Supplementary Information for
"Database and deep-learning scalability of anharmonic phonon properties by automated brute-force first-principles calculations"

Masato Ohnishi[1,2,*] Tianqi Deng[3,4], Pol Torres[5], Zhihao Xu[6], Terumasa Tadano[7], Haoming Zhang[3,4], Wei Nong[8], Masatoshi Hanai[9], Zhiting Tian[10], Ming Hu[11], Xiulin Ruan[12], Ryo Yoshida[2,13], Toyotaro Suzumura[9], Lucas Lindsay[14], Alan J. H. McGaughey[15], Tengfei Luo[6,16], Kedar Hippalgaonkar[8,17,18], and Junichiro Shiomi[1,2,19,20]

[1] Institute of Engineering Innovation, The University of Tokyo, Tokyo 113-0032, Japan
[2] The Institute of Statistical Mathematics, Research Organization of Information and Systems, Tachikawa, Tokyo 190-0014, Japan
[3] State Key Laboratory of Silicon and Advanced Semiconductor Materials, School of Materials Science and Engineering, Zhejiang University, Hangzhou 310027, China
[4] Key Laboratory of Power Semiconductor Materials and Devices of Zhejiang Province, Institute of Advanced Semiconductors, ZJU-Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou 311200, China
[5] Eurecat, Technology Centre of Catalonia, Unit of Applied Artificial Intelligence, Cerdanyola del Vallès, 08290, Spain
[6] Department of Aerospace and Mechanical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA
[7] Research Center for Magnetic and Spintronic Materials, National Institute for Materials Science, Tsukuba 305-0047, Japan
[8] School of Materials Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore
[9] Information Technology Center, The University of Tokyo, Tokyo 113-0032, Japan
[10] Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, New York 14853, USA
[11] Department of Mechanical Engineering, University of South Carolina, Columbia, SC 29201, USA
[12] School of Mechanical Engineering and Birck Nanotechnology Center, Purdue University, West Lafayette, IN 47907, USA
[13] The Graduate University for Advanced Studies, SOKENDAI, Tachikawa, Tokyo, 190-8562, Japan
[14] Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[15] Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA
[16] Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556, USA
[17] Institute of Materials Research and Engineering, Agency for Science Technology and Research, Innovis, Singapore 138634, Singapore
[18] Institute for Functional Intelligent Materials, National University of Singapore, Singapore 117544, Singapore
[19] Department of Mechanical Engineering, The University of Tokyo, Tokyo 113-0032, Japan
[20] RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan

I. FORMULATIONS FOR LATTICE THERMAL CONDUCTIVITY

Lattice thermal conductivity was calculated by solving Peierls-Boltzmann transport equation[1] for the particle-like contribution and the Wigner transport formula[2] for the wave-like contribution, both implemented in the ALAMODE package[3,4]. Considering the second-order perturbation within the relaxation approximation, the phonon linewidth due to the three-phonon scattering for phonon mode q is derived as

$$\Gamma_q = \frac{\pi}{16N} \sum_{q_1,q_2} |V_3(-q,q_1,q_2)|^2 \left[(n_1+n_2+1)\delta(\omega_q-\omega_1-\omega_2) - 2(n_1-n_2)\delta(\omega_q-\omega_1+\omega_2)\right], \quad \text{(S1)}$$

where the subscripts (i = 1, 2) denote phonon modes contributing to the scattering of the target mode q, n_i is the Bose-Einstein distribution function, ω_i is the phonon frequency, N is the number of $\boldsymbol{q}$ points, and $\pm q = (\pm\boldsymbol{q}, s)$ with $\boldsymbol{q}$ and s being the wavevector and branch index, respectively. The three-phonon coupling matrix element V_3 is given by

$$V_3(q,q_1,q_2) = \left(\frac{\hbar}{\omega\omega_1\omega_2}\right)^{\frac{1}{2}} \times \sum_{\boldsymbol{R}_i l_i p_i} \Psi^{p_0 p_1 p_2}_{0 l_0, \boldsymbol{R}_1 l_1, \boldsymbol{R}_2 l_2} \times \frac{\boldsymbol{e}^{p_0}_{l_0}(q)\boldsymbol{e}^{p_1}_{l_1}(q_1)\boldsymbol{e}^{p_2}_{l_2}(q_2)}{\sqrt{M_{l_0}M_{l_1}M_{l_2}}} \times \exp[i(\boldsymbol{q}\cdot\boldsymbol{R}_0 + \boldsymbol{q_1}\cdot\boldsymbol{R}_1 + \boldsymbol{q_2}\cdot\boldsymbol{R}_2)], \quad \text{(S2)}$$

where $\hbar$ is the reduced Planck constant, $\boldsymbol{R}_i$ is the position of the primitive cell, l_i is the atom index, p_i is the direction of the displacement of atom l_i, M is the atomic mass, Ψ is the cubic force constants, and $\boldsymbol{e}(q)$ is the eigenvector of the mode q. The phonon lifetime due to three phonon scattering τ_{pp} is given by $\tau_{\mathrm{pp}}(q) = 1/(2\Gamma_q)$. The total phonon lifetime of mode q is obtained with Matthiessen's rule: $\tau_q^{-1} = \tau_{q,\mathrm{pp}}^{-1} + \tau_{q,\mathrm{iso}}^{-1}$, where $\tau_{q,\mathrm{iso}}^{-1}$ is the scattering rate due to isotopes[5]. Finally, the Peierls term is calculated as $\kappa_{\mathrm{p}}^{\alpha\beta,\mathrm{scp}}(T) = (NV)^{-1}\sum_q c_q(T) v_q^{\alpha}(T) v_q^{\beta}(T)\tau_q(T)$. Here, V is the volume of the primitive unit cell, c is the mode specific heat, v is the group velocity, and α and β are the Cartesian directions.

In the automated workflow, the coherence contribution (κ_{c}), corresponding to the nondiagonal terms of the heat flux operator, was also considered. Its contribution can be obtained as

$$\kappa_{\mathrm{c}}^{\alpha\beta} = \frac{\hbar^2}{k_{\mathrm{B}}T^2VN} \sum_{\boldsymbol{q}} \sum_{s_1\neq s_2} \frac{\omega_1+\omega_2}{4} \mathrm{V}_{12}^{\alpha}(\boldsymbol{q}) \mathrm{V}_{21}^{\beta}(\boldsymbol{q}) \frac{\omega_1 n_1(n_1+1) + \omega_2 n_2(n_2+1)}{(\omega_1-\omega_2)^2 + (\Gamma_1+\Gamma_2)^2} (\Gamma_1+\Gamma_2), \quad \text{(S3)}$$

where k_{B} is the Boltzmann constant. The subscripts (1 and 2) denote the phonon modes $(\boldsymbol{q}, s_1)$ and $(\boldsymbol{q}, s_2)$, respectively. The generalized group velocity operator V(q) is given as

$$\mathrm{V}_{12}(\boldsymbol{q}) = \frac{1}{\omega_1+\omega_2} \left\langle \boldsymbol{e}(q_1) \middle| \frac{\partial D(\boldsymbol{q})}{\partial \boldsymbol{q}} \middle| \boldsymbol{e}(q_2) \right\rangle, \quad \text{(S4)}$$

where $D(\boldsymbol{q})$ is the dynamical matrix.

II. ELEMENTAL OCCURRENCE FREQUENCY

The elemental occurrence frequency in the materials for which anharmonic properties were calculated in the anharmonic phonon property database (APDB) is illustrated in the periodic table shown in Fig. S1. The number shown for each element represents the number of materials that include that element. For example, the most frequently occurring element, oxygen, is found in 2,425 out of the 6,113 materials in the current version of the APDB.

This figure shows that chalcogens (group 16) and alkali metals (group 1) are the two most frequently occurring groups of elements. It also indicates that elements from groups 1–2 and 11–17 are commonly present in the APDB materials, whereas transition metals occur less frequently—likely due to the absence of magnetic materials in the current version of the APDB. Group 18 elements are included only as single-element systems.

Overall, the APDB contains a broad range of elements, which is important for improving data quality and enabling its application to diverse fields—not only for thermal properties, but also for various other material properties, including mechanical, electronic, electrical, optical, and magnetic characteristics.

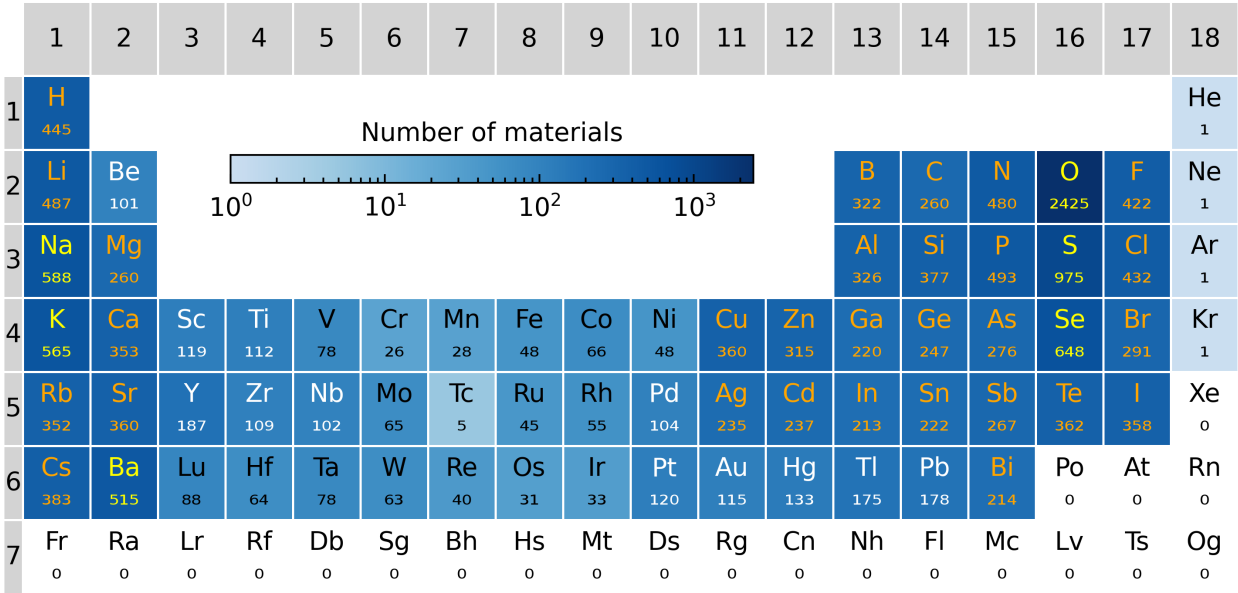


Fig. S1. Elemental occurrence frequency in the anharmonic phonon property database (APDB). The number in each cell indicates how frequently each element appears in the materials included in the APDB, with the magnitude visualized using a logarithmic blue color scale. Element symbols are colored according to their occurrence: yellow for elements appearing in more than 500 materials, orange for more than 200, white for more than 100, and black for all others.

III. LIST OF MATERIALS WITH HIGH AND LOW THERMAL CONDUCTIVITIES

The target materials consist of non-metallic, non-magnetic materials from the Materials Project[6–8], as well as those from Phonondb[9], totaling 21,452 materials. To construct APDB, we developed an automated calculation software package, auto-kappa, for anharmonic phonon properties. So far, we have calculated the anharmonic phonon properties of approximately 6,000 materials, while harmonic phonon properties have been computed for an even larger set of materials.

Among the calculated high-κ materials, the following materials, including their polymorphs, are novel or have been rarely discussed as high-κ materials: triclinic $Hg(BiS_2)_2$ (943 $Wm^{-1}K^{-1}$), cubic HC (306 $Wm^{-1}K^{-1}$), cubic BiB (235 $Wm^{-1}K^{-1}$), and trigonal $CsHoS_2$ (340 $Wm^{-1}K^{-1}$). Only 90 materials (1.5%) has a κ_{lat} exceeding 100 $Wm^{-1}K^{-1}$, as shown in Table S1. On the other hand, discovering novel materials with low thermal conductivity appears to be much easier. In APDB, more than 1,500 materials (>28%) have κ_{lat} below 1.0 $Wm^{-1}K^{-1}$. Even when limited to binary compounds, 187 materials (3.1%) have a $\kappa_{lat} \leq 1.0\ Wm^{-1}K^{-1}$. Considering that the thermal conductivities of typical thermoelectric materials Bi_2T_3 and PbTe are 1.2[10] and 2.2[11–13] $Wm^{-1}K^{-1}$, the constructed database provides an exciting opportunity to search for materials with low-κ for applications such as thermal insulators and thermoelectric materials.

Phonon properties of the fifty-five high-κ material ($\kappa_{lat} \geq 200\ Wm^{-1}K^{-1}$) and nineteen low-$\kappa$ materials ($\kappa_{lat} \leq 0.1\ Wm^{-1}K^{-1}$) are shown in Fig. S1 and Fig. S2, respectively.

Table S1. Number of materials exhibiting high and low thermal conductivity in APDB. The lattice thermal conductivity is evaluated as $\kappa_{lat} = (\kappa_{xx} + \kappa_{yy} + \kappa_{zz})/3$ for all materials, including anisotropic ones, for simplicity.

High thermal conductivity		**Low thermal conductivity**	
Range of κ_{lat} ($Wm^{-1}K^{-1}$)	**Number of materials**	**Range of κ_{lat} ($Wm^{-1}K^{-1}$)**	**Number of materials**
>1,000	9 (0.15%)	<0.1	19 (0.32%)
>500	20 (0.34%)	<0.2	197 (3.3%)
>200	55 (0.93%)	<0.5	910 (16%)
>100	90 (1.5%)	<1.0	1,656 (28%) (187 for binary)

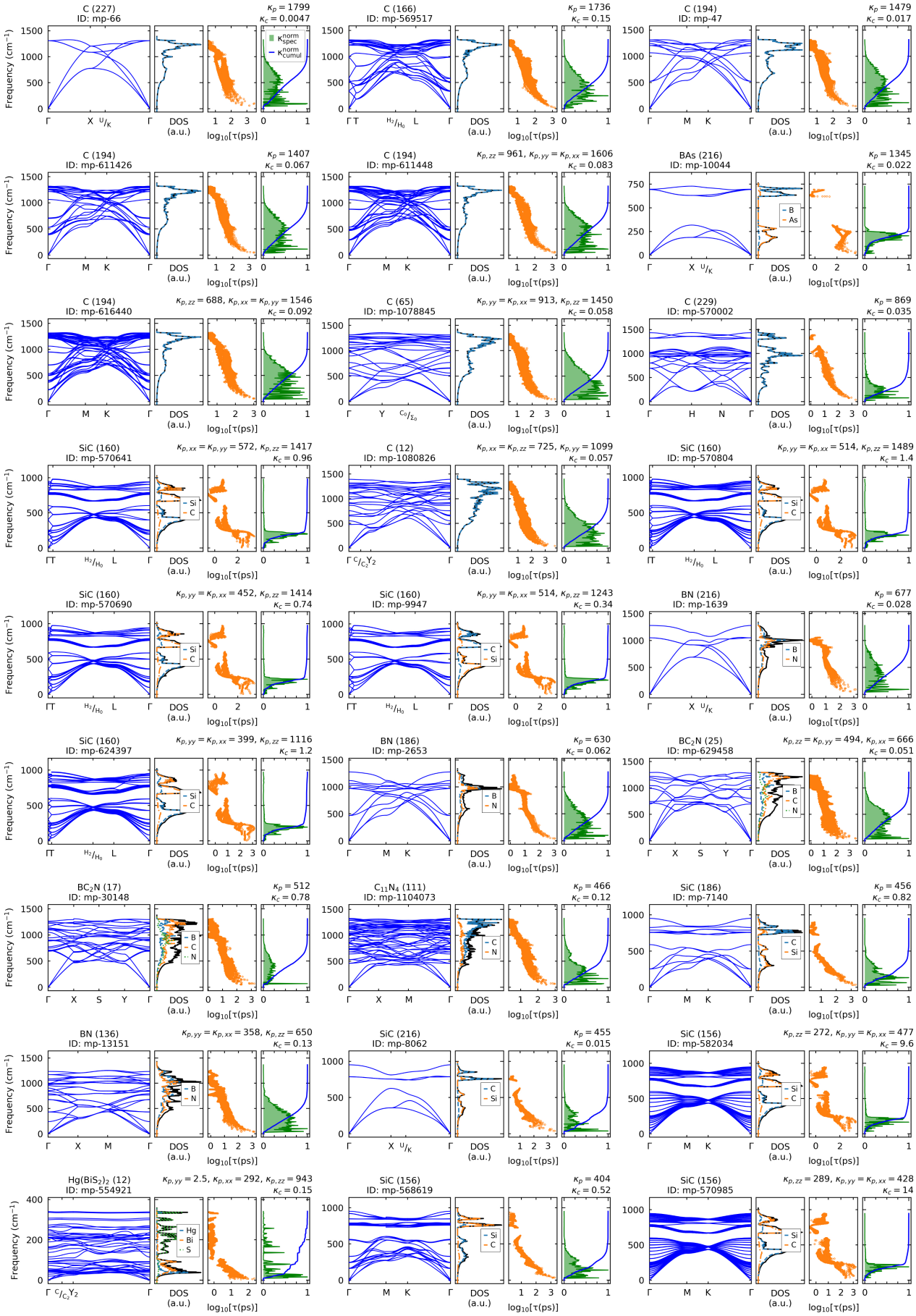


Fig. S2 (1/2). List of phonon properties for materials with high thermal conductivity

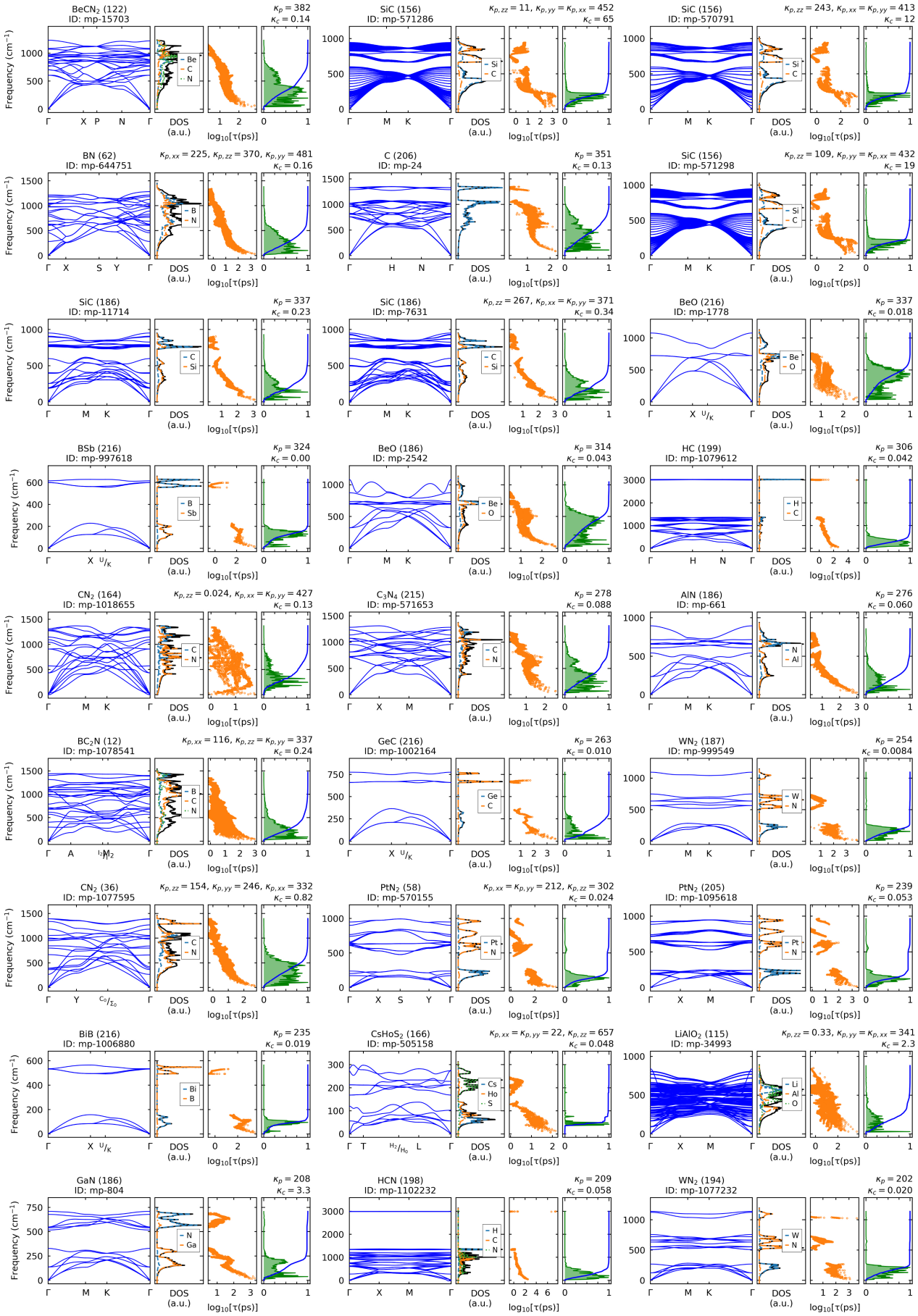


Fig. S2 (2/2). List of phonon properties for materials with high thermal conductivity.

Fig. S2. List of phonon properties for materials with high thermal conductivity ($> 200\,\mathrm{Wm^{-1}K^{-1}}$), ordered by magnitude. Each panel includes phonon dispersion, density of states (DOS), phonon lifetime (in logarithmic scale), and spectral (green) and cumulative (blue) thermal conductivity for Peierls component (κ_p), which are normalized by their respective maximum and total values. The chemical formula, space group number (in parentheses), and Materials Project ID are shown in the top-left corner, while the magnitudes of the Peierls (κ_p) and coherent (κ_c) components are displayed in the top-right corner of each panel.

1

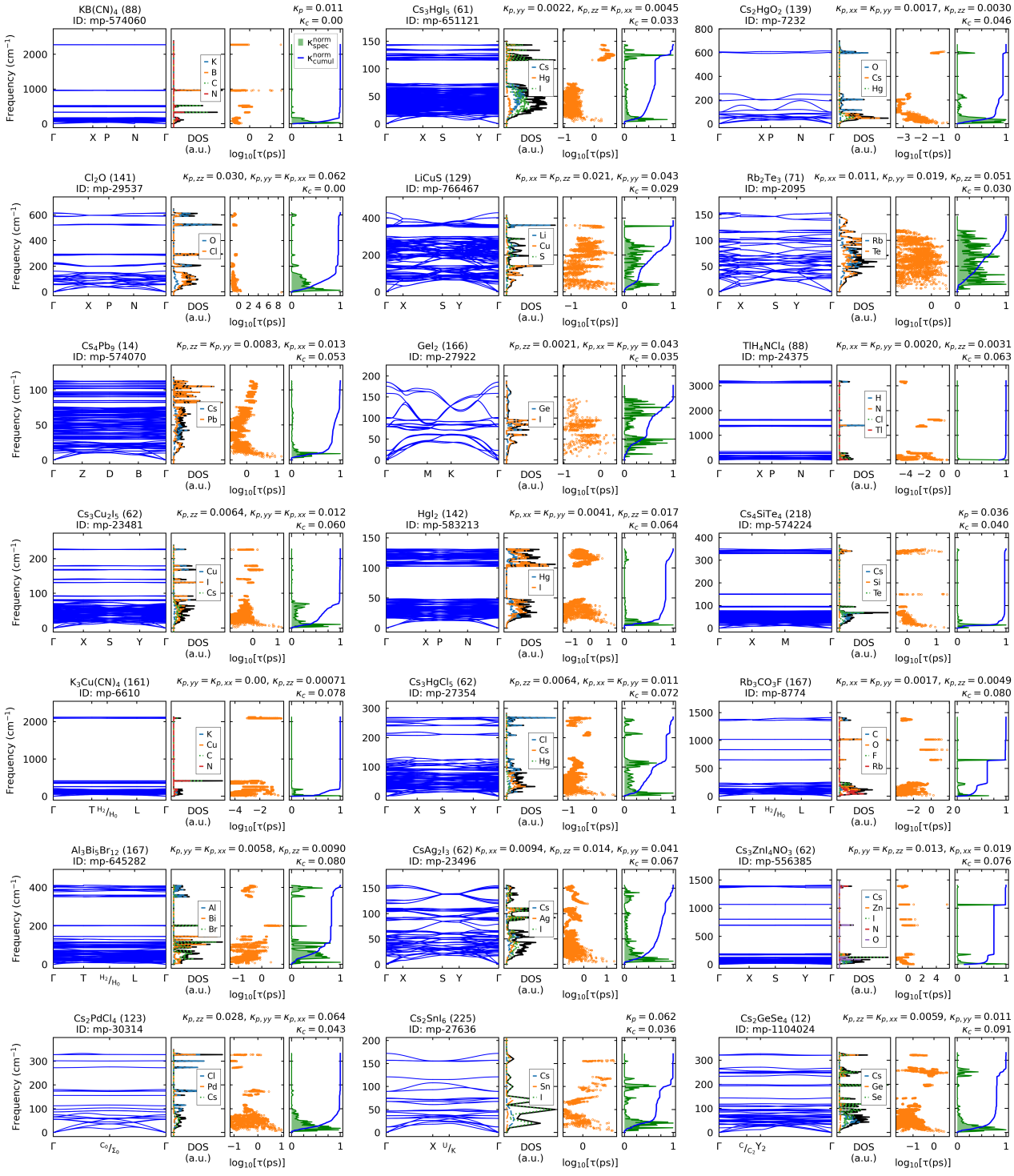


Fig. S3. List of phonon properties for materials with low thermal conductivity (< 0.1 $Wm^{-1}K^{-1}$), arranged in ascending order. See the caption of Fig. S1 for a detailed description of the figure.

IV. COMPARISON BETWEEN PARTICLE AND COHERENT CONTRIBUTIONS TO HEAT TRANSPORT

Lattice thermal conductivity is governed by both particle-like phonon gas transport and wave-like tunneling mechanisms[2,14]. In the early stages of first-principles phonon analysis, calculating the coherent component arising from wave-like tunneling was challenging. However, since Simoncelli *et al.* proposed a unified theory (also known as the Wigner heat transport formula) to evaluate both the Peierls component (κ_p) from particle-like phonons and the coherent component (κ_c)[2], this approach has been widely adopted for analysis[15].

Among the materials calculated in APDB, we identified that materials with a large coherent component (κ_c) are polymorphs of SiC and related structures with the space group *P3m1* (156). While the relative contribution of κ_c to total heat transport remains small due to the relatively large κ_p, the absolute magnitude of κ_c exceeds 50 $Wm^{-1}K^{-1}$. This significant coherent contribution arises from densely packed phonon branches, which result from the large number of atoms in the primitive cell. Their thermal conductivities (κ_p and κ_c) and phonon dispersions are shown in Fig. S3.

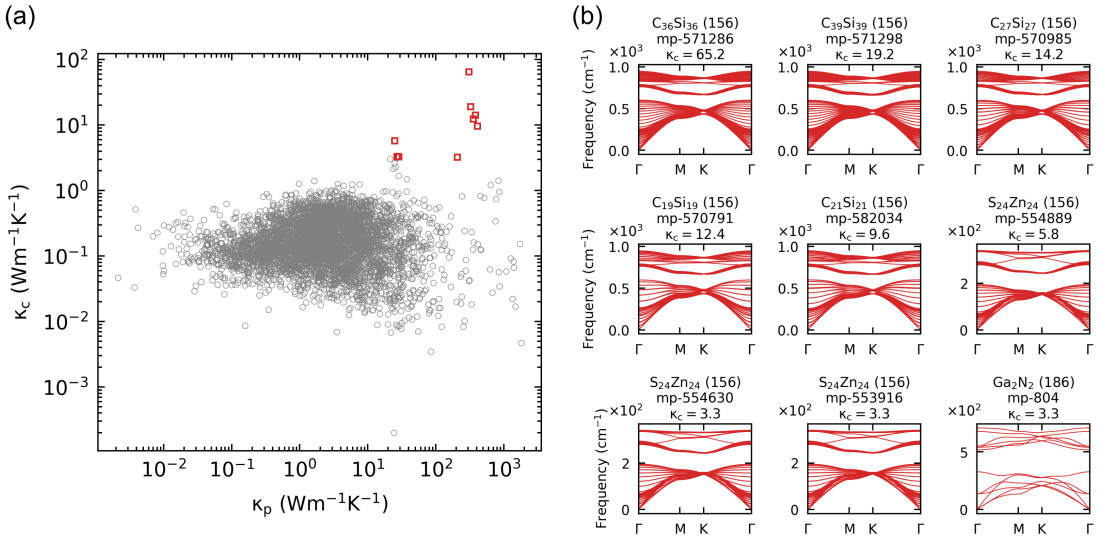


Fig. S4. Materials exhibiting large coherent thermal conductivity. (a) Distribution of Peierls (κ_p) and coherent (κ_c) thermal conductivity. Red markers indicate the top nine materials with the highest κ_c. (b) Phonon dispersion for these nine materials.

V. PRECAUTIONS FOR AUTOMATED CALCULATIONS

In high-throughput calculations, there is always a trade-off between computational accuracy and speed. Based on our knowledge and experience with first-principles phonon calculations, we have aimed to ensure the highest possible computational accuracy in the automated calculations conducted in this study. However, we found that some materials still exhibit implausible data, as shown in Fig. S4.

Implausible data led to excessively high κ_p, exceeding a few thousand $Wm^{-1}K^{-1}$. In most cases, this overestimation arises from excessively long phonon lifetimes in isolated flat branches or low-frequency acoustic modes, as well as from excessively high group velocities. To eliminate the excessively long lifetimes, the increase in **q**-mesh density and consideration of four-phonon interaction[16] are necessary. The computational cost of four-phonon interactions is excessively high, but machine-learning prediction[17] may be an effective strategy for accurately predicting high κ_p values. In contrast, excessively large group velocities are likely due to insufficient structural optimization. For example, implausible changes in phonon branches can be observed around the T point of Sr_2CuBrO_2 in Fig. S4. While these implausible data can be automatically handled in future versions of the software, they were manually excluded from the discussion in this study.

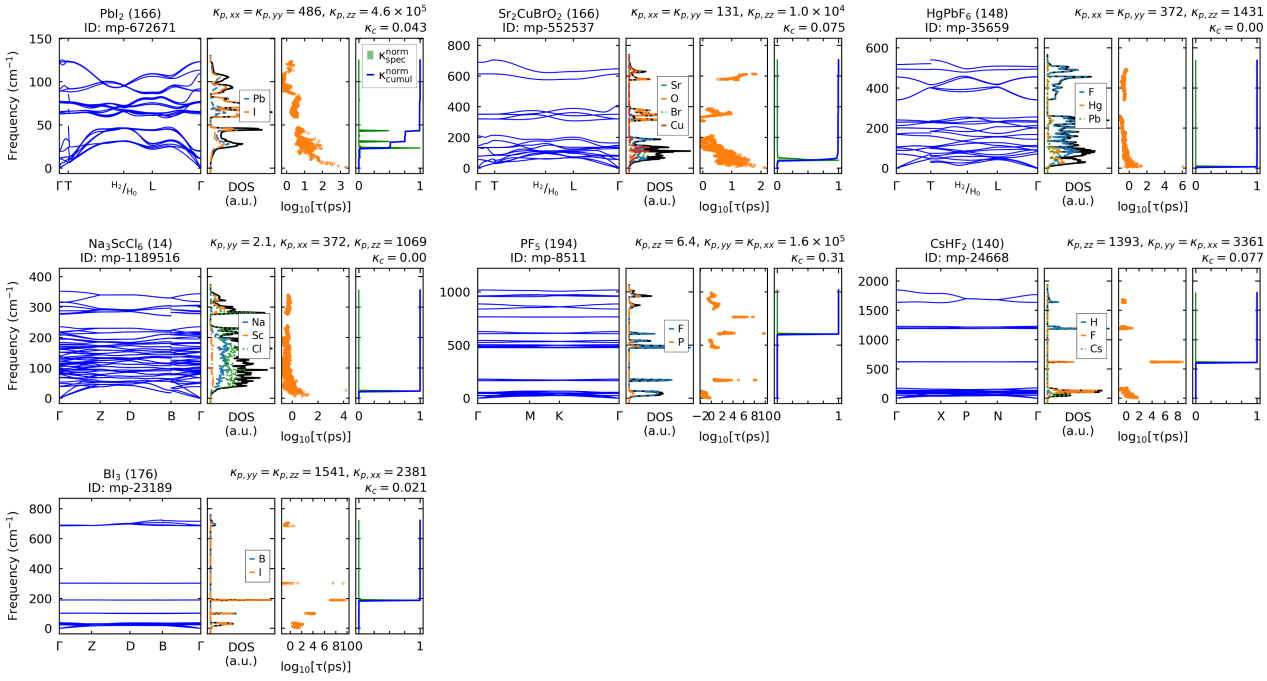


Fig. S5. Results of the automated calculations, including implausible data.

VI. MACHINE LEARNING-BASED PREDICTION OF PHONON PROPERTIES

Using the APDB, we predicted the Peierls lattice thermal conductivity (κ_{p}) and its cumulative value with respect to the mean free path (κ_{cumul}), as mentioned in the main text. In addition to these properties, we also conducted the prediction for the coherent component (κ_{c}), maximum phonon frequency (ω_{max}), and spectral κ_{p} as a function of frequency ($\kappa_{\mathrm{spec}}(\omega)$). For training of these properties, we used $\log_{10}(\kappa_{\mathrm{c}}+\varepsilon)$ (where $\varepsilon = 0.001\,\mathrm{Wm^{-1}K^{-1}}$), $\log_{10}\omega_{\mathrm{max}}$, and $\kappa_{\mathrm{spec}}(\omega)$ normalized by its maximum value, respectively. For $\kappa_{\mathrm{spec}}^{\mathrm{norm}}(\omega)$, the data were prepared from 0 to ω_{max} with 51 points. In the study on phonon density of states (DOS) prediction[18], which we referred to for the prediction method using the Euclidean neural networks (e3nn)[19], DOS data were discretized into 51 points from 0 to 1,000 cm^{-1}. In contrast, in our prediction, the frequency was normalized by its maximum value because the maximum frequency does not reach 1,000 cm^{-1} in most materials. Spectral thermal conductivity as a function of the absolute frequency can be obtained by combining the predicted values of ω_{max} and $\kappa_{\mathrm{spec}}^{\mathrm{norm}}(\omega/\omega_{\mathrm{max}})$. For the machine learning technique, CGCNN[20] was applied to the prediction of κ_{c} and ω_{max} and e3nn[19] was applied to κ_{spec}.

We observed deep learning scaling laws for these three phonon properties, similar to those for κ_{p} and $\kappa_{\mathrm{cumul}}^{\mathrm{norm}}(\Lambda)$, as discussed in the main text. In every case, the prediction accuracy increased as the training dataset size increased. The scaling factor was 0.06 to 0.3, which are compatible magnitude as those observed for large language model (0.095)[21] and deep learning for materials in a previous study (0.21)[22]. Notably, since the prediction accuracy for ω_{max} was remarkably high (with an MAE of 0.030 on a logarithmic scale), integrating the predictions of ω_{max} and $\kappa_{\mathrm{spec}}^{\mathrm{norm}}(\omega/\omega_{\mathrm{max}})$ may provide more valuable information than directly predicting $\kappa_{\mathrm{spec}}^{\mathrm{norm}}(\omega)$, which approaches zero at high frequencies.

While the prediction accuracy for $\kappa_{\mathrm{spec}}(\omega)$ (with an MAE of 0.073) was slightly lower than that for $\kappa_{\mathrm{cumul}}(\Lambda)$ (with an MAE of 0.068), APDB enables accurate prediction of the spectral anharmonic property, as shown in Fig. S5. The figure illustrates that spectral thermal conductivity was excellently predicted for half of the materials (green and blue), whereas predicting spectra with multiple peaks remains challenging, as indicated by the orange and red lines in Fig. S5.

Consequently, these data clearly show that APDB opens up multiple possibilities for predicting not only phonon properties but also their interactions with other quasiparticles or related properties. At the same time, further increasing the dataset size or developing advanced machine learning techniques for phonon properties remains essential.

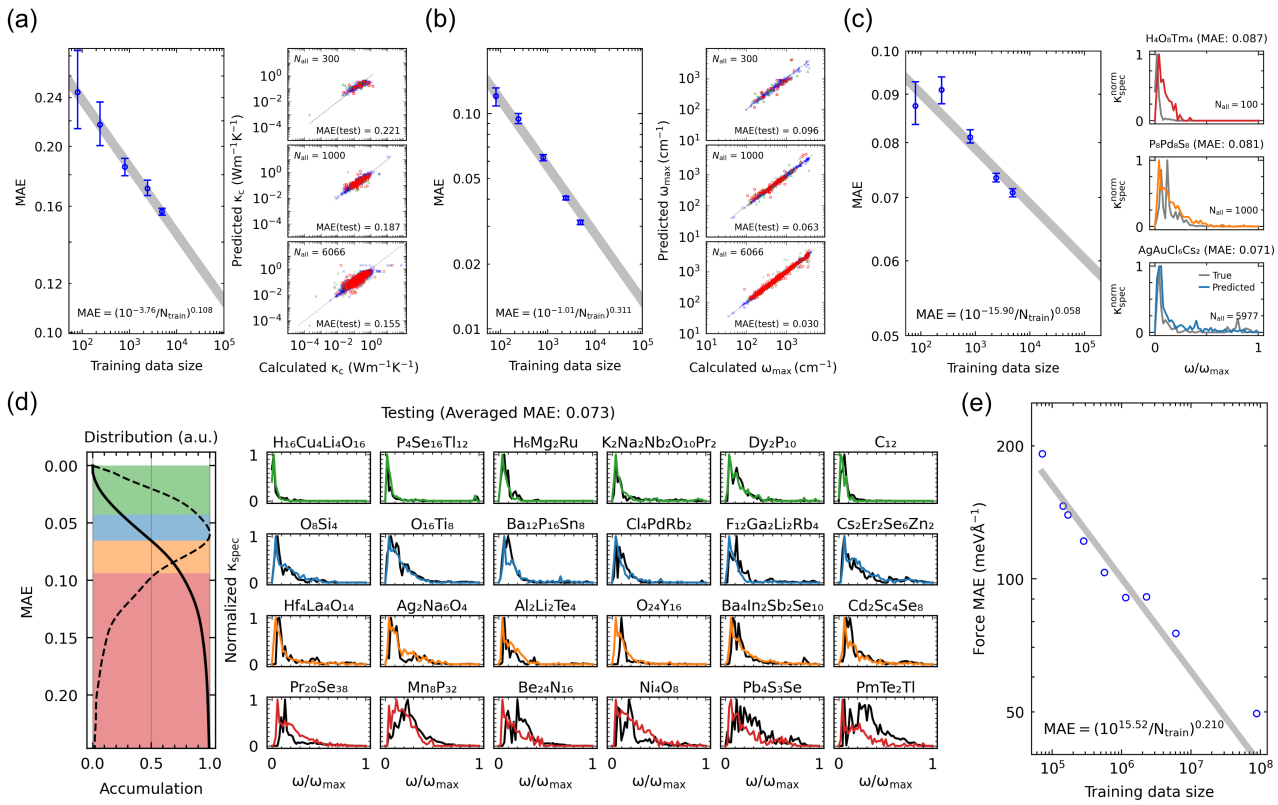
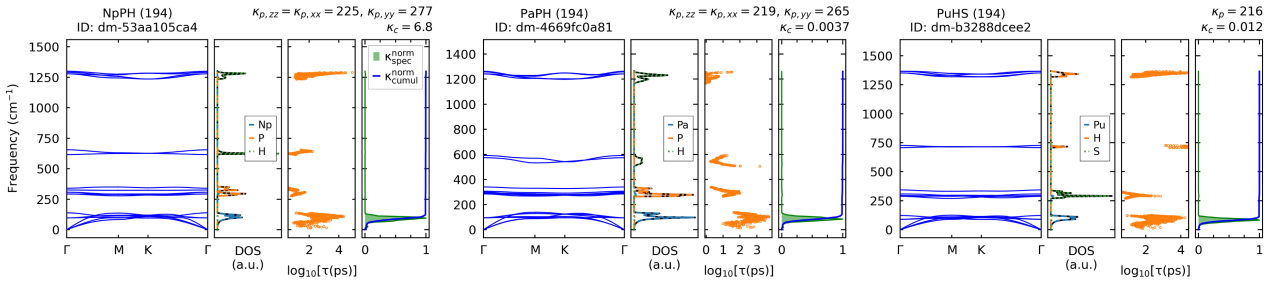


Fig. S6. Deep learning scaling law for phonon properties with respect to the training data size. (a) Coherent thermal conductivity (κ_c), (b) maximum frequency (ω_{max}), and (c) normalized spectral Peierls thermal conductivity (κ_{spec}^{norm}) were predicted. The MAEs were evaluated with $\log_{10}(\kappa_c + \varepsilon)$, where $\varepsilon = 0.001$ $Wm^{-1}K^{-1}$, $\log_{10}\omega_{max}$, and κ_{spec}^{norm}, respectively. (d) κ_{spec}^{norm} prediction for the test data using all available data ($N_{all} \approx 6{,}000$). See the caption of Fig. 3 in the main text for details.

VII. LIST OF SCREENED MATERIALS WITH HIGH AND LOW THERMAL CONDUCTIVITY

Using the prediction models built with APDB, we screened materials with high and low κ_p from the GNoME database, as discussed in the main text. Through this screening, we identified three materials with $\kappa_{lat} > 200\ Wm^{-1}K^{-1}$ and nine materials with $\kappa_{lat} < 0.2\ Wm^{-1}K^{-1}$. Their phonon properties are presented in Fig. S6.

(a)



(b)

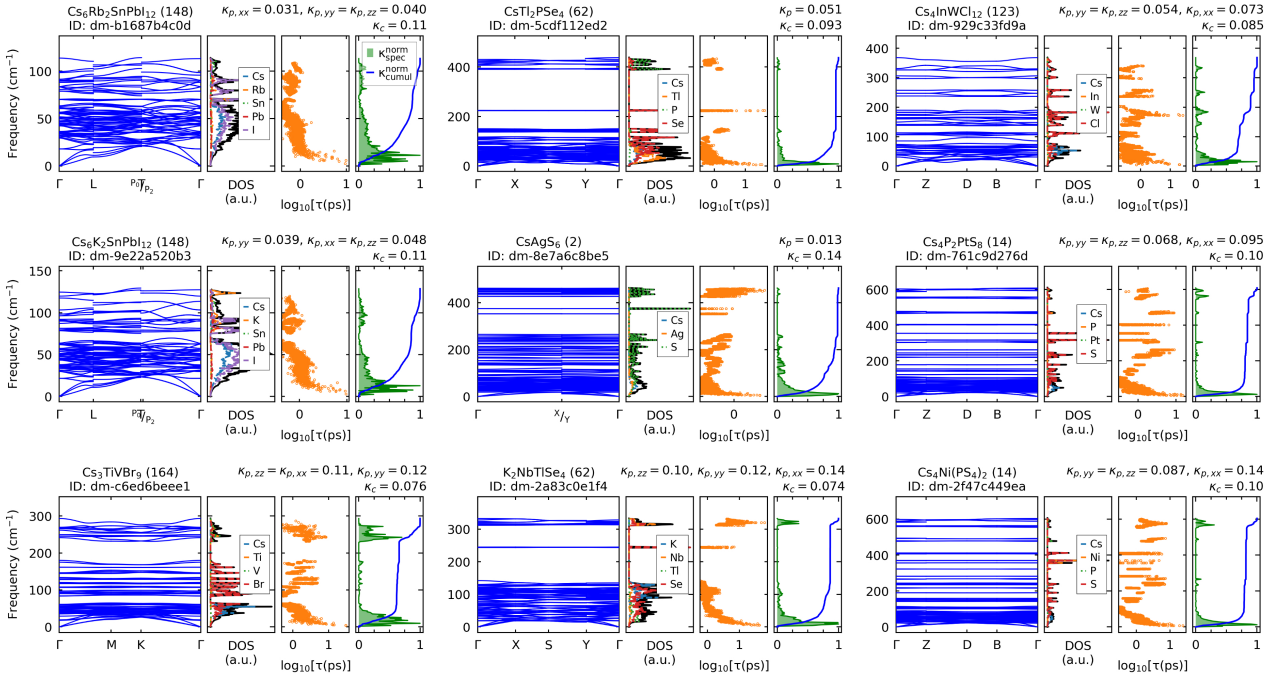


Fig. S7. Novel materials exhibiting extreme thermal conductivity. (a) A list of materials with $\kappa_{lat} > 200\ Wm^{-1}K^{-1}$ and (b) materials with $\kappa_{lat} < 0.2\ Wm^{-1}K^{-1}$. The materials were identified through screening of the GNoME database[22].

REFERENCES

1. Ziman, J. M. *Electrons and Phonons: The Theory of Transport Phenomena in Solids*. (Oxford University Press, 2001). doi:10.1093/acprof:oso/9780198507796.001.0001.

2. Simoncelli, M., Marzari, N. & Mauri, F. Unified theory of thermal transport in crystals and glasses. *Nat. Phys.* **395**, 1–813 (2019).

3. 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).

4. Tadano, T. & Tsuneyuki, S. First-Principles Lattice Dynamics Method for Strongly Anharmonic Crystals. *J. Phys. Soc. Jpn.* **87**, 041015 (2018).

5. Tamura, S. Isotope scattering of dispersive phonons in Ge. *Phys. Rev. B* **27**, 858–866 (1983).

6. Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).

7. 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).

8. Ong, S. P. *et al.* The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles. *Comput. Mater. Sci.* **97**, 209–215 (2015).

9. Togo, A. Phonondb. https://github.com/atztogo/phonondb.

10. Goldsmid, H. J. The Thermal Conductivity of Bismuth Telluride. *Proc. Phys. Soc. Sect. B* **69**, 203–209 (1956).

11. Akhmedova, G. A. & Abdinov, D. Sh. Effect of thallium doping on the thermal conductivity of PbTe single crystals. *Inorg. Mater.* **45**, 854–858 (2009).

12. Sootsman, J. R., Chung, D. Y. & Kanatzidis, M. G. New and Old Concepts in Thermoelectric Materials. *Angew. Chem. Int. Ed.* **48**, 8616–8639 (2009).

13. El-Sharkawy, A. A., El-Azm, A. M. A., Kenawy, M. I., Hillal, A. S. & Abu-Basha, H. M. Thermophysical properties of polycrystalline PbS, PbSe, and PbTe in the temperature range 300–700 K. *Int. J. Thermophys.* **4**, 261–269 (1983).

14. Allen, P. B., Feldman, J. L., Fabian, J. & Wooten, F. Diffusons, locons and propagons: Character of atomic vibrations in amorphous Si. *Philosophical Magazine B* **79**, 1715–1731 (2009).

15. Ohnishi, M., Tadano, T., Tsuneyuki, S. & Shiomi, J. Anharmonic phonon renormalization and thermal transport in the type-I $Ba_8Ga_{16}Sn_{30}$ clathrate from first principles. *Phys. Rev. B* **106**, 024303 (2022).

16. Feng, T., Lindsay, L. & Ruan, X. Four-phonon scattering significantly reduces intrinsic thermal conductivity of solids. *Phys. Rev. B* **96**, 161201 (2017).

17. Feng, T. & Ruan, X. Quantum mechanical prediction of four-phonon scattering rates and reduced thermal conductivity of solids. *Phys. Rev. B* **93**, 045202 (2016).

18. Chen, Z. *et al.* Direct Prediction of Phonon Density of States With Euclidean Neural Networks. *Adv. Sci.* **8**, 2004214 (2021).

19. Geiger, M. & Smidt, T. e3nn: Euclidean Neural Networks. *arXiv* (2022) doi:10.48550/arxiv.2207.09453.

20. Xie, T. & Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction

of Material Properties. *Phys. Rev. Lett.* **120**, 145301 (2018).

21. Kaplan, J. *et al.* Scaling Laws for Neural Language Models. *arXiv* (2020) doi:10.48550/arxiv.2001.08361.

22. Merchant, A. *et al.* Scaling deep learning for materials discovery. *Nature* **624**, 80–85 (2023).