A force-matching Stillinger-Weber potential for MoS_2: Parameterization and Fisher information theory based sensitivity analysis

Mingjian Wen,[1] Sharmila N. Shirodkar,[2,a)] Petr Plecháč,[3] Efthimios Kaxiras,[2,4] Ryan S. Elliott,[1] and Ellad B. Tadmor[1,b)]
[1]*Department of Aerospace Engineering and Mechanics, University of Minnesota, Minneapolis, Minnesota 55455, USA*
[2]*John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*
[3]*Department of Mathematical Sciences, University of Delaware, Newark, Delaware 19716, USA*
[4]*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*

(Received 3 October 2017; accepted 2 December 2017; published online 22 December 2017)

Two-dimensional molybdenum disulfide (MoS_2) is a promising material for the next generation of switchable transistors and photodetectors. In order to perform large-scale molecular simulations of the mechanical and thermal behavior of MoS_2-based devices, an accurate interatomic potential is required. To this end, we have developed a Stillinger-Weber potential for monolayer MoS_2. The potential parameters are optimized to reproduce the geometry (bond lengths and bond angles) of MoS_2 in its equilibrium state and to match as closely as possible the forces acting on the atoms along a dynamical trajectory obtained from *ab initio* molecular dynamics. Verification calculations indicate that the new potential accurately predicts important material properties including the strain dependence of the cohesive energy, the elastic constants, and the linear thermal expansion coefficient. The uncertainty in the potential parameters is determined using a Fisher information theory analysis. It is found that the parameters are fully identified, and none are redundant. In addition, the Fisher information matrix provides uncertainty bounds for predictions of the potential for new properties. As an example, bounds on the average vibrational thickness of a MoS_2 monolayer at finite temperature are computed and found to be consistent with the results from a molecular dynamics simulation. The new potential is available through the OpenKIM interatomic potential repository at https://openkim.org/cite/MO_201919462778_000. *Published by AIP Publishing.*
https://doi.org/10.1063/1.5007842

I. INTRODUCTION

Since the discovery of graphene,[1] two-dimensional (2D) nanostructures have attracted significant interest for potential applications in areas including electronics, sensors, and photonics due to their astounding electronic,[2–5] optical,[6,7] thermal,[8,9] and mechanical[10–12] properties. Monolayer MoS_2 is a member of the transition metal dichalcogenides family of 2D materials whose bulk counterpart has long been used for dry lubrication and as a catalyst in industry.[11,13] Monolayer MoS_2 is a direct gap semiconductor with a bandgap of 1.8 eV,[5] making it an ideal candidate for electronic and optoelectronic applications where a nonzero band gap is mandatory.

The most accurate methods for simulating the response of MoS_2 are first principles (*ab initio*) approaches, which involve solving the Schrödinger equation for the quantum system of nuclei and electrons. However, such methods are computationally very expensive and are typically limited to systems consisting of at most a few hundred atoms due to hardware and algorithmic limitations. Empirical interatomic potentials provide an alternative approach. Even complex interatomic potentials are far less expensive to compute than a first principles approach, which enables simulations of problems that are inaccessible to quantum calculations, such as fracture and atomic diffusion.[14–16]

To date, several empirical potentials for MoS_2 have been proposed. The earliest published in 1975 is a valence force field (VFF) model by Wakabayashi *et al.*[17] in which the potential energy was decomposed into harmonic components. The interlayer interaction was assumed to be due to an axially symmetric force between sulfur atoms of neighboring layers, and the intralayer interaction was assumed to be associated with the stretching and bending of Mo-S bonds. The potential parameters were optimized to reproduce the phonon spectrum obtained from inelastic neutron scattering. Liang *et al.*[18] developed a second-generation reactive empirical bond-order (REBO) potential for the Mo-S system using the master formula underlying the Abell,[19] Tersoff,[20–22] and REBO[23] potentials with an additional Lennard-Jones[24,25] potential to describe the weak interlayer van der Waals interactions. This potential was fit to a training set of the energy, bond length, and bond stiffness of Mo-Mo, Mo-S, and S-S systems with the main objective to reproduce the structural and elastic properties of MoS_2. Jiang *et al.*[26,27] developed two Stillinger-Weber[28] (SW) potentials for monolayer MoS_2. The first one[26] considered all available two-body and three-body interactions in monolayer MoS_2 and was fit to the same phonon spectrum used in the VFF model.[17] In the second parameterization,[27]

a)Present address: Material Science and Nanoengineering, Rice University, Houston, Texas 77005, USA.
b)Author to whom correspondence should be addressed: tadmor@umn.edu

Mo-Mo and S-S two-body interactions were neglected, and the potential was fit to bond lengths and bond angles from experiments and first principles calculations, and to energies predicted by the VFF model.[17] Very recently, a reactive force field (ReaxFF) potential was developed by Ostadhossein *et al.*[29] to study energetics and reaction mechanisms in single- and multi-layer MoS_2. The ReaxFF potential was fit to a training set of energies, geometries, and charges derived from first principles density functional theory (DFT) calculations for both clusters and periodic systems.

An important characteristic of an empirical potential is its *transferability*, i.e., its ability to accurately predict behavior that it was not fit to reproduce. Experience has shown that transferability tends to be improved when the training set is extended beyond a small set of selected material properties. In the *force-matching method* proposed by Ercolessi and Adams,[30] this is done by fitting the potential to the forces acting on the atoms along a dynamical trajectory obtained from *ab initio* molecular dynamics (AIMD). Transferability is likely enhanced by this procedure since the potential is exposed to a much larger cross-section of configuration space. An additional advantage is that the issue of insufficient training data that may occur in traditional fitting procedures can be resolved since as many configurations as needed can be readily generated.

Recent developments in sensitivity analysis of stochastic systems based on relative entropy measures and Fisher information theory[31–33] have led to a deeper understanding of the force-matching methodology. It is now recognized that force matching is equivalent to relative entropy minimization provided that the training set of forces is obtained from a trajectory that samples the appropriate distribution function.[34] This improves the transferability of the potential since it can be shown that minimizing relative entropy also bounds the uncertainty in predictions of other observables.[35] The statistical mechanics approach to force matching was originally studied for equilibrium conditions[36,37] and later extended to nonequilibrium steady states.[35] This generalization allows for the treatment of driven systems subject to external conditions, such as thermal gradients and deformations.

In this paper, we apply an information-theoretic based force-matching approach to retrain the SW potential of Jiang *et al.*[26,27] We find that this significantly improves the accuracy of the potential for a variety of properties. In addition, the information theory analysis yields (1) the uncertainty in the fitting parameters (i.e., the confidence with which the parameters are determined from a given training set); and (2) the sensitivity of the potential's predictions on its parameters (i.e., how variations in the parameters affect the results).[38,39]

The paper is structured as follows: In Sec. II, we introduce the SW potential for MoS_2. This is followed by a description of the parameterization process in Sec. III. In Sec. IV, we present the fitting results and test the potential predictions for several properties of interest. An information-based sensitivity analysis is carried out in Sec. V. We summarize our results in Sec. VI.

II. STILLINGER-WEBER POTENTIAL

The SW potential was originally introduced to model bulk silicon.[28] The innovation in this potential was the inclusion of a three-body term to penalize configurations away from the tetrahedral ground state structure of Si. The potential was later extended to other tetrahedral material systems including Ge,[40] III-V compound semiconductors,[41] and compounds of the major II-VI elements Zn, Cd, Hg, S, Se, and Te.[42] It has been adapted for monolayer MoS_2 and monolayer black phosphorus that do not have a tetrahedral structure.[26,27]

The total SW potential energy $\mathcal{V}$ of a system consisting of N atoms is

$$\mathcal{V} = \sum_{i=1}^{N}\sum_{j>i}^{N}\phi_2(r_{ij}) + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{\substack{k>j\\k\neq i}}^{N}\phi_3(r_{ij}, r_{ik}, \beta_{jik}), \tag{1}$$

where the two-body interaction takes the form

$$\phi_2(r_{ij}) = \epsilon_{IJ}\hat{A}_{IJ}\left[B_{IJ}\left(\frac{r_{ij}}{\sigma_{IJ}}\right)^{-p_{IJ}} - \left(\frac{r_{ij}}{\sigma_{IJ}}\right)^{-q_{IJ}}\right] \times \exp\left(\frac{1}{r_{ij}/\sigma_{IJ} - a_{IJ}}\right), \tag{2}$$

and the three-body term is

$$\phi_3(r_{ij}, r_{ik}, \beta_{jik}) = \epsilon_{JIK}\hat{\lambda}_{JIK}\left[\cos\beta_{jik} - \cos\beta^0_{jik}\right]^2 \times \exp\left(\frac{\hat{\gamma}_{IJ}}{r_{ij}/\sigma_{IJ} - a_{IJ}} + \frac{\hat{\gamma}_{IK}}{r_{ik}/\sigma_{IK} - a_{IK}}\right), \tag{3}$$

in which $r_{ij} = \|\boldsymbol{r}_i - \boldsymbol{r}_j\|$ is the bond length between atoms i and j, β_{jik} is the bond angle formed by bonds i–j and i–k with the vertex at atom i, and β^0_{jik} is the corresponding predetermined reference angle. The potential parameters are $\epsilon, \hat{A}, B, p, q, \sigma, a, \hat{\lambda}, \hat{\gamma}$. Both the two-body and three-body terms are designed to be identically zero at the cutoff radius $r^{\text{cut}} = a\sigma$. The parameters depend on the species of the interacting atoms, which are indicated by uppercase subscripts. For example, ϵ_{IJ} is the parameter ϵ for the pairwise interaction between atom i of species I and atom j of species J.

Equations (2) and (3) can be recast in a form in which all parameters are independent and the dependence on the cutoff radius is made explicit. We define $A_{IJ} = \epsilon_{IJ}\hat{A}_{IJ}$, $\hat{\lambda}_{JIK} = \epsilon_{JIK}\hat{\lambda}_{JIK}$, $\gamma_{IJ} = \sigma_{IJ}\hat{\gamma}_{IJ}$, and $r^{\text{cut}}_{IJ} = a_{IJ}\sigma_{IJ}$, then

$$\phi_2(r_{ij}) = A_{IJ}\left[B_{IJ}\left(\frac{r_{ij}}{\sigma_{IJ}}\right)^{-p_{IJ}} - \left(\frac{r_{ij}}{\sigma_{IJ}}\right)^{-q_{IJ}}\right] \times \exp\left(\frac{\sigma_{IJ}}{r_{ij} - r^{\text{cut}}_{IJ}}\right), \tag{4a}$$

$$\phi_3(r_{ij}, r_{ik}, \beta_{jik}) = \lambda_{JIK}\left[\cos\beta_{jik} - \cos\beta^0_{jik}\right]^2 \times \exp\left(\frac{\gamma_{IJ}}{r_{ij} - r^{\text{cut}}_{IJ}} + \frac{\gamma_{IK}}{r_{ik} - r^{\text{cut}}_{IK}}\right). \tag{4b}$$

The new parameters are $A, B, p, q, \sigma, \lambda, \gamma$ along with the cutoff radii and equilibrium angles. Note that when $r > r^{\text{cut}}$,

both ϕ_2 and ϕ_3 vanish. For MoS_2, we add an additional cutoff $r_{JK}^{\text{cut}*}$ for bond j–k in ϕ_3, that is, ϕ_3 vanishes as well when $r_{jk} > r_{JK}^{\text{cut}*}$. This will be explained in Sec. II A.

Based on the work of Jiang *et al.*,[26,27] two-body bond stretching (or compression) is considered for three types of interaction, i.e., $IJ \in \{\text{Mo}-\text{Mo}, \text{Mo}-\text{S}, \text{S}-\text{S}\}$ in (4a). For three-body bond bending, only interactions of type S-Mo-S (Mo is the species of the vertex atom) and Mo-S-Mo (S is the species of the vertex atom) are considered, i.e., in (4b) $JIK \in \{\text{S}-\text{Mo}-\text{S}, \text{Mo}-\text{S}-\text{Mo}\}$. Consequently, there are only two λ parameters ($\lambda_{\text{S-Mo-S}}$ and $\lambda_{\text{Mo-S-Mo}}$) and a single γ parameter ($\gamma = \gamma_{\text{Mo-S}} = \gamma_{\text{S-Mo}}$). We will denote the set of all parameters as $\boldsymbol{\theta}$ in the following discussion.

A. Cutoffs and bond angles

The crystal structure of monolayer MoS_2 is shown in Fig. 1. It consists of a monatomic Mo plane sandwiched between two monatomic S planes. Mo and S atoms occupy alternating corners of a hexagon to form a honeycomb structure. A unit cell, the green shaded region in Fig. 1(a), consists of one Mo atom and two S atoms. The in-plane zero-temperature equilibrium lattice constant of the relaxed structure obtained using the first principles code SIESTA[43] is $a_0 = 3.20$ Å, and the vertical separation between S layers is $b_0 = 3.19$ Å. Each Mo atom is surrounded by six first-nearest-neighbor S atoms and each S atom is connected to three first-nearest-neighbor Mo atoms.

The cutoff r_{IJ}^{cut} is set to the second-nearest neighbor distance of the corresponding IJ species. As an example, consider the calculation of $r_{\text{S-S}}^{\text{cut}}$. Referring to Fig. 1(b), the nearest neighbors of atom S2 are S1, S3, and S5. In fact $d_{\text{S2-S1}} = d_{\text{S2-S3}} = a$ is slightly larger than $d_{\text{S2-S5}} = b$; however, we ignore this small difference and treat all these atoms as the first neighbor ring for the purpose of determining the cutoff and the parameter $\sigma_{\text{S-S}}$ (as explained below). The second neighbors of S2 are atoms S4 and S6. Therefore, $r_{\text{S-S}}^{\text{cut}} = d_{\text{S2-S4}} = \sqrt{a^2 + b^2} = 4.51956$ Å. The other two cutoffs are determined in a similar fashion: $r_{\text{Mo-Mo}}^{\text{cut}} = \sqrt{3}a = 5.54660$ Å and $r_{\text{Mo-S}}^{\text{cut}} = \sqrt{4a^2/3 + b^2/4} = 4.02692$ Å.

As pointed out by Zhou *et al.*,[42] the SW two-body and three-body functions decay to close to zero at a distance smaller than the cutoff radius due to the presence of the exponential terms. This is demonstrated in Fig. 2 for the $\phi_2(r)$

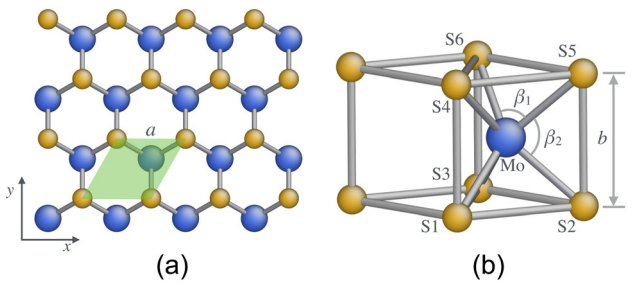

FIG. 1. The crystal structure of monolayer MoS_2. (a) Top view, where the green shaded region depicts a unit cell. (b) Oblique view of the shaded unit cell in (a). Each Mo atom is surrounded by six first-nearest-neighbor S atoms and each S atom is connected to three first-nearest-neighbor Mo atoms. Images rendered with AtomEye.[44]

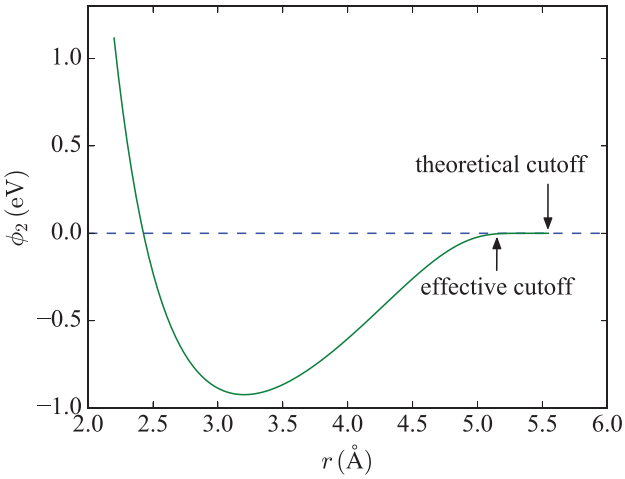

FIG. 2. Pair function $\phi_2(r)$ of the SW potential, whose effective cutoff is much smaller than the theoretical one. The parameters used in the plot are $A = 4.15$ eV, $B = 0.44$, $p = 5$, $q = 0$, $\sigma = 2.85$ Å, and $r^{\text{cut}} = 5.55$ Å.

function for the sample parameters listed in the caption. It is clear that while the theoretical cutoff is 5.55 Å, the potential energy becomes negligibly small beyond an *effective* cutoff of about $r = 5.10$ Å. If desired, this characteristic can be employed to expedite molecular dynamics (MD) simulations by using the effective cutoff rather than the theoretical one to compute neighbor lists.[42]

In the three-body term ϕ_3, an additional cutoff is employed to exclude certain interactions. Each Mo atom is surrounded by six first-nearest-neighbor S atoms, resulting in three different types of S-Mo-S angles after accounting for symmetry [β_1, β_2, and ∠S2-Mo-S4 in Fig. 1(b)]. While β_1 and β_2 are almost of the same value, ∠S2-Mo-S4 is much larger. Because (4b) only allows for one equilibrium angle, it is desirable to exclude large-angle interactions of the third type so that the equilibrium structure of MoS_2 can be correctly described. (Note that this is not a concern for Mo-S-Mo interactions, since there is only one type of Mo-S-Mo bond with a bond angle equal to β_1.) Following Refs. 26 and 27, and the GULP package,[45,46] in addition to the two cutoffs r_{IJ}^{cut} and r_{IK}^{cut} included in (4b), a new cutoff $r_{JK}^{\text{cut}*}$ is applied to r_{jk} when J and K are S atoms. For $r_{jk} > r_{JK}^{\text{cut}*}$, the 3-body term involving atoms j, i, and k is ignored. We take the additional cutoff to be $r_{\text{S-S}}^{\text{cut}*} = 3.86095$ Å corresponding to the average of the first- and second-nearest-neighbor distances of S-S bonds. This cutoff allows for bond angle interactions of types β_1 and β_2, but ∠S2-Mo-S4 type interactions will be excluded. We note that this introduces a discontinuity in the potential energy since the 3-body term is abruptly removed at $r_{jk} = r_{JK}^{\text{cut}*}$. The maximum discontinuity can be 2.67 eV, but this occurs when atom i is located in the middle between atoms j and k, which is far from the equilibrium structure. As long as the system is not subjected to extreme deformations far from the equilibrium ground state, the discontinuity will be mild if encountered and should not adversely affect molecular simulations.

Given $a_0 = 3.20$ Å and $b_0 = 3.19$ Å, it is straightforward to show that the angles in Fig. 1(b) are $\beta_1 = 81.92^\circ$ and $\beta_2 = 81.61^\circ$. Since the angles are quite close, we choose to

use the same β^0 as the reference angle for both S-Mo-S and Mo-S-Mo three-body interactions. We set $\beta^0 = 81.79°$, which is the value of both β_1 and β_2 if a_0 and b_0 are equal.

B. Predetermined parameters and constraints

Aside from the cutoff radii specified in Sec. II A, the SW potential for MoS_2 has 18 parameters: three values each for A, B, p, q, σ, two for λ, and one for γ. It is non-trivial to fit so many parameters at once given that the potential is highly nonlinear. To facilitate the fitting process and make it more robust, some parameters are determined *a priori* and the values of some others are constrained.

In other parameterizations of the SW potential,[26–28,42] the exponents q and p were taken to be 0 and 4, respectively. Here, we take $q = 0$, but allow p to be a fitting parameter that can only take on integer values.

In the original SW potential for Silicon,[28] Stillinger and Weber determined σ by requiring $r_m = 2^{1/6}\sigma$, where r_m is the distance at which $\phi_2(r)$ reaches its minimum.[47] In this work, σ is obtained in the same way. Given the lattice constants of the relaxed MoS_2 structure, $a_0 = 3.20$ Å and $b_0 = 3.19$ Å, the equilibrium bond lengths can be computed as $d_{\mathrm{Mo-Mo}} = 3.20$ Å, $d_{\mathrm{Mo-S}} = 2.44$ Å, and $d_{\mathrm{S-S}} = 3.19$ Å. Thus, we have $\sigma_{\mathrm{Mo-Mo}} = 2.85295$ Å, $\sigma_{\mathrm{Mo-S}} = 2.17517$ Å, and $\sigma_{\mathrm{S-S}} = 2.84133$ Å.

As in Ref. 27, we require that in the ground state structure all bonds are at their equilibrium lengths and all angles are at their equilibrium values, i.e., $(\partial\phi_2/\partial r)|_{r=d} = 0$ and $(\partial\phi_3/\partial\beta)|_{\beta=\beta^0} = 0$. The latter is satisfied automatically, and the former leads to a constraint relating B, p, q and σ

$$B = \frac{q(d/\sigma)^{-1-q}(d - r^{\mathrm{cut}})^2 + (d/\sigma)^{-q}\sigma^2}{p(d/\sigma)^{-1-p}(d - r^{\mathrm{cut}})^2 + (d/\sigma)^{-p}\sigma^2} = \frac{1}{pd^{-p-1}\sigma^{p-1}(d - r^{\mathrm{cut}})^2 + d^{-p}\sigma^p}, \quad (5)$$

where d is the equilibrium bond length computed above and in the last equality $q = 0$ was used.

Accounting for the preset parameters and applying the constraint in (5), the parameters left to be determined are $\boldsymbol{\theta} = \{A, p, \lambda, \gamma\}$. This is a small subset of all the parameters, which greatly helps with the fitting process described next.

III. PARAMETERIZATION

In the force-matching method,[30,35] a parameterized interatomic potential is fit to first principles forces for a training set of atomic configurations. If the configurations in the training set are obtained by sampling a thermodynamic ensemble (e.g., the canonical *NVT* ensemble or the isothermal-isobaric *NPT* ensemble), then the parameterization not only optimizes the forces but all observables that are defined as averages over the stationary distribution.[35] This significantly enhances the transferability of the potential. In this work, the training set is generated from a long thermostatted trajectory in the *NPT* ensemble from an AIMD simulation.

The force-matching method has been widely used since its introduction, and there are a number of open source implementations available. The *potfit* program originally developed by Brommer *et al.*[48–50] is widely used in solid-state physics. In Ref. 16, we adapted *potfit* to be compliant with the KIM Application Programming Interface (API) standard (part of the Knowledgebase of Interatomic Models Project,[51] which is an online suite of open source tools for molecular simulation of materials. https://openkim.org). This makes it possible to use *potfit* to fit any KIM potential model, not just those prebuilt in the *potfit* code. A difficulty associated with potential fitting is that interatomic potentials are nonlinear functions that are often "sloppy" in the sense that their predictions are insensitive to certain combinations of their parameters.[52] Transtrum *et al.*[53,54] augmented the Levenberg–Marquardt (LM) algorithm with a geodesic acceleration adjustment to improve convergence of sloppy models. The geodesic LM minimization algorithm was incorporated into KIM-potfit,[16] and in this work, it is used to optimize the potential parameters. In subsections III A and III B, we discuss the choice of training set and the cost function.

A. Training set

The training set trajectory was obtained by AIMD using the DFT code SIESTA.[43] The interactions between ionic cores and valence electrons were modeled by a double zeta polarized basis set and norm-conserving pseudopotential[55] constructed within the Troullier-Martins formalism.[56] The exchange-correlation energy of the electrons is treated within the Generalized Gradient Approximated (GGA) Perdew-Burke-Ernzerhof functional.[57] An energy cutoff of 90 Ry was used for the representation of charge density and potentials. Brillouin zone integration was carried out at a single k-point (Γ point).[58]

In the training set calculations, periodic boundary conditions were applied in all three directions with a vacuum of 30 Å perpendicular to the MoS_2 layer to minimize interactions between periodic images. The training set was constructed as follows. First, the equilibrium MoS_2 lattice structure was obtained by performing a full relaxation of a single unit cell allowing the cell to change its volume and shape until all stress components were less than 0.1 kBar and allowing the atoms to move until all forces were less than 0.03 eV/Å. The relaxed unit cell has an in-plane lattice constant of $a_0 = 3.20233$ Å and the separation between the two sulfur layers is $b_0 = 3.18928$ Å (see Fig. 1). Second, a rectangular block supercell was constructed with in-plane dimensions of 25.61 Å × 33.28 Å (corresponding to 8 × 12 relaxed unit cells) consisting of $N = 288$ atoms: 96 Mo atoms and 196 S atoms. Third, AIMD simulations were performed using the supercell under *NPT* conditions with a pressure of $p = 0$ and temperature of $T = 750$ K.[59] The atoms were initially assigned random velocities drawn from the Maxwell-Boltzmann distribution with the temperature equal to twice the target temperature. The system was then integrated in time for 3000 steps with a time step of $\Delta t = 0.7$ fs. The first 1000 steps were discarded to allow the system to equilibrate. In the subsequent 2000 steps, the atom coordinates $\boldsymbol{r}$ and the forces on the atoms $\boldsymbol{f}^0$ were recorded in the training set. Thus, the training set is $\{(\boldsymbol{r}_m, \boldsymbol{f}_m^0)\}_{m=1}^M$, where $M = 2000$, $\boldsymbol{r}_m \in \mathbb{R}^{3N}$, and $\boldsymbol{f}_m^0 \in \mathbb{R}^{3N}$.

B. Cost function

KIM-*potfit* defines a "cost function" quantifying the difference between the training set data and interatomic potential predictions and uses minimization algorithms to reduce the cost as much as possible. For a training set consisting of M configurations, the force-matching cost function is

$$C(\boldsymbol{\theta}) = \sum_{m=1}^{M} \frac{1}{2} w_m \|\boldsymbol{f}(\boldsymbol{r}_m;\boldsymbol{\theta}) - \boldsymbol{f}_m^0\|^2, \tag{6}$$

where $\boldsymbol{\theta}$ are the potential parameters and w_m is the weight for configuration m. We take all weights to be unity since all forces in a configuration and all configurations are equally important. The forces on the atoms in configuration m predicted by the potential are given by the negative gradient of the potential energy with respect to the atom coordinates, $\boldsymbol{f}(\boldsymbol{r}_m;\boldsymbol{\theta}) = (-\partial\mathcal{V}/\partial\boldsymbol{r})|_{\boldsymbol{r}_m}$, and $\boldsymbol{f}_m^0$ are the corresponding reference forces in the training set obtained from the AIMD trajectory. Note that $\boldsymbol{f}(\boldsymbol{r}_m;\boldsymbol{\theta}) \in \mathbb{R}^{3N}$ is the concatenated vector of the forces acting on the N atoms in the configuration.

The objective is to determine the optimal set of parameters $\boldsymbol{\theta}$ that minimizes the cost function C.

IV. RESULTS AND PREDICTIONS

The SW potential parameters were obtained by minimizing the cost function in (6) for the training set described in Sec. III A using the geodesic LM algorithm implemented in KIM-*potfit*.[16] The initial guesses of the parameters were taken from Ref. 26. The fitted parameters are listed in Tables I and II. We denote the new SW potential as SW-FM (Stillinger-Weber Force Matching) for later use in comparison.

To test the accuracy of the SW-FM potential, we computed the temperature dependence of the lattice constants and stiffness of MoS_2. These properties are important for the design of MoS_2 based electronic devices, since internal stress or strain due to thermal expansion can degrade performance or even cause damage.[13] The calculations were performed using the classical atomic simulation code LAMMPS[60,61] (which is compatible with the KIM API). For all simulations, periodic boundary conditions were applied in all directions, with a spacing of 40 Å in the direction perpendicular to the MoS_2 layer to isolate it from its periodic images. The simulation setup and results for the different properties are described below.

TABLE I. Fitted SW potential parameters in the two-body term ϕ_2.

	Interaction		
Parameter	Mo-Mo	Mo-S	S-S
A (eV)	3.9781804791	11.3797414404	1.1907355764
B	0.4446021306	0.5266688197	0.9015152673
p	5	5	5
q	0	0	0
σ (Å)	2.85295	2.17517	2.84133
r^{cut} (Å)	5.54660	4.02692	4.51956

TABLE II. Fitted SW potential parameters in the three-body term ϕ_3.

$\lambda_{\text{S-Mo-S}} = 7.4767529158$ eV	$\lambda_{\text{Mo-S-Mo}} = 8.1595181220$ eV
$\gamma = 1.3566322033$ Å	$\beta^0 = 81.7868^\circ$
$r^{\text{cut}}_{\text{Mo-S}} = 4.02692$ Å	$r^{\text{cut}*}_{\text{S-S}} = 3.86095$ Å

A. Lattice constants and cohesive energy

The zero-temperature equilibrium lattice constants and cohesive energy of MoS_2 were obtained by minimizing the energy of a single unit cell using conjugate gradients with energy and force tolerances of 10^{-10} eV and 10^{-10} eV/Å, respectively. The results for the SW-FM potential along with other potentials and DFT results are listed in Table III. As expected, both a_0 and b_0 agree with the SIESTA predictions since the potential σ parameters were preset to reproduce the equilibrium structure as explained in Sec. II B. The cohesive energy per unit cell E_c predicted by SW-FM is in good agreement with SIESTA (and other DFT) results.

The cohesive energy versus lattice constant curves plotted in Fig. 3 for different potentials and SIESTA show the effect of stretching and compressing MoS_2, which can be important for practical applications due to prestraining or rippling. The points were computed by creating a unit cell with in-plane lattice constant a and relaxing the unit cell atoms in the out-of-plane direction. The results show that SW-FM agrees with DFT SIESTA results across the entire range of stretching and compression (about ±15%), whereas other potentials agree either in tension or in compression, but not both.

B. Elastic constant

The zero-temperature elastic constants were computed using LAMMPS by finite difference, $C_{11} = \Delta\sigma_1/\Delta\epsilon_1$ and $C_{12} = (\Delta\sigma_1/\Delta\epsilon_2 + \Delta\sigma_2/\Delta\epsilon_1)/2$ in Voigt notation, where $\Delta\sigma$

TABLE III. Equilibrium lattice constants a_0 and b_0 (Å), cohesive energy E_c per unit cell (eV), and elastic constants C_{11} and C_{12} (N/m) for SW-FM, other potentials in the literature, and first principles results.

Method	a_0	b_0	E_c	C_{11}	C_{12}
SW-FM	3.19702	3.19386	15.28	119.2	41.0
SW-Jiang 2013	3.09368	3.18216	12.76	140.8	52.7
SW-Jiang 2015	3.11072	3.12898	3.72	105.0	28.7
REBO	3.16752	3.24248	21.48	154.4	45.8
ReaxFF[a]	3.19	3.11	15.20	205.1	81.6
SIESTA (GGA: PBE)	3.20	3.19	15.90	…	…
VASP (GGA: PW91)[b]	3.20	3.13	15.55	…	…
VASP (LDA)[c]	3.11	3.11	19.05	…	…
VASP (GGA: PBE)[c]	3.19	3.13	15.21	…	…
VASP (LDA)[c]	3.13	3.12	18.75	…	…
VASP (LDA)[d]	…	…	…	140.0	40.0
VASP (GGA: PBE)[d]	…	…	…	130.0	40.0
VASP (GGA: PBE)[e]	…	…	…	132.7	33.0

[a]Reference 29.
[b]Reference 62.
[c]Reference 63.
[d]Reference 11.
[e]Reference 64.

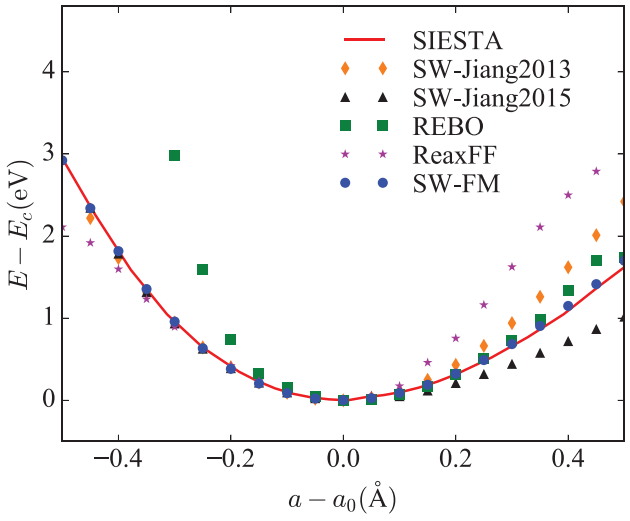

FIG. 3. Energy of unit cell as a function of in-plane lattice constant a. The data points are shifted such that all minima coincide. E_c and a_0 for each potential are listed in Table III.

and $\Delta\epsilon$ are the stress and strain induced by infinitesimally displacing atoms from their equilibrium positions.

Due to symmetry C_{11} and C_{22} are the same, which means that orientation (at least the armchair and zigzag directions) does not affect the elastic behavior of MoS_2. The results for C_{11} and C_{12} are listed in Table III. The SW-FM predictions are in good overall agreement with DFT, comparable with the other potentials (except for the ReaxFF potential, which appears to overestimate the elastic constants).

C. In-plane linear thermal expansion coefficient (LTEC)

The in-plane linear thermal expansion coefficient (LTEC) α_L of MoS_2 provides a measure of the temperature dependence of the lattice constants. There are two main methods for calculating LTEC from MD simulations. First, in the direct method, the LTEC is computed from its definition by taking the first derivative of lattice constant with respect to temperature at constant pressure

$$\alpha_L = \frac{1}{a}\frac{\partial a}{\partial T}\bigg|_p. \tag{7}$$

Second, in the fluctuation method, the LTEC is computed as an ensemble average of the covariance of the Hamiltonian $\mathcal{H}$ and the volume V[65]

$$\alpha_L = \frac{1}{2k_B T^2 \langle V\rangle}[\langle \mathcal{H}V\rangle - \langle \mathcal{H}\rangle\langle V\rangle], \tag{8}$$

where $\langle\cdot\rangle$ denotes a phase average and k_B is the Boltzmann constant. A detailed derivation of (8) is given in Appendix A.

To generate the data for both methods, a series of isothermal-isobaric (*NPT*) MD simulations were performed with a configuration of 1200 atoms (400 Mo and 800 S) at different temperatures and zero pressure. The equations of motion were integrated using a velocity-Verlet algorithm with a time step of 1 fs. The system was initially maintained at constant temperature using a Langevin thermostat for 10^6 time steps. Then, a Berendsen barostat was added and the system was evolved for another 10^6 time steps. This equilibration phase effectively dissipates lattice phonons generated by the initial conditions. Finally, the system was switched to an isothermal-isobaric (*NPT*) ensemble for 10^7 time steps using a Nose-Hoover thermostat and barostat to control the temperature and pressure with damping coefficients of 0.01 fs^{-1} and 0.001 fs^{-1}, respectively.

At a given temperature, the equilibrium supercell size in the x direction, L_x, was computed by averaging the instant cell size values. The equilibrium lattice constant defined in (7) follows as $a = L_x/c$, where c is the number of unit cells along the x direction in the supercell. [For example, if we use the system depicted in Fig. 1(a), c equals 4.][66] The equilibrium lattice constant at different temperatures is plotted in Fig. 4(a). To obtain the corresponding LTEC curve using the direct method in (7), it is necessary to obtain the slope of the lattice constant curve. To this end, we use the Gaussian process regression (GPR) method implemented in Scikit-learn[67,68] to fit the lattice constant data [blue line in Fig. 4(a)].[69] The LTEC is then computed from (7) by using finite differences to compute the derivative

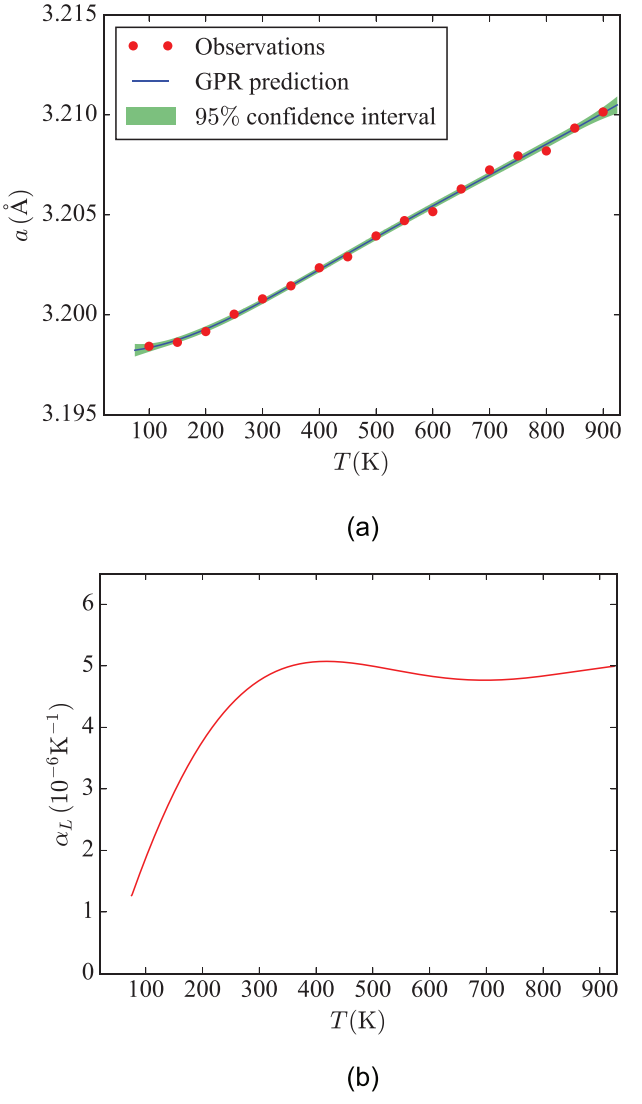

FIG. 4. (a) Equilibrium lattice constant a, and (b) the corresponding LTEC α_L computed using the direct method for the SW-FM potential.

of the GPR curve. To use GPR, it is necessary to provide lattice constant uncertainty. To compute this uncertainty at a given temperature, ten subsets, each with 5×10^5 simulation steps, were drawn randomly and independently from the simulation trajectory. The mean lattice constant was computed for each subset, and from this set of values, the standard deviation was obtained and used as the uncertainty. The result is plotted in Fig. 4(b). The numerical values of the standard deviations are too small to be seen, so they are not depicted in the figures.

For the fluctuation method, α_L was computed using (8). The resulting LTEC α_L values are plotted in Fig. 5. Similar to the direct method, GPR was used to fit the data with uncertainties computed using the same procedure described above. In the figure, the uncertainties are shown as error bars, and the 95% confidence interval predicted by GPR is plotted as the shaded green region.

Both the direct and fluctuation methods using SW-FM show that the LTEC α_L increases quickly at low temperatures and saturates at about 400 K. These results are in agreement with quasiharmonic DFT predictions,[9,13] and classic MD prediction[70] using the REBO potential.[18] We also computed the temperature dependence of the in-plane lattice constant using the original SW potentials in Refs. 26 and 27. The former[26] predicts that the lattice constant decreases with increasing temperature, resulting in a negative LTEC α_L. The latter[27] predicts a positive increasing LTEC α_L in the temperature range 0 to 900 K, that is, the LTEC α_L does not saturate at high temperature as observed by SW-FM and other sources as described above.

V. SENSITIVITY ANALYSIS

We perform a sensitivity analysis of the SW-FM potential based on Fisher information theory extended to path-space distributions.[35,71] This analysis has two objectives. First, it provides an estimate for the uncertainty in the SW-FM potential parameters, that is, how well the parameters are identified from the training set. Second, it provides an estimate for the uncertainty in the predictions of the SW-FM potential for new properties. The analysis is based on the path-space Fisher information matrix (FIM), which is an extension of the traditional FIM and plays a similar role.[33] For the force-matching potential fitting problem considered in this work, the path-space FIM has the following form:[72]

$$F_{ij}(\boldsymbol{\theta}) = \frac{1}{2k_B T\eta}\mathbb{E}_{\text{eq}}\left[\frac{\partial \boldsymbol{f}(\boldsymbol{r};\boldsymbol{\theta})}{\partial\theta_i}\cdot\frac{\partial \boldsymbol{f}(\boldsymbol{r};\boldsymbol{\theta})}{\partial\theta_j}\right] \approx \frac{1}{2k_B T\eta}\frac{1}{M}\sum_{m=1}^{M}\frac{\partial \boldsymbol{f}(\boldsymbol{r}_m;\boldsymbol{\theta})}{\partial\theta_i}\cdot\frac{\partial \boldsymbol{f}(\boldsymbol{r}_m;\boldsymbol{\theta})}{\partial\theta_j}. \quad (9)$$

The expected value $\mathbb{E}_{\text{eq}}[\cdot]$ denotes averaging with respect to an equilibrium (stationary) distribution of the observed dynamics. This is approximated by ergodic averaging on an equilibrated trajectory where M is the number of sampled configurations in the training set. The damping coefficient η is associated with the thermostat used to control temperature in the AIMD simulation. In our simulations $\eta = 0.02\,\text{fs}^{-1}$.

The path-space FIM was evaluated for the fitted parameters given in Tables I and II. The derivatives of the SW-FM force $\boldsymbol{f}$ with respect to the potential parameters $\boldsymbol{\theta}$ were calculated by finite difference using Ridders' method.[73,74]

A. Parameter uncertainty

The diagonal elements of the inverse FIM provide a lower bound on the variance of any unbiased estimator $\hat{\boldsymbol{\theta}}$ for the parameters, known as the Cramér-Rao bound[38,39]

$$\text{Var}_{\boldsymbol{\theta}}[\hat{\theta}_i] \geq [\boldsymbol{F}^{-1}]_{ii}. \quad (10)$$

In cases where the magnitudes of the model parameters differ greatly (as in our case where the parameters range over more than an order of magnitude), it is helpful to perform a relative parameter analysis by using the logarithm of the parameters instead of the parameters themselves. Defining $\tilde{\theta}_i = \log\theta_i$, it can be shown (see Appendix B) that the FIM in the logarithm parameter space is

$$\tilde{F}_{ij} = \theta_i F_{ij}\theta_j. \quad (11)$$

In terms of the logarithmic parameter space FIM, the Cramér-Rao bound in (10) becomes (see Appendix B)

$$\text{Var}_{\boldsymbol{\theta}}\left[\hat{\theta}_i/\theta_i\right] \geq [\tilde{\boldsymbol{F}}^{-1}]_{ii}. \quad (12)$$

This serves as an estimate for the uncertainty in the obtained parameters in a fractional sense.

The diagonal elements of the inverse FIM, $[\tilde{\boldsymbol{F}}^{-1}]_{ii}$, are plotted in Fig. 6. We see that all elements are within two orders of magnitude of each other and there are no parameters with extremely low values compared with the rest. This suggests that all parameters in the potential are identified, and there is no cause to simplify the model by removing undetermined parameters.

Examining the results more closely, we see that for the two-body interaction parameters (A, B, p, and σ), the lower bounds for the standard deviation of the logarithmic parameters $\tilde{\boldsymbol{\theta}}$ associated with Mo-S interactions are smaller than their Mo-Mo and S-S counterparts, which loosely indicates

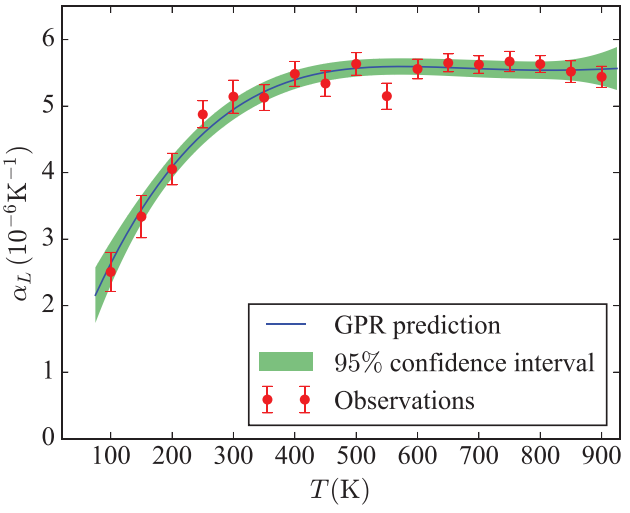

FIG. 5. LTEC α_L computed using the fluctuation method for the SW-FM potential. The line is a GPR fit to the data.

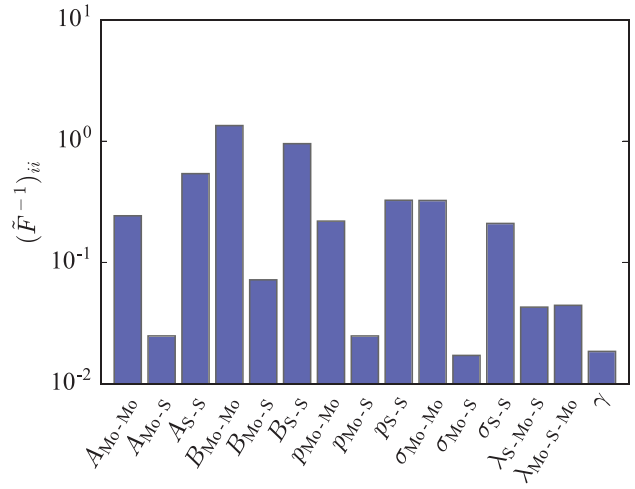

FIG. 6. The diagonal elements of the inverse FIM in logarithmic space.

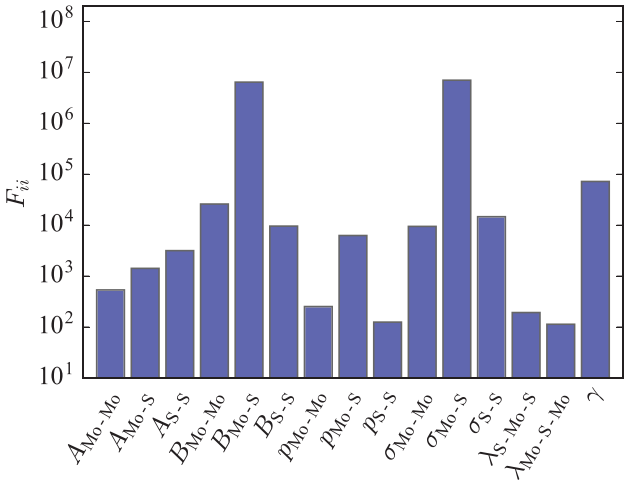

FIG. 7. The diagonal elements of the FIM.

that the Mo-S parameters are better determined. This is consistent with our knowledge of bonding in MoS_2, where the Mo-S bonds are shortest and expected to be strongest.[75]

B. Model uncertainty

The diagonal elements of the FIM provide an upper bound on the uncertainty due to variations in parameters in any observable predicted by the model that is obtained by averaging with respect to an equilibrium distribution in phase space. The bound is[32,33,35]

$$|\mathbb{E}_{\boldsymbol{\theta}+\boldsymbol{e}_i\Delta\theta_i}[\phi] - \mathbb{E}_{\boldsymbol{\theta}}[\phi]| \le \mathrm{Std}_{\boldsymbol{\theta}}[\phi]|\Delta\theta_i|\sqrt{F_{ii}}, \tag{13}$$

where $\mathbb{E}_{\boldsymbol{\theta}}[\phi]$ is the expectation of ϕ using the parameter set $\boldsymbol{\theta}$, $\mathrm{Std}_{\boldsymbol{\theta}}[\phi]$ is the corresponding standard deviation, $\Delta\theta_i$ is a perturbation of parameter θ_i, $\boldsymbol{e}_i$ is a unit vector of dimension P (where P is the number of parameters) with the ith component equal to one and all others zero. Thus, the diagonal of the FIM provides an upper bound on the uncertainty of the predictions of the potential to its fitting parameters. The larger F_{ii}, the more sensitive the predictions are to parameter θ_i.

The diagonal elements of the FIM, F_{ii}, are plotted in Fig. 7. We see that the SW-FM potential is most sensitive to $\sigma_{\text{Mo-S}}$ and least sensitive to $\lambda_{\text{Mo-S-Mo}}$. The ratio of F_{ii} for these two parameters is on the order of 10^5. In particular for the two-body interaction parameters (B, p, and σ), as we noted above, the potential is more sensitive to parameters associated with Mo-S interactions than to those associated with Mo-Mo and S-S interactions.

As an example for the bound in (13), we take the observable $\mathbb{E}_{\boldsymbol{\theta}}(\phi)$ to be the mean thickness $\bar{t}$ of a MoS_2 sheet

$$\bar{t} = \mathbb{E}_{\boldsymbol{\theta}}\left[\frac{1}{N}\left(\sum_{i=1}^{N} z_i^{\text{top}} - \sum_{i=1}^{N} z_i^{\text{bot}}\right)\right], \tag{14}$$

N is the number of atoms in each sulfur layer, z_i^{top} is the coordinate perpendicular to the MoS_2 plane of atom i in top layer, and z_i^{bot} is similarly defined for the bottom layer. The mean thickness $\bar{t}$ was computed by performing an MD simulation at $T = 750$ K using LAMMPS with the same setup used for calculations of the LTEC in Sec. IV C, however under *NVT* conditions. The simulation was repeated ten times with different initialization of atom velocities to compute the standard deviation that appears on the right-hand side of (13).

Equation (13) was evaluated for the parameters associated with the maximum and minimum diagonal FIM elements (see Fig. 7), $\sigma_{\text{Mo-S}}$ and $\lambda_{\text{Mo-S-Mo}}$, respectively. These are the parameters with respect to which observables will be most and least sensitive.

To evaluate (13), the mean thickness in (14) was computed for the parameter set $\boldsymbol{\theta}$ in Tables I and II, and also for $\boldsymbol{\theta} + \boldsymbol{e}_i\Delta\theta_i$ for the two studied parameters with $\Delta\theta_i = 0.01\theta_i$.

For parameter $\theta_i = \sigma_{\text{Mo-S}}$, we find

$$\begin{aligned}\text{Left of (13)} &= |3.21887 - 3.19900| = 0.01987\\ \text{Right of (13)} &= 0.00087 \times (0.01 \times 2.17517)\\ &\quad \times \sqrt{7.02436 \times 10^6} = 0.05016,\end{aligned} \tag{15}$$

and for parameter $\theta_i = \lambda_{\text{Mo-S-Mo}}$, we find

$$\begin{aligned}\text{Left of (13)} &= |3.19874 - 3.19900| = 0.00026\\ \text{Right of (13)} &= 0.00087 \times (0.01 \times 8.15952)\\ &\quad \times \sqrt{1.13313 \times 10^2} = 0.00076.\end{aligned} \tag{16}$$

We see that as expected (13) is satisfied for both parameters, and at least in this case, the bounds are rather tight. Thus, (13) can be used to estimate the reliability of a model in making new predictions.

VI. SUMMARY

The accuracy of an interatomic potential is critical to atomistic simulations of 2D monolayer MoS_2. We have parameterized a SW potential for MoS_2 using the force-matching method (SW-FM), where the potential parameters were optimized to match as closely as possible a training set of forces generated from a SIESTA AIMD trajectory at $T = 750$ K. The cutoffs and the reference bond angles were determined from the geometry of relaxed monolayer MoS_2 structure predicted by SIESTA. The equilibrium bond lengths and bond angles are prebuilt into the potential by applying

appropriate constraints to the potential parameters. In this way, the relaxed structure of monolayer MoS_2 is guaranteed to have the correct geometry.

To test the accuracy of the fitted potential, it was used to compute the lattice constants, cohesive energy versus lattice constant curve, elastic constants, and in-plane linear thermal expansion coefficient. Our validation tests show that:

1. The SW-FM potential correctly predicts the equilibrium lattice constants, cohesive energy, and energy versus lattice constant curves.
2. The elastic constant C_{11} is a bit underestimated compared with the first principles predictions, but the overall predictions for C_{11} and C_{12} are good.
3. The in-plane linear thermal expansion coefficient α_L, computed using both the direct method and the fluctuation method, increases rapidly at low temperature and saturates at high temperature in agreement with first principles calculations and classical computations using the REBO potential.

A Fisher information theory based sensitivity analysis shows that all the potential parameters are well identified. The potential is most sensitive to parameters associated with two-body Mo-S interactions, and less sensitive to Mo-Mo and S-S interactions. The analysis also provides an analytical upper bound on the uncertainty in any phase average predictions that the potential makes due to small changes in its parameters. This is demonstrated by example for the mean thickness of a MoS_2 sheet at finite temperature. The change in mean thickness computed by MD is found to be tightly bound by the analytical expression. The Fisher information theory based sensitivity analysis described in this paper is general and can be applied to potentials for other materials as long as the training set for the force-matching method is obtained from a dynamical trajectory sampling a distribution.

For the properties computed in this work, we find that the SW-FM potential outperforms the previous SW potentials for MoS_2[26,27] on which this work is based, and has comparable accuracy to the REBO[18] and ReaxFF[29] reactive potentials. Thanks to its simple functional form, MD simulations with SW-FM are significantly faster than with REBO or ReaxFF. We note that SW-FM is parameterized only for monolayer MoS_2 in the 2H phase, and thus should not be used for other phases of MoS_2 (e.g., bulk MoS_2 or the monolayer 1T phase).

The SW-FM potential is archived in OpenKIM.[51,76] It can be used with any KIM compliant molecular simulation code including LAMMPS,[60,61] ASE,[77–79] IMD,[80,81] DL_POLY,[82] and GULP.[45,46] See Appendix C for details.

ACKNOWLEDGMENTS

This research was partly supported by the Army Research Office (W911NF-14-1-0247) under the MURI program, and the National Science Foundation (NSF) under Grant Nos. PHY-0941493 and DMR-1408211. We thank Dr. Douglas E. Spearot for providing the LAMMPS implementation of the MoS_2 REBO potential and Dr. Adri van Duin for sharing the MoS_2 ReaxFF parameters. The authors wish to acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing resources that contributed to the results reported in this manuscript. The AIMD simulations were carried out on the Odyssey cluster of the Research Computing Group at Harvard University, and at the Extreme Science and Engineering Discovery Environment (XSEDE), supported by NSF Grant No. ACI-1053575.

APPENDIX A: FLUCTUATION METHOD TO COMPUTE LTEC

The partition function for the isothermal-isobaric (*NPT*) ensemble is[83]

$$Q(N,p,T) = C\int_0^\infty \mathrm{d}V \int_\Gamma \mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}\, e^{-\beta(\mathcal{H}+pV)}, \tag{A1}$$

where C is a normalization constant, $\beta = 1/(k_B T)$, k_B is the Boltzmann constant, $\mathcal{H}$ is the Hamiltonian, p is pressure, V is volume, and the two integrations are over volume space and phase space Γ. The macroscopic observable associated with a phase function A can be obtained as

$$\begin{aligned}\langle A\rangle &= \frac{C}{Q}\int_0^\infty \mathrm{d}V \int_\Gamma \mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}\, e^{-\beta(\mathcal{H}+pV)}A \\ &= CQ^{-1}\int_{\boldsymbol{x}} \mathrm{d}\boldsymbol{x}\, e^{-\beta(\mathcal{H}+pV)}A,\end{aligned} \tag{A2}$$

where in the last equality, we have rewritten the integration over volume space and phase space as $\int_{\boldsymbol{x}}$ for brevity.

Taking volume V as the phase function A, the derivative of the observable $\langle V\rangle$ with respect to β is

$$\begin{aligned}\left.\frac{\partial\langle V\rangle}{\partial\beta}\right|_{N,p} &= \frac{\partial(CQ^{-1})}{\partial\beta}\int_{\boldsymbol{x}} \mathrm{d}\boldsymbol{x}\, e^{-\beta(\mathcal{H}+pV)}V + CQ^{-1}\frac{\partial\int_{\boldsymbol{x}} \mathrm{d}\boldsymbol{x}\exp[-\beta(\mathcal{H}+pV)]V}{\partial\beta} \\ &= -CQ^{-2}\frac{\partial Q}{\partial\beta}\int_{\boldsymbol{x}} \mathrm{d}\boldsymbol{x}\, e^{-\beta(\mathcal{H}+pV)}V + CQ^{-1}\int_{\boldsymbol{x}} \mathrm{d}\boldsymbol{x}\frac{\partial\exp[-\beta(\mathcal{H}+pV)]}{\partial\beta}V \\ &= C^2Q^{-2}\int_{\boldsymbol{x}} \mathrm{d}\boldsymbol{x}\, e^{-\beta(\mathcal{H}+pV)}(\mathcal{H}+pV)\int_{\boldsymbol{x}} \mathrm{d}\boldsymbol{x}\, e^{-\beta(\mathcal{H}+pV)}V - CQ^{-1}\int_{\boldsymbol{x}} \mathrm{d}\boldsymbol{x}\, e^{-\beta(\mathcal{H}+pV)}(\mathcal{H}+pV)V \\ &= \left[CQ^{-1}\int_{\boldsymbol{x}} \mathrm{d}\boldsymbol{x}\, e^{-\beta(\mathcal{H}+pV)}(\mathcal{H}+pV)\right]\left[CQ^{-1}\int_{\boldsymbol{x}} \mathrm{d}\boldsymbol{x}\, e^{-\beta(\mathcal{H}+pV)}V\right] - CQ^{-1}\int_{\boldsymbol{x}} \mathrm{d}\boldsymbol{x}\, e^{-\beta(\mathcal{H}+pV)}(\mathcal{H}V+pV^2) \\ &= (\langle\mathcal{H}\rangle+\langle pV\rangle)\langle V\rangle - \langle\mathcal{H}V\rangle - \langle pV^2\rangle = \langle\mathcal{H}\rangle\langle V\rangle - \langle\mathcal{H}V\rangle + \langle pV\rangle\langle V\rangle - \langle pV^2\rangle,\end{aligned} \tag{A3}$$

where in the third equality, we used $\partial Q/\partial\beta = -C\int_{\boldsymbol{x}} d\boldsymbol{x}\, e^{-\beta(\mathcal{H}+pV)}(\mathcal{H}+pV)$, and in the second to last equality, we used (A2). The volumetric thermal expansion coefficient is

$$\alpha_V = \frac{1}{V}\frac{\partial V}{\partial T}\bigg|_{N,p} = \frac{1}{\langle V\rangle}\frac{\partial\langle V\rangle}{\partial\beta}\bigg|_{N,p}\frac{\partial\beta}{\partial T} = -k_B\beta^2\frac{1}{\langle V\rangle}\frac{\partial\langle V\rangle}{\partial\beta}\bigg|_{N,p}. \tag{A4}$$

At $p=0$, plugging (A3) into (A4), we obtain

$$\alpha_V = k_B\beta^2\frac{1}{\langle V\rangle}[\langle\mathcal{H}V\rangle - \langle\mathcal{H}\rangle\langle V\rangle]. \tag{A5}$$

It is seen that the volumetric thermal expansion coefficient $\alpha_V$is related to the covariance of the Hamiltonian $\mathcal{H}$ and the volume V.

Next, we get the LTEC α_L from α_V. For a 2 D material (e.g., MoS_2), assume $V = hL^2$, where L is the in-place dimension, and h is the out-of-place dimension independent of L. Then, the LTEC α_L is

$$\alpha_L = \frac{1}{L}\frac{\partial L}{\partial T}\bigg|_{N,p} = \left(\frac{V}{h}\right)^{-1/2}\frac{\partial L}{\partial V}\frac{\partial V}{\partial T}\bigg|_{N,p} = \frac{1}{2}\frac{1}{V}\frac{\partial V}{\partial T}\bigg|_{N,p} = \frac{1}{2}\alpha_V. \tag{A6}$$

APPENDIX B: SENSITIVITY ANALYSIS IN LOGARITHM PARAMETER SPACE

Defining $\tilde{\theta}_i = \log\theta_i$, we have

$$\partial\boldsymbol{f}/\partial\tilde{\theta}_i = \partial\boldsymbol{f}/\partial\theta_i\cdot\partial\theta_i/\partial\tilde{\theta}_i = \theta_i\partial\boldsymbol{f}/\partial\theta_i. \tag{B1}$$

The FIM in logarithm parameter space is

$$\begin{aligned}\tilde{F}_{ij}(\tilde{\boldsymbol{\theta}}) &= \frac{1}{2k_BT\eta}\mathbb{E}_{\text{eq}}\left[\frac{\partial\boldsymbol{f}(\boldsymbol{r};\boldsymbol{\theta})}{\partial\tilde{\theta}_i}\cdot\frac{\partial\boldsymbol{f}(\boldsymbol{r};\boldsymbol{\theta})}{\partial\tilde{\theta}_j}\right]\\ &= \theta_i\left(\frac{1}{2k_BT\eta}\mathbb{E}_{\text{eq}}\left[\frac{\partial\boldsymbol{f}(\boldsymbol{r};\boldsymbol{\theta})}{\partial\theta_i}\cdot\frac{\partial\boldsymbol{f}(\boldsymbol{r};\boldsymbol{\theta})}{\partial\theta_j}\right]\right)\theta_j = \theta_iF_{ij}\theta_j,\end{aligned} \tag{B2}$$

where in the second and third equality, we used (B1) and (9), respectively.

For an unbiased estimator $\hat{\boldsymbol{\theta}}$ of the model parameters, the Cramér-Rao bound is expressed as

$$\text{Cov}_{\boldsymbol{\theta}}[\hat{\boldsymbol{\theta}}] \geq \boldsymbol{F}^{-1}(\boldsymbol{\theta}). \tag{B3}$$

Let $\boldsymbol{D} = \text{diag}(\boldsymbol{\theta})$ be the diagonal matrix generated from vector $\boldsymbol{\theta}$. Pre-multiplying both sides of (B3) by $\boldsymbol{D}^{-1}$ and post-multiplying by $\boldsymbol{D}^{-\text{T}}$ gives

$$\begin{aligned}\boldsymbol{D}^{-1}\text{Cov}_{\boldsymbol{\theta}}[\hat{\boldsymbol{\theta}}]\boldsymbol{D}^{-\text{T}} &= \boldsymbol{D}^{-1}\mathbb{E}\left[(\hat{\boldsymbol{\theta}}-\mathbb{E}[\hat{\boldsymbol{\theta}}])(\hat{\boldsymbol{\theta}}-\mathbb{E}[\hat{\boldsymbol{\theta}}])^{\text{T}}\right]\boldsymbol{D}^{-\text{T}}\\ &= \mathbb{E}\left[\boldsymbol{D}^{-1}(\hat{\boldsymbol{\theta}}-\mathbb{E}[\hat{\boldsymbol{\theta}}])(\hat{\boldsymbol{\theta}}-\mathbb{E}[\hat{\boldsymbol{\theta}}])^{\text{T}}\boldsymbol{D}^{-\text{T}}\right]\\ &= \mathbb{E}\left[\left(\boldsymbol{D}^{-1}(\hat{\boldsymbol{\theta}}-\mathbb{E}[\hat{\boldsymbol{\theta}}])\right)\left(\boldsymbol{D}^{-1}(\hat{\boldsymbol{\theta}}-\mathbb{E}[\hat{\boldsymbol{\theta}}])\right)^{\text{T}}\right]\\ &= \text{Cov}_{\boldsymbol{\theta}}[\boldsymbol{D}^{-1}\hat{\boldsymbol{\theta}}],\end{aligned} \tag{B4}$$

and

$$\boldsymbol{D}^{-1}\boldsymbol{F}^{-1}(\boldsymbol{\theta})\boldsymbol{D}^{-\text{T}} = \left(\boldsymbol{D}^{\text{T}}\boldsymbol{F}(\boldsymbol{\theta})\boldsymbol{D}\right)^{-1} = \tilde{\boldsymbol{F}}(\tilde{\boldsymbol{\theta}})^{-1}, \tag{B5}$$

where we used (B2).

Because $\boldsymbol{D}$ is a non-negative matrix (i.e., each element of $\boldsymbol{D}$ is non-negative), the inequality (B3) still holds when pre-multiplied by D^{-1} and post-multiplied by $\boldsymbol{D}^{-\text{T}}$. Therefore

$$\boldsymbol{D}^{-1}\text{Cov}_{\boldsymbol{\theta}}[\hat{\boldsymbol{\theta}}]\boldsymbol{D}^{-\text{T}} \geq \boldsymbol{D}^{-1}\boldsymbol{F}^{-1}(\boldsymbol{\theta})\boldsymbol{D}^{-\text{T}}, \tag{B6}$$

and using (B4) and (B5), we have

$$\text{Cov}_{\boldsymbol{\theta}}[\boldsymbol{D}^{-1}\hat{\boldsymbol{\theta}}] \geq \tilde{\boldsymbol{F}}(\tilde{\boldsymbol{\theta}})^{-1}. \tag{B7}$$

The ith diagonal element of this inequality is

$$\text{Var}_{\boldsymbol{\theta}}\left[\hat{\theta}_i/\theta_i\right] \geq \left[\tilde{\boldsymbol{F}}^{-1}\right]_{ii}. \tag{B8}$$

APPENDIX C: USING THE OPEN KNOWLEDGEBASE OF INTERATOMIC MODELS (OPENKIM)

The Open Knowledgebase of Interatomic Models (OpenKIM) (https://openkim.org) is an open-source, publicly accessible repository of classical interatomic potentials, as well as their predictions for material properties that can be visualized and compared with first-principles data. Interatomic potentials stored in OpenKIM that are compatible with the KIM application programming interface (API) are called "KIM Models." KIM Models will work seamlessly with a variety of major simulation codes that are compatible with this standard including LAMMPS,[60,61] ASE,[77–79] IMD,[80,81] DL_POLY,[82] and GULP.[45,46]

As an example, we describe how a KIM Model would be used with LAMMPS. In LAMMPS, reactive interatomic potentials are specified using the pair_style command. LAMMPS has a "pair_style kim" option for using KIM Models. To use KIM Models with LAMMPS, perform the following steps:

1. Install the KIM API (see instructions at https://openkim.org/kim-api/);
2. Download and install the desired potential from https://openkim.org/ (see instructions that come with the API);
3. Enable KIM Models in LAMMPS by typing: "`make yes-kim`" and then compiling LAMMPS.

In a LAMMPS input script, a KIM Model is then selected in the same way as other LAMMPS potentials. For example, the potential developed in this paper can be used with the following two commands:

```
pair_style   kim   LAMMPSvirial   Three_Body_Stillinger_Weber_MoS__MO_201919462778_000
pair_coeff   *  *   Mo S
```

The advantage of releasing the potential as a KIM Model (as opposed to just a file compatible with LAMMPS or another code), is that the model will work with not just LAMMPS, but other major codes as noted above. In addition, a KIM Model has a "KIM ID" that can be cited in publications. The KIM ID provides a unique permanent link to the archived content and includes a three-digit version number to keep track of modifications. For example, a modification to the model parameters would lead to a version upgrade (or a new forked model if appropriate). Citing a KIM ID in a publication makes it possible for the reader to download the exact potential used in the reported simulation and to reproduce the results. The KIM ID for the model described in this paper is "MO_201919462778_000" (the last three digits are the version number).

[1]K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, Science **306**, 666 (2004).
[2]K. Novoselov, A. K. Geim, S. Morozov, D. Jiang, M. Katsnelson, I. Grigorieva, S. Dubonos, and A. Firsov, Nature **438**, 197 (2005).
[3]Y. Zhang, Y.-W. Tan, H. L. Stormer, and P. Kim, Nature **438**, 201 (2005).
[4]S. Lebegue and O. Eriksson, Phys. Rev. B **79**, 115409 (2009).
[5]K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, Phys. Rev. Lett. **105**, 136805 (2010).
[6]E. Hendry, P. J. Hale, J. Moger, A. K. Savchenko, and S. A. Mikhailov, Phys. Rev. Lett. **105**, 097401 (2010).
[7]Z. Yin, H. Li, H. Li, L. Jiang, Y. Shi, Y. Sun, G. Lu, Q. Zhang, X. Chen, and H. Zhang, ACS Nano **6**, 74 (2012).
[8]D. Yoon, Y.-W. Son, and H. Cheong, Nano Lett. **11**, 3227 (2011).
[9]C. Sevik, Phys. Rev. B **89**, 035422 (2014).
[10]C. Lee, X. Wei, J. W. Kysar, and J. Hone, Science **321**, 385 (2008).
[11]R. C. Cooper, C. Lee, C. A. Marianetti, X. Wei, J. Hone, and J. W. Kysar, Phys. Rev. B **87**, 035423 (2013).
[12]C. Zhi, Y. Bando, C. Tang, H. Kuwahara, and D. Golberg, Adv. Mater. **21**, 2889 (2009).
[13]L. F. Huang, P. L. Gong, and Z. Zeng, Phys. Rev. B **90**, 045409 (2014).
[14]Y. Mishin, D. Farkas, M. J. Mehl, and D. A. Papaconstantopoulos, Phys. Rev. B **59**, 3393 (1999).
[15]M. Wen, S. M. Whalen, R. S. Elliott, and E. B. Tadmor, Modell. Simul. Mater. Sci. Eng. **23**, 074008 (2015).
[16]M. Wen, J. Li, P. Brommer, R. S. Elliott, J. P. Sethna, and E. B. Tadmor, Modell. Simul. Mater. Sci. Eng. **25**, 014001 (2017).
[17]N. Wakabayashi, H. G. Smith, and R. M. Nicklow, Phys. Rev. B **12**, 659 (1975).
[18]T. Liang, S. R. Phillpot, and S. B. Sinnott, Phys. Rev. B **79**, 245110 (2009).
[19]G. Abell, Phys. Rev. B **31**, 6184 (1985).
[20]J. Tersoff, Phys. Rev. Lett. **56**, 632 (1986).
[21]J. Tersoff, Phys. Rev. B **37**, 6991 (1988).
[22]J. Tersoff, Phys. Rev. B **38**, 9902 (1988).
[23]D. W. Brenner, Phys. Rev. B **42**, 9458 (1990).
[24]J. E. Jones, Proc. R. Soc. A **106**, 441 (1924).
[25]J. E. Jones, Proc. R. Soc. A **106**, 463 (1924).
[26]J.-W. Jiang, H. S. Park, and T. Rabczuk, J. Appl. Phys. **114**, 064307 (2013).
[27]J.-W. Jiang, Nanotechnology **26**, 315706 (2015).
[28]F. H. Stillinger and T. A. Weber, Phys. Rev. B **31**, 5262 (1985).
[29]A. Ostadhossein, A. Rahnamoun, Y. Wang, P. Zhao, S. Zhang, V. H. Crespi, and A. C. van Duin, J. Phys. Chem. Lett. **8**, 631 (2017).
[30]F. Ercolessi and J. B. Adams, Europhys. Lett. **26**, 583 (1994).
[31]R. A. Fisher, Philos. Trans. R. Soc. London, Ser. A **222**, 309 (1922).
[32]Y. Pantazis and M. A. Katsoulakis, J. Chem. Phys. **138**, 054115 (2013).
[33]P. Dupuis, M. A. Katsoulakis, Y. Pantazis, and P. Plechác, SIAM/ASA J. Uncertainty Quantif. **4**, 80 (2016).
[34]E. Kalligiannaki, V. Harmandaris, M. A. Katsoulakis, and P. Plechác, J. Chem. Phys. **143**, 084105 (2015).
[35]V. Harmandaris, E. Kalligiannaki, M. Katsoulakis, and P. Plecháč, J. Comput. Phys. **314**, 355 (2016).
[36]P. Español and I. Zuniga, Phys. Chem. Chem. Phys. **13**, 10538 (2011).
[37]L. Lu, J. F. Dama, and G. A. Voth, J. Chem. Phys. **139**, 121906 (2013).
[38]A. van den Bos, *Parameter Estimation for Scientists and Engineers* (John Wiley & Sons, Hoboken, 2007).
[39]H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, 1999).
[40]Z. Jian, Z. Kaiming, and X. Xide, Phys. Rev. B **41**, 12915 (1990).
[41]M. Ichimura, Phys. Status Solidi A **153**, 431 (1996).
[42]X. W. Zhou, D. K. Ward, J. E. Martin, F. B. van Swol, J. L. Cruz-Campa, and D. Zubia, Phys. Rev. B **88**, 085309 (2013).
[43]J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, J. Phys.: Condens. Matter **14**, 2745 (2002).
[44]J. Li, Modell. Simul. Mater. Sci. Eng. **11**, 173 (2003).
[45]J. D. Gale, J. Chem. Soc., Faraday Trans. **93**, 629 (1997).
[46]See http://nanochemistry.curtin.edu.au/gulp/ for "General utility lattice program (GULP)" (2017).
[47]This equation comes from the Lennard-Jones potential,[24,25] $\Phi(r) = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6]$, whose minimum is at $r_m = 2^{1/6}\sigma$. The two-body term $\phi_2(r)$ of the SW potential is based on the Lennard-Jones potential with an additional exponential cutoff.
[48]P. Brommer and F. Gähler, Philos. Mag. **86**, 753 (2006).
[49]P. Brommer and F. Gähler, Modell. Simul. Mater. Sci. Eng. **15**, 295 (2007).
[50]P. Brommer, A. Kiselev, D. Schopf, P. Beck, J. Roth, and H.-R. Trebin, Modell. Simul. Mater. Sci. Eng. **23**, 074002 (2015).
[51]E. B. Tadmor, R. S. Elliott, J. P. Sethna, R. E. Miller, and C. A. Becker, JOM **63**, 17 (2011).
[52]J. J. Waterfall, F. P. Casey, R. N. Gutenkunst, K. S. Brown, C. R. Myers, P. W. Brouwer, V. Elser, and J. P. Sethna, Phys. Rev. Lett. **97**, 150601 (2006).
[53]M. K. Transtrum, B. B. Machta, and J. P. Sethna, Phys. Rev. Lett. **104**, 060201 (2010).
[54]M. K. Transtrum, B. B. Machta, and J. P. Sethna, Phys. Rev. E **83**, 036701 (2011).
[55]D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).
[56]N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
[57]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[58]Brillouin zone integration using a single k-point is generally inaccurate. However due to the relatively large supercell used here, the size of the reciprocal vectors are relatively small, and thus the sampling grid in reciprocal space is dense. The resulting accuracy is considered adequate, especially considering the high cost of the AIMD calculations that would greatly increase with a denser k-point grid.
[59]We also tried to fit the potential at other temperatures, but found only slight differences between the fitted potential parameters and the resulting predictions for material properties. We therefore only include the results for $T = 750$ K.
[60]S. Plimpton, J. Comput. Phys. **117**, 1 (1995).
[61]See http://lammps.sandia.gov for "Large-scale atomic/molecular massively parallel simulator (LAMMPS)" (2017).
[62]C. Ataca, H. Sahin, E. Akturk, and S. Ciraci, J. Phys. Chem. C **115**, 3934 (2011).
[63]Y. Ding, Y. Wang, J. Ni, L. Shi, S. Shi, and W. Tang, Physica B **406**, 2254 (2011).
[64]D. Cakir, F. M. Peeters, and C. Sevik, Appl. Phys. Lett. **104**, 203110 (2014).
[65]M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1989).
[66]We verified that upon heating the MoS_2hexagonal lattice expands uniformly. Therefore it does not matter whether the lattice is oriented in the zigzag or armchair direction along the x-axis when computing the lattice constant a, since both will give the same result.
[67]F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, J. Mach. Learn. Res. **12**, 2825 (2011).

[68]See http://scikit-learn.org for "Scikit-learn: Machine learning in Python" (2017).
[69]Polynomials are often used to fit lattice constant data in order to compute the LTEC, for example in Ref. 70. However, we found that in our case polynomials were far too sensitive. A small perturbation in one of the lattice constant data points can lead to a completely different LTEC curve.
[70]P. Anees, M. Valsakumar, and B. Panigrahi, Appl. Phys. Lett. **108**, 101902 (2016).
[71]M. A. Katsoulakis and P. Plecháč, J. Chem. Phys. **139**, 074115 (2013).
[72]A. Tsourtis, Y. Pantazis, M. A. Katsoulakis, and V. Harmandaris, J. Chem. Phys. **143**, 014116 (2015).
[73]C. Ridders, Adv. Eng. Software **4**, 75 (1982).
[74]W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, Cambridge, 2007).
[75]Y. Li, Y.-L. Li, C. M. Araujo, W. Luo, and R. Ahuja, Catal. Sci. Technol. **3**, 2214 (2013).
[76]M. Wen, see https://openkim.org/cite/MO_201919462778_000 for "A three-body Stillinger-Weber model (parameterization) for monolayer MoS2" (2017).
[77]A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schtt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, J. Phys.: Condens. Matter **29**, 273002 (2017).
[78]See https://wiki.fysik.dtu.dk/ase/ for "ASE: The atomic simulation environment python library for working with atoms" (2017).
[79]M. Wen, see https://github.com/mjwen/kimpy for "kimpy: An ASE calculator based on OpenKIM Interatomic Models" (2017).
[80]J. Stadler, R. Mikulla, and H.-R. Trebin, Int. J. Mod. Phys. C **08**, 1131 (1997).
[81]See http://imd.itap.physik.uni-stuttgart.de for "IMD: The ITAP molecular dynamics program" (2017).
[82]See https://www.scd.stfc.ac.uk/Pages/DL_POLY.aspx for "DL_POLY classic molecular simulation package" (2017).
[83]M. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation* (Oxford University Press, Oxford, 2010).