

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/140544>

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

Combining phonon accuracy with high transferability in Gaussian approximation potential models

Janine George,¹ Geoffroy Hautier,¹ Albert P. Bartók,² Gábor Csányi,³ and Volker L. Deringer^{4, a)}

¹⁾*Institute of Condensed Matter and Nanosciences, Université catholique de Louvain, Chemin des Étoiles 8, 1348 Louvain-la-Neuve, Belgium*

²⁾*Department of Physics and Warwick Centre for Predictive Modelling, School of Engineering, University of Warwick, Coventry CV4 7AL, United Kingdom*

³⁾*Engineering Laboratory, University of Cambridge, Cambridge CB2 1PZ, United Kingdom*

⁴⁾*Department of Chemistry, Inorganic Chemistry Laboratory, University of Oxford, Oxford OX1 3QR, United Kingdom*

Machine learning driven interatomic potentials, including Gaussian approximation potential (GAP) models, are emerging tools for atomistic simulations. Here, we address the methodological question of how one can fit GAP models that accurately predict vibrational properties in specific regions of configuration space, whilst retaining flexibility and transferability to others. We use an adaptive regularization of the GAP fit that scales with the absolute force magnitude on any given atom, thereby exploring the Bayesian interpretation of GAP regularization as an “expected error”, and its impact on the prediction of physical properties for a material of interest. The approach enables excellent predictions of phonon modes (to within 0.1–0.2 THz) for structurally diverse silicon allotropes, and it can be coupled with existing fitting databases for high transferability across different regions of configuration space, which we demonstrate for liquid and amorphous silicon. These findings and workflows are expected to be useful for GAP-driven materials modeling more generally.

^{a)}Electronic mail: volker.deringer@chem.ox.ac.uk

I. INTRODUCTION

Vibrational properties on the atomic scale determine the thermal behavior of materials. Their knowledge is therefore of central importance in many fields of physics, materials science, and engineering. For example, a low thermal conductivity is a requirement for thermoelectric waste-heat recovery,¹ and the ability to predict this property based on highly accurate theoretical and computational methods can allow the community to discover possible new thermoelectric materials. Indeed, thermal conductivity and other macroscopic quantities can nowadays be computed from first principles, normally based on density-functional theory (DFT), although at a very substantial computational cost.²⁻⁷ Only very recently, larger DFT-computed databases of harmonic vibrational properties became available,^{8,9} but high-throughput predictions of more computationally expensive properties such as thermal conductivities have not been attempted to our knowledge. So far, computational searches for materials with low thermal conductivity have been performed based on computationally cheaper models such as the quasi-harmonic Debye model, that do not require the full *ab initio* computation of the thermal conductivity, or combined with global optimization techniques, that only require a limited amount of full *ab initio* computations of the thermal conductivity.^{10,11}

In an effort to sidestep the computational cost of DFT, machine learning (ML) based interatomic potential models are increasingly widely used in materials modelling.¹²⁻¹⁵ Based on a reference database of (typically) DFT data and a regression framework including artificial neural networks,¹⁶⁻¹⁹ kernel methods,²⁰⁻²² or linear fitting,²³⁻²⁵ they enable atomistic simulations at similar accuracy levels but at orders of magnitude lower computational cost. The high accuracy that ML potentials can reach for phonons was demonstrated a decade ago already,²⁶ and early applications to amorphous phases showcased the ability to treat large and structurally complex systems.^{27,28} To date, vibrational properties continue to be a sensitive test for the quality of a candidate potential,²⁹⁻³² because they give a direct and physically meaningful measure for how reliably the interatomic forces are predicted by any ML model. Notably, a recent benchmark study has shown that a range of different ML potential fitting frameworks can lead to phonon predictions in excellent agreement with DFT reference data.¹⁴

Beyond the prediction of harmonic phonons in crystals, ML potentials have begun to

be used for thermal properties including anharmonic effects as well. A proof-of-concept in the Gaussian Approximation Potential (GAP) framework, for the temperature-dependent phonon dispersion curves of Zr, was reported in 2018.³³ Skutterudite CoSb_3 (Ref. 34) and elemental metals³⁵ were studied using Moment Tensor Potentials (MTP) very recently, showing excellent agreement with DFT reference values. Three independent studies dealt explicitly with crystalline diamond-type (**dia**) silicon.^{36–38}

An aspect that so far has been rarely discussed in this context is the ability of materials to crystallize in different structures (“allotropes” for elements, “polymorphs” for compounds). The evaluation of phonon properties of silicon with ML potentials has focused, so far, almost exclusively on the most abundant structure, *viz.* the **dia** allotrope.^{36–38} There is, however, ample interest in others: clathrate-type structures have been extensively studied by DFT;^{39–41} many other possible silicon allotropes have been proposed.^{42–47} An open-framework structure, oS24, was synthesized by de-intercalation from $\text{Na}_4\text{Si}_{24}$;⁴⁸ it was later studied specifically with respect to its thermal properties, using DFT.⁴⁹ Other metastable silicon allotropes have been observed in laser-induced transformations.⁵⁰ An overview, including possible synthesis routes, was given recently.⁵¹

In the present work, we explore how phonon properties for a diverse ensemble of crystal structures (allotropes or polymorphs) can be described in the GAP framework. We discuss computational protocols by which reference databases can be assembled, and potentials fitted, to deal with a range of silicon allotropes. Our work outlines a general strategy for generating GAP models that can be interfaced to high-throughput materials workflows.

II. METHODOLOGY

A. Gaussian Approximation Potentials

Interatomic forces, on which all of the present study is based, were obtained in the GAP framework.²⁰ Initially, we tested the potential model for silicon developed by Bartók et al. (Ref. 30), which has been designed as a “general-purpose” interatomic potential for various applications in physics and materials science, and which has been extensively validated for physical properties of crystalline³⁰ and amorphous^{52,53} silicon. We refer to this potential as “GAP-18” in the following. In the present work, we describe two methodological

advances over that previous study. First, we developed new fitting databases by various strategies, with a specific view to describe vibrational properties, as detailed in the Results and Discussion section. Second, we use an atom-wise adaptive regularization scheme to improve the accuracy of the fit.

In brief, and using the notation of Ref. 30, the energy in GAP-18 is fitted as

$$E = \sum_{i < j} V^{(2)}(r_{ij}) + \sum_i \varepsilon_i, \quad (1)$$

where the first sum is a baseline pair potential to capture exchange repulsion at short interatomic distances, and the second sum is given by the Gaussian process regression itself. The atomic energy, ε_i , of a given atom, unknown from DFT but the key quantity in GAP, is

$$\varepsilon_i = \sum_s^M \alpha_s K(\mathcal{R}_i, \mathcal{R}_s), \quad (2)$$

based on a kernel (or similarity) function, K . The latter compares the local environment of the i -th atom in a given structure, \mathcal{R}_i , with all M environments in a sparse reference set, \mathcal{R}_s . In GAP-18, this kernel function is given by the Smooth Overlap of Atomic Positions (SOAP) formalism;⁵⁴ we use the same SOAP parameters as in Ref. 30, to ensure comparability.

The task in fitting a GAP model is therefore to find the regression coefficients, α . Assume we have a reference database of n structures, for each of which we know DFT-computed energies, E_j , and forces, $\mathbf{F}_{j,j'}$, with scalar components in x , y , and z direction for all j' atoms in the j -th structure. (We also add DFT-computed virial stresses, which does not change the approach but makes the expressions more bulky, so we only discuss the case of energies and forces below.) We collect all entries of our DFT reference database in a single vector, \mathbf{y} :

$$\mathbf{y} = \left(E_1, \dots, E_n, F_{1,1}^{(x)}, \dots, F_{n,j'}^{(z)} \right). \quad (3)$$

We then define another vector, \mathbf{y}' , which contains the atomic energies for all N atoms in the reference database, which we *do not* know from DFT:

$$\mathbf{y}' = (\varepsilon_1, \dots, \varepsilon_N). \quad (4)$$

We finally define a linear differential operator, \mathbf{L} , which connects the two:

$$\mathbf{y} = \mathbf{L}\mathbf{y}'. \quad (5)$$

fitting phonon properties based on simple supercell displacements and an adjusted regularization (expected error) in the fit; a similar idea has been used very recently in Ref. 36, but not yet explored in detail. We here set the atom-wise regularization for force components in $\mathbf{F}_i = (F_i^{(x)}, F_i^{(y)}, F_i^{(z)})$ according to

$$\sigma_F^{(i)} = \begin{cases} f \times |\mathbf{F}_i|, & \text{if } |\mathbf{F}_i| > F_{\min}; \\ f \times F_{\min}, & \text{else.} \end{cases} \quad (8)$$

Initially, we chose $f = 0.1$ and $F_{\min} = 0.01 \text{ eV \AA}^{-1}$. This value already leads to a lower bound of $\sigma_F = 0.001 \text{ eV \AA}^{-1}$, compared with $\sigma_F = 0.1 \text{ eV \AA}^{-1}$ for crystalline configurations in GAP-18.³⁰ We later vary f over a wide range of even smaller values.

C. Vibrational properties

Phonon and thermal property computations were carried out using phonopy⁶⁰ and phono3py⁶ with the finite displacement method. The size of the atomic displacements were 0.01 \AA (phonopy) and 0.03 \AA (phono3py), as per the standard setting. The thermal conductivity was calculated with the relaxation time approximation as implemented in phono3py.

To compute second-order and third-order force constants for optimized structures with the help of GAP potentials, a script was built that uses pymatgen 2019.12.22,⁶¹ Atomic Simulation Environment (ASE) 3.19.0,⁶² phonopy 2.4.2, phono3py 1.18.2 and quippy (including the GAP code, development version of 10 Jan 2020). The built-in geometry optimization of ASE was first used to optimize the crystal structures (including cell size) with the help of the GAP potentials. The forces on each atom were smaller than 10^{-5} eV/\AA . Phonon band structures were calculated along high-symmetry lines in reciprocal space; the latter were identified as described in Ref. 63 and as implemented in pymatgen. 51 points were calculated between each high-symmetry point. The supercells for the phonopy calculations were created based on the primitive cells and in such a way that each lattice parameter was larger than 15 \AA . To compare the DFT benchmark calculations of the harmonic phonons to GAP potentials, an RMS value was defined that compares all bands along each point in reciprocal space at which the bands were calculated.

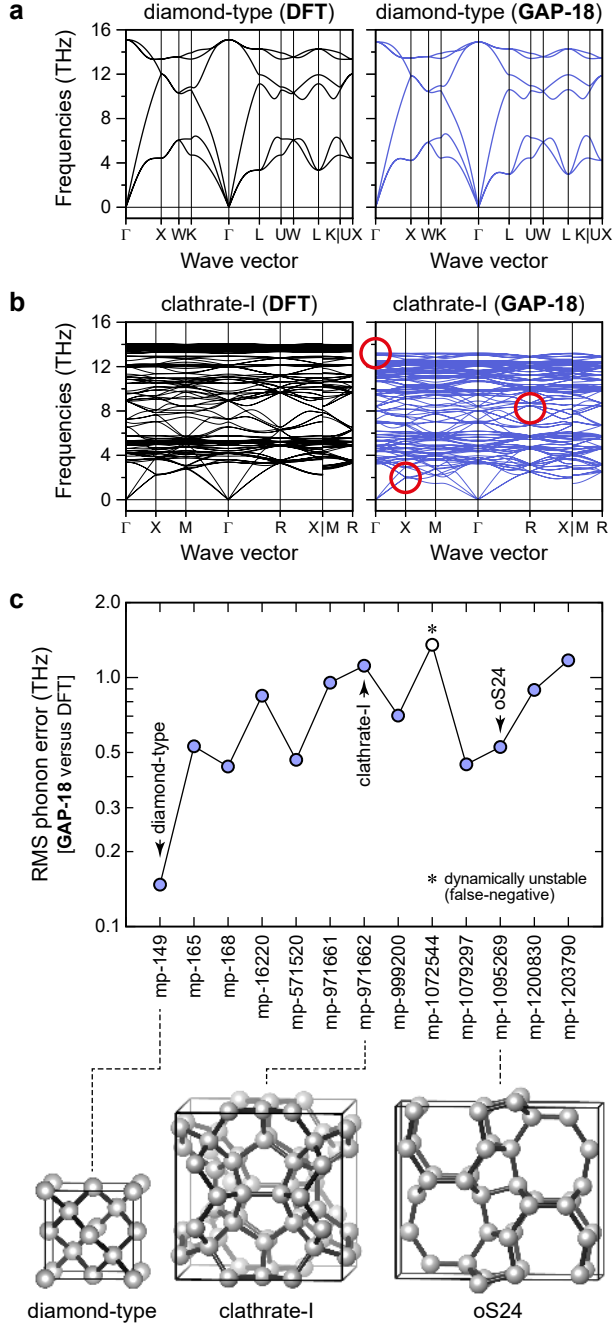


FIG. 1. The current state-of-the-art in predicting phonons with the GAP methodology, illustrated using the general-purpose GAP-18 model for silicon.³⁰ (a) Phonon dispersion curves for the **dia**-type structure, computed with DFT (*left*) and GAP-18 (*right*), showing excellent agreement. (b) Same for clathrate-I-type silicon. The general features of the phonon dispersion are reproduced as well, but there are several discrepancies in detail (highlighted in red), which are discussed in the text. (c) A survey of all ambient-pressure stable structures from the Materials Project (using the “mp” identifier) up to a reasonable size, and the root-mean-square error for their respective GAP-18 predicted phonon dispersions. Lines connecting data points are guides to the eye.

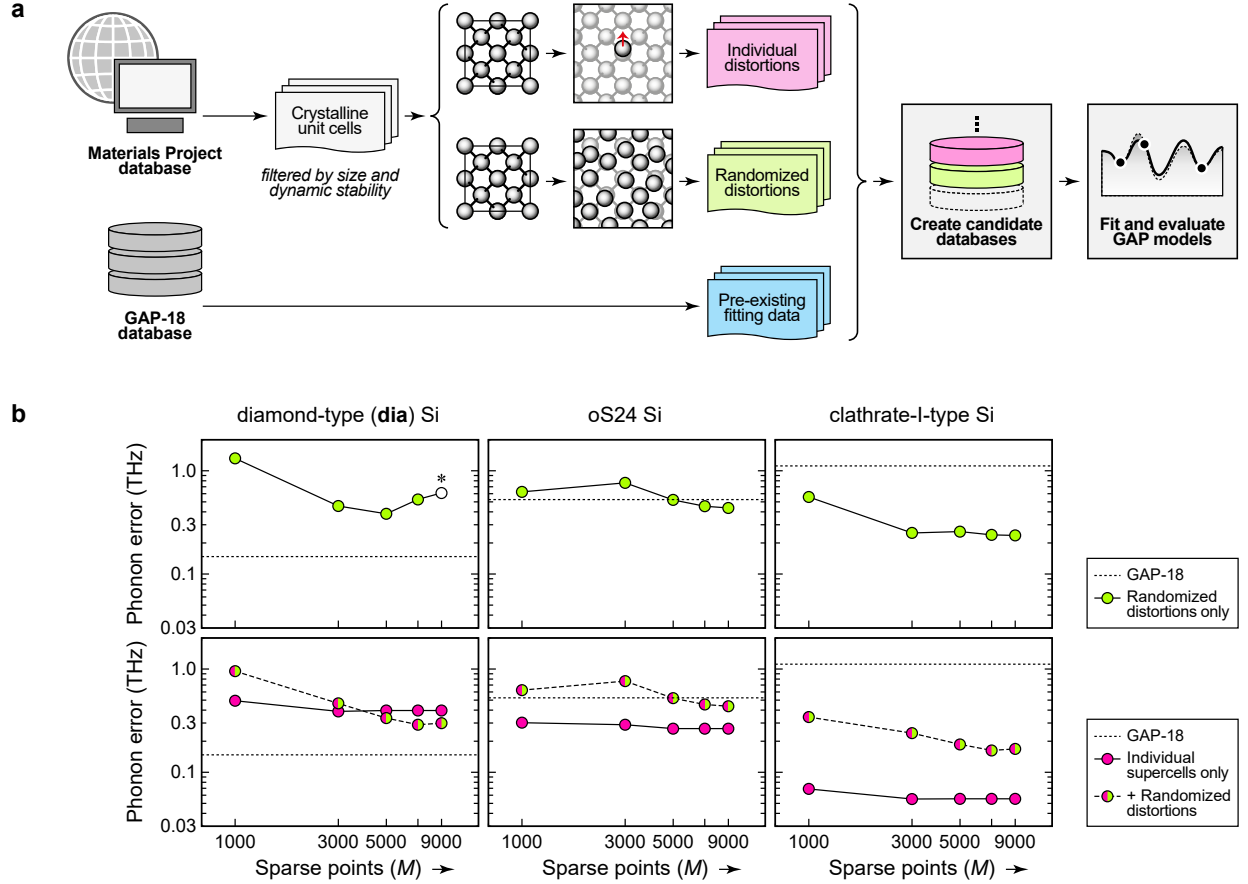


FIG. 2. (a) Strategies for building reference databases, shown by a simplified schematic. We extract crystal structures from the Materials Project database, with filtering as described in the text. We then create supercells based on these structures in which we either displace atoms individually, or all of them randomly, and compute DFT energies and forces for these supercells. Additionally, we investigate the effect of including the pre-existing GAP-18 database.³⁰ We then build various combinations of databases to which candidate GAP models are fitted. (b) Results for various database building schemes sketched above, shown as phonon-error “learning curves” for three representative Si allotropes. An asterisk (*) indicates a structure that is erroneously predicted as dynamically unstable. The reference values for GAP-18 are shown by dotted horizontal lines.

The supercells for the phono3py computations were based on the conventional cells and they were also built such that each lattice parameter was $> 15 \text{ \AA}$. The thermal conductivity for diamond was calculated for the conventional cell and with an $11 \times 11 \times 11$ q -point grid. The complete code is available *via* Github (https://github.com/JaGeo/Phonons_GAPs_silicon) and archived on zenodo.org (<https://doi.org/10.5281/zenodo.3924470>).

III. RESULTS AND DISCUSSION

A. State of the art

We begin by discussing the performance of GAP-18,³⁰ a general-purpose ML potential. We quantify how well it can predict the phonon dispersion relations for diamond-type (**dia**) silicon, which is abundantly represented in its reference database (489 pristine and 404 defective bulk cells, some containing > 200 atoms per cell, plus various surface configurations; Ref. 30)—but also for other allotropes which GAP-18 has not “seen”. We compute all reference phonon band structures at the same level of DFT, enabling direct benchmarking. In the present work, we consider only structures at ambient pressure, but we mention that high-pressure silicon allotropes have been successfully studied with ML potentials before.^{64,65}

For **dia**-Si (Fig. 1a), GAP-18 predicts phonons with practically quantitative accuracy, as previously shown in Ref. 30. This is consistent with the results reported in Ref. 36, where a more specialized GAP model was fitted to snapshots from DFT-based molecular dynamics (MD) simulations of pristine and defective **dia**-Si. At this level of quality, the computed phonon band structures from both methods are practically indistinguishable to the naked eye. To quantify this agreement, we determined the root-mean-square (RMS) error of phonon eigenvalues across the Brillouin zone, which amounted to 0.15 THz.

The situation is different for the clathrate-I-type structure which is not included in the GAP-18 reference database (Fig. 1b). Here, using the potential out-of-the-box still leads to a dynamically stable structure (*i.e.*, without any imaginary eigenvalues, which would conventionally be plotted as negative frequencies), and it does recover the general features. There are, however, notable quantitative differences, and some of them are highlighted in Fig. 1b by red circles. The highest-energy phonon mode at the zone center, Γ , is under-predicted by GAP by 0.8 THz (6%) whereas the lowest predicted band at the X point deviates by 0.3 THz (12%) from the DFT reference data. Especially the latter relative error is much larger (more than twice as large) than relative errors typically arising from DFT computations in comparison to experimental measurements: The mean relative error from DFT to experiment calculated from 53 materials was -3.6% in a recent high-throughput study.⁸ Errors arising from ML potentials should therefore be in a similar range or smaller. The red circle at the R point further highlights that the phonon frequencies calculated by

GAP-18 are smaller than the ones calculated by DFT. The RMS error is 1.11 THz, almost ten times that obtained for **dia**-Si.

To obtain a more comprehensive picture, we test the root-mean-square (RMS) error of phonon eigenvalues for a wide range of reasonably sized and dynamically stable structures from the Materials Project database⁶¹ (Fig. 1c). Out of 13 allotropes considered, all except one are correctly predicted to be dynamically stable. The exception is mp-1072544, a hypothetical structure.⁴² The phonon prediction error for most allotropes is about 0.5–1.0 THz, with clathrate-I being one of the more poorly described examples. This is still a remarkable quality for structures which have not been included in the GAP-18 reference database, such as clathrate-I or oS24.⁴⁸ But the error is clearly too large for *quantitative* studies or for a fully reliable assessment of dynamic stability, which one might wish to carry out when computationally screening large amounts of hypothetical structures.

B. Approaches for building fitting databases

We now perform a comprehensive study of how reference databases can be designed for fitting phonon-accurate GAPs if *no* prior such database exists. Figure 2a provides an overview. We extract structures from the Materials Project database and filter them according to a maximum system size and the criterion that the structure has not significantly changed after the initial structural optimization with CASTEP. This structural change was determined by the StructureMatcher routine implemented in pymatgen, using slightly tighter tolerances for matching structures (`ltol = 0.1`, `stol=0.1` and `angle_tol=3`) instead of the default parameters.⁶¹ We then create supercell models with atomic displacements and generate DFT reference data for those, using two different strategies. A common approach uses randomly displaced cells, creating several copies with lattice parameters scaled by a few percent, and atomic positions randomized with a standard deviation of 0.01 Å, for example (green in Fig. 2a). We also test a different approach (magenta): create supercells of the optimized crystal structures with (only) individual displacements and atom-wise regularization, as discussed in Sec. II. We finally fit candidate GAPs to the resulting databases, always using the same descriptors as in GAP-18, and evaluate their phonon errors by the same RMS measure as in Fig. 1.

The randomized structures, as used in previous GAPs (*e.g.*, Ref. 59), provide an accept-

able but not a truly reliable result for the prediction of phonons (Fig. 2b, top row). We study the evolution of the error as a function of the number of sparse points, M , used in the fit. The outcome of this procedure varies notably from structure to structure; the prediction quality is better than GAP-18 for clathrate-I, about on par for the open-framework oS24 allotrope, but *worse* than GAP-18 for diamond-type silicon. This is indicative of the fact that a similar (and rather small) number of 10 distorted supercells is used to represent each structure, with a view to keep the computational workload tractable even for more complex chemical systems. The number of **dia**-like atomic environments in our databases is therefore much smaller than in that of GAP-18. Interestingly, the **dia** “learning curve” does not converge with M as expected, instead leading to a false-positive prediction of dynamic instability (“*” in Fig. 2b).

An alternative strategy is to create individual displacements (only), in separate supercells, akin to the way that one would build supercells for phonon computations. Displacements are generated along all symmetry-inequivalent directions, which leads to only one supercell for several structures including **dia** ($Fd\bar{3}m$), but many supercells for other structures, depending on space-group symmetry. In this case (magenta in Fig. 2), the phonon accuracy is generally better than with randomized distortions only, and it plateaus after about 3,000 representative atoms—this is clearly expected because the total number of *relevant* force-component entries in \mathbf{y} (Eq. 3) is relatively small, mostly relating to the displaced atom in a supercell and those atoms in its local environment. There are, again, pronounced differences for the three allotropes, with clathrate-I again being most accurately described among those characterized in Fig. 2b.

We also test combined models, which include both Materials Project derived databases (and only those), indicated in Fig. 2b by mixed magenta/green symbols. This appears to even out the performance for the three different allotropes, increasing the phonon error for clathrate-I-type silicon, albeit not above 0.2 THz. The learning curves for the combined databases converge more slowly with M than those for only individual displacements.

Having assessed the qualitative and individual performance of the method for selected structures, we now re-visit the full range of relevant silicon allotropes, with results for RMS phonon errors collected in Fig. 3. To make the comparison easier, we always choose $M = 9000$, as in GAP-18.

Mirroring now more broadly what was already observed in Fig. 2b, the prediction quality

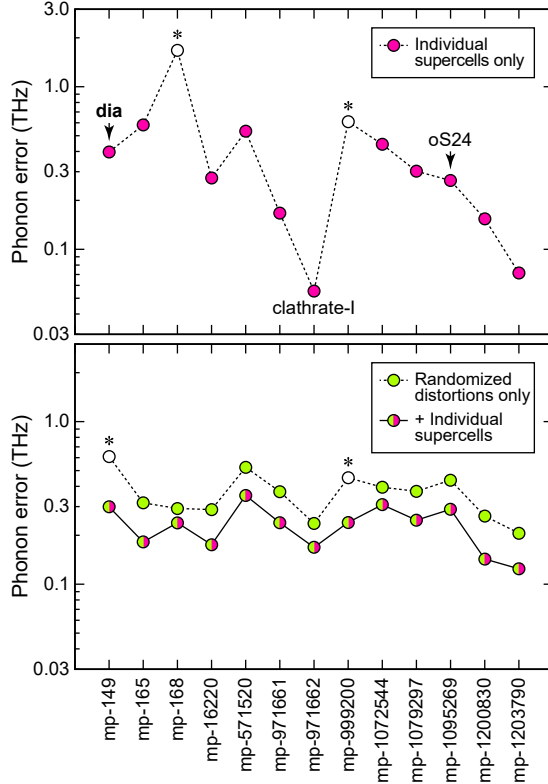


FIG. 3. Predicting phonon frequencies “from scratch” based on Materials Project entries. RMS phonon band structure errors are given for all relevant MP entries as in Fig. 1c, but now using new potentials, fitted to reference databases created as outlined in Fig. 2a. All potentials use $M = 9000$ sparse points. *Top*: Results from the fit with only individual supercells. Two structures are erroneously predicted to be dynamically unstable (*). *Bottom*: Same for a potential fitted to randomly distorted structures only (green) and to a combined database (green/magenta).

is overall quite scattered if only individual supercells are used to construct the reference database (top panel). For most of the silicon allotropes, the approach leads to a prediction error of better than 0.5 THz; clathrate-I turns out to be the best described of all of them. When randomized distortions are used, on the other hand, the distribution of prediction errors is more uniform (green points in the bottom panel in Fig. 3). Two of the 13 structures are erroneously predicted to be dynamically unstable with both approaches—one of them being **dia**-Si in the case of using only randomized distortions, as noted above.

Adding individual supercells to the randomized distortions, *i.e.* combining both strategies outlined at the top of Fig. 2a, appears to improve the results throughout. This is an important general finding: our individual-supercell strategy (magenta) can, and appar-

ently should, be combined with other ways of sampling configuration space when developing GAP fitting databases. We emphasize that this protocol can be fully automated, requiring only a choice of the magnitude of the displacements and of the number of supercells to be created per structure; it is therefore expected to be easily coupled to existing and new high-throughput databases of crystal structures. An open research question, beyond the scope of the present work, is how these automated potentials can accommodate previously unseen structures which are not drawn from the Materials Project or another database, but are discovered, for example, during a GAP-driven random search.⁶⁶

C. Optimized regularization

In this section, we analyze in more detail the effect of the atom-wise regularization on the GAP prediction of phonon band structures. This corresponds to the idea of making the fit “looser” (more flexible) or “tighter” (more accurate in the required regions of configuration space, but therefore less flexible), controlled by the single factor f which we use to scale all atom-wise force regularization parameters according to the absolute force on that atom (Eq. 8). We now perform GAP fits to versions of the “magenta” part of the database (Fig. 2a) in which we vary f over several orders of magnitude, and inspect its effect on the performance for **dia**, oS24, and clathrate-I-type silicon. We also test whether the additional inclusion of randomized distortions in the reference database would still improve the fit (as suggested by Fig. 3). We remind the reader that we had so far used $f = 0.1$ throughout, and we now tighten the regularization, down to $f = 0.0001$. These results are collected in Fig. 4.

For **dia** silicon, the phonon prediction error initially improves substantially when lowering f , *viz.* from about 0.4 to about 0.2 THz. However, for $f = 0.0001$, the error increases again, indicating a too low “expected error” for the input data. The quality of the fit can be even further improved, to better than 0.1 THz, if randomized distortions are added to the fitting database (note that we do not use atom-wise regularization for those configurations, instead employing the GAP-18 default of $\sigma_F = 0.1 \text{ eV \AA}^{-1}$, to retain higher flexibility for those more disordered configurations). For oS24, a qualitatively similar trend is observed in Fig. 4, although here the effect of too small f is more drastic: the resulting potential led to a different structure during the relaxation that must be performed prior to evaluating the phonons, and therefore no RMS error for the latter can be obtained. This example more

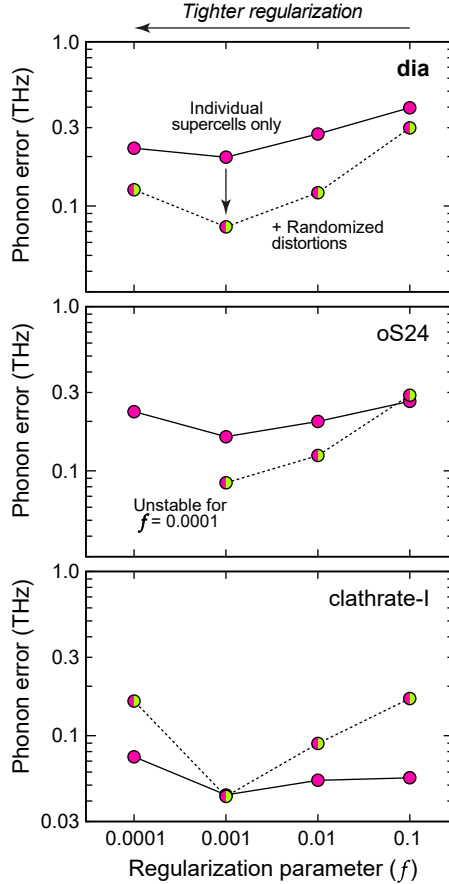


FIG. 4. Optimization of the regularization parameter f , as defined in Eq. 8. We performed series of fits, either with only the individually distorted Materials Project based supercells (magenta) or with the random displacements added (magenta / green). RMS errors for computed vibrational eigenvalues are given as before. The optimal value among those we tested, $f = 0.001$, is indicated by arrows. For oS24, the combined fit at $f = 0.0001$ leads to a change in structure during optimization, as a result of overfitting, and therefore no RMS error for the phonon band structure can be obtained. This indicates the reasonable limit within which f should be chosen.

generally emphasizes the need for a (reasonably large) regularization of the GAP fit to arrive at stable results in practice. For clathrate-I-type silicon, the data shown in Fig. 4 similarly suggest the existence of an optimal f , and here we observe an overall increase in phonon error when randomized configurations are added to the individually displaced supercells—a consequence of the “equalization” between different structures that had already been observed in the previous subsection. Taken together, these results suggest a choice of f between 0.01 and 0.001, but not smaller.

We note that ongoing work in the community aims to optimize GAP fits with regard to the basis functions and hyperparameters for the structural descriptor.^{67–69} The present study complements these efforts by probing the question how the regression task itself might be further optimized.

D. Transferability: Extending GAP-18

We now take our newly created reference databases, constructed from individual distortions with adaptive regularization (magenta in Fig. 2a) and randomized supercells (green in Fig. 2a), respectively, and combine them with the existing database to which the GAP-18 model had previously been fitted³⁰ (blue in Fig. 2a). We thereby aim to answer another more general question about GAPs—namely, how can existing potentials be extended? And how will these potentials retain their previously validated properties, *viz.* for GAP-18, the all-round accurate description of **dia**, liquid, and amorphous silicon?

Our combined fits use a total of $M = 12000$ sparse points, of which 9000 are drawn from the previous database, with distribution among different configuration types (liquid, amorphous, *etc.*) as in GAP-18,³⁰ and 3000 are drawn from a combined set of individually and randomly displaced supercells (Fig. 5a). We combine both new databases because Fig. 3 had indicated that doing this may lead to a more robust behaviour than using the individually displaced cells on their own. Based on our analysis of the regularization parameter (Fig. 4), we generate two candidate potentials with $f = 0.01$ or $f = 0.001$, respectively.

Figure 5b shows that both potentials (with added supercells, “+SCs”) reach a prediction quality within at least the 0.1–0.2 THz range, with even better performance for some structures. There is a further minor improvement when moving from $f = 0.01$ to $f = 0.001$, exemplified by an improved description of the **dia** structure with smaller f , in line with what the tests in Fig. 4 have shown. But both potentials would clearly seem useful for practical purposes when judged on the quality of their phonon predictions alone.

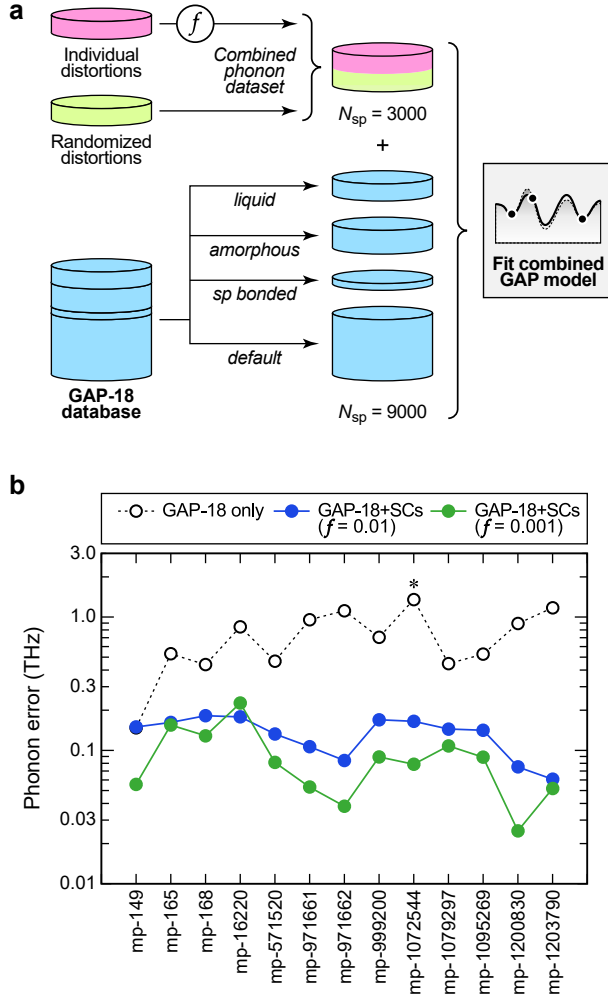


FIG. 5. (a) Building a reference database for the combined potential. Separate configuration type labels are set for the different parts of the GAP-18 database (blue), which are detailed in Ref. 30, and a new type is introduced for the datasets which we add in the present work (magenta/green). The number of sparse points, N_{sp} , for each configuration type is given; their sum is the total number of sparse points, $M = 12000$. (b) Prediction quality for phonons across the range of relevant structures – shown as in Fig. 3, but now comparing GAP-18 and extended potentials that are based on it. In both of these, the individually displaced and randomized supercells (“SCs”) have been added to the fit, with atom-wise regularization according to $f = 0.01$ and $f = 0.001$ (cf. Eq. 8), respectively. All structures are correctly predicted to be dynamically stable with both of the modified potentials.

TABLE I. Thermal conductivity, κ , for **dia**-Si as obtained from GAP-18 and extended versions of it, compared to DFT and experimental references at different temperatures.

Method	κ (W m ⁻¹ K ⁻¹)		
	300 K	600 K	900 K
Expt. (Ref. 70)	142.2	69.2	33.7
DFT (PBE, Ref. 49)	137.4		
DFT (PW91, this work)	126.4	57.1	37.4
GAP-18	110.1	50.4	33.1
GAP-18+SCs ($f = 0.01$)	110.2	50.5	33.1
GAP-18+SCs ($f = 0.001$)	53.7	26.5	17.7

To test transferability of the newly created potentials, we compared the calculated thermal conductivity for **dia** with DFT references and experimental data. We emphasize that by “transferability” we mean that across structures, not across chemical compositions: our claims are strictly related to silicon. The thermal conductivity is a property which is reasonably well described by GAP-18 and the question is therefore whether our modifications lead to a detriment in that performance. To calculate accurate thermal conductivity, third-order force constants have to be calculated. Here, this is done with the help of a finite displacement method where a pair of atoms within the supercell has to be displaced. Such cells are not part of our reference data, and therefore the potential needs to “pick up” the corresponding (anharmonic) physics from other cells in the database, such as MD snapshots. The previous GAP-18 model contains plenty of the latter, and therefore it arrives at acceptable values for the thermal conductivity in comparison to our DFT reference values (less than 15% deviation at each temperature, Tab. I). We mention in passing the good performance of independent, more specialized GAP and neural-network potential models for predicting κ .³⁶⁻³⁸

GAP-18+SCs ($f = 0.01$) shows the same results as the original GAP-18 within any reasonable accuracy, whereas $f = 0.001$ leads to an entirely unreliable prediction of κ . This is now a crucial point, because it indicates that the latter fit is too “tight” on the newly added structures (which correspond to harmonic phonons only), and therefore leads to overfitting behavior which manifests in poor prediction of forces in other regions of configuration space.

It also emphasizes that the prediction quality for κ is determined by the GAP-18 dataset. Follow-up work on optimized databases with regard to predicting thermal conductivity and related properties is currently being planned.

We finally address the question whether the new potentials can also be used in the initial design space of GAP-18, for example, for disordered phases. GAP-18 has enabled accurate atomistic studies of liquid and amorphous silicon^{52,53} and we therefore investigate whether the extended versions are still applicable to the same problems. We designed a test that critically assesses the energy and force accuracy during GAP-driven melt-quench simulations. We do this by running an example trajectory that is representative of “real-world” applications, collecting an ensemble of structures, and benchmarking energies and forces along the trajectory against DFT data.

The starting point (Fig. 6a) is a 512-atom supercell of diamond-type silicon in which 12 atoms are randomly removed, leading to defects including a vacancy cluster. We run a GAP-MD simulation over several hundreds of thousands of steps, in the NPT ensemble as implemented in LAMMPS,⁷¹ similar to Ref. 52. At every 10 ps, we take a structural snapshot and compute its energies and forces using CASTEP⁵⁶ at the PW91 level,⁵⁷ using the same settings as those used for generating the GAP fitting data.

The energy error (Fig. 6b) is instructive to watch for the various new GAPs. It traces the GAP-18 result almost perfectly for $f = 0.01$, but fluctuates much more strongly with $f = 0.001$, already indicating that this model might not completely recover the behavior of the initial potential on which it is based. However, the *absolute* energy error is superficially not bad, and this has an important message for the benchmarking of GAPs: the average energy error itself might not provide sufficient information as to whether or not a given candidate potential is physically meaningful. Indeed, the averaged energy error appears to be *best* for $f = 0.001$ in the final part of the liquid trajectory, *viz.* 8 meV per atom compared to 12 for both other potentials (Table II). A strikingly different result, however, is seen in the force components (Fig. 6c), where $f = 0.01$ recovers the behavior of GAP-18 whereas $f = 0.001$ leads to more than triple the error.

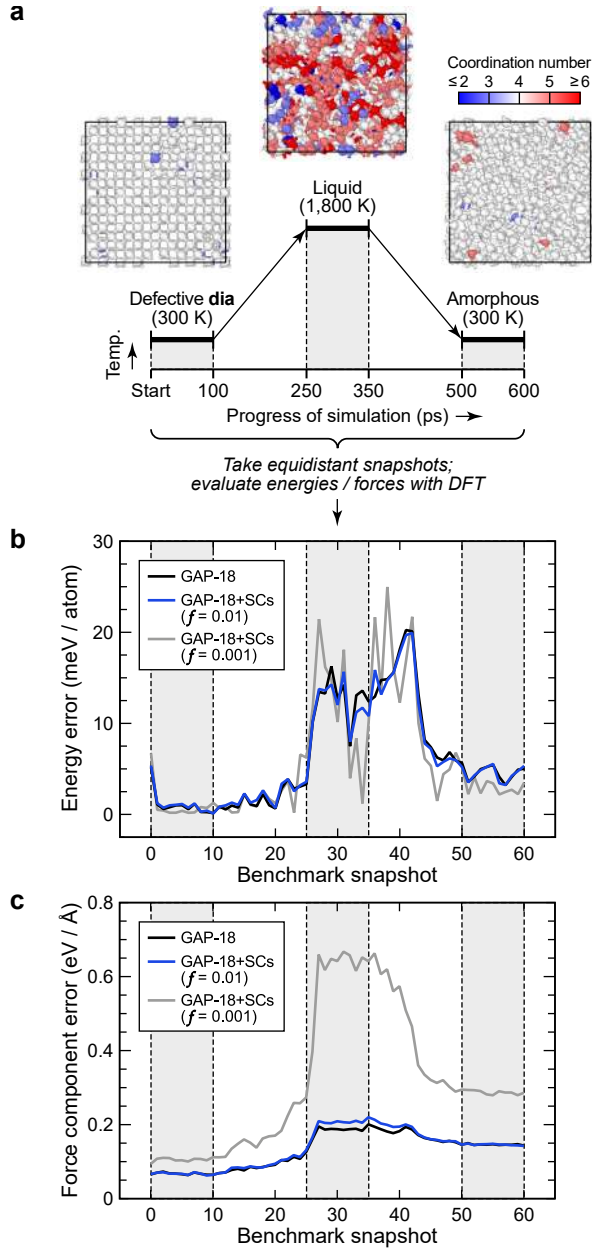


FIG. 6. Assessing the transferability of extended potentials by using a systematic benchmark for disordered structures. (a) Overview of the protocol: a defective diamond-type structure (500 atoms) is thermalized in constant-pressure molecular-dynamics simulations driven by GAP-18, heated to 1,800 K and cooled again (at 10^{13} K s $^{-1}$) to form a quenched amorphous state, similar to Ref. 52. From this trajectory, we take structural snapshots every 10 ps (10,000 simulation steps), and evaluate their energies and forces with DFT. (b) Prediction error for the per-atom energy from various GAP models, given as absolute error versus DFT. (c) Same but now for the forces, with errors given as RMS over all Cartesian force components in a given structure.

TABLE II. Numerical errors for the melt–quench test, as illustrated in Fig. 6a, as a concise quality measure for the different potentials. Errors are averaged over the last five snapshots for each respective part of the trajectory.

	Defective dia (300 K)	Liquid (1,800 K)	Amorphous (300 K)
<i>Energy error (meV / atom)</i>			
GAP-18	0.5	12.2	4.3
+SCs ($f = 0.01$)	0.6	11.5	4.2
+SCs ($f = 0.001$)	0.6	8.4	2.6
<i>Force component error (eV / Å)</i>			
GAP-18	0.07	0.19	0.15
+SCs ($f = 0.01$)	0.07	0.21	0.14
+SCs ($f = 0.001$)	0.11	0.65	0.29

This difference between the potentials not only manifests in different numerical errors, but it qualitatively affects whether they are in any way useful in practice. We repeated the diagnostic MD simulation as sketched in Fig. 6a, but now using either of the extended potential models and otherwise similar settings. The results of this experiment are summarized in Fig. 7. The $f = 0.01$ model led to a correct description both of the liquid and the quenched amorphous structures, indicated by radial distribution functions which are practically superimposable with results for the GAP-18 simulation (Fig. 7b). In stark contrast, the potential fitted with $f = 0.001$ led to an abrupt dissociation of the structure during heating—resulting in a strongly under-coordinated structure, indicated by blue coloring, with only one cluster of atoms, shown on the right-hand side of Fig. 7a. This structure persisted during the entire simulation, with only minimal atomic motion even at 1,800 K. This exercise underlines the need for careful testing and validation of ML potentials in computational practice.

In summary, our extended potential with $f = 0.01$ does not lead to any notable loss in quality for disordered silicon (Table II) whilst improving the description of phonons by almost an order of magnitude (Fig. 5).

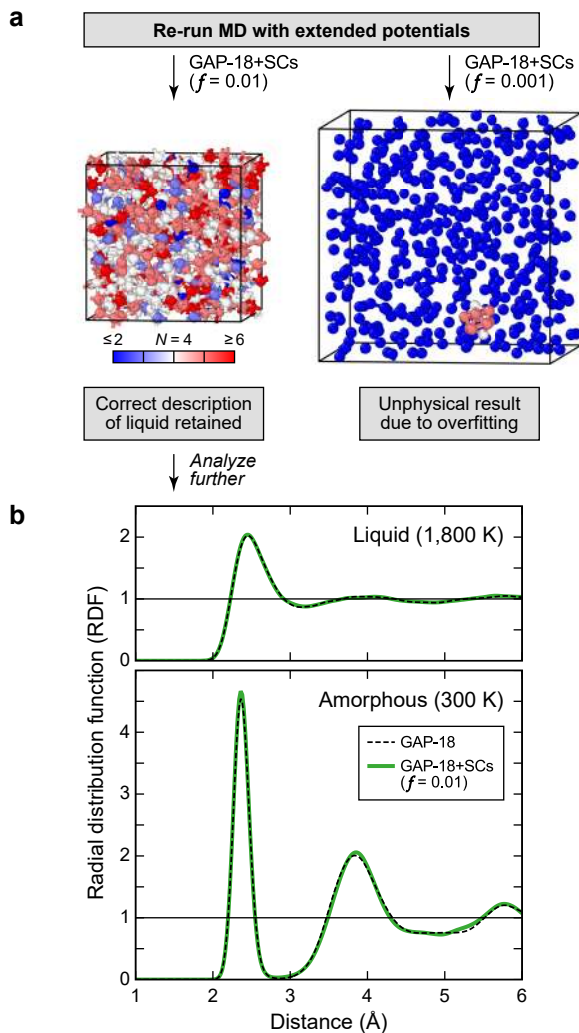


FIG. 7. Assessing the transferability of extended potentials by using them to run diagnostic MD simulations of liquid and amorphous silicon. (a) Snapshots after 300 ps of simulation time, which correspond to melted Si in the scheme of Fig. 6a. The potential with $f = 0.01$ reproduces the behavior of GAP-18, whereas that with $f = 0.001$ leads to an unphysical result during heating which then does not change substantially for the remainder of the simulation. Note that the simulations were done with variable cell volume, and therefore the structure on the right has strongly (erroneously) expanded compared to the correct one on the left. (b) Computed radial distribution functions for the last 50 ps of the liquid and amorphous parts of the trajectory, respectively; in both cases, the result from the extended GAP-18+SCs ($f = 0.01$) potential (solid green lines) is indistinguishable from that of GAP-18 (dashed black lines).

IV. CONCLUSIONS

The ability to compute accurate vibrational and thermal properties with ML potentials is an important result for two reasons. First, it has provided a useful testing ground for the way that reference databases for such potentials can be constructed with minimal computational effort—the main focus here being on the ability to treat various crystal structures *at the same time*. This may lead to a better understanding of the role and physical meaning of the “expected error” built into the GAP fit. Second, it is a step forward towards automated computational materials science, enabling the possibility of high-throughput vibrational property prediction especially when coupled to efficient workflows.^{72–74} It is expected that the approach outlined in the present work can be readily applied to other SOAP-GAP models, and in fact extended to other fitting frameworks—essentially to any material which is amenable to an accurate ML potential fit. It is noted that these frameworks typically rely on the use of local structural descriptors, and that ionic compounds with long-range Coulomb interactions might require additional methodological work.^{75,76}

While most computational work for silicon to date has focused on the diamond-type allotrope, future searches for other structures and thereby synthesis targets are expected to be productive. The theory-guided discovery and, ultimately, experimental realization of new materials has now been achieved in many cases⁷⁷ and it might be accelerated by ML-driven phonon computations as presented here. For the specific example of group-14 elements, we mention the successful synthesis of clathrate-II-type germanium⁷⁸ and also the wealth of intercalated clathrate compounds,⁷⁹ which may now be addressed with similar computational methodology.

ACKNOWLEDGEMENTS

J.G. acknowledges support within the HPC-Europa3 programme (INFRAIA-2016-1-730897), with the support of the EC Research Innovation Action under the H2020 Programme, and from the European Unions Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 837910. V.L.D. acknowledges a Leverhulme Early Career Fellowship. We are grateful for computational support from the UK national high performance computing service, ARCHER, for which access was obtained

via the UKCP consortium and funded by EPSRC grant ref EP/P022561/1. The authors would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work (<http://dx.doi.org/10.5281/zenodo.22558>). We also acknowledge the Consortium des Équipements de Calcul Intensif en Fédération Wallonie Bruxelles (CÉCI) for computational resources. Structural drawings were created using VESTA⁸⁰ and OVITO.⁸¹ We thank Atsushi Togo for helpful input.

CONFLICTS OF INTEREST

A.P.B. and G.C. are listed as inventors on a patent filed by Cambridge Enterprise Ltd. related to SOAP and GAP (US patent 8843509, filed on 5 June 2009 and published on 23 September 2014). The other authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

J.G. and V.L.D. initiated and coordinated the project. J.G. designed and implemented the workflows for phonon and thermal property analyses. G.C. and A.P.B. proposed the force-proportional atom-wise regularisation; A.P.B. implemented it in code; G.C. provided advice on fitting. V.L.D. developed the reference databases and fitted the modified potentials. J.G. and V.L.D. wrote the paper with input from all authors.

DATA AVAILABILITY

The data that support the findings of this study are openly available in a Zenodo repository (<https://doi.org/10.5281/zenodo.3924402>).

REFERENCES

- ¹G. J. Snyder and E. S. Toberer, *Nat. Mater.* **7**, 105 (2008).
- ²P. Carrier, R. Wentzcovitch, and J. Tsuchiya, *Phys. Rev. B* **76**, 064116 (2007).
- ³R. P. Stoffel, C. Wessel, M.-W. Lumey, and R. Dronskowski, *Angew. Chem. Int. Ed.* **49**, 5242 (2010).

- ⁴O. Hellman, P. Steneteg, I. A. Abrikosov, and S. I. Simak, *Phys. Rev. B* **87**, 104111 (2013).
- ⁵J. M. Skelton, S. C. Parker, A. Togo, I. Tanaka, and A. Walsh, *Phys. Rev. B* **89**, 205203 (2014).
- ⁶A. Togo, L. Chaput, and I. Tanaka, *Phys. Rev. B* **91**, 094306 (2015).
- ⁷K. Mizokami, A. Togo, and I. Tanaka, *Phys. Rev. B* **97**, 224306 (2018).
- ⁸G. Petretto, S. Dwaraknath, H. P. Miranda, D. Winston, M. Giantomassi, M. J. Van Setten, X. Gonze, K. A. Persson, G. Hautier, and G.-M. Rignanese, *Sci. Data* **5**, 180065 (2018).
- ⁹<http://phonondb.mtl.kyoto-u.ac.jp/>.
- ¹⁰C. Toher, J. J. Plata, O. Levy, M. de Jong, M. Asta, M. B. Nardelli, and S. Curtarolo, *Phys. Rev. B* **90**, 174107 (2014).
- ¹¹A. Seko, A. Togo, H. Hayashi, K. Tsuda, L. Chaput, and I. Tanaka, *Phys. Rev. Lett.* **115**, 205901 (2015).
- ¹²J. Behler, *Angew. Chem. Int. Ed.* **56**, 12828 (2017).
- ¹³V. L. Deringer, M. A. Caro, and G. Csányi, *Adv. Mater.* **31**, 1902765 (2019).
- ¹⁴Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood, and S. P. Ong, *J. Phys. Chem. A* **124**, 731 (2020).
- ¹⁵T. Mueller, A. Hernandez, and C. Wang, *J. Chem. Phys.* **152**, 050902 (2020).
- ¹⁶J. Behler and M. Parrinello, *Phys. Rev. Lett.* **98**, 146401 (2007).
- ¹⁷N. Artrith and J. Behler, *Phys. Rev. B* **85**, 045439 (2012).
- ¹⁸J. S. Smith, O. Isayev, and A. E. Roitberg, *Chem. Sci.* **8**, 3192 (2017).
- ¹⁹L. Zhang, J. Han, H. Wang, R. Car, and W. E, *Phys. Rev. Lett.* **120**, 143001 (2018).
- ²⁰A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, *Phys. Rev. Lett.* **104**, 136403 (2010).
- ²¹T. D. Huan, R. Batra, J. Chapman, S. Krishnan, L. Chen, and R. Ramprasad, *npj Comput. Mater.* **3**, 37 (2017).
- ²²S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, *Sci. Adv.* **3**, e1603015 (2017).
- ²³A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, *J. Comput. Phys.* **285**, 316 (2015).
- ²⁴A. Seko, A. Takahashi, and I. Tanaka, *Phys. Rev. B* **92**, 054113 (2015).

- ²⁵A. Shapeev, *Multiscale Model. Simul.* **14**, 1153 (2016).
- ²⁶H. Eshet, R. Z. Khaliullin, T. D. Kühne, J. Behler, and M. Parrinello, *Phys. Rev. B* **81**, 184107 (2010).
- ²⁷G. C. Sosso, D. Donadio, S. Caravati, J. Behler, and M. Bernasconi, *Phys. Rev. B* **86**, 104301 (2012).
- ²⁸G. C. Sosso, V. L. Deringer, S. R. Elliott, and G. Csányi, *Mol. Sim.* **44**, 866 (2018).
- ²⁹P. Rowe, G. Csányi, D. Alfè, and A. Michaelides, *Phys. Rev. B* **97**, 054303 (2018).
- ³⁰A. P. Bartók, J. Kermode, N. Bernstein, and G. Csányi, *Phys. Rev. X* **8**, 041048 (2018).
- ³¹L. Zhang, D.-Y. Lin, H. Wang, R. Car, and W. E, *Phys. Rev. Mater.* **3**, 023804 (2019).
- ³²M. R. G. Marques, J. Wolff, C. Steigemann, and M. A. L. Marques, *Phys. Chem. Chem. Phys.* **21**, 6506 (2019).
- ³³X. Qian and R. Yang, *Phys. Rev. B* **98**, 224108 (2018).
- ³⁴P. Korotaev, I. Novoselov, A. Yanilkin, and A. Shapeev, *Phys. Rev. B* **100**, 144308 (2019).
- ³⁵V. Ladygin, P. Korotaev, A. Yanilkin, and A. Shapeev, *Comput. Mater. Sci.* **172**, 109333 (2020).
- ³⁶H. Babaei, R. Guo, A. Hashemi, and S. Lee, *Phys. Rev. Materials* **3**, 074603 (2019).
- ³⁷E. Minamitani, M. Ogura, and S. Watanabe, *Appl. Phys. Express* **12**, 095001 (2019).
- ³⁸X. Qian, S. Peng, X. Li, Y. Wei, and R. Yang, *Mater. Today Physics* **10**, 100140 (2019).
- ³⁹X. Tang, J. Dong, P. Hutchins, O. Shebanova, J. Gryko, P. Barnes, J. K. Cockcroft, M. Vickers, and P. F. McMillan, *Phys. Rev. B* **74**, 014109 (2006).
- ⁴⁰A. J. Karttunen, T. F. Fässler, M. Linnolahti, and T. A. Pakkanen, *Inorg. Chem.* **50**, 1733 (2011).
- ⁴¹A. J. Karttunen and T. F. Fässler, *ChemPhysChem* **14**, 1807 (2013).
- ⁴²H.-Y. Zhao, J. Wang, Q.-M. Ma, and Y. Liu, *Phys. Chem. Chem. Phys.* **15**, 17619 (2013).
- ⁴³A. Mujica, C. J. Pickard, and R. J. Needs, *Phys. Rev. B* **91**, 214104 (2015).
- ⁴⁴M. Amsler, S. Botti, M. A. L. Marques, T. J. Lenosky, and S. Goedecker, *Phys. Rev. B* **92**, 014101 (2015).
- ⁴⁵V. A. Saleev, A. V. Shipilova, D. M. Proserpio, and G. Fadda, *Eur. Phys. J. B* **90**, 150 (2017).
- ⁴⁶L.-A. Jantke, S. Stegmaier, A. J. Karttunen, and T. F. Fässler, *Chem. Eur. J.* **23**, 2734 (2017).

- ⁴⁷C. He, X. Shi, S. J. Clark, J. Li, C. J. Pickard, T. Ouyang, C. Zhang, C. Tang, and J. Zhong, Phys. Rev. Lett. **121**, 175701 (2018).
- ⁴⁸D. Y. Kim, S. Stefanoski, O. O. Kurakevych, and T. A. Strobel, Nat. Mater. **14**, 169 (2015).
- ⁴⁹T. Ouyang, P. Zhang, H. Xiao, C. Tang, J. Li, C. He, and J. Zhong, J. Phys. D: Appl. Phys. **50**, 425501 (2017).
- ⁵⁰L. Rapp, B. Haberl, C. J. Pickard, J. E. Bradby, E. G. Gamaly, J. S. Williams, and A. V. Rode, Nat. Commun. **6**, 7555 (2015).
- ⁵¹B. Haberl, T. A. Strobel, and J. E. Bradby, Appl. Phys. Rev. **3**, 040808 (2016).
- ⁵²V. 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, J. Phys. Chem. Lett. **9**, 2879 (2018).
- ⁵³N. Bernstein, B. Bhattarai, G. Csányi, D. A. Drabold, S. R. Elliott, and V. L. Deringer, Angew. Chem. Int. Ed. **58**, 7057 (2019).
- ⁵⁴A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B **87**, 184115 (2013).
- ⁵⁵J. Quiñonero Candela and C. E. Rasmussen, J. Mach. Learn. Res. **6**, 1939 (2005).
- ⁵⁶S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson, and M. C. Payne, Z. Krist. **220**, 567 (2005).
- ⁵⁷J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992).
- ⁵⁸G. P. Francis and M. C. Payne, J. Phys.: Condens. Matter **2**, 4395 (1990).
- ⁵⁹V. L. Deringer and G. Csányi, Phys. Rev. B **95**, 094203 (2017).
- ⁶⁰A. Togo and I. Tanaka, Scripta Mater. **108**, 1 (2015).
- ⁶¹A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, *et al.*, APL Mater. **1**, 011002 (2013).
- ⁶²A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dulak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, *et al.*, J. Phys.: Condens. Matter **29**, 273002 (2017).
- ⁶³W. Setyawan and S. Curtarolo, Comput. Mater. Sci. **49**, 299 (2010).
- ⁶⁴J. Behler, R. Martoňák, D. Donadio, and M. Parrinello, Phys. Rev. Lett. **100**, 185501 (2008).
- ⁶⁵J. Behler, R. Martoňák, D. Donadio, and M. Parrinello, Phys. Status Solidi B **245**, 2618 (2008).

- ⁶⁶N. Bernstein, G. Csányi, and V. L. Deringer, *npj Comput. Mater.* **5**, 99 (2019).
- ⁶⁷M. J. Willatt, F. Musil, and M. Ceriotti, *Phys. Chem. Chem. Phys.* **20**, 29661 (2018).
- ⁶⁸M. A. Caro, *Phys. Rev. B* **100**, 024112 (2019).
- ⁶⁹L. Himanen, M. O. J. Jäger, E. V. Morooka, F. F. Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke, and A. S. Foster, *Comput. Phys. Commun.* **247**, 106949 (2020).
- ⁷⁰H. R. Shanks, P. D. Maycock, P. H. Siddles, and G. C. Danielson, *Phys. Rev.* **130**, 1743 (1963).
- ⁷¹S. Plimpton, *J. Comput. Phys.* **117**, 1 (1995).
- ⁷²S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, *Comput. Mater. Sci.* **68**, 314 (2013).
- ⁷³G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, and B. Kozinsky, *Comput. Mater. Sci.* **111**, 218 (2016).
- ⁷⁴K. Mathew, J. H. Montoya, A. Faghaninia, S. Dwarakanath, M. Aykol, H. Tang, I.-h. Chu, T. Smidt, B. Bocklund, M. Horton, *et al.*, *Comput. Mater. Sci.* **139**, 140 (2017).
- ⁷⁵N. Artrith, T. Morawietz, and J. Behler, *Phys. Rev. B* **83**, 153101 (2011).
- ⁷⁶S. A. Ghasemi, A. Hofstetter, S. Saha, and S. Goedecker, *Phys. Rev. B* **92**, 045131 (2015).
- ⁷⁷A. R. Oganov, C. J. Pickard, Q. Zhu, and R. J. Needs, *Nat. Rev. Mater.* **4**, 331 (2019).
- ⁷⁸A. M. Guloy, R. Ramlau, Z. Tang, W. Schnelle, M. Baitinger, and Y. Grin, *Nature* **443**, 320 (2006).
- ⁷⁹M. Beekman and G. S. Nolas, *J. Mater. Chem.* **18**, 842 (2008).
- ⁸⁰K. Momma and F. Izumi, *J. Appl. Crystallogr.* **44**, 1272 (2011).
- ⁸¹A. Stukowski, *Model. Simul. Mater. Sci. Eng.* **18**, 015012 (2010).