

Supplementary Information:

**Charting Lattice Thermal Conductivity for Inorganic Crystals
and Discovering Rare Earth Chalcogenides for Thermoelectrics**

Taishan Zhu,^{1,*} Ran He,^{2,*} Sheng Gong,^{1,*} Tian Xie,¹ Prashun
Gorai,³ Kornelius Nielsch,² and Jeffrey C. Grossman^{1,†}

¹*Department of Materials Science and Engineering,
Massachusetts Institute of Technology, Cambridge, MA 02139*

²*Leibniz Institute for Solid State and Materials Research, Dresden, 01069, Germany*

³*Department of Metallurgical and Materials Engineering,
Colorado School of Mines, Golden, CO 80401*

* These authors contributed equally.

† Corresponding author: jcg@mit.edu

I. METHODS

A. Transfer learning

In order to exploit knowledge learned from the larger calculated dataset and promote the learning of the small experimentally measured one, we develop a transfer learning scheme (see Fig. S1 (a)), which is based on the idea that correlated datasets share similar domain knowledge. In the main text, we extend our κ_C data to the ICSD set and learned information related to its structural chemistry. As mentioned, theoretical models are invaluable in that they contain the knowledge of κ , but due to simplicity they inevitably can only provide insufficient accuracy and limited universality. To take advantage of the knowledge learned from our larger high-throughput dataset, we develop a transfer learning framework demonstrated in Fig. S1 (a). This transfer learning scheme we used to predict experimental conductivity is a two-step modified CGCNN model: i) training a CGCNN model on our high-throughput κ_C dataset to extract knowledge, which has been done in the main text. ii) transferring the parameters of all layers from step i) to initialize a second CGCNN to transfer knowledge, and add one extra layer before the output layer to account for the difference between the two datasets. For the second step, we use the smaller κ dataset (132 entries, see Tab. S5) collected from experimental measurements in the literature. Since the experimental dataset is very small, in step ii), all the layers other than the last one are frozen to keep the pre-learned knowledge and reduce the degrees of freedom to suppress overfitting.

With this transfer learning scheme, we predict directly experimental values here using CGCNN, but with high MAEs (see Fig. S1 (b) and Tab. S1), due to small size of the experimental dataset, $< 10^3$ entries. The overall performance is compared with random forest and CGCNN in Fig. S1 (b), using different training datasets, and as can be seen our TL-CGCNN leads to the lowest MAE. Figure S1(c) plots the improvement for each data in the test set, defined by the absolute error difference between CGCNN and TL-CGCNN. It can be seen that the accuracy on the high- κ end ($\log \kappa > 1$) is improved, but the accuracy is deteriorated on the low- κ end, even though the overall performance is enhanced (see detailed analysis in the latent space in SI). Some example predictions in the high- κ limit from step ii), termed κ'_{exp} , can be found in Tab. 1. In the $\log \kappa < 1$ region, we recommend κ'_{exp} from random forest.

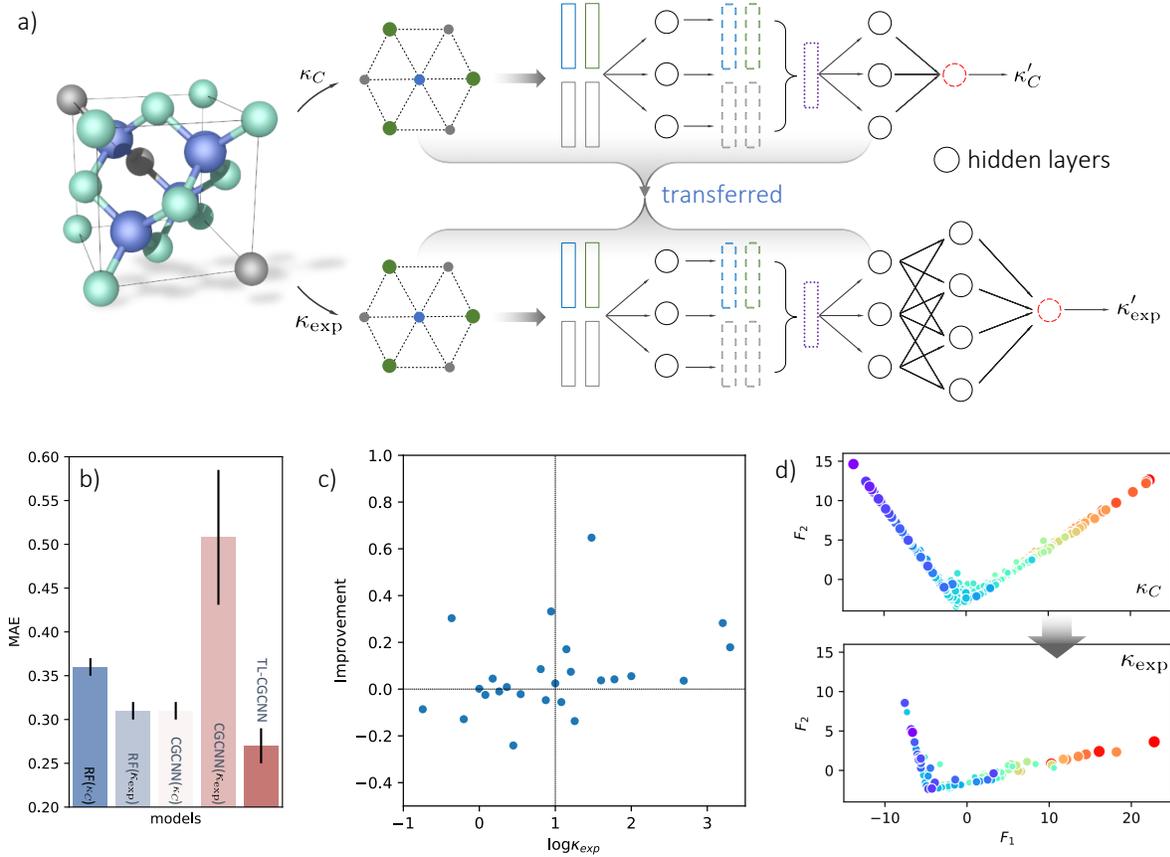


FIG. S1. Transfer learning based on CGCNN (TL-CGCNN). (a) This model learns high-throughput dataset κ_C and transfer the knowledge to learning κ_{exp} . (b) Comparison between different machine learning models, including random forest, CGCNN, and TL-CGCNN, trained on κ_C or κ_{exp} . TL-CGCNN exhibits the lowest MAE. (c) A closer look at the improvement of TL-CGCNN compared with CGCNN(κ_C) in prediction on the test set. The region of $\log \kappa > 1$ is systematically enhanced, while the $\log \kappa < 1$ region can be better or worse. (d) The distribution of the feature space \mathbf{V}_f projected onto two dimensions. The distribution and ranking of κ_C is generally smoother than κ_{exp} , and for κ_{exp} the upper end is smoother than the lower end.

B. Error of transfer learning

To understand the different performance in the high- and low- κ regions of the transfer learning model, we look into the space of crystal features in the neural networks. In Fig. S1(a), the network before the last hidden layer learns the feature vectors of materials \mathbf{V}_f , and the last operation from \mathbf{V}_f to output is simply a regression with softmax activation. Since

in TL-CGCNN we freeze \mathbf{V}_f and all layers before the extra layer due to the limited amount of data, we essentially use a one-layer neural network to fine tune κ_{exp} learnt from κ_C . We plot \mathbf{V}_f from the high-throughput and experimental datasets in Fig. S1(d). Interestingly, we observe a similar distribution between κ_{exp} and κ_C in the \mathbf{V}_f space, showing a strong correlation between the two datasets. However, in the high- κ region, κ_{exp} distributes more smoothly along the V-shape than in the low- κ region, which explains why TL-CGCNN performs better in the high- κ . Such issues in the low- κ region can be tackled from two aspects: i) more experimental data with low κ should be generated to better understand the κ_{exp} distribution, and ii) future high-throughput calculations should be refined to shrink the difference between κ_C and κ_{exp} , especially the outliers, in order to better sample the experimental \mathbf{V}_f space. The observation of data bias indicates the need to expand the current database. Instead of calculating hundreds of candidates in a certain material family each time, feature-space-based sampling techniques may be more computationally efficient to cover the material space.

C. Random forest, feature ranking, and dimension reduction

Random forest is an ensemble method that combines multiple decision trees. [3] This model has been used as both classifier and regressor for materials informatics. [4] In contrast to neural networks, random forest models are interpretable by providing an intrinsic metric to evaluate the importance of individual descriptors. We use this advantage of random forest in the main text to extract the important structural features that dominate κ . We use random forest implemented in scikit-learn [5]. The number of trees are set to 50 for all calculations, but the random states are randomly selected when studying the uncertainty in

TABLE S1. MAE of different models for predicting the experimental κ . Our CGCNN-based frameworks and random forest models predict κ directly from structure. In contrast, the Callaway and Slack models need extra calculations of bulk modulus and gruneisen parameter, which could be expensive. The MAE is 0.14 for Callaway [1] and 0.16 for Slack [2].

RF($\kappa_C \rightarrow \kappa'_{exp}$)	RF($\kappa_{exp} \rightarrow \kappa'_{exp}$)	CGCNN ($\kappa_C \rightarrow \kappa'_{exp}$)	CGCNN ($\kappa_{exp} \rightarrow \kappa'_{exp}$)	TL-CGCNN
0.36 \pm 0.01	0.31 \pm 0.01	0.31 \pm 0.01	0.51 \pm 0.08	0.27 \pm 0.02

predictions.

Dimension reduction has been performed through two approaches: i) principal component analysis (PCA) combined with t-distributed stochastic neighbor embedding (t-SNE). PCA is a linear reduction approach using singular value decomposition, and t-SNE converts similarities between data points to joint probabilities then minimizes the Kullback-Leibler divergence between the joint probabilities of the low-dimensional embedding and the high-dimensional data. We reduce the 154-dimensional feature space into 20 dimensions using PCA, then visualize the feature space in two dimensions by t-SNE analysis. In essence, this reduces our feature space into 2 dimensions, and enables direct visualization. PCA and t-SNE are both implemented in scikit-learn [5]. ii) Another approach is based on feature selection from random forest. Random forest ranks the importance of features, with which we could reduce the feature space till the performance (e.g., MAE) converges. This process could be more physics-based than the purely data-driven approach in i).

D. Hyperparameter optimization

In this work, we tune the following hyperparameters by grid-search: number of convolutional layers, length of atom feature vectors, length of hidden layer vectors, learning rate and type of optimizer, and for the last layer of the transfer learning scheme, the length of the layer and regularization term are taken into account. Descriptions of the hyperparameters for CGCNN are provided in Ref. [6]. For training the larger theoretical dataset, cross-validation is done by randomly selecting 20% of the data as the validation set, and for the small experimental dataset, a 5-fold cross-validation is used. In order to account for the random effect in training neural networks, for each parameter setting the training is repeated 20 times. We used both a Bayesian random search and deterministic grid search to optimize the hyperparameters, and the optimal hyperparameters used in this work are listed in Table S1.

E. First-principles validation

The κ_{DFT} values in Tab. 1 are calculated using a supercell perturbation method and the Botlzmman theory implemented in Phono3py. [7] Unit cell sizes are set to be greater

TABLE S2. The optimal hyperparameters used to train our neural networks.

	κ_C	κ_{exp}
number of convolutional layers	4	4
length of atom feature vectors	64	64
length of hidden layer vectors	128	16
learning rate	5×10^{-3}	1×10^{-2}
optimizer	Adam	Adam
length of the last layer		64
regularization term		0.1

than 10\AA , and the magnitude of atomic perturbation to be 0.005\AA . The force constants are extracted from density functional theory with plane-wave basis set and the projector augmented wave (PAW) method [8] through VASP. [9] Recommended PAW potentials are chosen for all the DFT calculations. We employ the generalized gradient approximation of Perdew, Burke, and Ernzerhof, [10] and uniform k-meshes with kpoint density greater than $700\text{ k-points}/\text{\AA}^{-3}$. The plane wave energy cutoff is set to be 1.3 times the maximal ENMAX of elements in the unit cell. The convergence criteria for energy and ionic forces are set to 10^{-6} eV and $0.01\text{ eV}/\text{\AA}$, respectively.

F. Experimental measurement

The compounds synthesized in this work were obtained by solid-state mechanical alloying. In a typical experiment, 10 grams of raw elements with high purity ($\geq 99.9\%$) were weighed according to stoichiometry in an argon-filled glovebox with O_2 and H_2O concentration less than 1 ppm. The weighted elements were subject to high energy ball milling (SPEX 8000D) for 20 hours under argon protection. The obtained powders were then compacted by a spark plasma sintering (SPS, FCT GmbH). The detailed sintering conditions are listed in Table S3. The phase purity of the as-sintered compounds was characterized by a Bruker D8 Advance diffractometer (Co radiation). Pure phases were realized for the compounds synthesized in this work.

The synthesized specimens were examined by X-ray diffraction (XRD). As shown in the supporting information (Fig. S2), the patterns of the compounds synthesized in this work

Compounds	SPS conditions
Cu ₂ HfTe ₃	573 K, 50 MPa, 3 min
Cu ₃ VTe ₄	573 K, 50 MPa, 3 min
Cu ₃ VTe ₄	573 K, 50 MPa, 3 min
TaCoTe ₂	873 K, 45 MPa, 3 min
AgAlTe ₂	543 K, 45 MPa, 3 min
FeIn ₂ S ₄	773 K, 45 MPa, 3 min
TiFeCoGa	1173 K, 50 MPa, 2 min
Er ₂ Se ₃	1573 K, 50 MPa, 3 min
Er ₂ Te ₃	1523 K, 50 MPa, 3 min
Tb ₂ Te ₃	1073 K, 50 MPa, 3min
Dy ₂ Te ₃	1173 K, 50 MPa, 3 min
Ho ₂ Te ₃	1073 K, 50 MPa, 3 min

TABLE S3. The SPS conditions for the compounds synthesized in this work.

(except for TiFeCoGa) matches well with the recorded patterns from the ICSD database, indicating the formation of compounds with high phase-purity. For TiFeCoGa, which is not recorded on ICSD, the measured pattern was compared to the simulation by assuming a quaternary Heulser crystal structure with space group Fm-3m (225) where the Ti, Fe, Co, and Ga atoms occupy the Wyckoff positions 4b ($\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$), 4c ($\frac{1}{4}, \frac{1}{4}, \frac{1}{4}$), 4d ($\frac{3}{4}, \frac{3}{4}, \frac{3}{4}$), and 4a (0, 0, 0), respectively. The simulation almost reproduced the experimental pattern of TiFeCoGa, yet some minor peaks cannot be observed experimentally. The slight mismatch between simulation and experiment might originate from atomic disorder, which contributed to the reduction of lattice thermal conductivity. Indeed, as shown in Table 1, the experimental lattice thermal conductivity of TiFeCoGa is lower than the predictions.

The thermal conductivities were calculated as a multiplication of thermal diffusivity, mass density, and specific heat. The thermal diffusivities were measured by a Laserflash (LFA1000 Linseis); the mass densities were measured by using the Archimedes principle; the specific heats were calculated from the Dulong-Petit law. The measurement uncertainties are 2% in mass density and 4% in diffusivity.

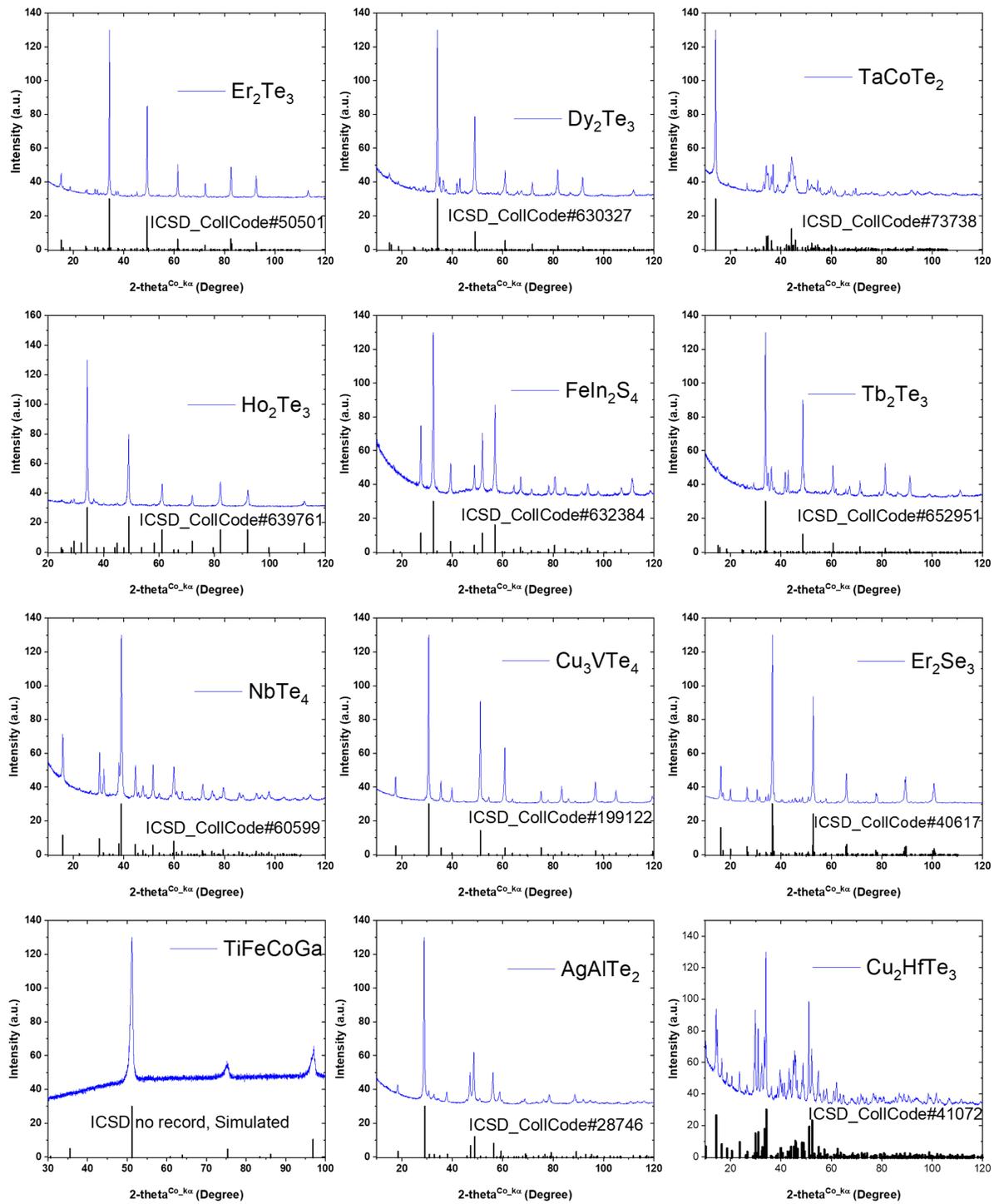


FIG. S2. X-ray powder diffraction (XRD) of all measured crystals in this work.

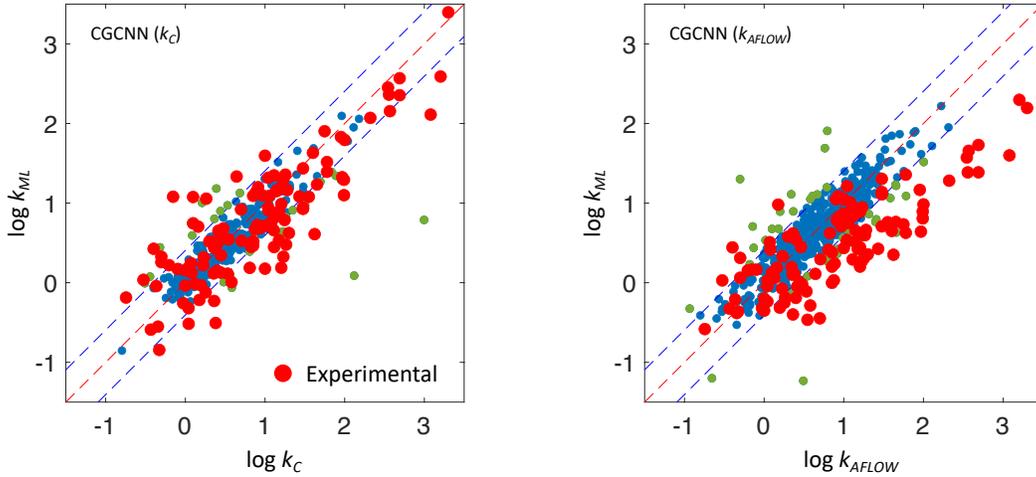


FIG. S3. Comparison of experimental measurements (κ_{exp} , red dots) and κ from high-throughput datasets (κ_C and κ_{AFLOW}) to machine learning predictions (κ_{ML} , the blue points are within a factor of 2, and the green points are outliers). high-throughput dataset aligns well with experiments.

II. RESULTS FROM DIFFERENT MACHINE LEARNING MODELS

We have trained CGCNN, TL-CGCNN and random forest models for learning κ from both high-throughput and experimental data sets. CGCNN has also been used to learn from AFLOW data. A detailed comparison is shown in Fig. S3. These diagonal plots compare the machine learning models (y axis) with given data sets (x axis). It is interesting to note the systematic underestimate from the model learnt from AFLOW data. In contrast, the model from our high-throughput data set align well with the diagonal. Table S1 lists the MAE from different machine learning models trained on the experimental data set, compared to the Callaway [1] and Slack [2] models. Our machine learning models consistently lead to lower error for the experimental data set.

III. EXPERIMENTAL DATA AND COMPARISON WITH MACHINE LEARNING PREDICTIONS

We collected 132 experimental κ_{exp} values from the literature, tabulated in Tab. S5.

TABLE S4. The 132 experimentally measured κ_{exp} .

Material	κ	Material	κ	Material	κ
C	2235	VFeSb	13	Cu ₂ ZnSiTe ₄	2.22
BN	1600	GaCuSe ₂	12.9	GaCuTe ₂	2.2
BA _s	1200	TiCoSb	12	PbSe	2
BN	760	Fe ₂ O ₃	11.3	LiBr	1.83
SiC	490	CdGeP ₂	11	NaI	1.8
BP	490	Li ₂ O	11	InTe	1.7
BeO	370	ZnGeAs ₂	11	CuI	1.68
SiC	360	CoSb ₃	10	Bi ₂ Te ₃	1.6
AlN	350	MnO	10	KSbS ₂	1.573
BP	350	SrO	10	SnTe	1.5
GaN	210	Mg ₂ Ge	9.3	AgGaS ₂	1.45
GaP	100	TiNiSn	9.3	Bi ₂ Se ₃	1.34
AlAs	98	NiSnZr	8.8	Cu ₃ SbSe ₃	1.29
SnO ₂	98	SrTiO ₃	8.5	Cu ₂ Se	1.25
InP	93	Mg ₂ Si	8.2	CuBr	1.25
AlP	90	SiO ₂	8	Ca ₅ Al ₂ Sb ₆	1.2
MgO	60	MgSe	7.69	YbFe ₄ Sb ₁₂	1.18
ZnO	60	CdTe	7.5	AgBr	1.1
AlSb	56	KCl	7.1	CsI	1.1
GaAs	45	Mg ₂ Sn	7.1	Bi ₂ Te ₃	1.1
CdGeAs ₂	42	NaCl	7.1	AgCl	1
GaSb	40	HfNiSn	6.7	CsCl	1
ZnGeP ₂	35.5	KF	6.43	TlCl	0.94
Al ₂ O ₃	30	GaCuS ₂	5.09	CsBr	0.94
CaO	30	Cu ₂ O	5	CuCl	0.84
InAs	30	CdSe	4.4	Bi ₂ O ₃	0.8
NiO	30	RbBr	3.8	NaSbTe ₂	0.75
ZnS	27	SrIn ₂ O ₄	3.64	CuCl	0.70

TABLE S5. The 132 experimentally measured κ_{exp} (continued).

Material	κ	Material	κ	Material	κ
InSb	20	Ba ₂ SnO ₄	3.64	NaBiTe ₂	0.64
ZnSe	19	ZnSb	3.5	AgBiSe ₂	0.63
NaF	18.4	KBr	3.4	SnSe	0.62
ZnTe	18	HgSe	3	TlBr	0.58
LiF	17.6	PbS	2.9	AgSbTe ₂	0.55
HfCoSb	17	NaBr	2.8	Cu ₃ SbSe ₃	0.50
Sc ₂ O ₃	17	RbCl	2.8	CuBiS ₂	0.50
CdS	16	KI	2.6	Tl ₉ BiTe ₆	0.47
Cr ₂ O ₃	16	HgTe	2.5	CsPbI ₃	0.45
Sb ₃ Ir	16	TePb	2.5	AgSbSe ₂	0.44
LiH	15	Sb ₂ Te ₃	2.4	CsPbBr ₃	0.43
ZrCoSb	15	PbTe	2.4	Sb ₂ O ₃	0.4
CoO	14	BaO	2.3	CsSnI ₃	0.37
Ga ₂ O ₃	14	RbI	2.3	Tl ₃ VSe ₄	0.30
ZnSiAs ₂	14	RbF	2.27	Gd ₁₁₇ Co ₅₆ Sn ₁₁₂	0.28
AlCuO ₂	13	Ba ₃ In ₂ O ₆	2.22	CsAg ₅ Te ₃	0.18

IV. GENERALIZED VAN-ARKELE TRIANGLES

Van-Arkel triangles have been used to characterize the bonding nature of binary compounds, in terms of the average and difference of element electronegativity (χ_a). This construction matches seamlessly with both the ranked features from random forest and the theoretical consideration from lattice dynamics. Therefore, we extend the average-difference in electronegativity to mean-variance in broadly all the ranked features. Except for the mean-variance of χ_a shown in Fig. 2(f), the other 9 plots, presented in Fig. S6, expand to other structural and elemental properties. Some properties, such as bond length (L_B), covalent radius (r_a) and mass (m) are features that clearly divide low- κ from high- κ materials. Several other features that are conventionally deemed reasonable proxy quantities (e.g. bond angle θ_B and coordination number CN) are however less obvious than the above

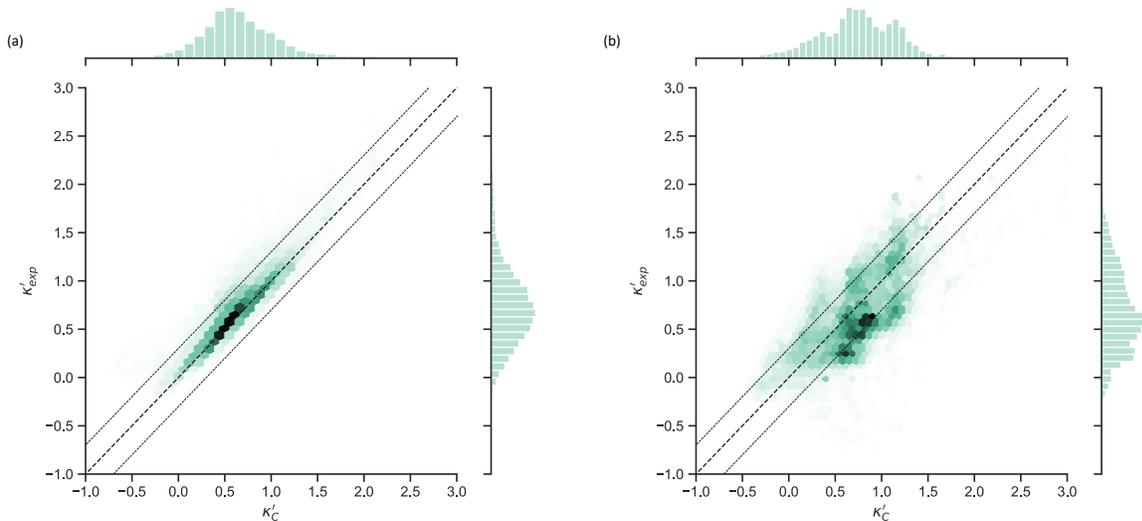


FIG. S4. Comparison between predicted experimental values and Callaway-model values from (a) TL-CGCNN and (b) random forest. TL-CGCNN gives more consistent predictions than random forest model. This should not be surprising, since TL-CGCNN tranfers the learnt knowledge from Callaway model, while random forest learnt Callaway and experimental datasets independently.

features.

V. ANOMALOUS BI DOPING

Note that Bi has one valance electron less than Te, however, increasing the content of Bi yields n-type properties instead of p-type, as shown by the Seebeck coefficients. To understand the role of Bi for optimizing the thermoelectric performance of REX compounds, we show the X-ray diffraction (XRD) patterns of the $\text{Er}_2\text{Te}_{3-x}\text{Bi}_x$ series in Fig. S7(a). The pristine compound, Er_2Te_3 , possesses a single phase that is crystallized in space group Fddd (No. 70). Upon the substitution of even 3.33% Bi into the Te sites ($x=0.1$), several extra diffraction peaks are recognized as elemental Bi. Therefore, the exact Bi contents that are solved into the Er_2Te_3 lattice is unknown but should not be higher than 3.33%. The other unsolved Bi elements exist as secondary phases, as indicated by the Energy Dispersive X-Ray (EDX) mapping of $\text{Er}_2\text{Te}_{2.4}\text{Bi}_6$ in Fig. S7(b). The unsolved Bi yields main compositions that are anion-deficient (i.e., Te-deficient), which explains the strengthened n-type transport behaviors when the Bi concentration is higher. We note that creating anion-deficiency

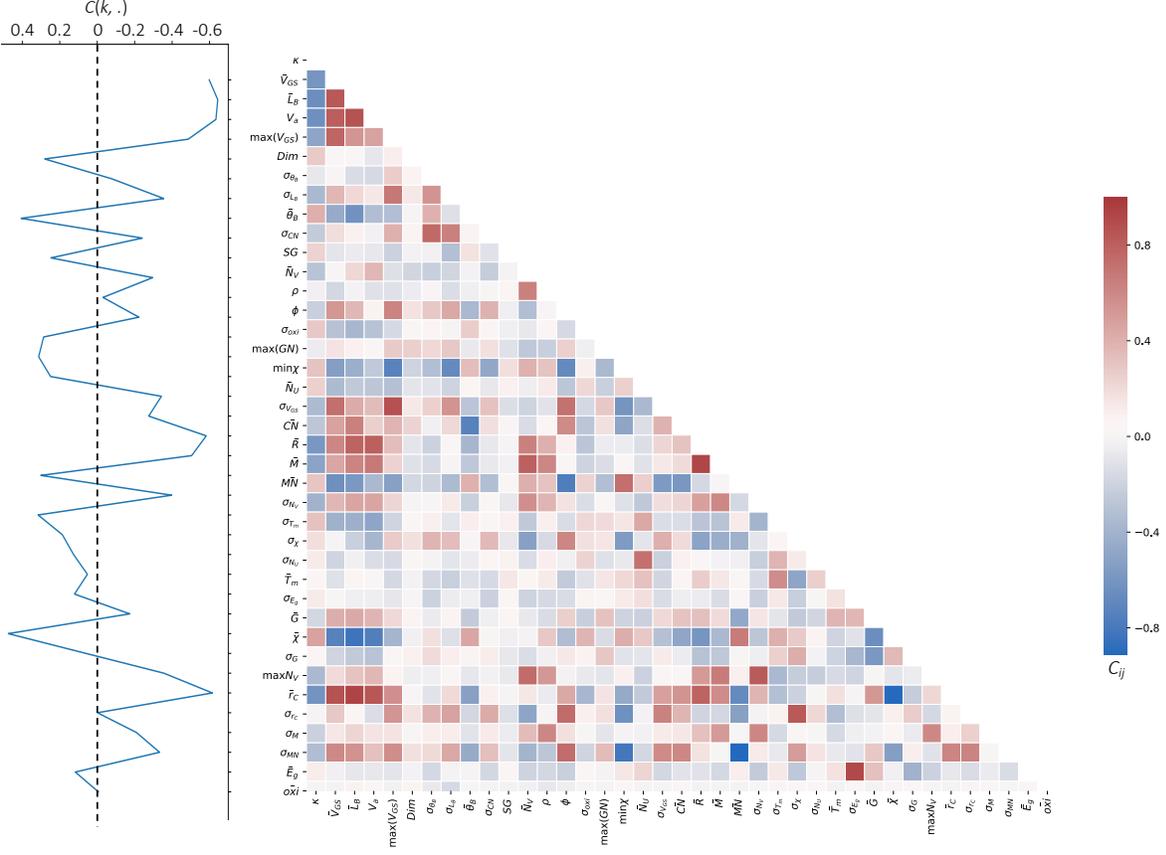


FIG. S5. Correlation between important features. The color bar is correlation value, and line plot on the left is the correlation between features and κ . This is compared to the learnt important features in the main text, showing the mined knowledge is significant.

compounds is a possible approach, as demonstrated in a similar composition $\text{Gd}_2\text{Se}_{2.98}$. [11] Overall, although it demands further investigations to clarify how these Bi precipitates affect exactly the thermoelectric transport properties, we experimentally demonstrate the promising potential of the REX family with vast members as high-temperature thermoelectric materials.

-
- [1] S. A. Miller, P. Gorai, B. R. Ortiz, A. Goyal, D. Gao, S. A. Barnett, T. O. Mason, G. J. Snyder, Q. Lv, V. Stevanovic, and E. S. Toberer, *Chemistry of Materials* **29**, 2494 (2017).
- [2] L. Chen, H. Tran, R. Batra, C. Kim, and R. Ramprasad, *Computational Materials Science* **170**, 109155 (2019).

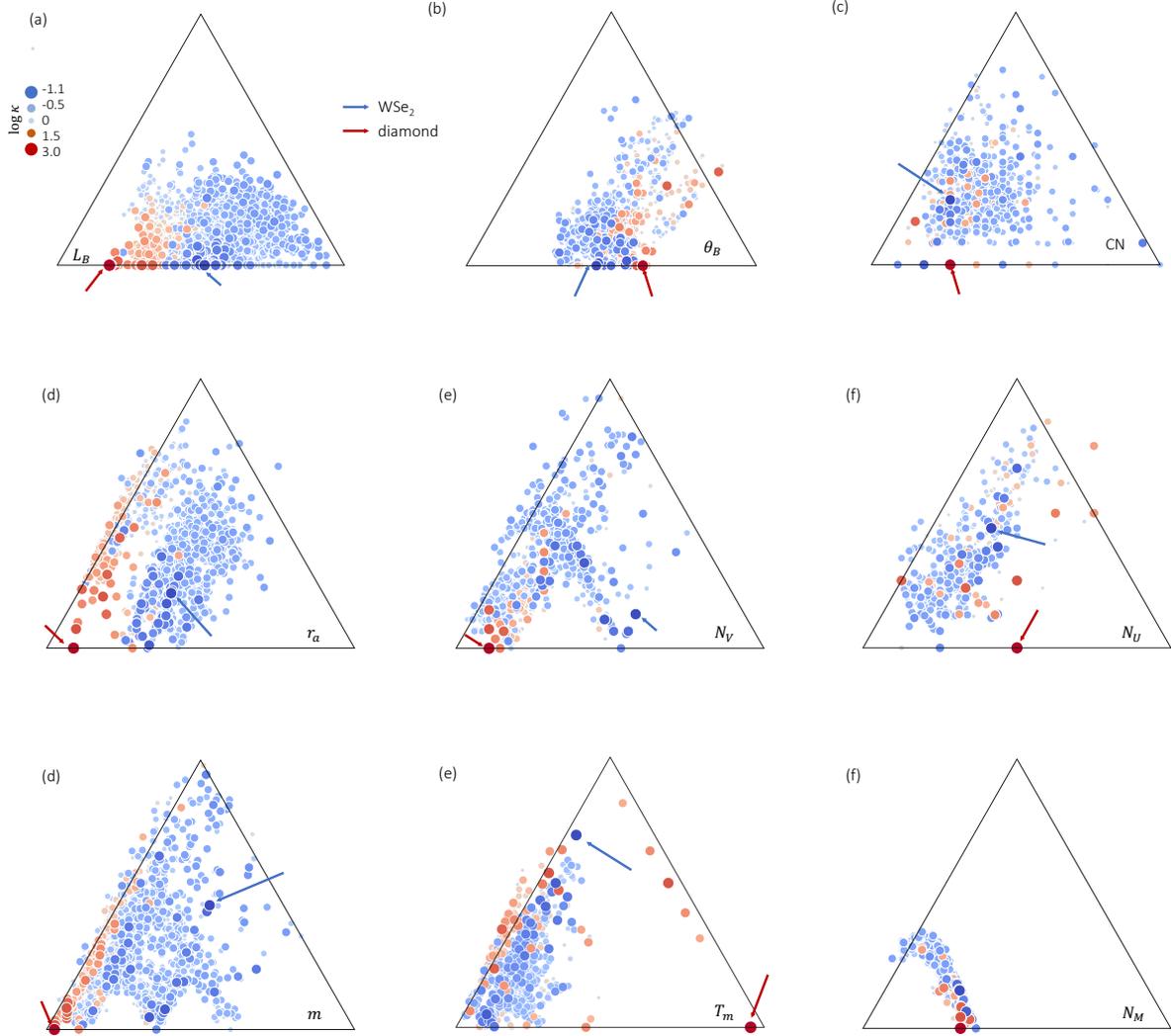


FIG. S6. van-Arkel-type triangles generalized to bond length L_B , bond angle θ_B , coordination number CN , covalent radius r_a , number of valence electrons N_V and number of unfilled valence N_U , atomic weight m , melting temperature of elemental solids T_m , and Mendeleev number N_M . Including V_{GS} and χ_a shown in the main text, L_B , θ_B , r_a , m are promising proxy for more expensive bonding strength and anharmonicity. Together, these quantities can be used to quickly estimate κ directly from structure.

- [3] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, *Journal of Chemical Information and Computer Sciences* **43**, 1947 (2003).
- [4] J. Carrete, W. Li, N. Mingo, S. Wang, and S. Curtarolo, *Physical Review X* **4**, 011019 (2014).
- [5] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher,

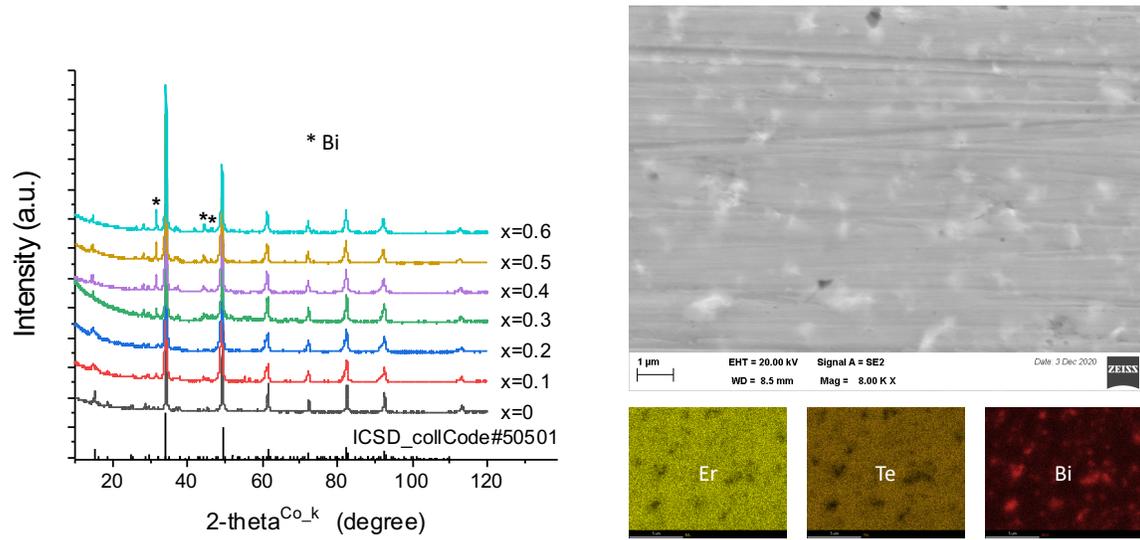


FIG. S7. (a) XRD patterns of compound series $\text{Er}_2\text{Te}_{3-x}\text{Bi}_x$ with $x=0, 0.1, 0.2, 0.3, 0.4, 0.5$ and 0.6 . (b) EDX mapping of $\text{Er}_2\text{Te}_{2.4}\text{Bi}_{0.6}$.

- M. Perrot, and E. Duchesnay, *Journal of Machine Learning Research* **12**, 2825 (2011).
- [6] T. Xie and J. C. Grossman, *Physical Review Letters* **120**, 145301 (2018).
- [7] A. Togo, L. Chaput, and I. Tanaka, *Physics Review B* **91**, 094306 (2015).
- [8] P. E. Blöchl, *Physical Review B* **50**, 17953 (1994).
- [9] G. Kresse and J. Furthmüller, *Physical Review B* **54**, 11169 (1996).
- [10] J. P. Perdew, K. Burke, and M. Ernzerhof, *Physical Review Letters* **77**, 3865 (1996).
- [11] F. Qin, S. A. Nikolaev, A. Suwardi, M. Wood, Y. Zhu, X. Tan, U. Aydemir, Y. Ren, Q. Yan, L. Hu, *et al.*, *Chemistry of Materials* (2020).