

UCLA

UCLA Previously Published Works

Title

Lattice Anharmonicity and Thermal Conductivity from Compressive Sensing of First-Principles Calculations

Permalink

<https://escholarship.org/uc/item/4s97m6km>

Journal

Physical Review Letters, 113(18)

ISSN

0031-9007

Authors

Zhou, Fei

□□

Nielson, Weston

et al.

Publication Date

2014-10-31

DOI

10.1103/physrevlett.113.185501

Peer reviewed

Lattice anharmonicity and thermal conductivity from compressive sensing of first-principles calculations

Fei Zhou(周非)

Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California 94550, USA

Weston Nielson, Yi Xia, and Vidvuds Ozoliņš

*Department of Materials Science and Engineering,
University of California, Los Angeles, California 90095-1595, USA*

(Dated: September 23, 2014)

First-principles prediction of lattice thermal conductivity κ_L of strongly anharmonic crystals is a long-standing challenge in solid state physics. Making use of recent advances in information science, we propose a systematic and rigorous approach to this problem, compressive sensing lattice dynamics (CSLD). Compressive sensing is used to select the physically important terms in the lattice dynamics model and determine their values in one shot. Non-intuitively, high accuracy is achieved when the model is trained on first-principles forces in *quasi-random* atomic configurations. The method is demonstrated for Si, NaCl, and $\text{Cu}_{12}\text{Sb}_4\text{S}_{13}$, an earth-abundant thermoelectric with strong phonon-phonon interactions that limit the room-temperature κ_L to values near the amorphous limit.

PACS numbers: 63.20.Ry, 63.20.dk, 66.70.-f

To a large extent, thermal properties of crystalline solids are determined by the vibrations of their constituent atoms. Hence, an accurate description of lattice dynamics is essential for fundamental understanding of the structure, thermodynamics, phase stability and thermal transport properties of solids. The seminal work of Born and Huang [1] forms the theoretical basis of our understanding of harmonic vibrations and their relation to elastic properties. With the advent of efficient density-functional theory (DFT) based methods for solving the Schrödinger’s equation, several *ab initio* methods for studying *harmonic* phonon properties of solids have been proposed, such as the frozen phonon approach [2, 3], supercell small displacement method [4, 5] and the density-functional perturbation theory (DFPT) [6]. Due to these developments, *ab initio* calculations of the harmonic phonon dispersion curves and phonon mode Grüneisen parameters have become routine.

A systematic approach to *anharmonicity* has been more difficult to develop. Anharmonic effects are key to explaining phenomena where phonon-phonon collisions become important, such as in the study of lattice thermal conductivity, κ_L , a key quantity for optimizing the performance of electronic materials, thermal coatings, and thermoelectrics [7]. For weakly anharmonic systems, the interaction processes involving three phonons are dominant, such as decay of a phonon into two lower-energy phonons or combining two phonons to create a higher-energy phonon. Their effects on phonon frequencies and lifetimes can be evaluated using the first-order perturbation theory (PT) [8, 9], and κ_L can then be obtained by either using the relaxation time approximation or solving the Boltzmann transport equation [10]. The computational feasibility and physical accuracy of these methods are well established [11–14]. Unfortunately, PT

tends to be computationally expensive for solids with large, complex unit cells, and its ability to handle strong anharmonicity is insufficient, especially when the harmonic phonon dispersion contains imaginary frequencies or when phonon scattering becomes so intense that κ_L saturates at its theoretical minimum [15]. Many interesting and technologically relevant materials belong to this class, e.g. ferroelectrics and thermoelectrics with ultra-low κ_L [16]. In these cases, a general and efficient non-perturbative approach that can accurately describe four-phonon and higher-order interactions, is needed. The “ $2n + 1$ ” theorem of DFPT [17] can be used to calculate the 4th- and higher-order terms, but the computations are cumbersome and require specialized codes which, to the best of our knowledge, are not available for $n > 1$.

In this paper, we introduce an approach to building lattice dynamical models which can treat compounds with large, complex unit cells and strong anharmonicity, including those with harmonically unstable phonon modes. Our approach, compressive sensing lattice dynamics (CSLD), determines anharmonic force constants from standard DFT total energy calculations. We utilize compressive sensing (CS), a technique recently developed in the field of information science for recovering sparse solutions from incomplete data [18], to determine which anharmonic terms are important and find their values simultaneously. A non-intuitive prescription based on CS to generate DFT training data is given. We show that CSLD is efficient, general and robust through a few prototypical case studies.

The starting point is a Taylor expansion of the total energy in powers of atomic displacements,

$$V = V_0 + \Phi_{\mathbf{a}}u_{\mathbf{a}} + \frac{\Phi_{\mathbf{ab}}}{2}u_{\mathbf{a}}u_{\mathbf{b}} + \frac{\Phi_{\mathbf{abc}}}{3!}u_{\mathbf{a}}u_{\mathbf{b}}u_{\mathbf{c}} + \dots \quad (1)$$

where $u_{\mathbf{a}} \equiv u_{a,i}$ is the displacement of atom a at a lat-

tice site \mathbf{R}_a in the Cartesian direction i , the 2nd-order expansion coefficients $\Phi_{\mathbf{ab}} \equiv \Phi_{ij}(ab) = \partial^2 V / \partial u_a \partial u_b$ determine the phonon dispersion in the harmonic approximation, and $\Phi_{\mathbf{abc}} \equiv \Phi_{ijk}(abc) = \partial^3 V / \partial u_a \partial u_b \partial u_c$, etc., are third- and higher-order anharmonic force constant tensors (FCTs). The linear term with $\Phi_{\mathbf{a}}$ is absent if the reference lattice sites represent mechanical equilibrium, and the Einstein summation convention over repeated indices is used throughout the paper.

Systematic fitting or direct calculation of the higher-order anharmonic terms in Eq. (1) is challenging due to combinatorial explosion in the number of tensors $\Phi(a_1 \cdots a_n)$ with increasing order n and maximum distance between the sites $\{a_1, \dots, a_n\}$. Since it is not *a priori* obvious where to truncate this expansion, one needs to rely on physical intuition, which can only be gained on a case-by-case basis through time-consuming cycles of model construction and cross-validation. As a result, anharmonic FCTs have been calculated only for relatively simple crystals and weak anharmonicity [14, 19, 20].

We have recently shown that a similar problem in alloy theory, the cluster expansion (CE) method for configurational energetics [21, 22], can be solved efficiently and accurately using compressive sensing [23, 24]. CS has revolutionized information science by providing a mathematically rigorous recipe for reconstructing S -sparse models (i.e., models with S nonzero coefficients out of a large pool of possibles, N , when $S \ll N$) from a set of only $O(S)$ data points [25–27]. Given training data, CS automatically picks out the relevant expansion coefficients and determines their values *in one shot* by applying a mathematical technique which in essence is the Occam’s razor for model choice. To see how this applies to lattice dynamics, we write down the force-displacement relationship for Eq. (1):

$$F_{\mathbf{a}} = -\Phi_{\mathbf{a}} - \Phi_{\mathbf{ab}} u_{\mathbf{b}} - \Phi_{\mathbf{abc}} u_{\mathbf{b}} u_{\mathbf{c}} / 2 - \dots \quad (2)$$

The forces can be obtained from first-principles calculations using any general-purpose DFT code for a set of L atomic configurations in a supercell. This establishes a linear problem $\mathbf{F} = \mathbb{A}\Phi$ for the unknown FCTs, where

$$\mathbb{A} = \begin{bmatrix} -1 & -u_{\mathbf{b}}^1 & -\frac{1}{2}u_{\mathbf{b}}^1 u_{\mathbf{c}}^1 & \dots \\ \dots & \dots & \dots & \dots \\ -1 & -u_{\mathbf{b}}^L & -\frac{1}{2}u_{\mathbf{b}}^L u_{\mathbf{c}}^L & \dots \end{bmatrix} \quad (3)$$

will be referred to as the sensing matrix. Its elements are products of atomic displacements corresponding to distinct terms in the force expansion Eq. (2); different training configurations are labeled by superscript $u_{\mathbf{b}}^i$. Each row corresponds to a calculated force component on one of the atoms, and the total number of rows is $M = 3LN_{\text{at}}$, where N_{at} is the number of atoms in the supercell. Columns corresponds to N FCT components, which are arranged in a vector Φ . In practice, N can far exceed M , which makes the linear problem Eq. (3) underdetermined. A reasonable approach would be to

choose Φ so that it reproduces the training data \mathbf{F} to a given accuracy with the smallest number of nonzero FCT components, i.e., by minimizing the so-called ℓ_0 norm of the solution. Unfortunately, this is an intractable (“NP-hard”) discrete optimization problem.

CS solves the underdetermined linear problem in Eq. (3) by minimizing the ℓ_1 norm of the coefficients, $\|\Phi\|_1 \equiv \sum_I |\Phi_I|$, while requiring a certain level of accuracy for reproducing the data. The ℓ_1 norm serves as an approximation to the ℓ_0 norm and results in a computationally tractable convex optimization problem. Mathematically, the solution is found as

$$\begin{aligned} \Phi^{\text{CS}} &= \arg \min_{\Phi} \|\Phi\|_1 + \frac{\mu}{2} \|\mathbf{F} - \mathbb{A}\Phi\|_2^2 \\ &= \arg \min_{\Phi} \sum_I |\Phi_I| + \frac{\mu}{2} \sum_{ai} (F_{ai} - A_{ai,J}\Phi_J)^2 \end{aligned} \quad (4)$$

where the second term is the usual sum-of-squares Euclidian ℓ_2 norm of the fitting error for the training data (in this case, DFT forces). The ℓ_1 term drives the model towards solutions with a small number of nonzero elements, and the parameter μ is used to adjust the relative weights of the ℓ_1 and ℓ_2 terms (see below). CSLD has several advantages over other methods for building models of lattice dynamics: it does not require prior physical intuition to pick out potentially relevant FCTs, the fitting procedure is very robust with respect to both random and systematic noise [26], and it gives an efficient prescription for generating training data.

A full account of the technical details of our approach will be given in a separate publication, and here we only describe the key features. Higher values of μ will produce a least-squares like fitting at the expense of denser FCTs that are prone to over-fitting, while small μ will produce very sparse under-fitted FCTs, degrading the quality of the fit. The optimal μ value that produces a model with the highest predictive accuracy lies in-between and can be determined by monitoring the predictive error for a leave-out subset of the training data not used in fitting [23]. The predictive accuracy of the resulting model is then validated on a third, distinct set of DFT data, which we refer to as the “prediction set”. Space group symmetry and translational invariance conditions are used to reduce the number of independent FCT elements [8]; the latter are also important for momentum conservation according to the Noether’s theorem. These linear constraints are applied algebraically by constructing a null-space matrix. For polar insulators, the long-range Coulomb interactions can be treated separately by calculating the Born effective charges and dielectric tensors. The long-range contributions are then subtracted from \mathbf{F} , ensuring that the remaining FCTs are short-ranged [28].

A key ingredient of CSLD is the choice of atomic configurations for the training and prediction sets. It is intuitively appealing to use snapshots from *ab initio* molecular dynamics (AIMD) trajectories since they represent

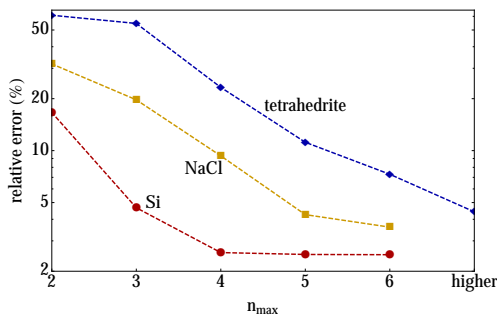


FIG. 1. Relative CSLD prediction error of force components from AIMD snapshots at 300 K not involved in fitting. For each system, a fitting of up to 6th order was performed, and errors using FCTs up to n_{\max} ($= 2, \dots$) are shown. Higher order expansions are used in tetrahedrite [30].

physically relevant low-energy configurations. However, these configurations give rise to strong cross-correlations between the columns of \mathbb{A} (i.e., high mutual coherence of the sensing matrix [29]), which decreases the efficiency of CS due to the difficulty of separating correlated contributions to \mathbf{F} from different FCTs. One of the most profound results of CS is that a near-optimal signal recovery can be realized by using sensing matrices \mathbb{A} with *random* entries that are independent and identically distributed (i.i.d.) [18]. In this case, the contributions from different FCTs are uncorrelated and can be efficiently separated using Eq. (4). For the discrete orthogonal basis in the CS cluster expansion [23, 24], i.i.d. sensing matrices \mathbb{A} could be obtained by enumerating all ordered structures up to a certain size and choosing those with correlations that map most closely onto quasi-random vectors. This strategy is difficult to adapt for CSLD since the Taylor expansion employs non-orthogonal and unnormalized basis functions of a continuous variable, u^n . To solve this conundrum, we combine the physical relevance of MD trajectories with the mathematical efficiency of CS by adding random displacements (~ 0.1 Å) to each atom in well-spaced MD snapshots. To guarantee that all the terms in the ℓ_1 norm have the same unit of force, Eq. (2) is scaled by $\Phi \rightarrow \Phi u_0^{n-1}$ and $u \rightarrow u/u_0$, where n is the order of the FCT and u_0 is a “maximum” displacement chosen to be on the order of the amplitude of thermal vibrations. This procedure was found to significantly decrease cross-correlations between the columns of \mathbb{A} and resulted in stable fits.

We begin by demonstrating the accuracy of our approach for two relatively simple cases, Si and NaCl. DFT calculations were performed using the Perdew-Becke-Ernzerhof (PBE) functional [31] and projector-augmented wave (PAW) potentials [32] as implemented in the VASP code [33]. An overview of the predictive accuracy of CSLD is shown in Fig. 1. We included up to the 6-th order FCTs for a total of 712 (Si) and 1375 (NaCl) symmetrically distinct elements. CS using Eq. (4) found

258 and 199 non-zero FCT elements, respectively. The errors decrease when higher order FCT parameters are considered. Anharmonic terms account for an increasing amount of the improved accuracy in Si, NaCl and tetrahedrite (to be discussed later), reflecting increasing anharmonicity. Phonon dispersion curves (Supplemental Material) using the CSLD pair force constants are in excellent agreement with experiment, validating our method on the harmonic level. We then used first-order PT [8, 9] to calculate phonon lifetimes of Si (Supplemental Material), which are in excellent agreement with other first-principles PT based studies [14, 20]. Lattice thermal conductivity κ_L of Si (Fig. 2a) was obtained with the ShengBTE code [34] and found to be in good agreement with experimental data [35], validating the numerical accuracy of our third-order FCTs.

To test the performance of CSLD in calculating κ_L of strongly anharmonic solids, a custom lattice molecular dynamics (LMD) program was developed with Eq. (1) as the potential. Multiple methods were implemented for calculating κ_L , including the Green-Kubo linear response formula [36, 37], reverse non-equilibrium MD (RNEMD) [38] and homogenous non-equilibrium MD (HNEMD) proposed by Evans [39]. While all methods yielded similar results, we found after extensive testing that HNEMD was the most efficient. In HNEMD, the equations of motion are modified so that the force on atom a is given by

$$\mathbf{F}_a = F_a - \sum_b \mathbf{F}_{ab} (\mathbf{r}_{ab} \cdot \mathbf{F}_e) + \frac{1}{N} \sum_{b,c} \mathbf{F}_{bc} (\mathbf{r}_{bc} \cdot \mathbf{F}_e), \quad (5)$$

where F_a is the unmodified force calculated from Eq. (2) and \mathbf{F}_{ab} is the force on atom a due to b [40]. The external field \mathbf{F}_e has the effect of driving higher energy (hotter) particles with the field and lower energy (colder) particles against the field, while a Gaussian thermostat is used to remove the heat generated by \mathbf{F}_e . Using linear response, the average heat flux is given by

$$\langle \mathbf{J}(t) \rangle = -\beta V \int_0^t ds \langle \mathbf{J}(t-s) \otimes \mathbf{J}(0) \rangle \cdot \mathbf{F}_e. \quad (6)$$

As $\mathbf{F}_e \rightarrow 0$ and $t \rightarrow \infty$, one recovers the Green-Kubo formula [36, 37]. For cubic systems the external field can be set to $\mathbf{F}_e = (0, 0, F_z)$, and we get

$$\kappa_L = \frac{V}{k_B T^2} \int_0^\infty dt \langle J_z(t) J_z(0) \rangle = \lim_{F_z \rightarrow 0} \frac{-\langle J_z(\infty) \rangle}{T F_z}. \quad (7)$$

The process then involves a series of simulations at varying external fields \mathbf{F}_e and constant T , with a simple linear extrapolation to zero field resulting in the true κ_L .

Simulations were performed for NaCl between 100 and 300 K, with system sizes ranging from 512 to 4096 atoms.

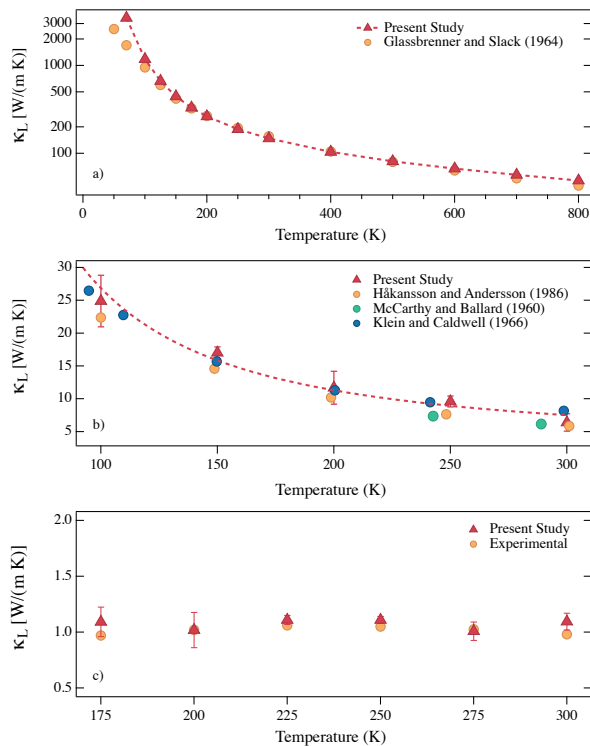


FIG. 2. Temperature dependence of calculated κ_L (triangles) for (a) Si from Boltzman transport equations using 3rd order FCTs; (b) NaCl and (c) $\text{Cu}_{12}\text{Sb}_4\text{S}_{13}$ by MD simulation using up to 6th-order FCTs. The dashed line is an inverse power fit of the calculation, and circles are experimental data from Ref. 35, Refs. 41–43 and Ref. 44, respectively.

The lengths of the simulations ranged from 100 ps to 1 ns and all used a timestep of 1 fs. At least four different values for F_z were taken at a given T . The results obtained are shown for NaCl in Fig. 2(b). Very good agreement is seen between the calculated and experimental values across the entire temperature range tested.

Finally, CSLD is applied to study anharmonic phonon dynamics and κ_L in $\text{Cu}_{12}\text{Sb}_4\text{S}_{13}$, a parent compound for the earth-abundant natural mineral tetrahedrite, which was recently shown to be a high-performance thermoelectric [44]. One of its key advantages is an exceptionally low κ_L , experimentally found to be $\lesssim 1$ W/(m K) in phase and compositionally pure samples [44]. Furthermore, our previous calculation found several harmonically unstable phonon modes, pointing to very strong anharmonicity [44]. $\text{Cu}_{12}\text{Sb}_4\text{S}_{13}$ has a body-centered cubic (space group $I43m$) structure with 29 atoms in the primitive cell, a large number that complicates the computation of FCTs using existing methods. For example, there are 188 distinct atomic pairs within a radius of $a = 10.4$ Å, 116 triplets within $a/2$, etc. Taking into account the 3^n elements of each tensor, the number of unknown coefficients is very large (55584 in our setting including up to 6th-order terms). After symmetrization, this is reduced to

$N = 3188$, which still represents a formidable numerical challenge.

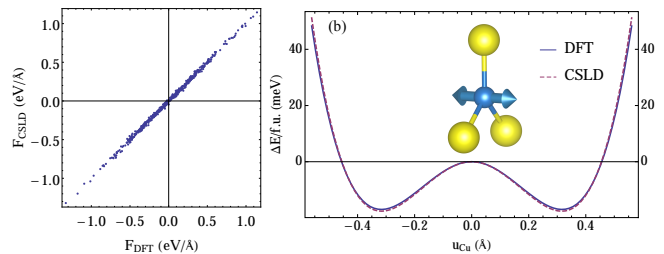


FIG. 3. Comparison of CSLD predictions with DFT data for tetrahedrite: (a) force at 300 K, (b) relative energy per formula unit of an unstable optical mode involving out-of-plane displacements of trigonally coordinated copper atoms (blue) bonded to sulfur (yellow sphere). DFT and CSLD are shown as solid and dashed lines, respectively.

High order interactions were added [30] to ensure highly accurate forces (see also Fig. 1). A large number of non-zero FCT elements (2101) were obtained by CSLD. Fig. 3a shows the overall accuracy of the model over a prediction set from *ab initio* MD snapshots at 300 K. The root-mean-square error of the predicted force components is 0.02 eV/Å, or 4%. The phonon dispersion calculated from the pair FCTs features unstable modes and is in good agreement with our previous DFPT calculations (see Supplemental material) [44], again validating CSLD at the harmonic level. Figure 3b shows the DFT potential energy surface (solid line) along an unstable Γ point mode involving displacement of trigonally coordinated Cu atoms (inset). The double-well behavior points to strong 4th-order anharmonicity. Our CSLD model (dashed line) is able to reproduce the potential energy to an absolute accuracy of 2 meV.

The HNEMD method was used to calculate κ_L of $\text{Cu}_{12}\text{Sb}_4\text{S}_{13}$, employing the same approach as described for NaCl above. All simulations were done with a supercell of 464 atoms and a minimum of 4 separate external fields at each temperature. The HNEMD results are compared with the experimental κ from Lu *et al.* [44] with electronic contributions subtracted in Fig. 2(c). Once again, very good agreement is seen across the entire temperature range tested. This example shows that CSLD extends the accuracy of DFT to treat lattice dynamics of compounds with large, complex unit cells and strong anharmonic effects previously beyond the reach of non-empirical studies.

In conclusion, CSLD is a powerful tool for highly anharmonic lattice dynamics in complex materials based on the robust and mathematically rigorous framework of compressive sensing and compressive sampling. The main advantage of CSLD over the current methods is that it is widely applicable, computationally efficient, systematically improvable and straight-forward to implement. Importantly, it works with general-purpose DFT codes

and can be used in an automated manner, with minimal human intervention. This technical development is a big step towards systematic, automated calculations of thermal transport properties for a wide variety of crystalline compounds, enabling computational design and discovery of new high-performance materials. Beyond lattice thermal conductivity, we expect CSLD to be useful in a wide range applications where strong anharmonicity plays a key role, such as ferroelectric phase transitions and temperature induced structural phase transformations, including martensitic transformations.

The authors gratefully acknowledge discussions with B. Sadigh and financial support for general method development from the National Science Foundation under Award No. DMR-1106024. CSLD and LMD code development and calculations related to $\text{Cu}_{12}\text{Sb}_4\text{S}_{13}$ were performed as part of the Center for Revolutionary Materials for Solid State Energy Conversion, an Energy Frontier Research Center funded by the US DOE, Office of Science, Basic Energy Sciences under Award No. DE-SC0001054. We are grateful to X. Lu and D. T. Morelli for providing low-temperature κ_L data for tetrahedrite. Part of the work by F.Z. was performed under the auspices of the US DOE by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. We used computing resources at NERSC, which is supported by the US DOE under Contract No. DE-AC02-05CH11231.

-
- [1] M. Born and K. Huang, *Dynamical Theory of Crystal Lattices*, International series of monographs on physics (Oxford University Press, Oxford, 1954).
- [2] H. Wendel and R. M. Martin, Phys. Rev. Lett. **40**, 950 (1978).
- [3] K. M. Ho, C. L. Fu, and B. N. Harmon, Phys. Rev. B **29**, 1575 (1984).
- [4] K. Kunc and R. M. Martin, Phys. Rev. Lett. **48**, 406 (1982).
- [5] K. Parlinski, Z.-Q. Li, and Y. Kawazoe, Phys. Rev. Lett. **78**, 4063 (1997).
- [6] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Gianozzi, Rev. Mod. Phys. **73**, 515 (2001).
- [7] M. Zebarjadi, K. Esfarjani, M. S. Dresselhaus, Z. F. Ren, and G. Chen, Energy Environ. Sci. **5**, 5147 (2012).
- [8] G. K. Horton and A. A. Maradudin, eds., *Dynamical Properties of Solids: Crystalline solids, fundamentals* (North-Holland, Amsterdam, 1974).
- [9] A. A. Maradudin, A. E. Fein, and G. H. Vineyard, Phys. Stat. Solidi B **2**, 1479 (1962).
- [10] M. Omini and A. Sparavigna, Phys. Rev. B **53**, 9064 (1996).
- [11] A. Debernardi, S. Baroni, and E. Molinari, Phys. Rev. Lett. **75**, 1819 (1995).
- [12] D. A. Broido, M. Malorny, G. Birner, N. Mingo, and D. A. Stewart, Appl. Phys. Lett. **91**, 231922 (2007).
- [13] J. Garg, N. Bonini, B. Kozinsky, and N. Marzari, Phys. Rev. Lett. **106**, 045901 (2011).
- [14] K. Esfarjani, G. Chen, and H. T. Stokes, Phys. Rev. B **84**, 085204 (2011).
- [15] D. G. Cahill, S. K. Watson, and R. O. Pohl, Phys. Rev. B **46**, 6131 (1992).
- [16] M. D. Nielsen, V. Ozolins, and J. P. Heremans, Energy Environ. Sci. **6**, 570 (2013).
- [17] X. Gonze and J. P. Vigneron, Phys. Rev. B **39**, 13120 (1989).
- [18] E. Candès and M. Wakin, IEEE Signal Proc. Mag. **25**, 21 (2008).
- [19] K. Esfarjani and H. T. Stokes, Phys. Rev. B **77**, 144112 (2008).
- [20] O. Hellman and I. A. Abrikosov, Phys. Rev. B **88**, 144301 (2013).
- [21] J. Sanchez, F. Ducastelle, and D. Gratias, Physica A **128**, 334 (1984).
- [22] D. de Fontaine, in *Solid State Physics*, Vol. 47, edited by H. Ehrenreich and D. Turnbull (Academic, New York, 1994).
- [23] L. J. Nelson, G. L. W. Hart, F. Zhou, and V. Ozoliņš, Phys. Rev. B **87**, 035125 (2013).
- [24] L. J. Nelson, V. Ozoliņš, C. S. Reese, F. Zhou, and G. L. W. Hart, Phys. Rev. B **88**, 155105 (2013).
- [25] E. J. Candès and T. Tao, IEEE Trans. Inform. Theory **51**, 4203 (2005).
- [26] E. J. Candès, J. K. Romberg, and T. Tao, Comm. Pure Appl. Math. **59**, 1207 (2006).
- [27] E. J. Candès, J. Romberg, and T. Tao, IEEE Trans. Inform. Theory **52**, 489 (2006).
- [28] X. Gonze and C. Lee, Phys. Rev. B **55**, 10355 (1997).
- [29] D. L. Donoho and X. Huo, IEEE Trans. Inform. Theory **47**, 2845 (2001).
- [30] A series expansion in orthogonal polynomials and their values determined simultaneously with the FCTs (details in a forthcoming paper).
- [31] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
- [32] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
- [33] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
- [34] W. Li, J. Carrete, N. A. Katcho, and N. Mingo, Comp. Phys. Commun. **185**, 1747 (2014).
- [35] C. J. Glassbrenner and G. A. Slack, Phys. Rev. **134**, A1058 (1964).
- [36] M. S. Green, J. Chem. Phys. **22**, 398 (1954).
- [37] R. Kubo, J. Phys. Soc. Jap. **12**, 570 (1957).
- [38] F. Müller-Plathe, J. Chem. Phys. **106**, 6082 (1997).
- [39] D. J. Evans, Phys. Lett. A **91**, 457 (1982).
- [40] Contributions from third- and higher-order interactions to \mathbf{F}_{ab} were obtained by partitioning the energy evenly among all interacting atoms, including repeated sites.
- [41] B. Håkansson and P. Andersson, J. Phys. Chem. Solids **47**, 355 (1986).
- [42] M. V. Klein and R. F. Caldwell, Rev. Sci. Instrum. **37**, 1291 (1966).
- [43] K. A. McCarthy and S. S. Ballard, J. Appl. Phys. **31**, 1410 (1960).
- [44] X. Lu, D. T. Morelli, Y. Xia, F. Zhou, V. Ozolins, H. Chi, X. Zhou, and C. Uher, Adv. Energy Mater. **3**, 342 (2012).

Supplemental Material

A. The procedure of model building

Our CSLD method for lattice anharmonicity and thermal conductivity calculations follow the procedure as outlined in Fig. S1. Note that in this procedure:

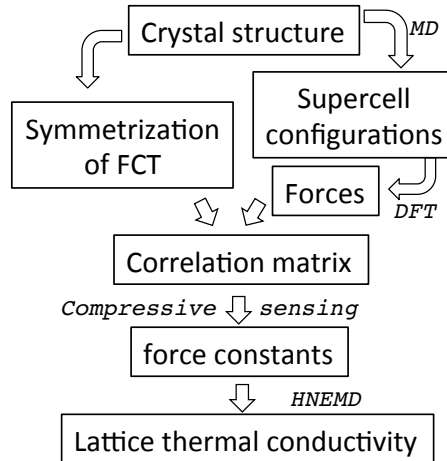


FIG. S1. Flow chart of calculating force constants and lattice thermal conductivity.

- The supercell structures used for training the lattice dynamical models are obtained as follows:
 1. *Ab initio* or classical molecular dynamics at high temperature is performed to obtain supercell snapshots. In this work we run Born-Oppenheimer *ab initio* MD at around 600-800 K for the case studies. Relatively small cutoff energy and few K-points may be used in these MD simulations since we are primarily interested in training structures, not accurate energy and forces. In each material, 5–15 snapshots were taken with time intervals of 100 ~ 150 fs from the MD trajectory.
 2. Random small displacements ($\lesssim 0.1 \text{ \AA}$) are given to each atom in the snapshot. These displacements constitute the training structure.
- Similar to the supercell small displacement method for phonons, high quality DFT calculations are performed for each snapshot to obtain atomic forces. Care was taken to ensure they are well converged with respect to K-point sampling, plane-wave cutoff energy, etc.
- The correlation matrix is computed from the structure-force relationship taking into account the symmetry properties and constraints of the force constants, e.g. space group symmetry and translational invariance.
- The pre-conditioned split-Bregman compressive sensing algorithm was used to fit the very large number of parameters and enhance the numerical stability of fitting force constants. The fitting accuracy is validated with a separate set of “hold-out” structures that are not included in the training set. We monitor the prediction accuracy over a range of μ values to determine the optimal μ .
- The obtained FCT’s are also applied to predict a third, independent set of structures/force components to check the accuracy.

B. Simple example of the compressive sensing approach

We use the compressive sensing approach to treat the under-determined linear problem $\mathbf{f} = \mathbf{A}\mathbf{u}$ while minimizing the ℓ_1 norm $\|\mathbf{u}\|_1 \equiv \sum_i |u_i|$. At the very core of the compressive sensing approach, one makes the assumption that the solution vector is sparse, or has few nonzero components. The ℓ_1 norm is an effective constraint to direct the search for optimal fitting towards the most sparse solution. To illustrate how compressive sensing finds an optimal

sparse solution, examine the trivially simple, underdetermined system of $7x + 10y = 20$, as shown in Fig. S2. This example was previously used to demonstrate the Bayesian compressive sensing approach for cluster expansion [L. J. Nelson, V. Ozolins, C. S. Reese, F. Zhou, and G. L. W. Hart, Phys. Rev. B 88, 155105 (2013)]. Minimization of the ℓ_2 norm by staying on the straight line corresponds to a dense solution: both x and y are non-vanishing (filled circle in Fig. S2a). Minimizing the ℓ_1 norm $|x| + |y|$ corresponds to an optimal sparse solution $\{x = 0, y = 2\}$ (filled diamond in Fig. S2b). Since we are seeking a solution with as few nonzero components as possible, this solution is obviously preferable. Note that the other sparse solution $\{x = 20/7, y = 0\}$ has a larger ℓ_1 norm of $20/7$ and is therefore not optimal.

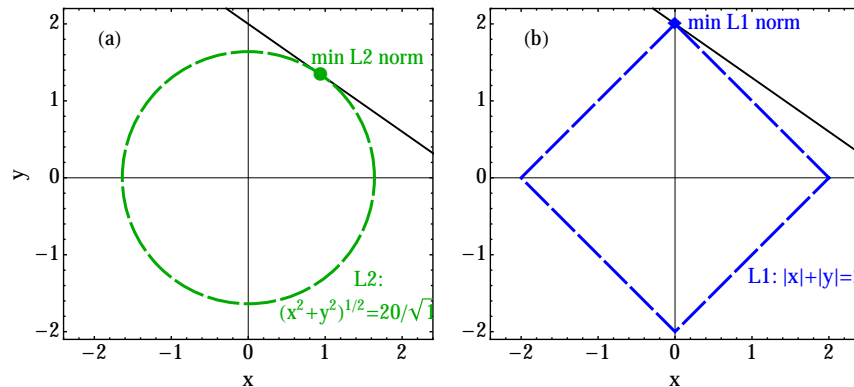


FIG. S2. An example to compare of the iso-surface of (a) the ℓ_2 norm (Euclidian norm $\sqrt{x^2 + y^2}$) and (b) ℓ_1 norm ($|x| + |y|$) in 2D, shown as dashed lines. The under-determined linear equation is $7x + 10y = 20$, designated by the solid straight line. The filled point where the iso-surface touches the straight line is the solution with the smallest ℓ_2 or ℓ_1 norm.

A large number of CS algorithms have been developed thanks to very active method development efforts in the CS community. To clearly illustrate how a sparse solution is obtained, we consider here a conceptually simple iterative algorithm called fixed point continuation (FPC) [E. T. Hale, W. Yin, and Y. Zhang, SIAM Journal on Optimization 19, 1107 (2008)] rather than other methods that are more efficient but complicated.

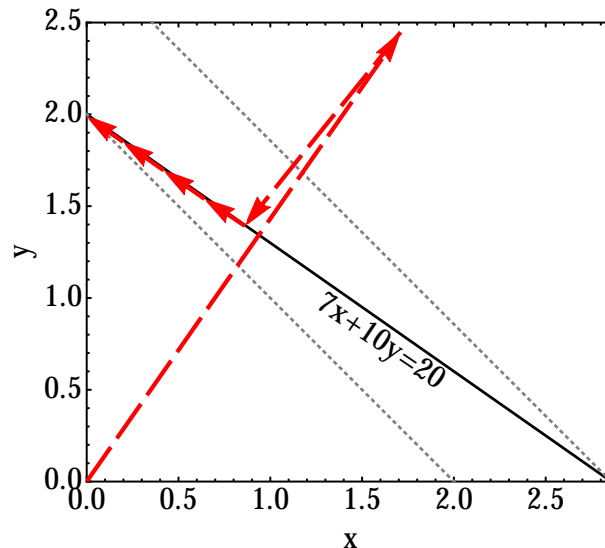


FIG. S3. Application of the FPC algorithm to the same example. The arrows designate the iterative search process, starting from $(0, 0)$, which optimizes the objective function to find the optimal solution vector $(0, 2)$. For clarity only 7 points along the iterative process are shown. With the first two points, FPC mainly optimizes fitting to the line and moves almost perpendicular to it. The rest of the points show minimization of the ℓ_1 norm of the solution point by moving towards $\{0, 2\}$, as can be seen with the help of the gray dotted lines representing the ℓ_1 norm iso-surfaces at $|x| + |y| = 2$ and $20/7$, respectively.

- The objective function consists of an ℓ_1 term of the solution vector plus an ℓ_2 term for the fitting error:

$$\frac{1}{\mu} \|\mathbf{u}\|_1 + \frac{1}{2} \|\mathbf{f} - \mathbb{A}\mathbf{u}\|_2^2$$

- Input sensing matrix \mathbb{A} ($M \times N$ dimensional), \mathbf{f} (M dimensional), μ

1. Initialize solution vector $\vec{u}^0 = \mathbf{0}$, step size $\tau = \min(1.999, -1.665M/N + 2.665)$. Normalize \mathbb{A} so that $\alpha_A = \max(\text{eigenvalues}(\mathbb{A}^T \mathbb{A})) \leq 1$. This may easily be accomplished by dividing both \mathbb{A} and \vec{f} by $\sqrt{\alpha_A}$.
2. for $k = 0, 1, 2, \dots$
 - (a) $\vec{g}^k := \mathbb{A}^T(\mathbb{A}\vec{u}^k - \vec{f})$, where g is the gradient of the ℓ_2 term, i.e. the direction of steepest descent to minimize the ℓ_2 term in the objective function
 - (b) $u_n^{k+1} := \text{shrink}(u_n^k - \tau g_n^k, \tau/\mu)$. In this step, the new solution vector u^{k+1} takes a step in the $-g$ direction and then gets “shrunk” to minimize the ℓ_1 norm (see below)
 - (c) break if converged
3. end for

- The shrinkage operator is defined as $\text{shrink}(y, \alpha) := \text{sign}(y) \max(|y| - \alpha, 0)$. It decreases y 's absolute value by α and sets y to zero if $|y| \leq \alpha$. Shrinkage is a **critical** step in the FPC algorithm to get a sparse solution vector. Parameters that are negligibly small are truncated to zero in the final solution.

- Convergence is reached when the gradient g drops below the shrinkage threshold and the change in the solution vector u^k is sufficiently small.

The FPC algorithm is applied to the simple problem $7x + 10y = 20$ as shown in Fig. S3. The iterative optimization process attempts to minimize both the ℓ_2 error of fitting ($\|\mathbf{f} - \mathbb{A}\mathbf{u}\|_2^2$) by moving perpendicular towards the solid line (most notably at the first two arrows of Fig. S3) and staying along the solid line, and the ℓ_1 norm of the solution vector ($\|\mathbf{u}\|_1 = |x| + |y|$) by shrinking x and y . The combined result is that x is shrunk to zero, leaving $y = 2$ (Fig. S3).

C. Additional case study results

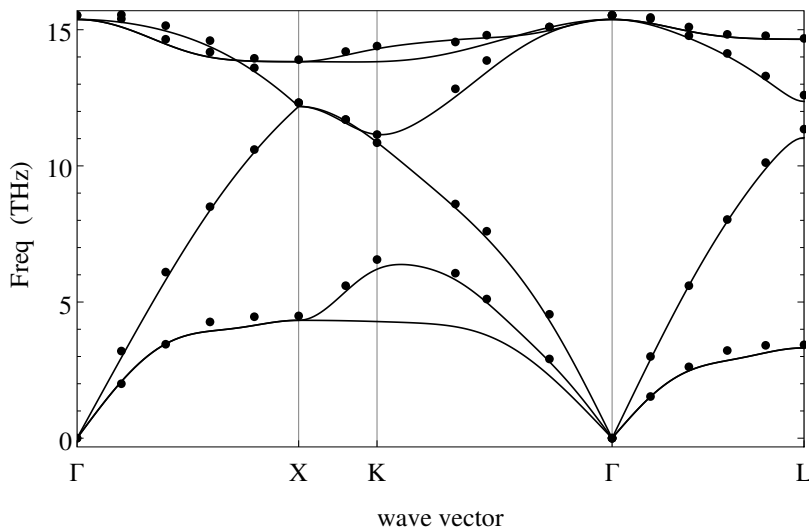


FIG. S4. Phonon dispersion curve of Si using harmonic force constants extracted by CSLD. The dots are inelastic neutron scattering data from G. Dolling, *Inelastic Scattering of Neutrons in Solids and Liquids* (IAEA, Vienna, 1963).

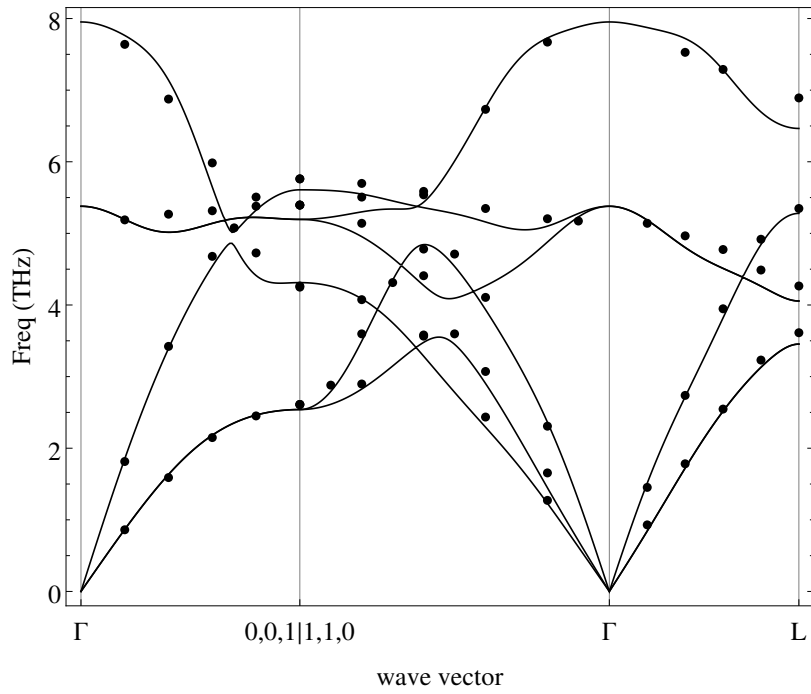


FIG. S5. The same for NaCl. The dots are inelastic neutron scattering data from G. Raunio, L. Almqvist and R. Stedman, *Phys. Rev.*, **178**, 1496. (1969). Non-analytical corrections for long-range dipole-dipole interactions have been added using the Parlinski scheme to reproduce the LO-TO splitting. [K. Parlinski, Z. Q. Li, and Y. Kawazoe, *Phys. Rev. Lett.* 81, 3298 (1998)].

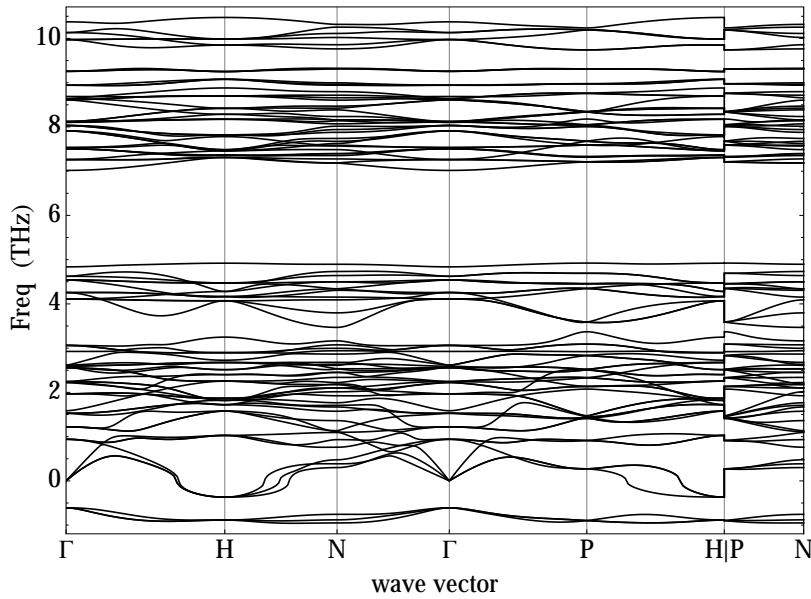


FIG. S6. The same for $\text{Cu}_{12}\text{Sb}_4\text{S}_{13}$. Since it is a metallic system, non-analytical correction was not necessary. Note that this is calculated with PBE while our previous results were obtained with DFPT using the local density approximation (LDA) [X. Lu, D. T. Morelli, Y. Xia, F. Zhou, V. Ozolins, H. Chi, X. Zhou, and C. Uher, *Adv. Energy Mater.* 3, 342 (2012)]

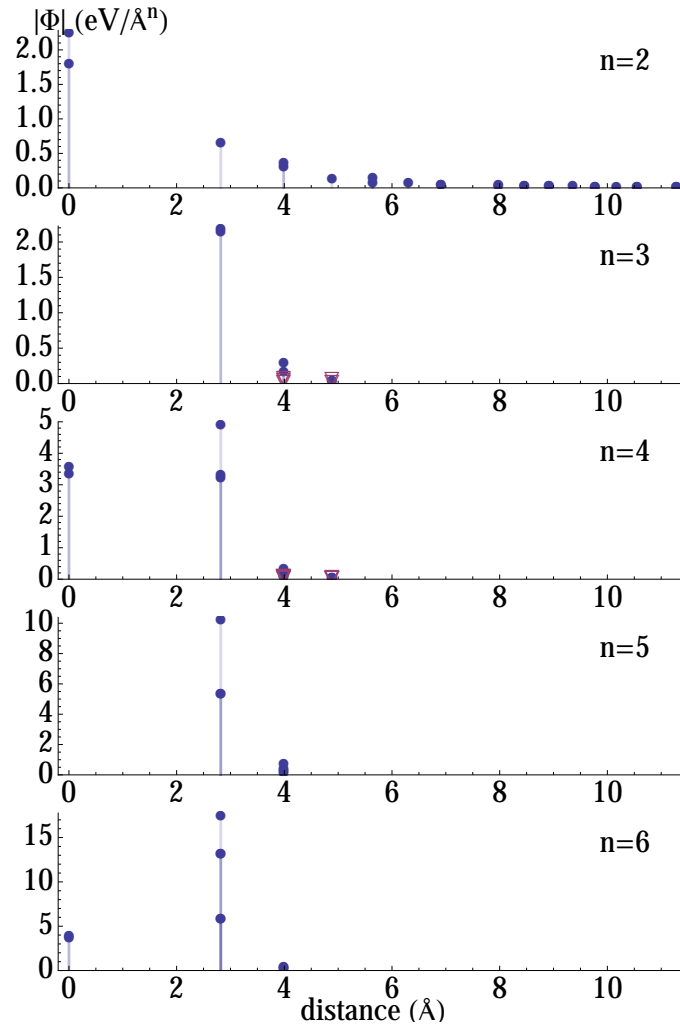


FIG. S7. Frobenius norm of calculated n -th order FCT for NaCl ($n = 2-6$) vs. interaction distance defined as the maximum distance between interacting atoms. The filled dots represent one (e.g. Φ_{aa} and Φ_{aaaa}) or two-body (e.g. Φ_{ab} and Φ_{aaaabb}) FCTs, while the open triangles represent three body interactions, e.g. Φ_{abc} and Φ_{aaabc} . Zero distance implies one-body interactions. Note the absence of odd-order one-body FCTs Φ_{aaa} and Φ_{aaaaa} due to symmetry. Clearly the short range interactions dominate, particularly for higher order anharmonic terms, the most significant of which are nearest-neighbor Na-Cl interactions. Multi-body Na-Cl-Na and Cl-Na-Cl interactions (purple triangles) are found to be very weak.

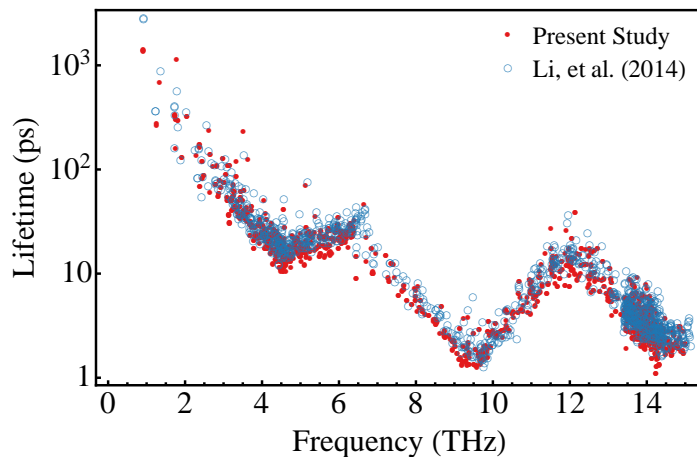


FIG. S8. Phonon mode lifetimes in silicon at $T = 300$ K calculated on a $19 \times 19 \times 19$ mesh of wave vectors. Data from literature were also shown for comparison [W. Li, J. Carrete, N. A. Katcho, and N. Mingo, *Comp. Phys. Commun.* 185, 1747 (2014)]

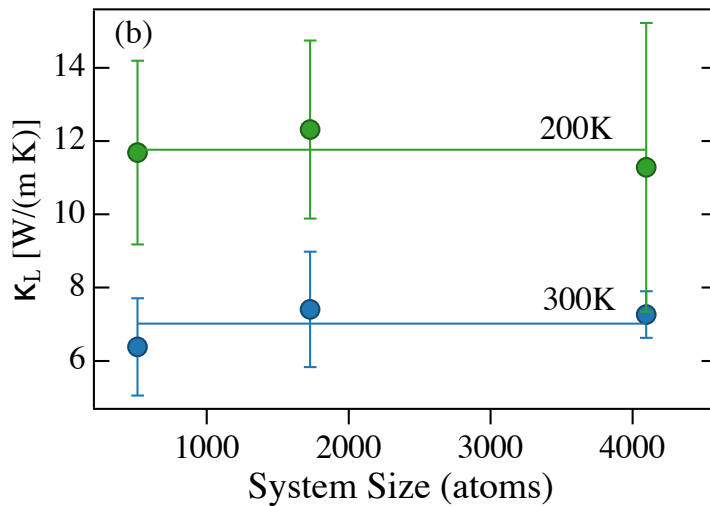


FIG. S9. Calculated κ_L of NaCl vs. the size of the simulation cell. There is no discernible size-dependence in the range of supercells tested.

D. Performance of LMD

LMD has been parallelized to take advantage of large computational systems. The following figures were obtained with sodium chloride system consisting of 1440 atoms. We were not able to perform a single DFT calculation with the computational resources available to us.

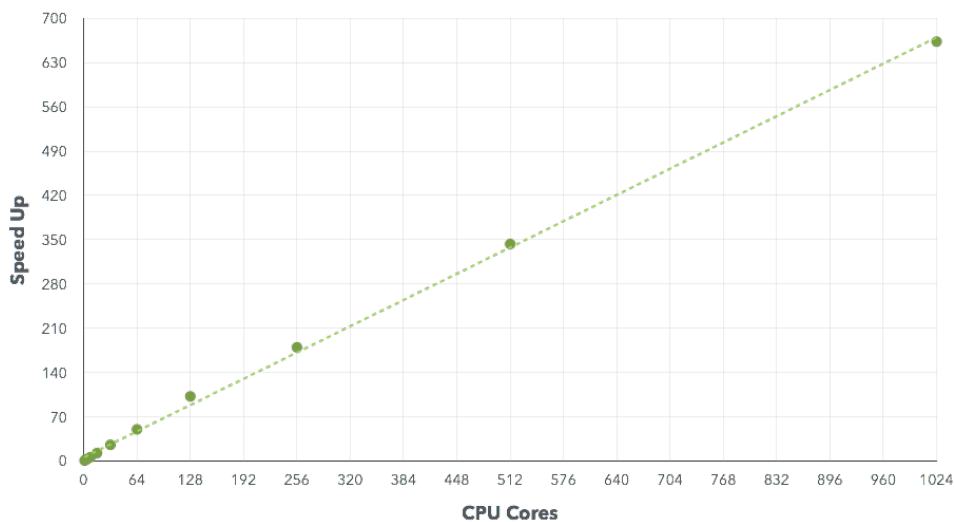


FIG. S10. LMD exhibits linear performance scaling with increasing parallelization across CPUs.

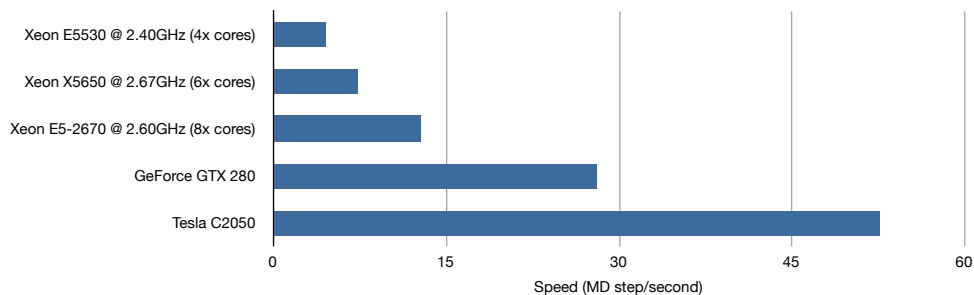


FIG. S11. LMD performance on a range of different hardware and varying levels of parallelization. The bottom two devices are graphics processing units (GPUs), both of which outperform conventional CPUs by a large margin.