Supporting Information

Higher-order phonon scattering and lattice thermal conductivity prediction via machine learning

Shengxian Liu, Kan Tang, Jian Zhou* and Zhimei Sun*

[1]School of Materials Science and Engineering, Beihang University, Beijing 100191, China.

*Corresponding authors: jzhou@buaa.edu.cn (J.Zhou), zmsun@buaa.edu.cn (Z. M. Sun)

The weighted 3ph and 4ph phase space is written as[1]:

$$WP_3(\lambda)^{(+)} = \frac{1}{N}\sum_{\lambda'\lambda''} \frac{n^0_{\lambda'} - n^0_{\lambda''}}{\omega_\lambda \omega_{\lambda'} \omega_{\lambda''}} \delta(\omega_\lambda + \omega_{\lambda'} - \omega_{\lambda''}) \quad \text{* MERGEFORMAT (1)}$$

$$WP_3(\lambda)^{(-)} = \frac{1}{N}\sum_{\lambda'\lambda''} \frac{n^0_{\lambda'} + n^0_{\lambda''} + 1}{2\omega_\lambda \omega_{\lambda'} \omega_{\lambda''}} \delta(\omega_\lambda - \omega_{\lambda'} - \omega_{\lambda''}) \quad \text{* MERGEFORMAT (2)}$$

$$WP_4(\lambda)^{(++)} = \frac{1}{N^2}\sum_{\lambda'\lambda''\lambda'''} \frac{(1+n^0_{\lambda'})(1+n^0_{\lambda''})n^0_{\lambda'''}}{2n^0_\lambda} \frac{\delta(\omega_\lambda + \omega_{\lambda'} + \omega_{\lambda''} - \omega_{\lambda'''})}{\omega_\lambda \omega_{\lambda'} \omega_{\lambda''} \omega_{\lambda'''}} \text{* MERGEFORMAT (3}$$

$$WP_4(\lambda)^{(+-)} = \frac{1}{N^2}\sum_{\lambda'\lambda''\lambda'''} \frac{(1+n^0_{\lambda'})n^0_{\lambda''}n^0_{\lambda'''}}{2n^0_\lambda} \frac{\delta(\omega_\lambda + \omega_{\lambda'} - \omega_{\lambda''} - \omega_{\lambda'''})}{\omega_\lambda \omega_{\lambda'} \omega_{\lambda''} \omega_{\lambda'''}} \text{* MERGEFORMAT (4)}$$

$$WP_4(\lambda)^{(--)} = \frac{1}{N^2}\sum_{\lambda'\lambda''\lambda'''} \frac{n^0_{\lambda'}n^0_{\lambda''}n^0_{\lambda'''}}{6n^0_\lambda} \frac{\delta(\omega_\lambda - \omega_{\lambda'} - \omega_{\lambda''} - \omega_{\lambda'''})}{\omega_\lambda \omega_{\lambda'} \omega_{\lambda''} \omega_{\lambda'''}} \text{* MERGEFORMAT (5)}$$

where n^0_λ is the phonon Bose-Einstein distribution at equilibrium, ω_λ is the phonon frequency for a certain mode λ.

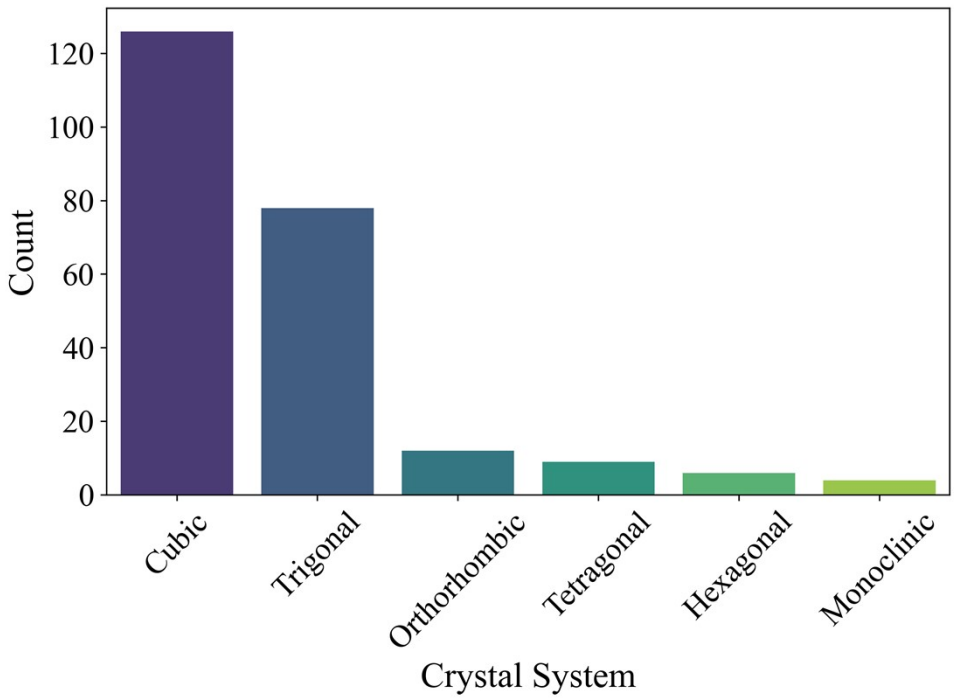


Fig. S1 Distribution of Materials Across Different Crystal Systems

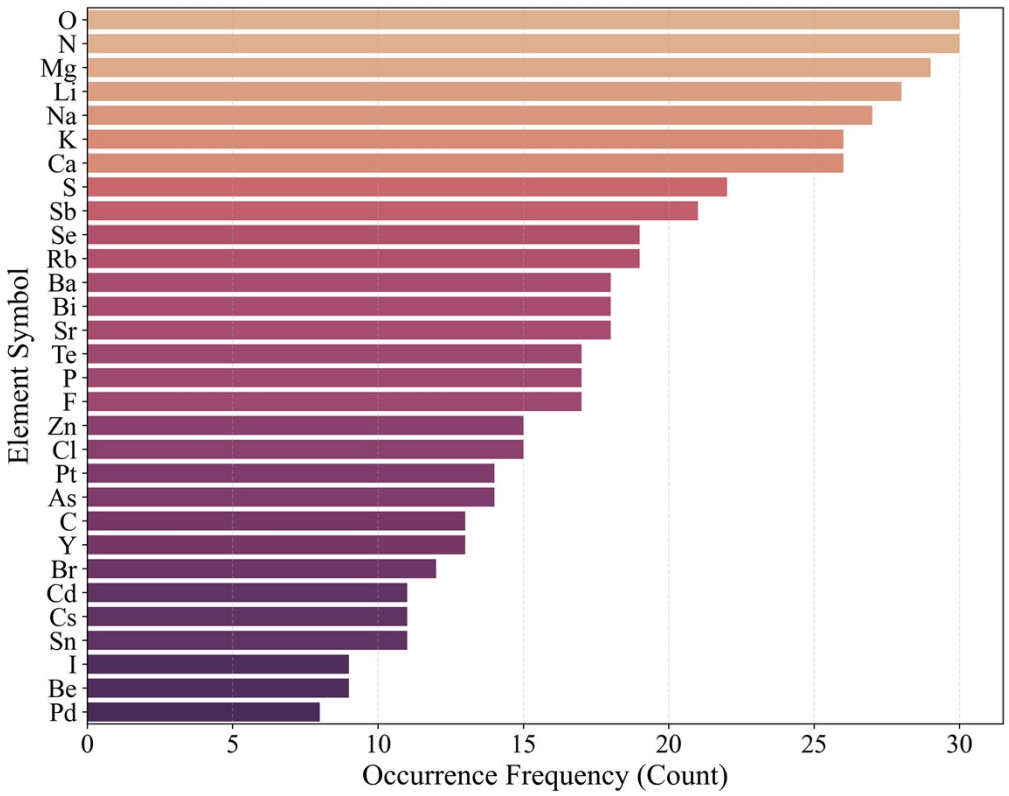


Fig. S2 Elemental Composition Analysis: Frequency of Element Occurrence

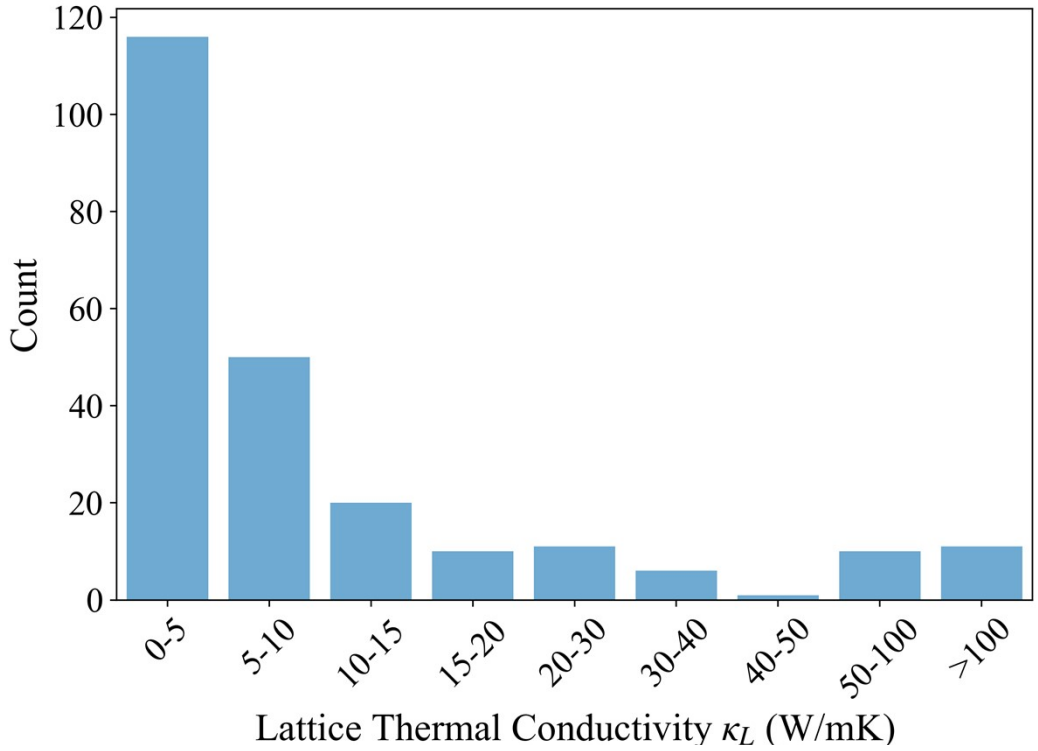


Fig. S3 Statistical Distribution of Calculated κ_{3ph}

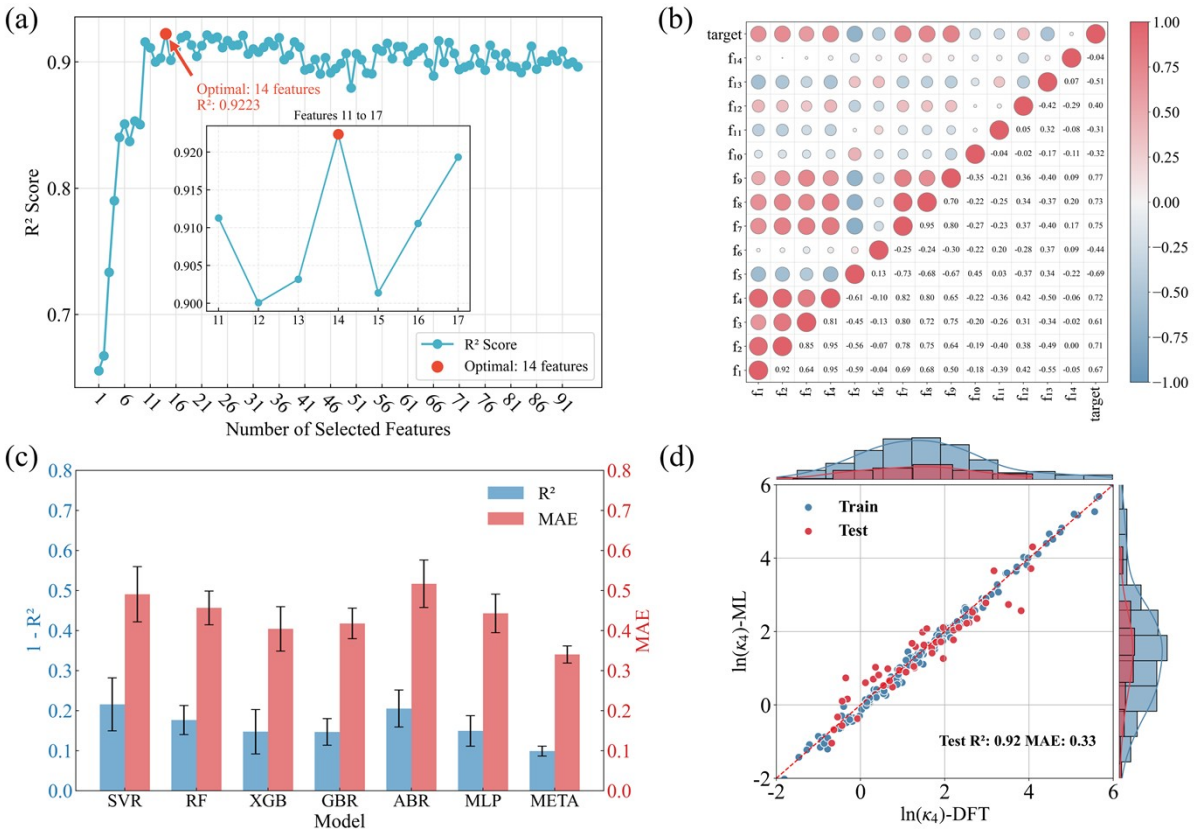


Fig. S4 (a) Recursive feature elimination results identify 14 optimal features achieving an R^2 score of 0.9223. (b) Heatmap of the Pearson correlation between the selected 14 features and κ_{3ph}. (c) 5-fold cross-validation 1-R^2 and MAE scores of six traditional ML models and the META model. (d) Scatter plots comparing the predicted values of the κ_{4ph} model and the DFT calculated values on the training and test sets.

The κ_{4ph} model identifies 14 optimal features through recursive feature elimination (RFE). Inter-feature correlation analysis removes the highly correlated features f_1, f_2, and f_8, resulting in a final optimized feature set of 11 features. Under 5-fold cross-validation, the META model achieves an average R^2 of 0.90 and a mean absolute error (MAE) of 0.34. The best-performing fold is shown in **Fig. S4d**, with an R^2 of 0.92 and an MAE of 0.35.

Table S1 Categorization and summary of the initial 90 candidate descriptors used in the machine learning workflow.

Category	Features	n
Components	Atomic radii, Electronegativity, Melting, Boiling…	62
Structure	Volume, Spacegroup, Lattice constant…	10
DFT calculations	B, G, v, Grüneisen, κ_{slack}, $f_{a,\,max}$…	18

To ensure model transparency, 90 initial descriptors were grouped into three functional categories. The Components set (62 features) covers intrinsic elemental properties such as atomic radii and electronegativity. The Structure set (10 features) captures lattice geometry and symmetry, including volume and lattice constants. The DFT calculations set (18 features) includes primary physical properties derived from basic DFT calculations, such as elastic moduli (B, G), $f_{a,\,max}$, and the Grüneisen parameter, providing a robust physical foundation for the machine learning model.

Table S2 Initial feature set for the κ_{3ph} model.

f_1	E	Young's modulus
f_2	κ_{slack}	Semi-empirical formula for κ_L
f_3	V_{prim}	Volume of the primitive cell
f_4	Grüneisen	Grüneisen constant
f_5	$f_{a,\,max}$	The maximum frequency of the acoustic branch
f_6	$v_{a,\,average}$	The average group velocity of the acoustic phonon
f_7	$f_{o,\,min}$	The minimum frequency of the optical branch
f_8	f_{AO2}	The band gap between the acoustic and optical branches
f_9	Φ_{min}	The minimum electronegativity among elements

f_{10}	$E_{\text{p, min}}$	The minimum energy of P-orbital
f_{11}	$n_{\text{p-unfilled, mean}}$	The average number of unfilled p-electron states
f_{12}	MN_{mean}	The average Mendeleev number among elements
f_{13}	$V_{\text{mol, min}}$	The minimum molar volume of all elemental solid
f_{14}	$V_{\text{mol, mode}}$	The mode of the molar volume of all elemental solid

Table S3 Initial feature set for the $\kappa_{4\text{ph}}$ model.

f_1	B	Bulk Modulus
f_2	E	Young's modulus
f_3	V_{prim}	Volume of the primitive cell
f_4	κ_{slack}	Semi-empirical formula for κ_{L}
f_5	$\kappa_{\text{experience}}$	Semi-empirical formula for κ_{L}
f_6	Grüneisen	Grüneisen constant
f_7	$f_{\text{a, max}}$	The maximum frequency of the acoustic branch
f_8	$v_{\text{a, mean}}$	The average group velocity of the acoustic phonon
f_9	$f_{\text{o, min}}$	The minimum frequency of the optical branch
f_{10}	d_{std}	The standard deviation of all atomic distances
f_{11}	EA_{std}	The standard deviation of electron affinities among elements
f_{12}	MN_{mean}	The average Mendeleev number among elements
f_{13}	$V_{\text{mol, min}}$	The minimum molar volume of all elemental solid
f_{14}	I_{std}	The standard deviation of ionization energies among elements

The κ_{slack} and $\kappa_{\text{experience}}$ is written as:

$$\kappa_{\text{slack}} = A(\gamma)\frac{\overline{M}\theta_{\text{D}}^3\delta}{\gamma^2 n^{2/3}T}(T \geqslant \theta_{\text{D}})^2 \qquad \text{* MERGEFORMAT (6)}$$

$$\kappa_{\text{experience}} = A_1\frac{\overline{M}v_{\text{s}}^{\ 3}}{V^{2/3}n^{1/3}} + A_2\frac{v_{\text{s}}}{V^{2/3}}(1-\frac{1}{n^{2/3}})^3 \qquad \text{* MERGEFORMAT (7)}$$

where $A(\gamma) = 2.43\times10^7/(1-0.514/\gamma+0.228/\gamma^2)$, $\overline{M}$ is the average atomic mass, δ^3 is the average volume per atom in the unit cell, θ_{D} is the Debye temperature, T is the absolute

temperature, and γ is the Grüneisen parameter.

Where $\overline{M}$, V, and n are the average atomic mass, the average volume per atom, and the number of atoms in the primitive cell of each material, respectively. And v_s is the speed of sound, which can be estimated by $(B/d)^{1/2}$, where B and d are the bulk modulus and density of the material, respectively. A_1 and A_2 are fitting parameters.

Table. S4 Comparison of optimized input features, DFT-calculated κ_{3ph}, and META model predicted values for the 3ph scattering model, where mp-id/Ref. denotes the material ID in the Materials Project database or the corresponding reference, E is the Young's modulus, κ_{slack} is the Slack equation-derived thermal conductivity, V_{prim} is the primitive cell volume, Grüneisen is the Grüneisen parameter, $f_{a,\ max}$ and $f_{o,\ min}$ are the maximum acoustic and minimum optical frequencies, f_{AO2} is the acoustic-optical band gap, Φ_{min} is the minimum electronegativity, $E_{p,\ min}$ is the minimum P-orbital energy, $n_{p\text{-}unfilled,\ mean}$ is the average number of unfilled p-electron states, MN_{mean} is the average Mendeleev number, $V_{mol,\ min}$ and $V_{mol,\ mode}$ are the minimum and mode of elemental molar volumes, and $\kappa_{3ph,DFT}$ and $\kappa_{3ph,ML}$ represent the lattice thermal conductivity from DFT calculations and META model predictions, respectively. For these 15 data points, the MAE is 0.51.

Formula	mp-id/Ref	E (GPa)	κ_{slack} (W/m·K)	V_{prim} (Å^3)	Grüneisen	$f_{a,\ max}$ (THz)	$f_{o,\ min}$ (THz)	f_{AO2} (THz)
$Mg(BeAs)_2$	mp-865185	113.28	9.07	80.59	1.81	4.28	1.82	0.43
$Ca(MgAs)_2$	mp-9564	87.20	8.65	118.59	1.73	3.59	2.23	0.62
Ca_2AsI	mp-28554	46.69	3.66	126.11	1.80	2.16	1.22	0.57
LiMgBi	mp-570213	53.05	4.70	79.29	1.75	2.56	5.10	1.99
NbFeSb	mp-9437	207.42	19.42	52.79	1.89	5.59	5.12	0.92
NaI	mp-23268	21.66	1.83	69.55	1.83	2.28	3.04	1.34
Li_2S	mp-1153	79.12	14.08	46.21	1.69	8.07	7.47	0.93
NbRuSb	mp-505297	175.41	13.55	59.21	1.98	4.95	4.72	0.95
ScNiSb	mp-3432	162.17	18.94	57.31	1.78	4.71	4.88	1.04
$PbTiO_3$	mp-20459	96.45	4.97	68.78	1.81	2.57	2.96	1.15
$LaGaTe_2$	[4]	37.38	1.28	119.90	2.10	1.74	1.05	0.60
$LaInTe_2$	[4]	43.63	1.79	125.61	2.01	1.74	1.28	0.74

$LaRbTe_2$	4	40.33	1.87	137.10	1.97	2.01	1.55	0.77
Li_2InBi	5	70.76	5.10	89.38	1.75	1.99	1.38	0.69
Li_2TlBi	5	83.73	6.11	92.38	1.74	1.66	1.55	0.94

73

Formula	Φ_{min}	$E_{p,\ min}$	$n_{p\text{-unfilled},\ mean}$	MN_{mean}	$V_{mol,\ min}$ (cm^3/mol)	$V_{mol,\ mode}$ (cm^3/mol)	$\kappa_{3ph,\ DFT}$ (W/m·K)	$\kappa_{3ph,\ ML}$ (W/m·K)
$Mg(BeAs)_2$	1.31	0.00	1.20	81.00	4.85	4.85	5.955	5.845
$Ca(MgAs)_2$	1.00	0.00	1.20	68.00	12.95	12.95	4.558	3.710
Ca_2AsI	1.00	0.00	1.00	54.50	12.95	26.20	1.450	1.297
LiMgBi	0.98	0.00	1.00	57.33	13.02	13.02	3.917	2.796
NbFeSb	1.60	0.00	1.00	67.33	7.09	7.09	17.367	17.863
NaI	0.93	0.00	0.50	54.00	23.78	23.78	1.096	1.457
Li_2S	0.98	0.00	0.67	39.33	13.02	13.02	12.275	13.753
NbRuSb	1.60	0.00	1.00	67.67	8.17	8.17	12.514	13.619
ScNiSb	1.36	0.00	1.00	58.00	6.59	6.59	12.744	12.880
$PbTiO_3$	1.54	0.00	2.00	87.20	10.64	17.36	2.749	2.493
$LaGaTe_2$	1.10	0.00	2.25	74.50	11.80	20.46	0.320	0.710
$LaInTe_2$	1.10	0.00	2.25	74.00	15.76	20.46	0.391	0.756
$LaRbTe_2$	0.82	0.00	1.00	56.50	20.46	20.46	1.061	0.986
Li_2InBi	0.98	0.00	2.00	47.50	13.02	13.02	1.550	1.855
Li_2TlBi	0.98	0.00	2.00	47.25	13.02	13.02	2.360	1.755

74

75

76

Table. S5. Comparison of optimized input features, DFT-calculated $\kappa_{4ph,DFT}$, and META model predicted values for the 4ph scattering model, where mp-id denotes the material ID in the Materials Project database, κ_{slack} and $\kappa_{experience}$ are the semi-empirical lattice thermal conductivities, V_{prim} is the primitive cell volume, Grüneisen is the Grüneisen parameter, $f_{a,\ max}$ and $f_{o,\ min}$ are the maximum acoustic and minimum optical frequencies, d_{std} is the standard deviation of all atomic distances, EA_{std} is the standard deviation of electron affinities, MN_{mean} is the average Mendeleev number, $V_{mol,\ min}$ is the minimum elemental molar volume, I_{std} is the standard deviation of ionization energies, and $\kappa_{4ph,DFT}$ and $\kappa_{4ph,ML}$ represent the lattice thermal conductivity from DFT calculations and META model predictions, respectively. For these 6 data points, the MAE is 0.33.

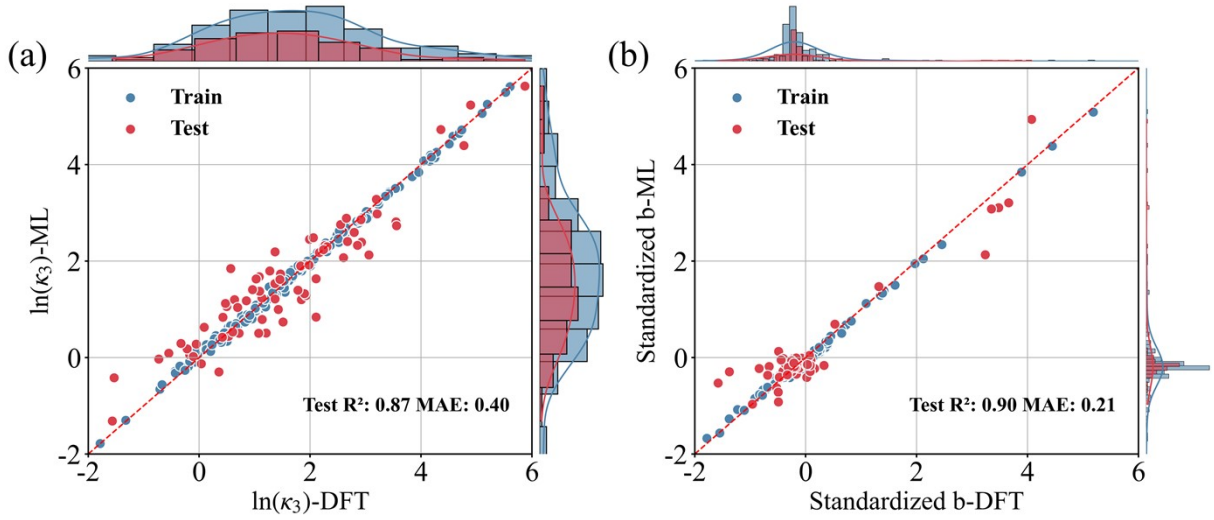
Formula	mp-id	κ_{slack} (W/m·K)	$\kappa_{experience}$ (W/m·K)	V_{prim} (Å^3)	Grüneisen	$f_{a,\ max}$ (THz)	$f_{o,\ min}$ (THz)
$Mg(BeAs)_2$	mp-865185	9.07	6.33	80.59	1.81	4.28	1.82
Ca_2AsI	mp-28554	3.66	5.04	126.11	1.80	2.16	1.22
NbFeSb	mp-9437	19.42	36.59	52.79	1.89	5.59	5.12
NaI	mp-23268	1.83	2.34	69.55	1.83	2.28	3.04
Li_2S	mp-1153	14.08	10.28	46.21	1.69	8.07	7.47
NbRuSb	mp-505297	13.55	37.89	59.21	1.98	4.95	4.72

Formula	d_{std}	EA_{std}	MN_{mean}	$V_{mol,\ min}$ (cm^3/mol)	I_{std}	$\kappa_{4ph,\ DFT}$ (W/m·K)	$\kappa_{4ph,\ ML}$ (W/m·K)
$Mg(BeAs)_2$	0.19	0.63	81.00	4.85	1.36	4.423	4.361
Ca_2AsI	0.22	1.24	54.50	12.95	3.50	1.170	0.990
NbFeSb	0.00	0.39	67.33	7.09	1.00	13.320	14.088

NaI	0.00	1.26	54.00	23.78	14.08	0.979	1.164
Li_2S	0.00	0.69	39.33	13.02	24.66	11.476	10.709
NbRuSb	0.00	0.06	67.67	8.17	1.12	11.352	11.362

89

90



91

92 **Fig. S5** (a) Scatter plots comparing the predicted values of the dual-label regression model and

93 the DFT calculated values on the training and test sets.

94

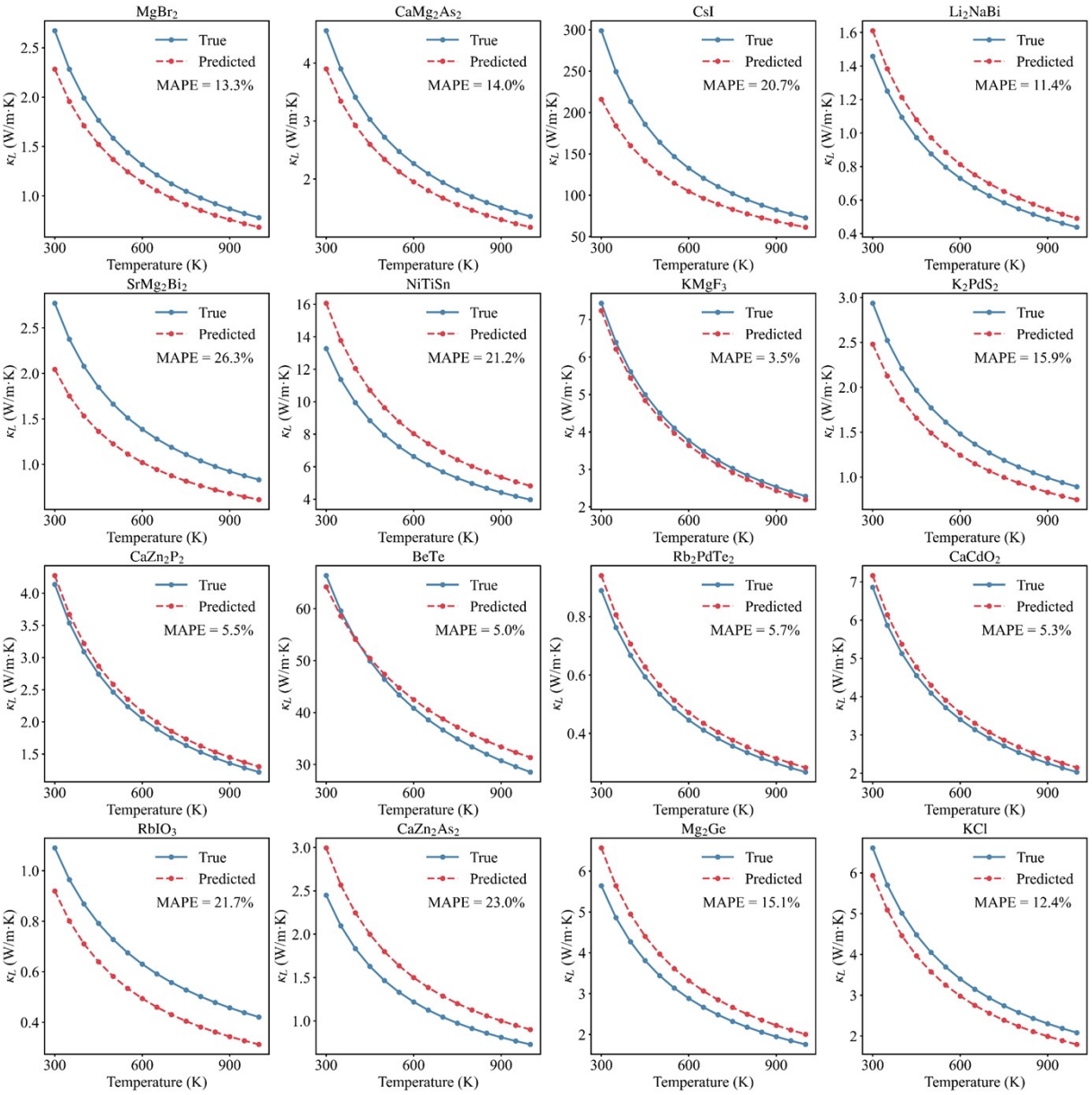


Fig. S6 Comparison plot of the κ_{3ph} curve predicted by machine learning and the κ_{3ph} curve calculated by DFT for a variety of material systems.

Table S6 Phonon-related descriptors.

1	$f_{a,\,max}$	The maximum frequency of the acoustic branch
2	$f_{o,\,min}$	The minimum frequency of the optical branch
3	f_{O1}	The frequency difference of the 456th optical branch
4	f_{O2}	The maximum phonon band gap between phonon branches
5	f_{AO1}	$f_{o,\,min} - f_{a,\,max}$
6	f_{AO2}	$f_{o,\,min} / f_{a,\,max}$
7	f_{mean}	The weighted average frequency of phonons

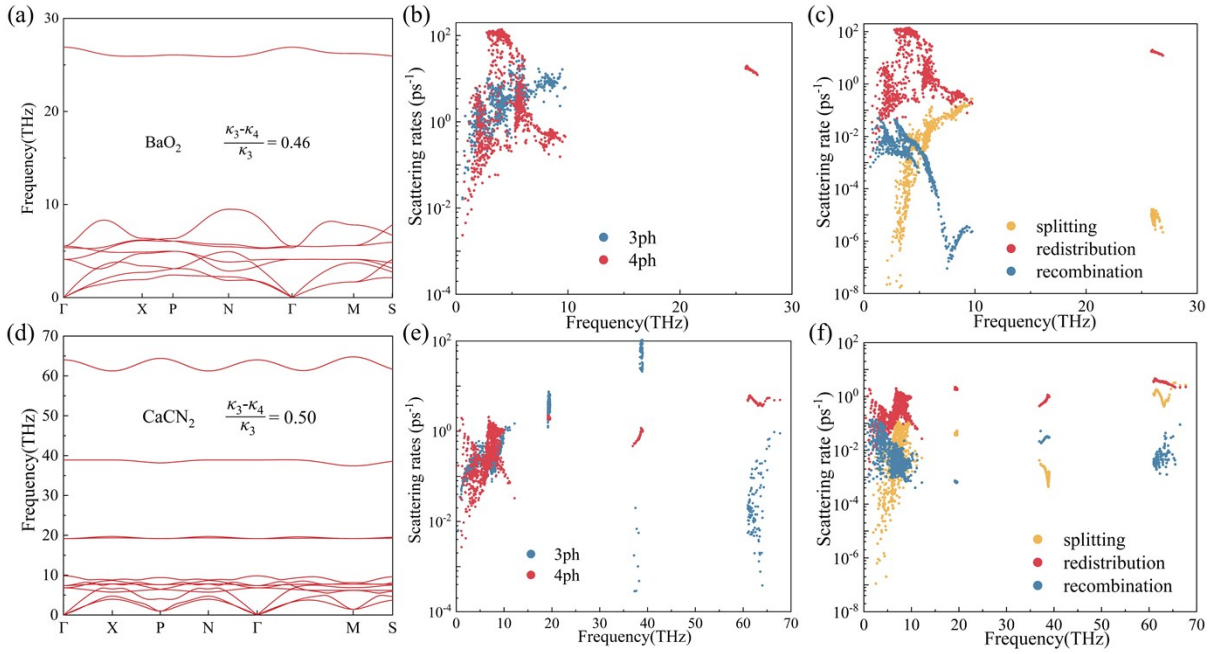


Fig. S7 Phonon dispersions for (a) BaO_2 and (d) $CaCN_2$. 3ph and 4ph scattering rates for (b) BaO_2 and (e) $CaCN_2$. 4ph scattering rates of different scattering process in (c) BaO_2 and (f) $CaCN_2$.

Table S7 Calculated lattice thermal conductivity (κ_{3ph} and κ_{4ph}), material ID in the Materials Project database (mp-id), maximum acoustic branch frequency ($f_{a,max}$), and relative four-phonon reduction (κ_{34}) for materials exhibiting significant four-phonon scattering effects with $\kappa_{34} > 0.4$.

formula	mp-id	$f_{a,max}$ (THz)	κ_{3ph} (W/mK)	κ_{4ph} (W/mK)	κ_{34}
BaLiAs	mp-10616	2.56	2.15	1.12	0.48
CsK_2Sb	mp-581024	1.25	0.21	0.11	0.48
GaN	mp-804	7.22	113.66	66.03	0.42
CuBr	mp-22913	3.54	0.84	0.43	0.49
$RbIO_3$	mp-27193	2.01	1.09	0.62	0.43
$NaBiTe_2$	mp-1221102	1.94	0.65	0.37	0.43
K_2PtC_2	mp-976876	2.65	7.43	3.97	0.47

$CaCN_2$	mp-4124	6.85	17.33	8.70	0.50
KCl	mp-23193	3.72	6.61	3.60	0.46
Na_2CN_2	mp-541989	5.00	4.27	2.06	0.52
BaO_2	mp-1105	4.18	1.35	0.73	0.46
$MgCN_2$	mp-9166	9.07	24.36	12.24	0.50
$NaSbTe_2$	mp-1220883	2.18	0.84	0.50	0.40
$RbBrO_3$	mp-28872	2.09	0.82	0.28	0.66
$LiRhO_2$	mp-14115	7.96	5.36	3.08	0.43
Rb_2PtC_2	mp-10919	1.94	8.23	4.42	0.46
LiBeSb	mp-1100398	4.55	108.63	45.34	0.58
CsI	mp-614603	1.55	1.23	0.72	0.41
Li_3Bi	mp-23222	2.47	1.14	0.65	0.43
CdSe	mp-1070	1.87	3.03	1.43	0.53
Li_2CN_2	mp-9610	8.44	25.36	8.96	0.65
AlSb	mp-2624	4.44	64.70	32.46	0.50
AlN	mp-661	12.85	164.20	96.16	0.41
RbBr	mp-22867	2.21	3.31	1.75	0.47
ZnO	mp-2133	4.88	34.72	19.64	0.43
$RbBiS_2$	mp-30041	2.19	0.80	0.46	0.42
K_2CN_2	mp-10408	3.32	1.65	0.70	0.58

Reference

1 Y. Wu, S. Dai, L. Ji, Y. Ding, S. Zeng, J. Yang and L. Zhou, *Phys. Rev. B*, 2024, **110**, 104308.

2 S. L. Shindé and J. S. Goela, Eds., *High Thermal Conductivity Materials*, Springer New York, New York, NY, 2006.

3 J. Yan, P. Gorai, B. Ortiz, S. Miller, S. A. Barnett, T. Mason, V. Stevanović and E. S. Toberer, *Energy Environ. Sci.*, 2015, **8**, 983–994.

4 S. Liu, Y. Huang, K. Tang, J. Zhou and Z. Sun, *Chem. Eng. J.*, 2025, **520**, 166247.

5 J. He, Y. Xia, S. S. Naghavi, V. Ozoliņš and C. Wolverton, *Nat. Commun.*, 2019, **10**, 719.