THE JOURNAL OF PHYSICAL CHEMISTRY C
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCC
Article

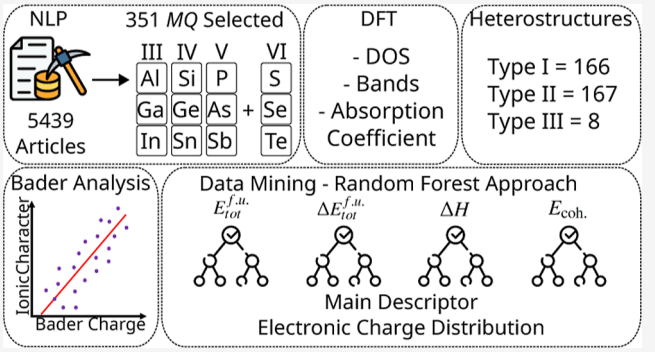
Bridging Text Mining and Quantum Simulations for the Design of 2D Monochalcogenide Materials

Mateus B. P. Querne, Marco A. M. T. Machado, Ronaldo C. Prati, Natan M. Regis, Matheus P. Lima, and Juarez L. F. Da Silva*

Cite This: https://doi.org/10.1021/acs.jpcc.5c04625
Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The formulation of a structured framework dedicated to the systematic review of the literature, identification of potential compounds, theoretical quantum chemistry characterization, and assessment of the significance of stability descriptors is essential for accelerating the discovery of two-dimensional materials for technological applications. In this work, we selected the two-dimensional monochalcogenides (MQ), which have attracted increased interest due to their potential applications in future devices, as prototype materials for the present investigation. Our framework started with a natural language processing analysis of more than 5400 articles, revealing a growing research interest in the two-dimensional MQ compounds, especially in electronics, energy, and fundamental studies. This led to the selection of 27 diverse MQ compounds across 13 distinct two-dimensional structural phases for density functional theory calculations, resulting in an extensive database of physicochemical properties. We evaluated formation enthalpies, uncovering clear stability trends (e.g., stability declines with heavier chalcogens, dynamic robustness of GeSe), and evaluated equilibrium lattice parameters, noting predictable expansions and significant anisotropies. Bader charge analysis offered insights into charge-transfer and ionicity trends. Our electronic structure analysis identified band gaps (direct or indirect, depending on the M element group and the weight of the chalcogen) and optical absorption anisotropies significantly influenced by crystal symmetry and spin orbit coupling. Importantly, band alignment calculations classified all possible heterojunctions as type I, II, or III, underscoring their extensive potential for future optoelectronic device development. Additionally, we incorporated machine learning, using a random forest approach, along with our density functional theory calculations to accurately forecast trends in energetic properties. This analysis identified the electronic charge as a highly significant stability descriptor among the 34 descriptors, underscoring its importance for future machine learning endeavors. Hence, this study offers a streamlined framework for the characterization and discovery of promising two-dimensional materials, highlighting the synergy between data-driven analysis and quantum-based simulations.



1. INTRODUCTION

The isolation of a monolayer of graphene[1,2] has instigated the emergence of a new research field focused on two-dimensional (2D) materials, propelled by graphene's exceptional properties, which include an extremely high surface-to-volume ratio,[3] macroscopic expressions of its hexagonal lattice symmetry,[4] extraordinary charge carrier mobility,[5] and other notable attributes. However, the absence of an intrinsic band gap in graphene[6] constrains its utility in semiconductor applications, presenting a significant obstacle to the development of practical devices. Consequently, the exploration of alternative 2D materials with tunable physicochemical properties has become crucial for the advancement of the discipline.

In this context, transition-metal dichalcogenides (TMDs), which are 2D materials consisting of a transition-metal atom (M) interposed between two layers of chalcogen (Q = S, Se, Te) species (MQ_2), have emerged as promising semiconductor candidates to complement carbon-based 2D materials. For example, TMDs effectively address the electronic deficiencies inherent in graphene while demonstrating additional functional characteristics such as superconductivity,[7] complex reciprocal space topology,[8] among other properties of high interest. Among the MQ_2 compounds, certain materials, such as MoS_2 and WS_2, are particularly significant[9,10] due to their applicability in diverse fields, including biomedicine,[11] sensing,[12] photovoltaics,[13] etc.

Received: July 3, 2025
Revised: September 4, 2025
Accepted: October 1, 2025

An additional category of 2D semiconductors encompasses the monochalcogenides,[14,15] which have attracted considerable research attention due to their unique physical−chemical properties. These compounds are defined by a generalized chemical formula MQ, where M represents a cationic element derived from groups III, IV, or V of the periodic table, while Q is a chalcogen element. The compositional adaptability provided by the simple 1:1 ratio facilitates the creation of a diverse range of 2D materials. For example, MQ are characterized by substantial chemical tunability[16] that allows the alteration of electronic and optical properties by doping or alloying; remarkable carrier mobility[17] crucial for the development of high-performance electronic circuits; and robust optical absorption,[18] making them promising contenders for photodetector and energy harvesting technologies.

Experimental investigations have highlighted the potential of MQ compounds. For example, Wong et al.[19] reported ultrahigh plasticity in InSe, while Chang et al.[20] elucidated its synthesis through chemical vapor deposition. In addition to this, Hu et al.[21] developed GaSe nanosheets employing a combination of mechanical exfoliation and solvent-assisted techniques, highlighting its promise for photodetector applications. Furthermore, Vaughn et al.[22] achieved the synthesis of GeS and GeSe nanosheets using a one-pot method, illustrating their aptitude for light-harvesting applications. In a similar study, Zhang et al.[23] produced SnS nanosheets by physical vapor deposition, while Huang et al.[24] employed liquid phase exfoliation to generate SnSe, thus illustrating its relevance in the fields of optoelectronics and photovoltaics.

In theoretical explorations, density functional theory (DFT) calculations have become essential to elucidate the structure−property relationships of 2D materials. For example, Akay and Kocak[25] analyzed the temperature-dependent carrier mobility in GeSe, highlighting its potential as a high-performance nanomaterial. Xu et al.[26] investigated the electrocaloric effect in group IV monochalcogenides (GeS, GeSe, SnS, and SnSe), demonstrating their promise for solid-state cooling applications. Similarly, Demirci et al.[27] showed that group III monochalcogenides possess highly suitable electronic and optical properties for nanoelectronics and photonics. More recently, Sheng and Wang[28] identified TlX (where X = S, Se, Te) as a compelling class of valleytronic semiconductors, expanding the range of potential 2D materials for next-generation electronic applications. Collectively, these studies underscore the versatility of DFT in guiding the design and discovery of novel 2D materials with tailored functionalities.

Therefore, in recent 20 years, the volume of publications on 2D materials, using theoretical, experimental, or hybrid approaches, has grown at an unprecedented rate. This rapid expansion reflects both the increasing interest in novel low-dimensional systems and the accelerated development of computational and experimental methodologies. However, the sheer number of studies now presents a significant challenge: keeping up with the latest discoveries, assessing the reliability of reported results, and identifying genuinely promising materials have become increasingly difficult tasks. Traditional literature surveys are often insufficient to maintain an up-to-date understanding, particularly given the multidimensional nature of material properties, synthesis methods, and application domains. Thus, the rapidly increasing number of publications on 2D materials requires the development of scalable computational tools for systematic literature characterization, data extraction, and data analysis.[29]

To mitigate these limitations, natural language processing (NLP) techniques have been implemented for automated literature extraction, facilitating faster retrieval and broader analytical insights.[30,31] This methodology also helps prioritize materials with significant experimental and theoretical potential. To further elucidate the relationships between structure and properties, explainable artificial intelligence (XAI) tools have been integrated into the process.[32] Although machine learning (ML) models provide predictive capabilities, their "black-box" nature can obscure comprehension of the underlying physical principles. XAI techniques, particularly SHAP values (SHapley Additive exPlanations),[33] are used to interpret model outputs by ascribing predictions to specific atomic-scale features. By amalgamating NLP-derived bibliometric insights with feature attribution via XAI, we have established a comprehensive framework for the rational discovery of 2D monochalcogenides.

In this study, we integrate NLP algorithms with DFT calculations to explore the physicochemical characteristics of 2D MQ materials. Through NLP-guided screening, we have identified 13 crystal structures, pertaining to space groups (SGs) *Aem*2, *C*2/*m*, $P\bar{1}\alpha$, $P\bar{1}\beta$, $P2_1/c$, $P\bar{3}m1\alpha$, $P\bar{3}m1\beta$, $P4/nmm$, $P\bar{6}m2$, *Pbcm*, *Pmmn*, $Pmn2_1$, and *Pmna*. Additionally, we considered 27 MQ compositions, with M corresponding to Al, Ga, In, Si, Ge, Sn, P, As, Sb, and Q aligning with S, Se, Te. DFT calculations were executed to assess their relative energetic stability, which was further corroborated by phonon dispersion analysis to ensure structural stability.

Our results indicate that group III monochalcogenides stabilized predominantly in the crystal phases of the space groups $P\bar{1}\beta$ and $P\bar{6}m2$, while group IV elements exhibit a preference for $P\bar{3}m1\alpha$, $Pmn2_1$, and *Pmna*. In contrast, the elements of group V are predisposed to favor $P\bar{1}\beta$ and $P2_1/c$ SGs as their ground states. Bader charge analysis demonstrates that charge transfer from M to Q diminishes sequentially along the chalcogen series (S → Se → Te). Most compounds are characterized by indirect band gaps within the energy range of 0.21−3.23 eV, which corroborates the formation of 166 type I, 177 type II, and 8 type III heterojunctions. Furthermore, optical analyses reveal a redshift in absorption peak positions that correlates with an increase in the chalcogen atomic number, accompanied by pronounced anisotropy in group V compounds. Moreover, data mining algorithms have identified critical atomic-scale descriptors that impact energetic properties, including total energy, relative energy, formation enthalpy, and cohesive energy. An observed correlation between charge transfer and the ionic character indicates a maximum ionic contribution at 21%. Notably, both site and interstitial charge densities have been recognized as the predominant descriptors dictating energetic trends.

2. THEORETICAL APPROACH AND COMPUTATIONAL DETAILS

This study focuses on the characterization and design of 2D monochalcogenides for optoelectronic applications by integrating multiple computational methodologies: (i) A review of the literature based on NLP to identify the most relevant compositions and potential 2D structures for monochalcogenides; (ii) DFT[34,35] calculations to determine the most stable structural configurations; (iii) DFT characterization of the most significant physicochemical properties pertinent to optoelectronic applications; (iv) ML models application in a set of descriptors to identify their impact on the material stability. Each step, along with additional details, is shown in the computational

https://doi.org/10.1021/acs.jpcc.5c04625
J. Phys. Chem. C XXXX, XXX, XXX−XXX

protocol depicted in Figure 1. Subsequently, we provide a summary of the techniques employed in this research.

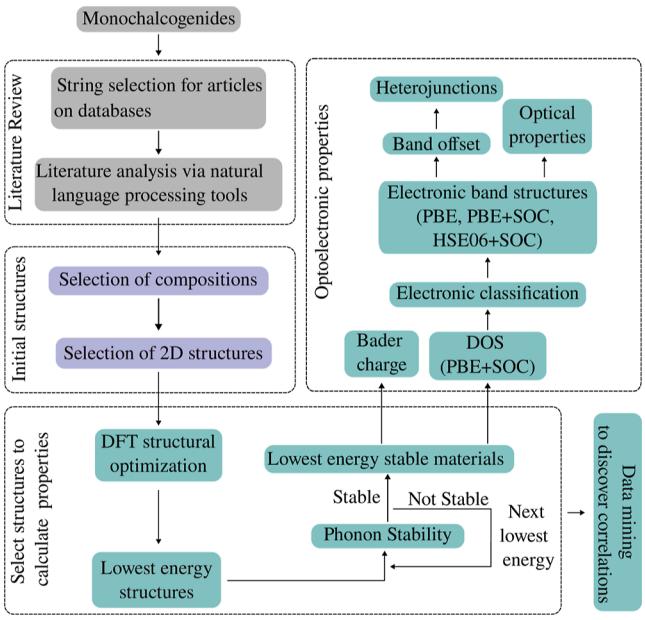


Figure 1. Flowchart illustrating the sequence of steps in our theoretical investigation. The process encompasses a thorough literature review, selection of 2D monochalcogenide compounds, rigorous energy and stability screening, and characterization of the most important physical−chemical properties.

2.1. Total Energy Calculations. Our calculations are based on the DFT[34,35] framework, which is widely used in the characterization and design of new materials.[36,37] The projector augmented-wave (PAW) method,[38,39] as implemented in the version 5.4.4 of the Vienna Ab initio Simulation Package (VASP),[40,41] was used to describe the interactions between the core and valence electrons. Structural optimizations of the 2D structures were performed using the generalized gradient approximation (GGA) as proposed by Perdew−Burke−Ernzerhof (PBE)[42] for the exchange−correlation energy functional. The characterization of the electronic properties, including the density of states and band structures, was conducted incorporating spin−orbit coupling (SOC) effects for the valence states, which are crucial for accurately determining electronic properties in compounds consisting of heavy element species. To improve the accuracy of fundamental band gaps, we used the hybrid exchange−correlation functional proposed by Heyd−Scuseria−Ernzerhof (HSE06)[43] and applied scissors operators with SOC effects over the PBE approximation to evaluate the optical properties of the dielectric tensor.[44,45]

A total energy convergence criterion of 1×10^{-6} eV was used to ensure convergence of the Kohn−Sham (KS) self-consistent cycle, while the equilibrium structures for the investigated 2D systems consider a force threshold for atomic forces in each atom of 0.01 eV Å^{-1} by minimizing the forces and the stress tensor simultaneously, but constraining the lattice parameter perpendicular to the monolayer plane. The plane-wave energy cutoff was set as twice the maximum recommended value among the PAW projectors used in each compound to minimize errors from the Pulay stress[46] in structural optimizations. However, for static calculations, a lower cutoff was used, set approximately 12.5% higher than the one recommended by the PAW projectors.

A vertical vacuum distance of 20 Å was applied to prevent spurious interactions, while the Monkhorst−Pack scheme defines the **k**-points and weights sampling the reciprocal space considering a **k**-length of R_k = 35 Å. To calculate phone band structures, we considered the supercell approach as implemented in the phonopy code.[47] The tight convergence criterion necessary for these calculations considers 1×10^{-8} eV to break the KS cycle, while the size of the supercell ranged from $2 \times 2 \times 1$ to $4 \times 4 \times 1$ depending on the structure with atomic displacements of 0.01 Å for the numerical evaluation of the Hessian matrix. The electronic Supporting Information file includes detailed information on all selected PAW projectors, along with additional relevant computational parameters.

2.2. Physical−Chemical Descriptors. To identify the physicochemical descriptors that most significantly influence the energetic stability of 2D monochalcogenides, we performed a correlation analysis between a broad set of structural and electronic features derived from first-principles calculations. Although our study focuses on monochalcogenides, the methodology is general and may be extended to other families of 2D materials. The physicochemical descriptors were extracted from DFT simulations considering PBE functional and include lattice parameters, bond lengths, bond angles, valence electron configurations, and charge distribution metrics such as Bader charges and interstitial electron densities, among others. In parallel, four energetic descriptors were considered: total energy, relative energy (with respect to the lowest-energy polymorph), formation enthalpy, and cohesive energy. The numerical values for all physicochemical and energetic descriptors are provided in Supporting Information File 2 (in CSV format), while Supporting Information File 1 contains detailed definitions and descriptions of each descriptor.

2.3. Explainable Machine Learning Analysis. To analyze the relationship between descriptors and energetic and electronic properties, we computed SHAP values[33] using a predictive model. SHAP values are a powerful tool for interpreting the output of machine learning models, especially when dealing with properties that have many nonlinearities, such as structural, energetic, electronic, and excitonic properties. Unlike correlation, which only measures the strength and direction of a linear relationship between two variables, SHAP values provide a detailed understanding of a model's predictions by quantifying the contribution of each feature to the final output. This is particularly useful in materials science, where the relationship between structural properties and material behavior can be complex and nonlinear. Using SHAP values, researchers can gain insight into how different structural parameters impact the properties of a material.

SHAP values are based on game theory, specifically the Shapley value concept. This concept was developed to fairly distribute the payouts among players in a cooperative game. In the context of machine learning, the features of a data set are considered players, and the prediction function is the game. The Shapley value for a feature is calculated as the average marginal contribution of that feature in all possible permutations of the other features. The marginal contribution of a feature is the difference between the prediction with and without that feature. By averaging the marginal contributions across all permutations, the SHAP values provide a unique and consistent way to quantify the importance of each feature in making a specific prediction. This approach ensures that the Shapley values satisfy

https://doi.org/10.1021/acs.jpcc.5c04625
J. Phys. Chem. C XXXX, XXX, XXX−XXX

several desirable properties, such as local accuracy, missingness, and consistency, making them a reliable and interpretable measure of the importance of features with respect to the target variables.

In our analysis, we have chosen to use the Random Forest algorithm as the predictive model to determine SHAP values. This decision is based on several factors, including robustness, relatively few parameters, and good performance of Random Forest on tabular data. Random Forest is an ensemble learning method that combines multiple decision trees to improve prediction accuracy and reduce overfitting. Its robustness makes it a reliable choice for handling complex and noisy data, while its simplicity in terms of hyperparameters means that it is relatively easy to tune and optimize. Furthermore, Random Forest has been shown to perform well on a wide range of tabular data sets, making it a versatile tool for predictive modeling. Using Random Forest as the predictive model, we can ensure that our analysis is accurate and reliable, providing valuable insights into the relationships between descriptors and energetic and electronic properties.

In addition to using Random Forest as the predictive model, we also ensure the validity and reliability of our analysis by computing the performance using data from the out-of-sample. Specifically, we used stratified cross-validation, where the folds were stratified according to the unitary chemical composition. This approach ensures that each fold is a representative sample of the overall data set with a similar number of unitary chemical composition in each fold. This reduces the risk of bias and improves the generalizability of the results. By evaluating the performance using out-of-sample data, we can obtain a more accurate estimate of the model's predictive ability, which in turn increases our confidence in the SHAP values and the insights they provide.

3. RESULTS AND DISCUSSION

3.1. Bibliometric Analysis Using Natural Language Processing. To review the literature, we query the Web of Science database[48] using the terms "(AlS OR AlSe OR AlTe OR GaS OR GaSe OR GaTe OR InS OR InSe OR InTe OR SiS OR SiSe OR SiTe OR GeS OR GeSe OR GeTe OR SnS OR SnSe OR SnTe OR PS OR PSe OR PTe OR AsS OR AsSe OR AsTe OR SbS OR SbSe OR SbTe) AND ("2D materials" OR "two-dimensional materials" OR "2D layered materials")", and retrieve all metadata information, including title, authors, abstract, publication outlet, etc. This query returned a total of 5439 articles published before September 2023. Upon analyzing the retrieved data, we notice that several papers are outside our scope. The reason is the natural ambiguity of the language mistake due to some chemical compound formula being also a word with meaning (e.g., the chemical formula GaTe and the word gate).

To address this issue, we developed an NLP classifier to separate the articles that were suitable for our analysis from the others. To this end, the title and abstract are converted to numerical vectors using sentence transformers (ST).[49] ST is a natural language processing model that excels in understanding the semantic meaning and relationships within the text using the transformer architecture.[50] The output of ST is a dense vector suitable for downstream tasks, such as text classification. Then, we use the Sentence Transformer Fine Tuning model (SetFit)[51] to classify the articles into relevant and irrelevant ones. SetFit uses a contrastive loss function to fine-tune the ST using pairs of relevant and irrelevant examples. The computation of the contrastive loss function depends on the relevance of the pair of articles. For relevant (positive) pairs, the loss is simply the squared distance between their vectors, creating an attractive force to pull them closer. For irrelevant (negative) pairs, a repulsive force is applied only if their distance is less than a predefined margin (m); if they are already far enough apart, their loss contribution is zero.[51] For our study, we used the default value $m = 0.5$ of the parameter. The training process then minimizes the sum of these competing forces, effectively organizing the feature space by clustering relevant items and separating irrelevant ones. One positive characteristic of SetFit is that it requires few labeled examples to be trained. By forming pairs, the number of training examples grows quadratically with the number of source documents, drastically reducing the need for many initial labeled documents.[51] We trained SetFit in two rounds. First, we manually selected five articles in each category (relevant and irrelevant). A first model is trained and used to classify all the retrieved articles. Then, we manually inspect ten articles classified as relevant or irrelevant, marking the incorrect classifications. A second model was trained and used to reclassify all the recovered articles. We manually checked for 20 random articles in each category and did not notice any further misclassifications. The final list contains 1257 articles.

Lastly, we apply several NLP techniques to analyze the list of relevant articles. We used the Google Gemini Pro large language model (LLM) to classify the articles according to the application area and type of study, as well as extract the mentions of monochalcogenides in the titles and abstracts of the articles. Gemini LLM is comparable to GPT4.0 in handling complex language tasks,[52] and offers a free level of 1500 API calls per day. Figure 2 presents some insight into this analysis.

The top panel of Figure 2 shows a histogram of the articles according to the area of application. Articles are classified into eight main areas: biomedical applications, electronics and devices, energy conversion and storage, fundamental studies and characterization, heterostructures and composites, optoelectronics, synthesis and processing, and tribology and mechanical applications. The definition of each category is included in the Supporting Information. In general, we can observe a growing trend in publications related to monochalcogenides. The areas of electronics and devices, energy conversion and storage, synthesis and processing, and fundamental studies and characterization also present a growing trend over the years.

Regarding the type of study, we categorize articles in the categories computational studies, experimental studies, integrated experimental and computational studies, proof-of-concept and prototype demonstrations, and review and perspective articles. The definitions of the types of study are also in the Supporting Information. The bottom panel of Figure 2 shows a co-occurrence graph with respect to the type of study and the application area. In general, we can observe more experimental studies in the areas of optoelectronics, electronic devices, and fundamental studies and characterization. Computational studies show a high prevalence of fundamental studies and characterization, followed by electronic devices, heterostructures, and composite and energy storage and conservation. Mixed theoretical and experimental studies show a similar trend in experimental studies, while proof-of-concept and prototype demonstrations focus on electronic devices and optoelectronics. Reviews and perspectives have a more homogeneous distribution, although few articles are related to biomedical and mechanical applications.

https://doi.org/10.1021/acs.jpcc.5c04625
J. Phys. Chem. C XXXX, XXX, XXX−XXX

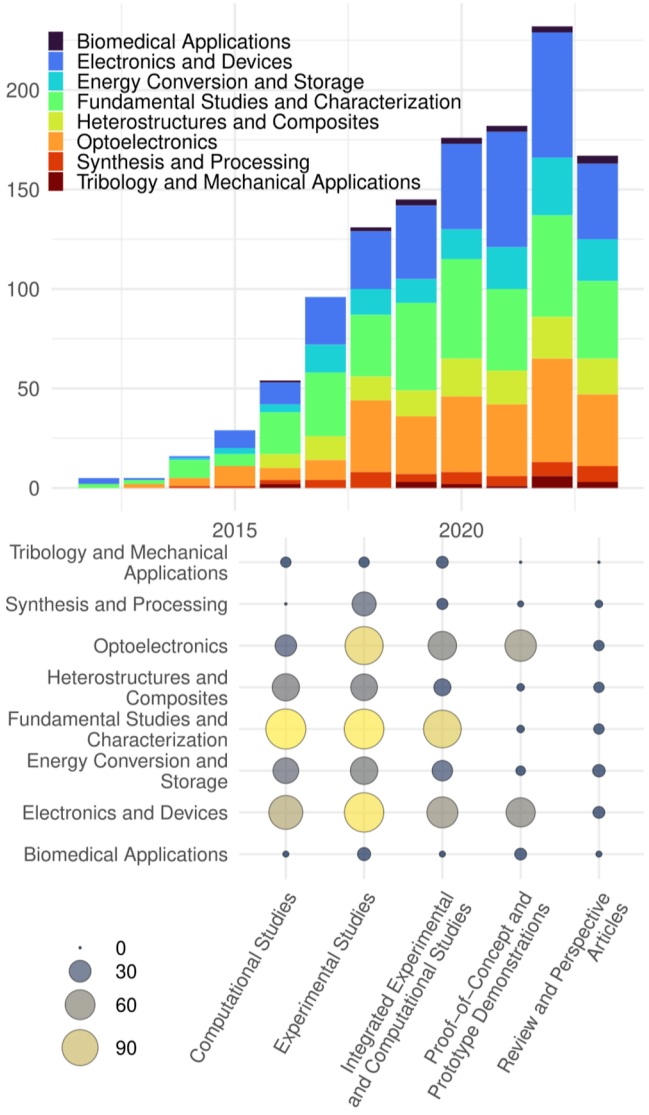


Figure 2. Histogram illustrating the number of articles published over the years, categorized by application area, showing the growth and shifts in research focus (top). Co-occurrence analysis of application areas and types of studies, highlighting the interdisciplinary nature of research in monochalcogenides (bottom).

Finally, Figure 3 illustrates the temporal distribution of monochalgonide occurrences. To optimize space, the figure includes only those compounds that feature in a minimum of ten articles. It is evident that there has been an upward trend in both the quantity and variety of compounds analyzed over the years, with a notable predominance of GaSe, and InSe. Additional insights are provided in the Supporting Information, where a topic modeling map delineates the main themes of the articles.

3.2. Selection of Two-Dimensional Structures. We systematically investigated a set of 27 monochalcogenide compounds with the general formula MQ, where M corresponds to elements from groups III (Al, Ga, In), IV (Si, Ge, Sn), and V (P, As, Sb) of the periodic table, while Q represents the chalcogen atoms S, Se, and Te. For each compound, we considered 13 candidates for the 2D structures identified in previous theoretical studies and material databases. These structures were categorized into three crystallographic families on the basis of their symmetry:

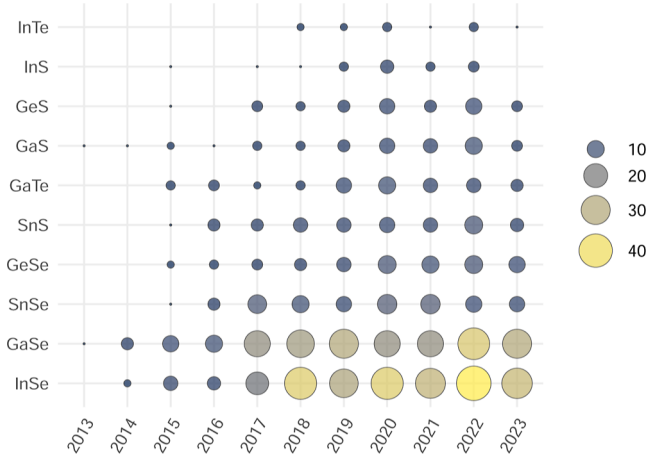


Figure 3. Temporal distribution of publications on 2D monochalcogenides (2013−2023). Only compounds with at least ten reported studies are included, revealing the growth of the field and the variety of monochalcogenides.

- **Hexagonal**: including $P\bar{3}m1\alpha$,[53] $P\bar{3}m1\beta$,[54] and $P\bar{6}m2$;[54]
- **Orthorhombic and Triclinic**: comprising a variety of lower-symmetry structures such as $P\bar{1}\beta$, $P1\alpha$, $P2_1/c$, $P4/nmm$, $Pbcm$, $Pmmn$, $Pmn2_1$, and $Pmna$;[55]
- **Triclinic**: specifically, the $Aem2$ and $C2/m$ structures.[55]

Figure 4 illustrates the initial geometries (both top and side views) of all these prototype structures, providing insight into their dimensionality and structure organization. The hexagonal structures exhibit high symmetry and layered arrangements, resembling the well-known trigonal prismatic or octahedral coordination seen in TMD compounds. In particular, the $P\bar{3}m1\alpha$ phase displays a compact, highly symmetric honeycomb pattern, while the β variant introduces a distortion along the out-of-plane direction.

The orthorhombic family exhibits a diverse set of structural features. For example, $Pmn2_1$ and $Pmna$ show strong in-plane anisotropy and puckered geometries, characteristic of materials such as GeS and SnSe, which are known for their anisotropic transport and ferroelectric properties. Structures like $P2_1/c$ and $P\bar{1}\beta$ possess more complex unit cells with multiple layers and nontrivial stacking, often associated with stereochemically active lone pairs from the elements of group IV and V. The presence of such lone pairs can lead to significant structural distortions, particularly in heavier compounds. In contrast, the triclinic structures $Aem2$ and $C2/m$ are characterized by low symmetry and distorted lattices. These geometries accommodate flexible coordination environments and may stabilize compounds under specific electronic configurations or pressure conditions.

3.3. Energetic Stability via Formation Enthalpy. To determine the energetic stability of the compounds, the formation enthalpy (ΔH) is calculated as follows

$$\Delta H = (E_{tot}^{MQ} - n_M\mu_M - n_Q\mu_Q)/N_{fu} \quad (1)$$

where, E_{tot} is the total energy of the MQ monolayer, while n_M and n_Q denote the atom count in the unit cell from the group III−V (M) and chalcogen (Q), respectively. The terms μ_M and μ_Q represent the chemical potential (total energy) related to the lowest energy bulk structures for the constituent elements: group III Al (fcc, $Fm\bar{3}m$), Ga (orthorhombic, $Cmce$), and In (tetragonal, $I4/mmm$); group IV elements (Si, Ge, Sn) with diamond-like cubic configurations; group V P (orthorhombic,

https://doi.org/10.1021/acs.jpcc.5c04625
J. Phys. Chem. C XXXX, XXX, XXX−XXX

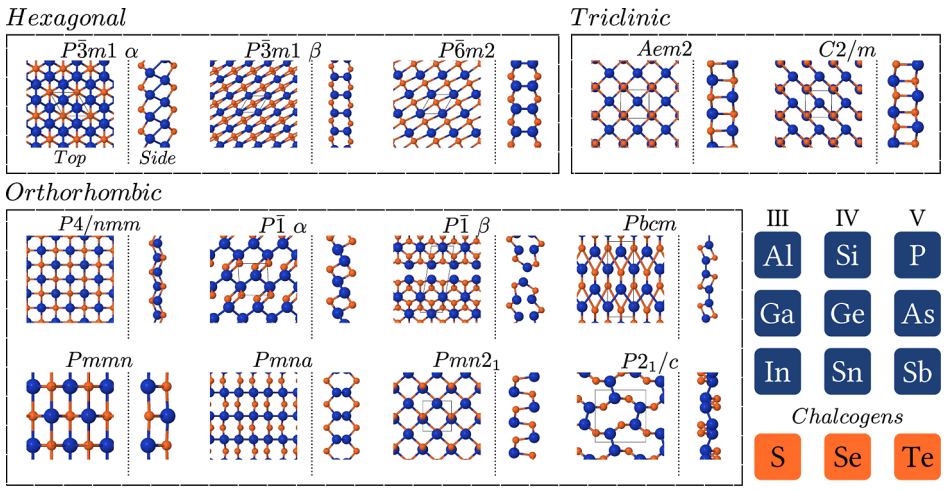


Figure 4. Ball and stick representations of 13 MQ structures with different space groups studied in this work.

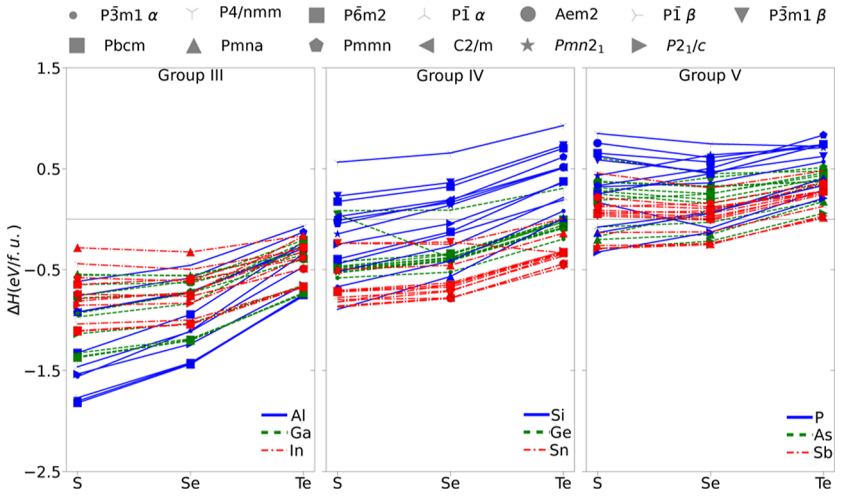


Figure 5. Formation enthalpy for (left panel) group III, (central panel) group IV, (right panel) group V MQ in the different space groups.

Cmce), As (orthorhombic, *Cmce*), and Sb (rhombohedral, $R\bar{3}m$); as well as chalcogens S (rhombohedral, $R\bar{3}$), Se (trigonal, $P3_121$), and Te (trigonal, $P3_121$). Moreover, N_{fu} denotes the number of formula units within the unit cell. According to the definition, negative values of ΔH signify exothermic formation and thermodynamically stable compounds relative to their bulk elemental states. Figure 5 provides an illustration of the calculated formation enthalpies for MQ compounds across different space groups, categorized by group III (left panel), group IV (central panel), and group V (right panel) M elements.

Several key trends emerge regarding the stability and structural preferences. For example, a significant trend observed across various material groups is the systematic increase in formation enthalpy, that is, it becomes less negative or more positive as the chalcogen element Q increases in atomic weight, moving sequentially from sulfur (S) to selenium (Se) to tellurium (Te). This indicates a general reduction in the thermodynamic stability of MS → MSe → MTe. This observed trend can be explained on the basis of the confluence of factors, including differences in electronegativity and variations in atomic size that directly affect the bond strength (equilibrium lattice constants). For example, sulfur exhibits the highest electronegativity among the chalcogens considered, i.e., S = 2.58, Se = 2.55, Te = 2.10 (Pauling scale).[56] The descending electronegativity from S to Te results in a diminished difference in electronegativity with the M element, which can lead to a weaker ionic character and an overall reduction in M−Q bond strength. In addition, the increasing atomic radius of the chalcogens in the group (S < Se < Te) affects both covalent and ionic bonds. The larger and more diffuse p-states of the heavier chalcogens (e.g., 5p-states of the Te versus 3p states of S) result in less effective orbital overlap for covalent bonding. With regard to ionic contributions, larger anion sizes augment interionic distances, thereby diminishing the magnitude of lattice energy. Both phenomena contribute to weaker overall bonds, thus resulting in less negative ΔH.

With respect to the impact of the element M within each classification, it is generally observed that among group III monochalcogenides, compounds incorporating the M element with the lowest atomic number (e.g., Al) tend to exhibit the highest stability (characterized by the most negative ΔH), while those with the highest atomic number (e.g., In) are the least stable. In contrast, in the cases of group IV and group V monochalcogenides, the opposite trend is manifest. Compounds containing the element M with the lowest atomic number (e.g., Si in group IV, P in group V) are typically less stable, while those with the highest atomic number (e.g., Sn in group IV, Sb in group V) demonstrate increased stability (more negative ΔH).

https://doi.org/10.1021/acs.jpcc.5c04625
J. Phys. Chem. C XXXX, XXX, XXX−XXX

This underscores a complex interaction between the electronic structure, atomic size, and bonding properties, which varies with the group. Furthermore, it is particularly notable for group III and V monochalcogenides that compounds based on S display a wider range of ΔH values across various polymorphs, in contrast to compounds based on Te, which exhibit a more restricted variation.

Group III monochalcogenides exhibit significant exothermic formation enthalpies. We observed a notable energetic competition among the $P\bar{1}\beta$, $P\bar{3}m1\beta$, and $P\bar{6}m2$ space groups. For example, the $P\bar{6}m2$ structure typically manifests the lowest energy for the majority of compounds within this group. However, exceptions such as AlTe and GaTe demonstrate a predilection for the $P\bar{1}\beta$ space group. It should be noted that all of these preferred structures remain exothermic ($\Delta H < 0$).

Most Group IV monochalcogenides exhibit exothermic formation, with the significant exception of SiTe in certain polymorphs considered. The preference for different space groups is contingent on the M element. In particular, compounds based on Si (e.g., SiS, SiSe) exhibit a predilection for the *Pmna* structure, whereas those based on Ge (e.g., GeS, GeSe) as well as SnTe, demonstrate a preference for the $P\bar{3}m1\alpha$ structure. In contrast, SnS and SnSe exhibit maximal stability in the $Pmn2_1$ structure. Almost all stable configurations of Group IV compounds show exothermic characteristics.

In the context of Group V monochalcogenides, those compounds composed of lighter chalcogens (S, Se) are predominantly exothermic. The $P2_1/c$ structure emerges as the preferred configuration for the majority of compounds. A notable deviation is SbTe, which demonstrates a preference for the $P\bar{1}\beta$ structure. A critical observation for this group is that all compounds based on Te (i.e., MTe in conditions where M = P, As, Sb) are projected to be endothermic ($\Delta H > 0$), thereby indicating a thermodynamic instability when compared to decomposition into elemental bulk phases under standard conditions.

3.4. Dynamical Stability via Phonons. Phonon band structures for the configurations of each MQ compound, characterized by the lowest formation enthalpy, were computed to assess their dynamical stability. Structures such as $P\bar{3}m1\alpha$, $P\bar{6}m2$, and $Pmn2_1$ contain 4 atoms within the unit cell, resulting in 12 phonon branches (3 acoustic and 9 optical); $P2_1/c$, *Pmna*, and a variant $Pmn2_1$ incorporate 8 atoms within the unit cell, giving rise to 24 branches (3 acoustic and 21 optical); meanwhile, $P\bar{1}\beta$ comprises 12 atoms per unit cell, producing 36 phonon branches (3 acoustic and 33 optical). The outcomes for MSe are illustrated in Figure 6, whereas the phonon-band structures for the remaining systems are included in the Supporting Information.

According to the widely recognized stability criterion that allows imaginary acoustic modes up to 10% of the maximal phonon frequency,[53,57] most of the compounds evaluated demonstrate dynamical stability. However, an exception is observed in the *Pmna* phase of SiTe. This compound exhibits significant imaginary frequencies within the acoustic region, particularly along the Γ−X and X−S directions, implying robust dynamical instabilities. Such characteristics typically imply the presence of soft-phonon modes that may induce structural distortions or phase transitions under conditions of finite temperature or stress. In light of this instability, the second most stable polymorph of SiTe, corresponding to the $P\bar{3}m1\alpha$ phase, was analyzed. This structure does not exhibit imaginary frequencies throughout the Brillouin zone. The phonon

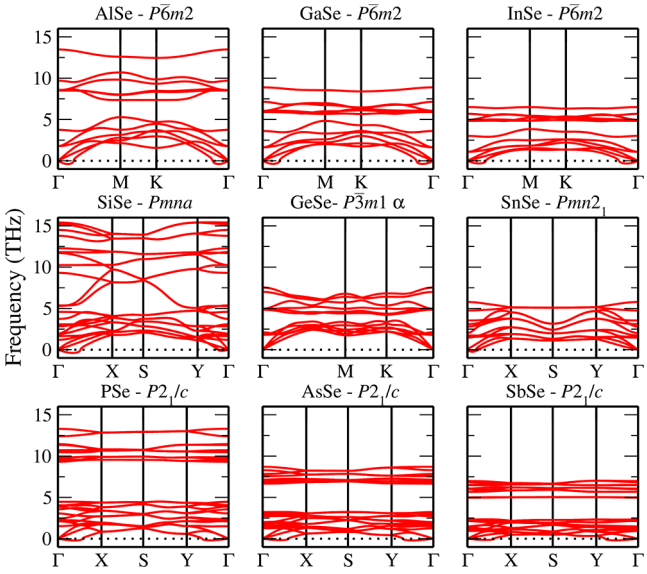


Figure 6. Stable lowest energy phonon band structures for the MSe compounds.

modes are well behaved, with clearly defined acoustic and optical branches, attesting to their dynamical stability. In particular, the optical branches are comparatively flat, suggesting localized vibrational modes that may impact the thermal transport properties. These phonon-band structures are accessible in the Supporting Information.

An evident pattern emerges across the various main groups of the periodic table: the maximum phonon frequency significantly declines as the atomic number (and consequently the atomic mass) of the M atom increases within a specific group. This is apparent when comparing AlSe, GaSe, and InSe (Group 13), where the maximum frequency diminishes from approximately 15 THz with 10 THz. Similarly, in Group 14 (SiSe, GeSe, SnSe) and Group 15 (PSe, AsSe, SbSe), the maximum frequencies consistently decrease with increasing atomic mass, despite differences in their respective space groups. This decrease in frequency is mainly attributed to the inverse relationship between vibrational frequency and atomic mass, as heavier atoms resonate at lower frequencies. Moreover, as the maximum frequency declines, the optical phonon branches tend to exhibit increased flatness, indicating an improved localization of the vibrational modes. Concerning dynamical stability, it is observed that for most systems, there are subtle imaginary frequencies in proximity to the Γ-point, which are generally of very small magnitude. In particular, GeSe emerges as the only compound devoid of imaginary frequencies throughout its Brillouin zone, corroborating its exceptional dynamical stability. This suggests that while minor instabilities might be present in other compounds near the center of the zone, GeSe maintains a particularly robust and stable crystal structure.

3.5. Characterization of the Equilibrium Lattice Parameters. Table 1 summarizes the optimized equilibrium lattice parameters associated with the most stable phases for each of the MQ monochalcogenides. These parameters are compared with values documented in the literature, encompassing theoretical predictions as well as experimental findings, when available. Values in bold represent experimental data reported in the literature.[20] In general, the computational results derived from our study demonstrate a strong concordance with previous reports, with discrepancies generally remaining below 2%,

https://doi.org/10.1021/acs.jpcc.5c04625
J. Phys. Chem. C XXXX, XXX, XXX−XXX

Table 1. Lattice Parameters for the Stable Lowest Energy MQ[a]

MQ	phase	a_0 (Å)	b_0 (Å)	ΔH (eV/f.u.)	$\overline{q}_B^M$ (e)	E_g^{PBE} (eV)	$E_g^{PBE+SOC}$ (eV)	$E_g^{HSE06+SOC}$ (eV)
AlS	$P\bar{6}m2$	3.58 (3.57[27])		−1.82	1.50	2.10	2.09	2.86
AlSe	$P\bar{6}m2$	3.77 (3.78[27])		−1.44	1.40	2.01	2.00	2.70
AlTe	$P\bar{1}\beta$	4.12	12.09	−0.76	0.52	1.90	1.84	2.49
GaS	$P\bar{6}m2$	3.64 (3.64[27])		−1.37	0.76	2.33	2.33	3.23
GaSe	$P\bar{6}m2$	3.82 (3.82[27])		−1.20	0.62	1.82	1.80	2.63
GaTe	$P\bar{1}\beta$	4.14	12.13	−0.74	0.43	1.46	1.38	1.99
InS	$P\bar{6}m2$	3.92 (3.94[27])		−1.11	0.82	1.66	1.66	2.50
InSe	$P\bar{6}m2$	4.08 (**4.05**[20])		−1.04	0.70	1.39	1.37	2.10
InTe	$P\bar{6}m2$	4.38 (4.40[27])		−0.67	0.52	1.35	1.25	1.88
SiS	*Pmna*	3.99	6.65	−0.89	1.21	0.45	0.45	1.18
SiSe	*Pmna*	4.00	7.03	−0.57	0.95	0.63	0.62	1.35
SiTe	$P\bar{3}m1\alpha$	3.89 (3.88[53])		0.08	0.36	0.00	0.14	0.26
GeS	$P\bar{3}m1\alpha$	3.65 (3.66[53])		−0.58	0.69	0.73	0.72	1.14
GeSe	$P\bar{3}m1\alpha$	3.81 (3.80[53])		−0.38	0.53	0.54	0.53	0.94
GeTe	$P\bar{3}m1\alpha$	4.06 (4.06[53])		−0.03	0.30	0.58	0.54	0.91
SnS	$Pmn2_1$	4.08 (4.07[58])	4.30 (4.24[58])	−0.87	0.98	1.53	1.45	2.04
SnSe	$Pmn2_1$	4.28 (4.30[58])	4.44 (4.36[58])	−0.78	0.84	0.99	0.93	1.38
SnTe	$P\bar{3}m1\alpha$	4.33 (4.33[53])		−0.48	0.48	0.58	0.50	0.81
PS	$P2_1/c$	6.16	6.71	−0.32	0.52	1.89	1.89	2.85
PSe	$P2_1/c$	6.54	6.82	−0.13	0.23	1.66	1.65	2.53
PTe	$P2_1/c$	7.09	7.01	0.21	−0.10	1.69	1.60	2.35
AsS	$P2_1/c$	6.49	7.12	−0.29	0.55	1.22	1.21	2.03
AsSe	$P2_1/c$	6.84	7.14	−0.22	0.35	0.97	0.96	1.70
AsTe	$P2_1/c$	7.37	7.29	0.06	0.08	0.96	0.91	1.55
SbS	$P2_1/c$	6.99	7.66	−0.23	0.73	1.06	0.99	1.65
SbSe	$P2_1/c$	7.35	7.67	−0.25	0.57	0.75	0.70	1.34
SbTe	$P\bar{1}\beta$	6.00	10.86	0.01	0.35	0.17	0.13	0.21

[a]The values in parentheses are reported in the literature, with the values reported in bold being experimental results. Formation enthalpy (ΔH). Bader charge corresponding to the atoms of group III, IV, and V ($\overline{q}_B^M$). Band gap considering PBE functional (E_g^{PBE}), PBE + SOC functional ($E_g^{PBE+SOC}$), HSE06 + SOC functional ($E_g^{HSE06+SOC}$).

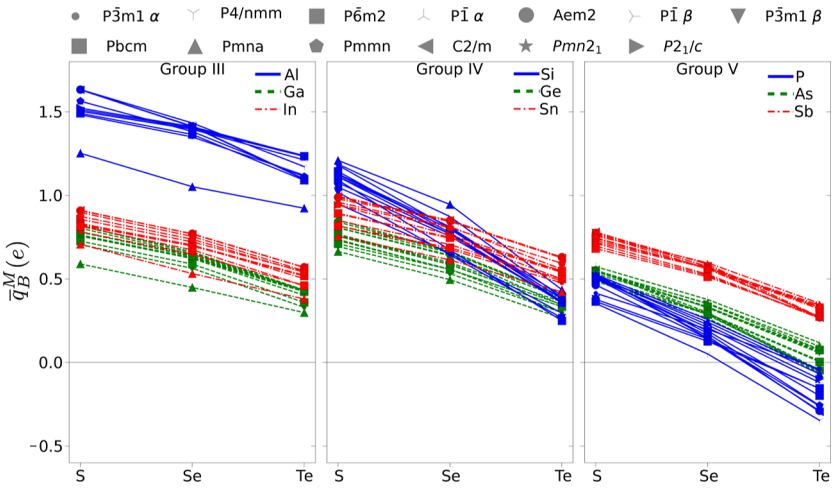


Figure 7. Average Bader charge ($\overline{q}_B^M$). Groups III, IV, and V compounds are shown on the left, central, and right panels, respectively.

particularly for compounds in the group III and group IV categories.

An evident and methodical trend is discernible in the progression of the lattice parameters as a function of the atomic species. For a constant chalcogen Q, the lattice constant a_0 shows an increase concomitant with the atomic number of the group III, IV or V element (i.e., Al → Ga → In, Si → Ge → Sn and P → As → Sb). This trend is in alignment with the increment in atomic radii down the groups of the periodic table, resulting in larger interatomic distances and consequently extended lattice vectors in the optimized structures. For example, within the group-III chalcogenides with S, a_0 extends from 3.58 Å for AlS to 3.92 Å for InS, coherently following the trend in atomic radii (Al: 1.18 Å, Ga: 1.26 Å, In: 1.44 Å).

Similarly, for a given metal M, the lattice parameter exhibits a systematic increase from S to Se to Te. This progression is attributed to the enlargement of the atomic size and the decreasing electronegativity of chalcogens: S (1.05 Å), Se (1.20

https://doi.org/10.1021/acs.jpcc.5c04625
J. Phys. Chem. C XXXX, XXX, XXX−XXX

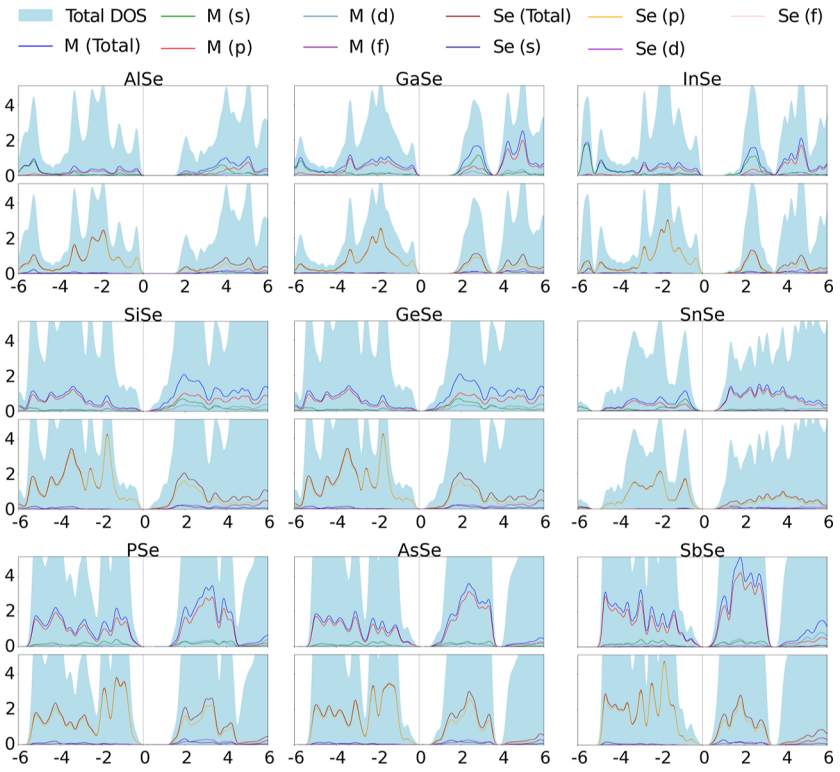


Figure 8. Density of states calculated for the stable lowest energy MSe compounds. On the top panel of each subplot we have the M contribution, while on the bottom panel we have the Se contribution. The other systems are accessible in the Supporting Information.

Å) and Te (1.38 Å). With the inclusion of heavier chalcogens, the bond lengths extend, resulting in the expansion of the unit cell. For instance, the lattice constant of SnQ in the $Pmn2_1$ phase increases from 4.08 Å (SnS) to 4.28 Å (SnSe) to 4.33 Å (SnTe).

The structural phase also plays a crucial role in the values of the lattice parameters. For example, high-symmetry phases such as $P\bar{6}m2$ and $P\bar{3}m1$ tend to produce more isotropic and compact in-plane lattice constants, while lower-symmetry orthorhombic and triclinic phases such as $Pmn2_1$, $Pmna$, or $P2_1/c$ exhibit anisotropy in a_0 and b_0 due to puckering or layer distortions. This is particularly evident for group V chalcogenides, which predominantly adopt the phase $P2_1/c$. These compounds show significantly larger and anisotropic lattice constants, as seen in SbSe with a_0 = 7.35 Å and b_0 = 7.67 Å. These features can be attributed to the presence of stereochemically active lone pairs in heavier group-V elements, which promote distorted bonding environments and structural buckling. For group V, the literature review did not yield theoretical or experimental lattice parameters. However, the alignment of our results with existing data for group III and IV compounds, together with the observed systematic trends across different groups and the chalcogen series, substantiates the reliability and physical significance of our predictions for these systems.

3.6. Charge Transfer Analysis via Bader Charges. Figure 7 shows the Bader charge averaged for each species in monochalcogenides where M belongs to Groups III, IV, or V ($\overline{q}_B^M$), evaluated for their lowest-energy configurations. The Bader charge quantitatively represents the extent of charge transfer that occurs from the metal (M) to the chalcogen (Q), thus providing an indirect measure of the ionicity associated with the M−Q bond. A clear and systematic trend is observed across all groups: for a specific M, $\overline{q}_B^M$ decreases as the chalcogen changes from S to Se to Te. This behavior is consistent with the decreasing electronegativity of chalcogen atoms (Pauling scale: S = 2.58, Se = 2.55, Te = 2.10),[56] which weakens the driving force for charge transfer and results in a more covalent character for heavier chalcogens. For example, in group III, $\overline{q}_B^M$ drops from 1.46 e in AlS to 0.73 e in InS, and continues to decrease for heavier chalcogens.

The evolution of $\overline{q}_B^M$ between each group is also notable. In group III, the Bader charge decreases from Al to Ga, reflecting the reduction in the effective nuclear charge and electronegativity (Al = 1.61, Ga = 1.81), which weakens the donor ability of the metal. Interestingly, the charge increases slightly from Ga to In, which could be attributed to the larger orbital extension and polarizability of indium, allowing greater interaction with the chalcogen p-states. In group IV, a nonmonotonic trend is observed. The Bader charge generally decreases from Si to Ge, and remains relatively stable or slightly increases toward Sn, especially for heavier chalcogens. This reflects a subtle balance between increasing the atomic radius (leading to a more diffuse electron density and less charge localization) and decreasing the electronegativity (Si = 1.90, Ge = 2.01, Sn = 1.96), which affects the charge distribution in these layered compounds.

The elements of Group V exhibit a different behavior: $\overline{q}_B^M$ increases consistently from P to As to Sb for all chalcogenides. This trend is counterintuitive considering that electronegativity decreases in this direction (P = 2.19, As = 2.18, Sb = 2.05). However, the increase in Bader charge can be attributed to structural distortions in the $P2_1/c$ phase and the presence of stereochemically active lone pairs, particularly in heavier pnictogens such as Sb. These lone pairs influence electron localization and can promote enhanced charge asymmetry,

https://doi.org/10.1021/acs.jpcc.5c04625
J. Phys. Chem. C XXXX, XXX, XXX−XXX

particularly in distorted or low-symmetry structures. Another important point is that the magnitude of $\bar{q}_B^M$ is strongly dependent on the crystal structure. High-symmetry phases such as $P\bar{6}m2$ and $P\bar{3}m1$ typically show more pronounced charge transfer due to higher orbital overlap and uniform coordination environments. In contrast, low-symmetry phases like $P2_1/c$ and $Pmn2_1$ can suppress charge transfer due to bond anisotropy and local distortions, as observed for some group V and group IV chalcogenides.

Our calculated values are in excellent agreement with available Bader charge analyzes from the literature. For example, in the $P\bar{3}m1\alpha$ phase, the calculated values for GeS and GeSe are 0.69 e and 0.53 e, respectively, closely matching those reported by Querne et al.[59] Likewise, for the $P\bar{6}m2$ phase, our values of 1.46 e for AlS, 0.75 e for GaS, and 0.73 e for InS are consistent with the data of Demirci et al.[27] These results further validate the accuracy of our computational approach and reinforce the chemical trends discussed above.

3.7. Orbital-Resolved Density of States. Figure 8 illustrates the total density of states (DOS) for the lowest energy configurations of the MSe compounds, alongside the local DOS contributions from the M species, representative of group III, IV, or V elements, in addition to Selenium (Se). The top panels of each subplot delineate the contribution of the M atom, while the bottom panels reveal the contribution of the Se atom. From our analyses, we have identified consistent trends across all MSe compounds concerning the composition of their electronic states, which will be discussed below.

The maximum valence band (VBM) is primarily composed of the p-states associated with the chalcogen atom (Se). Specifically, the Se p-states dominate the DOS near the VBM, underscoring their fundamental role in the valence band bonding framework. The Se s-states, having lower energy, contribute to the more profound regions of the valence band, as expected. Regarding the M atom, its s-states are responsible for deeper valence band energies, whereas its p-states are located nearer to the VBM, and therefore participate in hybridization with the Se p-states to establish covalent bonding. This orbital configuration is emblematic of materials in which the chalcogen p-states delineate the peak of the valence band.

In contrast, the conduction band minimum (CBM) generally manifests contributions from both M and Se atoms. With regard to the M atom, the lowest energy states in the conduction band predominantly originate from its s orbitals, whereas its p-states increasingly contribute at elevated energies within the conduction band. Similarly, for the Se atom, its p-states also contribute to the CBM, and notably, the contribution from Se d becomes pronounced at higher energies within the conduction band. This suggests the involvement of higher-energy unoccupied states in the conduction band.

An essential trend identified across various groups pertains to the development of the electronic band gap and the nature of orbital contributions. As one progresses through a group in the periodic table (for instance, from Al to In in Group 13, Si to Sn in Group 14, or P to Sb in Group 15), there is an increase in the atomic size of the M atom, inducing a general decrease in the band gap magnitude. This phenomenon is directly attributable to the augmented atomic size and diminished effective orbital overlap, which decreases the energy splitting between the bonding and antibonding states. In compounds such as SnSe and SbSe, the band gap becomes significantly narrow, indicative of an increase in metallic characteristics down the group. Moreover, for the heavier elements in Group 14 (Ge, Sn) and Group 15 (As, Sb), their respective d-states become energetically more accessible. This results in more pronounced contributions from M d-states in the DOS, particularly within the conduction band, which affect the electronic structure and potentially the transport properties.

3.8. Electronic Band Structure: Influence of Symmetry and Spin−Orbit Coupling. Figure 9 depicts the electronic band structures for each of the lowest-energy stable MSe compounds. We considered the PBE + SOC+χ electronic band structure calculated through the scissors operator (χ) of the PBE + SOC band gap, where $\chi = E_g^{HSE06+SOC} - E_g^{PBE+SOC}$, this approach has already been validated in previous studies reported in the literature.[60] Group III structures AlSe, GaSe, and InSe are $P\bar{6}m2$ structures and have indirect band gaps with the VBM along the **K**−**Γ** direction, while the CBM is located at the **K** point for the Al compounds and at **Γ** for the other compounds for this space group. Group IV structure SiSe is part of the *Pmna* space group and has direct band gap at **Γ** point, as for other *Pmna* structures, while GeSe and others $P\bar{3}m1\alpha$ structures have indirect band gaps as for GeS, with a VBM displacement up to 2 eV as already reported in the literature.[53] Furthermore, SnSe compound belongs to $Pmn2_1$ space group and also has indirect band gaps with the CBM and the VBM alongside **Γ**−**X** and **Y**−**Γ** respectively, as previously reported in the literature.[53] Finally, group V $P\bar{1}\beta$ PSe, AsSe, and SbSe compounds have direct bang gaps at the **Γ** point.

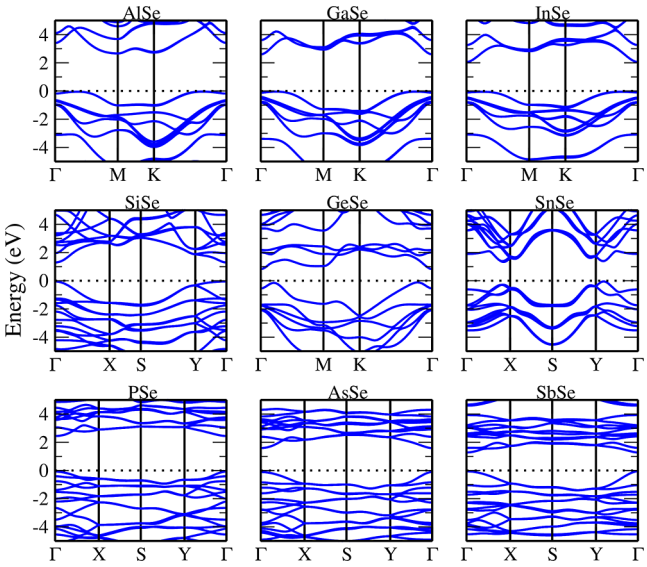


Figure 9. PBE + SOC + χ electronic bandstructures for the stable lowest energy MSe compounds. The other systems are accessible in the Supporting Information.

The structures $P\bar{1}\beta$ have indirect, but near direct, band gaps for all compounds except GaTe, with the CBM and VBM located at the **Γ** or **Y**-point for AlTe and GaTe, and near **M** for SbTe. Furthermore, we investigated the effects of SOC and observed reductions in band gaps up to 0.10 eV due to the presence of heavier chalcogens, especially Sb and Te atoms. However, SiTe was initially metallic and when submitted to the SOC effects it opened the band gap to 0.14 eV when only considering the PBE + SOC functional and 0.26 eV when considering the HSE06 + SOC functional, the behavior for this

https://doi.org/10.1021/acs.jpcc.5c04625
J. Phys. Chem. C XXXX, XXX, XXX−XXX

compound is already known.[61] Table 1 presents the band gap values for various functionals.

3.9. Band Alignment and Heterojunction Classification of Monochalcogenides via Anderson's Rule. The energy values for the VBM and CBM were also calculated relative to the vacuum level of the materials, as depicted in Figure 10, where the band gap values are also presented. The

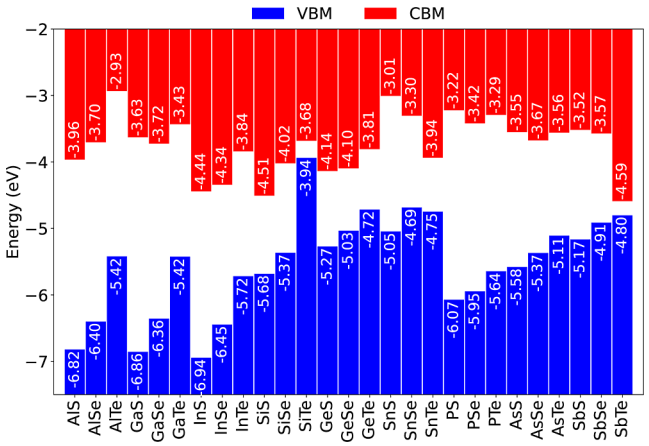


Figure 10. Band offset for the stable lowest energy monochalcogenides calculated using PBE + SOC + χ exchange−correlation functional.

band gap diminishes when M is maintained and Q is altered (S → Se → Te), with the exception of Si-based compounds, where the band gap initially increases from S → Se and subsequently decreases from Se → Te. The most significant reductions in band gap percentage occur when various space groups are present for the same M, specifically Sn (60%), Si (81%), and Sb (88%). Conversely, the Al-based materials exhibit the most Q resistance, showing only a 13% reduction in the band gap.

The distinct configurations of the VBM and the CBM facilitate the use of Anderson's rule[62] to discern various types of heterojunction between two compounds. These heterojunctions can be categorized into three distinct types: Type I heterojunctions (straddling gaps), in which the CBM of compound *A* is below and the VBM above that of compound B. These are pertinent for optical devices due to the high propensity for charge carrier recombination.[63,64] Type II heterojunctions (staggered gap) arise when the CBM of compound A is higher than its VBM and lower than the CBM of compound B, while the VBM of compound A is lower than that of compound B.[65] The condition where the VBM of compound A surpasses the CBM of compound B typifies a Type III (broken gap) heterojunction, exploitable in tunnel field effect transistors due to charge tunneling from the VBM to the CBM.[64] Through the application of Anderson's rule to the examined systems, we identified 351 heterojunctions, classified as 166 type I, 177 Type II, and 8 type III as seen in Figure 11.

3.10. Anisotropic Optical Absorption: Directional Dependence and Chalcogen-Induced Redshift. Figure 12 shows the absorption coefficient $\alpha(\hbar\omega)$, derived from the dielectric tensor, for the lowest energy compounds along the orientations *xx* (solid line) and *yy* (dashed line). These calculations incorporate the application of a scissors operator correction derived from the disparity between the HSE06 + SOC and PBE + SOC band gaps. A significant redshift in the onset of absorption is discerned as one progresses from lighter to heavier chalcogens (S → Se → Te), as demonstrated by the

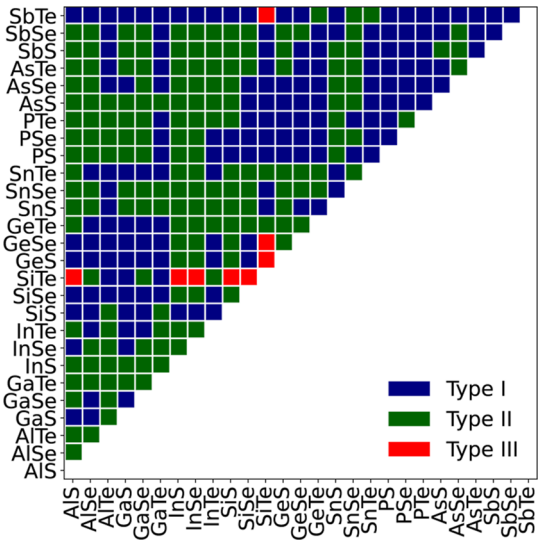


Figure 11. Heterojunctions classification for the combination of groups III−V monochalcogenides monolayers obtained through the application of the Anderson's rule in the HSE06 + SOC band offset.

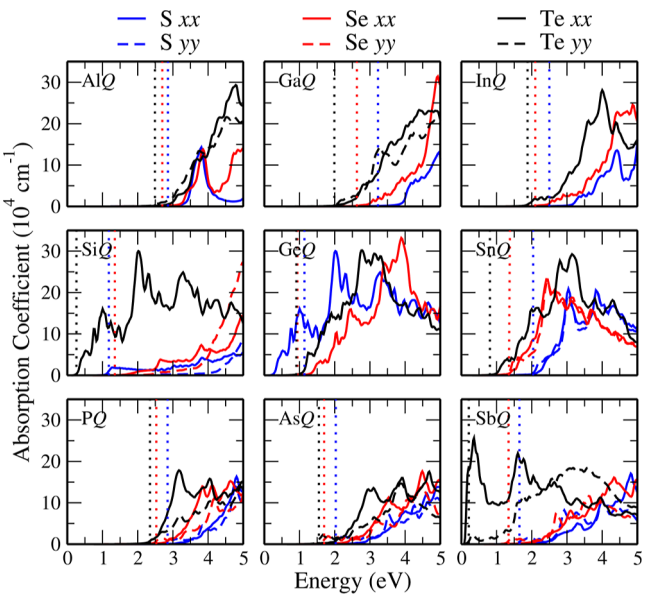


Figure 12. Absorption coefficient for the stable lowest energy compounds. The solid line represents the absorption coefficient for the *xx* direction and the dashed line for the *yy* direction. The hexagonal structures have the same values for both directions, therefore only one is shown. The dotted lines indicate the fundamental band gap of the compounds.

transition of the absorption edge from InS to InTe. This observable trend persists across all the studied compounds and shows a correlation with the reduction in the band gaps as the atomic number of the chalcogen increases.

The absorption spectrum of InS is predominantly located within the ultraviolet region, while InTe demonstrates extended absorption into the visible spectrum, making it more appropriate for applications in optoelectronics. In the case of orthorhombic compounds, exemplified by PTe, pronounced anisotropy is observed: the *xx* component presents a more pronounced absorption peak in the visible range as opposed to the *yy* direction. Such a directional dependence is indicative of low-symmetry crystal structures, contrary to the behavior in

https://doi.org/10.1021/acs.jpcc.5c04625
J. Phys. Chem. C XXXX, XXX, XXX−XXX

hexagonal structures, where the *xx* and *yy* responses exhibit near-equivalence due to the presence of in-plane isotropy.

Beyond these examples, general trends emerge across the families of compounds. For example, SiS and SiSe exhibit relatively low overall absorption coefficients, with *xx* absorption dominating in the visible range, but *yy* becoming more significant at higher energies. Group V materials, particularly those with P, As, and Sb, present the highest degrees of anisotropy in the optical response, with sharp directional variation in their spectra. In contrast, materials such as AlTe, SnS, and SnSe demonstrate a nearly isotropic behavior with similar absorption in both directions. In particular, low-band gap materials like SiTe and SbTe reach high absorption values even at energies around 0.50 eV, making them promising candidates for near-infrared applications.

3.11. Explainable Machine Learning Results. In this study, we used Random Forest models to predict energies based on descriptors. To assess the performance of these models, we utilized both resubstitution and stratified cross-validation techniques. The resubstitution method evaluates the model's performance using the same training data, while cross-validation involves training and testing the model on multiple random subsets of the training data. As described earlier, we used stratified cross-validation, where train and test subsets are drawn so that the proportion of Unitary Chemical Composition is preserved. This approach provides a more robust estimate of the model's ability to generalize to unseen data. Furthermore, for the cross-validation method, we adjust the parameter number of estimators via internal cross-validation.

Table 2 shows the coefficient of determination (R^2) for the Random Forest models, calculated using resubstitution and cross-validation techniques. R^2 ranges from 0 to 1, with higher values indicating that the model explains a greater proportion of the variance in the dependent variable. The resubstitution R^2 score and the cross-validation R^2 score provide insights into the model's ability to fit the data and generalize to unseen data, respectively. As can be seen in the table, the resubstitions R^2 are higher than 0.95 for all four energies.

Table 2. Random Forest Model Performance: Resubstitution and Cross-Validation R^2 Scores

energy	resubstition	cross-validation
$E_{\mathrm{tot}}^{\mathrm{f.u.}}$	0.9948	0.9547
$\Delta E_{\mathrm{tot}}^{\mathrm{f.u.}}$	0.9704	0.6776
ΔH	0.9924	0.9240
$E_{\mathrm{coh.}}$	0.9948	0.9585

The cross-validation R^2 is higher than 0.9 for all but relative energy, with a $R^2 \approx 0.68$. One potential reason is that, since the relative energy is determined by subtracting one energy value from another, the instance with the lowest value (zero) might be absent from the training set. This situation could make it more challenging for the model to accurately predict relative energies. Figure 13 displays a scatter plot of real versus predicted values, showcasing the model constructed using cross-validation. Overall, the models effectively capture the general trend of energy values, indicating that the insights derived from the SHAP values (discussed next) are relevant and valuable.

For the model calculated using resubstitution, we utilized SHAP values for eXplainable Artificial Intelligence (XAI). SHAP provides insights into the importance and contribution of each descriptor in predicting energies, enhancing the model's transparency and interpretability. Figure 14 shows a beeswarm plot of SHAP values. This plot is a visualization used to display the distribution of SHAP values for each feature for the most important features of the model for a data set. For each feature row, the SHAP value is represented by a single dot on each feature row. Color is used to display the original value of a feature using a single scale of low to high values. The vertical line at $x = 0$ represents the median SHAP value for a given feature in all instances in the data set. The points on each side of the vertical line suggest that the corresponding feature tends to have a positive or negative impact on the predicted energy. For example, for the total energy ($E_{\mathrm{tot}}^{\mathrm{f.u.}}$), shown in the upper left panel of Figure 14, the most important feature is the interstitial charge (q_{I}), with low values contributing to predict a lower energy value and medium to large values contributing positively. Similarly, for the formation of the enthalpy (ΔH), shown in the lower left panel of Figure 14, the most important characteristic is the group in the periodic table of the element M, with elements of the group III involved in the reduction of the prediction of ΔH, group V increased in the prediction of ΔH and group IV between.

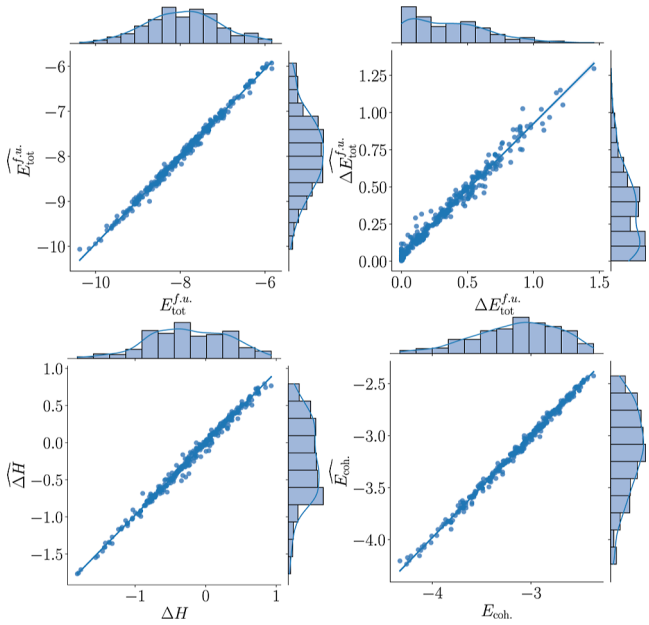


Figure 13. Real data points (x-axis) and the predicted data points (y-axis). The marginal histograms in the top and right panels depict the distribution of the real and predicted data.

4. INSIGHTS ON ELECTRONEGATIVITY AND ENERGETIC STABILITY FROM MACHINE LEARNING

The character of the atomic bonds determines relevant features for 2D materials, although it should be diverse and difficult to characterize.[66] In this spirit, Figure 15 depicts the correlation between the ionic character[67] (IC) of the M−Q atomic bond calculated as $\mathrm{IC} = \{1 - \exp[-(0.25)(X_{\mathrm{M}} - X_{\mathrm{Q}})^2]\} \times 100$ (X_{M} and X_{Q} are the electronegativities for the elements), and the average Bader charge transfer from M to Q for the most stable investigated materials It is observed that the IC of the compounds increases up to 21% while the $\overline{q}_{\mathrm{B}}^{\mathrm{M}}$ increases up to 1.5 e. This correlation is not completely linear as it has some deviations. Despite achieving high $\overline{q}_{\mathrm{B}}^{\mathrm{M}}$ values, the bond character in monochalcogenides is still the most predominant covalent

https://doi.org/10.1021/acs.jpcc.5c04625
J. Phys. Chem. C XXXX, XXX, XXX−XXX

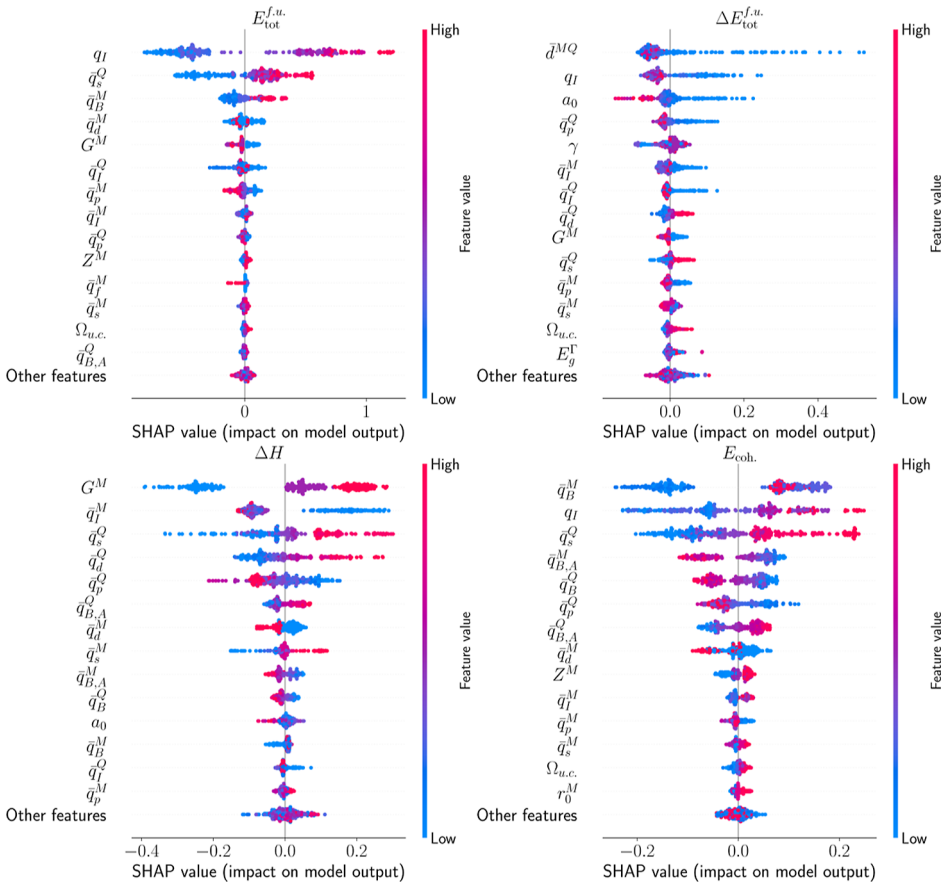


Figure 14. SHAP values for descriptors and energetic and electronic properties—positive values (to the right of the vertical axis) indicate an above-average contribution, while negative values (to the left of the vertical axis) indicate a below-average contribution to the target energy. The color of the points corresponds to the value of the descriptor, with darker colors indicating higher values.

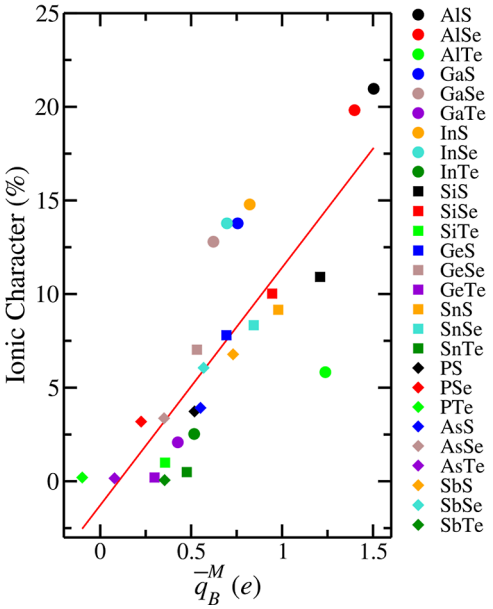


Figure 15. Ionic character versus Bader charge for all stable lowest energy compounds. The trend is represented by the red line.

(since only 21% is ionic). We identified that Te-based compounds have the lowest ionic character among all compounds and while varying M through groups III → IV → V makes the ionic character decrease.

The Random Forest model was able to capture the trend to predict the different energies calculated in this study. However, we noticed a spread of the values when comparing the real values and the predicted values, especially for the relative energy. This spread raises concerns on the ability to use machine learning models to predict properties values in desirable accuracy, making the use of quantum simulations essential to obtain accurate properties of these systems.

Furthermore, the SHAP analysis for the energetic descriptors revealed that the most influential features in predicting the energy values are related to the electronic charge distribution. Notably, the interstitial charge consistently ranks among the top two most impactful features across all energy-related targets. Based on this observation, we recommend that future investigations of 2D materials systematically incorporate detailed analyses of electronic charge distribution, as these descriptors provide essential insights into the energetic stability and bonding characteristics of low-dimensional systems.

5. CONCLUSION

Our computational investigation started with a systematic bibliographic analysis using NLP techniques, which extracted 1257 articles related to 2D MQ compounds from an initial database that exceeded 5400 entries. This analysis identified a pronounced and escalating research interest in monochalcogenides, particularly in domains such as electronics and devices, energy conversion, and storage, as well as fundamental studies. Furthermore, NLP techniques were used for categorization by

https://doi.org/10.1021/acs.jpcc.5c04625
J. Phys. Chem. C XXXX, XXX, XXX−XXX

application domain and study type, demonstrating that experimental studies are predominant in optoelectronics and electronics, while computational methodologies are more prevalent in fundamental characterizations. The temporal analysis of the references indicated a diversification and an increase in the number of compounds under investigation, with materials such as GaSe, InSe and SnS showing marked prominence.

For the selected 27 MQ compounds, DFT calculations were performed across 13 distinct 2D crystal structures, which generated a large database of physical−chemical properties. This extensive computational effort allowed for a detailed exploration of their fundamental characteristics. For example, from the formation enthalpy we identified clear trends: stability generally decreases as the chalcogen becomes heavier. For group III M elements, stability decreases down the group (e.g., Al > Ga > In), whereas for groups IV and V, it increases down the group (e.g., Si < Ge < Sn). Most of the group V tellurides were determined to be endothermic, indicating potential instability. The consideration of dynamical stability, assessed by phonon calculations, corroborated that most low energy configurations exhibit stability, with GeSe in the $P\bar{3}m1\alpha$ phase demonstrating exceptional dynamical robustness. In particular, the *Pmna* phase of SiTe was an exception, exhibiting significant imaginary frequencies, which led to further analysis of its more stable $P\bar{3}m1\alpha$ polymorph, highlighting the importance of thorough stability assessments.

Our evaluation of equilibrium lattice parameters revealed strong concordance with previously reported values, demonstrating predictable expansions concomitant with increasing atomic radii of both M and Q elements. Notable anisotropies were observed in lower-symmetry phases, especially evident in group V compounds within the $P2_1/c$ structure, indicating complex structural behavior. Bader charge analysis, which measures charge transfer from M to Q, indicated a general decrease in ionicity with heavier chalcogens, attributed to decreased differences in electronegativity. Variations in Bader charge on the M atom were observed across different groups: a decrease from Al to Ga, followed by an increase to In for group III; a nonlinear trend for group IV; and a consistent increase within group V elements. This pattern was attributed to structural distortions and the effects of single pairs of structurally active individuals, as exemplified in phases such as $P2_1/c$, emphasizing the intricate relationship between structure and electronic properties.

For MSe-based compounds, the VBM predominantly consists of Se p-states, while the CBM contains contributions from both M and Se s- and p-states. Descending a group for the M element typically results in a decrease in the band gap, a valuable trend for material design. Electronic band structures exhibit substantial variation according to crystal symmetry; for example, structures $P\bar{6}m2$ (e.g., AlS) typically possess indirect band gaps, while structures *Pmna* (e.g., SiS) and $P2_1/c$ (e.g., AsS) manifest direct band gaps at the Γ-point. SOC generally leads to a reduction in band gaps, particularly up to 0.10 eV for InTe, although it engenders a gap opening in the initially metallic SiTe, demonstrating its crucial role in determining electronic properties. Analyzes of optical absorption reveal a redshift with heavier chalcogens, correlated with the reduction of the band gap, and significant anisotropy in lower symmetry structures, such as orthorhombic PTe, which is important for optoelectronic applications.

The calculation of band alignments employing HSE06 + SOC levels relative to the vacuum level (electrostatic potential) provided the opportunity for the classification of prospective heterojunctions according to Anderson's rule. Among the identified pairings 351, 166 were type I, 177 type II, and 8 type III, demonstrating considerable potential for application in electronic and optoelectronic devices. The band gap typically decreases as the chalcogen atom Q becomes heavier (S → Se → Te) for a constant M, with compounds based on Si deviating from this consistent trend. An examination of the bonding characteristics revealed that the ionic nature of the M−Q bond, derived from differences in electronegativity, correlates with the Bader charge of the M atom ($\overline{q}_B^M$). Although the ionic character can approximate 21% as $\overline{q}_B^M$ increases, the bonding within these compounds remains predominantly covalent. Compounds containing tellurium exhibit the least ionic character, providing insights into their chemical stability and electronic behavior.

Finally, random forest models were used to forecast several energy-related properties, including total energy per formula unit, relative energy, formation enthalpy, and cohesive energy, utilizing elemental and structural descriptors. These models demonstrated elevated R^2 scores (resubstitution >0.95; cross-validation >0.9) for most of the energy properties, with the exception of relative energy, which was ≈0.68, thus indicating a high predictive capacity. Specifically, for the formation enthalpy, the elemental group of the element M also played a significant role. Although these models successfully captured general trends, the observed discrepancy between predicted and actual values, particularly for relative energy, emphasizes that quantum simulations remain essential for achieving high precision in predicting properties of these 2D materials. These findings strongly suggest that electronic charge distributions serve as essential descriptors for subsequent machine learning inquiries in this domain, paving the way for accelerated materials discovery.

■ ASSOCIATED CONTENT

Data Availability Statement

All DFT calculations were performed using the VASP[40,41] software. For Bader charge analysis, the data generated by VASP was processed using the Bader charge analysis code[68] developed by the Henkelman Research Group at the University of Texas at Austin. Phonon calculations were performed using VASP in conjunction with Phonopy,[47] which was responsible for generating supercells from the unit cell of each compound of lowest energy, as well as preparing the necessary input files for the VASP simulations. In addition, custom Python scripts were developed to automate key processes in this study, available in the SAbiM Group repository (https://github.com/Sabim-UFSCar/projects). These scripts were used for the following purposes: Generating input files and directory structures for each simulation; checking for errors, warnings, and potential issues in the simulations; extracting and analyzing simulation results. Python was chosen for these tasks due to its compatibility with tools such as VASPKIT.[69] py4vasp and the Atomic Simulation Environment,[70] which streamline data processing and analysis. Throughout all stages of this study, rigorous verification procedures and detailed logarithms were implemented to ensure the efficiency of the analysis of the 351 initial structures and the reliability of the results. A spreadsheet detailing 30 specific descriptors is provided along with the values of four target properties used in the machine learning analysis for

https://doi.org/10.1021/acs.jpcc.5c04625
J. Phys. Chem. C XXXX, XXX, XXX−XXX

optimized structures. In addition, three custom Python scripts developed to facilitate simulation automation and data analysis within this study are also included.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcc.5c04625.

Additional computational details, computational convergence tests, data used for the figures, and complementary analyzes are reported (PDF)

(CSV)

(ZIP)

AUTHOR INFORMATION

Corresponding Author

Juarez L. F. Da Silva − *São Carlos Institute of Chemistry, University of São Paulo, 13560-970 São Carlos, São Paulo, Brazil;* orcid.org/0000-0003-0645-8760; Email: juarez_dasilva@iqsc.usp.br

Authors

Mateus B. P. Querne − *Department of Physics, Federal University of São Carlos, 13565-905 São Carlos, São Paulo, Brazil;* orcid.org/0000-0002-9240-7101

Marco A. M. T. Machado − *Department of Physics, Federal University of São Carlos, 13565-905 São Carlos, São Paulo, Brazil;* orcid.org/0009-0003-4468-8679

Ronaldo C. Prati − *Center of Mathematics, Computer Science, and Cognition, Federal University of ABC, 09210-580 Santo André, São Paulo, Brazil;* orcid.org/0000-0001-8597-4987

Natan M. Regis − *Department of Physics, Federal University of São Carlos, 13565-905 São Carlos, São Paulo, Brazil;* orcid.org/0000-0003-1097-3036

Matheus P. Lima − *Department of Physics, Federal University of São Carlos, 13565-905 São Carlos, São Paulo, Brazil;* orcid.org/0000-0001-5389-7649

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcc.5c04625

Funding

The Article Processing Charge for the publication of this research was funded by the Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior (CAPES), Brazil (ROR identifier: 00x0ma614).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge the support from FAPESP (São Paulo Research Foundation) and Shell with project numbers 2017/11631-2 and 2018/21401-7, and the strategic importance of the support provided by ANP (Brazil's National Oil, Natural Gas and Biofuels Agency) through the R&D levy regulation. The authors also thank the infrastructure provided to our computer cluster by the Department of Information Technology—Campus São Carlos. M.P.L. acknowledges financial support from the CNPq (Brazilian National Council for Scientific and Technological Development) Grant Number 314169/2023-7 and FAPESP with project number 2024/08610-7. M.B.P.Q. acknowledges financial support from FAPESP with project number 2025/10506-6. We acknowledge the use of advanced language models for their assistance in English-language editing, grammar revision, and text refinement.

ABBREVIATIONS

NLP natural language processing; DFT density functional theory; GGA generalized gradient approximation; PBE Perdew−Burke−Ernzerhof; HSE06 Heyd−Scuseria−Ernzerhof; VASP Vienna Ab initio Simulation Package; PAW projected augmented-wave

REFERENCES

(1) Novoselov, K. S.; Geim, A. K.; Morozov, S. V.; Jiang, D.; Zhang, Y.; Dubonos, S. V.; Grigorieva, I. V.; Firsov, A. A. Electric Field Effect in Atomically Thin Carbon Films. *Science* **2004**, *306*, 666−669.

(2) Novoselov, K. S. Nobel Lecture: Graphene: Materials in the Flatland. *Rev. Mod. Phys.* **2011**, *83*, 837−849.

(3) Madurani, K. A.; Suprapto, S.; Machrita, N. I.; Bahar, S. L.; Illiya, W.; Kurniawan, F. Progress in Graphene Synthesis and its Application: History, Challenge and the Future Outlook for Research and Industry. *ECS J. Solid State Sci. Technol.* **2020**, *9*, 093013.

(4) Li, Z.; Qian, H.; Wu, J.; Gu, B.-L.; Duan, W. Role of Symmetry in the Transport Properties of Graphene Nanoribbons under Bias. *Phys. Rev. Lett.* **2008**, *100*, 206802.

(5) Bolotin, K.; Sikes, K.; Jiang, Z.; Klima, M.; Fudenberg, G.; Hone, J.; Kim, P.; Stormer, H. Ultrahigh electron mobility in suspended graphene. *Solid State Commun.* **2008**, *146*, 351−355.

(6) Randviir, E. P.; Brownson, D. A.; Banks, C. E. A decade of graphene research: production, applications and outlook. *Mater. Today* **2014**, *17*, 426−432.

(7) Shi, W.; Ye, J.; Zhang, Y.; Suzuki, R.; Yoshida, M.; Miyazaki, J.; Inoue, N.; Saito, Y.; Iwasa, Y. Superconductivity Series in Transition Metal Dichalcogenides by Ionic Gating. *Sci. Rep.* **2015**, *5*, 12534.

(8) Lin, Z.; Qin, W.; Zeng, J.; Chen, W.; Cui, P.; Cho, J.-H.; Qiao, Z.; Zhang, Z. Competing Gap Opening Mechanisms of Monolayer Graphene and Graphene Nanoribbons on Strong Topological Insulators. *Nano Lett.* **2017**, *17*, 4013−4018.

(9) Choi, W.; Choudhary, N.; Han, G. H.; Park, J.; Akinwande, D.; Lee, Y. H. Recent development of two-dimensional transition metal dichalcogenides and their applications. *Mater. Today* **2017**, *20*, 116−130.

(10) Besse, R.; Lima, M. P.; Da Silva, J. L. F. First-Principles Exploration of Two-Dimensional Transition Metal Dichalcogenides Based on Fe, Co, Ni, and Cu Groups and Their van der Waals Heterostructures. *ACS Appl. Energy Mater.* **2019**, *2*, 8491−8501.

(11) Garcia, J. H.; Vila, M.; Cummings, A. W.; Roche, S. Spin transport in graphene/transition metal dichalcogenide heterostructures. *Chem. Soc. Rev.* **2018**, *47*, 3359−3379.

(12) Kumar, R.; Goel, N.; Hojamberdiev, M.; Kumar, M. Transition metal dichalcogenides-based flexible gas sensors. *Sens. Actuators A Phys.* **2020**, *303*, 111875.

(13) Zhou, Z.; Lv, J.; Tan, C.; Yang, L.; Wang, Z. Emerging Frontiers of 2D Transition Metal Dichalcogenides in Photovoltaics Solar Cell. *Adv. Funct. Mater.* **2024**, *34*, 2316175.

(14) Chang, K.; Parkin, S. S. P. Experimental formation of monolayer group-IV monochalcogenides. *J. Appl. Phys.* **2020**, *127*, 220902.

(15) Rahman, S.; Sharme, R. K.; Terrones, M.; Rana, M. M. Recent Progress on Layered Sn and Pb-Based Mono Chalcogenides: Synthesis, Structure, Optical, and Thermoelectric Properties and Related Applications. *Nanomater* **2024**, *14*, 1530.

(16) Hu, Z.; Ding, Y.; Hu, X.; Zhou, W.; Yu, X.; Zhang, S. Recent progress in 2D group IV−IV monochalcogenides: synthesis, properties and applications. *Nanotechnology* **2019**, *30*, 252001.

(17) Feng, W.; Zheng, W.; Cao, W.; Hu, P. Back Gated Multilayer InSe Transistors with Enhanced Carrier Mobilities via the Suppression of Carrier Scattering from a Dielectric Interface. *Adv. Mater.* **2014**, *26*, 6587−6593.

(18) Lei, S.; Ge, L.; Liu, Z.; Najmaei, S.; Shi, G.; You, G.; Lou, J.; Vajtai, R.; Ajayan, P. M. Synthesis and Photoresponse of Large GaSe Atomic Layers. *Nano Lett.* **2013**, *13*, 2777−2781.
(19) Wong, L. W.; Yang, K.; Han, W.; Zheng, X.; Wong, H. Y.; Tsang, C. S.; Lee, C.-S.; Lau, S. P.; Ly, T. H.; Yang, M.; Zhao, J. Deciphering the ultra-high plasticity in metal monochalcogenides. *Nat. Mater.* **2024**, *23*, 196−204.
(20) Chang, H.-C.; Tu, C.-L.; Lin, K.-I.; Pu, J.; Takenobu, T.; Hsiao, C.-N.; Chen, C.-H. Synthesis of Large-Area InSe Monolayers by Chemical Vapor Deposition. *Small* **2018**, *14*, 1802351.
(21) Hu, P.; Wen, Z.; Wang, L.; Tan, P.; Xiao, K. Synthesis of Few-Layer GaSe Nanosheets for High Performance Photodetectors. *ACS Nano* **2012**, *6*, 5988−5994.
(22) Vaughn, D. D. I.; Patel, R. J.; Hickner, M. A.; Schaak, R. E. Single-Crystal Colloidal Nanosheets of GeS and GeSe. *J. Am. Chem. Soc.* **2010**, *132*, 15170−15172.
(23) Zhang, Z.; Yang, J.; Zhang, K.; Chen, S.; Mei, F.; Shen, G. Anisotropic photoresponse of layered 2D SnS-based near infrared photodetectors. *J. Mater. Chem. C* **2017**, *5*, 11288−11293.
(24) Huang, Y.; Li, L.; Lin, Y.-H.; Nan, C.-W. Liquid Exfoliation Few-Layer SnSe Nanosheets with Tunable Band Gap. *J. Phys. Chem. C* **2017**, *121*, 17530−17537.
(25) Akay, D.; Kocak, M. B. Controlling the Intrinsic Charge Carrier Properties of Two-Dimensional Monochalcogenides (GeSe). *ACS Omega* **2024**, *9*, 42778−42785.
(26) Xu, X.; Li, M.; Lin, X.; Wang, H. Giant reversible electrocaloric effect in monolayer group-IV monochalcogenides. *Phys. Rev. B* **2024**, *110*, 064103.
(27) Demirci, S.; Avazl ı, N.; Durgun, E.; Cahangirov, S. Structural and electronic properties of monolayer group III monochalcogenides. *Phys. Rev. B* **2017**, *95*, 115409.
(28) Sheng, K.; Wang, Z.-Y. Robust spin-valley coupling in hexagonal monolayers of Tl-based monochalcogenides. *J. Mater. Chem. C* **2024**, *12*, 12275−12281.
(29) Olivetti, E. A.; Cole, J. M.; Kim, E.; Kononova, O.; Ceder, G.; Han, T. Y.-J.; Hiszpanski, A. M. Data-driven materials research enabled by natural language processing and information extraction. *Appl. Phys. Rev.* **2020**, *7*, 041317.
(30) Brito, A. C. M.; Oliveira, M. C. F.; Oliveira, O. N.; Silva, F. N.; Amancio, D. R. Network Analysis and Natural Language Processing to Obtain a Landscape of the Scientific Literature on Materials Applications. *ACS Appl. Mater. Interfaces* **2023**, *15*, 27437−27446.
(31) Prati, R. C. Recent Advances in Natural Language Processing in Chemistry and Materials Science. *J. Braz. Chem. Soc.* **2025**, *36*, No. e20250067.
(32) Arrieta, A. B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R. others Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82−115.
(33) Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**, arXiv:1705.07874.
(34) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864−B871.
(35) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133−A1138.
(36) Lejaeghere, K.; Bihlmayer, G.; Björkman, T.; Blaha, P.; Blügel, S.; Blum, V.; Caliste, D.; Castelli, I. E.; Clark, S. J.; Dal Corso, A.; de Gironcoli, S.; Deutsch, T.; Dewhurst, J. K.; Di Marco, I.; Draxl, C.; Dułak, M.; Eriksson, O.; Flores-Livas, J. A.; Garrity, K. F.; Genovese, L.; Giannozzi, P.; Giantomassi, M.; Goedecker, S.; Gonze, X.; Grånäs, O.; Gross, E. K. U.; Gulans, A.; Gygi, F.; Hamann, D. R.; Hasnip, P. J.; Holzwarth, N. A. W.; Iuşan, D.; Jochym, D. B.; Jollet, F.; Jones, D.; Kresse, G.; Koepernik, K.; Küçükbenli, E.; Kvashnin, Y. O.; Locht, I. L. M.; Lubeck, S.; Marsman, M.; Marzari, N.; Nitzsche, U.; Nordström, L.; Ozaki, T.; Paulatto, L.; Pickard, C. J.; Poelmans, W.; Probert, M. I. J.; Refson, K.; Richter, M.; Rignanese, G.-M.; Saha, S.; Scheffler, M.; Schlipf, M.; Schwarz, K.; Sharma, S.; Tavazza, F.; Thunström, P.; Tkatchenko, A.; Torrent, M.; Vanderbilt, D.; van Setten, M. J.; Van Speybroeck, V.; Wills, J. M.; Yates, J. R.; Zhang, G.-X.; Cottenier, S. Reproducibility in density functional theory calculations of solids. *Science* **2016**, *351*, aad3000.
(37) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, *65*, 1501−1509.
(38) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50*, 17953−17979.
(39) Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-wave Method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 1758−1775.
(40) Kresse, G.; Hafner, J. Ab initio molecular dynamics for open-shell transition metals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *48*, 13115−13118.
(41) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 11169−11186.
(42) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(43) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **2003**, *118*, 8207−8215.
(44) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Fractional charge perspective on the band gap in density-functional theory. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *77*, 115123.
(45) Crowley, J. M.; Tahir-Kheli, J.; Goddard, W. A. Resolution of the Band Gap Prediction Problem for Materials Design. *J. Phys. Chem. Lett.* **2016**, *7*, 1198−1203.
(46) Francis, G. P.; Payne, M. C. Finite basis set corrections to total energy pseudopotential calculations. *J. Condens. Matter Phys.* **1990**, *2*, 4395.
(47) Togo, A.; Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **2015**, *108*, 1−5.
(48) Clarivate Web Of Scicence. https://www.webofscience.com/wos/woscc/basic-search, 2024; (accessed 07/18/2024).
(49) Reimers, N.; Gurevych, I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*; Association for Computational Linguistics, 2019.
(50) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. *Attention is All You Need*; Advances in Neural Information Processing Systems, 2017; Vol. *30*.
(51) Tunstall, L.; Reimers, N.; Jo, U. E. S.; Bates, L.; Korat, D.; Wasserblat, M.; Pereg, O. Efficient Few-Shot Learning Without Prompts. *arXiv* **2022**, arXiv:2209.11055. https://arxiv.org/abs/2209.11055
(52) Rane, N.; Choudhary, S.; Rane, J. Gemini versus ChatGPT: applications, performance, architecture, capabilities, and implementation. *J. Appl. Artif. Intell.* **2024**, *5*, 69−93.
(53) Querne, M. B. P.; Bracht, J. M.; Da Silva, J. L. F.; Janotti, A.; Lima, M. P. Crystal structure and electrical and optical properties of two-dimensional group-IV monochalcogenides. *Phys. Rev. B* **2023**, *108*, 085409.
(54) Özdamar, B.; Özbal, G.; Çınar, M. N.; Sevim, K.; Kurt, G.; Kaya, B.; Sevinçli, H. Structural, vibrational, and electronic properties of single-layer hexagonal crystals of group IV and V elements. *Phys. Rev. B* **2018**, *98*, 045431.
(55) Zhou, J.; Shen, L.; Costa, M. D.; Persson, K. A.; Ong, S. P.; Huck, P.; Lu, Y.; Ma, X.; Chen, Y.; Tang, H.; Feng, Y. P. 2DMatPedia, an open computational database of two-dimensional materials from top-down and bottom-up approaches. *Sci. Data* **2019**, *6*, 86.
(56) National Center for Biotechnology Information Electronegativity in the Periodic Table of Elements. https://pubchem.ncbi.nlm.nih.gov/periodic-table/electronegativity, 2024; Accessed: 12/26/2024.
(57) Inui, G. K.; Silveira, J. F. R. V.; Dias, A. C.; Besse, R.; Da Silva, J. L. F. Ab initio screening of two-dimensional CuQx and AgQx chalcogenides. *J. Phys.: Condens. Matter* **2022**, *34*, 305703.

https://doi.org/10.1021/acs.jpcc.5c04625
J. Phys. Chem. C XXXX, XXX, XXX−XXX

(58) Gomes, L. C.; Carvalho, A. Phosphorene analogues: Isoelectronic two-dimensional group-IV monochalcogenides with orthorhombic structure. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2015**, *92*, 085406.
(59) Querne, M. B. P.; Dias, A. C.; Janotti, A.; Da Silva, J. L. F.; Lima, M. P. Tuning Excitonic Properties of Monochalcogenides via Design of Janus Structures. *J. Phys. Chem. C* **2024**, *128*, 12164−12177.
(60) Bracht, J. M.; Querne, M. B. P.; Da Silva, J. L. F.; Lima, M. P. Theoretical Investigation of Stacked Two-Dimensional Transition-Metal Dichalcogenide Materials: The Role of Chemical Species and Number of Monolayers. *ACS Omega* **2025**, *10*, 8922−8934.
(61) Ma, Y.; Kou, L.; Dai, Y.; Heine, T. Proposed two-dimensional topological insulator in SiTe. *Phys. Rev. B* **2016**, *94*, 201104.
(62) Anderson, R. L. Germanium-Gallium Arsenide Heterojunctions [Letter to the Editor]. *IBM J. Res. Dev.* **1960**, *4*, 283−287.
(63) Wang, C.; Ni, H.; Dai, J.; Liu, T.; Wu, Z.; Chen, X.; Dong, Z.; Qian, J.; Wu, Z. Comparison of highly active Type-I and Type-II heterojunction photocatalytic composites synthesized by electrospinning for humic acid degradation. *Chem. Phys. Lett.* **2022**, *803*, 139815.
(64) Ferdous, N.; Islam, M. S.; Park, J. A resilient type-III broken gap Ga_2O_3/SiC van der Waals heterogeneous bilayer with band-to-band tunneling effect and tunable electronic property. *Sci. Rep.* **2024**, *14*, 12748.
(65) Li, C.; Cao, Q.; Wang, F.; Xiao, Y.; Li, Y.; Delaunay, J.-J.; Zhu, H. Engineering graphene and TMDs based van der Waals heterostructures for photovoltaic and photoelectrochemical solar energy conversion. *Chem. Soc. Rev.* **2018**, *47*, 4981−5037.
(66) Pike, N. A.; Van Troeye, B.; Dewandre, A.; Petretto, G.; Gonze, X.; Rignanese, G.-M.; Verstraete, M. J. Origin of the counterintuitive dynamic charge in the transition metal dichalcogenides. *Phys. Rev. B* **2017**, *95*, 201106.
(67) Callister, W. D.; Rethwisch, D. G. *Materials Science and Engineering: An Introduction*; Wiley, 2014.
(68) Tang, W.; Sanville, E.; Henkelman, G. A grid-based Bader analysis algorithm without lattice bias. *J. Phys.: Condens. Matter* **2009**, *21*, 084204.
(69) Wang, V.; Xu, N.; Liu, J.-C.; Tang, G.; Geng, W.-T. VASPKIT: A user-friendly interface facilitating high-throughput computing and analysis using VASP code. *Comput. Phys. Commun.* **2021**, *267*, 108033.
(70) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; Hermes, E. D.; Jennings, P. C.; Jensen, P. B.; Kermode, J.; Kitchin, J. R.; Kolsbjerg, E. L.; Kubal, J.; Kaasbjerg, K.; Lysgaard, S.; Maronsson, J. B.; Maxson, T.; Olsen, T.; Pastewka, L.; Peterson, A.; Rostgaard, C.; Schiøtz, J.; Schütt, O.; Strange, M.; Thygesen, K. S.; Vegge, T.; Vilhelmsen, L.; Walter, M.; Zeng, Z.; Jacobsen, K. W. The atomic simulation environment−a Python library for working with atoms. *J. Phys.: Condens. Matter* **2017**, *29*, 273002.

https://doi.org/10.1021/acs.jpcc.5c04625
J. Phys. Chem. C XXXX, XXX, XXX−XXX