, **16**, 15215–15225.
- [7] M. Lemon, F. G. Harvel, R. N. Gannon, P. Lu, S. P. Rudin and D. C. Johnson, *Journal of Vacuum Science & Technology A*, 2023, **41**, 022203.
- [8] M. Aykol, S. S. Dwaraknath, W. Sun and K. A. Persson, *Science advances*, 2018, **4**, eaaq0148.