

Machine Learning a General-Purpose Interatomic Potential for Silicon

Albert P. Bartók

*Scientific Computing Department, Science and Technology Facilities Council,
Rutherford Appleton Laboratory, Didcot, OX11 0QX, United Kingdom*

James Kermode

*Warwick Centre for Predictive Modelling, School of Engineering,
University of Warwick, Coventry, CV4 7AL, United Kingdom*

Noam Bernstein

*Center for Materials Physics and Technology, U.S. Naval Research Laboratory,
Washington, D.C. 20375, USA*

Gábor Csányi

*Engineering Laboratory, University of Cambridge,
Trumpington Street, Cambridge, CB2 1PZ, United Kingdom*

 (Received 26 May 2018; revised manuscript received 11 October 2018; published 14 December 2018)

The success of first-principles electronic-structure calculation for predictive modeling in chemistry, solid-state physics, and materials science is constrained by the limitations on simulated length scales and timescales due to the computational cost and its scaling. Techniques based on machine-learning ideas for interpolating the Born-Oppenheimer potential energy surface without explicitly describing electrons have recently shown great promise, but accurately and efficiently fitting the physically relevant space of configurations remains a challenging goal. Here, we present a Gaussian approximation potential for silicon that achieves this milestone, accurately reproducing density-functional-theory reference results for a wide range of observable properties, including crystal, liquid, and amorphous bulk phases, as well as point, line, and plane defects. We demonstrate that this new potential enables calculations such as finite-temperature phase-boundary lines, self-diffusivity in the liquid, formation of the amorphous by slow quench, and dynamic brittle fracture, all of which are very expensive with a first-principles method. We show that the uncertainty quantification inherent to the Gaussian process regression framework gives a qualitative estimate of the potential's accuracy for a given atomic configuration. The success of this model shows that it is indeed possible to create a useful machine-learning-based interatomic potential that comprehensively describes a material on the atomic scale and serves as a template for the development of such models in the future.

DOI: [10.1103/PhysRevX.8.041048](https://doi.org/10.1103/PhysRevX.8.041048)

Subject Areas: Computational Physics,
Condensed Matter Physics,
Materials Science

I. INTRODUCTION

A. Background

First-principles molecular simulation, based on various approximations of the electronic-structure theory, is the workhorse of materials modeling. For example, density-functional theory (DFT), a prominent method, is indicated

as the topic of about 19 000 papers published in 2017 according to the Web of Science database. However, because of the combination of its computational expense and unfavorable scaling, simulations that require thousands of atoms and/or millions of energy or force evaluations are carried out not using electronic-structure methods but *empirical analytical potentials* (also known as *force fields* in the chemistry literature). These are parametrized approximations of the Born-Oppenheimer potential energy surface (PES), the electronic ground-state energy viewed as a function of the nuclear positions [1].

The functional form of analytical potentials is typically simple, based partly on a combination of physical and

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

chemical intuition and partly on convenience. The parameters are usually optimized so that the model reproduces, as best as it is able, either some macroscopic observables or microscopic quantities such as total energies and/or forces, corresponding to selected configurations and calculated separately using an electronic-structure method. Unsurprisingly, while it is easy to find parameter sets that reproduce individual observables (e.g., the melting point corresponding to a particular composition or the binding energy of a crystalline structure), the simple functional forms postulated are not flexible enough to allow matching many properties simultaneously, and the potential energy surface is not accurate, which suggests that, when empirical analytical potentials are successful, there is a risk that this result is due to a fortuitous cancellation of errors and is then a case of getting the right answer for the wrong reasons. The practitioner is forced to select parameters subject to a trade-off between maximizing the accuracy for a few selected properties and transferability, i.e., avoiding a large or even qualitative error for a wide range of configurations and observables of interest [2,3]. This trade-off severely limits the predictive power of empirical analytical potentials when the very parameters that give sufficient accuracy for some known observables result in wildly varying predictions for new phases and properties whose prediction is the ultimate goal of the simulation.

Machine-learning (ML) methods provide a systematic approach to fitting functions in high-dimensional spaces without employing simply parametrized functional forms [4], thus opening up the possibility of creating interatomic potentials for materials with unprecedented accuracy for a wide range of configurations. The development in the past 10 yr required exploring a variety of ways to describe the chemical environment of atoms, the basis functions used to construct the potential, e.g., by various kernels or artificial neural network models, and the way such fits can be regularized, either by linear algebra or by various protocols of neural-network optimization [5,6]. The general approach is to define an atomic energy as a function of its local environment and fit this function in a 60–100-dimensional space corresponding to the relative positions of 20–30 nearby atoms. The challenge we take on here is to use this approach to develop a *general-purpose* interatomic potential, neither restricted to a narrow range of configurations or observables nor compromising accuracy in reproducing the reference potential energy surface. This approach requires adequately sampling enough of the space that is relevant to a wide range of atomistic simulations to interpolate it accurately and doing so in a computationally tractable manner.

The Achilles heel of machine-learning models, directly related to their flexibility, is their naturally much-reduced transferability: The flexible function representation is informed by a large training database, leading to a good fit for configurations near the database (in the space of the chosen representation) and progressively poorer away

from it. This limitation is often summarized by saying that high-dimensional fits are good at *interpolation* but less good at *extrapolation*, which can be viewed as another manifestation of the “curse of dimensionality”.

When first encountering the ML potential approach, one might wonder why such high-dimensional fits work at all, given that it is impractical to thoroughly sample the full space (e.g., on a grid). It is an empirical observation that they often do, so the real question is, what are the special properties of potential energy surfaces that make them amenable to such approximations? *Regularity* is almost certainly one of these, the mathematical concept encompassing the colloquial idea of a potential varying smoothly as a function of the atomic positions. Indeed, the regular kernels that are used (and the corresponding regular activation functions in artificial neural networks) define the length scales over which predictions are interpolated rather than extrapolated. This definition can also point towards explaining why some methods work better than others: Kernels that better capture the inherent regularity of the underlying function interpolate better and extrapolate farther. Another property is that the configurations that are likely to arise in an atomistic simulation actually occupy a volume of configuration space that is much smaller than the full space. Consequently, a database derived from configurations found in reference atomistic simulations is sufficient for fitting an interatomic potential, so long as it includes not only the low-energy configurations but also nearby high-energy ones to constrain the potential at the *boundary* of the region that will be explored when it is used.

Thus, the trade-off made by empirical analytical potentials (viz. between accuracy and transferability) is now replaced by another trade-off: that between transferability and database size, because high accuracy is possible but only near the training set. In order to achieve the promised wider impact in materials modeling, it is desirable to explore this trade-off. In particular, is it possible to create a training database of manageable size that covers almost all relevant configurations of a material and, thus, a potential for future larger length-scale simulations? Or will such models always be confined to a narrow set of atomic configurations, with every new scientific question necessitating a new fit trained on a problem-specific database of first-principles calculations? Since the notion of nearness is intimately tied to the representation, in this paper, we explore this question for a particular case, a kernel-based fit using the previously introduced smooth overlap of atomic positions (SOAP) kernel [7] and the Gaussian approximation potential (GAP) framework [8]. (Everywhere in this paper, when we refer to GAP models, we mean a Gaussian process regression model using the SOAP kernel, although, of course, other kernels and also combinations of different kernels can be used within the GAP framework and have indeed been used for other systems [9–12].)

An obvious alternative approach to the transferability problem is to give up on it entirely and accept that an interatomic potential will always be extremely narrowly confined to its training database. One can then develop algorithms that actively *adapt or grow* the training database during the course of a simulation [13–20]. The obvious disadvantage is that an electronic-structure method always has to be part of the simulation, to be called upon to calculate new target data as and when necessary. The efficacy of this approach then depends on what algorithm is used to detect that the simulation has strayed into parts of configuration space not sufficiently well covered by the current database and precisely what subsequent action is taken.

B. State of the art

The following is a brief review of the recent works in the emergent field of interatomic potential construction using high-dimensional nonparametric fits. Although fits to the potential energy surface of molecules and small molecule clusters have a much longer history [21–34], here we limit our scope to include only efforts that model strongly bound materials in the condensed phase. On the one hand, material models generally have a number of critical requirements that differentiate them from molecular models: (i) The potential must be reactive, i.e., needs to describe the forming and breaking of many covalent bonds, often simultaneously; (ii) a wide range of neighbor configurations need to be covered, including radical changes in neighbor count. Comprehensive models also need to (iii) cover multiple phases, e.g., metallic and insulating, solid and liquid, etc. On the other hand, many works cited below (and the present work) consider only one type of element, which allows the consideration of only relatively short-range interactions, because the absence of charge transfer obviates the need to describe long-range electrostatic effects.

Modeling the short-range interactions with artificial neural networks (NNs) really took off about a decade ago [35], starting with the bulk phases of silicon [36,37], with many more to follow: describing some silicon defects [38], the graphite-diamond transition [39], bulk zinc oxide [40], copper with some defects [41], the phase-change material GeTe in its various phases [42], various ionic solids [43,44], Li-Si alloys [45–47], bulk TiO₂ [48], alloys [49,50], Ta₂O₅ [51], Li₃PO₄ [52], gold clusters [53], graphene [54], and various surfaces [55–59]. Fitting NN potentials is beginning to be combined with the combinatorial structure search [60–62].

Kernel fitting is a different approach to high-dimensional interpolation, with origins in statistics [cf. kriging [63] and Gaussian process regression (GPR) [64]] and widely applied in numerical analysis and machine learning [65]. The key to its success is the choice of kernel and, through it, the basis functions employed. In the context of atomistic potentials, a significant step was the introduction of rotationally and permutationally invariant representations that

also vary smoothly with the coordination number, based on the spherical Fourier transform and the bispectrum constructed from it [66]. This representation was later simplified to the SOAP representation and kernel [7] and applied to tungsten [67], amorphous carbon [11,68], iron [69], graphene [12], and boron [70]. Using the original spherical bispectrum as a representation, potentials were made with linear regression for tantalum [71] and molybdenum [72], with quadratic regression for tantalum [73], and with a nonlinear kernel for bulk Li₂B₁₂H₁₂ [74]. Linear regression using yet another class of basis functions was introduced by Shapeev [75] and used to make a potential for Li [76]. Others use GPR with different representations to fit forces directly without constructing a potential, starting with a test for silicon [13], and more comprehensive models for aluminum [77–80].

Machine-learning methods and novel molecular representations and descriptors are also used for other regression tasks for molecules, using a variety of approaches to predict, e.g., atomization energies, atomic charges, NMR shifts, etc., [81–98], constructing molecular force fields [99–106], and even in combination with quantum mechanics/molecular mechanics (QM/MM) embedding methods [107].

C. A general potential

Here, we demonstrate that, using ML techniques, it is indeed possible to develop an accurate potential that spans a wide range of physically important structures and properties. Using silicon as an example, we create a potential and demand that it give reasonably accurate predictions for all configurations relevant to scientific questions within a wide temperature and pressure range, including surfaces, point and line defects, cracks, etc. Silicon is a good material for such a study for a number of reasons. First, it has a rich phase diagram with many stable and metastable crystal structures, as well as a wide variety of point and line defects and surface reconstructions. Second, the simulation community has extensive experience in understanding many aspects of its potential energy surface. Finally, there are a large number of empirical analytical potentials that have been constructed over the past decades whose successes and failures give a detailed picture of what it is about the potential energy surface that is relatively easy to get right and what the more difficult aspects are.

Indeed, many advances in materials simulation methodology over the past decades have been demonstrated first using silicon. Some of these are new approaches where silicon is used as a test system, including the Car-Parrinello method for *ab initio* molecular dynamics (MD) [108], maximally localized Wannier functions for analyzing electronic structure [109], concurrent coupling of length scales combining different simulation methods [110,111], and the learn-on-the-fly method for extending the timescale of *ab initio* MD [112]. Others use these new methods to explain experimentally observed phenomena in silicon,

e.g., using the DFT to study the 7×7 dimer-adatom-stacking-fault reconstruction of the Si (111) surface [113]. Silicon is also extensively used as a model system to understand fracture and, in particular, the interplay between brittle and ductile failure [114–118].

A large number of interatomic potentials have been developed for silicon with the intent of describing its bulk phases and defects. While there are too many publications to thoroughly review here, we discuss the most widely used and successful ones, to motivate our choices for comparison models in this work. By far the two most commonly used are those of Stillinger and Weber (SW) [119,120] and Tersoff [2,121–124]. Both include pair terms and three-body terms, the former defined in terms of bond lengths and bond angles, the latter in terms of a repulsive core and a bond-order-dependent attractive bonding interaction. Many other functional forms are also used, as reviewed, e.g., by Balamane, Halicioglu, and Tiller [125] and more recently by Purja Pun and Mishin [126]. While none produce sufficient improvement to lead to a significant adoption by the simulation community, recent attempts to add terms that depend on more than three-body interactions have been at least somewhat successful. These include the environment-dependent interatomic potential (EDIP) [127], modified embedded atom method (MEAM) [128,129], ReaxFF [115,130], and screened Tersoff [131,132]. EDIP uses the local coordination of each atom to approximate a bond order (a chemical concept that is also integral to the Tersoff potential) and change the preferred bond length, strength, and bond angle correspondingly. MEAM is an angle-dependent functional form that evolved out of the simpler embedded atom method, mainly used for metals. It was first applied to silicon with a first-neighbor cutoff distance by Baskes, Nelson, and Wright [133] and later reparametrized several times with different choices of fitting quantities and interaction cutoffs [129,134,135]. Here, we use the second-neighbor cutoff parametrization due to Lenosky *et al.* [129]. The ReaxFF form was originally developed in the context of computational chemistry to describe reactions of molecules, and the silicon potential we use [130] was previously used to simulate a brittle fracture [115]. The screened Tersoff form (TersoffScr) was developed by Pastewka *et al.*, who modified the Tersoff functional form with a screening term to improve its performance for fracture properties [131,132], where bonds are broken and formed. Finally, Purja Pun and Mishin took the modified Tersoff form developed by Kumagai *et al.* [136] and optimized it for a wide range of properties [126]. We compare the results of GAP to these interatomic potential models (EDIP, MEAM, Purja Pun, ReaxFF, SW, Tersoff, and TersoffScr) and also to the density-functional tight-binding (DFTB) method [137–139].

The inclusion of a tight-binding (TB) model in the above list is essential, because TB represents a middle ground between the DFT and interatomic potentials. The TB

approach is a minimal description of electronic structure [1], significantly cheaper than the DFT, yet still carrying the essentially quantum-mechanical nature of the electrons, giving a qualitatively robust description of their behavior in solids in a wide range of materials. Like interatomic potential models, TB can be easily implemented with a cost that is linear in the number of atoms. A lot of effort has gone into making accurate TB models [138,140–150], and, if they were clearly more accurate or transferable than conventional interatomic potentials, they might present the same trade-off between speed and accuracy as ML models, which are also significantly more computationally expensive than conventional interatomic potentials. However, such does not appear to be the case: The widely used TB model included here does not perform better on the whole than analytical potentials. For the DFTB calculations, we use the “pbc” parameter set and a k -point density of 0.007 \AA^{-1} for bulk configurations and 0.04 \AA^{-1} for others. To reduce the computational cost, no charge-self-consistency iterations are performed, since they are not expected to lead to substantial differences for this monatomic covalently bonded material.

Our focus is on creating *useful* models, and therefore the guiding principle is to create a data set and fitting protocol that is the *least* specific to the material and the observables as possible while still achieving the aims. On the one hand, we try to create tests that are as relevant to the materials modeler as possible, focusing on observables that are either directly comparable to experiments or at least generally agreed to be important for the understanding of material behavior. On the other hand, we think of the ML potential as an interpolation scheme for the reference DFT method, so, with a few exceptions, we compare the interatomic potential results to the DFT, rather than to an experiment, for example.

It is worth noting that the comparisons with analytical potentials we show in this paper are not meant as a definitive evaluation of their accuracy. Since the analytical potentials are fit to different sets of properties from different sources, their performance for any particular observable could very well be improved somewhat by refitting them to our DFT data; we do not see the relevance of doing that here and work with the published parametrizations, since the analytical potentials’ main advantages are simplicity, computational efficiency, and some transferability rather than ultimate accuracy. In the case of a machine-learned model, the unique selling point is the accuracy with which the target potential energy surface is matched, which is best demonstrated by comparing to DFT results. The route to an improved agreement with experimental observables is to improve the target, i.e., using a more accurate description of the electronic structure.

In contrast to many earlier works on materials modeling and machine learning, we do not emphasize learning outcomes in the statistical sense, using splits of the data

set into training, testing, and validation. This is partly because this has been done many times before for the same kernel, representation, and approach to parameter choice that we use here and partly because the SOAP kernel does not have hyperparameters that are worth optimizing: They are dictated by the length and energy scales inherent in atomic interactions, which are well known. (The other parameters in the regression correspond to accuracy targets on a few classes of configurations.) Ultimately, the paper is about the validation on material properties, and all those are based on atomic configurations which are themselves not in the training set.

Figure 1 provides an overview of many of the verification and validation tests carried out for our new silicon

GAP model in comparison to the empirical analytical models mentioned above. While the individual tests are discussed in more detail below, we present an overview here. The first three groups of quantities in the figure are verification tests, in the sense that they require accuracy on configurations which are directly represented in the training set. These are split into three classes of test: bulk properties, surfaces, and point defects. Bulk properties, namely, the bulk modulus B and diamond cubic elastic constants C_{11} , C_{12} , and C_{44} , are well reproduced by the GAP model with fractional errors relative to the DFT of less than 10%; none of the other interatomic potentials reach this accuracy, although in many cases they are fit to different training data (e.g., an experiment or simply other exchange-correlation

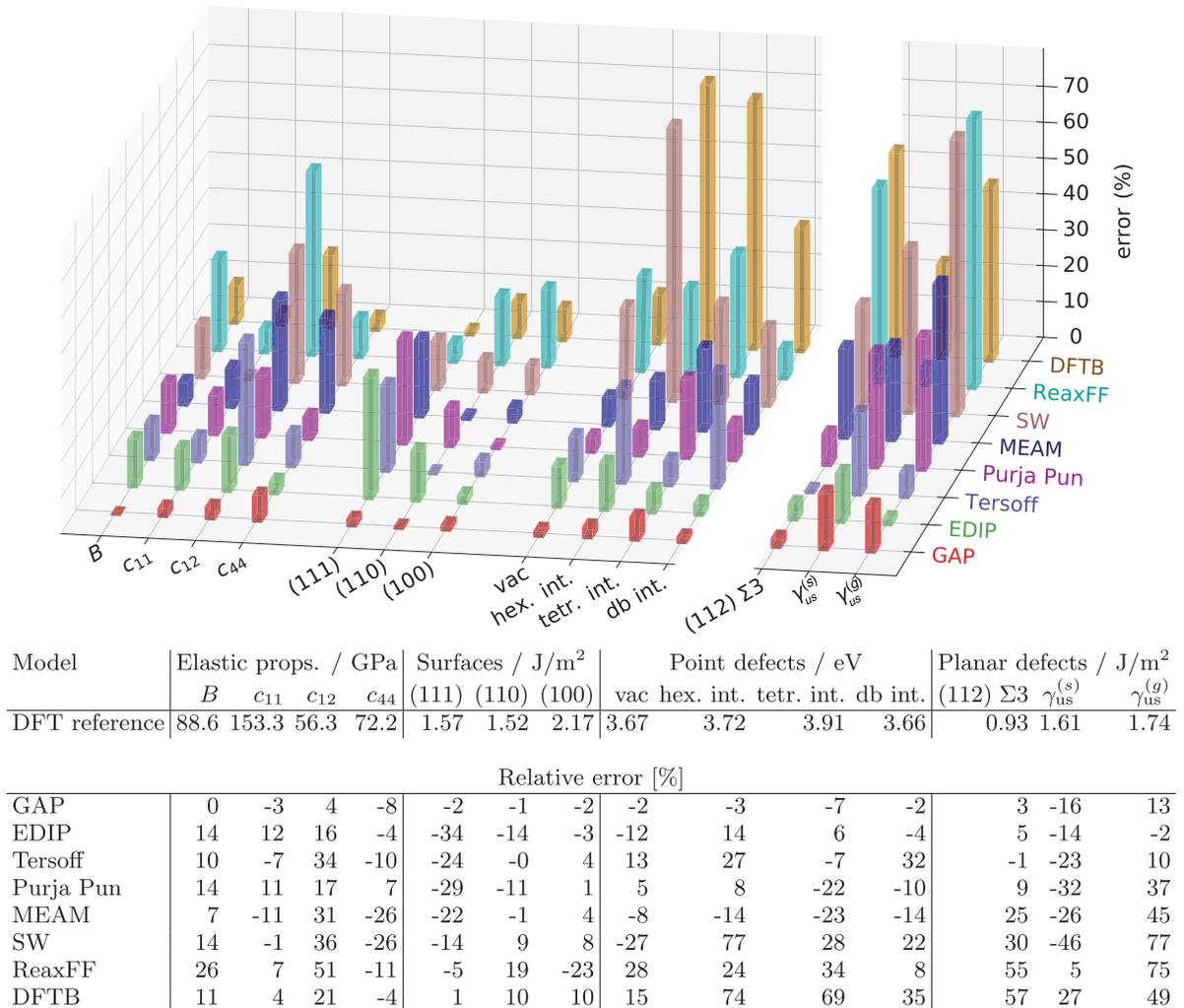


FIG. 1. Comparison of percentage errors made by a range of interatomic potentials for selected properties, with respect to our DFT reference. Those on the left of the break in the axis are interpolative, i.e., well represented within a training set of the GAP model: elastic constants (bulk modulus B , stiffness tensor components C_{ij}), unreconstructed (but relaxed) surface energies [(111), (110), and (100) low-index surfaces], point-defect formation energies (vacancy and hexagonal, tetrahedral, and dumbbell interstitials); while the planar defects to the right are extrapolative: (112) $\Sigma 3$ symmetric tilt grain boundary and unstable stacking-fault energies on shuffle plane $\gamma_{us}^{(s)}$ and glide plane $\gamma_{us}^{(g)}$. The first row in the corresponding table shows reference quantities computed with the DFT (units indicated in the header row).

functionals). The largest relative errors in bulk properties are typically made in the softest elastic constant C_{12} , with the EDIP model being the next most accurate after our new GAP model. The second class of verification tests demonstrates that the GAP model performs consistently at describing surface energies of the (111), (110), and (100) cleavage planes, with errors of around 2% with respect to our reference DFT calculations. Here, the scatter across the various other models is smaller than for bulk properties. For example, the (100) surface energy is, in general, well described by most models. For the third class of verification tests, formation energies of vacancy and interstitial point defects, we see a wide range of errors across the models evaluated. The new GAP model again predicts all these quantities within 10% of the reference DFT results. In general, for any particular property, there is often a model that provides an accurate description, but apart from the GAP, we are not aware of any model that provides uniform accuracy across the whole range of properties in Fig. 1. The typical spread between DFT exchange-correlation functionals for structural and vibrational properties of silicon is much smaller; e.g., Ref. [151] reports only around a 1% variation in Si lattice constants and fundamental phonon frequencies over a wide range of exchange-correlation functionals.

Moving to more stringent tests of the new model, we consider a set of planar defects which are not represented in the training set (right-hand group in Fig. 1), namely, the (112) Σ 3 symmetric tilt grain boundary, and unstable stacking-fault energies on the (111) shuffle plane $\gamma_{us}^{(s)}$ and (111) glide plane $\gamma_{us}^{(g)}$. For these tests, the accuracy of the GAP model is reduced but still within 20% of the DFT, comparing favorably with all other models, some of which include stacking-fault values in their training sets (e.g., EDIP). Moreover, the ability of the GAP model to provide an estimated error along with its predictions allows us to qualitatively assess the expected reliability of the model for particular classes of configurations. Figure 2

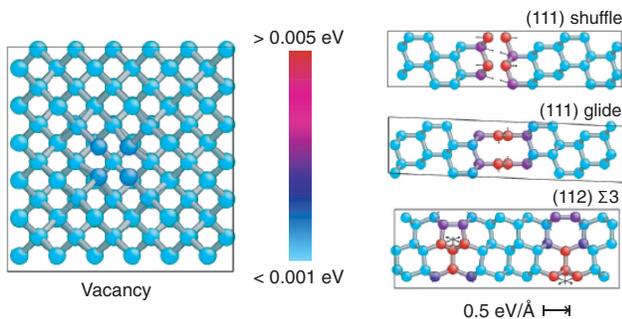


FIG. 2. Visualization of vacancy, (111) shuffle and glide unstable stacking faults, and (112) Σ 3 grain boundary configurations. Atoms are colored by the per-atom energy errors predicted by the GAP on the DFT-relaxed configurations, and arrows show the nonzero GAP forces.

shows the predicted errors for each atom in the vacancy, shuffle, glide, and grain boundary configurations. For the vacancy, the confidence of the model is high on all atoms (blue), and the corresponding accuracy with respect to DFT is high. The reduced confidence close to the planar defects (red atoms) is consistent with the larger errors made for these configurations and the fact that the database does not include any similar atomic environments.

The rest of the paper is organized as follows. In Sec. II, we give an overview of the potential fitting methodology and the construction of the database. In Sec. III, we report on extensive tests that serve to verify that those properties which the database is explicitly designed to capture are indeed correctly predicted. This testing includes equations of state, average structural properties of liquid and amorphous states, point-defect energetics, surface reconstructions, and crack-tip geometries. In Sec. IV, we validate the model by showing predictions for properties that are deemed fundamental for modeling this material but for which the database makes no special provision. These predictions include thermal expansion, di-interstitials, grain boundaries, and random structure search. We finally give a brief outlook in Sec. V.

II. METHODOLOGY

A. Potential fitting

The interatomic potential, even after assuming a finite interaction radius, is a relatively high dimensional function, with dozens of atoms affecting the energy and force on any given atom at the levels of tolerances we are interested in (around a meV per atom). However, much of the interaction energy (in absolute magnitude) is captured by a simple pair potential, describing exchange repulsion of atoms at close approach and potentially the chemical bonding in an average sense farther out. In anticipation of the kernel approach for fitting the interatomic potential, the pair potential also serves a useful purpose from the numerical-efficiency point of view, because the exchange repulsion it takes care of is a component of the potential that is very steep, in comparison with the bonding region, and such disparate energy scales are difficult to capture with a single kernel in high dimensions.

In the present case, we choose a purely repulsive pair potential, given by cubic splines that are fitted to the interaction of a pair of Si atoms, computed using DFT. This choice leaves the description of the attractive part entirely for the many-body kernel fit.

We start by giving a concise account of the GAP kernel-fitting approach, as we use it here. The total GAP model energy for our system is a sum of the predefined pair potential and a many-body term which is given by a linear sum over kernel basis functions [64]:

$$E = \sum_{i<j} V^{(2)}(r_{ij}) + \sum_i \sum_s^M \alpha_s K(\mathcal{R}_i, \mathcal{R}_s), \quad (1)$$

where i and j range over the number of atoms in the system, $V^{(2)}$ is the pair potential, r_{ij} is the distance between atoms i and j , K is a kernel basis function defined below, and \mathcal{R}_i is the collection of relative position vectors corresponding to the neighbors of atom i which we call a *neighborhood*. The last sum runs over a set of M *representative* atoms, selected from the input data set, whose environments are chosen to serve as a basis in which the potential is expanded (more on this below).

The value of the kernel quantifies the similarity between neighborhoods (in the Gaussian process literature, it is a covariance between values of the unknown function at different locations), which is largest when its two arguments are equal and smallest for maximally different configurations. The degree to which the kernel is able to capture the variation of the energy with the neighbor configuration determines how efficient the above fit is. The better the correspondence, the fewer representative configurations are needed to achieve a given accuracy. It also helps tremendously if exact symmetries of the function to be fitted are already built into the form of the kernel. For an interatomic potential, we need a kernel that is invariant with respect to the permutation of like atoms and 3D rotations of the atomic neighborhood. Note that translational invariance is already built in, because the kernel fit is applied to each atom individually—this very natural decomposition of the total energy is customary when fitting interatomic potentials and is directly analogous with the spatial decomposition of convolutional neural networks [152].

Here, we use the SOAP kernel [7,8]. We start by representing the neighborhood \mathcal{R}_i of atom i by its *neighbor density*,

$$\rho_i(\mathbf{r}) = \sum_{i'} f_{\text{cut}}(r_{ii'}) e^{-(\mathbf{r}-\mathbf{r}_{i'})/2\sigma_{\text{atom}}^2}, \quad (2)$$

where the sum ranges over the neighbors i' of atom i (including itself), f_{cut} is a cutoff function that smoothly goes to zero beyond a cutoff radius r_{cut} , and σ_{atom} is a smearing parameter, typically 0.5 Å. Invariance to rotations is achieved by constructing a Haar integral over the SO(3) rotation group [7]. The SOAP kernel between two neighbor environments is the integrated overlap of the neighbor densities, squared, and then also integrated over all possible 3D rotations:

$$\tilde{K}(\mathcal{R}_i, \mathcal{R}_j) = \int_{\hat{R} \in \text{SO}(3)} d\hat{R} \left| \int d\mathbf{r} \rho_i(\mathbf{r}) \rho_j(\hat{R}\mathbf{r}) \right|^2. \quad (3)$$

To obtain the final kernel, we normalize and raise to a small integer power:

$$K(\mathcal{R}_i, \mathcal{R}_j) = \delta^2 \left| \frac{\tilde{K}(\mathcal{R}_i, \mathcal{R}_j)}{\sqrt{\tilde{K}(\mathcal{R}_i, \mathcal{R}_i) \tilde{K}(\mathcal{R}_j, \mathcal{R}_j)}} \right|^\zeta \quad (4)$$

with $\zeta = 4$ in the present case. The δ hyperparameter corresponds to the energy scale of the many-body term, and we use $\delta = 3$ eV, commensurate with typical atomization energy per atom. The accuracy of the fit is not particularly sensitive to this parameter.

In practice, we do not evaluate the above integrals directly but expand the neighbor density in a basis of spherical harmonics $Y_{lm}(\hat{\mathbf{r}})$ and radial functions $g_n(r)$ (we use equispaced Gaussians, but the formalism works with any radial basis):

$$\rho_i(\mathbf{r}) = \sum_{nlm} c_{nlm}^i Y_{lm}(\hat{\mathbf{r}}) g_n(r). \quad (5)$$

The following spherical power spectrum vector (henceforth termed the ‘‘SOAP vector’’) is a unique rotationally and permutationally invariant description of the neighbor environment,

$$\tilde{\mathbf{p}}_{nl}^i = \sum_{m=-l}^l c_{nlm}^{i*} c_{nlm}^i \quad (6)$$

$$\mathbf{p}^i = \tilde{\mathbf{p}}^i / |\tilde{\mathbf{p}}^i|, \quad (7)$$

and the SOAP kernel can be written as its scalar product:

$$K(\mathcal{R}_i, \mathcal{R}_j) = \delta^2 |\mathbf{p}^i \cdot \mathbf{p}^j|^\zeta. \quad (8)$$

The coefficients α_s in Eq. (1) are determined by solving a linear system that is obtained when available data are substituted into the equation, as we detail below. In the present case, these data take the form of total energies and gradients (forces and stresses) corresponding to small- and medium-sized periodic unit cells, calculated using the density-functional theory.

We also need an algorithm to select the set of representative environments over which the sum in Eq. (1) is taken. This selection could be done by simple random sampling, but we find it advantageous to use this freedom to optimize the interpolation accuracy. One approach to this optimization is to maximize the dissimilarity between the elements of the representative set [153] such that the small number of environments best represents the variety of the entire set. Here, we use a matrix reconstruction technique called CUR matrix decomposition [154] and apply it to the rectangular matrix formed by the concatenation of SOAP vectors corresponding to all the neighbor environments appearing in the input data. The CUR decomposition leads to a low-rank approximation of the full kernel matrix using only a subset of its rows and columns [155].

There are two factors that complicate the determination of the vector of linear expansion coefficients, α . The first is that atomic energies are not directly available from the density-functional theory, and the second is the presence of gradients in the input data. The following treatment addresses both of these. We denote the number of atoms in the input database with N and define \mathbf{y} as the vector with D components containing the input data—all total energies, forces, and virial stress components in the training database— \mathbf{y}' as the vector with N components containing the *unknown* atomic energies of the N atomic environments in the database, and \mathbf{L} as the linear differential operator of size $N \times D$ which connects \mathbf{y} with \mathbf{y}' such that $\mathbf{y} = \mathbf{L}\mathbf{y}'$. After selecting M representative atomic environments (with $M \ll N$), the regularized least-squares solution for the coefficients in Eq. (1) is given by [156,157]

$$\alpha^* = [\mathbf{K}_{MM} + (\mathbf{L}\mathbf{K}_{NM})^T \Lambda^{-1} \mathbf{L}\mathbf{K}_{NM}]^{-1} (\mathbf{L}\mathbf{K}_{NM})^T \Lambda^{-1} \mathbf{y}, \quad (9)$$

where K_{MM} is the kernel matrix corresponding to the M representative atomic environments [with matrix elements from Eq. (8)], K_{NM} is the kernel matrix corresponding to the representative set and all of the N environments in the training data, and the elements of the diagonal matrix Λ^{-1} represent weights for the input data values. The Bayesian interpretation of the inverse weights is expected errors in the fitted quantities. While taking $\Lambda = \sigma_v^2 \mathbf{I}$ with an empirical value for σ_v would be sufficient to carry out the fit, this interpretation makes it straightforward to set sensible values. The expected errors are not just due to a lack of numerical convergence in the electronic-structure calculations but also include the *model error* of the GAP representation, e.g., due to the finite cutoff of the local environment. Our informed choices for these parameters are reported in Table I.

For several systems below, we include results on the *predicted error*, the measure of uncertainty intrinsic to our interpolated potential energy surface. These come from the Bayesian view of the above regression procedure, in which the data (and the predicted values) are viewed as samples from a Gaussian process whose covariance function is the chosen kernel function [64]. The mean of this Gaussian process is, of course, just the second term of the predicted energy, Eq. (1), and the predicted variance of the atomic energy for atom i is given by

$$K(\mathcal{R}_i, \mathcal{R}_i) - \mathbf{k}^T (\mathbf{K}_{MM} + \sigma_e \mathbf{I})^{-1} \mathbf{k}, \quad (10)$$

where the element s of the vector \mathbf{k} is given by $K(\mathcal{R}_i, \mathcal{R}_s)$, the covariance between the environment of atom i and the environments of the representative atoms s in the database. The above is a simplified error estimate, in which we regularize using the parameter σ_e , typically set to 1 meV [equal to the value used for the per-atom energy data

components of Λ for most of the database in Eq. (9)] rather than using the more complicated regularization as in Eq. (9). We interpret this variance as the (square of the) “one-sigma” error bar for the atomic energies.

B. Database

The database of atomic configurations (periodic unit cells) is described in Table I. It was built over an extended period, using multiple computational facilities. The kinds of configuration that we include are chosen using intuition and past experience to guide what needs to be included to obtain good coverage pertaining to a range of properties. The number of configurations in the final database is a result of somewhat *ad hoc* choices, driven partly by the varying computational cost of the electronic-structure calculation and partly by observed success in predicting properties, signaling a sufficient amount of data. Each configuration yields a total energy, six components of the stress tensor, and three force components for each atom. The database therefore has a total of 531 710 pieces of electronic-structure data. We represent the diversity of atomic neighborhoods using $M = 9000$ representatives [see Eq. (1)], and the number of these picked from each of the structure types by the CUR algorithm is also shown in the table.

We use the CASTEP software package [158] as our density-functional-theory implementation, and manual cross-checking is done to ensure that the calculations are consistent between different computers. The main parameters of the electronic-structure calculation are as follows: PW91 [159] exchange-correlation functional (the choice is motivated by the existence of a large-scale simulation of the melting point with this functional), 250 eV plane-wave cutoff (with finite basis corrections), Monkhorst-Pack k -point grids with 0.03 \AA^{-1} spacing (corresponding to a 60^3 grid in the primitive cell), ultrasoft pseudopotentials, and 0.05 eV smearing of the electronic band filling. The remaining numerical error is dominated by the finite k -point grid, leading to errors on the order of a few meV per atom. The reference data for testing purposes are calculated with the parameters kept the same, except for bulk energy-volume curves, which use a k -point spacing of 0.015 \AA^{-1} ; the reoptimization of IP minima of amorphous configurations (Table II) which use a k -point spacing of 0.07 \AA^{-1} ; and molecular dynamics of the liquid, whose parameters are given further below. The high k -point densities above might surprise some readers, but we find it necessary in order to converge values of the virial stress.

While we focus our efforts here on testing the GAP for its predictions for scientifically interesting observables, we also evaluate the global distribution of force errors relative to DFT calculations. The results for all the potentials evaluated on the GAP-fitting database, as well as for the GAP on a simple testing database (distinct from the fitting database), are shown in Fig. 3. The GAP shows much lower

TABLE I. Summary of the database for the silicon model. The first column shows the number of atoms in the periodic unit cells, and the second column shows the number of such unit cells in the database, while the third column is the product of the first two and, thus, shows the number of atoms (and, therefore, atomic environments) in the database for each structure type. The fourth column shows the number of representative atoms picked automatically from each structure type by the CUR algorithm (see the text). The last three columns show the regularization we use in the linear system (empty rows correspond to using the defaults, given at the top).

Structure type	No. atoms	No. structures	No. environments	No. representative atoms	σ_{energy}	σ_{force}	σ_{virial}
					default values:		
					0.001	0.1	0.05
Isolated atom	1	1	1	1			
Diamond	2	104	208	6			
	16	220	3520	53			
	54	110	5940	58			
β -Sn	128	55	7040	92			
	2	60	120	32			
	16	220	3520	51			
	54	110	5940	66			
	128	55	7040	157			
Simple hexagonal	1	110	110	13			
	8	30	240	15			
	27	30	810	42			
	64	53	3392	89			
Hexagonal diamond	4	49	196	7			
bcc	2	49	98	40			
bc8	8	49	392	66			
fcc	4	49	196	46			
hcp	2	49	98	28			
st12	12	49	588	94			
Liquid	64	69	4416	1114	0.003	0.15	0.2
	128	7	896	323			
Amorphous	64	31	1984	231	0.01	0.2	0.4
	216	128	27 648	1719			
Diamond surface (001)	144	29	4176	514			
Decohesion	32	11	352	28			
Diamond surface (110)	108	26	2808	338			
Decohesion	16	11	176	8			
Diamond surface (111)							
Decohesion	24	11	264	10			
Unreconstructed	96	47	4512	573			
Adatom	146	11	1606	62			
Pandey reconstruction	96	50	4800	632			
DAS 3×3 unrelaxed	52	1	52	6			
Diamond vacancy	63	100	6300	168			
	215	111	23 865	405			
Diamond divacancy	214	78	16 692	416			
Diamond interstitial	217	115	24 955	605			
Small (110) crack tip	200	7	1400	130			
Small (111) crack tip	192	10	1920	185			
Screw dislocation core	144	19	2736	124			
sp ² bonded	8	51	408	61			
sp bonded	4	100	400	392	0.01	0.2	0.4
Total		2475	171 815	9000			

force errors than any other potential tested, with a median of about 0.025 eV/Å, an order of magnitude smaller than for the analytical potentials. The testing database, which consists of a grain boundary, six di-interstitials, the unrelaxed and relaxed shuffle and glide generalized stacking-

fault paths, and an amorphous configuration, shows a very similar distribution of the force error, although the actual errors are strongly dependent on the type of geometry, so changing the proportions of each could change the resulting distribution somewhat.

TABLE II. Coordination statistics c_i (fraction of atoms with number of neighbors i within $r_c = 2.75 \text{ \AA}$, in percent) and energy per atom relative to the diamond structure (ΔE_{ac} , in eV) for amorphous structures resulting from quenching of the liquid. The energy difference evaluated using the interatomic potential is $\Delta E_{\text{ac}}^{\text{IP}}$, the energy difference of the interatomic-potential-relaxed structure evaluated (but not relaxed) using the DFT is $\Delta E_{\text{ac}}^{\text{eD}}$, and the DFT-evaluated energy difference of the DFT-relaxed structure starting from the interatomic potential structure is $\Delta E_{\text{ac}}^{\text{rD}}$. Most atoms in the MEAM structure have coordination ≥ 6 . The experimental defect density is from Ref. [160] and the energy from Ref. [161].

Model	c_3	c_4	c_5	$\Delta E_{\text{ac}}^{\text{IP}}$	$\Delta E_{\text{ac}}^{\text{eD}}$	$\Delta E_{\text{ac}}^{\text{rD}}$
Lit. exper.		≥ 99		0.137		
GAP	1.4	98.1	0.5	0.15	0.14	0.13
EDIP	0.5	94.4	5.1	0.22	0.22	0.19
Tersoff	0.0	98.1	1.9	0.22	0.18	0.17
MEAM	0.0	0.0	2.8	0.14	0.65	0.28
SW	2.3	75.5	21.8	0.20	0.29	0.23
ReaxFF	0.0	86.1	13.9	0.35	0.35	0.25

Note that the testing database for Fig. 3 is not the result of the usual random split into training and test sets but represents extrapolation into configurations entirely different from those in the training database, providing an even more stringent test than the usual split. Since the empirical analytical potentials are not fit to our database, the latter serves as a test for the potentials. It is remarkable how good the analytical potentials' predictions are for macroscopic properties, which are mostly energy differences, given the large force errors shown here.

C. Convergence

Since the principal goal of machine-learned interatomic potentials is to enable the prediction material properties by fitting the Born-Oppenheimer potential energy surface, it is interesting to consider the *convergence* of such a potential. The expectation is that a closer match of the potential energy surface will result in more accurate predictions. While a comprehensive convergence study is beyond the scope of this work, there are simple convergence

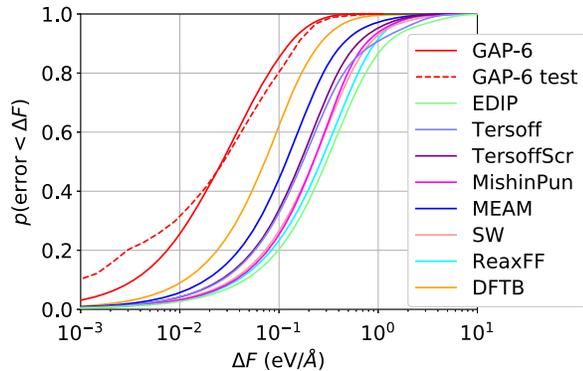


FIG. 3. Cumulative probability distribution of force component errors (relative to reference DFT calculations) for all potentials evaluated on the GAP model-fitting database (solid lines) and for the GAP model only on a separate testing database (dashed line).

parameters in the SOAP-GAP framework that directly control the trade-off between computational cost and accuracy of the fit. One is the number M of representative environments (effectively the number of basis functions in the regression), and the other is the truncation of the spherical harmonic and radial basis expansion of the atomic neighbor density [Eq. (5)]. Figure 4 shows the convergence of the SOAP or GAP model with respect to these. We use the Δ value of Lejaeghere *et al.* [162] to compute the error in the energy-volume curves for diamond and β -Sn with respect to our DFT reference, defined as

$$\Delta = \sqrt{\frac{\int_{0.94V_0}^{1.06V_0} [E^{\text{GAP}}(V) - E^{\text{DFT}}(V)]^2 dV}{0.12V_0}},$$

where E^{GAP} and E^{DFT} denote GAP and DFT energies, respectively, relative to the diamond energy minimum to allow comparison, V_0 is the DFT minimum-energy volume for each phase, and the integral is computed numerically by fitting cubic splines to 12 (E, V) pairs for each model. Good convergence can be seen with respect to both basis set size and the accuracy of the expansion of the atomic density, with a precision of the order of a meV (for β -Sn; for diamond, another order of magnitude better), indicating that GAP reproduces the target DFT energy surface better than the typical variability between DFT codes of the order of $\Delta = 1 \text{ meV}$ reported in Ref. [162].

In principle, a Gaussian process regression model should be able to converge to a given target function with arbitrary accuracy as the database size grows. However, in this case, the only remaining physical approximation is the finite cutoff of the interatomic potential, which means that the force on an atom that is computed using our DFT engine is not strictly a function of the finite neighborhood of the atom. From the point of view of a model with a finite cutoff, the target function appears to have a finite amount of uncertainty, and this uncertainty is taken into account when

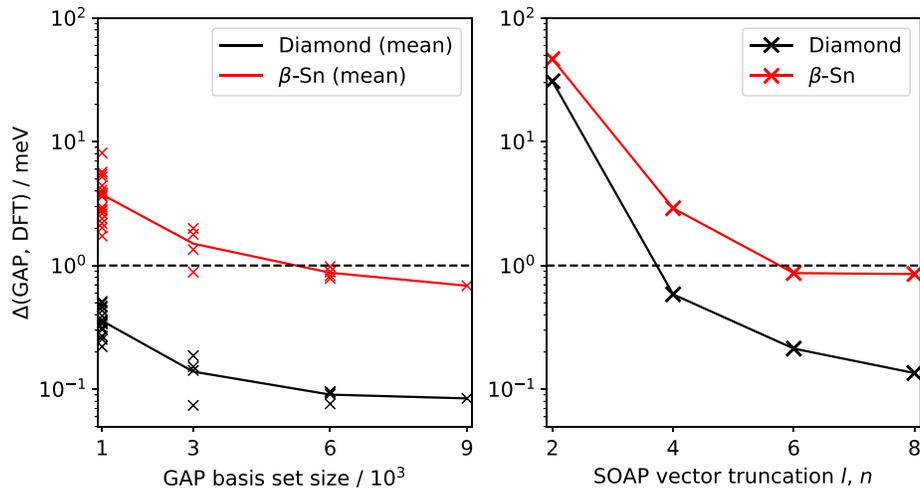


FIG. 4. Error of the SOAP or GAP model based on the Δ value of Ref. [162] with respect to the DFT for diamond (black) and β -Sn (red) structures. The left shows the error as a function of the number of basis functions used in the Gaussian process regression. Because we use a stochastic algorithm to select which basis functions to use, multiple models are shown, which differ only in the pseudorandom seed. The right shows the error as a function of the length of the SOAP vector, which in our case is controlled by the truncation of the radial and spherical harmonic expansion of the atomic neighbor density. The horizontal dashed line corresponds to 1 meV per atom error, our default energy accuracy target.

fitting the model, as mentioned above. Indeed, previous investigations show that, with a cutoff of 5 Å, an error of 0.1 eV/Å on the forces is about what is to be expected for the diamond structure [163–165]. Note that it is possible to estimate the expected force error due to the finite cutoff directly from the DFT engine, because forces are themselves local quantities, as opposed to site energies and virial stress components, which are not observable directly.

It is noteworthy how much more accurate the potential is for the diamond structure than for β -Sn. Two factors contribute to this: First, there are many more diamondlike configurations in the database, particularly the configurations associated with various defects, and, second, the locality error is expected to be significantly larger for the β -Sn structure due to its metallic electron density of states. Small clusters represent a class of systems where locality is even worse, due to delocalized surface states and a fluctuating Fermi level, whose effects decay slowly with the inverse of the system size [166]. The type of bonding in small clusters varies wildly as the cluster size changes, with reactive, metallic ground states and inert, closed-shell “magic number” clusters competing energetically [167–169]. This variation suggests that interatomic potentials in general are not well suited to accurately model silicon clusters; hence, we decide to omit them from this study.

We do not claim that our database in the present work is complete in a mathematical sense (even within the restriction of the given cutoff) but that, for any particular application whose relevant configurations are well represented in the database, errors can be improved only by choosing a larger cutoff, which in turn might lead to the need to enlarge the database further.

D. Testing

A software testing framework was built to run tests of the potential using the atomic simulation environment (ASE) [170]. Each model and test is implemented as an independent PYTHON module, allowing all tests to be run with each model (similarly to the design of the OpenKIM project [171]). The model modules are simple, consisting of calls to existing ASE interfaces to QUIP [172] (GAP, DFTB, Stillinger-Weber, Tersoff, and MEAM), LAMMPS [173] (EDIP, Purja Pun, and ReaxFF), and ATOMISTICA [174] (TersoffScr). Reference DFT results are obtained using the same tests with a model based on the ASE interface to Castep and using the parameters discussed above. One advantage of this automated approach is that it ensures consistency in starting configurations, minimization algorithms, and the final test results that are shown in our figures. Another is that it enables automated rerunning of tests when changes are made, e.g., to the GAP training database, allowing incremental improvements to be assessed. The framework is available for download [175].

III. RESULTS: VERIFICATION

In this section, we report on a series of basic tests which the GAP model is designed to pass, because they correspond to configurations that are selected for inclusion in the database for the purpose of describing those very observables. We refer to these as “verification,” by analogy to the usage of the term in software engineering, where it refers to the confirmation that the software implements the specifications correctly.

It is important to note that, by the very nature of such data-driven models, in some sense the database (and the

corresponding models) will never be deemed completely final and definitive. By designating some tests as part of “verification,” we mean to be open about the fact that the database is amended qualitatively and quantitatively until these tests are passed to our satisfaction, and, therefore, these tests are in some sense merely the achievement of a good fit. In the following “validation” section (again by analogy to the use of the term in software engineering where it refers to the confirmation that the specifications describe a method that achieves the desired goal), we collect tests for which the database is not explicitly designed but concern observables that a good model for the material ought to be able to describe. We make no attempt to augment or modify the database in order to improve the results of those tests, and this could, and indeed should, be done in future work.

A. Bulk crystals

As an initial test, we calculate the energy vs volume for a number of bulk crystal structures for silicon, including the ground-state diamond structure, closely related hexagonal diamond, known high-pressure structures β -Sn, simple hexagonal (sh), bc8, and st12 structures, as well as even higher pressure phases, hexagonal close packed (hcp), body-centered cubic (bcc), and face-centered cubic (fcc). When calculating these curves with the DFT as well as DFTB and each interatomic potential, we deform the lattice to the target volume and relax it with respect to the unit cell shape and atomic position while approximately constraining the volume and also constraining the symmetry (using spglib [176]) to remain that of the initial structure.

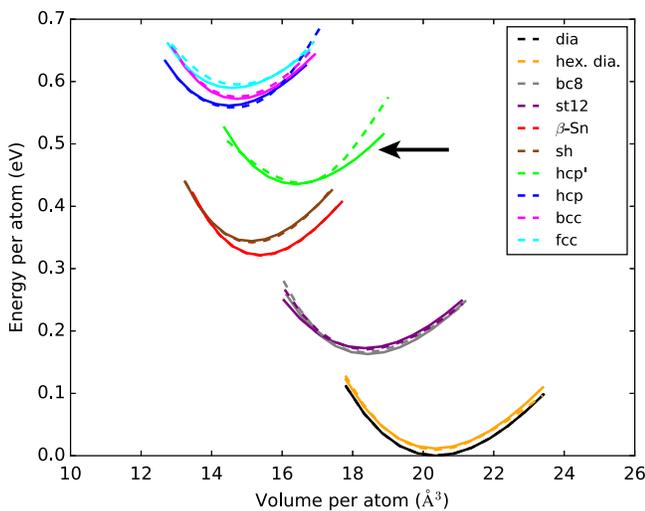


FIG. 5. Energy per atom vs volume per atom for various bulk crystal lattice structures computed using the DFT (solid lines) and GAP (dashed lines). The hcp’ structure (indicated by an arrow), which is not in the fitting database, has a substantially larger discrepancy between the DFT and GAP than any of the other structures, all of which are in the database.

We find that the hcp structure has two minima, the conventional one with $c/a \approx \sqrt{3/2}$ and another we label hcp’, which has a much lower $c/a < 1$.

The resulting $E(V)$ curves for each crystal structure calculated with the GAP and compared to our reference DFT calculations are shown in Fig. 5. The results are in excellent agreement for all structures tested, including minima positions (volume), depths (cohesive energy relative to the ground state), and curvatures (bulk modulus). The hcp’ structure, which is not in the fitting database, has a larger discrepancy than the other structures, although it is still in good agreement. A comparison of all the models for a few selected crystal lattices (diamond structure, β -Sn, and fcc) is shown in Fig. 6. Only the GAP is even qualitatively reproducing all three selected structures, and many of the models fail to reproduce even the first structure seen experimentally under applied pressure, β -Sn.

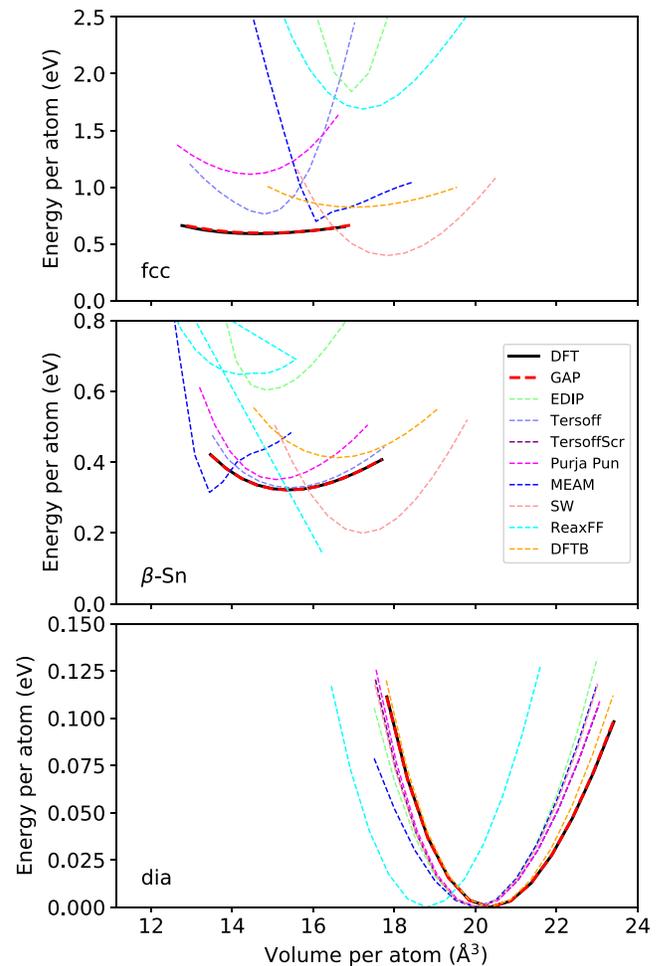


FIG. 6. Energy per atom relative to diamond structure vs volume per atom for fcc (top), β -Sn (middle), and diamond structure (bottom), computed with the DFT (black, solid line), GAP (red dashed line), and all other models (various colors, dashed lines). Note different y-axis ranges in each panel.

B. Liquid

To simulate the structure of liquid silicon with each interatomic potential and DFTB, we use constant-pressure ($P = 0$ GPa) molecular dynamics as implemented in the QUIP package through the quippy PYTHON interface [172]. A $2 \times 2 \times 2$ supercell of the eight-atom diamond cubic cell (64 atoms total) is heated from $T = 0$ K to $T = 5000$ K for rapid melting over 20 000 0.5 fs time steps and then equilibrated at $T = 2000$ K for 10 000 0.25 fs time steps. Structural data are gathered over an additional 5000 0.25 fs time steps. Reference DFT results are obtained from a similar MD simulation using the Castep software, averaging over 9700 0.25 fs time steps at $T = 2000$ K. For the electronic-structure calculations, a 200 eV plane-wave energy cutoff and a $2 \times 2 \times 2$ Monkhorst-Pack [177] k -point grid are used (equivalent to a k -point density of about 0.05 \AA^{-1}). The radial distribution function (RDF) and angular distribution function (ADF) are calculated and averaged using the tools included in QUIP.

The resulting structural quantities are shown in Fig. 7. The GAP RDF is in excellent agreement with the DFT result, including both peak heights and radii at all distances captured in the simulation cell. The DFTB is in comparably good agreement on this structural quantity, and the various interatomic potentials are in much worse agreement, with significant variation among them. The ADF proves to be an even more stringent test. Again, the GAP results are in excellent agreement with the DFT, showing a narrow peak at about 60° and a broader peak with similar height at about 100° . Most of the potentials greatly underestimate the height of the small-angle peak and overestimate the height of the large-angle peak. The only two that are qualitatively correct are EDIP and MEAM, but those both overestimate the depth of the trough separating the two peaks. Several issues with the analytical interatomic potentials may be the source of the differences. Some, e.g., Tersoff [2], greatly overestimate the melting point and are therefore strongly undercooled at $T = 2000$ K rather than an equilibrium liquid. In other cases, it is possible that the wide variety of curves observed is consistent with the hypothesized liquid-liquid phase transition between a high-coordination, high-density metallic phase and a low-coordination, low-density semiconductorlike phase [178]. Some of the potentials may simply be incorrectly predicting the low-coordination phase to be present at $T = 2000$ K and zero pressure, leading to a predominantly tetrahedral-like bond angle distribution.

In addition to the two structural quantities, we evaluate a dynamical quantity, the diffusivity of liquid Si, by carrying out variable cell size constant enthalpy MD simulations using the LAMMPS software [173,179] on a 512-atom cell for 10^5 1 fs time steps at temperatures ranging from about 1700 to 2200 K. The resulting diffusivity as a function of the temperature is shown in Fig. 8 and compared to the experimental results [180], DFT results [181] (using the PBE generalized gradient approximation

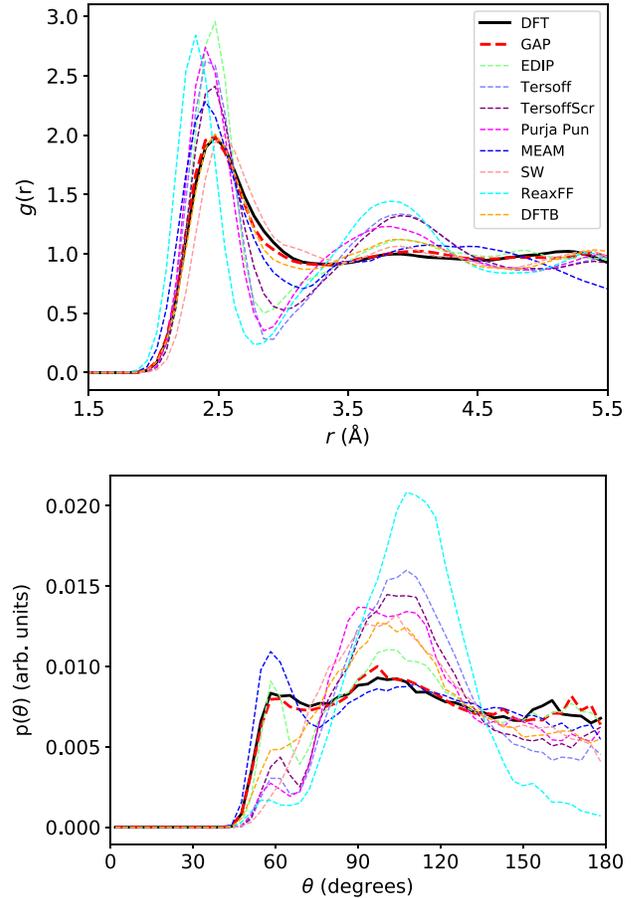


FIG. 7. Liquid silicon radial and angular structure from well-equilibrated constant temperature and pressure 64-atom samples at $P = 0$ GPa and $T = 2000$ K. Top: RDF. Bottom: ADF. The black solid line indicates DFT results, the red dashed line and symbols indicate GAP results, and dashed lines (various colors) indicate the DFTB and other interatomic potentials.

exchange-correlation functional, which is somewhat different from the PW91 functional we use to generate our fitting database), and previously published SW potential results [182–185]. The GAP results are in excellent agreement

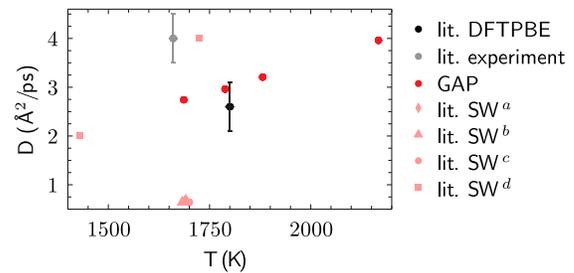


FIG. 8. Diffusivity of liquid silicon from literature DFT simulations [181] (black), literature experiment [180] (gray), GAP (red), and literature SW potential [Refs. [182–185] for (a)–(d), respectively, pink]. Error bars for GAP simulations are smaller than symbols on this scale and are not available for literature SW results.

with the DFT, and so both underestimate the experimental diffusivity. This difference relative to the experiment has previously been ascribed to the tendency for DFT to exaggerate the structure of the liquid [181], and so the similar diffusivities of the GAP and DFT are consistent with the similarities in their liquid RDF and ADF.

C. Amorphous phase

Amorphous silicon is an interesting tetrahedrally coordinated phase that forms upon various forms of processing, including ion implantation, low-temperature deposition, and rapid quenching from the melt. The last of these is commonly used in simulations, but it is challenging to reach experimentally relevant cooling rates using accurate methods such as the DFT. We therefore carry out zero-pressure variable cell volume (hydrostatic strain) simulations of the quenching of a 216-atom sample of liquid Si, cooled at 10^{12} K/s from 2000 to 500 K with a 1 fs time step (1.5×10^6 steps) using the LAMMPS software, and then relax to the local energy minimum with respect to atomic positions and cell size and shape. The initial configuration for all quenches is from a GAP equilibrated liquid at $T = 1800$ K, which is further equilibrated with each potential at $T = 2000$ K for an additional 10^5 time steps before cooling. As for the liquid above, for some potentials this initial thermodynamic state may be a strongly undercooled liquid due to their overestimation of the melting temperature.

The RDFs of the resulting structures are shown in Fig. 9, in comparison with experimental results [186] (since DFT results for comparable sizes or quench rates are not computationally feasible). The various interatomic models vary widely in the overall shape of their RDF, with GAP, EDIP, and Tersoff in best agreement with the experiment, showing a sharp first neighbor peak at about 2.35 \AA

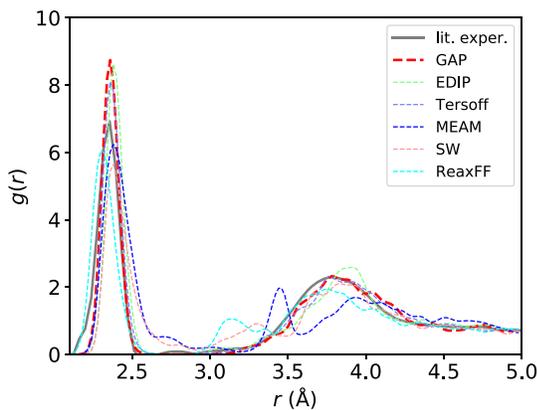


FIG. 9. RDF for a 216-atom amorphous configuration generated by cooling at 10^{12} K/s from 2000 to 500 K and then minimized, for GAP (red) and other interatomic potentials (other colors). Experimental results (gray) generated by ion implantation from Ref. [186] are shown for comparison.

and a broad second peak at about 3.8 \AA . These three models have essentially no atoms between the two peaks ($2.5 \text{ \AA} \lesssim r \lesssim 3.25 \text{ \AA}$). The other models show various qualitative problems, including smaller peaks between the two expected ones or an excess of atoms throughout the entire distance range between the first and second neighbor peaks. The corresponding coordination statistics (using $r = 2.75 \text{ \AA}$ as the nearest-neighbor distance cutoff) are shown in Table II. The GAP and Tersoff models have the lowest coordination defect concentration, significantly lower than the next-best model, EDIP, and closest to the experimental estimates of $\leq 1\%$ [160].

Table II also lists the amorphous-crystal energy difference ΔE_{ac} relative to the diamond structure. The obvious way to evaluate the energy difference for each structure is to use the same interatomic potential that is used to generate the structure, i.e., a calculation that is entirely self-consistent for that potential. This $\Delta E_{\text{ac}}^{\text{IP}}$ listed in the table shows the GAP with the closest value to the experiment [186] (excluding MEAM, which has a very unphysical structure), while other interatomic potentials result in higher energy differences. However, using the potential to evaluate the energy difference risks mixing up errors in the structure with errors in the energy difference given the structure, with the possibility of exaggerating or understating the stability of the amorphous structure, depending on the sign of the energy error. For an independent evaluation of the quality of the quenched a-Si structures, we evaluate their energies with the DFT ($\Delta E_{\text{ac}}^{\text{D}}$) and also further relax them with the DFT ($\Delta E_{\text{ac}}^{\text{rD}}$). Note that these calculations are done with a lower k -point density, 0.07 \AA^{-1} , due to the computational expense of the 216-atom cells. In general, the unrelaxed DFT energy shows a similar trend to the IP energy, except for SW and MEAM, where the IP energy greatly underestimates the (more reliable) DFT energy difference. Relaxing the structure leads to a small energy reduction for GAP as well as EDIP and Tersoff, indicating a structure that is relatively close to the nearest DFT local minimum, but much larger reductions for the other potentials.

All these DFT results show that quenching a liquid with GAP produces the most stable a-Si structure with the lowest energy difference relative to the diamond structure crystal as compared with the other interatomic potentials and that the GAP evaluated energy of this structure is in good agreement with the DFT. Further work at lower quench rates is required to generate structures that can be reasonably argued to be directly comparable to the experiment [187].

D. Phase diagram

The phase behavior corresponding to an interatomic potential is a useful benchmark: It not only informs the user about how realistic the model is but provides an indirect yet stringent test of the microscopic details of the PES.

The phase transitions result from a delicate balance between energetic and entropic effects and, for finite-temperature transitions, probe relatively high-energy configurations. To calculate the liquid-solid transition lines, we perform coexistence simulations for the diamond and simple hexagonal structure at a fixed pressure and enthalpy and measure the resulting average equilibrium temperature [188]. The diamond-liquid simulations contain 432 atoms, and the pressure is fixed at the values of 0, 4, and 8 GPa; the simple hexagonal-liquid system contains 1024 atoms, and the simulations are carried out at 8 and 12 GPa. To estimate the transition line between β -Sn and simple hexagonal phases, we run isothermic-isobaric molecular-dynamics simulations of both pure phases in a temperature range of 0–1000 K and a pressure range of 6–14 GPa and observe the transition (which occurs in both directions in all cases) by monitoring the Steinhardt bond-order parameters [189]. Finally, the transition line between diamond and β -Sn structures is determined by calculating the Gibbs free energy using the quasiharmonic approximation (QHA). We also establish that in these phases anharmonic contributions to the free-energy differences are negligible at 0 K. We use the LAMMPS package for the MD simulations and phonopy [190] for the phonon calculations. Figure 10 shows the calculated phase diagram, compared to the published DFT results for the diamond-liquid melting point [191] and our own calculations with the Castep program for the diamond/ β -tin and β -tin/simple hexagonal transition pressures at 0 K. For comparison, we also show the experimentally determined phase relations [192]. Note that the *Imma* phase is missing from the calculated phase diagram, which is due to the fact that both our DFT calculations and GAP model find the *Imma* phase to be metastable.

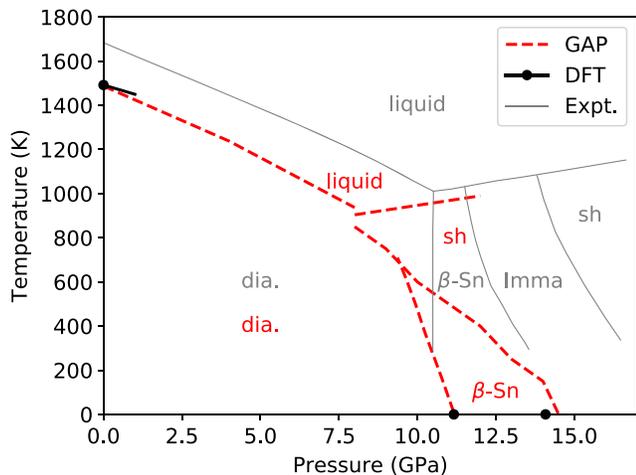


FIG. 10. Temperature-pressure phase diagram of silicon, computed with the GAP (red), compared to available DFT results (black) and experimental phase transitions (gray). The finite-temperature DFT value and slope are from Ref. [191], and the experimental lines are from Ref. [192].

E. Defects

1. Point defects

Several point defects are represented in the fitting database (Table I), and their formation energies would therefore be expected to be accurately reproduced by the GAP. Indeed, as Fig. 1 shows, the relative error for the vacancy and three interstitial positions, hexagonal, tetrahedral, and dumbbell, are all within at most 7% of the reference DFT values. The only other potential that is close to this level of accuracy is EDIP, with similar errors for all but the hexagonal interstitial, where it is off by 14%. All the other potentials, as well as DFTB, differ from our DFT calculations by tens of percent for at least some of the defects.

Since point defects control properties such as diffusivity in bulk silicon, their migration barriers are also of interest and as they represent bond breaking and formation processes, often present a challenge for interatomic potentials. Since the training database configurations come from finite-temperature MD, it could, in principle, include configurations near the barrier, but, since the system spends relatively little time near the energy saddle point, this is actually unlikely [52]. However, the hexagonal and tetrahedral interstitials are related by a short displacement, so one is typically a local minimum and the other a saddle point along an interstitial diffusion pathway. We find that the GAP preserves the DFT ordering, although the energy difference is underestimated, while the other potentials make much larger errors, many reversing the relative order of the two high-symmetry geometries. Two other related observables, the migration path of the vacancy and the formation energy of the fourfold defect [193] (the midpoint of the concerted-exchange diffusion mechanism [194,195]), which are not represented in the database, are discussed below in Secs. IV E and IV F.

2. Surfaces

Surfaces are a class of defects that have particular importance for the behavior of materials. Solids fail under tension by opening new surfaces, and it is on surfaces that reactions involving chemical species in the environment can take place, where special functional layers can form, e.g., by oxidation, and also where a crystal can grow under suitable conditions. Apart from useful applications, a rich complexity of bonding emerges on surfaces due to the subtle interplay of strain effects with the chemistry of dangling bonds. This complexity makes surface formation energies, and particularly the energies and geometries of various reconstructions, a sensitive test of the accuracy of an interatomic potential.

Figure 11 shows the energy as a function of separation as a gap is opened up in a unit cell that is long in one direction and has the dimensions of the minimal surface unit cell in the orthogonal plane. For the purposes of this test, the atomic positions are not relaxed but kept rigid relative to

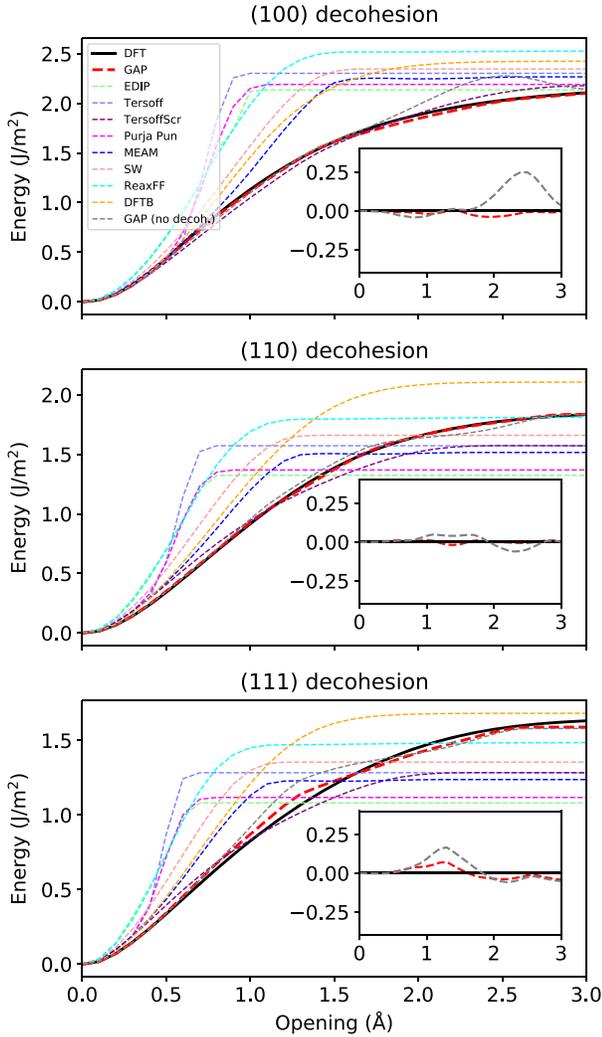


FIG. 11. Decohesion energy of diamond-structure silicon along various directions (labeled according to the orientation of the opening surface). Insets (same axes as the main plot) show errors with respect to the DFT for the current GAP (red) as well as for a previous version of the GAP (gray) with a fitting database that does not include any configurations along the separation path (or high-energy crystal-lattice structures) but only final fully separated surfaces.

one another as the gap is opened. All analytical potentials apart from the screened Tersoff show far too short a range—they plateau much earlier than the DFT, and, in fact, this observation is one of the motivating factors behind modifying the original Tersoff potential [131,132]. The right-hand-side limit corresponds to the unrelaxed surface energy in each case, a property in which the potentials show about 30% scatter. Note that, in the case of the (111) surface, the DFT is believed to overestimate the surface energy [196] and, e.g., Tersoff and its screened version are explicitly fit to reproduce the experimental value.

Note that the final version of the fitting database for the GAP presented here includes configurations along the

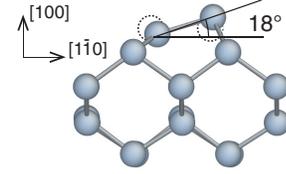


FIG. 12. Geometry of the 2×1 reconstruction of the (100) surface, showing the tilting of the surface dimers. The dashed circles show the surface dimer in the untilted (but rebonded) position.

separation path, in addition to fully separated surfaces. An earlier version of the GAP model [197] that does not include configurations from the separation path correctly reproduces the *fully separated* energy (since fully separated surfaces are included in the fitting) but not the intermediate energies, as shown in the insets in Fig. 11. The test results for the version of the potential without the decohesion path configurations (as well as fewer nondiamond crystal structures), listed in detail in Supplemental Material [198], are very close to the final GAP values, with most tested quantities differing by less than 1%. The only exceptions are the quantities directly related to configurations newly added to the database and a few other tests (described later in this subsection and in Sec. IV G) that are not explicitly fit [two di-interstitial formation energies that change by 3% and 10% and the (111) reconstructed surface energies that change by 2%–3%]. This example shows that the flexibility of the GAP functional form makes it possible to correct shortcomings by adding configurations to the database without significantly affecting accuracy for other configurations.

Figure 12 shows the geometry of the tilted-dimer 2×1 reconstruction, one of the low-energy configurations of the (100) surface, which forms spontaneously from the as-cut surface. In this reconstruction, the surface atoms dimerize to form additional bonds, and the dimers tilt (by 18° in our DFT calculations) due to a Jahn-Teller effect [199], which would seem to require an explicit description of the electrons. In fact, none of the analytical potentials reproduce the substantial tilting (zero tilt for all but EDIP, which tilts by 4°). Only the GAP, with its relatively long-range and flexible form, captures the tilting in reasonable agreement with the DFT (-2.5° error). The DFTB model, with its minimal description of the electronic structure, also shows the breaking of symmetry with a similar error on the resulting bond angle of about -2.3° .

The lowest energy configuration of the (111) surface is the famous 7×7 dimer-adatom-stacking-fault (DAS) reconstruction, already alluded to in the introduction. It is a rather complex structure, involving a 2D superlattice of ten-atom rings, connected by dimerized dislocation cores that separate triangles of stacking faults, half of which have extra atoms on top. A family of analogous structures can be defined by varying the number of dimers, $n \geq 3$, between the vertices of

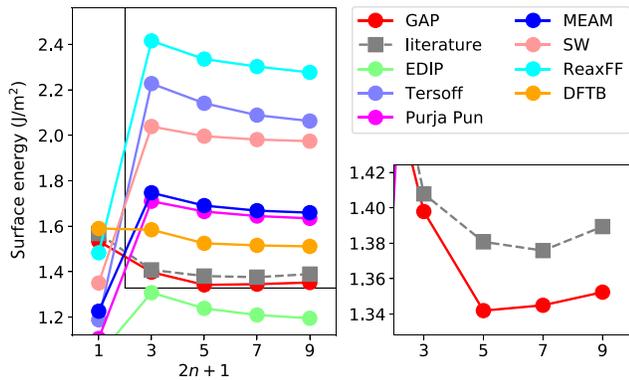


FIG. 13. Formation energy of the dimer-adatom-stacking-fault (DAS) reconstruction of the (111) surface for various surface unit cell sizes $2n + 1 = (3, 5, 7, 9)$ computed with different models. The value shown at $2n + 1 = 1$ corresponds not to a DAS reconstruction but rather to the unreconstructed surface. The box on the lower right is a magnified view that shows just the DFT and GAP results.

the superlattice, leading to the designation $(2n + 1) \times (2n + 1)$. As shown in Fig. 13, all analytical potentials predict these reconstructions to be higher in energy than the unreconstructed surface (shown in place of $n = 1$ for simplicity), and, furthermore, within the family of DAS structures, the energy goes down as n goes up. Computing accurate DFT energies is a nontrivial calculation, and its prediction that the 7×7 DAS structure is the lowest energy configuration is a significant early triumph of the DFT [113]. Here, our reference is a more recent careful determination of the DFT energies [200]. The DFTB model again stands out as qualitatively different from the analytical models but still fails to show quantitative agreement with the DFT. The GAP model, which includes in its training database just a single configuration of the 3×3 DAS structure (shown in Fig. 14), gives energies with an error below 0.05 J/m^2 (much smaller than a meV per atom over the supercell), correctly predicting the DAS family to be lower in energy than the unreconstructed surface and also giving an energy minimum.

The lowest energy structure for the present potential happens to be for $2n + 1 = 5$, within 0.01 J/m^2 of the 7×7 structure. The energy differences are much smaller

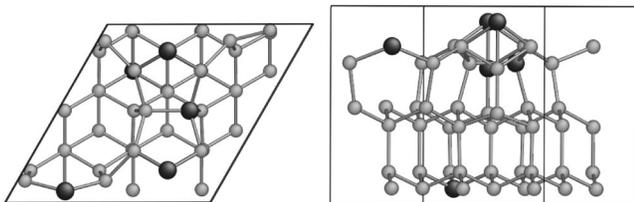


FIG. 14. Two views of the DAS 3×3 (111) surface reconstruction configuration that is in the database with atoms, marked with dark gray, whose environments are selected to be among the representative set for the purposes of defining the GAP model.

than the target (and assumed) error in the GAP model, and as such this level of detail is not robust: The earlier variant of the potential fitted to a slightly different database [in ways unrelated to the (111) surface] show the 7×7 DAS structure as the global minimum, as shown before [197]. What *is* robust is the relationship of the energies of the DAS family to other types of reconstructions and the upturn in energy for $n = 9$. Significantly more data relevant to these structures are needed in order to robustly capture the finest of relative energies within the DAS family.

Figure 14 shows which atoms are picked (automatically, by the CUR decomposition of the SOAP representation matrix, as mentioned above) to be part of the representative set: mostly those that are unique to the DAS family of reconstruction and do not appear elsewhere in the data set, i.e., the adatom, the atom just below it, one of the dimer atoms in at the boundary of the stacking fault, and one atom on the ten ring that surrounds the vertices of the surface unit cell.

F. Crack propagation

The atomic-scale details of crack propagation prove particularly challenging to model, since sufficient accuracy to describe bond-breaking processes must be combined with large model systems to avoid unrealistic strain gradients [201]. Interatomic potentials which provide an otherwise good description of the bulk and elastic properties of silicon (e.g., the Stillinger-Weber and Tersoff potentials) tend to overestimate the lattice-trapping barriers to brittle fracture, resulting in an overestimate of the fracture toughness as well as an erroneously ductile material response including features such as crack arrest and dislocation emission [116,117,202]. Progress has been made using reactive potentials such as ReaxFF [115] or with hybrid quantum or classical approaches where an *ab initio* crack-tip model is embedded within a larger classical model system [203–205]. The latter limits the applicability to timescales accessible to the DFT, making it extremely challenging to study processes such as thermally activated crack growth [206].

To test the accuracy of our new GAP model for fracture, we considered the well-studied (111)[$1\bar{1}0$] cleavage system, where fracture is known to exhibit a low-speed instability triggered by the formation of a crack-tip reconstruction [204]. We perform simulations of a 23496-atom model system of dimensions $600 \times 200 \times 3.86 \text{ \AA}^3$ using both molecular dynamics at 300 K with a range of strain rates between 10^{-6} and 10^{-4} fs^{-1} and quasistatic strain increments followed by relaxation. In all cases, the trajectories obtained are consistent with those expected from our earlier DFT-based hybrid simulations as reported in Ref. [204]. The GAP model predicts brittle fracture morphology with an atomically smooth fracture surface and the occasional formation of crack-tip reconstruction and subsequent surface steps in the “downward” [111] direction, in line with the results of our

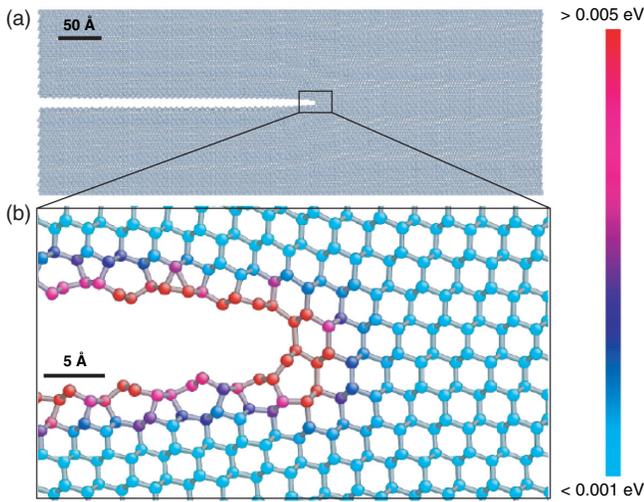


FIG. 15. (a) Snapshot from a quasistatic simulation of fracture in the Si(111)[110] cleavage system at a strain energy release rate of $G = 5.13 \text{ J/m}^2$. The model system contains 23 496 atoms and has dimensions $600 \times 200 \times 3.86 \text{ \AA}^3$. (b) A close-up of the crack tip, which undergoes a crack-tip reconstruction as previously reported in DFT-based hybrid simulations in Ref. [204]. Atoms are colored by the predicted error per atom of the GAP model, from blue (low) to red (high).

previous study. A snapshot from a quasistatic simulation showing the formation of the crack-tip reconstruction at a strain energy release rate of $G = 5.13 \text{ J/m}^2$ is illustrated in Fig. 15.

It is worth noting that atoms at the crack tip have rather unique neighbor environments, and, with nothing nearby in the database, the description of crack tips is not reasonable by an earlier version of the potential. But including just a handful (17) of crack tips in the database (using just an ~ 200 -atom unit cell) already leads to a potential that describes the subtle competition between bond breaking and bond rotation, and qualitatively correct surfaces and crack-tip reconstructions are obtained. The atoms are colored by the predicted error of the GAP model, showing high confidence in the bulk but significant predicted errors at the crack tip, which could be reduced by expanding the database.

IV. RESULTS: VALIDATION

In addition to the tests presented in the previous section, we tested quantities and configurations that are physically important but do not map so cleanly to particular geometries in the database. The first is a random structure search, which probes a very wide range of geometries, bonding topologies, and energies. The second is a test of the vibrational properties (harmonic phonons and anharmonic Grüneisen parameters) of the diamond structure, which are only implicitly included in the fit through the perturbed diamond configurations. Finally, two types of defects are

tested, a high-symmetry grain boundary and di-interstitials, which have geometries related to, but clearly different than, the defects in the fitting database.

A. Random structure search

The random structure search (RSS) [207,208] method provides a global test of the potential energy surface, including not only regions near the physically reasonable minima (i.e., typical bulk lattices with small distortions and defects that vary only locally from the bulk structure) but also much more distorted and correspondingly higher energy configurations. We carry out the RSS using the various interatomic potentials and DFT for eight-atom unit cells with constraints on the initial shape (close to cubic) and interatomic distances ($> 1.7 \text{ \AA}$) to exclude unphysically close atoms, relaxed with the two-point steepest-descent [209] method. The resulting distribution of configuration energy and volume are plotted in Fig. 16. The GAP results show a similar distribution to the DFT, with the diamond structure at the correct volume, a few structures with energies up to 0.2 eV per atom higher, mostly at comparable

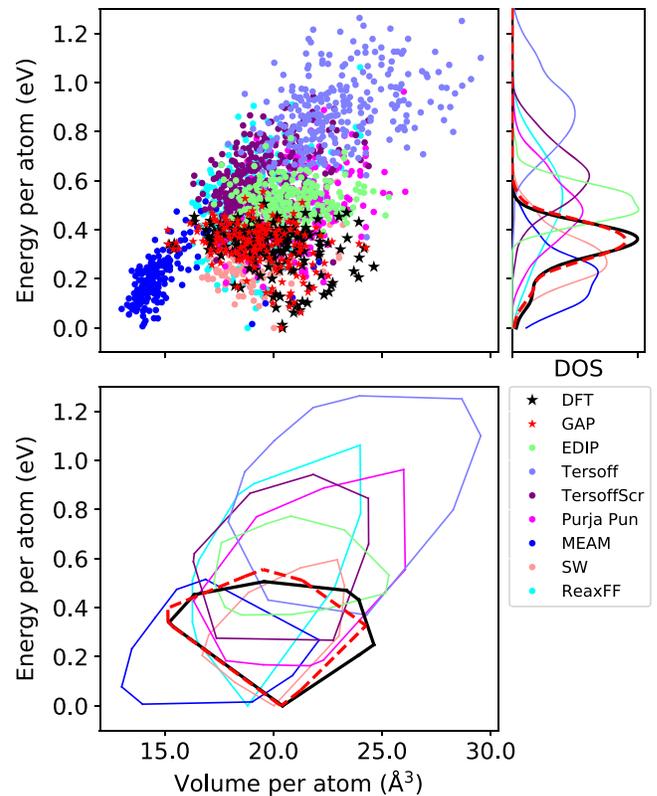


FIG. 16. Relaxed volumes and energies (relative to the diamond structure) for random structure search minima. The top left shows a scatter plot with the DFT (black stars), GAP (red stars), and various other interatomic potentials (various color circles). The top right shows the density of states for the minima. The bottom shows a convex hull surrounding all minima for each method with the same x axis and colors as the top.

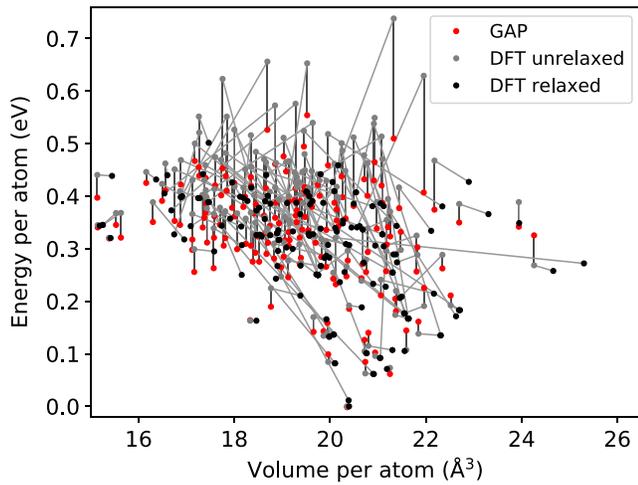


FIG. 17. Energies and volumes of relaxation of GAP RSS minima with DFT forces and stresses. Red circles indicate GAP minima, gray circles indicate DFT energies of GAP minima configurations, and black circles indicate the energy and volume of DFT-relaxed configurations starting from the corresponding GAP minimum.

or somewhat larger volumes (with one or two exceptions at a substantially smaller volume), and a large group at more than 0.2 eV per atom higher at a comparable or smaller volume. None of the interatomic potentials give a similar distribution, and some of the more sophisticated ones give drastically different distributions, including low energies for extremely small (25% below diamond structure) volumes or unphysical local minima at very high energies (>0.4 eV per atom). This discrepancy is in accordance with previous work, which shows that empirical force fields typically give a poor description of the energy landscape with many spurious local minima [210]. The densities of states of the minima found by each model, shown in the top right in Fig. 16, confirm the agreement between the DFT and GAP and the rather different population of minima found by the other potentials.

While the distribution of energies and volumes for the GAP-relaxed minima is similar to that of the DFT-relaxed ones, that does not necessarily mean that individual minima predicted by the GAP are also DFT minima. To test this assumption, we further relax the GAP minima using the DFT and plot the resulting positions on the (E, V) plane in Fig. 17. The plot shows that in many, although clearly not all, cases the GAP energy for GAP minimum configurations is close to the DFT energy for the same configuration, and further relaxation with the DFT does not change the volume or energy very much. The distribution of volume changes, shown in the top in Fig. 18, confirms that most volume changes are small, with 80% falling below 1 \AA^3 per atom or about 5% of the diamond structure volume. The corresponding changes in atomic positions shown in the middle in Fig. 18 show mostly small displacements,

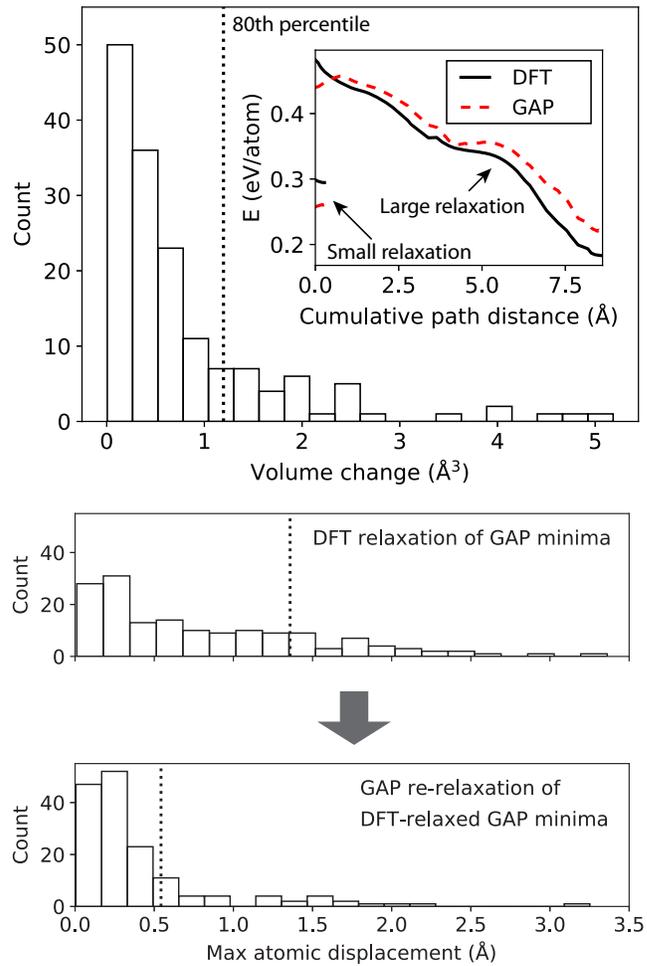


FIG. 18. The top shows the distribution of volume changes for the relaxation of GAP minima using the DFT. The dotted vertical line indicates the 80th percentile. The inset shows the energy per atom for trajectories generated by DFT relaxation starting from the GAP RSS minimum, computed with the DFT (black solid line) and GAP (red dashed line), for one configuration with a small volume change and one with a large volume change. The inset horizontal axis is the cumulative Euclidian distance along the relaxation path. The remaining panels show the distributions of atomic displacements for the DFT relaxation of the GAP minima (center) and the GAP rrelaxation of the DFT-relaxed GAP minima (bottom).

although 20% of the configurations have displacements larger than about 1.5 \AA . Rerelaxing these DFT minima using the GAP shows substantially smaller displacements, indicating that nearly all the DFT minima have nearby GAP minima.

A better understanding of the large volume changes that sometimes occur during DFT relaxation comes from calculating the GAP energy for the DFT relaxation trajectory configurations. Two examples of these energy variations are shown in the inset in Fig. 18. For the configuration with a small overall volume change, the DFT energy goes down a bit and quickly flattens as the DFT minimum is

reached, while the GAP energy on the same trajectory goes up a bit (as it must, since the initial configuration is a GAP local minimum) and also flattens. On the other hand, for the configuration with the large volume change, the DFT energy does not flatten immediately but instead goes down significantly. While the initial GAP local minimum is not also a DFT local minimum, the barrier that separates it from other minima on the GAP PES is small, and all of the configurations that the DFT minimization trajectory goes through have very similar energies (within 0.05 eV per atom) with the DFT and GAP. This similarity suggests that, while the GAP PES is not perfect, differing in the positions and height of some small energy barriers, the overall shape of the PES is, in fact, in good agreement with the DFT.

B. Phonons and thermal expansion

Vibrational properties probe the PES in the region close to the minima and influence the thermodynamic and transport behavior of the material. We compute the phonon and mode Grüneisen dispersion curves in the cubic diamond structure with the GAP, DFT, DFTB, and various interatomic potentials using phonopy [190]. The results are shown in Fig. 19. Even though the phonon frequencies are not included in the database explicitly, there is an excellent agreement between the DFT and GAP. The analytical

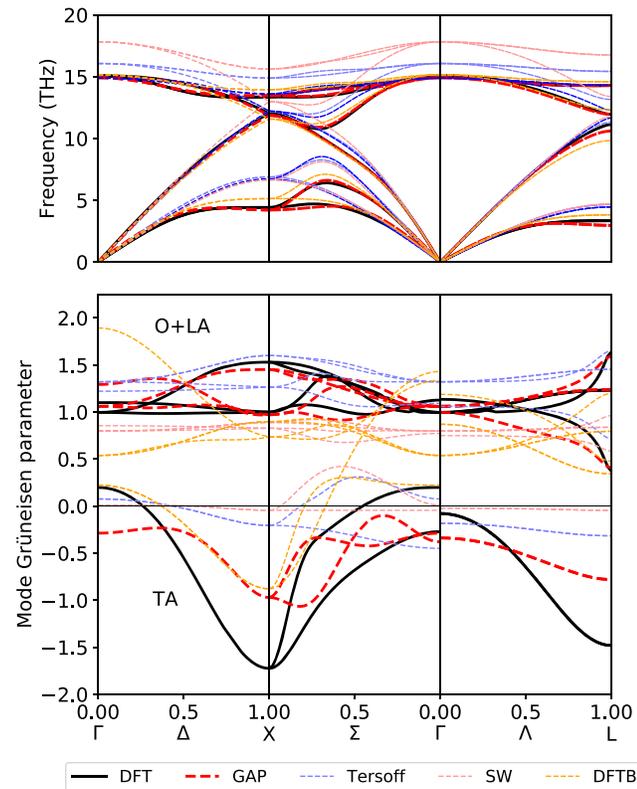


FIG. 19. Top: Phonon dispersion curves of cubic diamond silicon at 0 GPa with various models. Bottom: Dispersion curves of mode Grüneisen parameters.

interatomic potentials are generally in good qualitative agreement for the phonon spectrum, although they overestimate the acoustic branch zone edge and all optical branch frequencies, while the DFTB is in significantly better agreement with the DFT. Not unexpectedly, the slope of the dispersion curves at small k vectors is well reproduced, corresponding to the generally good agreement of the elastic constants of the examined potentials with the DFT. GAP results for the mode Grüneisen parameter are more mixed, with the transverse acoustic branch showing a large discrepancy, indicating that the force data of near-equilibrium crystalline configurations are not sufficient for the fitting procedure to resolve the anharmonicity of the PES. The analytical potentials and DFTB, on the other hand, differ qualitatively from the DFT calculation for all branches.

Diamond structure silicon displays negative thermal expansion at low temperatures [211] due to phonon entropic effects, and this behavior is reproduced by the DFT [212]. To benchmark this unusual feature and other thermal properties of the GAP, we calculate the thermal expansion and Grüneisen parameters using QHA and the heat capacity from the phonon frequencies [213]. Figure 20 shows the temperature dependence of the linear

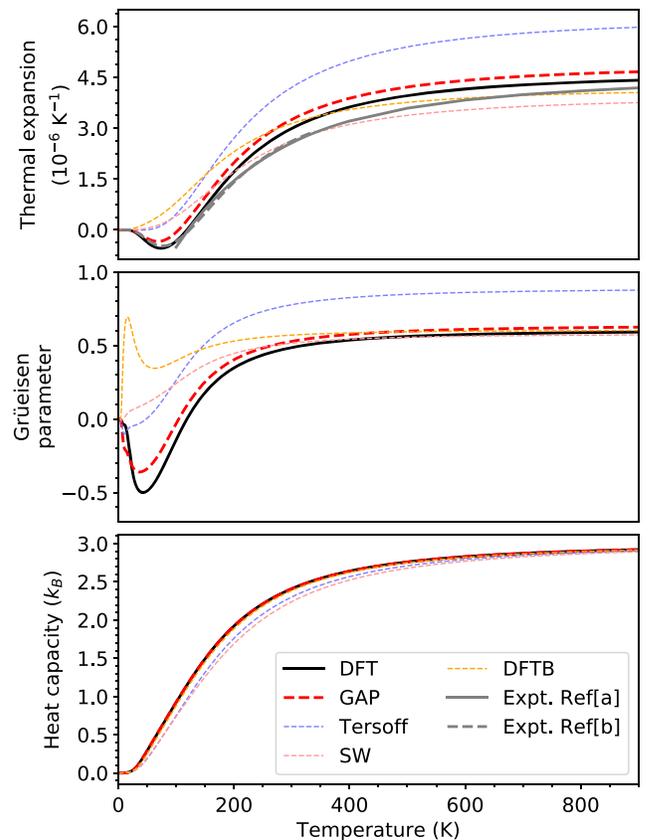


FIG. 20. Linear thermal expansion, Grüneisen parameter, and specific heat of silicon computed with various models, using the quasiharmonic approximation, compared to experimental results: (a) Lyon *et al.* [214] and (b) Okada and Tokumaru [211].

thermal expansion, the Grüneisen parameter, and the heat capacity of various models including the GAP and DFT. Even though the mode Grüneisen parameters of the DFT are not accurately reproduced by the GAP, these averaged thermodynamic properties show a reasonable agreement, including the temperature region with negative thermal expansion. Analytic potentials and the DFTB show a good agreement in the temperature dependence of the heat capacity, not surprisingly, as phonon frequencies are generally similar. However, the errors that the other tested models make in the mode Grüneisen parameters shown in the previous figures manifest themselves here in the form of qualitative errors in the thermal expansion, where none reproduce the negative values at a low temperature. At low temperatures, mostly low energy states are excited and, in particular, those in the transverse acoustic (TA) branch. The hardening of phonons, or, in other words, the strongly negative mode Grüneisen parameters, in the TA branch are associated with the negative thermal expansion of silicon [212]. The shift to large positive values of the TA mode Grüneisen parameters of the DFTB in the Λ direction is probably the reason for the spurious maximum in the Grüneisen parameter.

C. Generalized stacking faults

One type of planar defect that is important as a representative of dislocation properties, which control macroscopic behavior such as plasticity and fracture, is the generalized stacking-fault (GSF) surface [215]. This generalization of the conventional stacking fault, which has been studied extensively in silicon [216–219], is the energy as a function of an arbitrary in-plane shift between two blocks of otherwise undisturbed crystal. We focus on the diamond structure (111) planes, where the GSF can be introduced in glide planes (within a bilayer) or shuffle planes (between bilayers). We calculate the energy along high-symmetry paths on the GSF surface connecting equivalent representations of the ideal crystal in a 1×1 (111) surface cell with nine bilayers (18 atoms) in the normal direction, relaxing the atoms parallel to the surface normal. In each case, we choose the minimum barrier energy path: $[112]$ for the glide fault and $[110]$ for the shuffle fault. The resulting energies along the paths for the GAP and the reference DFT results are shown in Fig. 21. For both relaxed shuffle and glide GSFs, the GAP results are in good agreement with the DFT. The predicted errors shown in the plot are significant near the peaks of both paths, consistent with a relatively large disagreement with the DFT reference values and with the fact that neither type of configuration is included in the fitting database.

A comparison of the energy along one of the paths, glide plane relaxed, for all potentials, is shown in Fig. 22, and the corresponding fractional errors in the peak energy along all relaxed paths relative to the DFT for all the interatomic potentials are shown in Fig. 1. The results for the GAP

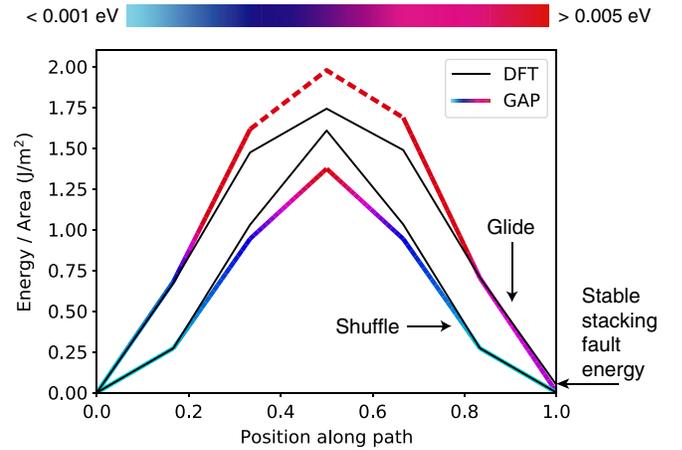


FIG. 21. Relaxed generalized stacking-fault energies along minimum barrier energy path directions ($[112]$ for glide and $[110]$ for shuffle) computed with the DFT (black solid lines) and the GAP. The thick curve showing the GAP model energies is colored according to the maximum per-atom predicted error of the GAP model and dashed where the predicted error exceeds the scale maximum of 5 meV per atom. The upper two curves correspond to the glide plane and the lower two to the shuffle plane.

show reasonable agreement with the DFT, similar to the best of the other interatomic potentials and much better than most.

The final point on the glide-plane GSF path is the conventional *stable* stacking-fault energy γ_{sf} , which is listed in Table III. The DFT reference value is small and positive, indicating that the hexagonal stacking is higher in energy than the cubic diamond structure. To get a sense of the scale, note that the glide curve does not quite reach zero

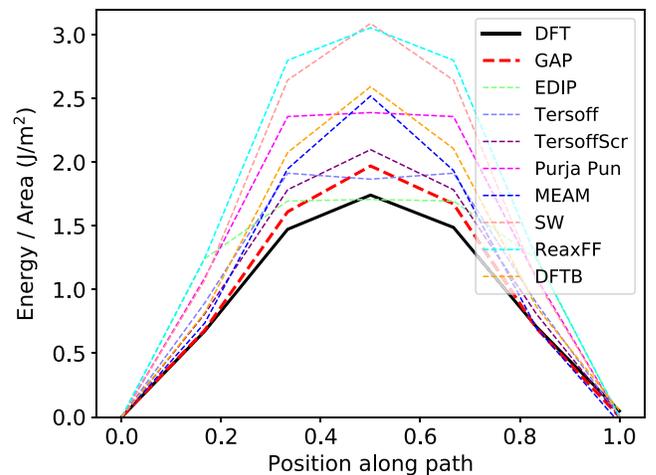


FIG. 22. Relaxed glide-plane generalized stacking-fault energies along the minimum barrier energy path $[112]$ direction computed with the DFT (solid lines), GAP (red dashed lines with symbols), and other interatomic potentials (other color dashed lines).

TABLE III. Stable stacking-fault energy γ_{sf} for each model.

Model	γ_{sf} (J/m ²)
DFT	0.047
GAP	0.002
EDIP	0.000
Tersoff	0.000
TersoffScr	0.001
Purja Pun	0.000
MEAM	-0.046
SW	0.000
ReaxFF	0.004
DFTB	0.052

at the right-hand side of Fig. 21: That mismatch corresponds to the stable stacking-fault energy. The GAP value is positive but much too small, indicating that the diamond structure is indeed the lowest-energy configuration but underestimating the energy difference. There are four atoms with a non-diamond-like second neighbor environment, and the DFT energy difference corresponds to a contribution of about 10 meV from each roughly in correspondence with the ~ 2.5 meV per atom predicted error (purple). The elevated predicted error shows that the GAP’s range and flexibility can distinguish these environments, and the γ_{sf} value could probably be improved by extending the database. While most potentials tested are short ranged and give exactly zero energy, ReaxFF has a similar value to the GAP, while MEAM gives a qualitatively incorrect negative γ_{sf} . The DFTB model is the only one that accurately reproduces the DFT value.

D. Grain boundary

Another class of planar defects that is not included in the fitting database are grain boundaries, which are the interfaces between identical crystal lattices in different orientations. As a simple example of these structures, we choose the (112) Σ 3 tilt boundary of the diamond structure, which can be represented by a relatively small unit cell and can therefore be efficiently computed with the DFT. We computed the energy per unit area of this grain boundary with the various interatomic potentials and the DFTB, as well as the DFT, using a cell with 48 atoms, which has a single interface unit cell and is about 27 Å long normal to the boundary. The resulting fractional errors relative to the DFT value are shown in Fig. 1, and the GAP force errors for the DFT-relaxed configuration are shown in Fig. 2. Despite the fact that the grain boundary structure is not in the fitting database, the GAP energy is in excellent agreement with the DFT. The difference between the DFT- and GAP-relaxed geometries is also small, as indicated by the small magnitudes of the GAP forces in the DFT-relaxed geometry (Fig. 2), and the corresponding displacements (not shown) are nearly imperceptible. The accuracy of the other interatomic potentials varies considerably, with some also in

very good agreement but others with very large energy errors relative to the DFT reference.

E. Fourfold defect

The point defect with the lowest formation energy in the diamond structure of silicon is the so-called “fourfold coordinated defect” [193], which is formed by a bond rotation followed by reconnecting all broken bonds. The energy barrier for the reverse process (i.e., annealing out this defect) is relatively small, and the GAP model does not, in fact, stabilize this defect, as shown in Fig. 23. Indeed, the database does not contain anything resembling the bond rotation process or the final defect structure, which is quantitatively shown by the predicted error. The energies of the GAP model agree very well with those of the DFT up to where the predicted error (taken as the maximum over all atoms) is lower than about 3 meV per atom and strongly deviate after that. Similarly to the planar defects, the predicted error gives a good qualitative indication of where the database is deficient and is in need of extension.

F. Vacancy migration

We compare the migration paths for vacancies in 63-atom diamond structure cells predicted by the various models, as a test of their ability to describe bond-breaking processes. The end points are relaxed with preconditioned the Limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) scheme [220] to a maximum force tolerance of 10^{-3} eV/Å, and the path is calculated as a linear

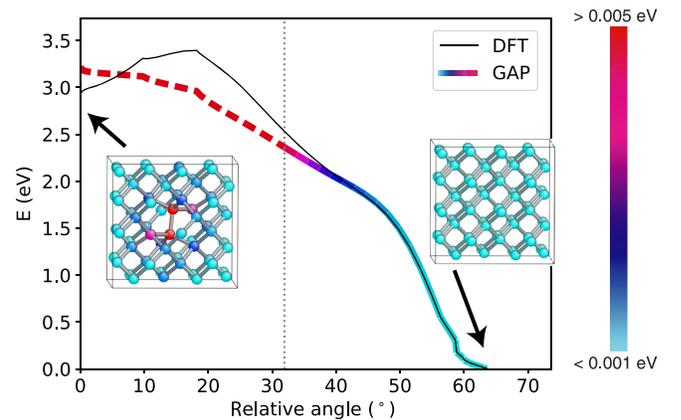


FIG. 23. Relaxation path of the GAP model showing the instability of the fourfold defect as a function of the angle of the rotating bond relative to its initial orientation. The left-hand side of the plot corresponds to the local minimum of the fourfold defect for the DFT model. The black curve shows the energy of this path evaluated with the DFT (which is *not* a DFT minimum energy path but, of course, still shows a barrier). The thick curve shows the GAP model energies, colored according to the maximum per-atom predicted error of the GAP model and is dashed where the predicted error exceeds the scale maximum of 5 meV per atom.

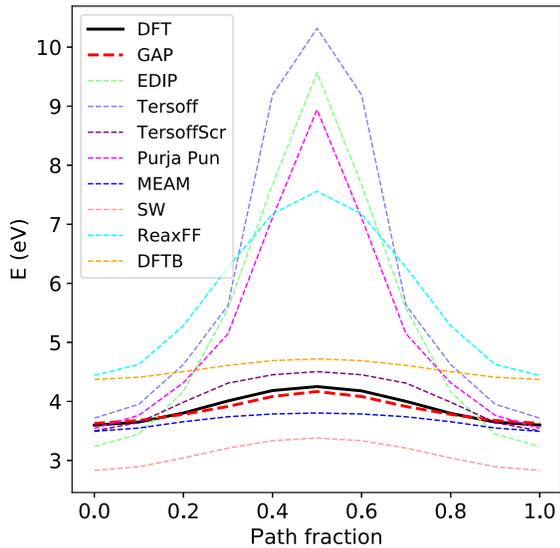


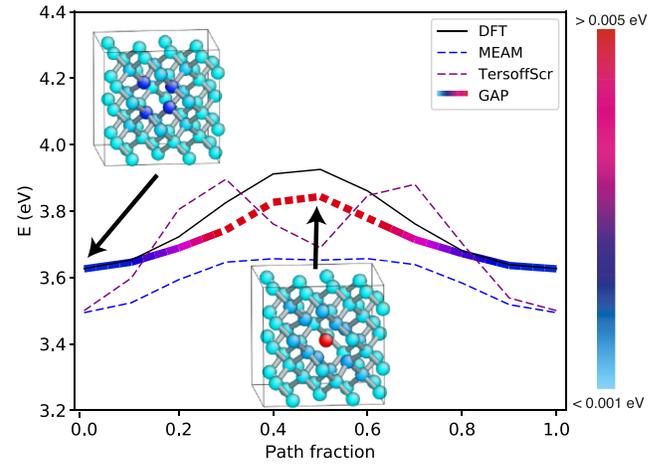
FIG. 24. Unrelaxed energy profiles for the migration of a vacancy, consisting of a series of linearly interpolated configurations between relaxed vacancy end points, in a 63-atom diamond cell for all models. Energies are calculated relative to a bulk crystal with the same number of atoms so that the end points indicate the relaxed vacancy formation energy.

interpolation between the two relaxed end points. The intermediate configurations are not relaxed (as in, e.g., the nudged elastic band method [221]), because features in the PES of many of the potentials lead to ill-behaved paths, similar to the inconsistencies previously noted for the Tersoff potential [222]. The results shown in Fig. 24 indicate the wide variability in the quality of the predictions from the interatomic potentials in comparison to the DFT, with many of the models significantly over- or underestimating both the formation energy and the migration barrier for vacancies.

For the GAP, MEAM, and TersoffScr, which produce formation energies and barriers close to the DFT, the minimum energy path (MEP) is determined using the nudged elastic band (NEB) [221] algorithm as implemented in ASE using nine intermediate images between minima. The results shown in Fig. 25 show that the GAP produces the most accurate MEP in comparison to the DFT, albeit with an underestimate of the barrier. TersoffScr predicts a local minimum at the split vacancy configuration. The insets show the per-atom GAP predicted errors at one of the minima and close to the saddle point; the model is more confident near the minima, since it is trained on similar configurations, while saddlelike configurations are not present in the training data. The predicted error here again provides a useful guide to the expected reliability of the model, with good agreement with the DFT where it is low and decreasing agreement where it exceeds 3 meV per atom.

G. Di-interstitials

Although configurations including simple point defects, such as the monovacancy and the interstitial, are part of our



Model	Formation / eV	Error	Barrier / eV	Error
DFT	3.63	—	0.30	—
MEAM	3.49	-3.7%	0.16	-45.6%
TersoffScr	3.50	-3.5%	0.40	32.5%
GAP	3.63	-0.0%	0.22	-27.0%

FIG. 25. Comparison of the DFT, GAP, MEAM, and TersoffScr models for the migration of a vacancy. Lines show NEB minimum energy pathways for a 63-atom cell, with the thick GAP line and inset images colored by the predicted error, becoming dashed where the maximum error exceeds 0.005 eV/Å, showing that the model is more confident at the minima than the saddle.

training database, the di-interstitial provides an interesting test case of transferability to new defect types. The atomic neighbor environments involved in the di-interstitial are clearly different from anything that is explicitly included in the database. Figure 26 shows the percentage error that various interatomic potentials make in the formation energy of the di-interstitial for six different conformations. We used 66 atoms in the unit cell, including the two extra that are added to a conventional 64-atom cubic unit cell.

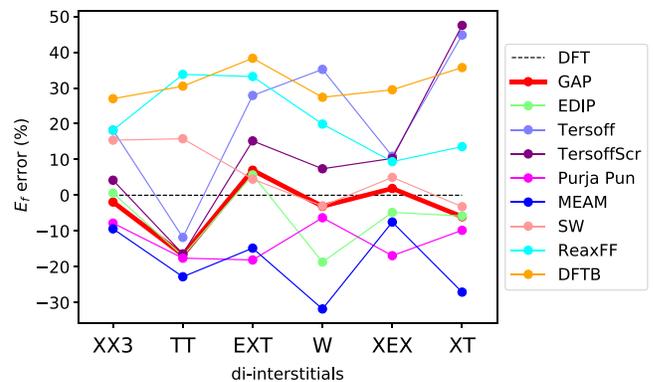


FIG. 26. Percentage error in the formation energy of di-interstitials (I_2) in various configurations for a variety of interatomic potentials, relative to the DFT.

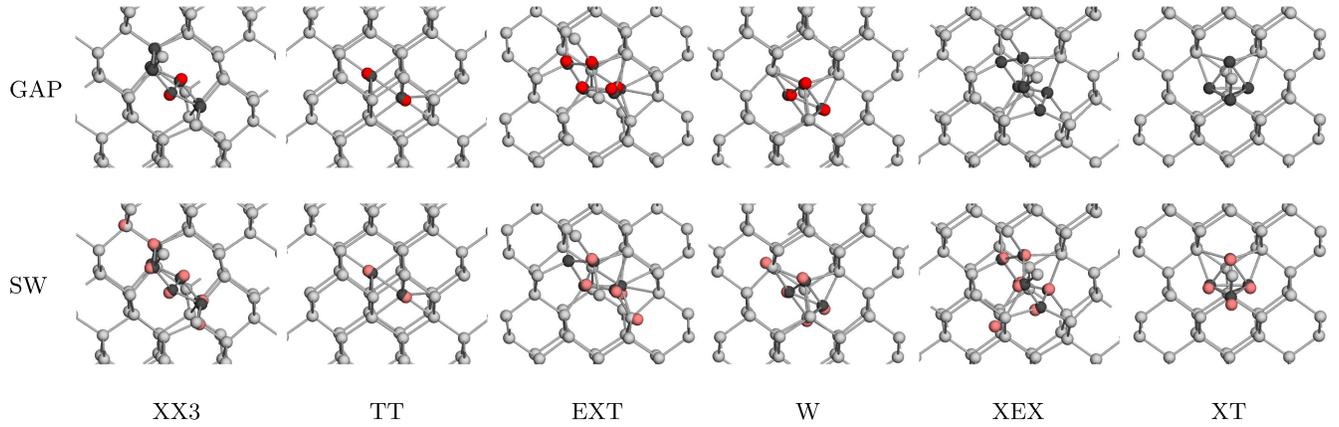


FIG. 27. Visualization of the atomic configurations of the relaxed di-interstitial structures with the GAP (top row) and Stillinger-Weber (bottom row). Dark gray spheres show reference DFT-relaxed positions of defect atoms (those that are significantly different from perfect lattice position or topology), light gray spheres show reference DFT-relaxed positions of other atoms, and colored spheres show interatomic-potential-relaxed positions of defect atoms that are more than 0.1 Å from the corresponding DFT atom position.

The starting positions [223] are relaxed with each potential, as well as the DFT, and the final energies from each local minimum compared.

The results show that most potentials struggle with this property. EDIP, e.g., which performs relatively well for the monovacancy and the single interstitial, makes up to 20% error here. The Stillinger-Weber model, on the other hand, which makes errors of over 50% for the single point defects, looks rather better here. MEAM, ReaxFF, and DFTB perform poorly, similarly to the case of single point defects. It is also clear from the plot that all potentials struggle with the TT configuration (even the best result is over 10%), including the GAP (which otherwise has errors less than about 6%). Figure 27 shows the corresponding relaxed geometries using the GAP and Stillinger-Weber model, as well as the DFT, with colored markings for significant deviations. The Stillinger-Weber model, despite its competitive energy accuracy, shows many more distorted geometries.

V. CONCLUSION

The benefit of the nonparametric approach for creating interatomic potentials, as presented here, is first and foremost its accuracy in matching the target potential energy surface. This enhanced accuracy is not just a quantitative improvement but actually leads to a *qualitatively* better potential: The GAP model for silicon presented here provides a uniformly high accuracy across a wide range of properties and systems including bulk structures, point, and plane defects tested while maintaining useful properties of conventional interatomic potentials: locality and linear scaling computational cost. This achievement requires an accurate description of the energetics of a wide range of configurations, including fully bonded systems as well as bond breaking. The ability of the Si GAP to accurately

describe both the energy and forces during the bond-breaking process, including surface decohesion, unstable stacking-fault minimum energy paths, and point-defect migration barriers, is an especially important point which has been challenging for interatomic potentials to achieve due to their limited variational freedom and short range. Such a comprehensive description of silicon has never before been achieved, despite many efforts, with analytical potentials. The probabilistic nature of the Gaussian process technique allows for uncertainty quantification (and similar measures are possible to obtain using other nonparametric fitting techniques), which is useful in assessing when configurations are encountered that are too far from the training set and are likely to have large errors. Initial efforts here strongly suggest that the accuracy can be further improved for specific properties by adding relevant configurations to the database without compromising the accuracy for other properties.

There are, of course, limitations of the specific potential presented here. The major ones are that (i) we focused on only ambient and moderate pressures here, and a comprehensive silicon potential would be expected to reproduce the large variety of high-pressure phases at finite temperatures; (ii) clusters are included neither in the database nor the tests: These are known to display strong quantum mechanical nonlocality, demonstrated by the existence of magic numbers [167–169], and we do not expect to accurately capture the properties of such small clusters using a local model; (iii) we limited ourselves to considering an elemental material. While it is true that increasing the number of different elements increases the dimensionality and therefore the complexity of the configuration space, several recent works found that including different atomic species does not qualitatively change the difficulty of the problem [153,224]. Another consequence of using

the elemental silicon condensed phase example is that no long-range interactions beyond 5 Å (such as Coulomb or van der Waals) are needed. Properly including a long-range interaction and integrating it with a high-dimensional fit of the short-range interactions is still an outstanding problem. Using a multiscale description to capture these is an alternative approach [225]. Another deficiency of the present potential is that the training database was assembled “by hand” using an *ad hoc* iterative process. It would be desirable to establish protocols that allow the essentially automated construction of databases suitable for predicting and studying specified macroscopic phenomena. There is every hope that the built-in uncertainty quantification can be used in the future to build much better databases and design algorithms that automatically select novel configurations encountered during a simulation for inclusion or even to generate new atomic configurations that are optimized to improve the database.

Finally, we are not making any claims about the optimality of the SOAP kernel and the corresponding basis functions. In particular, our implementation has a computational cost of around 100 ms per atom, and it is certainly possible that there are basis functions that are cheaper to calculate and better suited to the problem, so that fewer of them might be enough to achieve the same accuracy [71,75].

We believe that this potential, perhaps extended by the addition of particular geometries of interest or by a reevaluation of the reference database with more accurate methods, will enable a new and more quantitative approach to simulations of structural properties of silicon.

We are well aware that the merits of the silicon potential presented here will not satisfy all possible audiences. While it is undeniable that the potential is far more accurate and transferable than any before it, the intrinsic problem of transferability stemming from the high dimensionality of the fit has not gone away but is just shown to be less of a practical issue with a large database: The potential remains sensible and even competitive for configurations well away from the training set, such as the stacking faults and the fourfold coordinated defect. We hope that the present success in building a generally applicable potential will allow it to serve as a template for building such models for other materials, enabling scientifically and technologically relevant simulations that have thus far been limited by the trade-offs between accuracy and computational cost.

The potential is available for anyone to use and is provided in the form of an XML file for the QUIP code [172,175]. In addition to usage directly with QUIP, it can be used with the LAMMPS [173] software with the “`pair_style quip`” command, as well as from ASE [170] through QUIP’s `quippy` PYTHON module. The data file includes a copy of the training database structures and associated DFT data. The potential used throughout this paper has the unique label `GAP_2017_6_17_60_4_3_56_165`.

ACKNOWLEDGMENTS

N.B.’s work was supported by the Office of Naval Research through the U.S. Naval Research Laboratory’s core basic research program. J. R. K. acknowledges funding from the Engineering and Physical Sciences Research Council (EPSRC) under Grant No. EP/P002188/1 and the Royal Society under Grant No. RG160691. A. B. P. acknowledges a research fellowship from Magdalene College, Cambridge, and later he was supported by a Leverhulme Early Career Fellowship and the Isaac Newton Trust. A. B. P. also acknowledges support from the Collaborative Computational Project for NMR Crystallography (CCP-NC) and UKCP Consortium, both funded by the EPSRC (EP/M022501/1 and EP/P022561/1, respectively). An award of computer time was provided by the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program. This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract No. DE-AC02-06CH11357. We are grateful for computational support from the United Kingdom national high performance computing service ARCHER, for which access was obtained via the UKCP consortium and funded by EPSRC Grants No. EP/P022561/1 and No. EP/K014560/1. Additional computing facilities were provided by the Scientific Computing Research Technology Platform of the University of Warwick. The authors gratefully acknowledge useful discussion with L. Pastewka, V. Deringer, C. J. Pickard, and M. C. Payne. We value greatly discussions with Alessandro De Vita, our friend and mentor, who recently passed away unexpectedly.

-
- [1] M. W. Finnis, *Interatomic Forces in Condensed Matter* (Oxford University, New York, 2004).
 - [2] J. Tersoff, *Empirical Interatomic Potential for Silicon with Improved Elastic Properties*, *Phys. Rev. B* **38**, 9902 (1988).
 - [3] D. W. Brenner, *The Art and Science of an Analytic Potential*, *Phys. Status Solidi B* **217**, 23 (2000).
 - [4] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2016).
 - [5] J. Behler, *Perspective: Machine Learning Potentials for Atomistic Simulations*, *J. Chem. Phys.* **145**, 170901 (2016).
 - [6] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakithodi, and C. Kim, *Machine Learning and Materials Informatics: Recent Applications and Prospects*, *npj Comput. Mater.* **3**, 54 (2017).
 - [7] A. P. Bartók, R. Kondor, and G. Csányi, *On Representing Chemical Environments*, *Phys. Rev. B* **87**, 184115 (2013).
 - [8] A. P. Bartók and G. Csányi, *Gaussian Approximation Potentials: A Brief Tutorial introduction*, *Int. J. Quantum Chem.* **115**, 1051 (2015).
 - [9] A. P. Bartók, M. J. Gillan, F. R. Manby, and G. Csányi, *Machine-Learning Approach for One- and Two-Body Corrections to Density Functional Theory: Applications*

- to *Molecular and Condensed Water*, *Phys. Rev. B* **88**, 054104 (2013).
- [10] M. J. Gillan, Da. Alfè, A. P. Bartók, and G. Csányi, *First-Principles Energetics of Water Clusters and Ice: A Many-Body Analysis*, *J. Chem. Phys.* **139**, 244504 (2013).
- [11] V. L. Deringer and G. Csányi, *Machine Learning Based Interatomic Potential for Amorphous Carbon*, *Phys. Rev. B* **95**, 094203 (2017).
- [12] P. Rowe, G. Csányi, D. Alfè, and A. Michaelides, *Development of a Machine Learning Potential for Graphene*, *Phys. Rev. B* **97**, 054303 (2018).
- [13] Z. Li, J. R. Kermode, and A. De Vita, *Molecular Dynamics with On-the-Fly Machine Learning of Quantum-Mechanical Forces*, *Phys. Rev. Lett.* **114**, 096405 (2015).
- [14] A. Glielmo, P. Sollich, and A. De Vita, *Accurate Interatomic Force Fields via Machine Learning with Covariant Kernels*, *Phys. Rev. B* **95**, 214302 (2017).
- [15] E. V. Podryabinkin and A. V. Shapeev, *Active Learning of Linearly Parametrized Interatomic Potentials*, *Comput. Mater. Sci.* **140**, 171 (2017).
- [16] K. Gubaev, E. V. Podryabinkin, and A. V. Shapeev, *Machine Learning of Molecular Properties: Locality and Active Learning*, *J. Chem. Phys.* **148**, 241727 (2018).
- [17] L. Shen and W. Yang, *Molecular Dynamics Simulations with Quantum Mechanics/Molecular Mechanics and Adaptive Neural Networks*, *J. Chem. Theory Comput.* **14**, 1442 (2018).
- [18] T. L. Jacobsen, M. S. Jørgensen, and B. Hammer, *On-the-Fly Machine Learning of Atomic Potential in Density Functional Theory Structure Optimization*, *Phys. Rev. Lett.* **120**, 026102 (2018).
- [19] K. Miwa and H. Ohno, *Interatomic Potential Construction with Self-Learning and Adaptive Database*, *Phys. Rev. Mater.* **1**, 053801 (2017).
- [20] P. O. Dral, A. Owens, S. N. Yurchenko, and W. Thiel, *Structure-Based Sampling and Self-Correcting Machine Learning for Accurate Calculations of Potential Energy Surfaces and Vibrational Levels*, *J. Chem. Phys.* **146**, 244108 (2017).
- [21] T. B. Blank, S. D. Brown, A. W. Calhoun, and D. J. Doren, *Neural-Network Models of Potential-Energy Surfaces*, *J. Chem. Phys.* **103**, 4129 (1995).
- [22] A. Brown, B. J. Braams, K. Christoffel, Z. Jin, and J. M. Bowman, *Classical and Quasiclassical Spectral Analysis of CH₅⁺ Using an Ab Initio Potential Energy Surface*, *J. Chem. Phys.* **119**, 8790 (2003).
- [23] S. Lorenz, M. Scheffler, and A. Gross, *Descriptions of Surface Chemical Reactions Using a Neural Network Representation of the Potential-Energy Surface*, *Phys. Rev. B* **73**, 115431 (2006).
- [24] B. J. Braams and J. M. Bowman, *Permutationally Invariant Potential Energy Surfaces in High Dimensionality*, *Int. Rev. Phys. Chem.* **28**, 577 (2009).
- [25] G. I. Hawe and P. L. A. Popelier, *A Water Potential Based on Multipole Moments Trained by Machine Learning—Reducing Maximum Energy Errors*, *Can. J. Chem.* **88**, 104 (2010).
- [26] M. J. L. Mills and P. L. A. Popelier, *Intramolecular Polarizable Multipolar Electrostatics from the Machine Learning Method Kriging*, *Comput. Theor. Chem.* **975**, 42 (2011).
- [27] V. Babin, C. Leforestier, and F. Paesani, *Development of a “First Principles” Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient*, *J. Chem. Theory Comput.* **9**, 5395 (2013).
- [28] V. Babin, G. R. Medders, and F. Paesani, *Development of a “First Principles” Water Potential with Flexible Monomers. II: Trimer Potential Energy Surface, Third Virial Coefficient, and Small Clusters*, *J. Chem. Theory Comput.* **10**, 1599 (2014).
- [29] G. R. Medders, V. Babin, and F. Paesani, *Development of a “First Principles” Water Potential with Flexible Monomers. III: Liquid Phase Properties*, *J. Chem. Theory Comput.* **10**, 2906 (2014).
- [30] S. Manzhos, R. Dawes, and T. Carrington, *Neural Network-Based Approaches for Building High Dimensional and Quantum Dynamics-Friendly Potential Energy Surfaces*, *Int. J. Quantum Chem.* **115**, 1012 (2014).
- [31] T. J. Hughes, S. M. Kandathil, and P. L. A. Popelier, *Accurate Prediction of Polarised High Order Electrostatic Interactions for Hydrogen Bonded Complexes Using the Machine Learning Method Kriging*, *Spectrochim. Acta, Part A* **136**, 32 (2015).
- [32] T. Lenzen and U. Manthe, *Neural Network Based Coupled Diabatic Potential Energy Surfaces for Reactive Scattering*, *J. Chem. Phys.* **147**, 084105 (2017).
- [33] O. T. Unke and M. Meuwly, *A Toolkit for the Construction of Reproducing Kernel-Based Representations of Data: Application to Multi-Dimensional Potential Energy Surfaces*, *J. Chem. Inf. Model.* **57**, 1923 (2017).
- [34] T. L. Pham, H. Kino, K. Terakura, T. Miyake, K. Tsuda, I. Takigawa, and H. C. Dam, *Machine Learning Reveals Orbital Interaction in Materials*, *Sci. Technol. Adv. Mater.* **18**, 756 (2017).
- [35] J. Behler, *First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems*, *Angew. Chem., Int. Ed.* **56**, 12828 (2017).
- [36] J. Behler and M. Parrinello, *Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces*, *Phys. Rev. Lett.* **98**, 146401 (2007).
- [37] J. Behler, R. Martoňák, D. Donadio, and M. Parrinello, *Metadynamics Simulations of the High-Pressure Phases of Silicon Employing a High-Dimensional Neural Network Potential*, *Phys. Rev. Lett.* **100**, 185501 (2008).
- [38] E. Sanville, A. Bholoa, R. Smith, and S. D. Kenny, *Silicon Potentials Investigated Using Density Functional Theory Fitted Neural Networks*, *J. Phys. Condens. Matter* **20**, 285219 (2008).
- [39] R. Z. Khaliullin, H. Eshet, T. D. Kühne, J. Behler, and M. Parrinello, *Graphite-Diamond Phase Coexistence Study Employing a Neural-Network Mapping of the Ab Initio Potential Energy Surface*, *Phys. Rev. B* **81**, 100103 (2010).
- [40] N. Artrith, T. Morawietz, and J. Behler, *High-Dimensional Neural-Network Potentials for Multicomponent Systems: Applications to Zinc Oxide*, *Phys. Rev. B* **83**, 153101 (2011).
- [41] N. Artrith and J. Behler, *High-Dimensional Neural Network Potentials for Metal Surfaces: A Prototype Study for Copper*, *Phys. Rev. B* **85**, 045439 (2012).

- [42] G. C. Sosso, D. Donadio, S. Caravati, J. Behler, and M. Bernasconi, *Thermal Transport in Phase-Change Materials from Atomistic Simulations*, *Phys. Rev. B* **86**, 104301 (2012).
- [43] S. A. Ghasemi, A. Hofstetter, S. Saha, and S. Goedecker, *Interatomic Potentials for Ionic Systems with Density Functional Accuracy Based on Charge Densities Obtained by a Neural Network*, *Phys. Rev. B* **92**, 045131 (2015).
- [44] S. Faraji, S. A. Ghasemi, S. Rostami, R. Rasoulkhani, B. Schaefer, S. Goedecker, and M. Amsler, *High Accuracy and Transferability of a Neural Network Potential through Charge Equilibration for Calcium Fluoride*, *Phys. Rev. B* **95**, 104105 (2017).
- [45] N. Artrith, A. Urban, and G. Ceder, *Constructing First-Principles Phase Diagrams of Amorphous Li_xSi Using Machine-Learning-Assisted Sampling with an Evolutionary Algorithm*, *J. Chem. Phys.* **148**, 241711 (2018).
- [46] B. Onat, E. D. Cubuk, B. D. Malone, and E. Kaxiras, *Implanted Neural Network Potentials: Application to Li-Si Alloys*, *Phys. Rev. B* **97**, 094106 (2018).
- [47] E. D. Cubuk, B. D. Malone, B. Onat, A. Waterland, and E. Kaxiras, *Representations in Neural Network Based Empirical Potentials*, *J. Chem. Phys.* **147**, 024104 (2017).
- [48] N. Artrith and A. Urban, *An Implementation of Artificial Neural-Network Potentials for Atomistic Materials Simulations: Performance for TiO_2* , *Comput. Mater. Sci.* **114**, 135 (2016).
- [49] R. Kobayashi, D. Giofré, T. Junge, M. Ceriotti, and W. A. Curtin, *Neural Network Potential for Al-Mg-Si alloys*, *Phys. Rev. Mater.* **1**, 053604 (2017).
- [50] S. Hajinazar, J. Shao, and A. N. Kolmogorov, *Stratified Construction of Neural Network Based Interatomic Models for Multicomponent Materials*, *Phys. Rev. B* **95**, 014114 (2017).
- [51] W. Li, Y. Ando, and S. Watanabe, *Cu Diffusion in Amorphous Ta_2O_5 Studied with a Simplified Neural Network Potential*, *J. Phys. Soc. Jpn.* **86**, 104004 (2017).
- [52] W. Li, Y. Ando, E. Minamitani, and S. Watanabe, *Study of Li Atom Diffusion in Amorphous Li_3PO_4 with Neural Network Potential*, *J. Chem. Phys.* **147**, 214106 (2017).
- [53] S. Jindal, S. Chiriki, and S. S. Bulusu, *Spherical Harmonics Based Descriptor for Neural Network Potentials: Structure and Dynamics of Au_{147} Nanocluster*, *J. Chem. Phys.* **146**, 204301 (2017).
- [54] K. Ryczko, K. Mills, I. Luchak, C. Homenick, and I. Tamblyn, *Convolutional Neural Networks for Atomistic Systems*, *Comput. Mater. Sci.* **149**, 134 (2018).
- [55] K. Shakouri, J. Behler, J. Meyer, and G.-J. Kroes, *Accurate Neural Network Description of Surface Phonons in Reactive Gas-Surface Dynamics: $\text{N}_2 + \text{Ru}(0001)$* , *J. Phys. Chem. Lett.* **8**, 2131 (2017).
- [56] V. Quaranta, M. Hellström, and J. Behler, *Proton-Transfer Mechanisms at the Water-ZnO Interface: The Role of Presolvation*, *J. Phys. Chem. Lett.* **8**, 1476 (2017).
- [57] S. K. Natarajan and J. Behler, *Self-Diffusion of Surface Defects at Copper-Water Interfaces*, *J. Phys. Chem. C* **121**, 4368 (2017).
- [58] J. R. Boes and J. R. Kitchin, *Modeling Segregation on AuPd(111) Surfaces with Density Functional Theory and Monte Carlo Simulations*, *J. Phys. Chem. C* **121**, 3479 (2017).
- [59] B. Kolb, X. Luo, X. Zhou, B. Jiang, and H. Guo, *High-Dimensional Atomistic Neural Network Potentials for Molecule-Surface Interactions: HCl Scattering from Au(111)*, *J. Phys. Chem. Lett.* **8**, 666 (2017).
- [60] P. E. Dolgirev, I. A. Kruglov, and A. R. Oganov, *Machine Learning Scheme for Fast Extraction of Chemically Interpretable Interatomic Potentials*, *AIP Adv.* **6**, 085318 (2016).
- [61] V. L. Deringer, G. Csányi, and D. M. Proserpio, *Extracting Crystal Chemistry from Amorphous Carbon Structures*, *Chem. Phys. Chem.* **18**, 873 (2017).
- [62] V. L. Deringer, C. J. Pickard, and Gábor Csányi, *Data-Driven Learning of Total and Local Energies in Elemental Boron*, *Phys. Rev. Lett.* **120**, 156001 (2018).
- [63] M. L. Stein, *Interpolation of Spatial Data* (Springer Science, New York, 2012).
- [64] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT, Cambridge, MA, 2006).
- [65] V. N. Vapnik, *Statistical Learning Theory* (Wiley-Interscience, New York, 1998).
- [66] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, *Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons*, *Phys. Rev. Lett.* **104**, 136403 (2010).
- [67] W. J. Szlachta, A. P. Bartók, and G. Csányi, *Accuracy and Transferability of Gaussian Approximation Potential Models for Tungsten*, *Phys. Rev. B* **90**, 104108 (2014).
- [68] M. A. Caro, V. L. Deringer, J. Koskinen, T. Laurila, and G. Csányi, *Growth Mechanism and Origin of High sp^3 Content in Tetrahedral Amorphous Carbon*, *Phys. Rev. Lett.* **120**, 166101 (2018).
- [69] D. Dragoni, T. D. Daff, G. Csányi, and N. Marzari, *Achieving DFT Accuracy with a Machine-Learning Interatomic Potential: Thermomechanics and Defects in bcc Ferromagnetic Iron*, *Phys. Rev. Mater.* **2**, 013808 (2018).
- [70] V. L. Deringer, C. J. Pickard, and G. Csányi, *Data-Driven Learning of Total and Local Energies in Elemental Boron*, *Phys. Rev. Lett.* **120**, 156001 (2018).
- [71] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, *Spectral Neighbor Analysis Method for Automated Generation of Quantum-Accurate Interatomic Potentials*, *J. Comput. Phys.* **285**, 316 (2015).
- [72] C. Chen, Z. Deng, R. Tran, H. Tang, I.-H. Chu, and S. P. Ong, *Accurate Force Field for Molybdenum by Machine Learning Large Materials Data*, *Phys. Rev. Mater.* **1**, 043603 (2017).
- [73] M. A. Wood and A. P. Thompson, *Extending the Accuracy of the SNAP Interatomic Potential Form*, *J. Chem. Phys.* **148**, 241721 (2018).
- [74] K. Miwa and H. Ohno, *Interatomic Potential Construction with Self-Learning and Adaptive Database*, *Phys. Rev. Mater.* **1**, 053801 (2017).
- [75] A. V. Shapeev, *Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials*, *Multi-scale Model. Simul.* **14**, 1153 (2016).
- [76] E. V. Podryabinkin and A. V. Shapeev, *Active Learning of Linearly Parametrized Interatomic Potentials*, *Comput. Mater. Sci.* **140**, 171 (2017).

- [77] V. Botu and R. Ramprasad, *Learning Scheme to Predict Atomic Forces and Accelerate Materials Simulations*, *Phys. Rev. B* **92**, 094306 (2015).
- [78] V. Botu and R. Ramprasad, *Adaptive Machine Learning Framework to Accelerate Ab Initio Molecular Dynamics*, *Int. J. Quantum Chem.* **115**, 1074 (2015).
- [79] V. Botu, R. Batra, J. Chapman, and R. Ramprasad, *Machine Learning Force Fields: Construction, Validation, and Outlook*, *J. Phys. Chem. C* **121**, 511 (2017).
- [80] I. Kruglov, O. Sergeev, A. Yanilkin, and A. R. Oganov, *Energy-Free Machine Learning Force Field for Aluminum*, *Sci. Rep.* **7**, 8512 (2017).
- [81] M. Rupp, E. Proschak, and G. Schneider, *Kernel Approach to Molecular Similarity Based on Iterative Graph Similarity*, *J. Chem. Inf. Model.* **47**, 2280 (2007).
- [82] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, *Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning*, *Phys. Rev. Lett.* **108**, 058301 (2012).
- [83] G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, *Machine Learning of Molecular Electronic Properties in Chemical Compound Space*, *New J. Phys.* **15**, 095003 (2013).
- [84] K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O. A. von Lilienfeld, A. Tkatchenko, and K.-R. Müller, *Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies*, *J. Chem. Theory Comput.* **9**, 3404 (2013).
- [85] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, *Quantum Chemistry Structures and Properties of 134 Kilo Molecules*, *Sci. Data* **1**, 140022 (2014).
- [86] K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. von Lilienfeld, K.-R. Müller, and A. Tkatchenko, *Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space*, *J. Phys. Chem. Lett.* **6**, 2326 (2015).
- [87] T. Bereau, D. Andrienko, and O. A. von Lilienfeld, *Transferable Atomic Multipole Machine Learning Models for Small Organic Molecules*, *J. Chem. Theory Comput.* **11**, 3225 (2015).
- [88] P. O. Dral, O. A. von Lilienfeld, and W. Thiel, *Machine Learning of Parameters for Accurate Semiempirical Quantum Chemical Calculations*, *J. Chem. Theory Comput.* **11**, 2120 (2015).
- [89] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, *Big Data Meets Quantum Chemistry Approximations: The Δ -Machine Learning Approach*, *J. Chem. Theory Comput.* **11**, 2087 (2015).
- [90] M. Rupp, *Special Issue on Machine Learning and Quantum Mechanics*, *Int. J. Quantum Chem.* **115**, 1058 (2015).
- [91] F. A. Faber, A. Lindmaa, O. A. von Lilienfeld, and R. Armiento, *Machine Learning Energies of 2 Million Elpasolite (ABC2D6) Crystals*, *Phys. Rev. Lett.* **117**, 135502 (2016).
- [92] B. Huang and O. A. von Lilienfeld, *Communication: Understanding Molecular Representations in Machine Learning: The Role of Uniqueness and Target Similarity*, *J. Chem. Phys.* **145**, 161102 (2016).
- [93] F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley, and O. A. von Lilienfeld, *Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error*, *J. Chem. Theory Comput.* **13**, 5255 (2017).
- [94] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, *Quantum-Chemical Insights from Deep Tensor Neural Networks*, *Nat. Commun.* **8**, 13890 (2017).
- [95] P. Bleiziffer, K. Schaller, and S. Riniker, *Machine Learning of Partial Charges Derived from High-Quality Quantum-Mechanical Calculations*, *J. Chem. Inf. Model.* **58**, 579 (2018).
- [96] A. Grisafi, D. M. Wilkins, G. Csányi, and M. Ceriotti, *Symmetry-Adapted Machine-Learning for Tensorial Properties of Atomistic Systems*, *Phys. Rev. Lett.* **120**, 036002 (2018).
- [97] J. L. McDonagh, A. F. Silva, M. A. Vincent, and P. L. A. Popelier, *Machine Learning of Dynamic Electron Correlation Energies from Topological Atoms*, *J. Chem. Theory Comput.* **14**, 216 (2018).
- [98] G. Ferré, T. Haut, and K. Barros, *Learning Molecular Energies Using Localized Graph Kernels*, *J. Chem. Phys.* **146**, 114107 (2017).
- [99] J. S. Smith, O. Isayev, and A. E. Roitberg, *ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost*, *Chem. Sci.* **8**, 3192 (2017).
- [100] L. Zhang, J. Han, H. Wang, R. Car, and E. Weinan, *Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics*, *Phys. Rev. Lett.* **120**, 143001 (2018).
- [101] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, *SchNet—A Deep Learning Architecture for Molecules and Materials*, *J. Chem. Phys.* **148**, 241722 (2018).
- [102] F. A. Faber, A. S. Christensen, B. Huang, and O. A. von Lilienfeld, *Alchemical and Structural Distribution Based Representation for Improved QML*, *J. Chem. Phys.* **148**, 241717 (2018).
- [103] N. Lubbers, J. S. Smith, and K. Barros, *Hierarchical Modeling of Molecular Energies Using a Deep Neural Network*, *J. Chem. Phys.* **148**, 241715 (2018).
- [104] T. Bereau, R. A. DiStasio, Jr., A. Tkatchenko, and O. A. von Lilienfeld, *Non-Covalent Interactions across Organic and Biological Subsets of Chemical Space: Physics-Based Potentials Parametrized from Machine Learning*, *J. Chem. Phys.* **148**, 241706 (2018).
- [105] O. T. Unke and M. Meuwly, *A Reactive, Scalable, and Transferable Model for Molecular Energies from a Neural Network Approach Based on Local Information*, *J. Chem. Phys.* **148**, 241708 (2018).
- [106] F. Fracchia, G. Del Frate, G. Mancini, W. Rocchia, and V. Barone, *Force Field Parametrization of Metal Ions from Statistical Learning Techniques*, *J. Chem. Theory Comput.* **14**, 255 (2018).
- [107] J. Wu, L. Shen, and W. Yang, *Internal Force Corrections with Machine Learning for Quantum Mechanics/Molecular Mechanics Simulations*, *J. Chem. Phys.* **147**, 161732 (2017).

- [108] R. Car and M. Parrinello, *Unified Approach for Molecular Dynamics and Density-Functional Theory*, *Phys. Rev. Lett.* **55**, 2471 (1985).
- [109] N. Marzari and D. Vanderbilt, *Maximally Localized Generalized Wannier Functions for Composite Energy Bands*, *Phys. Rev. B* **56**, 12847 (1997).
- [110] J. Q. Broughton, F. F. Abraham, N. Bernstein, and E. Kaxiras, *Concurrent Coupling of Length Scales: Methodology and Application*, *Phys. Rev. B* **60**, 2391 (1999).
- [111] N. Bernstein and D. W. Hess, *Lattice Trapping Barriers to Brittle Fracture*, *Phys. Rev. Lett.* **91**, 025501 (2003).
- [112] G. Csányi, T. Albaret, M. C. Payne, and A. De Vita, “Learn on the Fly”: A Hybrid Classical and Quantum-Mechanical Molecular Dynamics Simulation, *Phys. Rev. Lett.* **93**, 175503 (2004).
- [113] I. Štich, M. C. Payne, R. D. Kingsmith, J. S. Lin, and L. J. Clarke, *Ab Initio Total-Energy Calculations for Extremely Large Systems—Application to the Takayanagi Reconstruction of Si(111)*, *Phys. Rev. Lett.* **68**, 1351 (1992).
- [114] T. Zhu, J. Li, and S. Yip, *Atomistic Configurations and Energetics of Crack Extension in Silicon*, *Phys. Rev. Lett.* **93**, 205504 (2004).
- [115] M. J. Buehler, A. C. T. Van Duin, and W. A. Goddard III, *Multiparadigm Modeling of Dynamical Crack Propagation in Silicon Using a Reactive Force Field*, *Phys. Rev. Lett.* **96**, 095505 (2006).
- [116] D. Holland and M. P. Marder, *Ideal Brittle Fracture of Silicon Studied with Molecular Dynamics*, *Phys. Rev. Lett.* **80**, 746(E) (1998).
- [117] D. Holland and M. Marder, *Ideal Brittle Fracture of Silicon Studied with Molecular Dynamics*, *Phys. Rev. Lett.* **80**, 746 (1998).
- [118] J. R. Kermode, T. Albaret, D. Sherman, N. Bernstein, P. Gumbsch, M. C. Payne, G. Csányi, and A. De Vita, *Low-Speed Fracture Instabilities in a Brittle Crystal*, *Nature (London)* **455**, 1224 (2008).
- [119] F. H. Stillinger and T. A. Weber, *Computer Simulation of Local Order in Condensed Phases of Silicon*, *Phys. Rev. B* **31**, 5262 (1985).
- [120] F. H. Stillinger and T. A. Weber, *Computer Simulation of Local Order in Condensed Phases of Silicon*, *Phys. Rev. B* **33**, 1451(E) (1986).
- [121] J. Tersoff, *New Empirical Model for the Structural Properties of Silicon*, *Phys. Rev. Lett.* **56**, 632 (1986).
- [122] J. Tersoff, *New Empirical Approach for the Structure and Energy of Covalent Systems*, *Phys. Rev. B* **37**, 6991 (1988).
- [123] J. Tersoff, *Modeling Solid-State Chemistry: Interatomic Potentials for Multicomponent Systems*, *Phys. Rev. B* **39**, 5566 (1989).
- [124] J. Tersoff, *Erratum: Modeling Solid-State Chemistry: Interatomic Potentials for Multicomponent Systems*, *Phys. Rev. B* **41**, 3248(E) (1990).
- [125] H. Balamane, T. Halicioglu, and W. A. Tiller, *Comparative Study of Silicon Empirical Interatomic Potentials*, *Phys. Rev. B* **46**, 2250 (1992).
- [126] G. P. P. Pun and Y. Mishin, *Optimized Interatomic Potential for Silicon and Its Application to Thermal Stability of Silicene*, *Phys. Rev. B* **95**, 224103 (2017).
- [127] J. F. Justo, M. Z. Bazant, E. Kaxiras, V. V. Bulatov, and S. Yip, *Interatomic Potential for Silicon Defects and Disordered Phases*, *Phys. Rev. B* **58**, 2539 (1998).
- [128] M. I. Baskes, *Modified Embedded-Atom Potentials for Cubic Materials and Impurities*, *Phys. Rev. B* **46**, 2727 (1992).
- [129] T. J. Lenosky, B. Sadigh, E. Alonso, V. V. Bulatov, T. D. de la Rubia, J. Kim, A. F. Voter, and J. D. Kress, *Highly Optimized Empirical Potential Model of Silicon*, *Model. Simul. Mater. Sci. Eng.* **8**, 825 (2000).
- [130] A. van Duin (personal communication).
- [131] L. Pastewka, P. Pou, R. Pérez, P. Gumbsch, and M. Moseler, *Describing Bond-Breaking Processes by Reactive Potentials: Importance of an Environment-Dependent Interaction Range*, *Phys. Rev. B* **78**, 161402(R) (2008).
- [132] L. Pastewka, A. Klemenz, P. Gumbsch, and M. Moseler, *Screened Empirical Bond-Order Potentials for Si-C*, *Phys. Rev. B* **87**, 205410 (2013).
- [133] M. I. Baskes, J. S. Nelson, and A. F. Wright, *Semiempirical Modified Embedded-Atom Potentials for Silicon and Germanium*, *Phys. Rev. B* **40**, 6085 (1989).
- [134] B.-J. Lee, *A Modified Embedded Atom Method Interatomic Potential for Silicon*, *CALPHAD: Comput. Coupling Phase Diagrams Thermochem.* **31**, 95 (2007).
- [135] S. Ryu, C. R. Weinberger, M. I. Baskes, and W. Cai, *Improved Modified Embedded-Atom Method Potentials for Gold and Silicon*, *Model. Simul. Mater. Sci. Eng.* **17**, 075008 (2009).
- [136] T. Kumagai, S. Izumi, S. Hara, and S. Sakai, *Development of Bond-Order Potentials That Can Reproduce the Elastic Constants and Melting Point of Silicon for Classical Molecular Dynamics Simulation*, *Comput. Mater. Sci.* **39**, 457 (2007).
- [137] D. Porezag, T. Frauenheim, T. Kohler, G. Seifert, and R. Kaschner, *Construction of Tight-Binding-Like Potentials on the Basis of Density-Functional Theory: Application to Carbon*, *Phys. Rev. B* **51**, 12947 (1995).
- [138] T. Frauenheim, F. Weich, T. Kohler, S. Uhlman, D. Porezag, and G. Seifert, *Density-Functional-Based Construction of Transferable Nonorthogonal Tight-Binding Potentials for Si and SiH*, *Phys. Rev. B* **52**, 11492 (1995).
- [139] <https://www.dftb.org/fileadmin/DFTB/public/slako/pbc/pbc-0-3.tar.xz>.
- [140] J. Gehrman, D. G. Pettifor, A. N. Kolmogorov, M. Reese, M. Mrovec, C. Elsässer, and R. Drautz, *Reduced Tight-Binding Models for Elemental Si and N, and Ordered Binary Si-N Systems*, *Phys. Rev. B* **91**, 054109 (2015).
- [141] W.-C. Lu, C.-Z. Wang, L.-Z. Zhao, W. Qin, and K.-M. Ho, *Three-Center Tight-Binding Potential Model for C and Si*, *Phys. Rev. B* **92**, 035206 (2015).
- [142] N. Bernstein, M. J. Mehl, D. A. Papaconstantopoulos, N. I. Papanicolaou, M. Z. Bazant, and E. Kaxiras, *Energetic, Vibrational, and Electronic Properties of Silicon Using a Nonorthogonal Tight-Binding Model*, *Phys. Rev. B* **62**, 4477 (2000).
- [143] C.-Z. Wang, B. C. Pan, and K.-M. Ho, *An Environment-Dependent Tight-Binding Potential for Si*, *J. Phys. Condens. Matter* **11**, 2043 (1999).

- [144] R. E. Cohen, L. Stixrude, and E. Wasserman, *Tight-Binding Computations of Elastic Anisotropy of Fe, Xe, and Si under Compression*, *Phys. Rev. B* **56**, 8575 (1997).
- [145] T. J. Lenosky, J. D. Kress, I. Kwon, A. F. Voter, B. Edwards, D. F. Richards, S. Yang, and J. B. Adams, *Highly Optimized Tight-Binding Model of Silicon*, *Phys. Rev. B* **55**, 1528 (1997).
- [146] N. Bernstein and E. Kaxiras, *Nonorthogonal Tight-Binding Hamiltonians for Defects and Interfaces in Silicon*, *Phys. Rev. B* **56**, 10488 (1997).
- [147] M. Menon and K. R. Subbaswamy, *Transferable Nonorthogonal Tight-Binding Scheme for Silicon*, *Phys. Rev. B* **50**, 11577 (1994).
- [148] I. Kwon, R. Biswas, C.-Z. Wang, K.-M. Ho, and C. M. Soukoulis, *Transferable Tight-Binding Models for Silicon*, *Phys. Rev. B* **49**, 7242 (1994).
- [149] J. L. Mercer, Jr. and M. Y. Chou, *Tight-Binding Model with Intra-Atomic Matrix Elements*, *Phys. Rev. B* **49**, 8506 (1994).
- [150] L. Goodwin, A. J. Skinner, and D. G. Pettifor, *Generating Transferable Tight-Binding Parameters: Application to Silicon*, *Europhys. Lett.* **9**, 701 (1989).
- [151] L. He, F. Liu, G. Hautier, M. J. T. Oliveira, M. A. L. Marques, F. D. Vila, J. J. Rehr, G.-M. Rignanese, and A. Zhou, *Accuracy of Generalized Gradient Approximation Functionals for Density-Functional Perturbation Theory Calculations*, *Phys. Rev. B* **89**, 064305 (2014).
- [152] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT, Cambridge, MA, 2016), <http://www.deeplearningbook.org>.
- [153] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, *Comparing Molecules and Solids across Structural and Alchemical Space*, *Phys. Chem. Chem. Phys.* **18**, 13754 (2016).
- [154] M. W. Mahoney and P. Drineas, *CUR Matrix Decompositions for Improved Data Analysis*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 697 (2009).
- [155] G. Imbalzano, A. Anelli, D. Giofre, S. Klees, J. Behler, and M. Ceriotti, *Automatic Selection of Atomic Fingerprints and Reference Configurations for Machine-Learning Potentials*, *J. Chem. Phys.* **148**, 241730 (2018).
- [156] E. Snelson and Z. Ghahramani, *Sparse Gaussian Processes Using Pseudo-Inputs*, in *Advances in Neural Information Processing Systems 18*, edited by Y. Weiss, B. Schölkopf, and J. C. Platt (MIT, Cambridge, MA, 2006), pp. 1257–1264.
- [157] J. Q. Quinonero-Candela and C. E. Rasmussen, *A Unifying View of Sparse Approximate Gaussian Process Regression*, *J. Mach. Learn. Res.* **6**, 1939 (2005).
- [158] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, and M. C. Payne, *First Principles Methods Using CASTEP*, *Z. Kristallogr.* **220**, 567 (2005).
- [159] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, *Atoms, Molecules, Solids, and Surfaces: Applications of the Generalized Gradient Approximation for Exchange and Correlation*, *Phys. Rev. B* **46**, 6671 (1992).
- [160] E. Nygren and M. J. Aziz, *Pressure-Enhanced Crystallization Kinetics of Amorphous Si and Ge—Implications for Point-Defect Mechanisms*, *J. Appl. Phys.* **70**, 5323 (1991).
- [161] S. Roorda, J. M. Poate, D. C. Jacobson, D. J. Eaglesham, and F. Spaepen, *Structural Relaxation and Defect Annihilation in Pure Amorphous-Silicon*, *Phys. Rev. B* **44**, 3702 (1991).
- [162] K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. Dal Corso, S. de Gironcoli, T. Deutsch, J. K. Dewhurst, I. Di Marco, C. Draxl, M. Dułak, O. Eriksson, J. A. Flores-Livas, K. F. Garrity, L. Genovese *et al.*, *Reproducibility in Density Functional Theory Calculations of Solids*, *Science* **351**, aad3000 (2016).
- [163] G. Csányi, T. Albaret, G. Moras, A. De Vita, and M. C. Payne, *Multiscale Hybrid Simulation Methods for Material Systems*, *J. Phys. Condens. Matter* **17**, R691 (2005).
- [164] N. Bernstein, J. R. Kermode, and G. Csányi, *Hybrid Atomistic Simulation Methods for Materials Systems*, *Rep. Prog. Phys.* **72**, 026501 (2009).
- [165] A. Peguiron, L. C. Ciacchi, A. De Vita, J. R. Kermode, and G. Moras, *Accuracy of Buffered-Force QM/MM Simulations of Silica*, *J. Chem. Phys.* **142**, 064116 (2015).
- [166] H. Chen, J. Lu, and C. Ortner, *Thermodynamic Limit of Crystal Defects with Finite Temperature Tight Binding*, *Arch. Ration. Mech. Anal.* **230**, 701 (2018).
- [167] J. Pan and M. V. Ramakrishna, *Magic Numbers of Silicon Clusters*, *Phys. Rev. B* **50**, 15431 (1994).
- [168] D. Tomanek and M. A. Schluter, *Calculation of Magic Numbers and the Stability of Small Si Clusters*, *Phys. Rev. Lett.* **56**, 1055 (1986).
- [169] J. C. Phillips, *Stability, Kinetics, and Magic Numbers of Si + n (n = 7–45) Clusters*, *J. Chem. Phys.* **83**, 3330 (1985).
- [170] 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. Bergmann Maronsson *et al.*, *The atomic Simulation Environment—A Python Library for Working with Atoms*, *J. Phys. Condens. Matter* **29**, 273002 (2017).
- [171] E. B. Tadmor, R. S. Elliott, J. P. Sethna, R. E. Miller, and C. A. Becker, *knowledgebase of interatomic models (KIM)*, <https://openkim.org/>.
- [172] <https://github.com/libAtoms/QUIP>.
- [173] <http://lammmps.sandia.gov/>.
- [174] <https://github.com/Atomistica/atomistica>.
- [175] <https://doi.org/10.5281/zenodo.1250555>
- [176] Spglib version 1.9.9, <https://atztogo.github.io/spglib/>.
- [177] H. J. Monkhorst and J. D. Pack, *Special Points for Brillouin-Zone Integrations*, *Phys. Rev. B* **13**, 5188 (1976).
- [178] P. Ganesh and M. Widom, *Liquid-Liquid Transition in Supercooled Silicon Determined by First-Principles Simulation*, *Phys. Rev. Lett.* **102**, 075701 (2009).
- [179] S. Plimpton, *Fast Parallel Algorithms for Short-Range Molecular Dynamics*, *J. Comput. Phys.* **117**, 1 (1995).
- [180] P. G. Sanders and M. J. Aziz, *Self-Diffusivity of Liquid Silicon Measured by Pulsed Laser Melting*, *J. Appl. Phys.* **86**, 4258 (1999).
- [181] R. C. Remsing, M. L. Klein, and J. Sun, *Dependence of the Structure and Dynamics of Liquid Silicon on the Choice of*

- Density Functional Approximation*, *Phys. Rev. B* **96**, 024203 (2017).
- [182] J. Q. Broughton and X.-P. Li, *Phase Diagram of Silicon by Molecular Dynamics*, *Phys. Rev. B* **35**, 9120 (1987).
- [183] K. Kakimoto, *Molecular Dynamics Simulation of Mass Transfer in Molten Silicon*, *J. Appl. Phys.* **77**, 4122 (1995).
- [184] W. Yu, Z. Q. Wang, and D. Stroud, *Empirical Molecular-Dynamics Study of Diffusion in Liquid Semiconductors*, *Phys. Rev. B* **54**, 13946 (1996).
- [185] S. Sastry and C. A. Angell, *Liquid-Liquid Phase Transition in Supercooled Silicon*, *Nat. Mater.* **2**, 739 (2003).
- [186] K. Laaziri, S. Kycia, S. Roorda, H. Chicoine, J. L. Robertson, J. Wang, and S. C. Moss, *High Resolution Radial Distribution Function of pure Amorphous Silicon*, *Phys. Rev. Lett.* **82**, 3460 (1999).
- [187] V. L. Deringer, N. Bernstein, A. P. Bartók, M. J. Cliffe, R. N. Kerber, L. E. Marbella, C. P. Grey, S. R. Elliott, and G. Csányi, *Realistic Atomistic Structure of Amorphous Silicon from Machine-Learning-Driven Molecular Dynamics*, *J. Phys. Chem. Lett.* **9**, 2879 (2018).
- [188] J. R. Morris, C.-Z. Wang, K.-M. Ho, and C. T. Chan, *Melting Line of Aluminum from Simulations of Coexisting Phases*, *Phys. Rev. B* **49**, 3109 (1994).
- [189] P. Steinhardt, D. Nelson, and M. Ronchetti, *Bond-Orientalional Order in Liquids and Glasses*, *Phys. Rev. B* **28**, 784 (1983).
- [190] A. Togo and I. Tanaka, *First Principles Phonon Calculations in Materials Science*, *Scr. Mater.* **108**, 1 (2015).
- [191] D. Alfè and M. Gillan, *Exchange-Correlation Energy and the Phase Diagram of Si*, *Phys. Rev. B* **68**, 205212 (2003).
- [192] G. Voronin, C. Pantea, T. Zerda, L. Wang, and Y. Zhao, *In Situ X-Ray Diffraction Study of Silicon at Pressures up to 15.5 GPa and Temperatures up to 1073 K*, *Phys. Rev. B* **68**, 020102(R) (2003).
- [193] S. Goedecker, T. Deutsch, and L. Billard, *A Fourfold Coordinated Point Defect in Silicon*, *Phys. Rev. Lett.* **88**, 235501 (2002).
- [194] K. C. Pandey, *Diffusion without Vacancies or Interstitials: A New Concerted Exchange Mechanism*, *Phys. Rev. Lett.* **57**, 2287 (1986).
- [195] E. Kaxiras and K. C. Pandey, *Contribution of Concerted Exchange to the Entropy of Self-Diffusion in Si*, *Phys. Rev. B* **47**, 1659 (1993).
- [196] R. J. Jaccodine, *Surface Energy of Germanium and Silicon*, *J. Electrochem. Soc.* **110**, 524 (1963).
- [197] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, and M. Ceriotti, *Machine Learning Unifies the Modeling of Materials and Molecules*, *Sci. Adv.* **3**, e1701816 (2017).
- [198] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevX.8.041048> for GAP potential parameter files, testing framework scripts, and a table of test result comparisons between previous and current GAP models.
- [199] D. Haneman, *Surfaces of Silicon*, *Rep. Prog. Phys.* **50**, 1045 (1987).
- [200] S. D. Solares, S. Dasgupta, P. A. Schultz, Y.-H. Kim, C. B. Musgrave, and W. A. Goddard, *Density Functional Theory Study of the Geometry, Energetics, and Reconstruction Process of Si111 Surfaces*, *Langmuir* **21**, 12404 (2005).
- [201] E. Bitzek, J. R. Kermode, and P. Gumbsch, *Atomistic Aspects of Fracture*, *Int. J. Fract.* **191**, 13 (2015).
- [202] J. Swadener, M. Baskes, and M. Nastasi, *Molecular Dynamics Simulation of Brittle Fracture in Silicon*, *Phys. Rev. Lett.* **89**, 085503 (2002).
- [203] N. Bernstein and D. W. Hess, *Lattice Trapping Barriers to Brittle Fracture*, *Phys. Rev. Lett.* **91**, 025501 (2003).
- [204] J. R. Kermode, T. Albaret, D. Sherman, N. Bernstein, P. Gumbsch, M. C. Payne, G. Csányi, and A. De Vita, *Low-Speed Fracture Instabilities in a Brittle Crystal*, *Nature (London)* **455**, 1224 (2008).
- [205] A. Gleizer, G. Peralta, J. R. Kermode, A. De Vita, and D. Sherman, *Dissociative Chemisorption of O₂ Inducing Stress Corrosion Cracking in Silicon Crystals*, *Phys. Rev. Lett.* **112**, 115501 (2014).
- [206] J. R. Kermode, A. Gleizer, G. Kovel, L. Pastewka, G. Csányi, D. Sherman, and A. De Vita, *Low Speed Crack Propagation via Kink Formation and Advance on the Silicon (110) Cleavage Plane*, *Phys. Rev. Lett.* **115**, 135501 (2015).
- [207] C. J. Pickard and R. J. Needs, *High-Pressure Phases of Silane*, *Phys. Rev. Lett.* **97**, 045504 (2006).
- [208] C. J. Pickard and R. J. Needs, *Ab Initio Random Structure Searching*, *J. Phys. Condens. Matter* **23**, 053201 (2011).
- [209] J. Barzilai and J. M. Borwein, *Two-Point Step Size Gradient Methods*, *IMA J. Numer. Anal.* **8**, 141 (1988).
- [210] S. A. Ghasemi, M. Amsler, R. G. Hennig, S. Roy, S. Goedecker, T. J. Lenosky, C. J. Umrigar, L. Genovese, T. Morishita, and K. Nishio, *Energy Landscape of Silicon Systems and Its Description by Force Fields, Tight Binding Schemes, Density Functional Methods, and Quantum Monte Carlo Methods*, *Phys. Rev. B* **81**, 214107 (2010).
- [211] Y. Okada and Y. Tokumaru, *Precise Determination of Lattice Parameter and Thermal Expansion Coefficient of Silicon between 300 and 1500 K*, *J. Appl. Phys.* **56**, 314 (1984).
- [212] S. Biernacki and M. Scheffler, *Negative Thermal Expansion of Diamond and Zinc-Blende Semiconductors*, *Phys. Rev. Lett.* **63**, 290 (1989).
- [213] N. Mounet and N. Marzari, *First-Principles Determination of the Structural, Vibrational and Thermodynamic Properties of Diamond, Graphite, and Derivatives*, *Phys. Rev. B* **71**, 205214 (2005).
- [214] K. G. Lyon, G. L. Salinger, C. A. Swenson, and G. K. White, *Linear Thermal Expansion Measurements on Silicon from 6 to 340 K*, *J. Appl. Phys.* **48**, 865 (1977).
- [215] V. Vitek, *Intrinsic Stacking Faults in Body-Centred Cubic Crystals*, *Philos. Mag.* **18**, 773 (1968).
- [216] E. Kaxiras and M. S. Duesbery, *Free Energies of Generalized Stacking Faults in Si and Implications for the Brittle-Ductile Transition*, *Phys. Rev. Lett.* **70**, 3752 (1993).
- [217] Y.-M. Juan and E. Kaxiras, *Generalized Stacking Fault Energy Surfaces and Dislocation Properties of Silicon: A First-Principles Theoretical Study*, *Philos. Mag. A* **74**, 1367 (1996).
- [218] M. de Koning, A. Antonelli, M. Z. Bazant, E. Kaxiras, and J. F. Justo, *Finite-Temperature Molecular-Dynamics Study of Unstable Stacking Fault Free Energies in Silicon*, *Phys. Rev. B* **58**, 12555 (1998).

- [219] J. Godet, L. Pizzagalli, S. Brochard, and P. Beauchamp, *Comparison between Classical Potentials and Ab Initio Methods for Silicon under Large Shear*, *J. Phys. Condens. Matter* **15**, 6943 (2003).
- [220] D. Packwood, J. Kermode, L. Mones, N. Bernstein, J. Woolley, N. Gould, C. Ortner, and G. Csányi, *A Universal Preconditioner for Simulating Condensed Phase Materials*, *J. Chem. Phys.* **144**, 164109 (2016).
- [221] G. Henkelman and H. Jónsson, *Improved Tangent Estimate in the Nudged Elastic Band Method for Finding Minimum Energy Paths and Saddle Points*, *J. Chem. Phys.* **113**, 9978 (2000).
- [222] M. Posselt, F. Gao, and H. Bracht, *Correlation between Self-Diffusion in Si and the Migration Mechanisms of Vacancies and Self-Interstitials: An Atomistic Study*, *Phys. Rev. B* **78**, 035208 (2008).
- [223] V. A. Borodin and M. G. Ganchenkova (private communication).
- [224] N. Artrith, A. Urban, and G. Ceder, *Efficient and Accurate Machine-Learning Interpolation of Atomic Energies in Compositions with Many Species*, *Phys. Rev. B* **96**, 014112 (2017).
- [225] M. Hirn, S. Mallat, and N. Poilvert, *Wavelet Scattering Regression of Quantum Chemical Energies*, *Multiscale Model. Simul.* **15**, 827 (2017).