Title

A high-throughput framework for lattice dynamics

Journal

npj Computational Materials, 10(1)

ISSN

2057-3960

Authors

Zhu, Zhuoying
Park, Junsoo
Sahasrabuddhe, Hrushikesh
et al.

Publication Date

2024

DOI

10.1038/s41524-024-01437-w

npj | computational materials

Article

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

https://doi.org/10.1038/s41524-024-01437-w

A high-throughput framework for lattice dynamics

Zhuoying Zhu[1,6], Junsoo Park[1,2,6], Hrushikesh Sahasrabuddhe[1], Alex M. Ganose[1,3], Rees Chang[1,4], John W. Lawson[5] & Anubhav Jain[1]

We develop an automated high-throughput workflow for calculating lattice dynamical properties from first principles including those dictated by anharmonicity. The pipeline automatically computes interatomic force constants (IFCs) up to 4th order from perturbed training supercells, and uses the IFCs to calculate lattice thermal conductivity, coefficient of thermal expansion, and vibrational free energy and entropy. It performs phonon renormalization for dynamically unstable compounds to obtain real effective phonon spectra at finite temperatures and calculates the associated free energy corrections. The methods and parameters are chosen to balance computational efficiency and result accuracy, assessed through convergence testing and comparisons with experimental measurements. Deployment of this workflow at a large scale would facilitate materials discovery efforts toward functionalities including thermoelectrics, contact materials, ferroelectrics, aerospace components, as well as general phase diagram construction.

Owing to its fundamental nature, properties derived from lattice dynamics are consequential for many functional applications, such as thermoelectrics[1], contact materials, ferroelectrics[2], shape-memory alloys[3,4], and all areas where phase diagram is of important consideration. In particular, lattice dynamics can describe[5,6] key macroscopic thermal properties such as lattice thermal conductivity, thermal expansion coefficient, and phase stability at finite temperatures. Despite its practical utility, lattice dynamics has long been a difficult problem to computationally describe with enough accuracy and efficiency combined so as to be implemented in high-throughput. A general high-throughput workflow for lattice dynamics is essential to advance materials engineering on multiple fronts. In particular, such a framework could be used to establish thermal properties in materials databases such as the Materials Project (MP)[7], AFLOW[8], and OQMD[9], clearing novel, data-driven paths to materials innovation.

In practice, one of the most accurate routes to theoretically describing lattice vibration at the level of density-functional theory (DFT) is ab initio molecular dynamics (AIMD). Correction for quantum nuclear motion may further be done using path-integral molecular dynamics[10,11]. Due to the immense computational cost of running AIMD, however, it is unwieldy to be applied in high-throughput for a large-scale coverage of the material space. Molecular dynamics using machine-learned force fields have emerged in the past few years and have demonstrated near-AIMD accuracy on several occasions[12,13]. However, force field training itself still requires a considerable amount of ab initio calculations (at least hundreds of trajectory structures)[14].

A much more cost-effective way to describe lattice dynamics is in terms of interatomic force constants (IFCs). They are defined within the Taylor expansion of the total energy with respect to atomic displacements (u), whose derivative yields interatomic forces (F_i^a) of atom a (at its equilibrium site) in direction i as:

$$F_i^a = -\sum_{b,j} \Phi_{ij}^{ab} u_j^b - \frac{1}{2!}\sum_{bc,jk} \Phi_{ijk}^{abc} u_j^b u_k^c - \frac{1}{3!}\sum_{bcd,jkl} \Phi_{ijkl}^{abcd} u_j^b u_k^c u_l^d + \cdots , \quad (1)$$

where Φ represents interatomic force constants (IFCs) within a cluster of atoms (e.g., a, b) in their respective Cartesian displacement directions (e.g., i, j).

The second-order IFCs describe pairwise interactions and thus define harmonic phonons. These can be easily obtained using the now widely known methods such as density functional perturbation theory (DFPT)[15] and the finite-displacement method[16] at minimal to moderate computational expense. Macroscopic thermal properties such as the vibrational free energy, entropy, and heat capacity at the harmonic level can then be derived directly from the theory of phonons. Anharmonic IFCs of third-and-higher order serve as perturbation to the harmonic phonons. They are directly

[1]Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [2]KBR, Inc., Intelligent Systems Division, NASA Ames Research Center, Moffett Field, CA, USA. [3]Department of Chemistry, Molecular Sciences Research Hub, Imperial College London, Wood Lane, London, UK. [4]Department of Materials Science and Engineering, University of Illinois Urbana-Champaign, Champaign, IL, USA. [5]Intelligent Systems Division, NASA Ames Research Center, Moffett Field, CA, USA. [6]These authors contributed equally: Zhuoying Zhu, Junsoo Park. e-mail: ajain@lbl.gov

related to other critical thermal properties such as thermal expansion, thermal conductivity, and dynamical phase stabilization and provide corrections to the harmonic free energy. In contrast to the second-order IFCs, the direct calculation of anharmonic IFCs has long been formidable. Usually, DFPT is applied only for second order and could be done for third order[17], but is essentially infeasible beyond this due to the $2n+1$ theorem of the linear response theory that requires second-or-higher-order derivatives of electronic wave functions[18]. As for finite-displacement, the combinatorial explosion of tensor elements in Φ^d ($d \geq 3$) with increasing order of IFCs and growing numerical errors severely hampers its application to the anharmonic terms. The finite-displacement approach has been taken for third-order IFCs[19], but this is inadequate for fourth-order IFC. Utilizing polynomial machine learning potentials can reduce the computational costs by one order of magnitude compared with the conventional finite-displacement approach for lattice thermal conductivity calculations, as demonstrated recently by Togo et al.[20]. A high efficiency even at the third-order level is the preferred guarantee for a truly wide-scale deployment to small and large systems.

Alternatives for resolving the challenge of calculating anharmonic IFCs have emerged in the past decade. Sampling IFCs using high-information-density configurations can make the anharmonic simulations more achievable at a reasonable expense. This procedure is crucial for the subsequent analysis of the thermodynamic properties of materials. Available packages include CSLD[21], ALAMODE[22], TDEP[23,24] and HiPhive[25]. Mathematically, these aim to extract IFCs by minimizing $\| \mathbf{F} - \mathbb{A}\mathbf{\Phi} \|_2$ subject to physically inspired constraints, where $\mathbb{A}$ is the sensing matrix containing the elements of the atomic displacements constructed from sampled configurations and $\mathbf{F}$ is a vector of forces on all atoms from the configurations. This approach in principle allows for a one-shot fitting of Φ up to any desired order with relatively few training sets. Minimization algorithms used range from ordinary least-squares to sparse-recovery methods such as L_1 regularization. The emergence of efficient methods for anharmonic IFCs and the subsequent development of tools for using anharmonic IFCs to compute macroscopic thermal properties have paved the road for high-throughput computing. Nevertheless, several practical challenges related to automation remain, which we address in this paper.

Here we present an automated workflow designed to (i) provide a bridge between atomic quantum simulations at 0 K and macroscopic thermal properties at finite temperatures, (ii) help to realize large-scale automated anharmonic phonon calculations, and (iii) serve as a user-friendly pipeline with benchmarked input parameters. This comprehensive lattice dynamics workflow calculates beyond harmonic phonon properties of interest, including lattice thermal conductivity (LTC), coefficient of thermal expansion (CTE), finite-temperature phonon dispersions, and vibrational free energy at finite temperatures. It uses a suite of software packages and is optimized with suitable input parameters (e.g., supercell size, cutoff radius for fitting IFCs, etc) that balance accuracy and computational costs. Some immediate potential applications of the results are for (i) high-throughput calculations of thermoelectric figure of merit (zT), using combined anharmonic phonons and electron scattering properties obtained from the recent development of AMSET[26], (ii) phase transitions of metastable materials at finite temperatures, and (iii) searching for negative thermal expansion (NTE) materials.

We illustrate input parameter benchmarking, output accuracy evaluation, and performance speedup estimate of the workflow and showcase temperature-dependent thermal properties for selected metastable materials. To aid reproducibility, we present the specific parameters used in our workflow after systematic benchmarking, e.g., the selected fitting method of rfe and convergence supercell of ~20 Å. More detailed descriptions of embedded parameters are given in Tables 1 and 2. Our proposed workflow can achieve high accuracy for downstream results, including (i) an $R^2 > 0.9$ in thermal expansion coefficient and lattice thermal conductivity across more than 30 materials and (ii) phase transition temperatures with less than <10% error, after considering temperature-dependent free energy corrections. Notably, our approach requires 2-3 orders of magnitude less computational time compared with the finite-displacement method, making the widescale calculations of accurate thermal properties tractable. Finally, implementation of our proposed workflow in the open source package of atomate[27] makes it straightforward to perform these calculations without dedicated user tweaking of HiPhive fitting and DFT parameters, allowing for the approach to be automated. A comprehensive plan for future improvements in computational efficiency and accuracy is outlined in the Discussion section.

Table 1 | The cutoff radii for 2nd, 3rd and 4th order of force constants fitting used in the workflow according to the mean elemental period number of a compound

Fitting cutoffs (Å) for n^{th} order of force constants			
Elemental period number	**$n = 2$**	**$n = 3$**	**$n = 4$**
1	5/5.25/5.5	3/3.2/3.4	2.5/2.65
2	6/6.5/7	3.5/3.9/4.3	3/3.3/3.6
3	7/7.75/8.5	4.5/5.1/5.6	3.5/3.95/4.4
4	8/9/10	5.5/6.25/7	4/4.6/5.2
5	9/10.25/11.5	6/6.9/7.9	4.5/ 5.25
6	10/11.5/13	6.5/7.6/8.8	5/5.9
7	10.6	7/8.3/9.6	5.6/6.5

According to the mean elemental period number of the compound, up to 27 sets of cutoffs are used for IFC fitting and the best fit that leads to the lowest RMSE will be selected for the following procedures.

Table 2 | The key parameters implemented in the workflow

Step	Parameter category	Default values in the workflow	Related procedure
1	ENCUT	600 eV	VASP
1	Functional	PBEsol	VASP
1/2	Perturbing method	FixedDisp:[1-3-8-10] (in the unit of picometer)	VASP/HiPhive
1/2	Configs per disp	*nsites*/2 (cubic system); *nsites*/$2 \times \sqrt{48/(N_{symm})}$ (non-cubic system)	VASP/HiPhive
1/2	Supercell size	20 ± 2 Å, number of atoms > 150	VASP/HiPhive
2	Fitting method	Recursive Feature Elimination (RFE)	HiPhive
2	Cutoff radii	see Table 1	HiPhive

nsites: number of atoms in a primitive cell.
N_{symm}: number of space group symmetry operations.

Results

Overview of high-throughput framework

The lattice dynamics workflow (Fig. 1) involves accurate force constants fitting and thermal property calculations. The main steps and packages used for each step are given as follows:

Step 1: A stringent structure optimization of the initial primitive cell and self-consistent field (SCF) force calculations in small-displaced supercells. All DFT calculations are performed by the Vienna ab initio simulation package (VASP)[28–30], although in principle any other DFT package could be used.

Step 2: The harmonic and anharmonic force constants fitting (IFC fitting performed by the HiPhive[25] code). We select the HiPhive package for IFCs fitting due to its seamless Python-integrability and flexibility of fitting methods.

Step 3: Renormalization step (if applicable) to obtain the stable effective phonons at finite temperatures. Harmonic and anharmonic thermal properties are calculated by Phonopy[16] and Phono3py[31], respectively.

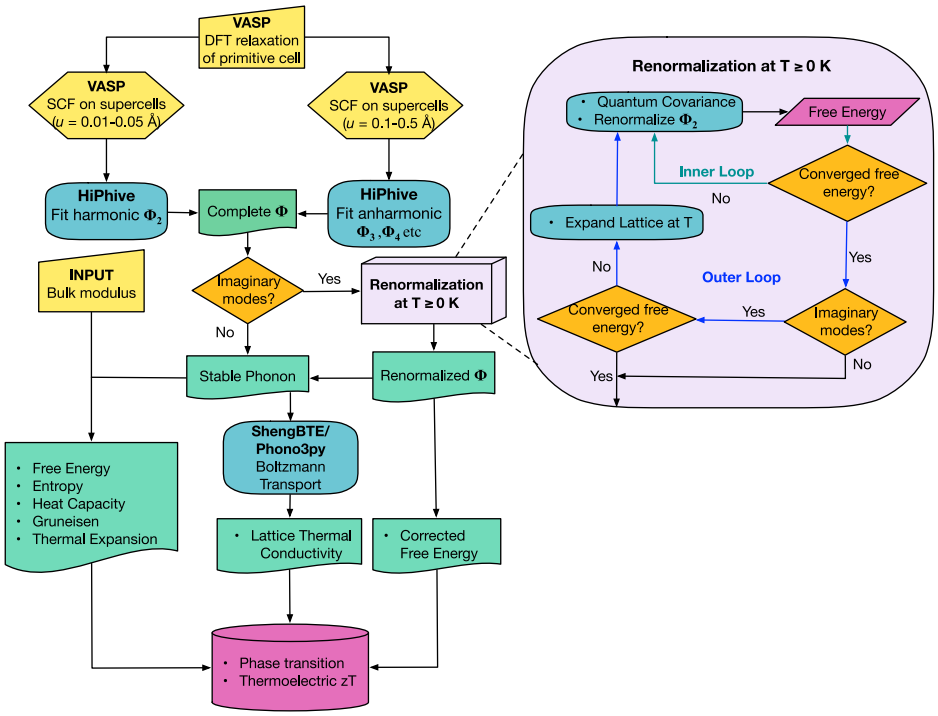

Fig. 1 | Overview of the lattice dynamics workflow. Preparatory steps and required manual inputs are colored in yellow. Note that bulk modulus as the input can either be retrieved from the Materials Project database or obtained from DFT calculations using the previously implemented workflow in atomate[60]. The core steps of the workflow are colored in fog blue. Decision points are colored in orange. Quantities of interest stored in the database are colored in ocean green. Other quantities (intermediate data or targeting results), which are not stored in the database, are colored in magenta.

Step 4: Lattice thermal conductivity from the Boltzmann transport equation calculated by ShengBTE[32,33] or Phono3py[34].

The overall workflow is written within the open source package of atomate[27], which works in conjunction with the Fireworks[35] package for job submission and management. Common structure generations and transformations are handled by pymatgen[36] and ASE[37] codes. Atomate streamlines the creation of standard workflows for various material properties with minimal input, usually only a crystal structure. The atomate package has been used in many downstream materials science studies to reproduce various types of simulations on users' systems of interest. Some of the capabilities added through the integration with atomate include:

- automated job submission and coordination at computing facilities
- job metadata storage in database
- simulation output parsing and storage in database
- job error recovery
- file I/O between simulations

Initial candidate structures are retrieved from the MP database[7].

Parameters benchmarking

We note that the overall workflow requires the selection of a multitude of input parameters. Such parameters are typically adjusted manually, which is infeasible for high-throughput calculations. Here, we provide a concise summary of the most critical parameters of our workflow with benchmarking, with the goal of enabling automatic selection of optimal inputs and ensuring consistency across large-scale calculations.

Functional. It is essential to obtain precise lattice parameters for accurate phonon calculations. In the workflow, all DFT calculations are performed using Perdew-Burke-Ernzerhof revised for solids (PBEsol)[38] exchange-correlation functional and projector-augmented wave pseudopotentials (PAW)[39,40]. PBEsol is selected over PBE because the latter tends to overestimate lattice parameters[38]. Furthermore, from the recent work by Ning et al., the PBE functional is shown to underestimate the phonon frequencies in their testing cases of GaAs, Fe, NiO[41]. The r^{2}SCAN functional is also a viable option, but it is not set as the default due to computational costs. According to the extensive benchmarking by Kingsbury et al.[42], meta-GGA (generalized gradient approximation) functionals such as r^{2}SCAN are usually 2-4 times more expensive than GGA functionals such as PBE and PBEsol.

Supercell size. Large supercells can describe higher degrees of disorder, allow larger cluster sampling cutoffs, and reduce errors due to the periodic boundary condition, but at the expense of higher computational costs. The upper bound of the cutoff radius (the largest interaction distance of any order IFC) is capped by the supercell size. In theory, accuracy should improve as the cutoff radii for IFC fitting and supercell sizes increase because a larger number of IFC components can more physically describe lattice dynamics (barring overfitting). For instance, Eriksson et al.[25] showed that the phonon modes of monolayer MoS_2 are converging when the second-order cutoff radius is larger than 9 Å, requiring a supercell lattice of at least 18 Å. However, larger supercells incur greater CPU and memory requirements.

To optimize the supercell size selection, we compared the root-mean-square-error (RMSE) of IFC fitting and the computational costs (VASP runtime) for 9 cubic structures including body-centered cubic (Ba, Ca, Sr), zincblende (GaAs, GaP, BP) and rocksalt (NaCl, MgO, BaO) in 3 supercell sizes (Fig. S2) and averaged the results for each size (Fig. 2a). For the consistency of the computational cost evaluation, we run all DFT calculations using 4 Cori-KNL nodes on NERSC (Intel Xeon Phi 7250 Processor: 68 cores@1.4GHz, 96GB DDR4, 16GB MCDRAM memory per node). The HiPhive step duration for different fitting methods (Fig. 2b) was done on 1 NERSC Perlmutter CPU node (2x AMD EPYC 7713 Milan CPUs: 64 cores per CPU@2.0GHz, and 512GB DDR4 memory).

We found that a lattice vector of around 20 Å is the optimal choice combining high accuracy and relatively low costs (average RMSE of 7.0 meV/Å and VASP runtime of 10.4 node hours). Utilizing a larger supercell of 24 Å yields a modestly reduced average RMSE (6.4 meV/Å) but comes at more than 2 times increase in computational demands compared with 20 Å. In a smaller supercell (~16 Å), the averaged RMSE increases marginally by 3.2 meV/Å, but crucially, certain materials (BaO, GaP, and GaAs) exhibit an over 30% RMSE increase. The individual RMSE and VASP runtime for each structure is presented in Supplementary Information Fig. S2. This inconsistency, coupled with unconverged anharmonic properties, especially for

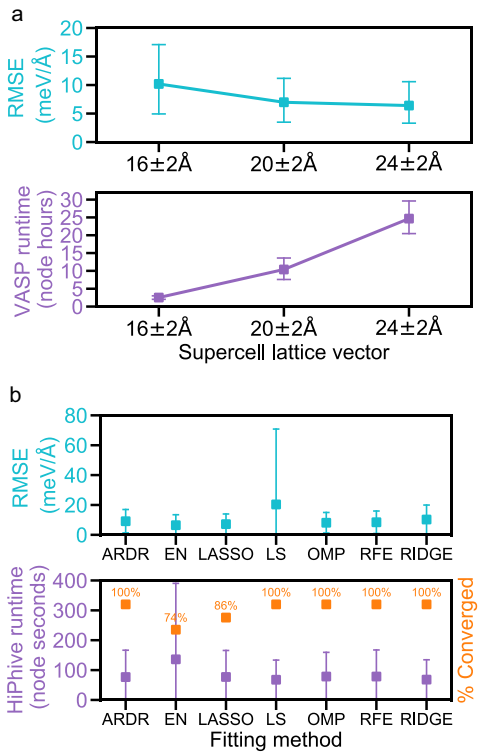

Fig. 2 | Supercell size and fitting methods benchmarking in terms of accuracy and computational costs. a The averaged RMSE (marked in blue squares) and VASP runtime (marked in purple squares) of 9 cubic structures at different supercell sizes (16, 20, 24 Å). The lines show the trend of RMSE and runtime as supercell size increases. The averaged numbers of atoms in supercells across the 9 cubic structures are 120 (16 Å), 229 (20 Å), and 388 (24 Å), respectively. **b** The averaged RMSE (marked in blue squares), HiPhive runtime (marked in purple squares), and the ratio of converged fittings (% converged, marked in orange squares) across various fitting methods for the same benchmark suite of 9 cubic structures. Here ARDR, EN, LASSO, LS, OMP, RFE, and RIDGE in fitting method represent Automatic Relevance Determination Regression, ElasticNet regularization, Least Absolute Shrinkage and Selection Operator, Least Squares, Orthogonal Matching Pursuit, Recursive Feature Elimination and Ridge regression, respectively.

BP and BaO as shown in Fig. S11(c, d), renders 16 Å an inadequate supercell lattice for reliable results. Thus, to strike a balance between accuracy and computational efficiency, a carefully benchmarked supercell size (lattice vectors 20 ± 2 Å and a minimum of 150 atoms per supercell) has been adopted in our workflow. It should be noted that, for low-symmetry lattices that are particularly obtuse or acute, we perform a Lenstra-Lenstra-Lovasz[43] lattice basis reduction to obtain a better set of orthogonal lattice vectors as the initial input structure.

Cutoff radius. Selecting the cutoff radius (maximum interaction distance between atoms in a cluster space) for different orders of IFC is critical for quality fitting. The number of IFC components grows combinatorially with cutoffs, which can be especially limiting for higher-order IFCs, even with implemented symmetry and sum rules constraints in HiPhive.[25,44]. This again requires a trade-off of accuracy and computational feasibility at the high-throughput level. If too large cutoffs are used, not only does computational demand quickly increase to an undesirable level, but the accuracy may also decrease due to parameter-equation imbalance by underfitting. If cutoffs are too small, on the other hand, lattice dynamics may not be captured to a sufficient level of physical accuracy.

To balance these considerations, we implemented an automatic cutoff generator based on the input structure's mean elemental period. Table 1 lists the cutoff values for 2nd, 3rd, and 4th order, corresponding to the mean elemental period number (abbreviated as the period number in Table 1, ranging from 1 to 7). A compound's mean elemental period is calculated according to each element's period number and the chemical stoichiometry. For a given compound, we explore 2 to 3 cutoff options in each order, resulting in (up to) 27 cutoff sets for [2nd, 3rd, 4th] IFCs fitting. All valid cutoff sets are calculated via HiPhive process; the best-fit is subsequently employed for the rest procedures of the workflow. The RMSE convergence test for 9 cubic systems (see Fig. S1 and S3) shows our cutoff generator leads to converged fitting results. Furthermore, the choice of a second-order cutoff significantly impacts the convergence of RMSE, restricted by supercell sizes. Effective coordination of supercell size and cutoff parameters is essential for achieving RMSE convergence. Taking cubic GaAs as an example (Fig. S1), in $4 \times 4 \times 4$ supercell (16 Å lattice), the RMSE fluctuates at ~ 20 meV/Å among 27 sets of cutoffs before converging to 10 meV/Å in $5 \times 5 \times 5$ supercell (20 Å lattice).

IFC Fitting Approach. We compare IFC fitting methods primarily focusing on three criteria: the error (RMSE), the cost (HiPhive runtime), and the convergence (% Converged). We applied each set of cutoffs for the same suite of 9 cubic structures as in Fig. 2a in a 20 Å lattice and measured the time of the HiPhive fitting. If it can not be finished within our wall time (2 hours here, which should be rather sufficient based on our experience), this single fitting is regarded as unconverged, and we compute the ratio of converged HiPhive fitting operations among 9 (materials) × 27 (cutoff sets) total runs for each fitting approach. From our benchmarking results depicted in Fig. 2b, least squares (LS) is the fastest method but it suffers from low accuracy and high uncertainty, evidenced by a high RMSE and large error range. This is unsurprising since IFC matrix is known to be sparse and near-sighted, and least-squares cannot favor sparse recovery[21]. All other methods are capable of favoring sparse recovery and offer obviously better RMSE than the least-squares. Among them, we find that 26% of the cases using the elastic net regularization and 14% of the cases using the least absolute shrinkage and selection operator (LASSO) fail to converge, which are substantial proportions. Recursive feature elimination (RFE) shows very similar RMSE and speed to LASSO (see Table S3) but with essentially 100% convergence. Automatic relevance determination regression (ARDR) has a bit higher RMSE than RFE. All things considered, we select RFE as the default IFC fitting method in the workflow (see Table 2), but the user may override this default with another method facilitated by our integration of importing methods from Scikit-learn[45]. We note that the use of RFE has also been recommended by the HiPhive developers[46].

There also exist multiple modes of IFC fitting in terms of the order in which they are fitted and to what displacements. The most prevalent practice is fitting all IFC orders in a single shot, but one could fit them separately, e.g. harmonic-first then anharmonic, or sequentially by order, using the same or separate training sets for each. Fitting all IFCs in one shot using all training configurations sounds most straightforward, and ideally, forces of the correct order would perfectly distribute to IFC parameters of their corresponding orders. In practice, however, the inevitably imperfect training set and fitting errors cannot guarantee this to a desirable extent. Fitting harmonic IFCs using large displaced training structures is problematic and lacks physical justification. Harmonicity is predominantly present in the region of small displacements, while anharmonicity becomes significant at larger displacements. Numerically, large atomic displacements in fitting harmonic IFCs at 0 K could fictitiously attribute what in reality are anharmonic contributions to harmonic IFC description, as has been observed by a recent systematic study of this matter[47]. As such, we decide to fit harmonic IFCs first to forces in the small-displacement set, and then fit anharmonic IFCs to the residual forces from the large-displacement set with the harmonic forces subtracted out. This decision of separate order, separate-training-set fitting affects our subsequent decision of the supercell perturbing method.

Perturbing methods. A significant advantage of the method used in our workflow lies in employing the high-information-density configurations (displacing each atom in a supercell), resulting in reduced computational

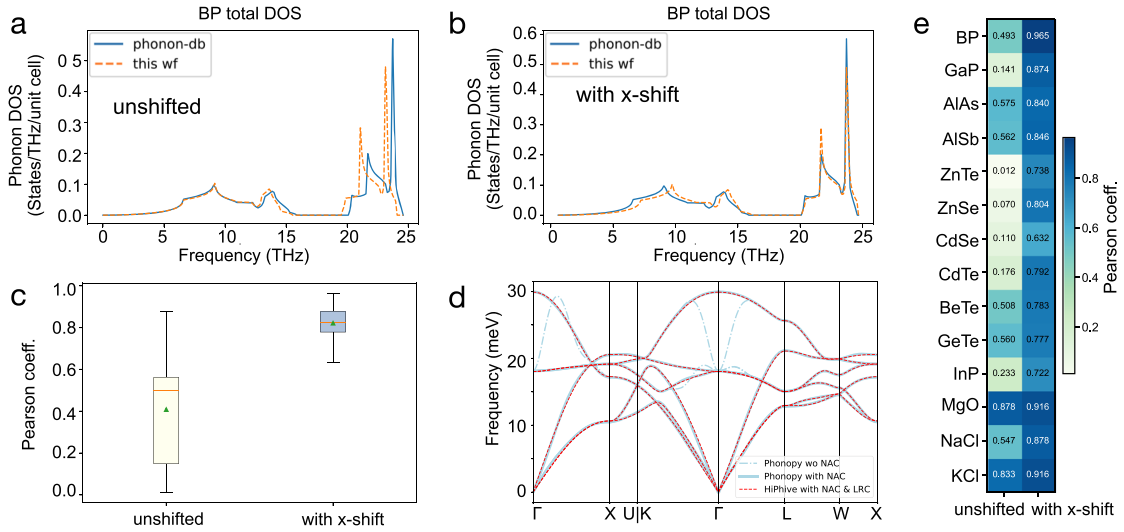

Fig. 3 | Accuracy benchmarking of phonon density of states and the non-analytic term corrected phonon dispersion of NaCl. a, **b** The phonon DOS of BP computed from our workflow and data from phonon-db[52] before any x-axis shifting (unshifted) and after aligning the highest peak (with x-shift). **c** The boxplot of Pearson coefficients shows the enhanced phonon DOS similarity after alignment (with x-shift) from 14 materials labeled in **e**. **d** Ionic crystal NaCl needs a combination of non-analytic correction (NAC) and long-range correction (LRC) to replicate the phonon dispersion computed by Phonopy (with NAC term). **e** The heatmap of Pearson coefficients among 14 materials (BP, GaP, AlAs, AlSb, ZnTe, ZnSe, CdSe, CdTe, BeTe, GeTe, InP, MgO, NaCl, KCl). All DOS plots are given in Fig. S7 and S8.

costs, particularly in terms of VASP runtime. Therefore, the method used to perturb the structure becomes crucial. We explored two approaches: the Monte Carlo (MC) Rattle method[25] and the standard Fixed Displacement (FixedDisp) method.

The MC-rattle method, as implemented in HiPhive, generates realistic random atomic displacements with a Monte Carlo step to avoid unrealistically small interatomic distances which may lead to large repulsive forces. Initially, we explored this method due to its capability to produce physically representative displacement patterns. However, we opted against its continued use for two primary reasons: (i) The difficulty of separating small and large displacement sets for the aforementioned separate fitting of harmonic and anharmonic IFCs and (ii) the reliance on material-dependent empirical amplitude for rattling. For (ii), we need to introduce a second scheme of tuning the standard deviation for MC rattle to ensure acceptable outputs as shown in Fig. S10, but this is difficult to automate. We note that the predicted thermal properties, including lattice thermal conductivity, are found highly sensitive to these rattling settings (see Fig. S13b for further details). Therefore, we have since opted to perturb supercells with consistent and fixed displacements regardless of material type.

The Fixed Displacement method, as implied by its name, generates displacements of a fixed, predetermined magnitude in random directions for each atom in a given supercell. This approach allows for a distinct separation for harmonic force constant extraction across different structure configurations. We designate displacements of ≤ 0.05 Å for harmonic IFC fitting and > 0.05 Å as anharmonic fitting. We benchmarked two sets of Fixed Displacements: (i) [0.01 Å, 0.03 Å, 0.1 Å, 0.3 Å] (labeled as [1-3-10-30] in the unit of picometer), and (ii) [0.01 Å, 0.03 Å, 0.08 Å, 0.1 Å] (labeled as [1-3-8-10]) (see Fig. S9) by comparing their RMSE from fitting and calculated CTE. The set of [0.01 Å, 0.03 Å, 0.08 Å, 0.1 Å] demonstrates significantly lower RMSE and more accurate CTE compared with experimental data. Consequently, we choose this set as the default displacements for perturbing structures and multiply the configuration number based on the primitive cell size and structure symmetry as discussed in the next session. It is worth noting that the default setting we suggested here [0.01 Å, 0.03 Å, 0.08 Å, 0.1 Å] is not the universally optimal perturbing distances for all materials. For strongly anharmonic materials, such as cubic ZrO_2[48], β-GeTe[49], BCC Zr[50] (as shown in Fig. 7), larger displacements up to 0.5 Å are necessary to avoid large RMSE in IFC fitting and unreliable anharmonic properties caused by insufficient small-displaced training dataset.

Number of configurations. To ensure parameter-equation balance for fitting accuracy, it is imperative to generate a sufficient number of training configurations for DFT force calculations. Lower-symmetry crystals generate more symmetry-independent IFC parameters and necessitate a more extensive sampling of atomic configurations to accurately describe IFC-governing phonon behaviors and vibrational properties, especially when handling anharmonicity. Our determination of configurations is based on two aspects: (i) Symmetry factor, which is set to 1 for the cubic system. Other crystal systems use the factor expressed as $\sqrt{48/N_{\mathrm{symm}}}$, where N_{symm} is the number of point group symmetry operations and 48 is the highest symmetry operation number for O_h point group, (ii) Site factor, which is set as the number of atoms in a primitive cell divided by 2. We multiply the two factors and round it to the nearest integer as the configuration number for each displacement. The default set of displacements are [0.01 Å, 0.03 Å, 0.08 Å, 0.1 Å], which we have discussed above. With this setting for the number of configurations, the calculated outputs (CTE and LTC) demonstrate a high level of accuracy ($R^2 > 0.9$ compared with experimental values, see details in Coefficient of Thermal Expansion, and Lattice Thermal Conductivity sections). We note that usually, a handful (< 10) of perturbed structures prove adequate for extracting higher-order force constants in simple cubic systems such as rocksalt and zincblende (Fig. S12 and Table S4).

Evaluation of phonon density of states

Harmonic phonon dispersion and density of states (DOS) has become routine and readily available to use in many software packages[51] such as Phonopy[16]. We evaluate the harmonic phonon DOS computed from our workflow, comparing it with phonon-db developed by Togo et al. (available in the Materials Data Repository[52]). The assessment is carried out utilizing the Pearson coefficient as a quantitative measure of spectral similarity. As part of the pre-processing procedure, the y-axis values of phonon DOS were scaled by the total area of DOS (normalization) for the unit cell basis. Notably, the direct comparison (labeled as unshifted in Fig. 3c) across 14 materials (BP, GaP, AlAs, AlSb, ZnTe, ZnSe, CdSe, CdTe, BeTe, GeTe, InP, MgO, NaCl, KCl) exhibits a modest correlation in the DOS plots obtained by our workflow and from the phonon-db. The DOS curves look similar but with an obvious major peak shift (Fig. S7). To determine the origin of the observed x-shift in the highest peaks, we compared all inputs provided in phonon-db. The root cause of the observed shift appears to be the

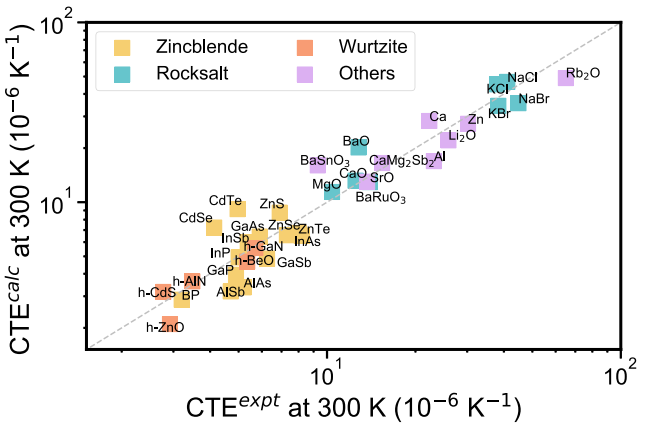

Fig. 4 | Comparison between the calculated linear coefficient of thermal expansion (CTE) and the experimental values at 300 K. The evaluated materials are among various crystal types including: Zincblende (GaAs, GaP, GaSb, BP, InAs, InSb, InP, ZnTe, ZnSe, ZnS, AlAs, AlSb, CdTe, CdSe), Rocksalt (MgO, BaO, CaO, SrO, KCl, KBr, NaCl, NaBr), Wurtzite (h-ZnO, h-CdS, h-BeO, h-GaN, h-AlN) and others (Al, Zn, Ca, Li_2O, Rb_2O, $BaSnO_3$, $BaRuO_3$, $CaMg_2Sb_2$). For anisotropic solids, such as Wurtzites, CTE(||) (to c-axis) is used for the plot. Table S5 lists more detailed data on CTE. $R^2 = 0.91$ from the linear regression fitting here among all the materials.

underestimation of lattice parameters in phonon-db, coupled with substantially lower energy cutoff values (ENCUT) for structure relaxation. In contrast, our workflow adopted a consistent ENCUT of 600 eV (as listed in Table 2), promoting enhanced accuracy and convergence. The volume difference of the relaxed structures due to the difference in choice of ENCUT is plotted in Fig. S6. The consequential adjustment in the x-shift (better alignment as shown in Fig. S8) improves the averaged Pearson coefficients, rising from 0.41 to 0.82 for the evaluated 14 materials (Fig. 3c, e).

In ionic or polar solids, long-range Coulomb interactions lead to a significant separation between the longitudinal optical (LO) and transverse optical (TO) phonon modes at and near the zone center ($\vec{q} \rightarrow 0$), leading to LO-TO splitting near the Γ point. To accurately represent the phonon dispersion in such systems, a non-analytic correction (NAC) term should be included, which accounts for the macroscopic electric field generated by the relative displacements of charged ions.

Within Phonopy, the NAC term is typically integrated into the calculation of the dynamical matrix by specifying the Born effective charges ($\mathcal{Z}^*_{i,\alpha\beta}$) and the electronic part of the static dielectric tensor ($\epsilon^{\infty}{}_{\alpha\beta}$)[53]. However, the direct application of the NAC term in the HiPhive fitting can not fully address the long-range Coulomb interactions arising from the macroscopic electric field. It shows discernible oscillations in phonon modes at the Γ point, as illustrated in Fig. S5b when applying the NAC only. To rectify this, it is crucial to exclude the long-range force contributions of 2nd-order force constants, akin to the method implemented in CSLD[54], followed by a refitting of the short-range force constants. Subsequently, the long-range contributions are reincorporated into the short-range IFCs and the overall IFCs yield a precise phonon dispersion with this long-range correction (LRC). Fig. 3d demonstrates the congruence of the harmonic phonon dispersion derived from HiPhive (with NAC and LRC) and that from Phonopy (with NAC) in NaCl. For less ionic (more covalent) crystals, the influence of the NAC term is marginal, as evidenced in BP (see Fig. S5a). We offer users the ability to enable the NAC-LRC correction feature in the workflow if they want to prevent LO-TO splitting in harmonic phonon dispersions. By default, this setting is turned off to conserve computational time due to its negligible impact on anharmonic properties such as CTE and LTC.

Coefficient of thermal expansion

The coefficient of thermal expansion describes the change of lattice parameter (linear CTE) or volume (volumetric CTE) as a function of temperature. The measurement of thermal expansion coefficients is pivotal in understanding the thermomechanical properties of materials in applications ranging from metal and alloys in aerospace to electronic components[55]. While the experimental approach for CTE measurements can be employed using either an absolute method or relative techniques regarding a reference material[56], theoretical studies have traditionally relied on the quasi-harmonic approximation (QHA)[57–59].

In this workflow, we compute the thermal expansion coefficient using the Grüneisen parameter through the associated Phono3py[31] routine (see Methods for more details). Quantities including 3rd-order IFCs, heat capacity, and bulk modulus are required to calculate the Grüneisen parameter and subsequently the linear CTE. Only the bulk modulus is retrieved from the MP database or determined from a separate workflow[60] while other intermediate data are obtained from this workflow. In Fig. 4, the calculated CTE for more than 30 materials covering Zincblende, Rocksalt, and Wurtzite crystal systems are compared with the experimental CTE (see detailed values in Supplementary Information Table S5). The data are presented in log-log scale to effectively capture the wide range of evaluated materials' CTE, spanning from 10^{-6} to 10^{-4} K^{-1}. The coefficient of determination (R^2) from linear regression among 35 materials' raw data (not log-log scale) is 0.91, indicating our computed CTE from the workflow is in excellent agreement with the experimental data.

Lattice thermal conductivity

We calculate lattice thermal conductivities in the reciprocal space using Boltzmann transport formalism and perturbation theory to phonon-phonon scattering[22,61]. We limit ourselves to three-phonon scattering for two reasons. First, the four-phonon scattering calculation is too demanding to be carried out at a large scale[54,61], which is our ultimate objective. Second, three-phonon scattering is sufficiently accurate in most cases[62], and even when its accuracy becomes insufficient, e.g. for highly anharmonic materials with very low LTC and strong 4th-order interaction, it is still sufficient for predicting qualitatively very low LTC for those materials[47]. These materials, not many in the grand scheme of things, can be subject to deeper, targeted studies involving four-phonon scattering once identified.

Here, we present a comprehensive analysis of LTC for 29 binaries and ternaries (mostly overlapped with the materials in the CTE section, the difference is due to the availability of experimental data for particular structures). As noted earlier, we found the thermal conductivity calculation is very sensitive to the supercell size: a smaller-than-recommended supercell (e.g., 16 ± 2 Å) can lead to an incorrect prediction of LTC (see Fig. S11). We store the 2nd and 3rd orders of IFCs in the proper formats for subsequent use in either ShengBTE or Phono3py for LTC calculations. The default choice is using ShengBTE since Phono3py[34] tends to slightly underestimate the LTC as compared to ShengBTE from our tests (Fig. S13a, S14a) and requires more computational time (Fig. S14b). However, for certain cases such as hexagonal Wurtzite materials in Fig. 5, in which a locally adaptive broadening algorithm in ShengBTE can not ensure convergence, we used Phono3py (tetrahedron method implemented for the Brillouin zone integration) to perform the LTC calculations in the workflow.

Figure 5 and Table S6 demonstrate our calculated thermal conductivity in this workflow is in remarkable agreement with the experimental values ($R^2 = 0.98$ for LTC data). The benchmarking results for both CTE and LTC as downstream outputs validate the accuracy of the anharmonic force constants produced in our workflow using the selected parameters.

Speedup Estimation

We benchmark the speedup of our workflow relative to the conventional finite-displacement method (configurations generated by Phono3py) by comparing the consumed node hours of DFT calculations. The primary focus is on the vastly reduced VASP runtime due to the reduced number of displaced-supercell configurations for quality results. Extra time is required for the IFC fitting and LTC calculations, even though the DFT process takes

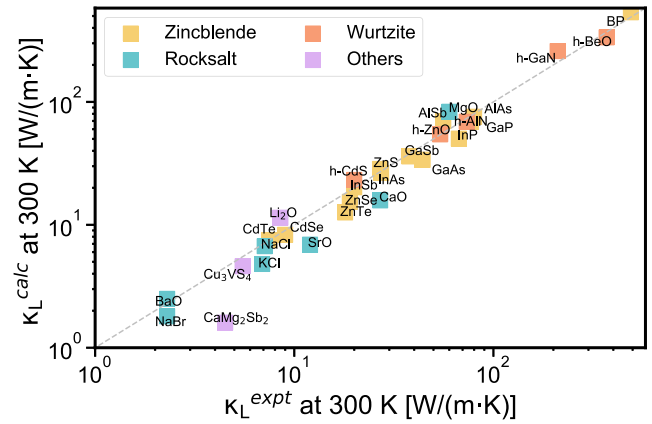

Fig. 5 | Comparison between the calculated lattice thermal conductivity (LTC) and the experimental values at 300 K. The evaluated materials are among various crystal types including Zincblende (GaAs, GaP, GaSb, BP, InAs, InSb, InP, ZnTe, ZnSe, ZnS, AlAs, AlSb, CdTe, CdSe), Rocksalt (MgO, BaO, CaO, SrO, KCl, NaCl, NaBr), Wurtzite (h-ZnO, h-CdS, h-BeO, h-GaN, h-AlN) and Others (Li_2O, $CaMg_2Sb_2$, Cu_3VS_4). Table S6 lists more detailed data on LTC. $R^2 = 0.98$ from the linear regression fitting here among all the materials.

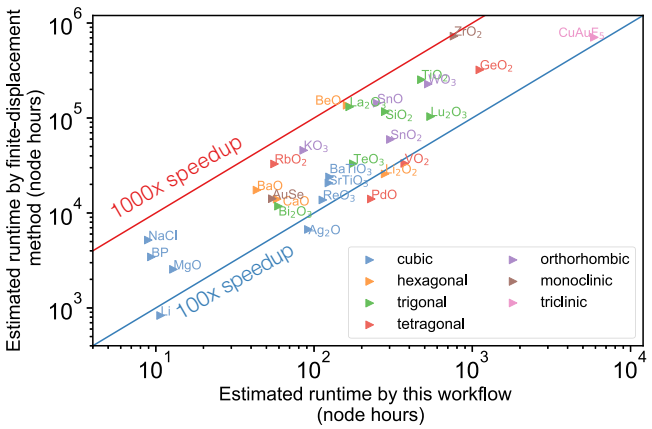

Fig. 6 | Performance speedup via estimated VASP runtime (in node hours) using this workflow compared with the conventional finite-displacement method for 3rd-order anharmonic IFCs extraction. The testing structures are binaries and ternaries in seven crystal systems (cubic, hexagonal, trigonal, tetragonal, orthorhombic, monoclinic, and triclinic). The VASP runtime is the dominant portion of workflow runtime (85% of the total time) based on our tests as shown in Fig. S14c. Supercell configurations for the finite-displacement method are generated by Phono3py package (usually ≥ 1000 configurations). The estimated runtime is calculated by multiplying the configuration number and the averaged static calculations from 3–5 random configurations. More details are given in Table S4.

up the majority of time (85%) in the workflow as shown in Fig. S14c. This assessment was conducted across a range of binary and ternary compounds, encompassing all seven distinct crystal systems. The conventional finite-displacement method requires many more configurations than we need in this workflow. We averaged five random samples (by Phono3py) and three configurations (by our workflow) to estimate the total VASP runtime. All the VASP calculations were done under the same supercomputer system setting (4 Cori-KNL nodes on NERSC, detailed node info: Intel Xeon Phi 7250 Processor: 68 cores@1.4GHz, 96GB DDR4, 16GB MCDRAM memory per node) to maintain consistency. The runtime, number of configurations, and supercell size are given in Table S4. Our workflow requires 2-3 orders of magnitude less computational time compared with the finite-displacement method as shown in Fig. 6, which finally makes widescale calculation of anharmonic thermal property calculations attainable.

Finite-temperature phonons

Although the 0 K phonons reasonably describe the lattice dynamics of many materials, they break down for materials with imaginary frequencies (unstable phonon modes). Such dynamical instability occurs when the harmonic potential is inverted in certain vibrational directions. Materials with dynamical instability can only manifest in practice if the temperature is high enough to increase atomic displacements beyond the inverted harmonic region, at which point the system can be stabilized by anharmonic potentials. This high-temperature lattice dynamics can yield effective harmonic phonons that are fully real to the indication of dynamical stability. In this workflow, we adopt a renormalization procedure[63] using 4th-order anharmonic force constants to efficiently implement finite-temperature phonon calculations. This method's advantage lies in not requiring any additional DFT calculations beyond those for the initial fitting for harmonic and anharmonic force constants. A detailed explanation of the renormalization step is provided in the Methods section.

We benchmark the effect of renormalization for three materials: ZrO_2, GeTe, and Zr. The tetragonal-to-cubic transition of ZrO_2 occurs at a very high temperature of 2650 K[64,65], and the harmonic phonon dispersion of cubic ZrO_2 at 0 K shows severe instability with very negative imaginary modes as shown in Fig. 7b. As we increase the renormalization temperature, we find the imaginary phonon modes at the X point turning real at 2500 K, which is very close to the experimental transition temperature of 2650 K. However, as we discuss in more detail later, phonons becoming real is a necessary but not sufficient condition for the phase's manifestation. In GeTe, a semiconductor that exhibits a reversible phase transition between a rhombohedral and cubic structure, our results indicate the imaginary modes of the cubic phase at the Γ point vanishing around 400 K. However, the cubic phase does not manifest experimentally until 650 ± 100 K[66]. In BCC Zr, we find the imaginary modes around the N point becoming real at around 700 K, but the experimental HCP-to-BCC phase transition temperature is much higher at around 1135 K[67]. For a phase to manifest, it must not only be dynamically stable (real phonons) but also have the lowest free energy among all competing phases. This is investigated in the next section.

Free energy and phase transition

For a phase to be the most stable and thus manifest experimentally, it must be both dynamically stable and should have a lower free energy than all other competing phases. Fig. 8 shows that our workflow can describe phase transition rather well with renormalized phonons and vibrational free energy.

Upon heating, Zr undergoes HCP-to-BCC phase transition at approximately 1135 K[67]. By the quasi-harmonic free energy of temperature-dependent (TD) phonons (F_{QHA}) only, the HCP phase yields lower free energy than the BCC phase from 0 K to 1500 K(ΔF always positive), failing to predict the HCP-to-BCC phase transition entirely. Once anharmonic free energy correction (F_{anh}) is applied (see Methods for details), the differential free energies (ΔF) of Zr can cross zero at around 1100 K (Fig. 8a), demonstrating a very accurate prediction for phase transition temperature. Since Zr is a metal, we separately calculated and included electronic free energy in plotting Fig. 8a, which is computed as

$$F_e = -k_B T \int D\left[f\log(f) + (1-f)\log(1-f)\right] dE \qquad (2)$$

where $f(E)$ is the Fermi-Dirac distribution and $D(E)$ is the electronic density of states calculated using the tetrahedron method[68]. Electronic free energy is not a part of the workflow, however.

In the case of ZrO_2 also, F_{QHA} alone fails to describe the tetragonal-to-cubic phase transition at all. The tetragonal phase always has a lower free energy than the cubic phase without any correction. If F_{anh} is accounted for, then the cubic phase has lower free energy than the tetragonal phase when it becomes dynamically stable between 2000 K and 2500 K. (see Fig. 8b). This is an intriguing outcome because a similar but more methodical study of ZrO_2 entirely within the self-consistent phonon theory did not predict the

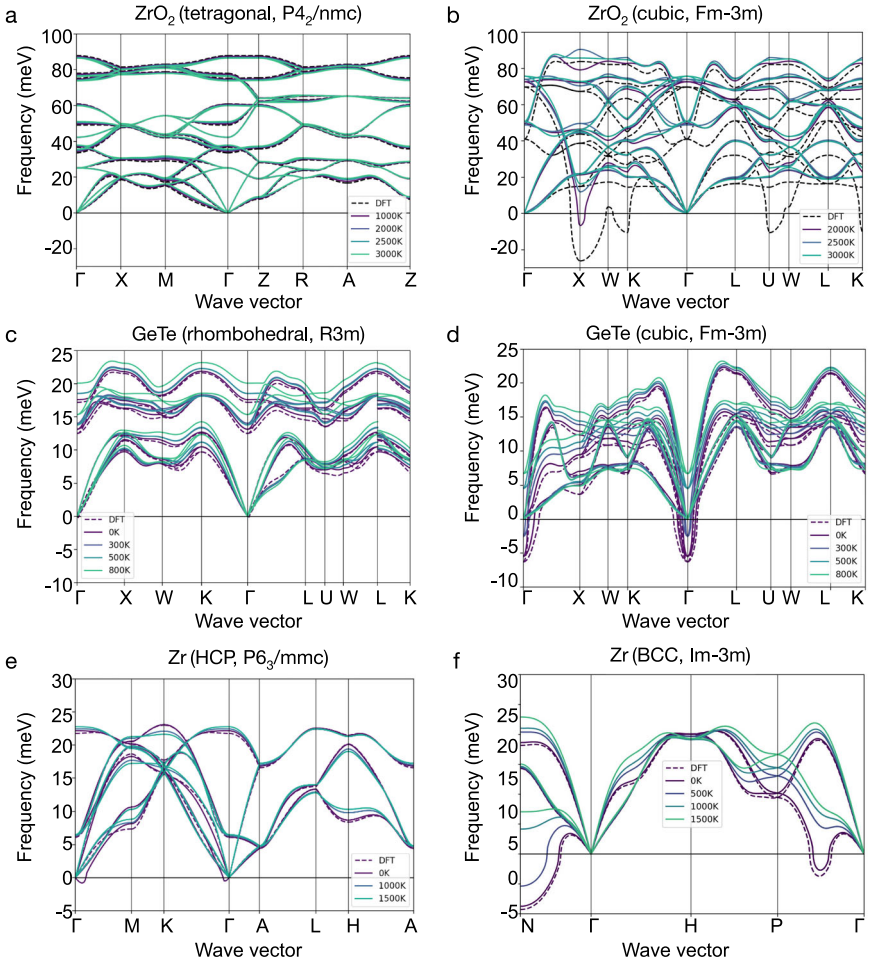

Fig. 7 | Finite-temperature phonon dispersions of ZrO_2, GeTe and Zr in different phases after renormalizations. a, **b** tetragonal and cubic ZrO_2, **c**, **d** rhombohedral and cubic GeTe, **e**, **f** HCP and BCC Zr. The dash lines labeled as DFT are harmonic phonon dispersions before renormalization. The temperatures at which we perform renormalizations are indicated in the legend. To keep the plot clean and uncluttered, only a subset of the renormalization temperatures is displayed here.

transition with the tetragonal phase always having a lower free energy than the cubic phase[69]. Its main differences from our study are the method of TD phonon generation and the application of an additional, higher-order free energy correction. The discrepancy between the two results is a testament that the temperature of phase transition, and whether or not it occurs, is sensitive to even relatively small uncertainties in energy (DFT, harmonic, or anharmonic).

In Fig. 8c, our free energy calculation with anharmonic correction predicts the rhombohedral-to-cubic transition temperature of GeTe at around 430 K, which is lower than reported value (650 K). QHA alone appears to provide a better prediction of 520 K even though it is a lower level of theory. Here too at work is the high sensitivity of phase transition temperature to small variations in energy. The free energy difference between the two phases is marginal, at less than 4 meV/atom from 300 K to 600 K. A tiny uncertainty in the computed energy values of ground state DFT energy or harmonic/anharmonic contribution to vibrational energy would disproportionately swing the critical temperature prediction. It should be highlighted that spontaneous anisotropy is a common feature in cubic IV-VI materials like GeTe[70], which could account for the observed shallow energy differences. Enhancing the prediction accuracy for GeTe would require more precise techniques for calculating both the vibrational free energy and the ground state total energy.

Discussion

Phonon anharmonicity can provide solutions to many aspects of fundamental problems such as the thermal properties, mechanical properties, and phase transition behavior in crystals. The emerging hybrid machine learning and DFT-based tools like the HiPhive package make high-order force constants extraction accessible with acceptable computational costs. However, any model with abundant parameters faces the dilemma of prediction accuracy and the consumed time of determining and benchmarking parameters for specific systems. Here, we incorporate all necessary steps into a smooth pipeline with pre-tested parameters from evaluations in diverse materials and demonstrated accuracy against experimental measurements. The ultimate goal of this work is to produce the same level of accurate results as direct calculations from Phonopy (harmonic phonons) and Phono3py (anharmonic properties) but with much less computational expense and at a large scale.

Next, we point out aspects of the workflow that can benefit from further computational efficiency or accuracy of property calculation in the near future. While the workflow in its current form serves as a good first step as evidenced by its performance across the ~ 30 compounds tested, more complex crystal structures will require additional testing.

The first set of improvements would come with the upcoming release of atomate2, a newer version of atomate (where this workflow is currently implemented) under development. In the atomate2 framework, the dynamic workflow that incorporates intermediate results and allows for post-submission updating is more feasible[71]. Ideally, the process of adding training supercells and fitting force constants should be iterative, aiming to reduce computational time while ensuring quality convergence. The training set can be augmented over multiple fitting trials until a desired accuracy is met. For example, atomate2 facilitates scaling the number of additional training structures and the order of anharmonic fitting based on harmonic fitting outcomes, allowing for dynamic and iterative addition of training structures until a desired numerical accuracy is met. Another significant advantage of atomate2 is its flexibility in swapping the DFT calculator for a surrogate model, such as a machine-learning (ML) model. This allows for running workflows with ML frameworks like MACE[72],

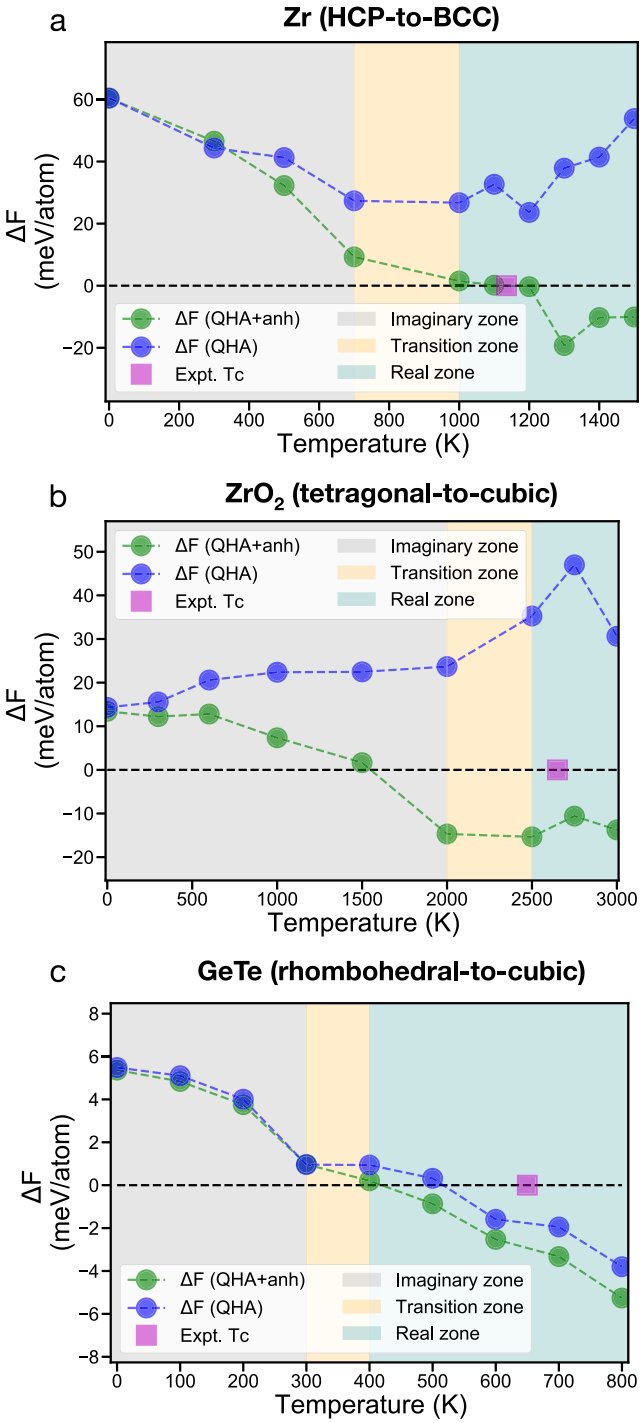

Fig. 8 | Vibrational free energy differentials (ΔF) determine phase transition temperatures for Zr, ZrO_2 and GeTe. a HCP-to-BCC phase transition in Zr, **b** tetragonal-to-cubic phase transition in ZrO_2, **c** rhombohedral-to-cubic phase transition in GeTe. The blue lines represent free energies calculated using the quasi-harmonic approximation, denoted as ΔF(QHA), whereas the green lines correspond to free energies that include an additional anharmonic correction, indicated as ΔF(QHA+anh). The experimental phase transition temperatures (Expt. T_c) are marked in magenta squares. We separate all temperatures into three regions depending on whether the phonon modes are imaginary or not. Temperatures are categorized into three distinct zones based on the phonon modes, as illustrated in Fig. 7. The "imaginary zone" encompasses the temperature range where phonon modes are imaginary. Conversely, the "real zone" refers to where phonon modes are entirely real. The region between the highest imaginary-mode temperature and the lowest real-mode temperature is defined as the "transition zone" here. In this zone, the imaginary modes gradually vanish and become real. The width of this "transition zone" is determined by the interval of renormalization temperatures. Here the "transition zone" is narrower than in Fig. 7 since we only show a subset of the renormalization temperatures in Fig. 7.

M3GNet[73], CHGNet[74], or other machine-learning force fields (MLFFs) supported by atomate2 without any modifications to the workflow. As universal MLFFs improve in accuracy, they become increasingly viable for use in this context. We note that this should also ease the transition of the workflow to other DFT packages such as ABINIT[75].

The second set of improvements could come by adopting more reliable physical methods to implement anharmonic free energy correction. At the level of perturbation theory, explicit forms of free energy correction due to 3rd-order anharmonicity have been implemented and used in the literature[76–79]. This would serve as a higher-order free energy correction to what is currently implemented in this workflow. At a different level, a method implementing thermodynamic integration (TI) using only IFCs with no need for molecular dynamics has demonstrated capability for accurate prediction of anharmonic free energy and phase transition across strongly anharmonic materials from shape-memory alloys to the materials presented in this paper[80]. This method uses about 3 ~ 4 orders of magnitude less computation time than TI performed with AIMD. If properly implemented, TI by design offers a complete description of anharmonic free energy whether or not a system is dynamically unstable by accounting for the actual potential difference between a temperature-dependent harmonic potential and the actual potential (or an approximation thereof). Though it would add to the computational expense beyond the essentially instantaneously calculable correction formula we adopt here, its computational demand is still within high-throughput applicability.

We anticipate our workflow would accelerate inorganic compounds phase transition studies, for example, the polymorphic transitions in $BaTiO_3$[81,82], $SrTiO_3$[83], $CsPbI_3$[84] and ZrS_2[85] by eliminating or reducing the need for ab initio molecular dynamics. Our implemented renormalization scheme relies solely on 0 K force calculations to generate temperature-dependent displaced configurations. Additionally, the workflow provides valuable CTE insights for thin film, and may be used in the future to facilitate the selection of substrate/film combinations that minimize CTE mismatch. It might also be applied to screen compounds with negative CTE values, such as $MZrF_6$ (M=Ca,Mg,Sr)[86], ZrW_2O_8[87], silicon clathrate frameworks[88], $LnFe(CN)_6$ (Ln = Y, La)[89]. For LTC calculations, our workflow boosts both accuracy and efficiency compared with previous models or methods[90–93]. In the future, our plan is to extend our investigated systems to more complex materials such as ZrCuSiAs and $CsAg_7S_4$[94].

In summary, we have developed a workflow that can efficiently extract high-order anharmonic IFCs and compute thermal properties within an automated pipeline with the input of merely the crystal structure and bulk modulus (for which other atomate routines exist[60]). Parameters are evaluated to optimize throughput and accuracy. The benchmarked parameters, as shown in Supplementary Tables S5 and S6, accurately capture third-order anharmonic properties across various materials. For the comparison between experimental and calculated CTE, R^2 values of 0.84, 0.89, and 0.91 were achieved for metals, semiconductors, and insulators, respectively, demonstrating high accuracy. Similarly, for LTC, R^2 values of 0.97 for metals and semiconductors, and 0.98 for insulators, highlight strong predictive capability. The computational costs primarily depend on crystal symmetry and primitive cell size, which affects the number of configurations needed for force constant fitting. For dynamically unstable materials, in which the harmonic approximation is not capable of computing thermal properties, we adopt a renormalization procedure to incorporate the temperature effect and stabilize the high-temperature phase. The phase transition temperatures after considering temperature-dependent free energy corrections achieve less than 10% error. Notably, our workflow attains the goal of 2-3 orders of magnitude speedup compared with the conventional finite-displacement method for 3rd-order IFC calculations. The emergence of this workflow makes high-throughput calculations accessible for general lattice dynamic properties. We will continue to improve our protocols based on the practice of large-scale computation for various materials and new emerging tools, and hope that this method facilitates the practical application of lattice dynamics workflows in both theoretical and applied studies of materials properties.

Methods

Grüneisen parameter and thermal expansion

Grüneisen parameter is a dimensionless quantity measuring the degree of phonon frequency shift with changes in volume. It is described by the following equation:

$$\gamma_{\lambda\mathbf{q}}(V) = -\frac{V}{\omega_{\lambda\mathbf{q}}(V)}\frac{\partial\omega_{\lambda\mathbf{q}}(V)}{\partial V}, \tag{3}$$

where ω is the phonon frequency at wave vector $\mathbf{q}$ and mode index λ. We adopt Phono3py[31] routine to obtain the mode Grüneisen parameters. Phono3py calculates the derivative of phonon frequency with respect to unit cell volume by computing the derivative of the dynamical matrix with 3rd-order IFCs and gain mode-Grüneisen parameters. Then a tensor of mode Grüneisen parameters is averaged based on their contributions to heat capacity to obtain the total Grüneisen parameter (also called the mean Grüneisen parameter) at each temperature according to the equation below:

$$\gamma = \frac{\sum_{\lambda\mathbf{q}}\gamma_{\lambda\mathbf{q}}c_{\lambda\mathbf{q}}}{\sum_{\lambda\mathbf{q}}c_{\lambda\mathbf{q}}} \tag{4}$$

Here, $c_{\lambda\mathbf{q}}$ is the contribution of each mode and wavevector to the constant-volume heat capacity.

The linear thermal expansion coefficient (α) describes how the dimensions of a material change as a function of temperature. For isotropic solids, volumetric thermal expansion β=3α. The volumetric thermal expansion can be directly computed from the total Grüneisen parameter using the following equation:

$$\beta = \frac{\gamma C_v}{VK}, \tag{5}$$

where C_v is heat capacity, primitive cell volume is V and bulk modulus is K. The only parameter that is not computable in the present workflow is K. However, MP already has elastic properties including K computed for many materials, and the atomate workflow for this is available[60] for materials without pre-computed K.

Lattice thermal conductivity

Lattice thermal conductivity is calculated within the unified Peierls-and-Wigner framework of Boltzmann transport, defined as[95]

$$\kappa_{\text{lat}} = \frac{\hbar^2}{k_B T^2 V N_{\mathbf{q}}}\sum_{\mathbf{q}}\sum_{\lambda\lambda'}\mathbf{v}_{\lambda\mathbf{q}}\otimes\mathbf{v}_{\lambda'\mathbf{q}}\frac{\omega_{\lambda\mathbf{q}}+\omega_{\lambda'\mathbf{q}}}{2}\frac{\omega_{\lambda\mathbf{q}}n_{\lambda\mathbf{q}}(n_{\lambda\mathbf{q}}+1)+\omega_{\lambda'\mathbf{q}}n_{\lambda'\mathbf{q}}\left(n_{\lambda'\mathbf{q}}+1\right)}{4(\omega_{\lambda\mathbf{q}}-\omega_{\lambda'\mathbf{q}})^2+\left(\tau_{\lambda\mathbf{q}}^{-1}+\tau_{\lambda'\mathbf{q}}^{-1}\right)^2}\left(\tau_{\lambda\mathbf{q}}^{-1}+\tau_{\lambda'\mathbf{q}}^{-1}\right) \tag{6}$$

where ω is the phonon frequency, $\mathbf{v}$ is the group velocity, τ^{-1} is the scattering rate, and n is the phonon occupation for phonon modes λ and λ'. The diagonal part, $\lambda = \lambda'$, denotes the Peierls term describing particle-like transport that reduces to the usual Boltzmann transport formulation, while the off-diagonal part, $\lambda\neq\lambda'$, denotes the Wigner term describing the coherent tunneling between distinct phonon modes. The Wigner contribution increases relative to the Peierl contribution as modes become closer in energy and their broadening due to scattering increases such that they significantly overlap. As such, the Wigner term manifests substantially in materials with very soft phonons and heavy anharmonic scattering, e.g. $CsPbBr_3$[95], $Cu_{12}Sb_3S_{13}$[96], and increasingly so at higher temperatures. Eq. (6) is evaluated using FourPhonon-ShengBTE[32,33] with in-house modification to incorporate the Wigner contribution (see Code Availability for details). Anharmonic scattering is limited to three-phonon processes arising from third-order IFCs as this is known to yield accurate κ_{lat} for most materials.

Finite-temperature lattice dynamics

Phonon renormalization. Harmonic phonons that are directly calculated using DFT-based methods at 0 K are effectively modified at higher temperatures as atomic displacements reach beyond the harmonic region. Most prominent of such an effect is the stabilization of otherwise dynamically unstable materials at 0 K (imaginary phonon frequencies), which may lead to phase transitions. This not only has a direct consequence for functionalities such as ferroelectricity and shape memory, but also affects the analysis of other thermal properties derived from phonons including lattice thermal conductivity. Several available methods can calculate effective TD phonons. SCAILD (Self-consistent ab initio Lattice Dynamics) iteratively updates harmonic frequencies via temperature-dependent, normal-mode displacements of atoms[97,98]. QSCAILD (Quantum SCAILD) improves upon it by using the so-called quantum covariance of atomic displacements accounting for quantum motion to sample atomic configurations[99,100], and use these to obtain effective harmonic IFC. SSCHA (Stochastic Self-consistent Harmonic Approximation) finds IFCs that minimize harmonic free energy at a given temperature[101]. TDEP (Temperature Dependent Effective Potential) uses AIMD simulations for sampling temperature-dependent configurations, followed by fitting effective IFCs[23,102,103]. Though these represent accurate and reliable approaches, all of them by their standard protocol invoke DFT calculations after every iteration of ensemble generation to compute energies and forces. For high-throughput deployment, a higher computational efficiency is essential.

The approach we adopt most closely follows the method of Refs. 63,104. Similar to some of the other methods listed above, configurations are statistically sampled from a Gaussian probability distribution defined by the quantum covariance of atomic displacements. Quantum covariance of a given pair of atoms a and b in their respective Cartesian directions i and j[99]:

$$\Sigma_{ab,ij} = \frac{1}{2\sqrt{M_a M_b}}\sum_{\lambda}\hat{\mathbf{e}}_{\lambda ai}\frac{2n_\lambda+1}{\omega_\lambda}\hat{\mathbf{e}}^*_{\lambda bj}, \tag{7}$$

where $\hat{\mathbf{e}}$ are phonon eigenvectors, and the temperature dependence enters through phonon mode population n_λ given by Bose-Einstein distribution. Once a temperature-dependent ensemble of configurations is sampled, we in effect use IFCs and not DFT to predict their forces. Specifically, by collecting the sampled configurations into a new displacement matrix $\mathbb{A}(T)$ matching the clusters present in the IFCs, we use the anharmonic IFCs to expediently predict anharmonic forces in those configurations via $\mathbf{F} = \mathbb{A}(T)\Phi$. We then obtain effective temperature-dependent harmonic IFCs, Φ_2^{TD}, that describe the computed forces. These amount to calculating[63]

$$\Phi_2^{\text{TD}} = \Phi_2 + \mathbb{A}_2^{-1}(T)\sum_{d\geq 4}\mathbb{A}_d(T)\Phi_d, \tag{8}$$

where the subscript denotes the order and, as $\mathbb{A}_2$ is usually not a square matrix, $\mathbb{A}_2^{-1}$ is the Moore-Penrose pseudoinverse. As Φ_2^{TD} and $\mathbb{A}(T)$ are interdependent, Eqs. (7) and (8) are iterated until the free energy converges to a desired threshold determined by temperature. This only involves diagonalization of the dynamical matrix constructed from Φ_2^{TD}, between two iterations, to obtain ω and $\hat{\mathbf{e}}$.

Because dynamical instabilities involve double-well potentials that are stabilized by fourth-order anharmonicity, it is critical to accurately obtain Φ_4 at minimum for the successful execution of the process. This iterative convergence routine is merged with the HiPhive codebase, where it resides under the Renormalization class. The primary advantages of this approach are the efficiency of statistical sampling of thermal ensemble via quantum covariance and the direct determination of Φ_2^{TD} by known IFCs without invoking DFT.

Thermal expansion. Thus far, for renormalization, volume expansion is not taken into account. If renormalization predicts completely real phonons at some temperature (T_r), the Grüneisen parameters and the coefficients of thermal expansion (CTE) are calculated straight from third-order IFCs. Once CTE values are accrued at multiple temperatures where phonon is real, the linear thermal expansion fraction at a given T^* can be computed by numerically integrating

$$\left(\frac{\Delta L}{L_0}\right)_{T^*} = \int^{T^*} \alpha(T) dT. \tag{9}$$

Note that anisotropy in thermal expansion is accounted for through the anisotropy of CTE denoted as α. Because we do not perform renormalization at sparse temperature points, and noting that CTE is generally not very sensitive to temperature (in other words, close to being constant), we expand the lattice by $\Delta L \approx \alpha(T)T$ at each temperature, and redo renormalization from the solution obtained at $V = V_0$. This second renormalization process tends to converge after much fewer iterations since the initial solution would already be quite close to the true solution, except that $V \neq V_0$ by only a small difference. The converged results with respect to thermal expansion are the final results at a given T^*.

Anharmonic Free Energy Correction. In the harmonic phonon theory, vibrational free energy and entropy respectively are

$$F = \frac{1}{N_{\mathbf{q}}} \sum_{\lambda \mathbf{q}} \frac{1}{2} \omega_{\lambda \mathbf{q}} + k_B T \cdot \log \left[1 - \exp\left(\frac{\omega_{\lambda \mathbf{q}}}{k_B T}\right)\right], \tag{10}$$

$$S = \frac{1}{N_{\mathbf{q}}} \sum_{\lambda \mathbf{q}} \frac{1}{T} \frac{\omega_{\lambda \mathbf{q}}}{\exp\left(\frac{-\omega_{\lambda \mathbf{q}}}{k_B T}\right) - 1} - k_B \cdot \log \left[1 - \exp\left(\frac{\omega_{\lambda \mathbf{q}}}{k_B T}\right)\right]. \tag{11}$$

These are the equations implemented in Phonopy, and can be used to describe harmonic properties of any real phonon spectrum, including volume-dependent phonon spectrum representing the effects of thermal expansion. However, when phonon frequencies themselves become explicitly temperature-dependent, $\omega \to \omega(T)$, Eq. (10) is no longer an accurate description of vibrational free energy, and must be corrected.

One way of correcting F comes from observing that, even with $\omega(T)$, entropy retains its microscopic definition of Eq. (11), but then Eq. (10) violates the fundamental thermodynamic relation $S = -\left(\frac{\partial F}{\partial T}\right)_V$. Instead, the new relation in $S = -\left(\frac{\partial F_h}{\partial T}\right)_V + \frac{1}{2}\frac{\partial \omega(T)}{\partial T} \coth\left(\frac{\omega(T)}{2k_B T}\right)$. In order to obey the fundamental relation, free energy must pick up a correction term[105,106], namely

$$F_{\text{anh}} = -\frac{1}{2} \sum_{\lambda \mathbf{q}} (\omega_{\lambda \mathbf{q}}(T) - \omega_{\lambda \mathbf{q},\text{h}}) \left(n_{\lambda \mathbf{q}}(T) + \frac{1}{2}\right), \tag{12}$$

where the subscript "h" indicates the original harmonic term. In strict terms, Eq. (12) is only applicable as is when $\omega_{\lambda \mathbf{q},\text{h}}$, the original harmonic frequencies at 0 K, are all real. To use it still in the presence of imaginary $\omega_{\lambda \mathbf{q},\text{h}}$, one could set those imaginary frequencies to 0 and sum over the frequency shift on the real axis. The self-consistent phonon theory, a many-body perturbation theory, provides the explicit correction due to 4th-order interaction (the so-called "loop" correction)[77,105],

$$F_{\text{anh}} = -\frac{1}{4} \sum_{\lambda \mathbf{q}} \frac{(\omega^2_{\lambda \mathbf{q}}(T) - \mathbb{C}^\dagger \omega^2_{\lambda \mathbf{q},\text{h}} \mathbb{C})}{\omega_{\lambda \mathbf{q}}(T)} \left(n_{\lambda \mathbf{q}}(T) + \frac{1}{2}\right), \tag{13}$$

where $\mathbb{C}$ is the unitary transformation matrix between the TD and bare harmonic eigenvectors. Eq. (13) can be directly applied even for imaginary $\omega_{\lambda \mathbf{q},\text{h}}$. Eqs. (12) and (13) are identical when $\mathbb{C} = \mathbb{I}$ and $\omega_{\lambda \mathbf{q}}(T) = |\omega_{\lambda \mathbf{q},\text{h}}|$, and Eq. (13) tends to one-half of Eq. (12) as $\omega_{\lambda \mathbf{q}}(T)$ increases. Refer to Fig. S15 for a graphical comparison. Both corrections are implemented in the workflow, but the free energy results shown in this paper use Eq. (13).

When the potential becomes increasingly anharmonic, however, both Eqs. (12) and (13) become increasingly insufficient accounts of anharmonic free energy. Eq. (12) can only trace out real frequency shifts, as mentioned, and Eq. (13) is based on the lowest-order perturbation theory. The correction due to 3rd-order interaction ("bubble") would serve as the next-order correction to the latter, but this is more complicated to compute[76,78] and still incomplete for very strongly anharmonic materials[69].

Data availability

The raw data for thermal conductivity and renormalization calculations are available in Zenodo (https://doi.org/10.5281/zenodo.10810590).

Code availability

The code for our lattice dynamics workflow is available at https://github.com/hackingmaterials/atomate. Related in-house modifications for renormalization and Wigner contribution for thermal conductivity are accessible at https://gitlab.com/jsyony37/hiphive and https://github.com/FourPhonon/FourPhonon, respectively.

Received: 14 March 2024; Accepted: 14 October 2024;
Published online: 14 November 2024

References

1. Sootsman, J. R., Chung, D. Y. & Kanatzidis, M. G. New and old concepts in thermoelectric materials. *Angew. Chem. Int. Ed.* **48**, 8616–8639 (2009).
2. Martin, L. W. & Rappe, A. M. Thin-film ferroelectric materials and their applications. *Nat. Rev. Mater.* **2**, 1–14 (2016).
3. Huang, W. On the selection of shape memory alloys for actuators. *Mater. Des.* **23**, 11–19 (2002).
4. Kadkhodaei, S. & van de Walle, A. First-principles calculations of thermal properties of the mechanically unstable phases of the PtTi and NiTi shape memory alloys. *Acta Mater.* **147**, 296–303 (2018).
5. Dove, M. T.*Introduction to Lattice Dynamics* (Cambridge University Press, 1993).
6. Ackland, G. J., Warren, M. C. & Clark, S. J. Practical methods in ab initio lattice dynamics. *J. Phys.: Condens. Matter* **9**, 7861 (1997).
7. Jain, A. et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 1002 (2013).
8. Curtarolo, S. et al. Aflowlib.org: A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58**, 227–235 (2012).
9. Kirklin, S. et al. The open quantum materials database (oqmd): assessing the accuracy of dft formation energies. *Npj Comput. Mater.* **1**, 5010 (2015).
10. Ceperley, D. M. Path integrals in the theory of condensed helium. *Rev. Mod. Phys.* **67**, 279–355 (1995).
11. Tolborg, K., Klarbring, J., Ganose, A. M. & Walsh, A. Free energy predictions for crystal stability and synthesisability. *Digit. Discov.* **1**, 586–595 (2022).
12. Zong, H., Pilania, G., Ding, X., Ackland, G. J. & Lookman, T. Developing an interatomic potential for martensitic phase transformations in zirconium by machine learning. *npj Comput. Mater.* **4**, 48 (2018).
13. Xie, Y. et al. Uncertainty-aware molecular dynamics from Bayesian active learning for phase transformations and thermal transport in SiC. *npj Comput. Mater.* **9**, 1–8 (2023).
14. Verdi, C., Karsai, F., Liu, P., Jinnouchi, R. & Kresse, G. Thermal transport and phase transitions of zirconia by on-the-fly machine-learned interatomic potentials. *npj Comput. Mater.* **7**, 156 (2021).

15. Baroni, S., de Gironcoli, S. & Corso, A. D. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.* **73**, 515–562 (2001).
16. Togo, A. & Tanakal, I. First principles phonon calculations in materials science. *Scr. Mater.* **108**, 1–5 (2015).
17. Paulatto, L., Mauri, F. & Lazzeri, M. Anharmonic properties from a generalized third-order ab initio approach: Theory and applications to graphite and graphene. *Phys. Rev. B* **87**, 214303 (2013).
18. Feng, T. & Ruan, X. Quantum mechanical prediction of four-phonon scattering rates and reduced thermal conductivity of solids. *Phys. Rev. B* **93**, 045202 (2016).
19. Plata, J. J. et al. An efficient and accurate framework for calculating lattice thermal conductivity of solids: Aflow–aapl automatic anharmonic phonon library. *npj Comput. Mater.* **3**, 45 (2017).
20. Togo, A. & Seko, A. On-the-fly training of polynomial machine learning potentials in computing lattice thermal conductivity. article arXiv:2401.17531, arXiv (2024). http://arxiv.org/abs/2401.17531.
21. Zhou, F., Nielson, W., Xia, Y. & Ozoliņš, V. Lattice anharmonicity and thermal conductivity from compressive sensing of first-principles calculations. *Phys. Rev. Lett.* **113**, 185501 (2014).
22. Tadano, T., Gohda, Y. & Tsuneyuki, S. Anharmonic force constants extracted from first-principles molecular dynamics: applications to heat transfer simulations. *J. Phys. Condens. Matter* **26**, 225402 (2014).
23. Hellman, O. & Abrikosov, I. A. Temperature-dependent effective third-order interatomic force constants from first principles. *Phys. Rev. B* **88**, 144301 (2013).
24. Knoop, F. et al. TDEP: Temperature dependent effective potentials. *J. Open Source Softw.* **9**, 6150 (2024).
25. Eriksson, F., Fransson, E. & Erhart, P. The hiphive package for the extraction of high?Order force constants by machine learning. *Adv. Theory Simul.* **2**, 1800184 (2019).
26. Ganose, A. M. et al. Efficient calculation of carrier scattering rates from first principles. *Nat. Commun.* **12**, 2222 (2021).
27. Mathew, K. et al. Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows. *Comput. Mater. Sci.* **139**, 140–152 (2017).
28. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993).
29. Kresse, G. & Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal amorphous-semiconductor transition in germanium. *Phys. Rev. B* **49**, 14251–14269 (1994).
30. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
31. Togo, A., Chaput, L. & Tanaka, I. Distributions of phonon lifetimes in brillouin zones. *Phys. Rev. B* **91**, 094306 (2015).
32. Li, W., Carrete, J., Katcho, N. A. & Mingo, N. Shengbte: A solver of the Boltzmann transport equation for phonons. *Comput. Phys. Commun.* **185**, 1747–1758 (2014).
33. Han, Z., Yang, X., Li, W., Feng, T. & Ruan, X. Fourphonon: An extension module to shengbte for computing four-phonon scattering rates and thermal conductivity. *Comput. Phys. Commun.* **270**, 108179 (2022).
34. Togo, A., Chaput, L., Tadano, T. & Tanaka, I. Implementation strategies in phonopy and phono3py. *J. Phys. Condens. Matter* **35**, 353001 (2023).
35. Jain, A. et al. Fireworks: a dynamic workflow system designed for high-throughput applications. *Concurr. Comp. - Pract. E.* **27**, 5037–5059 (2015).
36. Ong, S. P. et al. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).
37. Larsen, A. H. et al. The atomic simulation environment: a python library for working with atoms. *J. Phys. Condens. Matter* **29**, 273002 (2017).
38. Perdew, J. P. et al. Restoring the density-gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.* **100**, 136406 (2008).
39. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
40. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
41. Ning, J., Furness, J. W. & Sun, J. Reliable lattice dynamics from an efficient density functional approximation. *Chem. Mater.* **34**, 2562–2568 (2022).
42. Kingsbury, R. et al. Performance comparison of r 2 SCAN and SCAN metaGGA density functionals for solid materials via an automated, high-throughput computational workflow. *Phys. Rev. Mater.* **6**, 013801 (2022).
43. Schnorr, C. P. Progress on LLL and Lattice Reduction. In Nguyen, P. Q. & Vallée, B. (eds.) *The LLL Algorithm: Survey and Applications*, Information Security and Cryptography, 145–178 (Springer, Berlin, Heidelberg, 2010). https://doi.org/10.1007/978-3-642-02295-1_4.
44. Born, M., Huang, K. & Lax, M. Dynamical theory of crystal lattices. *Am. J. Phys.* **23**, 474–474 (1955).
45. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
46. Eriksson, F., Fransson, E. & Erhart, P. Efficient construction of linear models in materials modeling and applications to force constant expansions. *npj Comput. Mater.* **6**, 135 (2020).
47. Xia, Y. et al. First-principles calculations of lattice thermal conductivity in tl_3vse_4: Uncertainties from different approaches of force constants. *Phys. Rev. B* **108**, 184307 (2023).
48. Souvatzis, P. & Rudin, S. Dynamical stabilization of cubic ZrO_2 by phonon-phonon interactions: Ab initio calculations. *Phys. Rev. B–Condens. Matter Mater. Phys.* **78**, 184304 (2008).
49. Xia, Y. & Chan, M. K. Anharmonic stabilization and lattice heat transport in rocksalt β-gete. *Appl. Phys. Lett.* **113**, 193902 (2018).
50. Kadkhodaei, S. & Davariashtiyani, A. Phonon-assisted diffusion in bcc phase of titanium and zirconium from first principles. *Phys. Rev. Mater.* **4**, 043802 (2020).
51. Wang, Y., Shang, S.-L., Fang, H., Liu, Z.-K. & Chen, L.-Q. First-principles calculations of lattice dynamics and thermal properties of polar solids. *npj Comput. Mater.* **2**, 1–10 (2016). Number: 1 Publisher: Nature Publishing Group.
52. Togo, A. MDR phonon calculation database (2023). https://mdr.nims.go.jp/collections/8g84ms862.
53. Gonze, X. & Lee, C. Dynamical matrices, Born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory. *Phys. Rev. B* **55**, 10355–10368 (1997).
54. Zhou, F., Sadigh, B., Aberg, D., Xia, Y. & Ozoliņš, V. Compressive sensing lattice dynamics. ii. efficient phonon calculations and long-range interactions. *Phys. Rev. B* **100**, 184309 (2019).
55. Roy, R., Agrawal, D. K. & McKinstry, H. A. Very low thermal expansion coefficient materials. *Annu. Rev. Mater. Sci.* **19**, 59–81 (1989).
56. James, J. D., Spittle, J. A., Brown, S. G. R. & Evans, R. W. A review of measurement techniques for the thermal expansion coefficient of metals and alloys at elevated temperatures. *Meas. Sci. Technol.* **12**, R1 (2001).
57. Jin, H. M. & Wu, P. First principles calculation of thermal expansion coefficient: Part 1. Cubic metals. *J. J. Alloy. Compd.* **343**, 71–76 (2002).
58. Drebushchak, V. A. Thermal expansion of solids: review on theories. *J. Therm. Anal. Calorim.* **142**, 1097–1113 (2020).
59. Pike, N. A. & Løvvik, O. M. Calculation of the anisotropic coefficients of thermal expansion: A first-principles approach. *Comput. Mater. Sci.* **167**, 257–263 (2019).
60. de Jong, M. et al. Charting the complete elastic properties of inorganic crystalline compounds. *Sci. Data* **2**, 150009 (2015).

61. Zhou, F., Nielson, W., Xia, Y. & Ozoliņš, V. Compressive sensing lattice dynamics. i. general formalism. *Phys. Rev. B* **100**, 184308 (2019).
62. Xia, Y. et al. High-throughput study of lattice thermal conductivity in binary rocksalt and zinc blende compounds including higher-order anharmonicity. *Phys. Rev. X* **10**, 041029 (2020).
63. Xia, Y. Revisiting lattice thermal transport in pbte: The crucial role of quartic anharmonicity. *Appl. Phys. Lett.* **113**, 073901 (2018).
64. Kisi, E. H. & Howard, C. Crystal structures of zirconia phases and their inter-relation. *Key Eng. Mater.* **153**, 1–36 (1998).
65. Aldebert, P. & Traverse, J.-P. Structure and ionic mobility of zirconia at high temperature. *J. Am. Ceram. Soc.* **68**, 34–40 (1985).
66. Chattopadhyay, T., Boucherle, J. X. & vonSchnering, H. G. Neutron diffraction study on the structural phase transition in GeTe. *J. Phys. C: Solid State Phys.* **20**, 1431 (1987).
67. Stassis, C., Zarestky, J. & Wakabayashi, N. Lattice dynamics of bcc zirconium. *Phys. Rev. Lett.* **41**, 1726–1729 (1978).
68. Blöchl, P. E., Jepsen, O. & Anderson, O. Improved tetrahedron method for brillouin-zone integrations. *Phys. Rev. B* **49**, 3616–3621 (1994).
69. Tolborg, K. & Walsh, A. Exploring the high-temperature stabilization of cubic Zirconia from anharmonic lattice dynamics. *Cryst. Growth Des.* **23**, 3314–3319 (2023). Publisher: American Chemical Society.
70. Kimber, S. A. J. et al. Dynamic crystallography reveals spontaneous anisotropy in cubic GeTe. *Nat. Mater.* **22**, 311–315 (2023).
71. Rosen, A. S. et al. Jobflow: Computational workflows made simple. *J. Open Source Softw.* **9**, 5995 (2024).
72. Batatia, I., Kovacs, D. P., Simm, G. N. C., Ortner, C. & Csanyi, G. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In Oh, A. H., Agarwal, A., Belgrave, D. & Cho, K. (eds.) *Advances in Neural Information Processing Systems* (2022). https://openreview.net/forum?id=YPpSngE-ZU.
73. Chen, C. & Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput Sci.* **2**, 718–728 (2022).
74. Deng, B. et al. Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat. Mach. Intell.* **5**, 1031–1041 (2023).
75. Gonze, X. et al. The abinit project: Impact, environment and recent developments. *Comput. Phys. Commun.* **248**, 107042 (2020).
76. Goldman, V. V., Horton, G. K. & Klein, M. L. An improved self-consistent phonon approximation. *Phys. Rev. Lett.* **21**, 1527–1529 (1968).
77. Tadano, T. & Tsuneyuki, S. Self-consistent phonon calculations of lattice dynamical properties in cubic srtio$_3$ with first-principles anharmonic force constants. *Phys. Rev. B* **92**, 054301 (2015).
78. Oba, Y., Tadano, T., Akashi, R. & Tsuneyuki, S. First-principles study of phonon anharmonicity and negative thermal expansion in scf3. *Phys. Rev. Mater.* **3**, 033601 (2019).
79. Tadano, T. & Saidi, W. A. First-principles phonon quasiparticle theory applied to a strongly anharmonic halide perovskite. *Phys. Rev. Lett.* **129**, 185901 (2022).
80. Park, J., Wu, Z. & Lawson, J. W. Thermodynamic integration for dynamically unstable systems using interatomic force constants without molecular dynamics. arXiv arXiv:2403.08644 (2024).
81. Ab initio structural optimization at finite temperatures based on anharmonic phonon theory: Application to the structural phase transitions of batio3, volume = 106, shorttitle = Ab initio structural optimization at finite temperatures based on anharmonic phonon theory, https://link.aps.org/doi/10.1103/PhysRevB.106.224104
82. Ehsan, S., Arrigoni, M., Madsen, G. K. H., Blaha, P. & Tröster, A. First-principles self-consistent phonon approach to the study of the vibrational properties and structural phase transition of batio3. *Phys. Rev. B* **103**, 094108 (2021).
83. Tadano, T. & Tsuneyuki, S. Ab initio prediction of structural phase-transition temperature of srtio3 from finite-temperature phonon calculation. *J. Ceram. Soc. Jpn.* **127**, 404–408 (2019).
84. Gu, H.-Y., Yin, W.-J. & Gong, X.-G. Significant phonon anharmonicity drives phase transitions in CsPbI3. *Appl. Phys. Lett.* **119**, 191101 (2021).
85. Pandit, A. & Hamad, B. The effect of finite-temperature and anharmonic lattice dynamics on the thermal conductivity of ZrS2 monolayer: self-consistent phonon calculations. *J. Phys.: Condens. Matter* **33**, 425405 (2021).
86. Gupta, M. K., Singh, B., Mittal, R. & Chaplot, S. L. Negative thermal expansion behavior in mzrf6(M=Ca,Mg,Sr): Ab initio lattice dynamical studies. *Phys. Rev. B* **98**, 014301 (2018).
87. Gupta, M. K., Mittal, R. & Chaplot, S. L. Negative thermal expansion in cubic zrw2o8: Role of phonons in the entire Brillouin zone from ab initio calculations. *Phys. Rev. B* **88**, 014303 (2013).
88. Härkönen, V. J. & Karttunen, A. J. Ab initio lattice dynamical studies of silicon clathrate frameworks and their negative thermal expansion. *Phys. Rev. B* **89**, 024305 (2014).
89. Chang, D. et al. Negative thermal expansion properties and the role of guest alkali atoms in LnFe(CN)6 (Ln = Y, La) from ab initio calculations. *J. Phys. Chem. C.* **122**, 12421–12427 (2018).
90. Chen, W. et al. Understanding thermoelectric properties from high-throughput calculations: trends, insights, and comparisons with experiment. *J. Mater. Chem. C.* **4**, 4414–4426 (2016).
91. Zhu, H. et al. Computational and experimental investigation of tmagte2 and xyz2 compounds, a new group of thermoelectric materials identified by first-principles high-throughput screening. *J. Mater. Chem. C.* **3**, 10554–10565 (2015).
92. Sarikurt, S., Kocabaş, T. & Sevik, C. High-throughput computational screening of 2D materials for thermoelectrics. *J. Mater. Chem. A* **8**, 19674–19683 (2020).
93. Jia, T., Feng, Z., Guo, S., Zhang, X. & Zhang, Y. Screening promising thermoelectric materials in binary chalcogenides through high-throughput computations. *ACS Appl. Mater. Interfaces* **12**, 11852–11864 (2020).
94. He, J. et al. Accelerated discovery and design of ultralow lattice thermal conductivity materials using chemical bonding principles. *Adv. Funct. Mater.* **32**, 2108532 (2022).
95. Simoncelli, M., Marzari, N. & Mauri, F. Unified theory of thermal transport in crystals and glasses. *Nat. Phys.* **15**, 809–813 (2019).
96. Xia, Y., Ozoliņš, V. & Wolverton, C. Microscopic mechanisms of glasslike lattice thermal transport in cubic $cu_{12}sb_4s_{13}$ tetrahedrites. *Phys. Rev. Lett.* **125**, 085901 (2020).
97. Souvatzis, P., Eriksson, O., Katsnelson, M. I. & Rudin, S. P. Entropy driven stabilization of energetically unstable crystal structures explained from first principles theory. *Phys. Rev. Lett.* **100**, 095901 (2008).
98. Souvatzis, P., Eriksson, O., Katsnelson, M. I. & Rudin, S. P. The self-consistent ab initio lattice dynamical method. *Comput. Mater. Sci.* **44**, 888–894 (2009).
99. van Roekeghem, A., Carrete, J. & Mingo, N. Anomalous thermal conductivity and suppression of negative thermal expansion in scf_3. *Phys. Rev. B* **94**, 020303(R) (2016).
100. van Roekeghem, A., Carrete, J. & Mingo, N. Quantum self-consistent ab-initio lattice dynamics. *Comput. Phys. Commun.* **44**, 107945 (2021).
101. Errea, I., Calandra, M. & Mauri, F. Anharmonic free energies and phonon dispersions from the stochastic self-consistent harmonic approximation: Application to platinum and palladium hydrides. *Phys. Rev. B* **89**, 064302 (2014).
102. Hellman, O., Abrikosov, I. A. & Simak, S. I. Lattice dynamics of anharmonic solids from first principles. *Phys. Rev. B* **84**, 180301(R) (2011).
103. Hellman, O., Steneteg, P., Abrikosov, I. A. & Simak, S. I. Temperature dependent effective potential method for accurate free energy calculations of solids. *Phys. Rev. B* **87**, 104111 (2013).

104. Xia, Y. & Chan, M. Anharmonic stabilization and lattice heat transport in rocksalt β-gete. *Appl. Phys. Lett.* **113**, 193902 (2018).
105. Cowley, R. A. The lattice dynamics of an anharmonic crystal. *Adv. Phys.* **12**, 421–480 (1963).
106. Allen, P. B. Anharmonic phonon quasiparticle theory of zero-point and thermal shifts in insulators: Heat capacity, bulk modulus, and thermal expansion. *Phys. Rev. B* **92**, 064106 (2015).

Acknowledgements

This work was primarily supported by the Materials Project, funded by the U.S. Department of Energy under award DE-AC02-05CH11231 (Materials Project program KC23MP). J. P. acknowledges the support from the U.S. Department of Energy, Office of Basic Energy Sciences, Early Career Research Program. J. W. L. and J. P. also acknowledge funding by the Transformational Tools and Technologies (TTT) project of the Aeronautics Research Mission Directorate (ARMD) at the National Aeronautics and Space Administration (NASA). A. M. G. was supported by EPSRC Fellowship EP/T033231/1. This work used computational resources of the National Energy Research Scientific Computing Center (NERSC), a Department of Energy Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Author contributions

J.P. and A.J. conceived the ideas and designed the necessary steps of the workflow. Z.Z. performed the calculations, benchmarked the parameters used in the workflow by evaluating the output's accuracy, and finalized the workflow with optimized settings. J.P. developed the workflow in atomate by incorporating the finite-temperature renormalization process and the free energy calculations. A.M.G. and R.C. contributed to the initial codes for the workflow in atomate. H.S. adopted the workflow to atomate2 and tested the fitting methods. J.W.L. contributed to free energy calculations. Z.Z. and J.P. drafted the manuscript. A.J. supervised the research. All authors contributed to the discussions and provided feedback on the manuscript.

Competing interests

Dr. Alex M. Ganose is an Editorial Board Member for npj Computational Materials and was not involved in the editorial review or the decision to publish this article.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-024-01437-w.

Correspondence and requests for materials should be addressed to Anubhav Jain.