



OPEN

DATA DESCRIPTOR

# An open database of computed bulk ternary transition metal dichalcogenides

Scott E. Muller <sup>✉</sup>, Micah P. Prange, Zexi Lu, W. Steven Rosenthal & Jenna A. Bilbrey 

We present a dataset of structural relaxations of bulk ternary transition metal dichalcogenides (TMDs) computed via plane-wave density functional theory (DFT). We examined combinations of up to two chalcogenides with seven transition metals from groups 4–6 in octahedral (1T) or trigonal prismatic (2H) coordination. The full dataset consists of 672 unique stoichiometries, with a total of 50,337 individual configurations generated during structural relaxation. Our motivations for building this dataset are (1) to develop a training set for the generation of machine and deep learning models and (2) to obtain structural minima over a range of stoichiometries to support future electronic analyses. We provide the dataset as individual VASP xml files as well as all configurations encountered during relaxations collated into an ASE database with the corresponding total energy and atomic forces. In this report, we discuss the dataset in more detail and highlight interesting structural and electronic features of the relaxed structures.

## Background & Summary

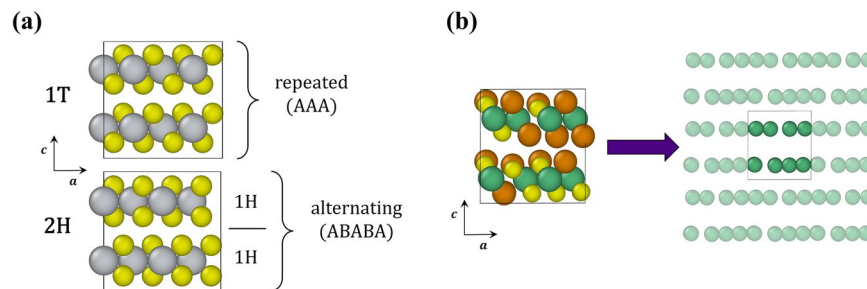
Transition metal dichalcogenides (TMDs) have fascinating electrical, optical, and mechanical properties<sup>1,2</sup>, with wide-ranging applications in sensing<sup>3</sup>, electronics<sup>4</sup>, batteries<sup>5</sup>, catalysis<sup>6,7</sup>, and pollution remediation<sup>8</sup>. The structure and properties of TMDs depend on the choice of metal and chalcogenide, as well as the crystal structure. Combinations of chalcogenides and/or metals lead to different behaviors that allow TMDs to be tailored for a desired structural phase or band gap<sup>9</sup>. The number of potential metal/chalcogenide combinations is challenging and expensive to explore by synthesis and experiment. The lack of data covering the full range of possible stoichiometries hinders data-driven scientific discovery. Computational modeling of TMD structures and calculation of their resulting properties will allow for downselection of potential TMD combinations for synthesis.

Pure TMDs have a stoichiometry of  $\text{MX}_2$ , where M is a transition metal and X is a chalcogenide. Ternary and quaternary TMDs, containing mixtures of metals and/or chalcogenides, have been investigated for improved structure and property control. For example, Zhou *et al.* synthesized 13 TMD alloys (11 ternary, 1 quaternary, and 1 quinary) via molten-salt-assisted CVD<sup>10</sup>. Susarla *et al.* showed that the bandgap of quaternary alloys of  $\text{Mo}_{1-\gamma}\text{W}_{\gamma}\text{S}_{2\delta}\text{Se}_{2(1-\delta)}$  could be tuned between 1.61 and 1.85 eV, with theoretical investigation demonstrating tuning between an even wider range of 1.60 and 2.03 eV<sup>11</sup>.

In addition to the large variety of possible elemental compositions, TMDs are polymorphic and thus form several polytypes distinguished by intralayer coordination and interlayer stacking. Here, we focus on the 1T and 2H polytypes, where the number denotes the number of layers present in a repeating unit, and the letter denotes the crystal system (tetragonal or hexagonal) with the transition metal sitting in either octahedral (1T) or trigonal prismatic (2H) sites<sup>12</sup>. These are shown in Fig. 1a. The relative stabilities of polymorphs vary across the TMD family, with the primary influence on stability being the variance in *d* orbital electrons depending on the constituent transition metal group in the periodic table; e.g., TMDs with group 4 and 6 metals often present in the 2H phase, while those with group 5 metals present as both 1T and 2H<sup>13</sup>. Metastable polytypes can be stabilized through various techniques, including the intercalation of ions between the van der Waals layers and reaction quenching<sup>14,15</sup>.

To add to the complexity of these systems, distorted polytypes resulting from intralayer symmetry breaking have also been observed experimentally. Lu *et al.* attributed the broken symmetry in 1T'  $\text{WTe}_2$  to charge density wave distortion, which causes the vertical displacement of transition metal atoms and is clearly observed along the [100] and [010] orientations<sup>16</sup>, demonstrated in Fig. 1b. Lai *et al.* observed a similar variation in

Pacific Northwest National Laboratory, Richland, WA, 99352, USA. ✉e-mail: [scott.muller@pnnl.gov](mailto:scott.muller@pnnl.gov)



**Fig. 1** Details of structure and composition for studied TMDs. **(a)** Structural depictions of 1T and 2H TMDs, where gray spheres represent a transition metal and yellow spheres represent a chalcogenide. **(b)** Structural depiction of a 1T' distorted structure, where green spheres represent a transition metal and yellow and orange spheres represent different chalcogenides. At the right we show the unit cell (inside the box) and its semi-transparent periodic images, with chalcogenides removed to show the broken symmetry in the 1T' polytype.

M-M distances in 1T' WS<sub>2</sub> structures<sup>17</sup>. The distorted 1T' phase of MoTe<sub>2</sub>—a semimetal with superconductivity—has been formed through laser-induced and thermal phase transitions<sup>18–21</sup>, and synthesized via chemical vapor deposition (CVD)<sup>22,23</sup> and molecular beam epitaxy (MBE)<sup>24</sup>. Zhang *et al.* found that a metal-to-insulator transition can be induced in W<sub>γ</sub>Mo<sub>1–γ</sub>Te<sub>2</sub> under ambient conditions resulting from a transition from the 2H (semiconducting) to 1T' (semimetallic) polytype, which demonstrates the correspondence between structural and electronic properties in these systems<sup>25</sup>.

The electronic and magnetic properties of TMDs are dependent on the composition, which can be studied by examining the variation in the computed electronic structure induced by changing the identity and concentrations of the atomic constituents. Many of the desirable properties in this class of materials arise from the transition metal *d* orbitals. The number and behavior of electrons occupying these orbitals are affected by the identity of the transition metal, with more *d* orbitals filled as one moves right on the periodic table. They are also affected by the chalcogen species, which interact with the transition metal through the valence *p* orbitals. These orbitals grow in size with the principal quantum number *n*. There is still a need for thorough computational analyses across a broad compositional space to understand underlying trends in these materials.

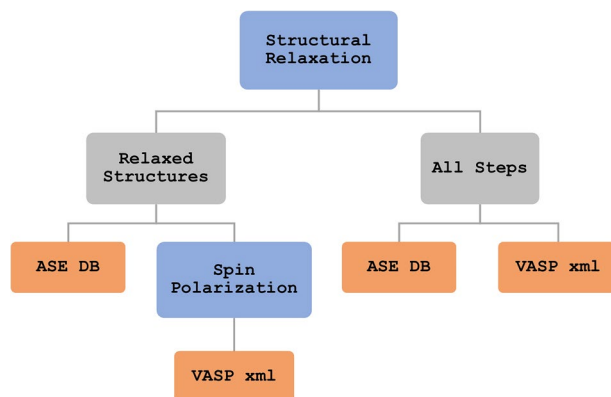
The diverse composition and varied structural and electronic properties make TMDs ripe for machine learning (ML) analysis. Sumi *et al.* used ML to predict the semiconductor-to-metal transition in MoX<sub>2–δ</sub>O<sub>δ</sub> bilayers (where X is a chalcogenide) and found that the concentration of chalcogen atoms in the interlayer van der Waals region was a leading predictor of the semiconductor-to-metal transition<sup>26</sup>. Zhao *et al.* developed a high-throughput screening approach combining first-principles computations with ML to screen for synthesizable 2D TMDs with effective Hg<sup>0</sup> sensing response<sup>27</sup>. Their approach relied on the use of AFLOW-ML, a software tool to apply ML-based property prediction algorithms to the user's dataset of interest<sup>28</sup>. Behler discusses the acceleration of atomic-scale simulations of condensed systems using neural network potentials (NNPs), noting the need for electronic structure data computed at high level of theory to serve as training sets<sup>29</sup>. For any ML approach, particularly neural networks, it is crucial that sufficient data is available for training, testing, and validation<sup>30</sup>.

The field of computational materials science has seen rapid advancements due to the availability of large datasets generated using high-throughput calculation methods and/or supercomputing resources. In particular, the National Institute of Standards and Technology (NIST)'s wide-spanning Materials Genome Initiative (MGI)<sup>31</sup>, and the associated Materials Project<sup>32</sup>, have laid the foundation for researchers to share simulation results in a way that supports FAIR data principles – Findability, Accessibility, Interoperability, and Reusability<sup>33</sup>. Repositories with strict computational requirements or more specialized information, such as the joint automated repository for various integrated simulations (JARVIS)<sup>34</sup>, are also available.

Supervised ML approaches, commonly applied in the materials science domain, require training datasets that encompass the configuration space of interest. Meanwhile, tuning the performance of TMD-based devices via constituent composition requires knowledge of the relationship among compositional variables, atomic structure, and functional properties. To support both needs, we produced a database of structurally relaxed ternary TMDs computed at the DFT level. We examined 1T and 2H ternary structures (MX<sub>2–δ</sub>Y<sub>δ</sub>) of seven transition metals (M = Ti, V, Nb, Mo, Hf, Ta, and W) with varying binary combinations of chalcogenides (X, Y = S, Se, and Te), comprising 672 unique stoichiometries. For a subset of the computed structures, we provide spin-polarization calculations to characterize their electronic structure. Our primary motivations for building this dataset are (1) to provide a rich dataset for training ML-based property prediction models and NNPs and (2) to obtain structural minima over a range of stoichiometries for further electronic analyses. In this report, we discuss the computational methods used to compute the structures and provide brief structural and energetic analyses of the computed ternary TMDs. The database and corresponding analysis scripts are openly available at <https://doi.org/10.6084/m9.figshare.21308157>.

## Methods

**Composition selection.** We consider ternary TMDs with the stoichiometry MX<sub>2–δ</sub>Y<sub>δ</sub>, where  $\delta$  is the ratio of Y atoms to M atoms. We include the chalcogens S, Se, and Te, along with seven transition metals from three different columns in the periodic table: two metals are from group 4 (Ti, Hf), three metals are from group 5 (V, Nb, Ta), and two metals are from group 6 (Mo, W). These transition metal groups tend to be stable in TMDs



**Fig. 2** Simulation workflow. Methods are in blue, outputs in gray, and data products in orange.

at ambient pressure and temperature<sup>35</sup>. We select only metals that are generally stable with all three chalcogenides, which precludes Cr from group 6<sup>36</sup>. We also neglect Zr, as TMDs from group 4 follow similar trends in behavior, and we already include the corresponding lighter and heavier metals (Ti and Hf, respectively)<sup>37</sup>.

We initialize each TMD as either the 1T or 2H polytype. While in binary form the 1T structure can be effectively simulated as a single layer, ternary TMDs will almost certainly break the symmetry that can be represented by one layer. Thus, we consider two-layered TMD simulation cells for both 1T and 2H structures. Note that finite size effects still exist due to artificial symmetry imposed by periodic boundary conditions of the simulation cells. Our TMD simulation cells consist of 48 atoms, with 24 atoms in each layer. The 2H structures use the same  $c$  lattice parameter as in their primitive unit cell, while the  $c$  lattice parameter is doubled for the 1T structures. From the primitive 2-layer structure, which consists of 6 atoms, our 48-atom simulation cell is formed by quadrupling the primitive cell in both the  $a$  and  $b$  directions, and then partitioning a nearly orthorhombic section of 48 atoms. This yields a simulation cell with dimensions  $a = (2\sqrt{3})a_0$ ,  $b = 2a_0$ , and  $c_{1T} = 2c_0$ , and  $c_{2H} = c_0$ , where  $a_0$  and  $c_0$  are the primitive lattice parameters. Chalcogenide lattice site occupations for mixed TMD structures are chosen in the manner of special quasirandom structures<sup>38</sup>. Note that there are many ways that specific chalcogenides may be distributed to lattice sites in mixed TMDs—a good review of structured TMDs was compiled by Zhang *et al.*<sup>39</sup> The chalcogenide stoichiometric ratio  $\delta$  ranges from 0 to 2 with minimum steps of 0.25 (corresponding to the replacement of 4 chalcogenide atoms in the simulation cell). For the 1T polytypes, we examine a reduced step size of 0.0625 (corresponding to the replacement of 1 chalcogenide atom in the simulation cell) from 0 to 0.5 and 1.5 to 2 to explore fine-grained effects of chalcogenide mixing.

**Simulation specifics.** We carry out positional relaxation of TMD structures using the Vienna Ab-initio Simulation Package (VASP)<sup>40,41</sup>, which solves the Kohn-Sham equations of density functional theory using plane-wave basis sets (defined by a 600 eV cutoff) in a periodic system. A conjugate gradient algorithm is applied to optimize structural parameters, including the internal coordinates and the cell vectors. We use the generalized gradient approximation (GGA) of the Perdew-Burke-Ernzerhof (PBE) functional<sup>42</sup> to compute the exchange-correlation potential. The projector augmented wave (PAW) method was used with parameters included in the VASP supplied POTCAR files. These exhibited 4 (Ti, Hf), 5 (V, Ta), or 6 (S, Se, Te, Mo, W) explicit electrons per atom. Nb exhibited 11 explicit electrons per atom because it also included the  $4p^6$  electrons in the valence electrons. Relaxations were performed with a Gamma-centered  $3 \times 6 \times 4$  Monkhorst-Pack grid to sample the Brillouin zone. Orbital occupancies were broadened with Gaussian smearing with a width of 0.01 eV. We include van der Waals corrections with zero damping using Grimme *et al.*'s DFT-D3 method<sup>43</sup>. We consider a structure to be in a relaxed state when the magnitude of the maximum force on any atom is less than 0.01 eV/Å. All input parameters for the VASP calculations can be retrieved from the datafiles supplied in MXY\_DB, with the simulation workflow shown in Fig. 2. For most  $\delta$  values only a single supercell was generated. An exception was  $\delta = 1.0$  where two supercells were generated for each structure, with the initial chalcogenide locations being identical but the chalcogenide types reversed. We compared our quasirandom structures to a large set of structures (200) with completely randomly assigned chalcogenides for specific  $\delta$  values and found that the energy of our original quasirandom structures were within a standard deviation of the mean energy value, thus representing a typical structure.

Spin-polarization calculations were carried out on a subset of the relaxed structures (316 total). We perform collinear spin-polarized calculations within the PBE +  $U$  formalism<sup>44</sup>, where the semi-local approximate DFT exchange-correlation potential is corrected with onsite Hubbard terms that cancel the tendency of GGA to over-delocalize electrons. Parameters for the  $U$  correction are taken from the Hubbard  $U$  (Atomic) values obtained by Kirchner-Hall *et al.*<sup>45</sup>. Due to the wide variety of electronic structures observed, a complete analysis of all the variation is beyond the scope of this manuscript, but could be undertaken with the data provided in MXY\_DB.

**Property calculations.** Along with the atomic coordinates, total energy ( $E_0$ ), and atomic forces, we also conduct analysis on some structural and energetic features for each relaxed structure. Structural features include interlayer spacing, cell lengths ( $a$ ,  $b$ ,  $c$ ) and angles. Energetic features include per-atom formation energy ( $E_{\text{form}}$ ), calculated as

		Energies in eV/atom		
		$E_{\text{form}}(1\text{T})$	$E_{\text{form}}(1\text{T}')$	$\Delta E_{\text{form}}(1\text{T} \rightarrow 1\text{T}')$
Group 4	TiS <sub>2</sub>	-6.798	-6.791	0.006
	TiSe <sub>2</sub>	-6.268	-6.259	0.009
	TiTe <sub>2</sub>	-5.646	-5.632	0.014
	HfS <sub>2</sub>	-7.855	-7.851	0.003
	HfSe <sub>2</sub>	-7.244	-7.240	0.004
	HfTe <sub>2</sub>	-6.533	-6.521	0.012
Group 5	VS <sub>2</sub>	-6.779	-6.777	0.002
	VSe <sub>2</sub>	-6.237	-6.205	0.032
	VTe <sub>2</sub>	-5.676	-5.706	<b>-0.030</b>
	NbS <sub>2</sub>	-7.358	-7.354	0.004
	NbSe <sub>2</sub>	-6.815	-6.812	0.003
	NbTe <sub>2</sub>	-6.211	-6.229	<b>-0.017</b>
	TaS <sub>2</sub>	-7.971	-7.970	0.001
	TaSe <sub>2</sub>	-7.354	-7.353	0.001
TaTe <sub>2</sub>	-6.698	-6.736	<b>-0.038</b>	
Group 6	MoS <sub>2</sub>	-7.246	-7.333	<b>-0.087</b>
	MoSe <sub>2</sub>	-6.721	-6.834	<b>-0.113</b>
	MoTe <sub>2</sub>	-6.165	-6.317	<b>-0.151</b>
	WS <sub>2</sub>	-7.854	-7.965	<b>-0.111</b>
	WSe <sub>2</sub>	-7.288	-7.441	<b>-0.153</b>
	WTe <sub>2</sub>	-6.650	-6.859	<b>-0.209</b>

**Table 1.** Energy of formation from elements of 1T and 1T' polytypes for binary TMDs. The rightmost column is given by  $\Delta E_{\text{form}}(1\text{T} \rightarrow 1\text{T}') = E_{\text{form}}(1\text{T}') - E_{\text{form}}(1\text{T})$ ; text in this column is bolded for binary TMDs with better 1T' stability.

$$E_{\text{form}} = \frac{1}{3N_{\text{M}}} \{E_0(\text{MX}_{2-\delta}\text{Y}_\delta) - N_{\text{M}}[E_{\text{M}} + (2 - \delta)E_{\text{X}} + \delta E_{\text{Y}}]\}, \quad (1)$$

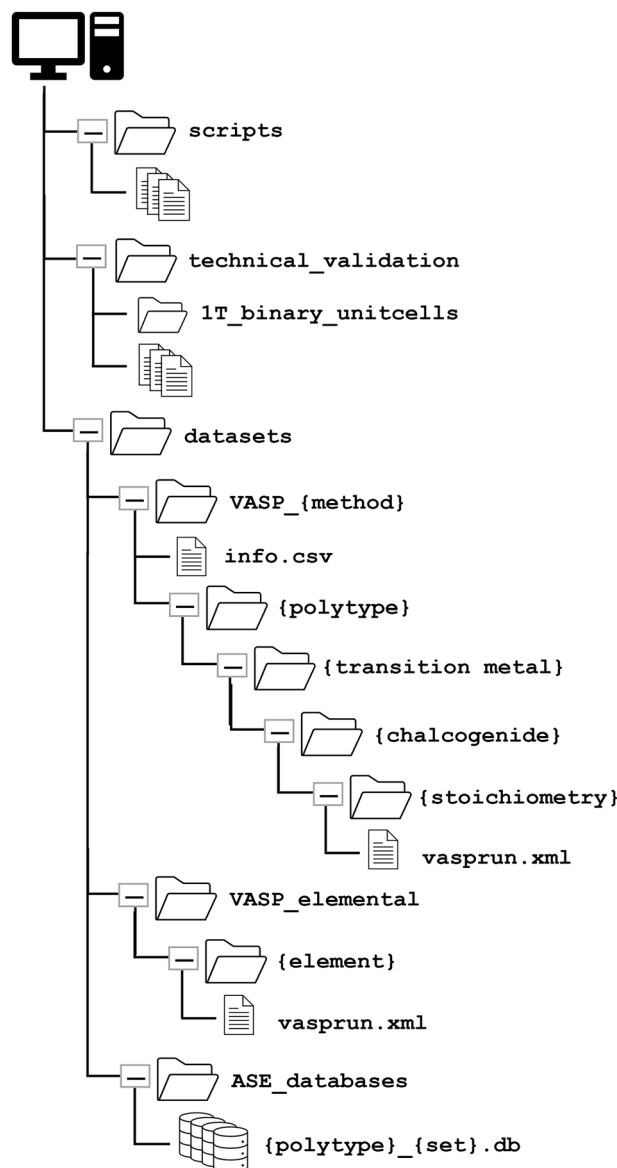
where  $E_{\text{M}}$ ,  $E_{\text{X}}$ , and  $E_{\text{Y}}$  are the per-atom energies of the elemental solid phases of the metal M and chalcogens X and Y and  $N_{\text{M}}$  is the number of metal atoms in the unit cell (16 in our case). We obtain mono-elemental structures from the Materials Project database<sup>32</sup> and relax them under the same approximations as used for relaxation of the ternary TMDs.

Binary TMDs have been experimentally observed with both octahedral (1T) and distorted octahedral (1T') coordination. To determine which polytype should be used in the calculation of  $E_{\text{form}}$ , we compute the energy-per-atom for binary structures with primitive 1T and 1T' unit cells, which lack the degrees of freedom to change phase during structural relaxations. The results of these calculations are shown in Table 1. The polytype with lower  $E_{\text{form}}$  is considered to be more stable for a given binary TMD. Binary TMDs with group 4 metals (Ti, Hf) were found to be more stable in the 1T polytype, while group 6 metals (Mo, W) produced greater stability with the 1T' polytype. The stability of binary TMDs with group 5 metals (V, Nb, Ta) was dependent on the chalcogenide present: 1T was more stable with S or Se and 1T' was more stable with Te. We note that the energy difference is small in many cases, indicating some degree of degeneracy in the polytype for TMD structures. We use the most stable structure to determine the initial pre-relaxed atomic arrangement for binary TMDs: when  $E(1\text{T}) > E(1\text{T}')$ , we use the 1T' polytype in the calculation of  $E_{\text{form}}$ , otherwise we use 1T.

### Data Records

We distribute MXY\_DB via Figshare<sup>46</sup>. Datasets include VASP xml files for the individual structural relaxations, VASP xml files for the spin-polarized calculations, and SQLite3 databases containing all structures of a given polytype, including intermediate steps taken during structural relaxation, generated using the atomic simulation environment (ASE) python library<sup>47</sup>. Each row in the database is composed of an ASE Atoms object, which has information on the elements, coordinates, cell, and periodic boundary conditions, along with a dictionary containing the total energy (eV) and forces on each atom (eV/Å). Python scripts used to collate the database and perform structural analysis are provided alongside the data.

The file structure of MXY\_DB is shown in Fig. 3. All the datasets are contained in the datasets folder. The simulation output is organized by calculation method (structural\_relaxation or spin\_polarization), polytype (1T or 2H), transition metal (Ti, V, Nb, Mo, Hf, Ta, or W), chalcogenide(s) (S, Se, Te, SSe, STe, or SeTe), and stoichiometry of the lowest-Z chalcogenide (numerical value). Each calculation method folder has a table containing information on the polytype, elements, stoichiometry, relative file path, and calculated properties to allow for search and filtering of the various calculations. VASP xml files for per-atom energies of the elemental solid phases used to compute  $E_{\text{form}}$  are provided in the VASP\_elemental folder, which is organized by element

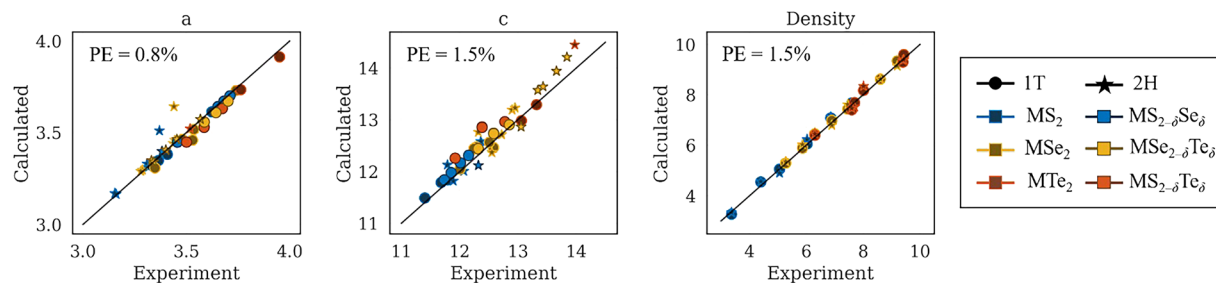


**Fig. 3** File structure of MXY\_DB.

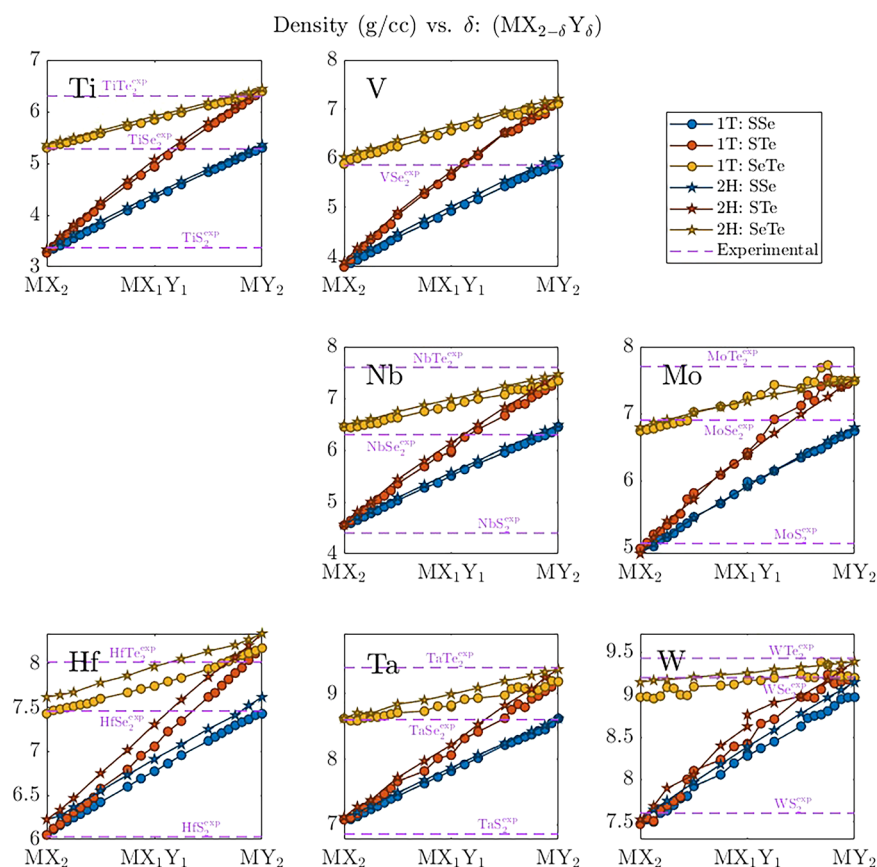
(Ti, V, Nb, Mo, Hf, Ta, W, S, Se, or Te). Four distinct ASE databases are in the ASE\_databases folder: minimized structures for 1T or 2H ( $\{\text{polytype}\}_{\text{relaxed}}.db$ ) and all structures generated during relaxation for 1T or 2H ( $\{\text{polytype}\}_{\text{all\_steps}}.db$ ). Python-based scripts used to generate the databases and analyze simulation data are in the scripts folder, along with a README.md describing each script. Information supporting our validation of the database, including unit cell computations for 1T and 1T' binary TMDs and experimental lattice parameters and densities, is included in the technical\_validation folder.

### Technical Validation

To validate our computed binary ( $\text{MX}_2$ ) and ternary ( $\text{MX}_{2-\delta}\text{Y}_\delta$ ) structures, we compare lattice parameters  $a$  and  $c$  and density against experimental values found in the literature. Experimental lattice parameters were obtained for 37 TMDs, 16 of which are ternary structures<sup>48,49</sup>. Comparisons between the experimental values and those arising from our relaxed structures are shown in Fig. 4. Note that we doubled the reported  $c$  values for 1T polytype structures to comparing to our 2-layer structures. Correlation between the experimental lattice parameters and our computed lattice parameters is very high, with percent errors of 0.8% and 1.5% and mean absolute errors (MAE) of 0.028 and 0.147 Å for  $a$  and  $c$ , respectively. While the MAE of  $a$  is the same for both 1T and 2H polytypes, the MAE of  $c$  is much lower for the 1T polytype: 0.070 versus 0.237 Å for 1T and 2H polytypes, respectively. Values for the experimental and computed primitive lattice parameters  $a$  and  $c$  for various structures in the database are given in MXY\_DB. The primitive lattice parameters for all structures in MXY\_DB are given in the info.csv file of the VASP\_structural\_relaxation dataset.

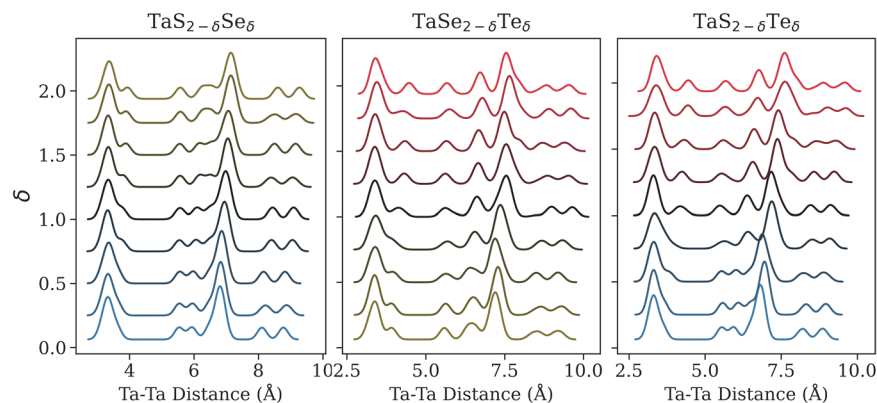


**Fig. 4** Comparison between experimental and computed parameters. These include lattice parameters  $a$  and  $c$  (Å) and density (g/cc) going from left to right. Percent error (PE) is shown in the upper left of each plot. Lattice parameters for 37 TMDs (16 ternary, 21 binary)<sup>48,49</sup> and densities of 20 TMDs (all binary)<sup>17,50</sup> were obtained from the literature (some by computation using lattice parameters). In instances where the literature value is reported for a single layer,  $c$  is doubled to compare with our two-layer cells, and where the polytype was not reported, we plot against calculated values for both polytypes.

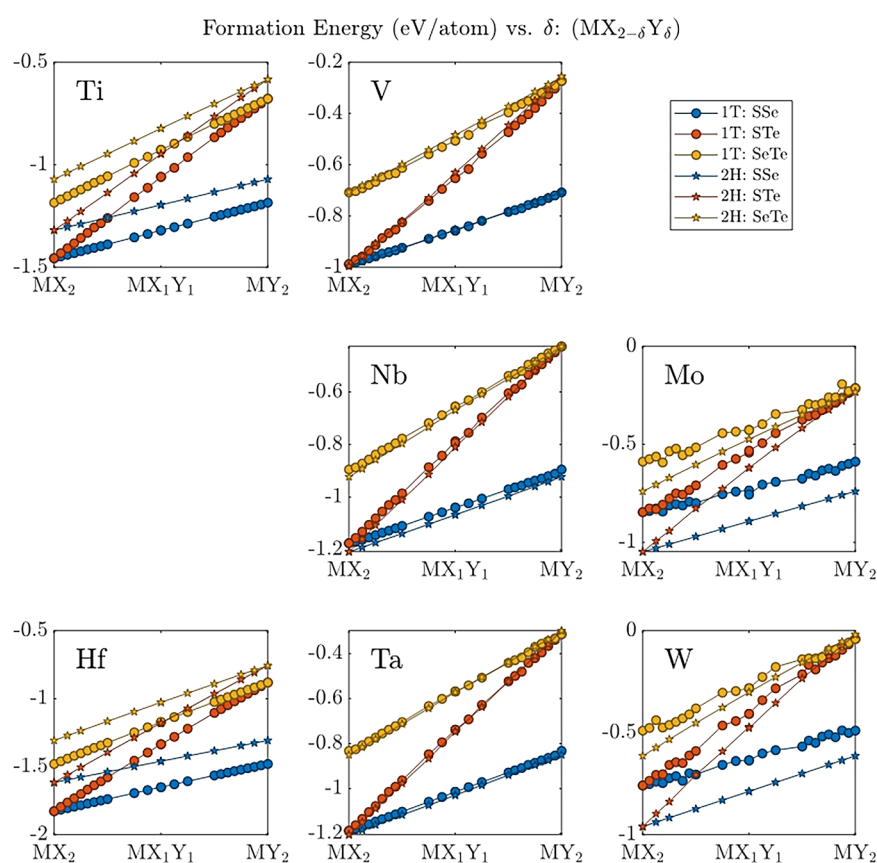


**Fig. 5** Computed densities of ternary TMDs. Density is in units of g/cc. Comparison is made to available experimental values (pink dotted lines, see *MX<sub>Y</sub>\_DB* for more details)<sup>17,50</sup>. The location of each plot corresponds to the location of the transition metal on the periodic table.

The densities of 14 binary TMDs were obtained from the CRC Handbook of Chemistry and Physics<sup>50</sup>, which didn't report polytype with their structures. We computed an additional 6 experimental densities using the 1T polytype lattice parameters reported by Lai *et al.*<sup>17</sup> with their corresponding stoichiometries. The only binary TMD structures for which we found no experimental density were  $\text{VS}_2$  and  $\text{VTe}_2$ . Figure 4 shows the density comparison between experimental and the corresponding relaxed binary structures; values for the experimental and computed densities (g/cc) are given in *MX<sub>Y</sub>\_DB*. No experimental density values for the ternary TMDs included in *MX<sub>Y</sub>\_DB* could be found in the literature. In addition, not all reported densities were associated with a polytype; therefore, we compare the reported value against both polytypes in the figure. Our computed values match well with experimental values, showing a MAE of 0.104 g/cc and MSE of 0.016 g/cc.

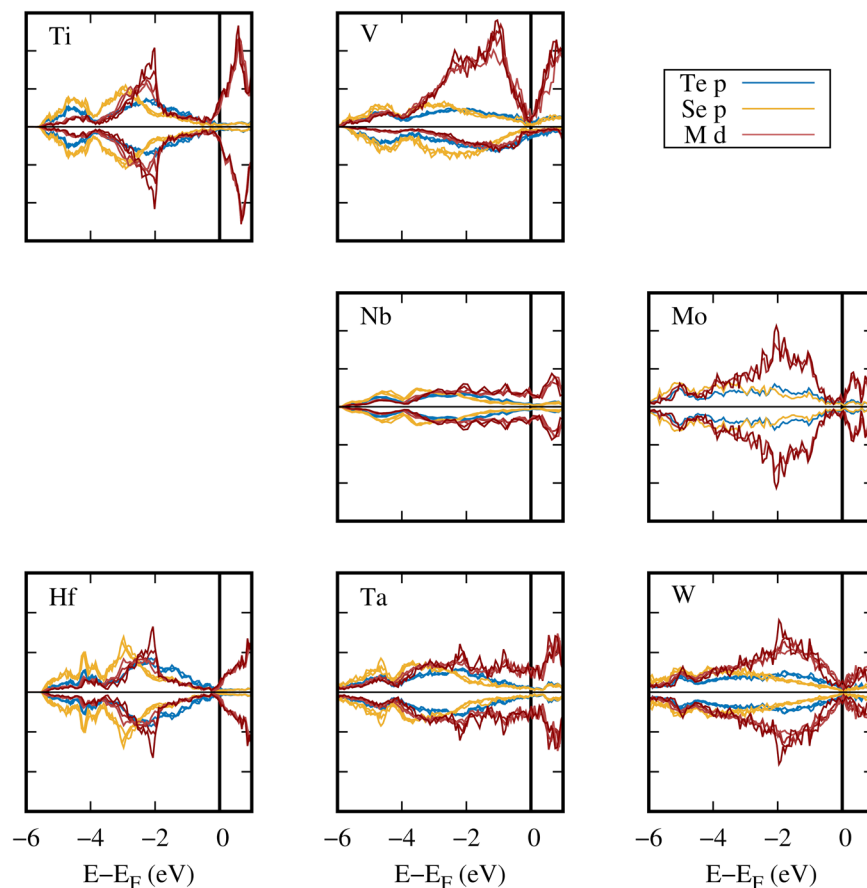


**Fig. 6** Radial distance function (RDF) of M-M distances for Ta-series ternary TMDs. The color gradient corresponds to  $\delta$ , where blue represents S, yellow Se, and red Te.



**Fig. 7** Formation energy vs.  $\delta$  for ternary TMDs in MXY\_DB. Formation energy is in units of eV/atom. The location of each plot corresponds to the location of its transition metal on the periodic table. The lighter chalcogenide is denoted as X, while the heavier chalcogenide is denoted as Y.

As the X:Y ratio in ternary TMDs increases, a linear change in density from that of  $\text{MX}_2$  to that of  $\text{MY}_2$  is expected. As shown in Fig. 5, the expected trend is observed for all metal-chalcogenide-polytype combinations examined in MXY\_DB. Slight deviations from linearity for individual structures are more frequent for structures containing group 6 metals (Mo and W). These deviations are the result of increased symmetry breaking during structural relaxations, leading to several local minima, and demonstrates the complicated dependency of structural properties on chalcogenide stoichiometry. In general, the high correlation between available experimental and our computed structural parameters, along with the linear transition in computed densities, indicates the quality of the structures in MXY\_TMD.



**Fig. 8** Electronic DOS of  $M\text{Se}_{2-\delta}\text{Te}_{\delta}$  for all  $\delta$ . The electronic DOS is projected onto chalcogenide  $p$  (Te in blue and Se in yellow) and transition metal  $d$  (red) orbitals.

### Usage Notes

VASP xml files can be read using common Python packages such as ASE (`ase.io.read('vasprun.xml')`) and pymatgen (`pymatgen.io.vasp.outputs.Vasprun('vasprun.xml')`), as well as visualization programs such as JMol and VMD. ASE databases can be read via `ase.db.connect('{polytype}_{set}.db')`. ASE Atoms objects can be converted to pymatgen Structure objects using `pymatgen.io.ase.AseAtomsAdaptor`.

The computational results provided in MXY\_DB can be used to examine structural changes, such as lattice distortions, stacking symmetry, interlayer distances, and distortion in the 1T polytype, arising from chalcogenide substitutions. For instance, stability of the 1T' phase is desired in order to manifest useful optical and electronic properties resulting from reduced symmetry in the distorted phase<sup>15</sup>. Phase classification is based on symmetries of the crystallographic cell, which are broken by the quasirandom substitution of chalcogenides in ternary TMDs. Ternary TMDs, instead, take on intermediate phases in high-dimensional space. Therefore, instead of direct classification of the phase, we identify the amount of distortion in the transition metal (M) lattice through examination of the radial distance function (RDF) of M-M distances, which provides a 1D descriptor of polytype distortion. Figure 6 shows the RDFs for the Ta-series 1T-like ternary TMDs, which demonstrate a transition from 1T to 1T' character with increased proportion of high-Z chalcogenide. The peak at 3.4 Å, corresponding to M-M distances in the first coordination shell, develops a shoulder with the addition of Se to  $\text{TaS}_{2-\delta}\text{Se}_{\delta}$ . As the transition progresses from Se to Te in  $\text{TaSe}_{2-\delta}\text{Te}_{\delta}$ , the shoulder separates into a distinct peak 4.5 Å. The distortion pattern involves the separation of metal layers along the  $b$  axis, leading to distinct M-M distances, as shown in Fig. 1b. A higher ratio of the first-to-second peak indicates fewer regions of distortion, while a larger separation represents increased magnitude of distortion between the regions. A Jupyter notebook demonstrating M-M RDF analysis, called `MM_distance_analysis.ipynb`, is included in the scripts folder.

The formation energy of each structure was calculated via Eq. (1) using energetics of the relaxed structures in VASP\_structural\_relaxations along with the elemental energies in VASP\_elemental. The raw values are included in info.csv under the Eform column and, for convenience, are plotted in Fig. 7. All systems in MXY\_DB have negative formation energies, indicating that the reactions to form them from elemental precursors are exothermic. All ternary systems in MXY\_DB show a near linear increase in formation energy with an increase in the proportion of heavier chalcogenide ( $\delta \rightarrow 2$  in  $\text{MX}_{2-\delta}\text{Y}_{\delta}$ ). Ternary TMDs with group 4 transition metals (Ti and Hf) show the 1T polytype to be more stable ( $\sim 0.2$  eV/atom) than the 2H polytype, while the opposite trend is observed for group 6 transition metals (Mo and W). Group 5 transition metals show near degeneracy of



the 1T and 2H polytypes, with V having a slight preference for 1T and Nb and Ta having a slight preference for 2H. These differences between groups are caused by interactions between the increasingly occupied transition metal  $d$  shells and the valence chalcogen  $p$  electrons as the transition metal changes from group 4 to 6.

As an illustration of the utility of MXY\_DB for analyzing compositional trends in material properties, we plot the electronic density of states (DOS) of  $\text{MSe}_{2-\delta}\text{Te}_\delta$  in Fig. 8. From this comparison, we see qualitatively different magnetic behaviors: there is spin polarization for  $M = V$  and not otherwise. Other interesting trends appear as the transition metal is varied: metal  $d$  hybridization with the chalcogen  $p$ -derived orbitals is much more pronounced for the 4- and 5 $d$  metals compared to the 3 $d$  metals, an effect we attribute to the larger  $d$  orbital radius, which results in increased spatial overlap with the valence  $p$  orbitals. The latter are split into two crystal field subbands the separation of which ranges from quite distinct (Hf, Ti, V) through modestly distinct (Nb, Ta) to barely discernible (W). The metal  $d$  DOS consists of components that share the same shape as the Te and Se upper crystal field valence subbands and excess states near the Fermi level. This behavior reflects covalent mixing and atomic-like occupation, respectively. The  $d$  orbitals of V adopt a local spin moment driven by onsite  $d$ - $d$  exchange interactions. An interesting feature is that the total  $d$  occupation on the V sites varies with the doping level  $\delta$ . Hence the existence of spin polarization can be adjusted by choice of transition metal, while the strength of the polarization can be controlled through the stoichiometry of the chalcogenide sublattice. The DOS at the Fermi level is also determined by the competition between onsite atomic effects and covalent interactions with the chalcogen layers resulting in a range of predicted metallicities from n-type semiconducting (Ti, Hf, Mo), through semimetallic (V, W), to metallic (Nb, Ta). We point out that these trends do not follow simple  $d$ -orbital filling arguments for  $M^{4+}$  transition metals that are commonly invoked<sup>13</sup>. The wide variety of electronic structures presented by the ternary TMDs in MXY\_DB provides a foundation for the further analysis of electronic trends resulting from tuning chalcogenide stoichiometries. Considering large regions of composition space consistently clarifies systematic physical effects that can be missed by studies that focus on only one material.

MXY\_DB is an open-source database of computed ternary TMD structures composed of 1T and 2H polytypes of seven transition metals (Ti, V, Nb, Mo, Hf, Ta, and W) with varying binary combinations of chalcogenides (S, Se, and Te). The database provides relaxed structures from which more detailed energetic analyses can be performed, a source for further data mining of the interesting structural properties present in ternary TMDs, and samples for training neural network potentials or other machine learning algorithms, among other functions invented in the creative researcher's mind. Ternary TMDs with a single chalcogenide component and two transition metal components and quaternary TMDs containing mixtures of three different chalcogenides are not included in this database, but represent possible directions for future computational examination by the methods described in this report. The database is openly available at <https://doi.org/10.6084/m9.figshare.21308157><sup>46</sup>.

### Code availability

Scripts used to collate the ASE database and perform structural analyses are located alongside the dataset. All scripts are written in Python (v3.9) and are described in the accompanying README.md.

Received: 12 October 2022; Accepted: 24 March 2023;

Published online: 30 May 2023

### References

- Manzeli, S., Ovchinnikov, D., Pasquier, D., Zayzev, O. V. & Kis, A. 2D transition metal dichalcogenides. *Nature Reviews Materials* **2**, 17033, <https://doi.org/10.1038/natrevmats.2017.33> (2017).
- Zhang, W., Zhang, Y., Qiu, J., Zhao, Z. & Liu, N. Topological structures of transition metal dichalcogenides: A review on fabrication, effects, applications, and potential. *InfoMat* **3**, 133–154, <https://doi.org/10.1002/inf2.12156> (2021).
- Wang, Y.-H., Huang, K.-J. & Wu, X. Recent advances in transition-metal dichalcogenides based electrochemical biosensors: A review. *Biosensors and Bioelectronics* **97**, 305–316, <https://doi.org/10.1016/j.bios.2017.06.011> (2017).
- Choi, W. *et al.* Recent development of two-dimensional transition metal dichalcogenides and their applications. *Materials Today* **20**, 116–130, <https://doi.org/10.1016/j.mattod.2016.10.002> (2017).
- Chen, B. *et al.* Transition metal dichalcogenides for alkali metal ion batteries: engineering strategies at the atomic level. *Energy & Environmental Science* **13**, 1096–1131, <https://doi.org/10.1039/c9ee03549d> (2020).
- Liang, Q., Zhang, Q., Zhao, X., Liu, M. & Wee, A. T. S. Defect Engineering of Two-Dimensional Transition-Metal Dichalcogenides: Applications, Challenges, and Opportunities. *ACS Nano* **15**, 2165–2181, <https://doi.org/10.1021/acsnano.0c09666> (2021).
- Wu, X., Zhang, H., Zhang, J. & Lou, X. W. Recent Advances on Transition Metal Dichalcogenides for Electrochemical Energy Conversion. *Advanced Materials* **33**, 2008376, <https://doi.org/10.1002/adma.202008376> (2021).
- Zhang, X. *et al.* Transition Metal Dichalcogenides for the Application of Pollution Reduction: A Review. *Nanomaterials* **10**, 1012, <https://doi.org/10.3390/nano10061012> (2020).
- Loh, L., Zhang, Z., Bosman, M. & Eda, G. Substitutional doping in 2D transition metal dichalcogenides. *Nano Research* **14**, 1668–1681, <https://doi.org/10.1007/s12274-020-3013-4> (2021).
- Zhou, J. *et al.* A library of atomically thin metal chalcogenides. *Nature* **556**, 355–359, <https://doi.org/10.1038/s41586-018-0008-3> (2018).
- Susarla, S. *et al.* Quaternary 2D Transition Metal Dichalcogenides (TMDs) with Tunable Bandgap. *Advanced Materials* **29**, 1702457, <https://doi.org/10.1002/adma.201702457> (2017).
- Naumis, G. G. in *Synthesis, Modeling, and Characterization of 2D Materials, and Their Heterostructures* 77–109 (Elsevier, 2020).
- Voiry, D., Mohite, A. & Chhowalla, M. Phase engineering of transition metal dichalcogenides. *Chemical Society Reviews* **44**, 2702–2712, <https://doi.org/10.1039/C5CS00151J> (2015).
- Liu, L. *et al.* Phase-selective synthesis of 1T' MoS<sub>2</sub> monolayers and heterophase bilayers. *Nature materials* **17**, 1108–1114, <https://doi.org/10.1038/s41563-018-0187-1> (2018).
- Tan, S. J. *et al.* Chemical stabilization of 1T' phase transition metal dichalcogenides with giant optical Kerr nonlinearity. *Journal of the American Chemical Society* **139**, 2504–2511, <https://doi.org/10.1021/jacs.6b13238> (2017).
- Lu, N. *et al.* Atomic and electronic structures of WTe<sub>2</sub> probed by high resolution electron microscopy and ab initio calculations. *The Journal of Physical Chemistry C* **120**, 8364–8369, <https://doi.org/10.1021/acs.jpcc.6b01044> (2016).
- Lai, Z. *et al.* Metastable 1T'-phase group VIB transition metal dichalcogenide crystals. *Nature Materials*, 1–8, <https://doi.org/10.1038/s41563-021-00971-y> (2021).

18. Cho, S. *et al.* Phase patterning for ohmic homojunction contact in MoTe<sub>2</sub>. *Science* **349**, 625–628, <https://doi.org/10.1126/science.aab3175> (2015).
19. Tan, Y. *et al.* Controllable 2H-to-1T' phase transition in few-layer MoTe<sub>2</sub>. *Nanoscale* **10**, 19964–19971, <https://doi.org/10.1039/c8nr06115g> (2018).
20. Berger, A. N. *et al.* Temperature-driven topological transition in 1T'-MoTe<sub>2</sub>. *npj Quantum Materials* **3**, <https://doi.org/10.1038/s41535-017-0075-y> (2018).
21. Keum, D. H. *et al.* Bandgap opening in few-layered monoclinic MoTe<sub>2</sub>. *Nature Physics* **11**, 482–486, <https://doi.org/10.1038/nphys3314> (2015).
22. Pace, S. *et al.* Synthesis of Large-Scale Monolayer 1T'-MoTe<sub>2</sub> and Its Stabilization via Scalable hBN Encapsulation. *ACS Nano* **15**, 4213–4225, <https://doi.org/10.1021/acsnano.0c05936> (2021).
23. Zhang, Q. *et al.* Iodine-Mediated Chemical Vapor Deposition Growth of Metastable Transition Metal Dichalcogenides. *Chemistry of Materials* **29**, 4641–4644, <https://doi.org/10.1021/acs.chemmater.6b05065> (2017).
24. Tang, S. *et al.* Electronic structure of monolayer 1T'-MoTe<sub>2</sub> grown by molecular beam epitaxy. *APL Materials* **6**, 026601, <https://doi.org/10.1063/1.5004700> (2018).
25. Zhang, C. *et al.* Charge mediated reversible metal–insulator transition in monolayer MoTe<sub>2</sub> and W<sub>x</sub>Mo<sub>1-x</sub>Te<sub>2</sub> alloy. *ACS Nano* **10**, 7370–7375, <https://doi.org/10.1021/acsnano.6b00148> (2016).
26. Chen, Q. *et al.* Composition-Gradient-Mediated Semiconductor–Metal Transition in Ternary Transition-Metal-Dichalcogenide Bilayers. *ACS Applied Materials & Interfaces* **12**, 45184–45191, <https://doi.org/10.1021/acsmi.0c13104> (2020).
27. Zhao, H. *et al.* Integration of machine learning approaches for accelerated discovery of transition-metal dichalcogenides as HgO sensing materials. *Applied Energy* **254**, 113651, <https://doi.org/10.1016/j.apenergy.2019.113651> (2019).
28. Gossett, E. *et al.* AFLOW-ML: A RESTful API for machine-learning predictions of materials properties. *Computational Materials Science* **152**, 134–145, <https://doi.org/10.1016/j.commatsci.2018.03.075> (2018).
29. Behler, J. First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angewandte Chemie International Edition* **56**, 12828–12840, <https://doi.org/10.1002/anie.201703114> (2017).
30. Wang, A. Y.-T. *et al.* Machine Learning for Materials Scientists: An Introductory Guide toward Best Practices. *Chemistry of Materials* **32**, 4954–4965, <https://doi.org/10.1021/acs.chemmater.0c01907> (2020).
31. De Pablo, J. J. *et al.* New frontiers for the materials genome initiative. *npj Computational Materials* **5**, <https://doi.org/10.1038/s41524-019-0173-4> (2019).
32. Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **1**, 011002, <https://doi.org/10.1063/1.4812323> (2013).
33. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018, <https://doi.org/10.1038/sdata.2016.18> (2016).
34. Choudhary, K. *et al.* The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. *npj Computational Materials* **6**, <https://doi.org/10.1038/s41524-020-00440-1> (2020).
35. Brune, V. *et al.* Influence of the choice of precursors on the synthesis of two-dimensional transition metal dichalcogenides. *Dalton Transactions* **50**, 12365–12385, <https://doi.org/10.1039/d1dt01397a> (2021).
36. Samadi, M. *et al.* Group 6 transition metal dichalcogenide nanomaterials: synthesis, applications and future perspectives. *Nanoscale Horizons* **3**, 90–204, <https://doi.org/10.1039/C7NH00137A> (2018).
37. Toh, R. J., Sofer, Z. & Pumera, M. Catalytic properties of group 4 transition metal dichalcogenides (MX<sub>2</sub>; M = Ti, Zr, Hf; X = S, Se, Te). *Journal of Materials Chemistry A* **4**, 18322–18334, <https://doi.org/10.1039/c6ta08089h> (2016).
38. Van de Walle, A. *et al.* Efficient stochastic generation of special quasirandom structures. *Calphad* **42**, 13–18, <https://doi.org/10.1016/j.calphad.2013.06.006> (2013).
39. Zhang, X., Lai, Z., Ma, Q. & Zhang, H. Novel structured transition metal dichalcogenide nanosheets. *Chemical Society Reviews* **47**, 3301–3338, <https://doi.org/10.1039/c8cs00094h> (2018).
40. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical review B* **54**, 11169, <https://doi.org/10.1103/PhysRevB.54.11169> (1996).
41. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical review b* **59**, 1758, <https://doi.org/10.1103/PhysRevB.59.1758> (1999).
42. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Physical review letters* **77**, 3865, <https://doi.org/10.1103/PhysRevLett.77.3865> (1996).
43. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *The Journal of chemical physics* **132**, 154104, <https://doi.org/10.1063/1.3382344> (2010).
44. Anisimov, V. I., Zaanen, J. & Andersen, O. K. Band theory and Mott insulators: Hubbard U instead of Stoner I. *Physical Review B* **44**, 943–954, <https://doi.org/10.1103/physrevb.44.943> (1991).
45. Kirchner-Hall, N. E., Zhao, W., Xiong, Y., Timrov, I. & Dabo, I. Extensive Benchmarking of DFT+U Calculations for Predicting Band Gaps. *Applied Sciences* **11**, 2395, <https://doi.org/10.3390/app11052395> (2021).
46. Muller, S., Prange, M., Lu, Z., Rosenthal, W. & Pope, J. MXY\_DB, An open database of computed bulk ternary transition metal dichalcogenides. *Figshare* <https://doi.org/10.6084/m9.figshare.21308157.v1> (2022).
47. Hjorth Larsen, A. *et al.* The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **29**, 273002, <https://doi.org/10.1088/1361-648x/aa680e> (2017).
48. Bastos, C. M., Besse, R., Da Silva, J. L. & Sipahi, G. M. Ab initio investigation of structural stability and exfoliation energies in transition metal dichalcogenides based on Ti-, V-, and Mo-group elements. *Physical Review Materials* **3**, 044002, <https://doi.org/10.1103/PhysRevMaterials.3.044002> (2019).
49. Lévy, F. & Hulliger, F. *Structural Chemistry of Layer-Type Phases*. Vol. 5 (Springer Netherlands, 1976).
50. Haynes, W. M. *CRC handbook of chemistry and physics*. (CRC press, 2014).

## Acknowledgements

The authors thank Dr. Amra Peles for insightful discussions into the examination of ternary TMDs. This research was supported by the Laboratory Directed Research and Development Program under the Chemical Dynamics Initiative at Pacific Northwest National Laboratory (PNNL), which is operated by Battelle for the U.S. Department of Energy under contract DE-AC06-76RLO. The research was performed using resources available through Research Computing at PNNL.

## Author contributions

S.M. performed and analyzed the structural relaxations and spin polarization calculations. M.P. performed and analyzed the spin polarization calculations. J.A.B. analyzed structural properties and curated the data. All authors contributed to writing of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to S.E.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



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

© Battelle Memorial Institute 2023