High Throughput Screening of Millions of van der Waals Heterostructures for Superlubricant Applications

Marco Fronzi,[1, *] Sherif Abdulkader Tawfik,[2] Mutaz Abu Ghazaleh,[1] Olexandr
Isayev,[3] David A.Winkler,[4, 5, 6, 7] Joe Shapter,[8, 9] and Michael J. Ford[1, †]

[1] *School of Mathematical and Physical Science, University of Technology Sydney, Ultimo 2007, New South Wales 2007, Australia*

[2] *School of Science Cluster, Department of Applied Chemistry & Environmental Science,*
RMIT University, 124 La Trobe St, Melbourne, Victoria 3000, Australia

[3] *Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry,*
UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC 27599, USA

[4] *Data61, Commonwealth Scientific and Industrial Research Organization, Pullenvale, Queensland 4069, Australia*

[5] *Monash Institute of Pharmaceutical Sciences, Monash University,*
381 Royal Parade, Parkville, Victoria 3052, Australia

[6] *Latrobe Institute for Molecular Science, La Trobe University,*
Kingsbury Drive, Bundoora, Victoria 3086, Australia

[7] *School of Pharmacy, The University of Nottingham, Nottingham NG7 2RD, UK*

[8] *Australian Institute for Bioengineering and Nanotechnology,*
The University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia

[9] *College of Science and Engineering, Flinders University,*
Bedford Park, Adelaide, South Australia 5042, Australia

(Dated: August 19, 2020)

The screening of novel materials is an important topic in the field of materials science. Although traditional computational modelling, especially first-principles approaches, is a very useful and accurate tool to predict the properties of novel materials, it still demands extensive and expensive state-of-the-art computational resources. Additionally, they can often be extremely time consuming. We describe a time and resource efficient machine learning approach to create a dataset of structural properties of 18 million van der Waals layered structures. In particular, we focus on the interlayer energy and the elastic constant of layered materials composed of two different 2-dimensional (2D) structures, that are important for novel solid lubricant and super-lubricant materials. We show that machine learning models can predict results of computationally expansive approaches (i.e. density functional theory) with high accuracy.

I. MAIN

Solid lubricants are materials that reduce friction and damage of contacting surfaces in relative motion. A good lubricant can be identified by two main properties: shear strength and abrasivity. The dynamics of solid lubrication show that sliding motion is followed by significant ductile shear of the solid lubricant. Therefore, the solid lubricant must have low shear strength, which occurs in the crystalline phase by slip along preferred crystallographic planes.[1] On the other hand, abrasivity is a relative property that is a function of the hardness ratio of the lubricant and the lubricated material. Typically, the lubricant particles should be softer than the contact material to avoid abrasions.[1] Clearly, thermo-chemical stability in the environment of the application is also essential. This is particularly important for high temperature applications but is equally important for moderate temperature applications to insure adequate storage stability and to avoid corrosion by atmospheric components such as oxygen and salt spray. According to this description, thermo-chemically stable materials, with a low interlayer energy and low degree of hardness, are good candidates for solid lubricant materials.

A significant drawback for the application of conventional structural materials concerns the strong anisotropic nature of friction in homogeneous and heterogeneous interfaces with respect to their relative orientation. Even when the interface is formed in an incommensurate ultra-low friction configuration, the contact surfaces have a tendency to rotate towards the aligned commensurate configuration during the sliding motion and eventually lock in a high friction state, which corresponds to a higher interaction energy.[2] Low-energy-interaction/high-shear-motion structures can be found in novel 2-dimensional (2D) van der Waals (vdW) layered structures.[3,4] Van der Waals forces differ from covalent and ionic bonding in that they are caused by correlations in the fluctuating polarizations of nearby particles, resulting in weak, long-range forces. The vdW strength of two contacting structures is a key requirement for lubricity/superlubricity behaviour.[5]

Although the mechanism of friction can be in general complicated to describe, a few works in the literature suggests a simplification can be done in specific cases. In the text we mention sheet silicates, whose basic structure is built up of regular sequences of tetrahedral and of octahedral (Si, Al, Mo) layers.[6,7] These structures share in common with the vdW hetero-structures the presence of in-plane strong atomic bonds and out-of -plane weak atomic bonds. In these particular class of materials, the friction mechanism is dominated by the intensity of the atomic bonds that hold together the layers. Furthermore, the inter-layer interaction strengths during relative sliding of layers in a 2D material was also found in correlation with the experimentally measured frictional forces using scanning probe techniques.[8,9] On the other hand, in most materials, including ceramics and inorganic composites, the hardness is directly dependent on elastic modulus. Elastic modulus is an intrinsic material property and fundamentally related to atomic bonding, whereas hardness is an engineering property and for some materials it can be related to yield strength. Hardness has strong usefulness in characterization of different types of microstructures, and in general, the trend of the indentation hardness against the elastic modulus for a large range of materials indicates that the two do increase together.[10] A simplified model can be formulated where the elastic and plastic deformation components are assumed to act in series, with two fundamental material parameters: an elastic modulus and a "resistance to plastic deformation". Within this approach, the indentation hardness is actually related to both of these parameters, a it is a function of both the elastic and plastic parts. In most materials, including ceramics and inorganic composites the contributions to the total deformation from elastic and plastic deformation can be similar and so the results from, for example, a series of

nanoindentation, indicate that the hardness is directly dependent on elastic modulus, that we approximate with the C_{33} (see Section III A 1).[11]

Currently, only a few Van der Waals structures, graphite, boron nitride and molybdenum disulfide, are used as dry lubricants. Recently, 6,138 potential single layer 2D materials have been identified and are available from the online database, https://2dmatpedia.org. These structures have been obtained by theoretical exfoliation and elemental substitution from a large number of inorganic bulk structures included in the Materials Project database (https://materialsproject.org/).[12,13] By a direct stacking, it is possible to generate 18,834,453 (FULL SET) unique novel bilayer heterostructures from these 2D monolayers ($N_b = N_m(N_m + 1)/2$).[14–16] Furthermore, previous research has shown that stacking processes in 2D materials are self-cleaning, resulting in near ideal hybrid 2D layered structures.[17,18] Due to the large number of structures and to the nature of the vdW forces that are responsible for the stacking, it is likely that within the generated heterostructures there will be many with low interaction energies, and eventually a desirable softness and temperature stability to be used in super-lubricant applications. However, due to the extremely large number of possible layer combinations, exhaustive experimental assessment is clearly infeasible. Traditional quantum chemical computational techniques, such density functional theory (DFT), can accurately predict the properties of such materials and can be used for the discovery of new materials, however, the computational demands of these calculations means the materials assessment process is still very slow.[19,20]

Here, we propose a time and resources efficient machine learning (ML) approach that, combined with a limited number of first-principles calculations, is able to calculate and predict the interlayer energy (IE) and the elastic constant (C_{33}) of a large number of layered heterostructures, expanding the capabilities of a canonical theory. We use DFT to predict the desired properties of a relatively small number of 2D layered hybrid materials. We leverage this smaller pool of results by using them to train supervised machine learning models. The models can then rapidly and reliably predict these quantities for a large number of structures (i.e. FULL SET) within the domain of applicability of the models. We use a Bayesian neural network (BNN) model, which importantly provides the confidence interval for each prediction point.[21,22] While such techniques are routine within the machine learning/artificial intelligence community theory, they are only just starting to be used widely by the materials science community, although some introduction started already in the late 90's.[23]

We calculated the interlayer energies (IE) and the elastic constant C_{33} for two representative subsets of all the possible combinations of two 2D materials, consisting of ~200 (a detailed description is in Section III A 1, and the complete list is reposted in SI).[24,25] Members of these subsets had an interlayer energy $E \leq -1.0$ eV/Å^2 and a maximum lattice constant mismatch of 2%, where a negative interlayer energy denotes the two monolayers are bound to each other. Due to the nature of the vdW forces, the interlayer energies of 2D heterostructures depend weakly on the twisting angle between the specific stacking configurations, typically by ~30 meV per unit formula of one reference layer.[26] Therefore, it is important to point out that any twist angle would not affect our conclusions. This makes the problem of finding low friction structures less difficult from a computational perspective.[27] Consequently, for simplicity, we set the twist angle between the two monolayers to be 0 degrees. Finally, the temperature stability of each bilayer was estimated by considering the minimum value of the decomposition energy of each monolayer comprising the heterostructure. The interlayer energies that we obtain agree well with other calculated values reported in the literature, whereas the C_{33} values fall in the range of other calculations, as shown in Table I. The significant deviations from the experimental values, that the literature reports only in a very limited number of works, can be attributed to

the lack of sufficient information to compare the experimental approach used and and theory, and possibly to the limits of the current vdW approximations. However, the values calculated here follow the same trend as the values calculated using different vdW approximations, which support the qualitative reliability of our calculation.[28–30] Here, we report only a few examples that highlight discrepancies and agreements with other works, however, a wider comparison can be found in the literature, particularly in the work of Björkman *et al.*[28,31–37]
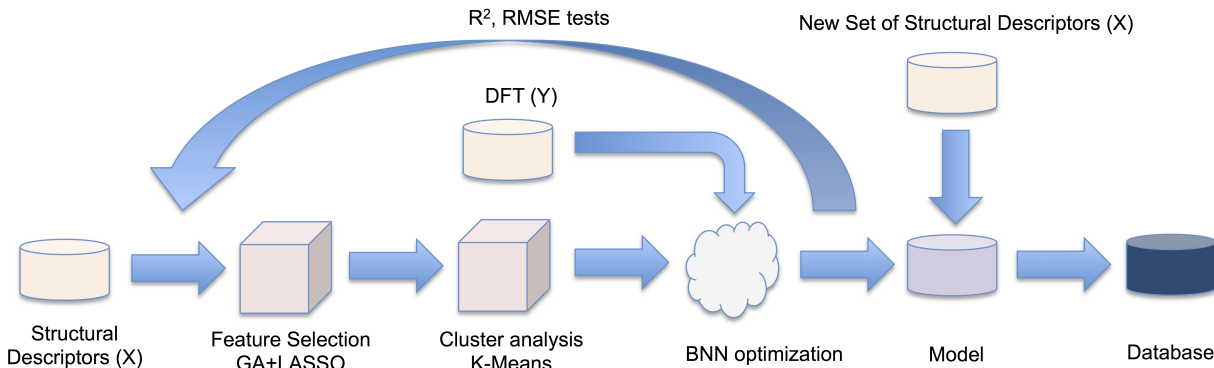

FIG. 1. (Color online) a) Schematic representation of the workflow used to create our BNN model and to extrapolate IE and C_{33} values using a set of descriptors for new structures.

The DFT results were used to train supervised machine learning (ML) models of interlayer energies and elastic constants. A schematic diagram of our combined DFT/machine learning approach is shown in Figure 1. The Bayesian neural networks that we employed require a target property (Y), calculated by means of DFT, and a list of descriptors (X) (mathematical objects that represent the molecular properties of the materials) for each bilayer. We generated descriptors for each monolayer using the method developed by one of us previously which have been shown to be useful in previous work.[38,39] Descriptors for each bilayer were obtained by adding the values of the descriptors for the two monolayers, as described by Tawfik *et al.*[39] The algorithm generated 2,764 descriptors for each monolayer. To avoid problems of overfitting due to this large number of descriptors, we selected a small subset of the most relevant features after using a combination of Genetic Algorithm (GA) search and LASSO regression. This reduced the number of descriptors to 42 for the interlayer energy model and 89 for the C_{33} model. The GA and LASSO eliminate irrelevant or low relevance features, which makes the ML models train more quickly and easier to interpret.[40] The same subset of relevant descriptors is used for predicting the training and test sets, and for generating the properties of the FULL SET.[38,39] The domain of applicability of the model to the large dataset is ensured by the equivalent range of the descriptor values in the large dataset and in the subset used in the training. To obtain the training and test set in a way that each subgroup contains representative structures of the total set, we perform a K-means cluster analysis. This will maximize the diversity of structures assigned to training and test sets while ensuring that the test sets are still within the domain of the models.

One of the main problems in the application of DFT calculations to heterostructures is the commensurability of the unit cells of the two monolayers. Calculations under periodic boundary conditions can only be performed if supercells of the single monolayer supercells have the same dimensions along the xy plane (where z is the stacking direction). This can be achieved by either increasing the size of the supercell by multiplying the single unit cell for a factor
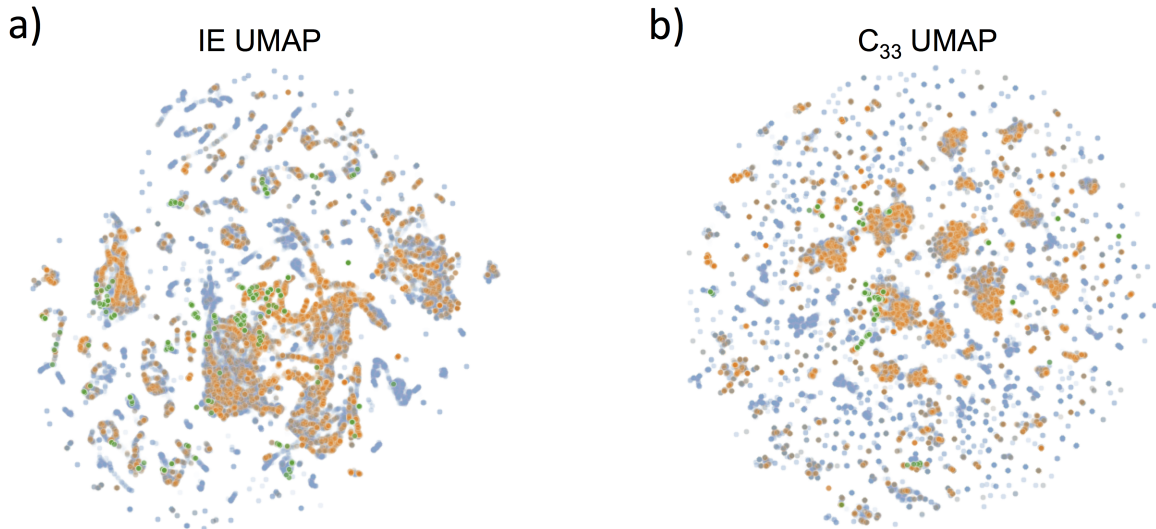
FIG. 2. (Color online) a) and b) represent the UMAPs of the IE and C_{33}, respectively. The blue, orange and green dots represent the total representative bilayers subset, the commensurate unit cells, and the DFT calculations, respectively.

until commensurability is reached, and/or applying a artificial strain to the individual supercells, which must be small enough so that the electronic structure and the corresponding properties of the monolayer are not significantly affected. These conditions significantly limit the number of heterostructures that can be used in DFT calculations. Here, the subset used in the training is represented in the uniform manifold approximation and projection (UMAP) in Figure 2, which shows how the DFT calculations are distributed over the commensurate bilayers, and over the whole dataset. The UMAP gives also additional information on the quasi-uniform distribution of the commensurate cells over the whole set. This was estimated to be $\sim 8.2\%$ of the total number of bilayers, considering two constraints: the lattice mismatch $L_m \leq 2\%$, and the number of atoms in the cell to be $N_a \leq 600$. The first constraint ensures a reliable outcome of DFT calculations with respect to a realistic scenario; the second constraint ensures that the DFT calculations can be performed within a reasonable computational time, which, in conventional DFT calculations, scales cubically with the number of Kohn Sham orbitals in target systems.[19,20,41]

After BNN optimization, performed using cross-validation subsets, we test the quality of the results by calculating R^2, Root Mean Square Error and Mean Absolute Error, as shown in Table II. Here, we consider Root Mean Square Error a better measure of model prediction power than R^2 as they have been shown to be less dependent on the number of training samples and complexity of the models.[45] Our implementation uses a dropout approach that is used during training for regularization, and during predictions, to obtain a statistical distribution of the response, which provide information on the uncertainty of the values, as described in Section III C. To further test the quality of our predictions we perform an additional test on a validation set, which was not used during BNN training.

Using this model we extrapolated the interlayer energy and C_{33} for the FULL SET and the complete list of values, together with the associated uncertainty, is available online (http://doi.org/10.26195/5dd36650d7e1e). Figures 3 a) and b) show the interlayer energy and C_{33} as a heatmap. Here, each axis contains the list of monolayers that forms the bilayers, and the axes are ordered to cluster structures with similar values of the interlayer energy. The order of

TABLE I. Interlayer energies (IE) (in eV/Å^2) and C_{33} (in GPa) measured and calculated by DFT (using vdW-VV10, vdW-DF1, vdW-DF2) as reported in the literature, and calculated in the present work using DFT (vdW-TS).

2D bilayer	Calculated IE	Measured Value	Calculated C_{33}	Measured Value
Graphene	-0.39[28]	-0.020 ± 0.001[34]	23.0[28]	$37-41$[37,42]
	-0.27[28]	-0.016 ± 0.001[32,33]	46.1[28]	
	-0.29[31]	-0.013 ± 0.005[43]	13[37]	
	-0.34[44]			
	-0.34*		15.19*	
h-BN	-0.36[28]		20.2[28]	
	-0.26[28]		41.2[28]	
			11[37]	
	-0.52*		12.10*	
MoS$_2$	-0.44[28]		24[28]	
			55.2[28]	
			49[37]	
	-0.37*		31.27*	

*This work.

TABLE II. R^2, Root Mean Squared Error (MRSE) and mean absolute error (MAE) on test, train, and validations set for interlayer energy and C_{33}. Values of RMSE and MAE are in eV/Å^2 for IE, and in GPa for C_{33}.

Set	R^2	RMSE	MAE
IE-BNN-Test-Set	0.80	0.055	0.035
IE-BNN-Train-Set	0.97	0.014	0.010
IE-BNN-Valid-Set	0.72	0.089	0.055
C_{33}-BNN-Test-Set	0.80	9.98	16.04
C_{33}-BNN-Train-Set	0.98	5.99	5.76
C_{33}-BNN-Valid-Set	0.73	11.89	20.65

the monolayer in the axes is the same in the interlayer energy and C_{33} plots, therefore the sparse clustering in C_{33} map, suggest a weak correlation between the two properties. Figures 3 c) and d) show the relative error calculated for the interlayer energy and C_{33}. The map suggests that our BNN is particularly inaccurate in predicting hard materials, indicated by the coloured areas, which represent a small fraction ($\leq 1\%$) of the overall results. However, the remainder of the map suggests a reliable prediction with an average accuracy of $\sim 4\%$ for the interlayer energy and $\sim 11\%$ for the C_{33}.

To estimate the correlation between the two variables, we calculate the Pearson coefficient, which evaluates the linear relationship between two variables, and the Spearman coefficient, which evaluates the monotonic relationship between two variables (see the Section III D for details). Considering the absolute value of the IE, we obtain a Pearson value 0.06 in ML and 0.05 in DFT, and Spearman value of 0.09 in ML and 0.08 in DFT, indicating an extremely
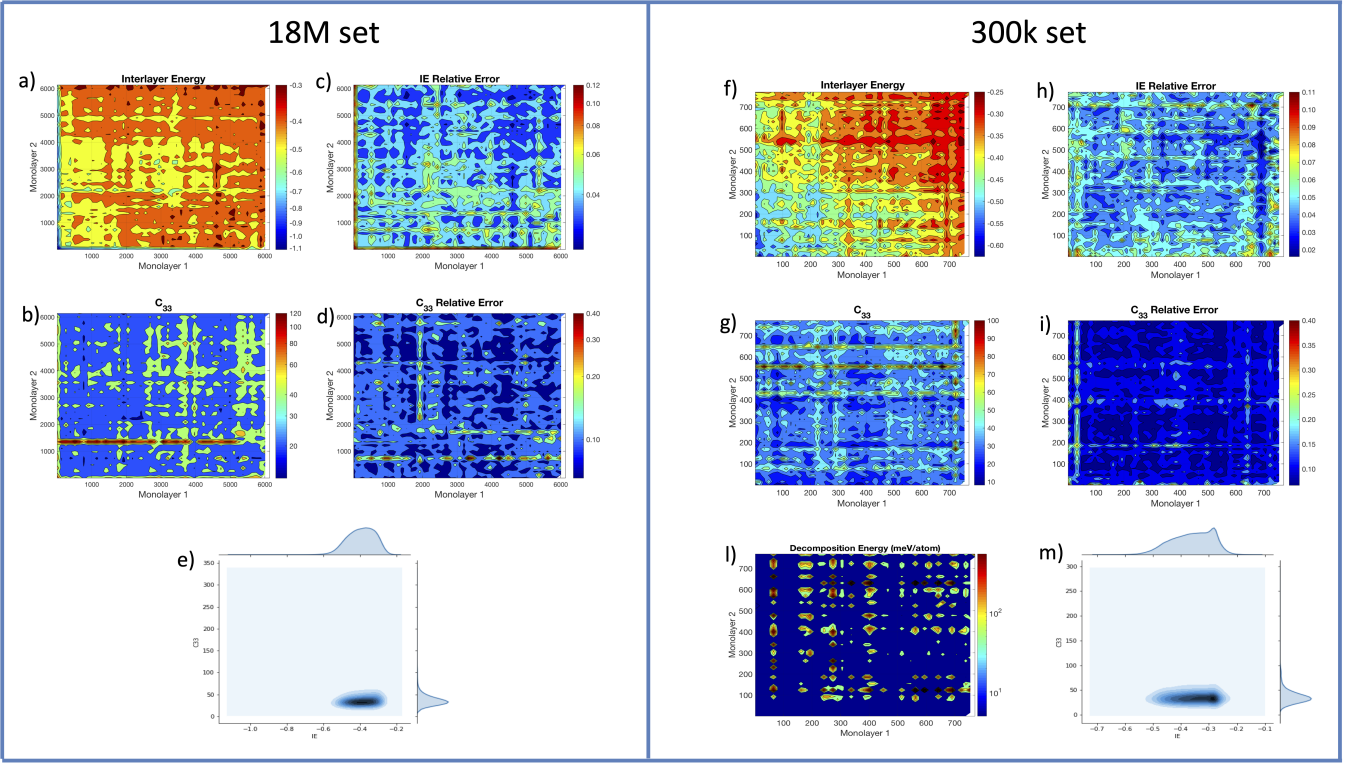
FIG. 3. (Color online) a) and b) Interlayer Energies and C_{33} of the FULL SET. c) and d) Relative Error of the Interlayer Energies and C_{33} of the FULL SET. e) Statistical distribution of IE and C_{33} values in the FULL SET. f) and g) Interlayer Energies and C_{33} of the SUBSET A of. h) and i) Relative Error of the Interlayer Energies and C_{33} of the SUBSET A. l) Temperature stability of the SUBSET A. m) Statistical distribution of IE and C_{33} values in the SUBSET A. Here, Interlayer Energies are expressed in eV/Å^2, the C_{33} in GPa and temperature stability in energy per atom (meV/atom). Absolute errors have been calculated as the standard deviation of the response distribution, using a dropout approach with probability 0.1. Detailed information can be found in Section III C. (The heatmaps have been generated by interpolating the function IE $= f(x,y)$ and the functions $C_{33} = f(x,y)$ so that the images can provide information by showing potential clustering).

weak relationship between IE and C_{33}. However, the same coefficients are higher in the subset of homo-bilayer, where the Pearson and Spearman values are 0.40 and 0.55. Although there is no correlation between the two quantities, we calculated that 90% of the bilayers has an IE between -0.51 eV/Å^2 and -0.28 eV/Å^2 and a C_{33} between 19.44 GPa and 63.44 GPa, as shown in Figures 3 e) and m) (values are reported in the SI). Due to the very large size of the dataset considered here, we can generalize this conclusion to all the possible van der Waals heterostructures formed by assembling any combination of two 2D material.

Although the 6,138 monolayers were theoretically predicted from the existing and thermodynamically relative bulk counterpart, the decomposition energy suggests that most of these structures are not stable, which implies that the stability of a large fraction of the FULL SET cannot be ensured at standard conditions.[13,47] Indeed, only 3,497 monolayers have a decomposition energy larger than 100 meV/atom, the typical threshold for metastability, which can generate 6,114,504 bilayers. However, is has been shown that structures with decomposition energies lower than this threshold are still stable.[48]

We continue our analysis on a subset of the FULL SET that have decomposition energy $\geq$0 meV/atom and exfo-

TABLE III. Monolayer frequency among the bilayers with small and large IE or C_{33} values, included in the SUBSET A. In this analysis, we consider the first 5000 biayers with IE ≥ -0.23 eV/Å^2 or C_{33} ≤ 15.00 GPa, and the first 1000 bilayers with IE ≤ -0.56 eV/Å^2 or C_{33} ≥ 100.00 GPa. The "1T" prefix denotes the 1T polymorph of transition metal chalcogenides. Here, we use the stoichiometric notation of the individual monolayers as in the https://materialsweb.org database.[46]

IE ≥ -0.23 eV/Å^2	OR	C_{33} ≤ 15.00 GPa
In_2S_2: 245		$ZrCd_2H_{12}O_6F_8$: 550
As_4S_6: 233		CdO: 329
$In_2Cl_2O_2$: 222		$Cd_2P_2S_6$: 267
Ga_2Se_2: 208		BN-1T: 251
In_2Se_2: 204		BN: 243
In_2Se_2-1T: 203		C_2: 229
Tc_4P_{16}: 194		$Cd_2Te_2Mo_2O_{12}$: 224
$Sc_2P_2S_8$: 186		$Cd_2P_2S_6$-1T: 217
Al_2Cl_6: 178		$Al_4Cd_2Cl_{16}$: 161
IE ≤ -0.56 eV/Å^2	OR	C_{33} ≥ 100.00 GPa
$Sr_2Ti_2Si_4O_{14}$:179		$HfFeCl_6$-1T: 89
$Ca_2La_2I_{10}$-1T: 121		BN: 83
$Ti_2Ge_2O_6$: 81		$Sr_2Ti_2Si_4O_{14}$: 75
$TmAgP_2Se_6$: 74		BN-1T: 44
$Ti_6H_4O_{14}$: 73		Si_4O_8: 40
Hf_3Te_2: 70		$Al_2Si_2H_4O_9$: 35
$Ca_2La_2I_{10}$: 54		$Ti_6H_4O_{14}$: 30
Ta_3TeI_7: 52		Si_4O_8-1T: 28
V_4F_{16}: 45		$Al_2Si_4O_{11}$: 26

liation energy ≤ 55 meV/atom, which is estimated to be lower than existing 2D materials, suggesting the possibility of exfoliation from bulk phases.[13,46,49–51] This new subset consists of 770 monolayers with a Gibbs free energy at zero temperature that indicates thermodynamic stability at standard conditions.[50,51] The monolayers were identified within the lattice of a large number of 3-dimensional stable structures possessing a layered geometry, and formed by the elements listed in the SI. The combinations of the 770 monolayers generate 296,835 (SUBSET A) of stable and manufacturable bilayers. Using our BNN model, we extrapolated the interlayer energy and C_{33} for the thermodynamically stable SUBSET A, together with the relative error and their temperature stability, shown in Figures 3 f)-l) (the complete list of values is reported available online (http://doi.org/10.26195/5dd36650d7e1e). Here, the quality of the results is assured by the fact that the descriptor values of the structures in the FULL SET and the SUBSET A lie within or close to the same range. Interestingly, although the IE and C_{33} values follow logarithmic distribution with respect of the monolayers composition in the FULL SET and a linear distribution in the SUBSET A, the calculated Pearson and Spearman coefficients for the SUBSET A are 0.07 and 0.09, respectively, confirming the lack of correlation between the two properties.

From a screening of the SUBSET A, we extracted the most common monolayers with a low absolute value of the interlayer energy (≥ -0.23 eV/Å^2) and low C_{33} (≤ 15.00 GPa), listed in Table III. Figure 3 g) shows areas, at around

550 on Monolayer2 axis, where a few monolayers, when coupled with a large fraction of the whole set of monolayers, will form a bilayer with a particularly high C_{33}. We found the most frequent monolayer to give high C_{33} to be HfFeCl$_6$-1T, which is the sixth most frequent monolayer with high C_{33} in the FULL SET. Interestingly, BN appear frequently in both high and low C_{33} bilayers. In conclusion, our IE predictions suggest the possible application of a significant fraction of the SUBSET A as super-lubricant, where the best one would be as a As$_4$S$_6$-In$_2$Se$_2$ together with its polymorph form As$_4$S$_6$-In$_2$Se$_2$-1T with and interlayer energy of -0.12eV/Å^2, which however is stable only up to 69K, while its polymorph form As$_4$S$_6$-In$_2$Se$_2$-1T, while having a very similar interlayer energy, is stable up to 928K. Furthermore, the predicted C_{33} value of the As$_4$S$_6$-In$_2$Se$_2$ is 39.49 GPa, which is relatively low, ensuring its wide applicability. On the other hand ZrCdH$_{12}$O$_6$F$_8$-Hf$_2$Br$_2$N$_2$-1T is the softest bilayer, with a C_{33} value of 4.04 GPa. Although ZrCdH$_{12}$O$_6$F$_8$-Hf$_2$Br$_2$N$_2$-1T has interlayer energy value of -0.36eV/Å^2, which lies in the middle of the our IE range, it represents the most universal dry lubricant. Our calculations show that there are 287,245 structure with decomposition energy ≥ 0 meV/atom, 60,910 of which are stable at room temperature, and 19,900 up to 1300K.

II. CONCLUSIONS

With the present work, we created a very large database of atomic properties (IE and C_{33}, together with the relative temperature stability) for a class of materials with growing technological and scientific interest. Furthermore, we demonstrated the potential of machine learning in amplifying the capabilities of conventional computational approaches used in materials discovery. The approach described, augmenting DFT calculations with machine learning models, is fully transferable to other the scenarios where a very large number of structures can be generated from combinations of a relatively small number of atomic structures.

III. METHODOLOGY

Machine Learning approach relies on high quality structured data, using a set of descriptors to indicate known properties, from which the algorithm will learn hidden patterns. In general, the problem can be reduced to the identification of a general non-linear function $Y = f(X)$, where Y is either the interlayer energy or the elastic constant (C_{33}), X represents the input space of descriptors, and f is the transfer function that links the descriptors to the response variable. The work-flow of our implementation is structured in three main parts, shown schematically in Figure 1 a), and concatenated as follows:

A) Data Collection

    A.1) Density functional theory calculations

    A.2) Structural descriptors

B) Data preparation

    B.1) Feature selection

    B.2) Cluster analysis

C) Bayesian neural network

D) Statistical analysis.

A. Data Collection

1. Density functional theory calculations

Each element of the FULL SET is uniquely identified by a set of descriptors, and therefore each bilayer can be represented as a point in this hyperspace. For a defined metric, structural similarities are clustered. Here, we selected 282 and 226 structures to calculate the interlayer energy and the C_{33}, respectively, that approximately represent each cluster and span over the whole hyperspace of the descriptor.

To calculate the properties of the bilayers by means of DFT, we used VASP within the GGA-PBE approximation where a Tkatchenko-Scheffler van der Waals correlation correction was applied.[52–54] A k-point space of $8\times8\times1$ for structures with atoms less than 10, and $3\times3\times1$ otherwise, and an energy cut-off is 520 eV. The energy minimization tolerance is 10^{-6} eV, and the force tolerance is 10^{-2} eV/Å. We use a dipole correction that falls like $1/L^3$ to correct unphysical interaction along the z direction. We calculated the interlayer energy as the difference between the total energy of the individual monolayers and the total energy of the bilayer, where a negative energy indicates attractive interaction, then normalize this quantity per unit area. For each bilayer we select the minimum energy structure among the possible configurations resulting from different monolayer orientations. The supercell size along the z-axis was chosen to be large enough to avoid interactions with replica of the layer in the periodic boundary conditions.[39] We want to point out that the characterization hetero-structures indicate how they are typically assembled in a A/B/A/B bulk-like structure and therefore, the A/B stacking configuration in vacuum does not seem to resemble a realistic scenario.[32,55] However, we calculated the IE in vacuum because a large distance between the bilayers may significantly minimize the fictitious interactions between replicas along the z-directions.[39] Then, we use the relevant information to build a bulk-like structure, where an elastic deformation of the cell that resemble realistic conditions allows us to extract the C_{33} values.[56] To validate the use two approaches for IE and C_{33}, we repeat the calculation of IE in a bulk-like cell, and we verify that the difference between the two values is less then 0.7%, which is relatively small and within average error for the values calculated by ML.

To calculate the value of the elastic constant, we consider its value along the z-axis of a bilayer (i.e. C_{33}). Due to the different forces acting in-plane and out-of-plane, we can approximate the hardness with the C_{33}.[10,11,24,25] We calculate the C_{33} by interpolating the interlayer energy change as a function of the interlayer distance, obtained by varying the supercell size along the z-axis. Here, the supercell dimension at the equilibrium along the z-axis doubles the equilibrium distance of the bilayers previously calculated, to resemble a two-component multi-layered structure, as schematically shown in the Fig.S2. When varying the supercell size, we remain in the elastic domain of the solid and a quadratic dependence of the total energy with respect to the strain is expected.[56] We calculated the error of 0.7% in the IE to correspond to an error of 1.1% in the C_{33}.

2. Structural descriptors

We obtained structural information for 6,138 monolayers from an online database (https://2dmatpedia.org/). The 6,138 monolayers were obtained from a large number of inorganic bulk structures available on the online database Materials Project (https://materialsproject.org/) by using a "top-down" approach where the bulk crystals are screened

for layered structures which are then theoretically exfoliated to 2D monolayers, and a "bottom-up" approach, in which elemental substitution is systematically applied to the unitary and binary 2D materials obtained from the top-down approach. To obtain feature vector for materials (X) we adapted a method we developed previously, which allows us to calculate materials fragment descriptors computed from the connectivity graph inside the unit cell.[38] Within this approach, each crystal structure is represented as a graph, with vertices decorated according to the reference properties of the atoms they represent, and each node is connected to its neighbour according to the Voronoi tessellation. The adjacency matrix of this graph determines the global topology for a given system, including interatomic bonds and contacts within a crystal. The final descriptor vector for the machine learning model is obtained by partitioning a full graph into subgraphs called fragments. Descriptors for each bilayer were obtained by adding the values of the descriptors for the two monolayers, as described by Tawfik *et al.*[39]

B. Data preparation

1. Features selection

The number of descriptors calculated for our structures is too large to be used for our calculations, as overfitting will occur. Feature selection involves choosing a subset of d features from a set of D features based using some optimization criterion, creating a more compact descriptor space X with as little performance loss as possible. The features removed should therefore be largely irrelevant for the calculation of a specific target property. For this purpose, we use a combination of Genetic Algorithm (GA) search and LASSO regression.[57] The idea of GA is to generate some random possible solutions, which represent different variables, to then combine the best solutions in an iterative process. The GA process tries to maximise a fitness function, that in our case is the LASSO function. We further screen the features using a LASSO regression analysis. The goal of LASSO regression is to obtain the subset of descriptors (X) that minimizes prediction error for a quantitative response variable (Y). The LASSO does this by imposing a constraint on the model parameters that causes regression coefficients for some variables to drop to zero. Variables with a regression coefficient equal to zero after the shrinkage process are excluded from the model. Variables with non-zero regression coefficients variables are most strongly associated with the response variable. The goal of the algorithm is to minimize:

$$L = \sum_{i=1}^{n}(y_i - \sum x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p}|\beta_j| \tag{1}$$

Where the tuning parameter, λ controls the strength of the penalty. Therefore, λ control the degree of elimination: When $\lambda = 0$, no parameters are eliminated. The estimate is equal to the one found with linear regression. As λ increases, more and more coefficients are set to zero and eliminated (theoretically, when $\lambda = \infty$, all coefficients are eliminated). As λ increases, bias increases. As λ decreases, variance increases.

2. *Cluster Analysis*

The choice of representative structures to be used to train our model is crucial for the quality of predictions. K-means is a method of vector quantization that is popular for cluster analysis in data mining.[58] K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, acting as a representative model of the cluster. This results in a partitioning of the data space into Voronoi cells. The best number of clusters k leading to the largest distance is not known a priori and must be computed from the data. The objective of K-means clustering is to minimize total intra-cluster variance, or, the squared error function:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} ||x_j^i - c_j||^2 \quad .$$

(2)

This procedure will maximize the diversity of structures assigned to training and test sets while ensuring that the test sets are still within the domain of the models. Here, we use the silhouette score to express the quality of the clustering. The silhouette score, with values between $+1$ and -1, is a measure to indicate how close each point in one cluster is to points in the neighbouring clusters, where $+1$ indicates that the sample is far away from the neighbouring clusters. A K-means analysis gives the best score when three sub-groups are formed for interlayer energy, with an average silhouette score of 0.78, and five for C_{33} silhouette score of 0.63. The training set contains 75% of the data and the test set 25%.

C. Bayesian neural networks

In the present work we use machine learning in a Bayesian framework in order to predict not only the transfer function and the property of a large number of structures, but also to give the confidence interval for each value.[59]

In the Bayesian point of view, regressions are formulated using probability distributions rather than point estimates. The target property or response, Y, is not estimated as a single value, but is assumed to be drawn from a probability distribution. The aim of Bayesian regressions is not to find the single "best" value of the model parameters, but rather to determine the posterior distribution for the model parameters.[60–63] Not only is the response generated from a probability distribution, but the model parameters are assumed to come from a distribution as well. The posterior probability of the model parameters is conditional upon the train inputs and outputs:

$$P(\beta|y, X) = \frac{P(y|\beta, X) \times P(\beta|X)}{P(y|X)} \quad ,$$

(3)

Here, P(β | y, X) is the posterior probability distribution of the model parameters given the inputs and outputs. This is equal to the likelihood of the data, P(y | β, X), multiplied by the prior probability of the parameters and divided by a normalization constant.

Here, we have a posterior distribution for the model parameters that is proportional to the likelihood of the data multiplied by the prior probability of the parameters. We can observe two primary benefits of Bayesian regressions. Priors: parameters distributions are included in the model. If these are unknown, we can use non-informative priors for the parameters such as a normal distribution. Posterior: the result of performing Bayesian regression is a distribution

of possible model parameters based on the data and the prior. This allows us to quantify our uncertainty about the model.

To implement this methodology, we use a dropout approach, which can be seen as a Bayesian approximation of a well known probabilistic model. Dropout is used in many models in deep learning as a way to avoid over-fitting by randomly creating deviations from the optimizaton pathway. In our implementation dropout approximately integrates over the weights in the model.[64–66] Basically, this is comparable to performing a number of stochastic passes through the network, and then averaging the results. This result has been presented in the literature before as model averaging. With dropout, we sample binary variables for every input point and for every network unit in each layer. Each binary variable takes value 1 with probability p_i for layer i. A unit is dropped to zero for a given input if its corresponding binary variable takes value 0. We use the same values in the backward pass propagating the derivatives to the parameters, obtaining a distribution over each descriptor, from which we can extrapolate the uncertainty associated to each prediction.

Although several approaches have been proposed to fix the number of nodes and hidden neurons in neural network, no general rule can be identified, and typically starting from basic consideration on the complexity of the problem, the architecture of the neural network is modified adding or removing neurons or layers until the NN gives desired values of a loss function.[67,68] Here, we follow the same approach and we tested several BNN configurations that includes from 64 to 254 nodes, including either one or two hidden layers, and the quality of the NN architecture is determined by the root mean square error (RMSE) and the Mean Absolute Error (MAE), and R^2.

We use a BNN with 2 hidden layers composed of 128 neurons each, with a ReLU transfer function in the hidden nodes, where the dropout probability is 0.1. The dropout creates a distribution over the calculated response, which is then averaged over 600 trial networks giving the response value and the associate standard deviation. In the SI we report the IE and C_{33} values, with relative error-bar, of the train and test set after BNN optimization.

D. Statistical analysis

To investigate correlations between the properties, we calculated the Pearson product moment correlation, which evaluates the linear relationship between two continuous variables, and Spearman rank-order correlation, which evaluates the monotonic relationship between two continuous or ordinal variables, as follows:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2(y_i - \bar{y})^2}} \quad , \tag{4}$$

where x_i and y_i are the two variables, and

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \quad , \tag{5}$$

where $d =$ is the pairwise distances of the ranks of the variables x_i and y_i, and $n =$ is the number of samples. Small coefficients indicate lack of correlation between two variables.

IV.   DATA AVAILABILITY

The complete IE, C_{33} and thermal stability data is available online (http://doi.org/10.26195/5dd36650d7e1e).

V.   CODE AVAILABILITY

Custom Python codes for data preprocessing and Bayesian neural network training and data extrapolation are available online (https://github.com/fronzi/projBNN).

VI.   COMPETING INTERESTS

The authors declare no competing interests.

VII.   ACKNOWLEDGEMENTS

* Corresponding author

  E-mail: marco.fronzi@uts.edu.au

† Corresponding author

  E-mail: Mike.Ford@uts.edu.au

[1] H. E. Sliney, ASLE Trans. **21**, 109 (1978).

[2] A. E. Filippov, M. Dienwiebel, J. W. M. Frenken, J. Kla, and M. Urbakh, Phys. Rev. Lett. **100**, 046102 (2008).

[3] K. S. Novoselov, A. Mishchenko, A. Carvalho, and A. H. C. Neto, Science **353**, 9439 (2016).

[4] A. K. Geim and I. V. Grigorieva, Nature **499**, 419 (2013).

[5] J. Wang, L. Li, Z. Shen, P. Guo, M. Li, B. Zhao, L. Fang, and L. Yang, Materials **11**, 2462 (2018).

[6] D. Moore and D. A. Lockner, J. Geoph. Res. **109**, B03401 (2004).

[7] R. Zhang and R. Cheung, *Two-dimensional Materials - Synthesis, Characterization and Potential Applications* (IntechOpen, 2016).

[8] T. Filleter, J. L. McChesney, A. Bostwick, E. Rotenberg, K. V. Emtsev, T. Seyller, K. Horn, and R. Bennewitz, Phys. Rev. Lett. **102**, 086102 (2009).

[9] P. Sheehan and C. M. Lieber, Science , 1158 (1996).

[10] M. L. Oyen, Journal of Biomechanics **39:14**, 2699 (2006).

[11] K. L. Johnson, *Contact Mechanics* (Cambridge University Press, 1987).

[12] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. Persson, APL Materials **1**, 011002 (2013).

[13] J. Zhou, L. Shen, M. D. Costa, K. A. Persson, S. P. Ong, P. Huck, Y. Lu, X. Ma, Y. Chen, H. Tang, and Y. P. F. Scientific, Scientific Data **6**, 86 (2019).

[14] L. A. Ponomarenko, A. K. Geim, A. A. Zhukov, R. Jalil, S. V. Morozov, K. S. Novoselov, I. V. Grigorieva, E. H. Hill, V. V. Cheianov, V. I. Falḱo, T. T. K. Watanabe, and R. V. Gorbachev, Nat. Phys. **7**, 958961 (2011).

[15] L. Britnell, R. V. Gorbachev, R. Jalil, B. D. Belle, F. Schedin, A. Mishchenko, T. Georgiou, M. I. Katsnelson, L. E. abd S. V. Morozov, N. M. Peres, J. Leist, A. K. Geim, and K. S. N. L. A. Ponomarenko, Science **335**, 947950 (2012).

[16] S. J. Haigh, A. Gholinia, R. Jalil, S. Romani, L. Britnell, D. C. Elias, K. S. Novoselov, L. A. Ponomarenko, A. K. Geim, and R. Gorbachev, Nat. Mater. **11**, 764767 (2012).

[17] T. Georgiou, R. Jalil, B. D. Belle, L. Britnell, R. V. Gorbachevn, S. V. Morozov, Y.-J. Kim, A. Gholinia, S. J. Haigh, O. Makarovsky, L. Eaves, L. A. Ponomarenko, A. K. Geim, K. S. Novoselov, and A. Mishchenko, Nat. Nanotechnol. **8** (2013).

[18] L. Wang, I. Meric, P. Y. Huang, Q. Gao, Y. Gao, H. Tran, T. Taniguchi, K. Watanabe, L. M. Campos, D. A. Muller, J. Guo, P. Kim, J. Hone, K. L. Shepard, and C. R. Dean, Science **342**, 614 (2013).

[19] S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).

[20] D. Bowler and T. Miyazaki, Rep. Prog. Phys. **75**, 036503 (2012).

[21] M. Correa, C. Bielza, and J. Pamies-Teixeirac, Expert Syst. Appl. **36**, 7270 (2009).

[22] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, Nature **559**, 547 (2018).

[23] H. K. D. H. Bhadeshia, ISIJ International **39**, 966 (1999).

[24] S. Bertolazzi, J. Brivio, and A. Kis, ACS Nano **5**, 9703 (2011).

[25] K. Liu and J. Wu, J. Mat. Res. **31**, 832 (2016).

[26] N. Lu, H. Guo, Z. Zhuo, L. Wang, X. Wu, and X. C. Zeng, Nanoscale **9:48**, 19131?19138 (2017).

[27] N. Lu, H. Guo, Z. Zhuo, L. Wang, X. Wu, and X. C. Zeng, Nanoscale **9**, 19131 (2017).

[28] T. Björkman, A. Gulans, A. V. Krasheninnikov, and R. M. Nieminen, Phys. Rev. Lett. **108**, 235502 (2012).

[29] C. R. C. Rêgo, L. N. Oliveira, P. Tereshchukand, and J. L. F. Da Silva, J. Phys.: Condens. Matter **27**, 415502 (2015).

[30] T. Björkman, A. Gulans, A. V. Krasheninnikov, and R. M. Nieminen, J. Phys.: Condens. Matt. **24**, 424218 (2012).

[31] S. L. S, J. Harl, T. Gould, J. G. Angyán, G. Kresse, and J. F. Dobson, Phys. Rev. Lett. **105**, 196401 (2010).

[32] Z. Liu, J. Z. Liu, Y. Cheng, and Q. Zheng, "Interlayer binding energy of graphite – a direct experimental determination," (2011), arXiv:1104.1469 [cond-mat.mes-hall].

[33] Z. Liu, J. Z. Liu, Y. Cheng, Z. Li, L. Wang, and Q. Zheng, Physical Review B **85** (2012), 10.1103/physrevb.85.205418.

[34] R. Zacharia, H. Ulbricht, and T. Hertel, Phys. Rev. B **69**, 155406 (2004).

[35] G. Graziano, J. Klimeš, F. Fernandez-Alonso, and A. Michaelides, J. Phys.: Condens. Matt. **24**, 424216 (2012).

[36] L. Wang, X. Zhou, T. Ma, D. Liu, L. Gao, X. Li, and J. Luo, Nanoscale **9**, 10846 (2017).

[37] H. Rydberg, M. Dion, N. Jacobson, E. Schröder, P. Hyldgaard, S. I. Simak, D. C. Langreth, and B. I. Lundqvist, Phys. Rev. Lett. **91**, 126402 (2003).

[38] O. Isayev, Nat. Comm. **8**, 15679 (2017).

[39] S. A. Tawfik, O. Isayev, C. Stampfl, J. Shapter, D. A. Winkler, and M. J. Ford, Adv. Theory Simul. **2**, 1800128 (2019).

[40] J. Cai, J. Luo, S. Wang, and S. Yang, Neurocomputing **300**, 70 (2018).

[41] S. Tamleh, G. Rezaei, and J. Jalilian, Phys. Lett. A **382**, 339 (2018).

[42] Landolt-Börnstein, http://link.springer.de (2003).

[43] L. X. Benedict, N. G. Chopra, M. L. Cohen, A. Zettl, S. G. Louie, and V. H. Crespi, Chem. Phys. Lett. **286**, 490 (1998).

[44] L. Spanu, S. Sorella, and G. Galli, Phys. Rev. Lett. **103**, 196401 (2010).

[45] D. Alexander, A. Tropsha, and D. A. Winkler, J. Chem. Inf. Model. **55**, 1316 (2015).

[46] "Materialsweb," Available at `https://materialsweb.org`.

[47] "http://2dmatpedia.org," Available at `http://http://2dmatpedia.org`.

[48] W. Sun, S. T. Dacek, S. P. Ong, G. Hautier, A. Jain, W. D. Richards, A. C. Gamst, K. A. Persson, and G. Ceder, Sci. Adv. **2**, 1 (2016).

[49] W. Wang, S. Dai, X. Li, J. Yang, D. J. Srolovitz, and Q. Zheng, Nat. Comm. **6** (2015).

[50] M. Ashton, R. G. Hennig, S. R. Broderick, K. Rajan, and S. B. Sinnott, Phys. Rev. B **94**, 054116 (2016).

[51] M. Ashton, J. Paul, S. B. Sinnott, and R. G. Hennig, Phys. Rev. Lett. **118**, 106101 (2017).

[52] G. Kresse and J. Hafner, J. Phys.: Condens. Matt. **6**, 8245 (1994).

[53] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).

[54] A. Tkatchenko and M. Scheffler, Phys. Rev. Lett. **102**, 073005 (2009).

[55] Q. Zheng and F. Z. Liu, Friction **2**, 182 (2014).

[56] P. Ravindran, L. Fast, P. A. Korzhavyi, and B. Johansson, J. Appl. Phys., Vol. 84, No. 9, 1 November 1998 **84:9**, 1 (1998).

[57] R. Tibshirani, Journal of the Royal Statistical Society **58**, 267 (1996).

[58] E. W. Forgy, Biometrics **21**, 768 (1965).

[59] S. Bergmann, S. Stelzer, and S. Strassburger, Journal of Simulation **8**, 76 (2014).

[60] D. J. C. MacKay, Comput. Neural Sys. **5**, 469 (1995).

[61] D. J. C. MacKay, Neural Comput. **4**, 415 (1992).

[62] D. J. C. MacKay, Neural Comput. **4**, 448 (1992).

[63] F. R. Burden and D. A. Winkler, J. Med. Chem. **42**, 3183 (1999).

[64] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," (2015), arXiv:1506.02142 [stat.ML].

[65] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Journal of Machine Learning Research **15**, 1929 (2014).

[66] D. Tran, M. W. Dusenberry, M. van der Wilk, and D. Hafner, "Bayesian layers: A module for neural network uncertainty," (2018), arXiv:1812.03973 [cs.LG].

[67] J. Sun, Neurocomputing **79**, 158163 (2012).

[68] Y. Lan, Y. C. Soh, and G. B. Huang, Neurocomputing **73**, 3191 (2010).